

Abderrahim Elmoataz Driss Mammass
Olivier Lezoray Fathallah Nouboud
Driss Aboutajdine (Eds.)

LNCS 7340

Image and Signal Processing

5th International Conference, ICISP 2012
Agadir, Morocco, June 2012
Proceedings



Springer

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Abderrahim Elmoataz Driss Mammass
Olivier Lezoray Fathallah Nouboud
Driss Aboutajdine (Eds.)

Image and Signal Processing

5th International Conference, ICISP 2012
Agadir, Morocco, June 28-30, 2012
Proceedings

Volume Editors

Abderrahim Elmoataz
Université de Caen Basse-Normandie
Caen, France
E-mail: abderrahim.elmoataz-billah@unicaen.fr

Driss Mammass
Université IbnZohr
Agadir, Morocco
E-mail: mammass@univibnzohr.ac.ma

Olivier Lezoray
Université de Caen Basse-Normandie
Caen, France
E-mail: olivier.lezoray@unicaen.fr

Fathallah Nouboud
Université du Québec à Trois-Rivières
Trois-Rivières, QC, Canada
E-mail: fathallah.nouboud@uqtr.ca

Driss Aboutajdine
Université Mohammed V – Agdal
Rabat, Morocco
E-mail: aboutaj@fsr.ac.ma

ISSN 0302-9743
ISBN 978-3-642-31253-3
DOI 10.1007/978-3-642-31254-0
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-31254-0

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.4.6-8, I.4, I.5.1-4, I.5, I.2.7, I.2.10, F.2, C.3

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

ICISP 2012, the International Conference on Image and Signal Processing, was the fifth ICISP conference, and was held in Agadir, Morocco. Historically, ICISP is a conference resulting from the actions of researchers from Canada, France and Morocco. Previous editions of ICISP were held in Trois-Rivières, Québec, (Canada 2010), in Cherbourg-Octeville (France, 2008) and in Agadir (Morocco, 2003 and 2001). ICISP 2012 was sponsored by EURASIP (European Association for Image and Signal Processing) and IAPR (International Association for Pattern Recognition).

The response to the call for papers for ICISP 2012 was encouraging. From 158 full papers submitted, 75 were finally accepted. The review process was carried out by the Program Committee members; all are experts in various image and signal processing areas. Each paper was reviewed by at least two reviewers, and also checked by the conference Co-chairs. The quality of the papers in these proceedings is attributed first to the authors, and second to the quality of the reviews provided by the experts. We would like to thank the authors for responding to our call, and we thank the reviewers for their excellent work. We were very pleased to be able to include in the conference program keynote talks by three world-renowned experts: Matti Pietikäinen, Director of Infotech Oulu Research Center, Finland; Denis Laurendeau, President of the International Association for Pattern Recognition (IAPR), Director of the REPARTI Research Center and Head of the Computer Vision and Systems Laboratory at Laval University, Quebec, Canada; and Saad Biaz, Professor, Computer Science and Software Engineering Department, Auburn University, USA.

We would also like to thank the members of the local committee for their advice and help. We are grateful to Springer's editorial staff for supporting this publication in the LNCS series. Finally, we were very pleased to welcome all the participants to this conference. For those who did not attend, we hope this publication provides a good view into the research presented at the conference, and we look forward to meeting you at the next ICISP conference.

April 2012

Abderrahim Elmoataz
Driss Mammass
Olivier Lezoray
Fathallah Nouboud
Driss Aboutajdine

ICISP 2012 Organization

General Chairs

Driss Mammass	Ibn Zohr University, Morocco
Abderrahim Elmoataz	Université de Caen Basse-Normandie, France

Program Committee Chairs

Fathallah Nouboud	Université du Québec à Trois-Rivières, Québec, Canada
Olivier Lezoray	Université de Caen Basse-Normandie, France
Driss Aboutajdine	Université Mohammed V- Agdal, Morocco

Local Organizing Committee

Hassan Douzi	Ibn Zohr University, Agadir, Morocco
Soufiane Idbraim	Ibn Zohr University, Agadir, Morocco
Abdelkarim Zatni	Ibn Zohr University, Agadir, Morocco
Youssef Es-saady	Ibn Zohr University, Agadir, Morocco
Mohamed El Hajji	Ibn Zohr University, Agadir, Morocco

Web Masters

Mohamed El Hajji	IRFSIC, Ibn Zohr University, Agadir, Morocco
Youssef Es-saady	IRFSIC, Ibn Zohr University, Agadir, Morocco

International Associations Sponsors

International Association for Pattern Recognition (IAPR)
European Association for Signal Processing (EURASIP)
Association of Research on Pattern Recognition and Imaging (ARPRI)
Pôle de Compétences Sciences Technologies de l'information et la
Communication (STIC)

Sponsoring Institutions

Université Ibn Zohr, Agadir, Morocco
Université du Québec à Trois-Rivières, Québec, Canada
Université de Caen Basse-Normandie, France
Faculté des sciences, Université Ibn Zohr, Agadir, Morocco
Ecole supérieur de technologie, Université Ibn Zohr, Agadir, Morocco

Program Committee

Ennaji Abdel	Université de Rouen, France
Shigeo Abe	Kobe University, Japan
Driss Aboutajdine	Université Mohamed V, Morocco
Jesus Angulo	MINES ParisTech, France
Antonis Argyros	University of Crete and FORTH-ICS, Greece
Sebastiano Battiato	University of Catania, Italy
George Bebis	University of Nevada, USA
Abdel Belaid	Université Vand.-Les-Nancy, France
Mostafa Bellafkih	INPT- Rabat, Morocco
Charles Beumier	Royal Military Academy, Belgium, Belgium
Guillaume-Alexandre Bilodeau	École Polytechnique de Montréal, Canada
Walter Blondel	INPL, France
Giuseppe Boccignone	Università degli Studi di Milano, Italy
Diego Borro	Ceit And Tecnun, Spain, Spain
Adrian Bors	University of York, UK
Alain Boucher	The Francophone Institute For Computer Science, Vietnam
Sébastien Bougleux	Université de Caen Basse-Normandie, France
Alexandra Branzan	University of Victoria, Canada
Xavier Bresson	City University of Hong Kong, China
Luc Brun	ENSICAEN, France
Gustavo Carneiro	The University of Adelaide, Australia
Emre Celebi	Louisiana State University in Shreveport, USA
Alain Chalifour	Université du Québec à Trois Rivières, Canada
Jocelyn Chanussot	GIPSA-Lab, Grenoble Institute of Technology, France
Christophe Charrier	Université de Caen Basse-Normandie, France
Xiaochun Cheng	Middlesex University, UK
Mohamed Cheriet	University of Quebec's École de technologie supérieure, Canada
Ronald Chung	The Chinese University of Hong Kong, Hong Kong
Laurent Cohen	Ceremade, France
Tomeu Coll	Universitat de les Illes Balears, Spain
Jose Crespo	Universidad Politécnica de Madrid, Spain
Kevin Curran	University of Ulster, UK
Jérôme Darbon	UCLA, USA
Marleen de Bruijne	Erasmus MC, The Netherlands
Farzin Deravi	University of Kent, United Kingdom, UK
François Deschenes	Université du Québec à Rimouski, Canada
Hassan Douzi	Ibn Zohr Universty, Morocco
Jean-Luc Dugelay	Company EURECOM Sophia Antipolis, France

Laurent Duval	IFP, France, France
Abdelaziz El Fazziki	UCAM (Marrakech), Morocco
Ayman El-Baz	University of Louisville, USA
Abderrahim Elmoataz	Université de Caen Basse-Normandie, France
Adrian Evans	University of Bath, UK
Christine Fernandez-Maloigne	Univ. Poitiers, France
Antonios Gasteratos	Democritus University of Thrace, Greece
Basilios Gatos	National Center for Scientific Research “Demokritos”, Greece
Theo Gevers	University of Amsterdam, The Netherlands
Abel Gomes	Beira Interior University, Portugal
Fabio Gonzalez	National University of Colombia, Bogota, Colombia
Michael Greenspan	Queen’s University, Canada
Metin Gurcan	OSU Columbus, USA
Edwin Hancock	York University, UK
Rachid Harba	Ecole Polytechnique de l’Univ. d’Orléans, France
Jon Yngve Hardeberg	Gjovik University College, France
Laurent Jacques	Université catholique de Louvain, Belgium
Stéphanie Jehan-Besson	Laboratoire GREYC CNRS UMR 6072, France
Xiaoyi Jiang	University of Münster, Germany, Germany
Pierre-Marc Jodoin	Université de Sherbrooke, Canada
Zoltan Kato	University of Szeged, Hungary
Mohamed Lamine Kherfi	Université de Québec à Trois Rivières, Canada
Dimitrios Kosmopoulos	NCSR Demokritos, Greece
Michal Kozubek	Masaryk University, Czech Republic, Czech Republic
Zakaria Lakhdari	Université de Caen, France
Denis Laurendeau	Université Laval, Canada
John Lee	UCL, Belgium
Sébastien Lefevre	Université de Bretagne-Sud, France
Anne-Claire Legrand	Université de Saint-Etienne - Laboratoire Hubert Curien, France
Olivier Lézoray	Université de Caen Basse-Normandie, France
Xuelong Li	University of London, UK
Chen Liming	Ecole Centrale de Lyon, France
Xiaoqiang Lu	Center for Optical Imagery Analysis and Learning, China
Yves Lucas	Orleans University, France
Rastislav Lukac	Sigma Corp. / Foveon, Inc., USA

Anant Madabhushi	Rutgers University, USA
Rémy Malgouyres	Univ – Clermont-Ferrand, France
Driss Mammass	Université Ibn Zohr, Morocco
Alamin Mansouri	Université de Bourgogne, France
Franck Marzani	Université de Bourgogne, France
Petr Matula	CMM, France
Brendan Mccane	University of Otago, New Zealand
Erik Meijering	Erasmus MC - University Medical Center Rotterdam, The Netherlands
Mahmoud Melkemi	Université Haute Alsace, France
Jean Meunier	Université de Montréal, Canada
François Meunier	Université du Québec à Trois-Rivières, Canada
Cyril Meurie	Univversité de Technologie de Belfort-Montbéliard / Laboratoire Set, France
Max Mignotte	Université de Montréal, Canada
Amar Mitiche	INRS-Enetgie, Matériaux et Télécommunications, Canada
El Yacoubi Mounim	Telecom Sud-Paris, France
Laurent Najman	Université Paris-Est, LIGM, Equipe A3SI, ESIEE Paris, France
Stéphane Nicolas	LITIS, Université de Rouen, France
Fathallah Nouboud	Université de Québec à Trois Rivières, Canada
Jean-Marc Ogier	Université de la rochelle, France
Yanwei Pang	Tianjin University, China
Ruven Pillay	C2RMF, France
Ioannis Pitas	University of Thessaloniki, Greece
Nasir Rajpoot	University of Warwick, UK
Eraldo Ribeiro	Florida Institute of Technology, USA
Audrey Roman	Université de Toulon, France
Eduardo Romero	National University of Colombia, Colombia
Christophe Rosenberger	GREYC - ENSICAEN, France
Gerald Schaefer	Loughborough University, UK
Sophie Schupp	Université de Caen, France
Lik-Kwan Shark	University of Central Lancashire, UK
Jialie Shen	Singapore Management University, Singapore
Robert Sitnik	Warsaw University of Technology, Poland
Bogdan Smolka	Silesian University of Technology, Poland
Jean-Luc Starck	CEA, France
Vinh-Thong Ta	LaBRI (Université de Bordeaux - CNRS - IPB), France
Salvatore Tabbone	LORIA-Nancy Université, France

Yi Tang	Chinese Academy of Sciences, China
Andrea Torsello	University of Venice, Italy
Sylvie Treuillet	University of Orleans, France
David Tschumperle	CNRS, France
Norimishi Tsumura	University of Chiba, France
Eiji Uchino	Yamaguchi University, Japan
Yvon Voisin	Université de Bourgogne, France
Liang Wang	University of Melbourne, Australia
Qi Wang	Xi'an Institute of Optics and Precision Mechanics of CAS, China
Michael Wilkinson	Johann Bernoulli Institute, University of Groningen, The Netherlands
Pingkun Yan	Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, China
Djemel Ziou	Université de Sherbrooke, Canada

Table of Contents

Multi/Hyperspectral Imaging

Bayesian Image Matting Using Infrared and Color Cues	1
<i>Layachi Bentabet and Hui Zhang</i>	
Salient Pixels and Dimensionality Reduction for Display of Multi/Hyperspectral Images	9
<i>Steven Le Moan, Ferdinand Deger, Alamin Mansouri, Yvon Voisin, and Jon Yngve Hardeberg</i>	
SVM and Haralick Features for Classification of High Resolution Satellite Images from Urban Areas	17
<i>Aissam Bekkari, Soufiane Idbraim, Azeddine Elhassouny, Driss Mammass, Mostafa El yassa, and Danielle Ducrot</i>	
Data Acquisition Enhancement in Shape and Multispectral Color Measurements of 3D Objects	27
<i>Grzegorz Mączkowski, Robert Sitnik, and Jakub Krzesłowski</i>	
Multi-model Approach for Multicomponent Texture Classification	36
<i>Ahmed Drissi El Maliani, Mohammed El Hassouni, Yannick Berthoumieu, and Driss Aboutajdine</i>	
Simultaneous Multispectral Imaging and Illuminant Estimation Using a Stereo Camera	45
<i>Raju Shrestha and Jon Yngve Hardeberg</i>	
Multisource Fusion/Classification Using ICM and DS _m T with New Decision Rule	56
<i>Azeddine Elhassouny, Soufiane Idbraim, Aissam Bekkari, Driss Mammass, and Danielle Ducrot</i>	

Image Filtering and Coding

Text Enhancement by PDE's Based Methods	65
<i>Zouhir Mahani, Jalal Zahid, Sahar Saoud, Mohammed El Rhabi, and Abdelilah Hakim</i>	
Kernel-Based Laplacian Smoothing Method for 3D Mesh Denoising	77
<i>Hicham Badri, Mohammed El Hassouni, and Driss Aboutajdine</i>	
Embedded Real-Time Video Processing System on FPGA	85
<i>Yahia Said, Taoufik Saidani, Fethi Smach, Mohamed Atri, and Hichem Snoussi</i>	

Edge Preserving Image Fusion Based on Contourlet Transform 93
Ashish Khare, Richa Srivastava, and Rajiv Singh

Selecting Vision Operators and Fixing Their Optimal Parameters
 Values Using Reinforcement Learning 103
Issam Qaffou, Mohamed Sadgal, and Aziz Elfazziki

A Phase Congruency Based Document Binarization 113
Hossein Ziaei Nafchi and Hamidreza Rashidy Kanan

Porting a H264/AVC Adaptive in Loop Deblocking Filter to a TI
 DM6437EVM DSP 122
Abdellah Skoudarli, Mokhtar Nibouche, and Amina Serir

Signal Processing 1

Methodology for Acoustic Characterization of a Labial Constraint in
 Speech Production 131
Leila Falek, Hocine Teffahi, and Amar Djeradi

Performance of OFDM in Radio Mobile Channel 142
Mohamed Tayebi and Mrahi Bouziani

Spatial Correlation Characterization for UWB Indoor Channel Based
 on Measurements 149
H. Chaibi, R. Saadane, My A. Faqihi, and M. Belkasm

Nonlinear Blind Source Separation Applied to a Simple Bijective
 Model 157
*Shahram Hosseini, Yannick Deville, Sonia El Amine, and
 Hicham Saylani*

Seismic Signal Discrimination between Earthquakes and Quarry Blasts
 Using Fuzzy Logic Approach 166
*El Hassan Ait Laasri, Es-Saïd Akhouayri, Dris Agliz, and
 Abderrahman Atmani*

Signal Processing 2

Ultra Wide-Band Channel Characterization Using Generalized Gamma
 Distributions 175
Zakaria Mohammadi, Rachid Saadane, and Driss Aboutajdine

Design of an Antenna Array for GNSS/GPS Network 183
Hocine Hamoudi, Boualem Haddad, and Philippe Lognonné

Blind Separation of Convolutional Mixtures of Non-stationary and
 Temporally Uncorrelated Sources Based on Joint Diagonalization 191
Hicham Saylani, Shahram Hosseini, and Yannick Deville

Maximizing Network Lifetime through Optimal Power Consumption in Wireless Sensor Networks	200
<i>El Abdellaoui Saïd, Fakhri Youssef, Debbah Merouane, and Aboutajdine Driss</i>	

Evolutionary Spectrum for Random Field and Missing Observations	209
<i>Rachid Sabre</i>	

Biometric

Iris-Biometric Fuzzy Commitment Schemes under Signal Degradation . . .	217
<i>C. Rathgeb and Andreas Uhl</i>	

Sfax-Miracl Hand Database for Contactless Hand Biometrics Applications	226
<i>Salma Ben Jemaa, Mayssa Frikha, Imen Moalla, Mohamed Hammami, and Hanene Ben-Abdallah</i>	

Spiral Cube for Biometric Template Protection	235
<i>Chouaib Moujahdi, Sanaa Ghouzali, Mounia Mikram, Mohammed Rziza, and George Bebis</i>	

Sparse Representation Based Classification for Face Recognition by k -LiMapS Algorithm	245
<i>Alessandro Adamo, Giuliano Grossi, and Raffaella Lanzarotti</i>	

3D Face Recognition Using an Expression Insensitive Dynamic Mask . . .	253
<i>Sadegh Salahshoor and Karim Faez</i>	

Score Fusion in Multibiometric Identification Based on Fuzzy Set Theory	261
<i>Khalid Fakhar, Mohammed El Aroussi, Mohamed Nabil Saidi, and Driss Aboutajdine</i>	

Security Analysis of Key Binding Biometric Cryptosystems	269
<i>Maryam Lafkih, Mounia Mikram, Sanaa Ghouzali, and Mohamed El Haziti</i>	

Watermarking and Texture

Improved Watermark Extraction Exploiting Undetermined Source Separation Methods	282
<i>Mohammed Khalil, Nawal EL Hamdouni, and Abdellah Adib</i>	

Texture Analysis for Trabecular Bone X-Ray Images Using Anisotropic Morlet Wavelet and Rényi Entropy	290
<i>Ahmed Salmi EL Boumnini El Hassani, Mohammed El Hassouni, Rachid Jennane, Mohammed Rziza, and Eric Lespessailles</i>	

Improving of Gesture Recognition Using Multi-hypotheses Object Association	298
<i>Sebastian Handrich, Ayoub Al-Hamadi, and Omer Rashid</i>	
An Improved Images Watermarking Scheme Using FABEMD Decomposition and DCT	307
<i>Noura Aherrahrou and Hamid Tairi</i>	
A Fragile Watermarking Scheme Based CRC Checksum and Public Key Cryptosystem for RGB Color Image Authentication	316
<i>Nour El-Houda Golea</i>	
Maximum Likelihood Estimation, Interpolation and Prediction for Fractional Brownian Motion	326
<i>Rachid Harba, Hassan Douzi, and Mohamed El Hajji</i>	
Gabor Filter-Based Texture Features to Archaeological Ceramic Materials Characterization	333
<i>Mohamed Abadi, Majdi Khoudeir, and Sylvie Marchand</i>	
RGB Color Distribution Analysis Using Volumetric Fractal Dimension	343
<i>Dalcimar Casanova and Odemir Martinez Bruno</i>	
 Segmentation and Retrieval	
Multiobjective Genetic Algorithm for Image Thresholding	352
<i>Layla Tahri and Mohamed Wakrim</i>	
Dual-Resolution Active Contours Segmentation of Vickers Indentation Images with Shape Prior Initialization	362
<i>Michael Gadermayr and Andreas Uhl</i>	
Matching Noisy Outline Contours Using a Descriptor Reduction Approach	370
<i>Saliha Aouat and Slimane Larabi</i>	
Brain MRI Image Segmentation in View of Tumor Detection: Application to Multiple Sclerosis	380
<i>Rabeb Mezgar, Mohamed Ali Mahjoub, Randa Salem, and Abdellatif Mtibaa</i>	
3D Shape Retrieval Using Bag-of-Feature Method Basing on Local Codebooks	391
<i>El Wardani Dadi, El Mostafa Daoudi, and Claude Tadonki</i>	

Segmentation of Prostate Using Interactive Finsler Active Contours and Shape Prior	397
<i>Foued Derraz, Abdelmalik Taleb-Ahmed, Azzeddine Chikh, Christina Boydev, Laurent Peyrodie, and Gerard Forzy</i>	
Tracking Moving Objects in Road Traffic Sequences	406
<i>Salma Kammoun Jarraya, Najla Bouarada Ghrab, Mohamed Hammami, and Hanene Ben-Abdallah</i>	
Eigen Combination of Colour and Texture Informations for Image Segmentation	415
<i>D. Attia, C. Meurie, and Y. Ruichek</i>	
A Graph Based Approach for Heterogeneous Document Segmentation	424
<i>Fattah Zirari, Driss Mammass, Abdellatif Ennaji, and Stephane Nicolas</i>	
Rotation Invariant Fuzzy Shape Contexts Based on Eigenshapes and Fourier Transforms for Efficient Radiological Image Retrieval	432
<i>Alaidine Ben Ayed, Mustapha Kardouchi, and Sid-Ahmed Selouani</i>	
Image Processing	
A 2D Rigid Point Registration for Satellite Imaging Using Genetic Algorithms	442
<i>Fatiha Meskine, Nasreddine Taleb, and Ahmad Imjabeer</i>	
Image Quality Assessment Measure Based on Natural Image Statistics in the Tetrolet Domain	451
<i>Abdelkaher Ait Abdelouahad, Mohammed El Hassouni, Hocine Cherifi, and Driss Aboutajdine</i>	
Real Time Door Access Event Detection and Notification in a Reactive Smart Surveillance System	459
<i>Gaetano Di Caterina, Nurulfajar Abd Manap, Masrullizam Mat Ibrahim, and John J. Soraghan</i>	
Optical Flow Estimation on Omnidirectional Images: An Adapted Phase Based Method	468
<i>Brahim Alibouch, Amina Radgui, Mohammed Rziza, and Driss Aboutajdine</i>	

DWT Based-Approach for Color Image Compression Using Genetic Algorithm	476
<i>Aldjia Boucetta and Kamal Eddine Melkemi</i>	

Pattern Recognition

Accelerator-Based Implementation of the Harris Algorithm	485
<i>Claude Tadonki, Lionel Lacassagne, Elwardani Dadi, and Mostafa El Daoudi</i>	
Writer Recognition on Arabic Handwritten Documents	493
<i>Chawki Djeddi, Labiba Souici-Meslati, and Abdellatif Ennaji</i>	
Outline Matching of the 2D Shapes Using Extracting XML Data	502
<i>Noreddine Gherabi and Mohamed Bahaj</i>	
Texture Classification Based on Lacunarity Descriptors	513
<i>João Batista Florindo and Odemir Martinez Bruno</i>	
Real-Time Fall Detection Method Based on Hidden Markov Modelling	521
<i>Alban Meffre, Christophe Collet, Nicolas Lachiche, and Pierre Gançarski</i>	
Extracting Buildings by Using the Generalized Multi Directional Discrete Radon Transform	531
<i>I. ELouedi, A. Hamouda, H. Rojbanı, R. Fournier, and A. Nait-Ali</i>	
Speaker Tracking Using Multi-modal Fusion Framework	539
<i>Anwar Saeed, Ayoub Al-Hamadi, and Michael Heuer</i>	
New Encoding Algorithm for Distributed Speech Recognition Based on DTFS Transform	547
<i>Azzedine Touazi and Mohamed Debyeche</i>	
Satellite Image Classification Using a Divergence-Based Fuzzy c-Means Algorithm	555
<i>Dong-Chul Park</i>	
Classifiers Combination for Arabic Words Recognition: Application to Handwritten Algerian City Names	562
<i>Soulef Nemouchi, Labiba Souici Meslati, and Nadir Farah</i>	
Robust Arabic Multi-stream Speech Recognition System in Noisy Environment	571
<i>Anissa Imen Amrous and Mohamed Debyeche</i>	

SVM Based GMM Supervector Speaker Recognition Using LP Residual Signal	579
<i>Dalila Yessad and Abderrahmane Amrouche</i>	
Plugin of Recommendation Based on a Hybrid Method for the Ranking of Documents in the E-Learning Platforms	587
<i>Hicham Moutachaouik, Hassan Douzi, Abdelaziz Marzak, Hicham Behja, and Brahim Ouhbi</i>	
Author Index	597

Bayesian Image Matting Using Infrared and Color Cues

Layachi Bentabet and Hui Zhang

Computer Science Department, Bishop's University,
J1M 1Z7, Sherbrooke, Québec, Canada
{layachi.bentabet, hui.zhang}@ubishops.ca

Abstract. In this paper, we propose a new matting solution that combines the use of color and infrared cameras for matting applications involving human actors. The infrared camera facilitates the extraction of the initial trimap and provides additional information for the matte estimation. The approach proposed in this paper differs from the techniques proposed in the literature in many aspects. It employs thermal information for human actors, which proves to be useful and effective for matting when combined with color information. It also introduces a new technique for automatic trimap construction that is based on the temperature's difference between the foreground actor and the background objects. Finally, the matting step is carried out using a Bayesian approach which combines the color and the infrared inputs into a single criterion. The matting results accuracy shows that our approach is capable of tackling digital image and video matting problems.

Keywords: Image matting, trimap construction, cues combination, infrared imagery, background subtraction.

1 Introduction

The matting problem of separating a non-rectangular foreground image from a background image is a classical problem in image processing and analysis [1][2]. A common example is a film frame where an actor is extracted from the background to later be placed on a different background. When the original image is of high resolution and/or contains motion blur and grain, as what is usually used in visual effects industry, the matting becomes an under determined problem, for which a unique solution cannot be found. Images of this type are currently matted by help of user input, a process that is time consuming. To increase the quality of the mattes shot against arbitrary backgrounds, and also to reduce the amount of human interaction required to generate them, several matting techniques make use of additional imagery information.

In blue screen approaches [3], the matting problem is simplified by using a constant color background which normally is blue. To have a unique alpha solution for the blue screen matting problem, there should be no pure blue color in the foreground image. Blue Screen matting is most popular design of matting in TV studios & movies. In [4], a flash matting approach is proposed to extract alpha matte

by using flash/non-flash image pairs. This technique is based on the observation that the most noticeable difference between the flash and no-flash image is the foreground object, if the background scene is sufficiently distant. The algorithm is strongly based on the assumption that the foreground objects become brighter with the flash whereas the background objects remain the same. This assumption does not happen all the time in real scenes. Other possibilities which may cause the failure of the matting could be the low reflectance of the foreground surfaces and pixels' saturation. This approach also assumes that the input image pair is pixel-aligned. Thus, it will fail when the fine foreground structures have moved in the time interval between the two images. Other techniques, such as [5], propose the use of a camera system that builds the alpha matte using the parallax motion between the frames.

In this paper, we propose a new matting system that combines the use of infrared and color information. The combination of color and infrared cues has been studied recently for target tracking applications [6]. Many algorithms focusing on the thermal domain have been explored. These methods are based on the assumption that the objects of interest appear at a contrast from their surroundings in the scene [7][8]. Our solution is based on the assumption that a human body emits more heat than a background containing non living objects. We will show that the use of infrared allows automatic extraction of the trimap and accurate estimation of the alpha matte

2 The Matting Problem

2.1 Matting Equations

An input image I is composed of a foreground component F and a background component B as indicated in equation (1), where for the i^{th} pixel of the input image I_i is:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i . \quad (1)$$

In the equation above, F_i is the foreground color, B_i is the background color and α_i is the pixel's foreground opacity. If $\alpha_i = 1$, the related pixel belongs to the foreground. If $\alpha_i = 0$, the related pixel belongs to the background. Otherwise we call it a mixed pixel. For a color image, C , equation (1) is generalized over RGB channels as follows:

$$C_R = \alpha_R F_R + (1 - \alpha_R) B_R \quad (2)$$

$$C_G = \alpha_G F_G + (1 - \alpha_G) B_G \quad (3)$$

$$C_B = \alpha_B F_B + (1 - \alpha_B) B_B . \quad (4)$$

In a color image, all quantities on the right-hand side of the above equations are unknown. We assume that alpha matte for red, green and blue channels are the same (*i.e.* $\alpha_R = \alpha_G = \alpha_B$). Therefore, for each pixel of a color image, there are three equations and seven unknowns. Thus matting is inherently an under-constrained problem.

2.2 Trimap Based Techniques

As mentioned before alpha matting is an ill-posed problem. Therefore, we need additional information about the image before to proceed with alpha estimation. Several approaches, such as Bayesian [9] and Robust matting [10], start by the user manually segmenting the input image into three regions, called trimap. A trimap is composed of three regions: a known foreground region, Ω_F , where $\alpha = 1$; a known background region, Ω_B , where $\alpha = 0$; and an unknown region, Ω_U , where $\alpha \in [0,1]$ (see figure 1). The foreground and background regions provide the additional information that is needed to estimate α in the unknown region.

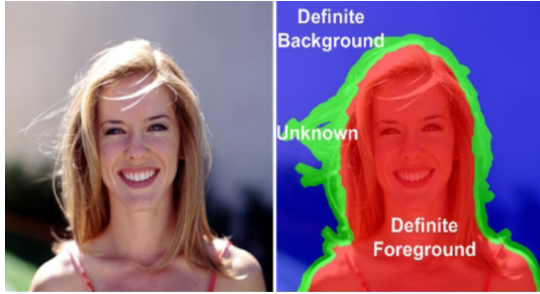


Fig. 1. Left image: original color image. Right image: user specified trimap.



Fig. 2. Left image: input image. Right image: input image with scribbles constraints.

Recent techniques proposed the use of user-interface scribbles [11] as shown in figure 2. Instead of marking the whole image into three regions, the user puts some scribbles according to the eigenvectors of a Laplacian measure.

3 Bayesian Image Matting Using Infrared and Color Cues

3.1 Overview

In our approach, we propose to use an infrared camera as an additional source of information. Our research assumes that the foreground component contains only living subjects such as humans or animals. Consequently, the use of infrared sensor

will help subtracting the background, and thus reduce the effect of similar colors in the background on the foreground matte. Instead of manually providing a trimap as most of the proposed methods in the literature, we use infrared images to automatically generate it. A general overview of our approach is presented in figure 3.

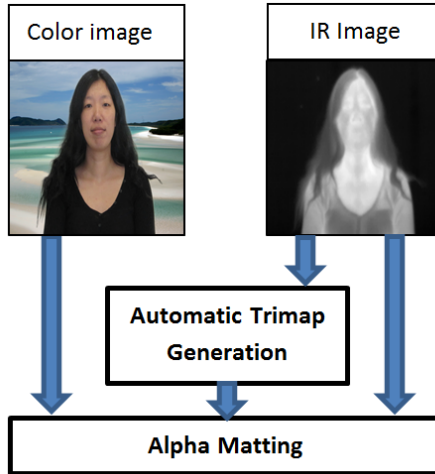


Fig. 3. Alpha matting using infrared and color sensors

The infrared and color images are assumed to be synchronized and spatially registered.

3.2 Automatic Trimap Generation

Our method starts with a foreground mask obtained by thresholding the entire infrared image. The threshold value is chosen according to the image histogram. As a result, the image domain is split into a foreground region and a region that contains both background and mixed pixels. In the next step, the unknown pixels are iteratively separated from the background.

The infrared level is characterized by the facts that: 1) it drops quickly at the border between foreground and background regions; 2) it decreases slowly for the pixels contained in the unknown region; and finally 3) it is low for the pixels in the background region. Given these considerations, our approach starts with a foreground mask that is iteratively propagated towards the background zone. At each iteration, new pixels are added to the unknown region if:

- 1) they are connected to the foreground or to the current unknown region, and,
- 2) their infrared levels are greater than a given threshold, and,
- 3) the difference between the infrared level of the new pixels and the infrared level of the pixels on the border of the unknown region is less than a given threshold.

Condition 1) ensures that the foreground and the unknown regions form together a single connected component. Condition 2) sets a threshold under which a pixel is considered as part of the background because of its low heat emission. Condition 3) reflects the smooth decrease of the infrared level in the unknown region. This corresponds to the human's temperature decrease as heat propagates in long thin parts such as the hair. The decrease rate is assumed to be lower than a given threshold. Therefore, if the transition at a given pixel is rapid, the pixel is considered as being part of the background. Our approach can be summarized as follows:

Let Ω_F^n and Ω_U^n be the foreground and unknown regions at iteration n . We define $\Omega_{F-U}^n = \Omega_F^n \cup \Omega_U^n$. The unknown region at iteration $n + 1$ is given by:

$$\Omega_U^{n+1} = \Omega_U^n \cup \partial\Omega_{F-U}^n. \quad (5)$$

where $\partial\Omega_{F-U}^n$ is the set of pixels that are added to the unknown region. If pixel $x \in \partial\Omega_{F-U}^n$ then:

- 1) $x \notin \Omega_{F-U}^n$ and
- 2) $IR(x) > T_1$ and
- 3) $\exists z \in (\Omega_{F-U}^n \cap N_8(x))$ such that $||IR(z) - IR(x)| < T_2$

$N_8(\cdot)$ is the set of 8-Neighbors, T_1 and T_2 are thresholds set by the user.

3.3 Joint Bayesian Matting

In our algorithm, we consider the IR information as an additional channel similar to the RGB channels of the color image C . We thus form a 4D image $I_{4D} = (R, G, B, IR)^T$. Image matting is then written and solved for this image using a Bayesian framework.

In Bayesian estimation, we find the most likely estimates for F_{4D} , B_{4D} and α , given the observation IR and C . We can express this as a maximization of the posterior probability (MAP) $P(\alpha, F_{4D}, B_{4D} | I_{4D})$, and then use Bayes's rule with the log likelihood as follows:

$$\underset{\alpha, F_{4D}, B_{4D}}{arg \max} P(\alpha, F_{4D}, B_{4D} | I_{4D}) \propto \underset{\alpha, F_{4D}, B_{4D}}{argmax} \{L(I_{4D} | \alpha, F_{4D}, B_{4D}) + L(F_{4D}) + L(B_{4D})\} \quad (6)$$

The log likelihood for alpha $L(\alpha)$ is assumed to be constant since we have no appropriate prior for α 's distribution. The first two log likelihoods on the right hand side of (6) are used to measure the fitness of solved variables (α, F_{4D}, B_{4D}) with respect to matting equations. We model these terms by measuring the difference between the observed I_{4D} and the image that would be predicted by the estimated F_{4D} , B_{4D} and α .

$$L(I_{4D} | \alpha, F_{4D}, B_{4D}) = -\| I_{4D} - \alpha F_{4D} - (1 - \alpha) B_{4D} \|^2 / \sigma_I^2$$

For $L(F_{4D})$ and $L(B_{4D})$, we follow the color sampling technique proposed in [9], where a group of nearby foreground and background pixels are collected to form an oriented Gaussian distribution. Thus:

$$\begin{aligned} L(F_{4D}) &= -(F_{4D} - \overline{F_{4D}})^T \Sigma_{F_{4D}}^{-1} (F_{4D} - \overline{F_{4D}}) \\ L(B_{4D}) &= -(B_{4D} - \overline{B_{4D}})^T \Sigma_{B_{4D}}^{-1} (B_{4D} - \overline{B_{4D}}) \end{aligned}$$

$\overline{F_{4D}}$ and $\overline{B_{4D}}$ are the mean values of the foreground and background components. $\Sigma_{F_{4D}}^{-1}$ and $\Sigma_{B_{4D}}^{-1}$ are the covariance matrices. The minimization steps of equation (6) are detailed in [9] and can be summarized by the following algorithm:

Step 1: Fix α to solve for F_{4D} and B_{4D} :

$$\begin{aligned} & \begin{bmatrix} \Sigma_{F_{4D}}^{-1} + I_{4D}\alpha^2/\sigma_{4D}^2 & I_{4D}\alpha(1-\alpha)/\sigma_{4D}^2 \\ I_{4D}\alpha(1-\alpha)/\sigma_{4D}^2 & \Sigma_{B_{4D}}^{-1} + I_{4D}(1-\alpha)^2/\sigma_{4D}^2 \end{bmatrix} \begin{bmatrix} F_{4D} \\ B_{4D} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{F_{4D}}^{-1}\overline{F_{4D}} + I_{4D}\alpha/\sigma_{4D}^2 \\ \Sigma_{B_{4D}}^{-1}\overline{B_{4D}} + I_{4D}(1-\alpha)/\sigma_{4D}^2 \end{bmatrix} \end{aligned}$$

Step 2: Fix F_{4D} and B_{4D} to solve for α :

$$\alpha = \frac{(I_{4D} - B_{4D}) \cdot (F_{4D} - B_{4D})}{\|F_{4D} - B_{4D}\|^2}$$

where I_{4D} is the 4×4 identity matrix. To maximize equation (6), we iteratively estimate α and (F_{4D}, B_{4D}) using steps 1 and 2 until changes between two successive iterations are negligible.

4 Experimental Results

The automatic trimap extraction method described in section 3.2 is applied on the infrared image of figure 4. The results are given in figure 5. The foreground is in white, the background in black and the unknown region in grey. These results show that our technique produces a single connected component for the three regions. The long hair details were successfully classified in the unknown region.

The joint Bayesian matting is applied on the unknown regions of the extracted trimaps and the results are presented in figure 5. The results demonstrate the ability of the developed technique to accurately estimate the alpha channel. As illustrated in figure 6, the level of detail for regions such as the hair is very accurate. It shows the transparency property of these areas, which is the main characteristic that allows a successful compositing of the actors on a new background.

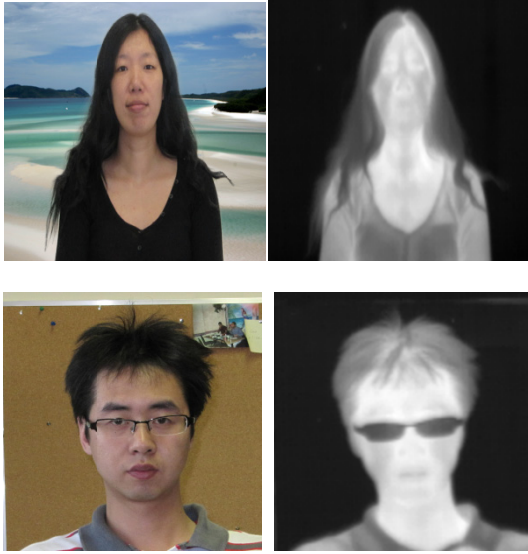


Fig. 4. Color and infrared image pairs



Fig. 5. Left: Automatic Trimap. Right: Alpha matte.

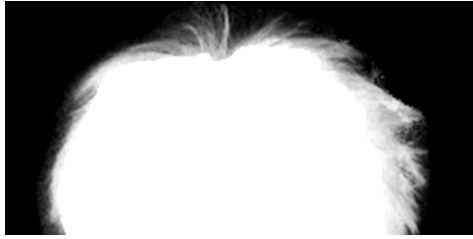


Fig. 6. Enlargement of the result in figure 4 that shows the details of the hair

5 Conclusion

In this paper, we have developed a Bayesian approach to solve the image matting problem. Though sharing a similar probabilistic view with [9], our approach differs in a number of key aspects. It uses MAP estimation to optimize color term and thermal term simultaneously. Also, the trimap is automatically generated from the thermal image. Our approach has an intuitive probabilistic motivation, is relatively easy to implement, and provide accurate matte results.

References

1. Fishkin, K., Barsky, B.: A family of new algorithms for soft filling. *ACM SIGGRAPH* 18(3), 235–244 (1984)
2. Porter, T., Duff, T.: Compositing digital images. *ACM SIGGRAPH* 18(3), 253–259 (1984)
3. Smith, A., Blinn, J.: Blue screen matting. *ACM SIGGRAPH* 30, 259–268 (1996)
4. Sun, J., Li, Y., Kang, S.-B., Shum, H.-Y.: Flash matting. *ACM SIGGRAPH* 25(3), 772–778 (2006)
5. Joshi, N., Matusik, W., Avidan, S.: Natural video matting using camera arrays. *ACM SIGGRAPH* 25(3), 779–786 (2006)
6. Airouche, M., Bentabet, L., Zelmat, M., Gao, G.: Pedestrian tracking using color, thermal and location cue measurements: a DSMT-based framework. *Machine Vision and Applications* (2011) (online first), doi:10.1007/s00138-011-0342-z
7. Yilmaz, A., Shafique, K., Shah, M.: Target tracking in airborne forward looking infrared imagery. *Image and Vision Computing* 21(7), 623–635 (2003)
8. Fernández-Caballero, A., Carlos-Castillo, J., Serrano-Cuerda, J., Maldonado-Bascón, S.: Real-time human segmentation in infrared videos. *Expert Systems with Applications* 38, 2577–2584 (2011)
9. Chuang, Y.-Y., Curless, B., Salesin, D.-H., Szeliski, R.: A bayesian approach to digital matting. *IEEE CVPR* 2, 264–271 (2001)
10. Wang, J., Cohen, M.: Optimized color sampling for robust matting. *IEEE CVPR*, 1–8 (2007)
11. Levin, A., Lischinski, D., Weiss, Y.: A closed form solution to natural image matting. *IEEE TPAMI* 30, 228–242 (2008)

Salient Pixels and Dimensionality Reduction for Display of Multi/Hyperspectral Images

Steven Le Moan^{1,2}, Ferdinand Deger^{1,2}, Alamin Mansouri¹,
Yvon Voisin¹, and Jon Y. Hardeberg²

¹ Laboratoire d'Electronique, Informatique et Image,
Université de Bourgogne, Auxerre, France

² The Norwegian Color Research Laboratory, Gjøvik University College, Norway

Abstract. Dimensionality Reduction (DR) of spectral images is a common approach to different purposes such as visualization, noise removal or compression. Most methods such as PCA or band selection use either the entire population of pixels or a uniformly sampled subset in order to compute a projection matrix. By doing so, spatial information is not accurately handled and all the objects contained in the scene are given the same emphasis. Nonetheless, it is possible to focus the DR on the separation of specific Objects of Interest (OoI), simply by neglecting all the others. In PCA for instance, instead of using the variance of the scene in each spectral channel, we show that it is more efficient to consider the variance of a small group of pixels representing several OoI, which must be separated by the projection. We propose an efficient method based on saliency to automatically identify OoI and extract only a few relevant pixels to enhance the separation foreground/background in the DR process.

1 Introduction

Dimensionality Reduction (DR) is a very common process in multi/hyperspectral imagery to project pixels to a space with a small number of attributes such as a three-dimensional color space (sRGB, HSV). To do so, many techniques were proposed, which are roughly divided into two categories: the ones which *transform* and the ones which *select* spectral channels. Even though Band Selection (BS) can be thought of as a generalization of transformation, they are based on two very different philosophies. Indeed, BS aims to preserve the physical meaning of spectral channels during the DR [1,2,3], whereas band transformation techniques such as Principal Components Analysis (PCA) [4,5], Independent Components Analysis [6] or true color [7,8], can mix channels to better fuse information along the spectrum. Evidently, the choice between these two approaches is application-driven.

The major drawback of most methods in the literature is that they are based on the assumption that all the pixels are part of the same population, i.e. performing a global mapping. Some approaches such as the Orthogonal Subspace Projection (OSP) [3] require a regular subsampling of the pixel population (down

to 1% without noticeable change, according to the authors) in order to alleviate their respective complexity. Scheunders [9] proposed to spatially divide the image into square blocks in order to achieve local mappings by means of PCA and Neural Network-based techniques. However, natural scenes are rich and complex, showing large contrasts among their constituents, therefore a more dedicated spatial partitioning would better take care of these properties.

In this paper, we propose to use a non-visual saliency detection [10] to extract relevant pixels, so that to respect the properties of the scene. Three sets of pixels are extracted: the salient ones, the surroundings and the background. Only a few pixels are then extracted, by means of PCA, so that to represent each of the first two sets aforementioned in the dimensionality reduction process.

The remainder of this paper is structured as follows: We first tackle the extraction of the representing set of pixels and present the results obtained, before conclusion.

2 Pixel Selection

As explained earlier, we aim to perform DR by means of a minimum-sized set of pixels to speed up the process, but not only. One of the tasks of DR is to convey, nay, enhance the relative discrepancies between the various Objects of Interest (OoI), contained in the input data. When it comes to images, it is generally equally weighted over the spatial dimensions, despite the rich and complex properties of natural scenes.

To obtain saliency maps from high dimensional images, we used the model that was previously introduced by Le Moan *et al.* [10]. It is inspired by the famous *Itti* model [11] and uses euclidean distance, spectral angle and Gabor filtering to compute low-resolution saliency maps. Figure 3 shows the results obtained on 4 images of the database introduced and used in the results section. It is important to note that these maps depict non-visual saliency, which can be seen as a measure of informative content, as they are computed regardless of the human visual system.

By thresholding these maps into three parts, we isolate different sets of pixels according to their respective contribution to the scene:

- The **salient pixels**, Ω_1 , are the pixels whose level of saliency is higher than a threshold T_{up} .
- The **surrounders**, Ω_2 , are the pixels whose level of saliency is lower than T_{up} and higher than T_{down} .
- The **background pixels**, Ω_3 , are all the rest.

Figure 1 shows an example of such segmentation on a natural scene, using different threshold values. The values of the optimal threshold are of course scene-dependent. We recommend to define them according to the separation of objects present in Ω_1 and Ω_2 . For example, the segmentation in Figure 1b would be a more relevant choice than the one in 1c, where the flower petals spread out

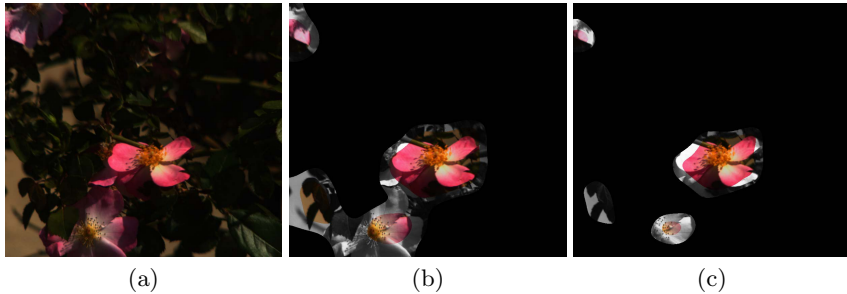


Fig. 1. Examples of saliency-based thresholding. Left: *true color* composite, Middle: $T_{up} = 0.3$ and $T_{down} = 0.1$, Right: $T_{up} = 0.5$ and $T_{down} = 0.3$. Saturated areas represent the salient pixels while surrounds are shown in grey and background in black.

on both Ω_1 and Ω_2 , which is undesirable. In this study, the optimal thresholds were defined manually for each scene.

In order to extract a set of representative pixels from each segment, we used PCA [12], over the spatial dimensions. During our experiments, we assessed that no more than five principal components are necessary to explain most of the data’s energy (more than 95%) and therefore to represent each Ω_1 and Ω_2 (as we disregard the background). Eventually, only 10 pixels are considered to compute the projection matrix.

Moreover, by mastering the number and type of objects present in the input data, one allows the latter algorithm to be more dedicated to conveying the discrepancies between, in our case, objects in Ω_1 and Ω_2 . Considering the relatively high computational complexity of PCA, we performed a random subsampling of 50 pixels in both groups. Moreover, resulting components are then normalized so that to fit the range [0..1]. Figure 2 shows an example of the principal components obtained.

3 Experiments and Results

3.1 Datasets

In this study, we used 4 natural scenes from the multispectral image database used in [13]. They contain 31 spectral channels each, covering the visible range of wavelengths (400-700nm). For more information about the acquisition system, calibration and processings, please refer to the database webpage¹.

3.2 Pre-processing and Normalization

In the raw reflectance data R_{raw} , all pixels above a threshold $\omega = \bar{R} + 3 * std(R)$ has been clipped to ω , to remove the influence of outliers and noisy pixels.

¹ http://personalpages.manchester.ac.uk/staff/david.foster/Hyperspectral_images_of_natural_scenes_02.html

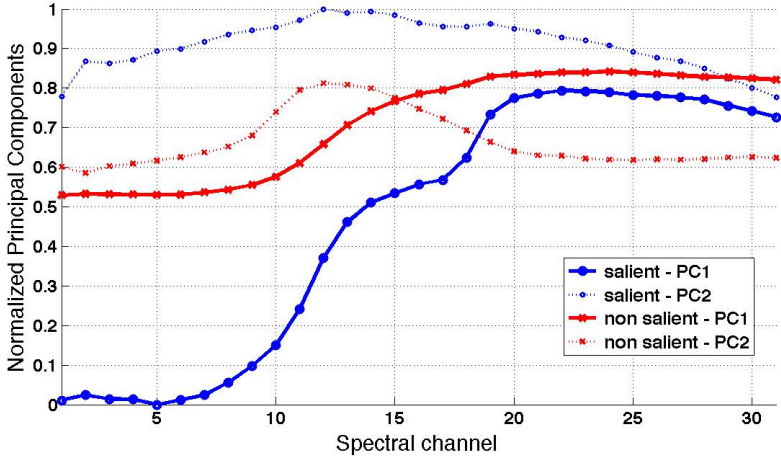


Fig. 2. Examples of (first and second) principal components obtained. Disks: representing Ω_1 and Crosses: representing Ω_2 . We can observe for instance that the first PCs (plain lines) are discriminable mostly in the first half of the image’s spectrum.

The result was divided by its maximal value so that it fits in the range [0..1]. Moreover, bands with average reflectance value below 2% and those with low correlation (below 0.8) with their neighboring bands have been removed, as suggested in [14].

3.3 Dimensionality Reduction Techniques

We selected three dimensionality reduction techniques to illustrate the proposed approach.

- Information-based Band Selection (**IBS**). We used the band selection approach that was used in [15] without the spectrum segmentation. It is based on a progressive research of dissimilar channels from single to third order.
- Orthogonal Subspace Projection-based Band Selection **OSPBS** [3] is a state-of-the-art band selection approach which consists of progressively selecting bands by maximizing their respective orthogonality.
- **PCA_{hsv}** is the traditional Principal Components Analysis of which components are mapped to the HSV color space, according to the normalization used in [5], without shifting the origin of the HSV cone.

Band selection approaches have been implemented in such a way that the band are eventually sorted by descending wavelength before mapping to sRGB.

3.4 Results

Figure 3 shows the *true color* composites of the images used in this study, as well as the corresponding saliency maps. Figures 4 to 6 show the results obtained by means of the different dimensionality reduction techniques, both by considering all the pixels in the image (or a uniform subsampling for OSPBS) and only a reduced set of pixels.

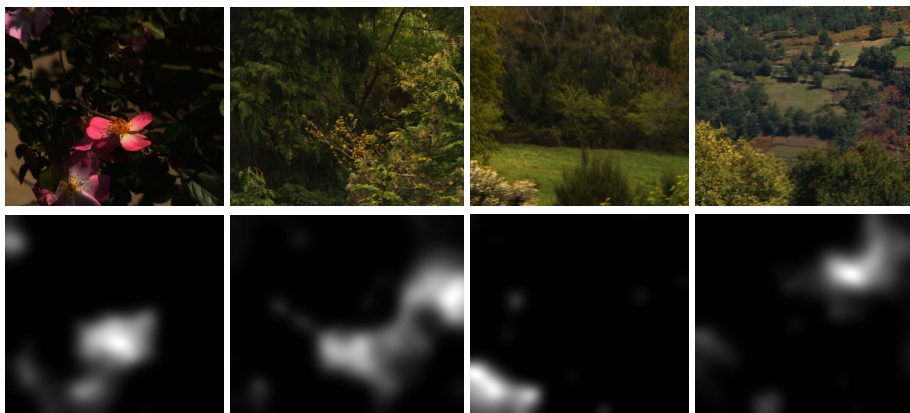


Fig. 3. *True color* composites (first row) and the corresponding saliency maps (computed from the high-dimensional datasets)

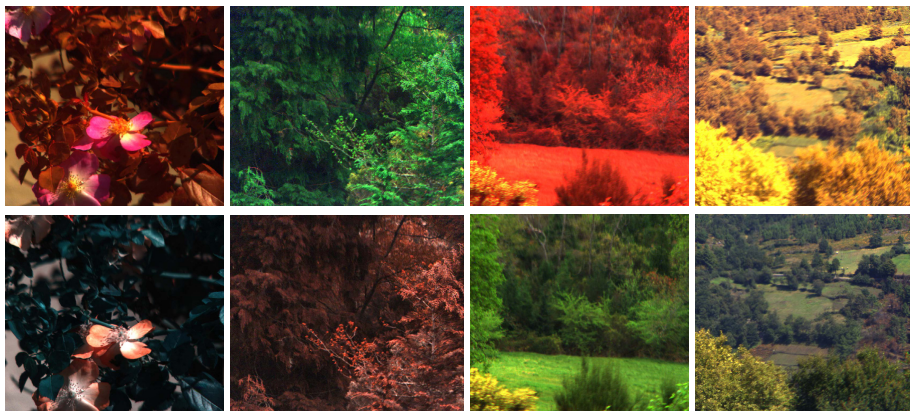


Fig. 4. IBS approach. First row: using all the pixels in the image. Second row: using a reduced set.

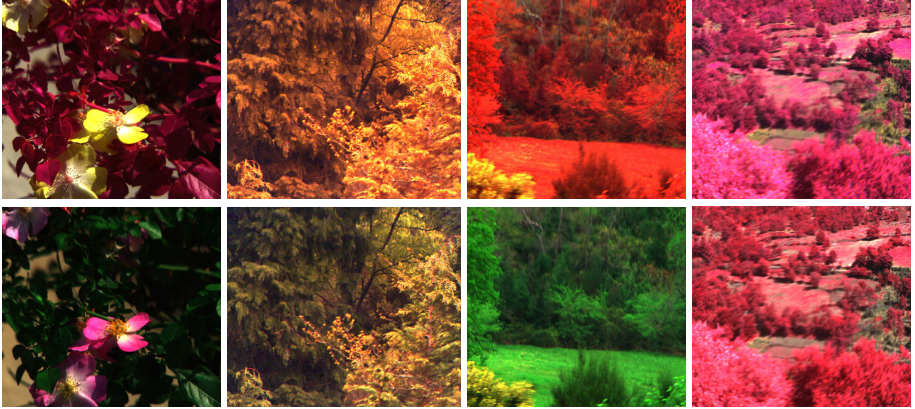


Fig. 5. OSPBS approach. First row: using a uniform subsampling of 1% of the image’s pixels. Second row: using a reduced set.

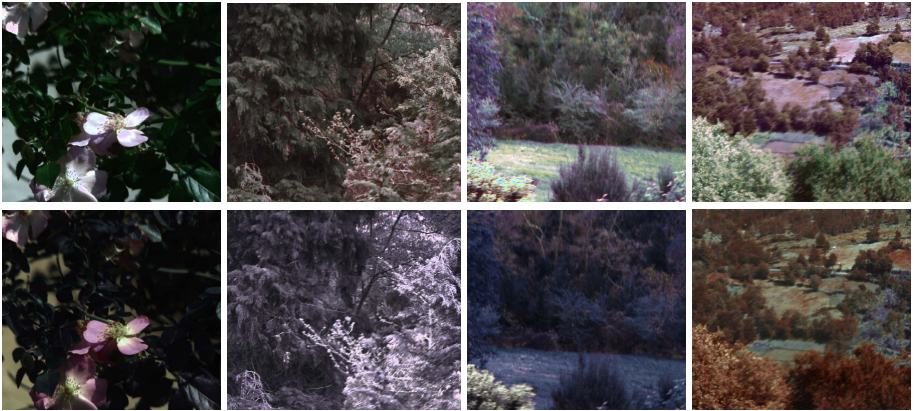


Fig. 6. PCA_{hsv} approach. First row: using all the pixels in the image. Second row: using a reduced set.

The optimal thresholds for each scene are given in table [11](#). **OSPBS** obtains their best results with a quite low upper threshold (0.3), while **IBS** and PCA_{hsv} perform better with a very reduced set of pixels. Overall, we observe that the most salient objects are emphasized, mostly because of a darkening or a diminution of contrast of their surroundings. Note that the natural rendering of these composites is considered outside the scope of this study, although it is very tempting to subjectively judge them according to this single feature.

In order to objectively evaluate the results, we used the color difference metric ΔE^* , which measures the Euclidean distance in the perceptually uniform color space CIELAB. Let ω_1 and ω_2 be two sets of 20 randomly selected pixels from

Ω_1 and Ω_2 , respectively. Now let $\bar{\Delta}_{12}$ be the average color difference between ω_1 and ω_2 , on a composite obtained with considering all pixels or a uniform subsampling and let $\bar{\Delta}'_{12}$ have the same definition but on a composite obtained by the proposed approach. We define the improvement of saliency $\delta_s = \bar{\Delta}'_{12} - \bar{\Delta}_{12}$. Table 2 shows the values obtained in this experiment.

Table 1. Optimal thresholds

	scene 1	scene 2	scene 3	scene 4
T_{up}	0.3	0.7	0.5	0.9
T_{down}	0.1	0.5	0.3	0.8

Table 2. Improvements of saliency δ_s . Difference of average Euclidean distance in CIELAB between Ω_1 and Ω_2 , using all the pixels versus using only a subset.

	scene 1	scene 2	scene 3	scene 4
IBS	18.8	4.0	44.9	23.7
OSPBS	13.9	26.1	32.5	2.1
PCA_{hsv}	20.5	42.8	47.3	9.0

Results show that there is an overall increase of conspicuity for the top salient objects. It is not surprising to see that the PCA is more sensitive to the pixel selection as it is more adaptive to the data and has more degrees of freedom than the BS techniques. However, it also shows less contrast in the background areas, due to the fact that these pixels are disregarded during the computation of the projection matrix. Scene 3 shows the best results, mainly because of the well-defined salient region on the bottom left side.

4 Conclusions

We introduced a new approach to perform dimensionality reduction in spectral images over a limited number of relevant pixels. By thresholding the saliency map of the high-dimensional image, we classify pixels according to their conspicuity in the scene, that we assume to be related to their overall relevance in a visualization task. Dimensionality reduction is then performed so that to focus on emphasizing these most important areas. Results show an increased conspicuity of the selected objects of interest, both objectively and subjectively. Yet, several challenges remain such as the efficient finding of optimal parameters for thresholding and the number of principal components to represent each set of pixels.

Acknowledgements. We would like to thank the the Regional Council of Burgundy for supporting this work.

References

1. Chang, C., Du, Q., Sun, T., Althouse, M.: A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Trans. on Geoscience and Remote Sensing* 37, 2631–2641 (1999)
2. Guo, B., Damper, R., Gunn, S., Nelson, J.: A fast separability-based feature-selection method for high-dimensional remotely sensed image classification. *Pattern Recognition* 41, 1670–1679 (2008)
3. Du, Q., Yang, H.: Similarity-based unsupervised band selection for hyperspectral image analysis. *Geoscience and Remote Sensing Letters* 5, 564–568 (2008)
4. Jia, X., Richards, J.: Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Trans. on Geoscience and Remote Sensing* 37, 538–542 (1999)
5. Tyo, J., Konsolakis, A., Diersen, D., Olsen, R.: Principal-components-based display strategy for spectral imagery. *IEEE Trans. on Geoscience and Remote Sensing* 41, 708–718 (2003)
6. Wang, J., Chang, C.: Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans. on Geoscience and Remote Sensing* 44, 1586–1600 (2006)
7. Poldera, G., van der Heijden, G.: Visualization of spectral images. In: *Proceedings of SPIE*, vol. 4553, p. 133 (2001)
8. Jacobson, N., Gupta, M.: Design goals and solutions for display of hyperspectral images. *IEEE Trans. on Geoscience and Remote Sensing* 43, 2684–2692 (2005)
9. Scheunders, P.: Multispectral image fusion using local mapping techniques. In: *International Conference on Pattern Recognition*, vol. 15, pp. 311–314 (2000)
10. Le Moan, S., Mansouri, A., Hardeberg, J., Voisin, Y.: Saliency in Spectral Images. In: Heyden, A., Kahl, F. (eds.) *SCIA 2011. LNCS*, vol. 6688, pp. 114–123. Springer, Heidelberg (2011)
11. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 20, 1254–1259 (1998)
12. Smith, L.: A tutorial on principal components analysis, vol. 51, p. 52. Cornell University, USA (2002)
13. Nascimento, S., Ferreira, F., Foster, D.: Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A* 19, 1484–1490 (2002)
14. Cai, S., Du, Q., Moorhead, R.: Hyperspectral imagery visualization using double layers. *IEEE Trans. on Geoscience and Remote Sensing* 45, 3028–3036 (2007)
15. Le Moan, S., Mansouri, A., Voisin, Y., Hardeberg, J.Y.: A constrained band selection method based on information measures for spectral image color visualization. *Transactions on Geoscience and Remote Sensing* 49, 5104–5115 (2011)

SVM and Haralick Features for Classification of High Resolution Satellite Images from Urban Areas

Aissam Bekkari¹, Soufiane Idbraim¹, Azeddine Elhassouny¹,
Driss Mammass¹, Mostafa El yassa¹, and Danielle Ducrot²

¹ IRF – SIC Laboratory, Faculty of Sciences B.P. 80 000 Agadir – Morocco
{a_bekkari, soufianeidbraim, info_azeddine}@yahoo.fr,
mammass@univ-ibnzohr.ac.ma, melyass@menara.ma

² Cesbio bpi 2801 31401 Toulouse cedex 9 France
danielle.ducrot@cesbio.cnes.fr

Abstract. The classification of remotely sensed images knows a large progress taking in consideration the availability of images with different resolutions as well as the abundance of classification's algorithms. A number of works have shown promising results by the fusion of spatial and spectral information using Support vector machines (SVM). For this purpose we propose a methodology allowing to combine these two informations using a combination of multi-spectral features and Haralick texture features as data source with composite kernel. The proposed approach was tested on common scenes of urban imagery. The results allow a significant improvement of the classification performances when compared with the two sets of attributes used separately. The experimental results indicate an accuracy value of 93.29% which is very promising.

Keywords: SVM, composite kernel, Haralick features, Satellite image, Spatial and spectral information, GLCM.

1 Introduction

With the commercial emergence of the optical satellite images of sub-metric resolution (Ikonos, Quickbird) the realization as well as the regular update of numerical maps with large scales becomes accessible and increasingly frequent [1].

Several classification algorithms have been developed since the first satellite image was acquired in 1972 [2-4]. Recently, some non-parametric classification techniques such as artificial neural networks, decision trees and Support vector machines (SVM) have been recently introduced.

SVM is a group of advanced machine learning algorithms that have seen increased use in land cover studies [5, 6]. One of the theoretical advantages of the SVM over other algorithms (decision trees and neural networks) is that it is designed to search for an optimal solution to a classification problem whereas decision trees and neural networks are designed to find a solution, which may or may not be optimal. This theoretical advantage has been demonstrated in a number studies where SVM generally produced more accurate results than decision trees and neural networks [7].

On other hand, the consideration of the spatial aspect in the spectral classification remains very important, for this case, Haralick described methods for measuring texture in gray-scale images, and statistics for quantifying those textures. It is the hypothesis of this research that Haralick's Texture Features and statistics as defined for gray-scale images can be modified to incorporate spectral information, and that these Spectral Texture Features will provide useful information about the image.

The proposed method consists in combining spatial and spectral information to obtain a better classification. We start with the extraction of spectral and spatial information. Then, we apply the SVM classification to the result file. Experimental results are provided and comparisons with a spectral classification and spatial classification are made to illustrate that the method is able to find better classes.

This paper is organized as follows. In the second section, we discuss the extraction of spatial and spectral information especially the Grey-Level Co-occurrence Matrix (GLCM) and Haralick texture features used in experimentations. In section 3, we give outlines on the used classifier: Support Vector Machines (SVM). In section 4, the results are presented with the used kernel defined as well as the stating of numerical evaluation. Finally, conclusions are given in section 5.

2 Extraction of Information and Classification

2.1 Spectral Information

The most used classification methods for the multispectral data consider especially the spectral dimension. The set of spectral values of each pixel is treated as a vector of attributes which will be directly employed as entry of the classifier. According to Fauvel [8] this allows a good classification based on the spectral signature of each area. However, this does not take in account the spatial information represented by the various structures in the image.

2.2 Spatial Information

Information in a remote sensed image can be deduced based on their textures. Many approaches were developed for texture analysis. Grey-Level Co-occurrence Matrix (GLCM) [9] is one of the most widely used methods, which is a powerful technique for measuring texture features; it contains the relative frequencies of the two neighbouring pixels separated by a distance on the image.

Haralick uses these matrices to develop a number of spatial indices that are easier to interpret. He assumed that the texture information is contained in the co-occurrence matrix, and texture features are calculated from it. A large number of textural features have been proposed starting with the original fourteen features (f_1 to f_{14}) described by Haralick et al [10], however only some of these features are in wide use. Wezcka et al [11] used four of Haralick features (f_1, f_2, f_5, f_8). Connors and Harlow [12] use five features (f_1, f_2, f_3, f_4, f_5). We found that these five features are commonly used seen that the fourteen are much correlated with each other, and that the five sufficed to give good results in classification [13].

In this work, we have used these five features: homogeneity (E), contrast (C), correlation (Cor), entropy (H) and local homogeneity (LH), and co-occurrence matrices are calculated for four directions: 0° , 45° , 90° and 135° degrees.

Let us recall their definitions:

$$E = \sum_i \sum_j (M(i, j))^2 \quad (1)$$

$$C = \sum_{k=0}^{m-1} k^2 \sum_{|i-j|=k} M(i, j) \quad (2)$$

$$Cor = \frac{1}{\sigma_i \sigma_j} \sum_i \sum_j (i - \mu_i)(j - \mu_j) M(i, j) \quad (3)$$

Where μ_i and σ_i are the horizontal mean and the variance, and μ_j and σ_j are the vertical statistics.

$$H = \sum_i \sum_j M(i, j) \log(M(i, j)) \quad (4)$$

$$LH = \sum_i \sum_j \frac{M(i, j)}{1 + (i - j)^2} \quad (5)$$

Each texture measure can create a new band that can be incorporated with spectral features for classification purposes.

2.3 SVM Classification

SVM is a group of advanced machine learning algorithms that have seen increased use in land cover studies; it generally produced more accurate results than other algorithms (decision trees and neural networks).

In this section we briefly describe the general mathematical formulation of SVMs introduced by Vapnik [14]. Starting from the linearly separable case, optimal hyperplanes are introduced. Then, the classification problem is modified to handle non-linearly separable data and a brief description of multiclass strategies is given.

2.3.1 Linear SVM

For a two-class problem in a n -dimensional space \mathbb{R}^n , we assume that l training samples $x_i \in \mathbb{R}^n$, are available with their corresponding labels $y_i = \pm 1$, $S = \{(x_i, y_i) \mid i \in [1, l]\}$. The SVM method consists of finding the hyperplane that maximizes the margin, i.e., the distance to the closest training data points for both classes [15]. Noting $w \in \mathbb{R}^n$ as the normal vector of the hyperplane and $b \in \mathbb{R}$ as the bias, the hyperplane H_p is defined as:

$$\langle w, x \rangle + b = 0, \forall x \in H_p \quad (6)$$

Where $\langle w, x \rangle$ is the inner product between w and x . If $x \notin H_p$ then $f(x) = \langle w, x \rangle + b$ is the distance of x to H_p . The sign of f corresponds to decision function $y = \text{sgn}(f(x))$.

Finally, the optimal hyperplane has to maximize the margin: $2/\|w\|$. This is equivalent to minimize $\|w\|/2$ and leads to the following quadratic optimization problem:

$$\min \left[\frac{\|w\|^2}{2} \right] \quad \text{subject to } y_i (\langle w, x_i \rangle + b) \geq 1 \quad \forall i \in [1, l] \quad (7)$$

For non-linearly separable data, the optimal parameters (w, b) are found by solving:

$$\min \left[\frac{\|w\|^2}{2} + C \sum_{i=1}^l \xi_i \right] \quad (8)$$

subject to $y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad \forall i \in [1, l]$

Where the constant C control the amount of penalty and ξ_i are *slack* variables which are introduced to deal with misclassified samples. This optimization task can be solved through its Lagrangian dual problem:

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

subject to $0 \leq \alpha_i \leq C \quad \forall i \in [1, l]$ (9)

$$\sum_{i=1}^l \alpha_i y_i = 0$$

Finally:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (10)$$

The solution vector is a linear combination of some samples of the training set, whose α_i is non-zero, called Support Vectors. The hyperplane decision function can thus be written as:

$$y_u = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i \langle x_u, x_i \rangle + b \right) \quad (11)$$

Where x_u is an unseen sample.

2.3.2 Non-linear SVM

Using the Kernel Method, we can generalize SVMs to non-linear decision functions. With this way, the classification capability is improved. The idea is as follows. Via a non-linear mapping Φ , data are mapped onto a higher dimensional space F :

$$\begin{aligned}\Phi: R^n &\rightarrow F \\ x &\mapsto \Phi(x)\end{aligned}\tag{12}$$

The SVM algorithm can now be simply considered with the following training samples: $\Phi(S) = \{(\Phi(x_i), y_i) \mid i \in [1, l]\}$. It leads to a new version of the hyperplane decision function where the scalar product is now: $\langle \Phi(x_i), \Phi(x_j) \rangle$. Hopefully, for some kernels function k , the extra computational cost is reduced to:

$$\langle \Phi(x_i), \Phi(x_j) \rangle = k(x_i, x_j)\tag{13}$$

The kernel function k should fulfill Mercer's conditions.

With the use of kernels, it is possible to work implicitly in F while all the computations are done in the input space. The classical kernels used in remote sensing are the polynomial kernel and the Gaussian radial basis function:

$$k_{poly}(x_i, x_j) = [(x_i \cdot x_j) + 1]^p\tag{14}$$

$$k_{gauss}(x_i, x_j) = \exp\left[-\gamma \|x_i - x_j\|^2\right]\tag{15}$$

In experiments we used Gaussian RBF kernel (15) which is commonly used in classification of remotely sensed images.

2.3.3 Multiclass SVMs

SVMs are designed to solve binary problems where the class labels can only take two values: ± 1 . For a classification of remotely sensed images, several classes are usually of interest. Various approaches have been proposed to address this problem [16]. They usually combine a set of binary classifiers.

Two main approaches were originally proposed for a k -classes problem.

- *One versus the Rest*: k binary classifiers are applied on each class against the others. Each sample is assigned to the class with the maximum output.
- *Pairwise Classification*: $k(k-1)/2$ binary classifiers are applied on each pair of classes. Each sample is assigned to the class getting the highest number of votes. A vote for a given class is defined as a classifier assigning the pattern to that class.

The *pairwise classification* has shown to be more suitable for large problems [15, 16]. Even though the number of the used classifiers is larger than for *the one versus the rest* approach, the whole classification problem is decomposed into much simpler ones. Therefore, this second approach was used in our experiments.

2.4 The Proposed Workflow

The proposed workflow has two main tasks, we start with the extraction of spectral information and spatial information and then the result will be used as an input to SVM classifier.

To use jointly spatial and spectral information, we chose to go through the definition of a kernel. In [17], several kernels are proposed to include spatial information. The weighted sums of kernels provide the best results for classification.

They also allow to control the influence of each type of information:

$$k_{\mu}(x, y) = \mu k_{spectral}(x, y) + (1 - \mu)k_{spatial}(x, y) \quad \text{with } 0 \leq \mu \leq 1 \quad (16)$$

The parameter μ will be chosen at the learning phase, it varied in steps of 0.1. For simplicity and for illustrative purposes, μ was the same for all the classes in our experiments. The penalization factor in the SVM was tuned in the range $C = \{10^{-1} \dots 10^7\}$. We use a RBF kernel (15) (with $\sigma = \{10^{-1} \dots 10^3\}$) for the two kernels. $k_{spectral}$ uses a spectral information while $k_{spatial}$ uses Haralick features.

3 Experimentations and Results

3.1 The Data

The first image used in classification is a sample of high resolution Quickbird satellite image. Its size is 240x360 pixels. It represents scene urban areas. We dispose of four spectral bands: blue, green, red and near infrared. We can see in Fig.1 (a) a representation of this image.

The second test image is another sample of Quickbird satellite image with exactly the same properties except the size, 500x280 pixels. The scene does contain also urban areas. The original image is represented in Fig.2 (a).

We will have two files for each image, "TrainFile.dat" and "TestFile.dat" respectively for learning and for classification, divided on six classes as described in Table 1.

3.2 The Results

The classification maps presented on (b) respectively in Fig. 1 and Fig. 2, are obtained when the classification is performed using the spatial information only (Haralick features). We can note the appearance of misclassifications. When the classification is performed using the spectral information only, we obtain the corresponding classification maps which are presented on (c) respectively in Fig. 1 and Fig. 2. These results appear as noisy as the spatial information that is not taken into account.

The fusion of the spectral and the spatial features give us the classification maps presented on (d) respectively in Fig. 1 and Fig. 2. The classification maps are less noisy and the classification performances are increased globally as well as almost all the classes. It matches well with an urban land cover map in terms of smoothness of the classes; and it also represents more connected classes.

Table 2 summarizes the results obtained using the SVM classification with Gaussian RBF kernel. These values were extracted from the confusion matrix. The overall accuracy is the percentage of correctly classified pixels. Kappa coefficient is another criterion classically used in remote sensing classification to measure the degree of agreement and takes into account the correct classification that may have been obtained “by chance” by weighting the measured accuracies.

The use of this composite kernel (16) gives good classification results for the overall accuracy and the Kappa coefficient. Moreover, with all of the accuracies over 90%, this composite kernel seems also promising for the classification of remotely sensed images.

Table 1. Different classes

<i>Class N°</i>	<i>Class name</i>	<i>Train samples</i>	
		<i>Image 1</i>	<i>Image 2</i>
1	<i>Asphalt</i>	1 592	753
2	<i>Green area</i>	2 252	1 680
3	<i>Tree</i>	880	519
4	<i>Soil</i>	176	1 387
5	<i>Building</i>	4 217	1 282
6	<i>Shadow</i>	1 280	808
Total		10 397	6 429

Table 2. Classification accuracies for the classified images

<i>Methods</i>	<i>Image 1</i>			<i>Image 2</i>		
	<i>SVM spatial</i>	<i>SVM spectral</i>	<i>SVM Spectral & spatial</i>	<i>SVM spatial</i>	<i>SVM spectral</i>	<i>SVM Spectral & spatial</i>
<i>Overall accuracy</i>	83.19%	87.27%	93.68%	85.24%	88.02%	92.90%
<i>Kappa coefficient</i>	0.85	0.89	0.93	0.86	0.89	0.92

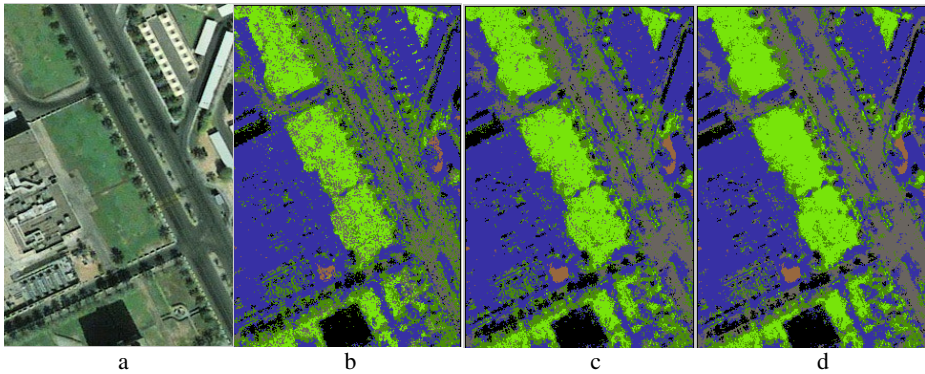


Fig. 1. (a) Original image, (b) Classification Map obtained with the classical RBF kernel using only spatial information, (c) Classification Map obtained with the classical RBF kernel using spectral information only and (d) map classification obtained with the proposed kernel

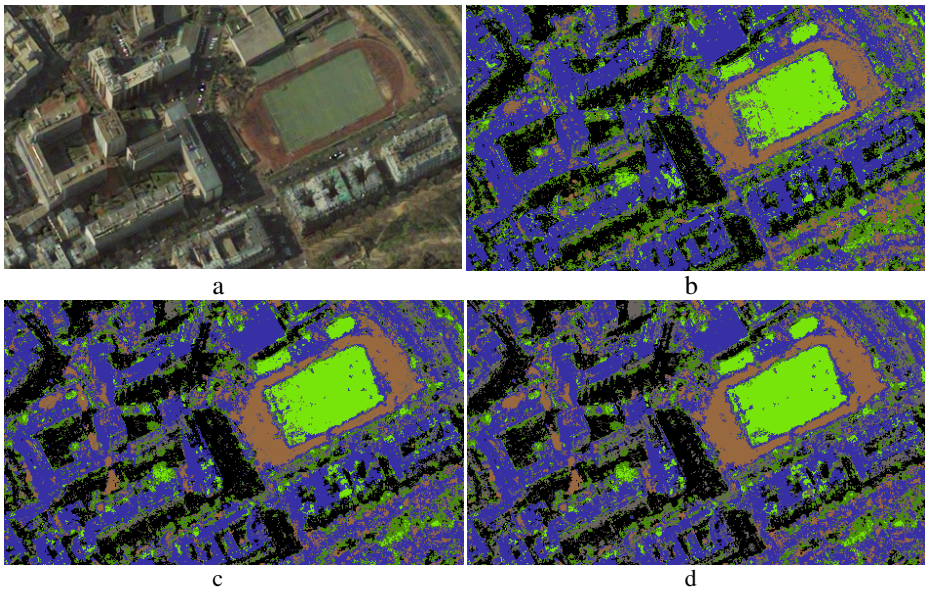


Fig. 2. (a) Original image, (b) Classification Map obtained with the classical RBF kernel using only spatial information, (c) Classification Map obtained with the classical RBF kernel using spectral information only and (d) map classification obtained with the proposed kernel

4 Conclusion

Addressing the classification of high resolution satellite images from urban areas, we have presented an algorithm taking simultaneously the spectral and the spatial information into account. This is achieved by concatenating the two vectors of attributes (the spectral values and the Haralick features).

This data combination allows a significant improvement of the classification performances when compared with the two sets of attributes used separately.

As a perspective of this work, we will be concentrating on the study of the kernel choice in order to determine the appropriate one, for this type of image classification.

Acknowledgments. This work was funded by CNRST Morocco and CNRS France Grant under “Convention CNRST CRNS” program SPI09/11.

References

1. Samson, C.: Contribution à la classification des images satellitaires par approche variationnelle et équations aux dérivées partielles: Thesis of doctorate, University of Nice-Sophia Antipolis (2000)
2. Townshend, J.R.G.: Land cover. *International Journal of Remote Sensing* 13, 1319–1328 (1992)
3. Hall, F.G., Townshend, J.R., Engman, E.T.: Status of remote sensing algorithms for estimation of land surface state parameters. *Remote Sensing of Environment* 51, 138–156 (1995)
4. Lu, D., Weng, Q.: A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28, 823–870 (2007)
5. Pal, M., Mather, P.M.: Support vector machines for classification in remote sensing. *International Journal of Remote Sensing* 26, 1007–1011 (2005)
6. Zhu, G., Blumberg, D.G.: Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. *Remote Sensing of Environment* 80, 233–240 (2002)
7. Scholkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., et al.: Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing* 45, 2758–2765 (1997)
8. Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R.: Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. In: *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007, Barcelona Spain* (2007)
9. Chiu, W.Y., Couloigner, I.: Evaluation of incorporating texture into wetland mapping from multispectral images. University of Calgary, Department of Geomatics Engineering, Calgary, Canada, *EARSel eProceedings* (2004)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. *IEEE Transactions on Systems Man and Cybernetics* (1973)
11. Weszka, J.S., Dyer, C.R., Rosenfeld, A.: A Comparative Study of Texture measures for Terrain Classification. *IEEE Transactions on Systems Man and Cybernetics* (1976)
12. Connors, R.W., Harlow, C.A.: A Theoretical Comparison of Texture Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1980)

13. Arvis, V., Debain, C., Berducat, M., Benassi, A.: Generalization of the cooccurrence matrix for colour images: application to colour texture classification. *Journal Image Analysis and Stereology* 23, 63–72 (2004)
14. Aseervatham, S.: Apprentissage à base de Noyaux Sémantiques pour le traitement de données textuelles: Thesis of doctorate, University of Paris 13 –Galilée Institut Laboratory of Data processing of Paris Nord (2007)
15. Bousquet, O.: Introduction au Support Vector Machines (SVM). Center mathematics applied, polytechnique school of Palaiseau (2001), <http://www.math.u-psud.fr/~blanchard/gtsvm/index.html>
16. Fauvel, M., Chanussot, J., Benediktsson, J.A.: A Combined Support Vector Machines Classification Based on Decision Fusion. In: *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2006, Denver, USA* (2006)
17. Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Francés, J., Calpe-Maravilla, J.: Composite kernels for hyperspectral image classification. *IEEE Geoscience Remote Sensing Letters* 3(1), 93–97 (2006)

Data Acquisition Enhancement in Shape and Multispectral Color Measurements of 3D Objects

Grzegorz Mączkowski, Robert Sitnik, and Jakub Krzesłowski

Institute of Micromechanics and Photonics,
Warsaw University of Technology
8 Boboli, Warsaw, PL 02-525
g.maczkowski@mchtr.pw.edu.pl
<http://ogx.mchtr.pw.edu.pl/>

Abstract. The paper presents application of a correction technique proposed for registering images by a CCD (Charged Coupled Device) camera. The device is installed in a 3D shape, angular reflectance distribution and multispectral color measurement system set up for digitization of cultural heritage objects. The procedure compensates for the camera noise and scene illumination non-uniformity according to previously published model. The paper describes the measurement system and provides analysis of data collected from measurements of the Color Checker reference target and uniform reference plane to evaluate enhancement of reflectance and shape accuracy after correction. Additionally a few examples of digitized objects are shown.

1 Introduction

Digitization of cultural heritage objects is recently becoming more available and well known in artifacts conservation. However, to better use its potential possibilities and to achieve truthful results, it is necessary to understand principles of image acquisition techniques and tune them, so that measurement uncertainty can be reduced.

The measurement process usually assumes determination of shape of the surface by the means of a 3D scanner and additionally its color using RGB or multispectral camera. Known solutions are based on laser scanning devices[1] as well as the ones using structured light projection technique[2,3] and usually a multi-band, self made camera with interference filters to separate spectral channels[1,4]. Some systems use a single detector for data acquisition, so that there is no need to manually align color texture to cloud of points which represents shape[3], but there are also solutions which require some manual adjustment[4]. Despite construction differences all devices known by the authors are based on non-contact data acquisition by capturing images with a digital camera. Therefore it is important to consider the capturing process and eliminate errors characteristic for digital image registration, such as noise and non-uniform illumination. Very thorough examination of this problems can be found in work[5].

This paper describes the application of method presented in work [5] to a developed integrated measurement system which includes 3D shape measurement system using structured light projection method, multispectral 10 band camera for color measurement and additional directional illumination setup for establishing surface reflectivity in a sense of BRDF (Bidirectional Reflectance Distribution Function) model. First the measurement setup is described, following brief outline of image acquisition method with some implementation details. After that evaluation results of shape and color measurement are described and commented.

2 Measurement Setup

The 3D shape measurement method uses the 3D Measurement with Algorithms of Directional Merging And Conversion (3DMADMAC) system [6]. This method of measurement is based on a structured light technique with digital sine patterns and Gray codes projection (Fig. 2a). The system consists of a Digital Light Projector (DLP) and an industrial CCD camera (Figure 2a). The 3DMADMAC system can be customized depending on end user requirements regarding size of measurement volume, amount of measurement points and duration of a single measurement. It also consists of a set of Software Development Kit tools which extend its functionality and automate all required measurement and data analysis algorithms.

Color of a measured surface is captured using a multi-spectral approach. The custom built camera was constructed to register images in 10 spectral bands with the aid of interference filters (Fig. 2b). The filter wheel is placed between the camera matrix and the lens. It has 11 slots, because additional empty window without a filter is necessary for performing shape and BRDF measurement which uses the same detector. The lens mount is located outside the case which allows for simple lens replacement according to required measurement conditions.

Normally the multispectral capture system uses analytic calibration procedure described in previous work [7]. It is based on capturing images of white reference plate for light source spectrum compensation and images of uniform background to compensate for illumination distribution and spectral filters' angular characteristic. Additionally transverse shifts of images due to positioning errors can be eliminated. For the purpose of the described analysis the spectral images calculation procedure was modified in order to properly perform new image acquisition process.

Another device is employed to measure angular reflectance distribution of surface. It comprises of a set of light sources distributed on a grid pattern and illuminating measurement volume. Each illuminator resembles a lambertian source and allows directional illumination of the investigated surface (Fig. 2b). Pictures are captured with all illuminators turned on sequentially which gives a collection of reflectance values in the function of illumination angle and serves as a BRDF estimation which leads to Phong parameters calculation [8]. The result is a BRDF function modeled with Phong parameters [9]. This paper however, will not further discuss analysis of angular reflectance data.

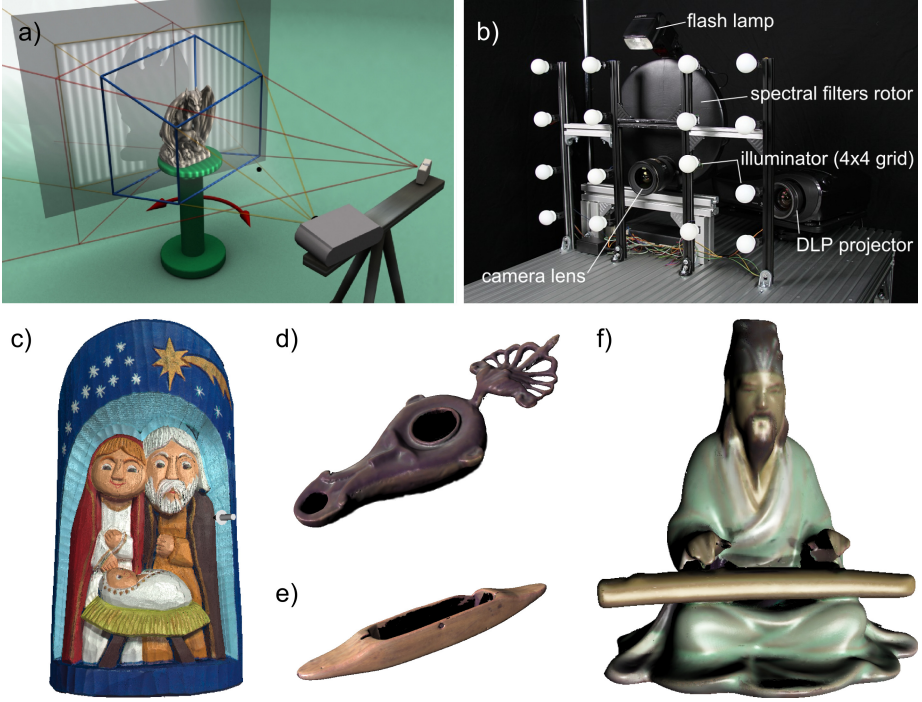


Fig. 1. Measurement system and virtual models of previously digitized objects: a) concept of the shape measurement device; b) constructed measurement setup; c) painted, wood-carved Nativity Scene figure; d) brass olive lamp; e) wooden shuttle on a loom; f) china figure

3 Data Acquisition Procedure

The goal of presented research was to implement and adopt a concept proposed by Mansouri et. al. in [5] of a digital camera calibration to reduce noise and non-uniformity of illumination. The purpose was to enhance data acquisition process in the case of the multispectral camera and 3D scanner. The proposed solution is based on the camera response in each pixel according to the following model described by equation (1).

$$[R] = [O] + [T] + [U \times S] \quad (1)$$

The camera response $[R]$ is a sum of a zero level (offset) image $[O]$, thermal signal proportional to acquisition time $[T]$ and a product of useful signal $[U]$ and sensor response $[S]$. Further reasoning leads to derivation of a formula for the useful

signal, knowing the offset, the thermal and the flat field $[F]$ characteristics, as in formula (2). For details see [5].

$$[U] = \frac{[R] - [O] - [T]}{[F] - [O] - [T_F]} \times F_{mean} \quad (2)$$

Implementation of the proposed model required acquisition of additional calibration data (offset, thermal and flat field images) and including them in data processing path. They are registered for every shutter value used in the measurement. Additionally the radiometric calibration procedure (RC) calculates mean values for flat field images which are afterwards used to rescale the useful images into proper range.

Because the software controlling the measurement system has a modular structure it was possible to implement an adapter for a detector module which is responsible for preparing raw images for further processing, without modification of the measurement head module. The only modification was made to the module, which calculates spectral reflectance, because it already used flat field image and it was not desirable to take it into account twice. Therefore the color measurement procedure benefits mainly due to camera noise reduction. The detector adapter module averages several frames to reduce temporal noise and applies compensation procedure mentioned above before sending data further along the processing path.

The camera installed in the measurement system is a Prosilica GE4900 device with 16Mpix sensor equipped with 12bit Analog-to-Digital Converter (ADC), which allows for increased signal quantization resolution, especially important for capturing spectral images. The camera was additionally tested for its linearity by the comparison of its response with a light meter. Results show a 0.99992 coefficient of cross correlation with the best fitted plane and root mean square error of 2.12% in relation to measured intensity range. Based on these results the decision was made to do not perform additional non-linearity correction.

Illumination for color measurement was provided by a digital light projector with a halogen lamp, which was also used for structured light projection for shape measurement. This allowed for a single procedure for flat field image acquisition for both measurement modules. Moreover the DLP projector's light flux is quite stable after sufficient warm up time, so there is no meaningful drift of spectral images lightness. Measurements were conducted in a controlled laboratory environment, after stabilization of thermal conditions for electronic devices and measurement volume.

4 Evaluation of Color Measurement Results

Evaluation of color measurement system accuracy was based on measurement of Color Checker target which was scanned with ($RC+$) and without radiometric calibration ($RC-$). Spectral reflectance data registered from both approaches was averaged for every color patch and compared with each other and with reference

measurement of the color target made with a Minolta CM-2600D spectrophotometer. Root mean square deviations from reference measurement for $RC+$ and $RC-$ were calculated for each color patch and are presented in Fig. 2. Measurement of two patches was not possible because of too low contrast of fringes projected during shape measurement. Consequently there was no sufficient phase information to calculate cloud of points for these patches. A few patches, indicated with a black border came out worse with the $RC+$ procedure, which is especially visible for the neutral ones. Our initial assumption was that it is the result of not implemented correction for camera non-linearity, but most likely this outcome is due to quantization noise which varies between spectral channels. It is the effect of a fixed shutter value for all channels and different transmission of spectral filters. Consequently the exposure of spectral images varies and so does the quantization noise. It is planned in the next step of research to implement constant exposure conditions by adjusting of shutter value for all spectral channels independently.

Fig. 3 shows spectral reflectance measurement results of sample color patches. Analysis of the reflectance plots leads to the conclusion that measurement errors are distributed along the whole registered spectrum and do not occur in any specific wavelength. However there is a trend that the difference between reference and estimated reflectance is bigger for small reflectance values, which agrees with the mentioned assumption that the quantization noise affects the camera response.

2.95 5.10	4.06 0.89	0.86 2.56	1.33 1.73	2.17 1.74	1.93 3.09
4.62 2.93	2.65 4.02	3.91 2.11	1.42 3.49	1.60 2.37	1.33 4.44
1.55 2.92	3.57 5.49	3.50 4.14	3.27 2.58	3.23 2.79	1.46 4.52
7.95 4.12	3.95 2.23	0.39 3.21	0.59 3.09	n.a.	n.a.

Fig. 2. RMS values for Color Checker patches measured with $RC+$ (bold font) and $RC-$ (normal font) procedure compared with the reference spectrophotometer measurement. Patches with black border are worse with $RC+$ comparing with $RC-$.

Comparison of color obtained in sRGB [10] color space shows that $RC+$ measurement gives more uniform output, with less chromatic noise and overall better hue than $RC-$ (Fig. 4).

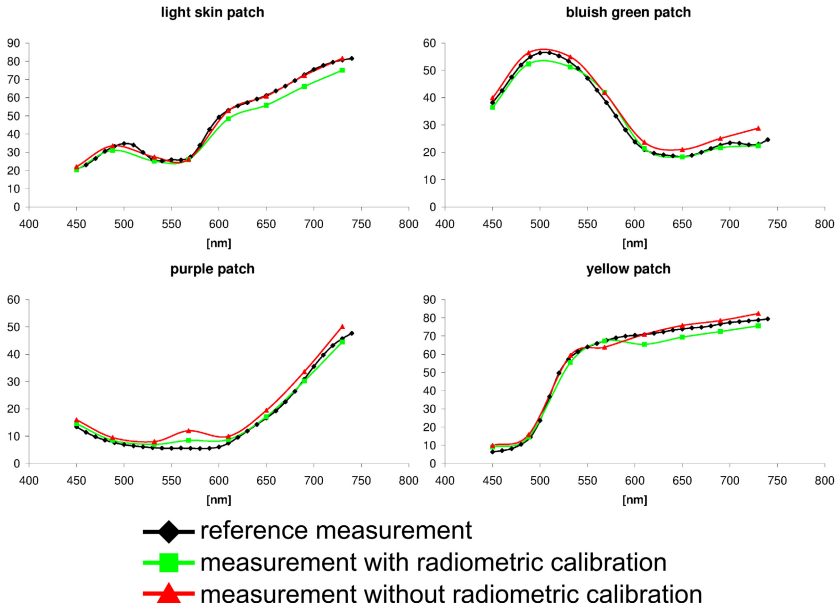


Fig. 3. Spectral reflectance plots for chosen color patches

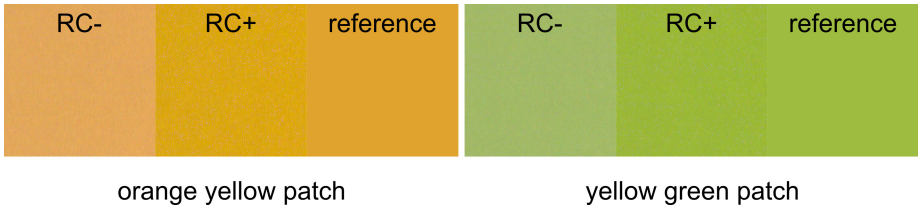


Fig. 4. Comparison of sRGB color values for exemplary target patches

5 Evaluation of Shape Measurement Results

The shape measurement uses sinusoidal fringe projection by the means of a DLP projector. Phase of fringes is established by the temporal phase shifting algorithm [11] in several reference planes within a measurement volume. Afterwards it is interpolated over the whole calibrated space. The algorithm gives very accurate results, but is sensitive to differences of lightness between consecutive fringe images and variations in phase shifts. In this case the latter are negligible, because the DLP projector has very high repeatability when the period of fringes is an integer multiple of number of projector pixels. Nevertheless image noise influences phase calculation accuracy and non-uniform illumination causes varying contrast of fringes within a single image. The same radiometric

calibration procedure was therefore applied to correct the fringe images to improve phase quality.

To evaluate results of this improvement two measurements of a uniform plane reference target of size 300×220 mm were conducted with and without radiometric calibration, so that deviations from an ideal fitted plane could be compared in both situations. The obtained clouds of points have on average 100 points per square millimeter. Fig. 5 visualizes error distribution after plane fitting to measurement data. It shows that in each interval of one, two and three standard deviations the plane fit error is 1.5 to 2 times smaller after the application of radiometric calibration procedure.

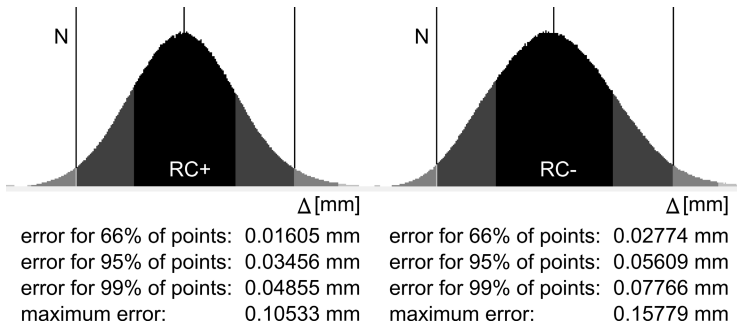


Fig. 5. Comparison of plane fit error distribution. The plot shows error value Δ around fitted plane versus number of points (N).

Fig. 6 shows comparison between registered corresponding fringe images and local fringes contrast values. It is visible that after radiometric correction contrast variation is less pronounced compared to the case without enhancement.

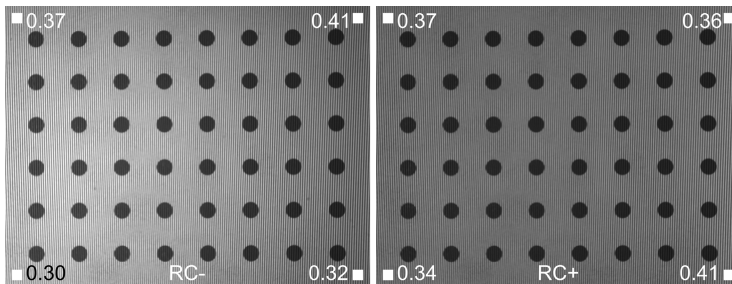


Fig. 6. Comparison of fringe images and their local contrast

6 Conclusions

Presented work shows application of a digital camera response model to image acquisition procedure in an integrated shape and reflectance measurement system. It shows that such enhancement leads to more accurate measurement results, especially for spectral and fringe images. It gives foundation for future research which may include application of this method for capturing images for BRDF measurement and implementation of constant exposure in different spectral channels which will further improve registered data.

It is necessary to point out that in case of 3D measurement the flat field image compensation is generally not sufficient to eliminate influence of uneven illumination, because in such case it is necessary to consider more complex illumination pattern. The measured object occupies certain volume in space, so points on its surface are placed in different distances from the detector and not on a single plane as it is assumed for flat field image compensation. Additionally different points on the measured surface may have different reflection properties, which will influence spectral acquisition. This introduces opportunity for future work, because solving the mentioned problems will be very important for faithful digitization of cultural heritage objects.

Acknowledgments. This work was performed under the grant No. PL0097 financed by the Norwegian Financial Mechanism and EEA Financial Mechanism (2004-2009) and partially under the statutory work of Warsaw University of Technology.

References

1. Tonsho, K., Akao, Y., Tsumura, N., Miyake, Y.: Development of gonio-photometric imaging system for recording reflectance spectra of 3d objects. In: Proc. SPIE, vol. 4663, pp. 370–378 (2002)
2. Sitnik, R., Mączkowski, G., Krzesłowski, J.: Calculation Methods for Digital Model Creation Based on Integrated Shape, Color and Angular Reflectivity Measurement. In: Ioannides, M., Fellner, D., Georgopoulos, A., Hadjimitsis, D.G. (eds.) EuroMed 2010. LNCS, vol. 6436, pp. 13–27. Springer, Heidelberg (2010)
3. Mansouri, A., Lathuiliere, A., Marzani, F., Voisin, Y., Gouton, P.: Toward a 3d multispectral scanner: an application to multimedia. *IEEE Multimedia* 14(1), 40–47 (2007)
4. Simon, C., Huxhagen, U., Mansouri, A., Heritage, A., Boochs, F., Marzani, F.: Integration of high resolution spatial and spectral data acquisition systems to provide complementary datasets for cultural heritage applications. In: Proc. SPIE, vol. 7531(1), p. 75310L (2010)
5. Mansouri, A., Marzani, F.S., Gouton, P.: Development of a protocol for ccd calibration: Application to a multispectral imaging system. *International Journal of Robotics and Automation* 20(2), 94–100 (2005)
6. Sitnik, R., Kujawska, M., Woznicki, J.: Digital fringe projection system for large-volume 360-deg shape measurement. *Opt. Eng.* 41, 443–449 (2002)

7. Mączkowski, G., Sitnik, R., Krzesłowski, J.: Integrated method for 3d shape and multispectral color measurement. *Journal of Imaging Science and Technology* 55(3), 030502–(10) (2011)
8. Mączkowski, G., Sitnik, R., Krzesłowski, J.: Integrated three-dimensional shape and reflection properties measurement system. *Appl. Opt.* 50, 532–541 (2011)
9. Phong, B.: Illumination for computer generated pictures. *Communications of the ACM* 18, 311–317 (1975)
10. IEC 61966-2-1:1999: Multimedia systems and equipment. Colour measurement and management. Colour management. Default RGB colour space. sRGB
11. Chen, F., Brown, G.M., Song, M.: Overview of three-dimensional shape measurement using optical methods. *Optical Engineering* 39(1), 10 (2000)

Multi-model Approach for Multicomponent Texture Classification

Ahmed Drissi El Maliani¹, Mohammed El Hassouni²,
Yannick Berthoumieu³, and Driss Aboutajdine¹

¹ LRIT, Unité Associée au CNRST (URAC 29),
Mohammed V University, Agdal, Morocco

² DESTEC, FLSHR, Mohammed V University, Agdal, Morocco

³ IMS- Groupe Signal- UMR 5218 CNRS, ENSEIRB, University Bordeaux, France

Abstract. This paper concerns multicomponent texture classification. The aim is to provide a flexible model when wavelet subband coefficients of components do not have the same distributions. Example of such case is when color textures are represented in a perceptual color space. In this kind of representation, the separability between luminance and chrominance components have to be considered in the modeling process. The contribution of this work consists in proposing a multi-model based characterization for this type of multicomponent images. For this, two models M_L and M_{C_r} are used in order to extract features from luminance and chrominance components, respectively. We discuss in detail and define the multi-model when textures are represented in the HSV color space as a special case of multicomponent analysis. Experimental results show that the proposed approach improves performances of the classification system when compared with existing methods.

Keywords: Multicomponent textures, Copula, Rao distance.

1 Introduction

Analysis of multicomponent images has become an important and very challenging task with the continuous advance of multimedia tools. Multicomponent image databases are bigger and need to convenient techniques in order to be managed. Many works emphasized that textured images are simple to model in the wavelets domain [8][9]. Distributions of the resulting subbands are characterized using well suited parametric models like generalized Gaussian distribution (GGD) [1] or Weibull distribution [2]. This marginal approach is convenient for unicomponent images such as grey level textures. But when retrieval or classification systems deal with multicomponent textures, the dependence across components have to be modeled using multivariate distributions as the multivariate generalized Gaussian distribution (MGGD) [3], or more recently using copula theory [11][4]. Such multivariate modeling leads to considerable enhancements when compared with the marginal modeling. However, in all aforementioned multivariate studies, the marginal behavior of the different components is not taken

into account along with modeling dependencies among those components. This fact make the modeling less pertinent, especially when components are separable, i.e do not have the same marginal distributions. An example of such case is when color textures are represented in a luminance-chrominance color space such as HSV (Hue Saturation Value). Luminance-chrominance or more specifically perceptual color spaces represent the most natural and intuitive way to describe color images. Researches on the human visual system revealed that the human eye percepts colors as a luminance and chrominance separated signals, and considers the chrominance as a Hue (pure color information) and saturation (level of intensity of the Hue). Modeling color textures in luminance-chrominance color spaces has been considered by many researchers as in [6] and [7]. However, in these works, one given model has been used to describe luminance and chrominance which are considered separable. For this, coming up with a statistical model that takes into account the different natures of luminance and chrominance channels seems to be a welcome advantage. Furthermore, in perceptual color spaces, the existence of a circular component (Hue) must also be taken into account in the feature extraction step.

Based on these assumptions, we propose a multi-model for color texture classification in the HSV color space. Wavelet subbands of the luminance and chrominance are characterized using two independent joint parametric models M_L and M_{C_r} , respectively. These latters repose on copula theory due to its ability to represent different marginals in one joint model which is the case of chrominance components.

2 Statistical Modeling of Multicomponent Textures

As said in the introduction, we detail a specific example of multicomponent images which is color textures when represented in HSV color space. In this section, we define the wavelet representation of color textures, we give a review of copulas and define the multi-model (models M_L and M_{C_r}).

2.1 DTCWT Representation of Color Textures in HSV Color Space

We consider the dual tree complex wavelet transform (DTCWT) in order to overcome disadvantages of the classic discrete wavelet transform (DWT) decomposition. DWT suffers from the lack of shift-invariance and directional selectivity, since it provides only three orientations at each decomposition level. DTCWT is based on two real wavelets to resort with complex wavelet coefficients, which will be shift-invariant. DTCWT provides six detail subbands per scale instead of three subbands in the case of DWT, which presents a rich directional selectivity.

Let us suppose a color texture I_M from the database. I_M is represented in the HSV perceptual color space. Let l , h and st , be the luminance (value or brightness), hue and saturation components of I_M respectively. We decompose each of those components via a DTCWT, and we call $l_k = l_{s;o}$, $h_k = h_{s;o}$, $st_k = st_{s;o}$ the wavelet subbands in scale s and orientation o respectively for the three components.

2.2 Copulas Theory

A copula is a multivariate cumulative distribution function (cdf), defined on the d -dimensional unit cube $[0; 1]^d$. Given a d -dimensional vector $X = [X_1, \dots, X_d]$ on the unit cube $[0; 1]$, with a cumulative distribution function F and marginal distributions F_1, \dots, F_d Sklar theorem [10] shows that there exist a d -dimensional copula C such that:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (1)$$

The joint PDF is then deduced uniquely from the margins and the copula density (dependence structure) as follows:

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i) \quad (2)$$

where $f_i, i = 1, \dots, d$, represent the marginal densities.

We use the Gaussian copula for its advantages in term of computation and naturality of the dependence structure. Copula density c_ϕ is then defined by:

$$c_\phi(u, \Sigma) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \quad (3)$$

with $\vartheta_i = \phi^{-1}(F_i(u_i))$, and ϕ represents the standard normal cumulative distribution function. Σ denotes the correlation matrix, and I denotes the d -dimensional matrix identity.

To estimate parameters of copula based models, we use the Inference From Margins (IFM) method [12].

2.3 Multivariate Model for Luminance

We study the spatial structure information for luminance via the model M_L . The dataset representing the neighborhood to be modeled by M_L is constructed from each subband l_k by moving a window of size $d = (2p + 1) \times (2q + 1)$ in an overlapping manner. Assuming the spatial homogeneity of subbands, we start from a reference coefficient $l_k(i, j)$ from the k_{th} subband l_k , and then we concatenate neighbors to have:

$$l_k = [l_k(i - p, i - q), \dots, l_k(i + p, i + q)]^T$$

Based on the Gaussian copula, the model M_L is defined by choosing appropriate pdf as a marginal for luminance:

$$f_{M_L}(l) = c_\phi(F_1(l_1), F_2(l_2), \dots, F_N(l_N)) \prod_{i=1}^N f_i(l_i) \quad (4)$$

that is:

$$f_{M_i}(l; w, \Sigma) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \prod_{i=1}^N f_i(l_i, w_i) \quad (5)$$

where $w=(w_1, w_2, \dots, w_N)$ denotes the parameters of luminance marginals, and N represents the number of detail subbands.

2.4 Bivariate Model for Chrominance

We model the chrominance by a bivariate model representing the correlation between chrominance subbands. Given the k_{th} subband, this dependency is represented by 2-dimensional vector $c_r^{(k)}$ as $c_r^{(k)} = [h_k, st_k]$. The model M_{C_r} is also defined reposing on the Gaussian copula:

$$f_{M_{C_r}}(c_r) = c_\phi(F_1(h), F_2(st))f_1(h)f_2(st) \quad (6)$$

that is:

$$f_{M_{C_r}}(c_r; w, \Sigma) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\vartheta^T(\Sigma^{-1} - I)\vartheta\right] \prod_{i=1}^2 f_i(c_i, w_i)$$

where $w=(w_1, w_2)$ the parameters of chrominance components marginals, and Σ the correlation matrix.

3 Classification Results

3.1 Experimental Setup

Texture classification was chosen as an application in order to validate the proposed multi-model. Experiment was carried out on 24 textures from the Vistex database [14] as shown in [7]. Each image of size 512×512 was divided into subimages of size 32×32 pixels. Then for each image, we consider 96 from the resulting 256 subimages as the training set, while the remaining 160 subimages are considered as the test set. For all textures of our database, every component of each subimage was normalized by subtracting its mean and dividing by its standard deviation, and then decomposed using a 2 scales DTCWT with a Q-shift (14,14) tap filter. Here, we use the K-nearest neighbor classifier which is the most simple and straightforward classification method. KNN is a kind of instance based classifiers, where the main idea is that an instance is classified reposing on a similarity measure, and is accorded the label of the majority of its K-nearest neighbor. Thus we need to a pertinent similarity measure that accounts of the multi-model approach.

3.2 Similarity Measurement Based on Rao Distance

Measuring similarity between two color textures in the database is done by measuring closeness of there luminance and chrominance components individually according to a luminance and a chrominance Rao distances, and then by calculating the overall similarity measure between color textures using a combination method as :

$$L = \lambda L_L + (1 - \lambda)L_{C_r} \quad (7)$$

where L_L represents the Rao distance between luminance models, and L_{C_r} the Rao distance between chrominance models. The coefficient λ is calculated empirically by considering the learning set as the test set and then returning value of λ that leads to the best classification rates. In [5], we determined a closed form expression of Rao distance between two copula based pdfs as the sum of the Rao distances between the two copulas and the Rao distances between the marginal distributions. Thus, the Rao distance between two copulas based probability density functions $f(x; \theta_1)$ and $f(x; \theta_2)$ is defined as follows:

$$L(f(x; \theta_1) || f(x; \theta_2)) = L_{Gauss}(f(x; \Sigma_1) || f(x; \Sigma_2)) + \sum_{i=1}^d L_{Margins}(f(x; w^{(1)}) || f(x; w^{(2)})) \quad (8)$$

that is:

$$L(f(x; \theta_1) || f(x; \theta_2)) = \left[\frac{1}{2} \sum_{i=1}^d (\ln r^i)^2 \right]^{1/2} + \sum_{i=1}^d L_{Margins}(f(x; w^{(1)}) || f(x; w^{(2)})) \quad (9)$$

where $r^i, i = 1, \dots, d$ represents the eigenvalues of $\Sigma_1^{-1} \Sigma_2$.

Expressions of $L_{Margins}$ for Weibull and Gamma marginals were provided in [5] [16], for the vonMises marginal expression of the Rao distance can be found in [15].

3.3 The Multi-model

To come up with most pertinent multi-model for the multicomponent textures, we test different combinations of luminance and chrominance models. We use copula based multivariate Gamma and copula based multivariate Weibull models for luminance, beside the copula based bivariate models {vonMises, Weibull} and {vonMises, Gamma} for chrominance.

-Model M_L

From equation (9), if we consider Gamma marginals for luminance, the pdf of M_L is presented as follows:

$$f_{M_L}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \times \left(\frac{\beta^{-\alpha}}{\Gamma(\alpha)}\right)^d \prod_{i=1}^d x_i^{\alpha-1} \exp\left[-\sum_{i=1}^d \left(\frac{x_i}{\beta}\right)\right] \quad (10)$$

with $\theta = (\alpha, \beta, \Sigma)$, α represents the shape parameter, β represents the scale parameter, and Σ denotes the covariance matrix. We call this model the copula based multivariate Gamma (*CopMGam*).

When, we consider Weibull marginals, the pdf of M_L is presented by:

$$f_{M_L}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \times \left(\frac{\tau}{\lambda}\right)^d \prod_{i=1}^d x_i^{\tau-1} \exp\left[-\sum_{i=1}^d \left(\frac{x_i}{\lambda}\right)^\tau\right] \quad (11)$$



Fig. 1. 24 texture classes from Vistex database

with $\theta = (\tau, \lambda, \Sigma)$, τ represents the shape parameter, λ represents the scale parameter, and Σ denotes the covariance matrix. In this case, we call the model M_L as copula based multivariate Weibull (*CopMWbl*).

-Model M_{C_r}

As already said, the chrominance model M_{C_r} accounts for the different natures of chrominance components. For this we use the property of merging different marginals when the model is based on copulas. We use two different marginals for the Hue and the Saturation components respectively. Then we use the Gaussian copula for the dependence structure. We test two couples of chrominance marginals.

Couple $\{vonMises, Weibull\}$, means that we use vonMises distribution for the Hue and Weibull for the Saturation. Thus, the probability density of the joint linear-circular model M_{C_r} is as follows:

$$f_{M_{C_r}}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\vartheta^T(\Sigma^{-1} - I)\vartheta\right] \times$$

$$\frac{\tau}{2\pi\lambda I_0(\nu)} \left(\frac{x_2}{\lambda}\right)^{\tau-1} \exp[\nu \cos(x_1 - \mu) - \left(\frac{x_2}{\lambda}\right)^\tau] \quad (12)$$

where $\theta = (\mu, \nu, \tau, \lambda, \Sigma)$ the hyperparameters of the chrominance model. μ and ν , are respectively the mean direction and the concentration parameters for the vonMises marginal, while τ and λ are respectively the shape and scale parameters of the Weibull marginal, Σ represents the covariance matrix.

If we consider $\{\text{vonMises}, \text{Gamma}\}$, the probability density of the joint linear-circular model M_{C_r} is defined as:

$$f_{M_{C_r}}(x, \theta) = \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \vartheta^T (\Sigma^{-1} - I) \vartheta\right] \times$$

$$\frac{\beta^{-\alpha}}{2\pi\Gamma(\alpha)I_0(\nu)} \left(\frac{x_2}{\beta}\right)^{\alpha-1} \exp[\nu \cos(x_1 - \mu) - \left(\frac{x_2}{\beta}\right)^\alpha] \quad (13)$$

where $\theta = (\mu, \nu, \alpha, \beta, \Sigma)$ the hyperparameters of the chrominance model. μ and ν , are respectively the mean direction and the concentration parameters for the vonMises marginal, while α and β are respectively the shape and scale parameters of the Gamma marginal, Σ represents the covariance matrix.

3.4 Results

We present results for different combinations of the multi-model:

- multi-model 1: CopMGam for luminance and the $\{\text{vonMises}, \text{Weibull}\}$ for chrominance.
- multi-model 2: CopMWbl for luminance and the $\{\text{vonMises}, \text{Weibull}\}$ for chrominance.
- multi-model 3: CopMGam for luminance and the $\{\text{vonMises}, \text{Gamma}\}$ for chrominance.
- multi-model 4: CopMWbl for luminance and the $\{\text{vonMises}, \text{Gamma}\}$ for chrominance.

Table 1, presents percentage classification of color textures using the four combinations of the multi-model in comparison with the MGGD based approach [3] and the copula based joint Weibull approach [4] in the RGB color space. It is to note that we consider the same aforementioned experience conditions for all methods. We can clearly observe from these results that considering multi-model for both luminance and chrominance information, beside accounting for the circular *Hue* component improves the classification rates, when compared with the uni modeling approach wether when components are characterized using MGGD or copula based joint Weibull models. A percentage classification of 94.37% is achieved using multi-model 1, while when luminance and chrominance components are characterized by the same model the rates are just 89.74% and 91.87% for models proposed in [3] and [4] respectively. We also deduce from Table 1 that the multi-model 1 leads to better rates, this is due to the ability of CopMGam in modeling spatial structure as stressed in [13] and the suitability of Weibull marginal for characterizing the Saturation component.

Table 1. Average classification rate using the multi-model method in comparison with existing approaches

Approach	Percentage classification (%)
multi-model 1	94.37
multi-model 2	93.48
multi-model 3	93.95
multi-model 4	93.15
MGGD/RGB [3]	89.74
CopWbl/RGB [4]	91.87

4 Conclusion

We proposed a multi-model characterization for multicomponent images and more specifically for color textures in perceptual color spaces. A model for luminance and another model for chrominance were used to consider the separability of these latters in such color spaces. We have also taken into account the angular nature of the Hue component in the modeling process. Results on the Vistex database show the superiority of the proposed approach in comparison with the existing statistical modeling approaches.

References

1. Do, M., Vetterli, M.: Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing* 11, 146–158 (2002)
2. Kwitt, R., Uhl, A.: Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval. In: 15th IEEE International Conference on Image Processing, ICIP 2008, pp. 933–936 (2008)
3. Verdoolaege, G., Scheunders, P.: Geodesics on the manifold of multivariate generalized gaussian distributions with an application to multicomponent texture discrimination. *Int. J. Comput. Vision* 95, 265–286 (2011)
4. Kwitt, R., Meerwald, P., Uhl, A.: Efficient texture Image Retrieval Using Copulas in a Bayesian Framework. *IEEE Transactions on Image Processing* 20, 2063–2077 (2010)
5. El Maliani, A.D., El Hassouni, M., Lasmar, N.-E., Berthoumieu, Y., Aboutajdine, D.: Color Texture Classification Using Rao Distance between Multivariate Copula Based Models. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011, Part II. LNCS, vol. 6855, pp. 498–505. Springer, Heidelberg (2011)
6. Kato, Z., Pong, T.C.: A Markov random field image segmentation model for color textured images. *Image and Vision Computing* 24(10), 1103–1114 (2006)
7. Qazi, I.U.H., Alata, O., Burie, J.C., Fernandez- Maloigne, C.: Color spectral analysis for spatial structure characterization of textures in IHLS color space. *Pattern Recognition* 43(3), 663–675 (2010)
8. Mallat, S.: *A Wavelet Tour of Signal Processing*, 3rd edn. The Sparse Way. Academic Press (2008)

9. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 837–842 (1996)
10. Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. *Publications de l'institut de Statistique de l'Université de Paris 8*, 229–231 (1959)
11. Sakji-Nsibi, S., Benazza-Benyahia, A.: Fast scalable retrieval of multispectral images with kullback-leibler divergence. In: *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010)*, Hong Kong, pp. 2333–2336 (2010)
12. Joe, H.: *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall (1997)
13. Stitou, Y., Berthoumieu, Y., Lasmar, N.: Copulas based multivariate gamma modeling for texture classification. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009, Taipei, Taiwan, April 19-24*, pp. 1045–1048 (2009)
14. MIT vision and modeling group, <http://vismod.media.mit.edu>
15. Ceolin, S., Hancock, E.R.: Characterising facial gender difference using fisher-rao metric. In: *Proceedings of the 2010 20th International Conference on Pattern Recognition, ICPR 2010*, pp. 4308–4311 (2010)
16. Reverter, F., Oller, J.M.: Computing the Rao distance for Gamma distributions. *Journal of Computational and Applied Mathematics* 157, 155–167 (2003)

Simultaneous Multispectral Imaging and Illuminant Estimation Using a Stereo Camera

Raju Shrestha and Jon Yngve Hardeberg

The Norwegian Color Research Laboratory, Gjøvik University College,
Gjøvik, Norway
raju.shrestha@hig.no
<http://www.colorlab.no>

Abstract. We propose here a novel approach to acquire a multispectral image and at the same time estimate the illuminant with the use of a stereo camera. Two images of a scene: one normal RGB and one filtered image with an appropriate optical filter selected from among readily available filters placed in front of a lens of the stereo camera are acquired. The spectral reflectance and/or color at each pixel on the scene are estimated from the corresponding outputs in the two images. In the mean time, the illuminant used during the image capture is estimated using chromagenic illuminant estimation method. Experiments with the simulated data show that this is a promising technique for simultaneous multispectral imaging and the illuminant estimation. Today's increasing commercial availability of digital stereo cameras makes the proposed solution a viable one for many applications.

1 Introduction

Multispectral imaging provides a solution to the limitations of conventional three channel (usually RGB) color imaging like metamerism and environment dependency. There are different types of multispectral imaging systems, most of them are filter-based which use additional filters to expand the number of color channels. The state-of-the art filter-based multispectral imaging systems [6,8,19] acquire images in multiple shots. Their use is, therefore, limited to static scenes only, thus making them less useful and less practical. Shrestha et al. [16,17] has made a comprehensive study and proposed a practical and feasible 6-channel multispectral imaging system with the use of a stereo camera. Depending upon the sensitivities of the two cameras in the stereo system, one or two appropriate optical filters are selected from among a set of readily available filters and placed in front of one or both lenses of the stereo camera, so that they will modify the sensitivities of one or two cameras to produce six channels (three each contributed from the two cameras) in the visible spectrum so as to give optimal estimation of the scene spectral reflectance and/or the color. This produces a 6-channel multispectral system.

Color constancy is yet another important issue in color imaging. It is the ability to account for the color of the light source which allows seeing the color of an object more or less the same under different lighting conditions. Human vision is said to be color constant as it has a natural tendency to correct for the effects of the color of the light source. Computational color constancy tries to emulate the color constancy in color imaging, and this is one of the fundamental requirements in many color imaging and computer vision applications. Computational color constancy models, in general, comprise of two steps: illuminant estimation and color correction. The illuminant estimation is the primary task in a computational color constancy algorithm. Knowing the estimated illuminant, the effects of the color of the illuminant are corrected to obtain the desired color constant image. Many methods have been proposed for the illuminant estimation, and these methods are typically based on the assumption of spatially uniform color of the light source across the scene. Some example methods are gray-world, max-RGB, a gamut based algorithm, neural networks, color-by-correlation, a Bayesian method. Yet another color constancy algorithm, known as the chromagenic color constancy, has been proposed by Finlayson et al. [4] that uses a special color filter which they named as chromagenic. This algorithm estimates the illuminant from two images: a normal RGB and a filtered RGB, of a scene. The algorithm has registration problems and also requires two shot images.

In this paper, we extend the multispectral imaging proposed by Shrestha et al. [16, 17] by making the system capable of acquiring multispectral image and at the same time estimating the illuminant under which the image has been acquired. For this, instead of two, only one filter is used in front of one of the lenses of the stereo camera. It thus acquires two images: one normal RGB image and one filtered image, in a single shot. The 6-channel multispectral image is estimated from these two images, and at the same time the illuminant is estimated using the same two images. The proposed system is thus capable not only of acquiring the multispectral image but also acquiring the normal RGB image, and at the same time capable of estimating the illuminant under which the images are taken. This gives users not only the flexibility to choose among a normal RGB and a multispectral image, but also to obtain a color constant image with the use of the estimated illuminant. The proposed system would therefore be useful in many applications where multispectral images and color constancy are applicable. We have performed experiments with the simulated data and they produce promising results.

After this introduction, we present the proposed system along with the method of multispectral imaging and the illuminant estimation in Section 2. We then present the experiments and results in Section 3. The results will be discussed next in Section 4, and finally we conclude the paper in Section 5.

2 Proposed System

The proposed system is constructed from a stereo camera and an appropriate optical filter in front of one of the lenses of the stereo camera (Fig. 1). Either a commercial stereo camera or two modern digital (RGB) cameras joined in a stereoscopic configuration can be used. An optimal filter is selected from the given set of filters through an exhaustive search. Since only one filter needs to be selected, the computational complexity is just $\mathcal{O}(n)$. We select a filter that produces minimum estimation errors (spectral or color) with regards to both the multispectral output and the color constancy output. We use here in this paper the minimum color estimation error as the criteria for the multispectral output for more accurate color reproduction, and use the most widely used median error [17] for more accurate illuminant estimation. However, depending on the application, spectral estimation error criteria could also be used. Section 3.2 describes the filter selection in the experiment in details.

The selected filter modifies the sensitivities of the filtered side of the camera producing six channels (three each contributed from the two cameras) in the visible spectrum. The system allows capturing two images, one normal RGB and one filtered RGB images of a scene. Furthermore, knowing the geometry of the stereo camera, not only the two images can be registered rather more precisely but also 3D information of the scene can be obtained. However 3D acquisition is beyond the scope of this paper. Among many registration techniques proposed in the literature, the phase-only correlation method (POC) [18] could be the one for precision registration. From the two registered images of a scene, the multispectral image is obtained and at the same the illuminant is estimated. The subsections below discuss the multispectral system model and the illuminant estimation with the proposed system.



Fig. 1. Illustration of a propose multispectral camera

2.1 Multispectral System Model

In order to model the proposed multispectral system, let $S = [s_R, s_G, s_B]$ denotes the matrix of spectral sensitivities of the three channels of the normal RGB camera of the stereo camera, and similarly $S^F = [s_R^F, s_G^F, s_B^F]$ is the matrix of the spectral sensitivities of the three channels of the filtered side of the stereo camera. Let t is the spectral transmittance of the selected filter, L is the spectral power distribution of the light source, and R is the spectral reflectance of the surface captured by the camera. As there is always acquisition noise introduced into the camera outputs, let n and n^F denote the noise vectors corresponding to the acquisition noise in the three channels of the normal and the filtered side of the stereo camera respectively. The camera responses of the normal (C) and the filtered (C^F) sides of the stereo camera are respectively given by:

$$C = S' \text{Diag}(L)R + n \quad (1)$$

$$\text{and } C^F = (S^F)' \text{Diag}(L)R + n^F, \quad (2)$$

where X' denotes the transpose of the matrix X . The combined response $\begin{bmatrix} C \\ C^F \end{bmatrix}$ of the two cameras gives six responses. The estimated reflectance (\tilde{R}) is obtained for the corresponding original reflectance (R) from these camera responses for the training and the test targets C_{train} and C respectively, using different spectral estimation methods. Training targets are the database of surface reflectance functions from which basis functions are generated and test targets are used to validate the performance of the device. Among many estimation algorithms proposed in the literature, we have investigated the performance of the proposed system with four different estimation methods: Imai and Berns (IB) [9], Linear Regression, Polynomial (PN) [3] and Neural Network (NN) [12]. These methods are described in details by Shrestha et al. [17].

The estimated reflectances are evaluated using spectral as well as colorimetric metrics. Two different metrics: GFC (Goodness of Fit Coefficient) and RMS (Root Mean Square) error have been used as spectral metrics, and ΔE_{ab}^* (CIELAB Color Difference) as the colorimetric metric. For the details on these metrics also, we refer to Shrestha et al. [17].

2.2 Illuminant Estimation

The two images of a scene allow estimating the illuminant as well. We use the chromagenic illuminant estimation method proposed by Finlayson et al. [4]. The chromagenic algorithm is based on the first approximation that the image formed by placing a colored filter in front of the camera is the same as changing the illumination impinging on the scene. In other words, the responses of the camera with and without the filter can be considered as the responses of a single surface under two different illuminants. When the same surfaces are viewed under two light sources, the corresponding camera responses, to a good approximation, can be related by a linear transform [11]. Therefore, if C and C^F denote the unfiltered and the filtered camera responses respectively, then these responses can be related by the following equation:

$$C^F = MC, \quad (3)$$

where M is a 3×3 linear transformation matrix. M depends on both the illuminant and the filter used, and it can be computed as:

$$M = C^F C^+. \quad (4)$$

The transformation matrix M can be described as the transform that maps, in a least square sense, unfiltered to filtered responses of the camera under a given illuminant. A linear model of illuminant change is not perfect and may result in large estimation errors. More accurate mapping with a reduced estimation error can be obtained with a convex relational model by expressing an RGB image as a convex combination of the training RGBs for each training light in turn. The chromagenic illuminant estimation method is, therefore, based on the assumption that we know all possible illuminants a priori. The transforms M_i

are different for different illuminants l_i ; the matrix M_i is determined for each of these illuminants. This property of chromagenic camera responses is used to identify the illuminant in a scene, i.e., to solve the color constancy problem. Let $l_i(\lambda)$, $i = 1, \dots, m$ are the spectral power distributions of the possible known illuminants, and $r_j(\lambda)$, $i = 1, \dots, n$ is the reflectances of the n representative real world surfaces. For each illuminant i , we determine the camera responses without and with the chromagenic filter: C_i , and C_i^F respectively, which are $3 \times n$ matrices whose j^{th} column contains the camera responses of the j^{th} surface under the i^{th} illuminant. The transformation matrix M_i for the i^{th} illuminant is obtained using Eq. 4.

For a given test illuminant, we select an illuminant $l_{\text{est}}(\lambda)$ from all plausible illuminants l_i as the estimated illuminant, which gives the minimum error:

$$\text{est} = \underset{i}{\text{argmin}}(e_i), \quad i = 1, \dots, m \quad (5)$$

where e_i is the fitting error that can be calculated as:

$$e_i = \|M_i C - C^F\|, \quad i = 1, \dots, m. \quad (6)$$

The illuminant estimation algorithms are evaluated using the same framework as proposed by Hordley and Finlayson 7. They recommended using the median angular error over the mean root mean square (RMS) error. Angular error is intensity independent and it has been widely used in the literature 11,7. Let $C_{l_{\text{est}}}$ and $C_{l_{\text{act}}}$ be the camera responses of a white reflectance under the estimated and the actual illuminant respectively, then the angular error e_{ang} is calculated as:

$$e_{\text{ang}} = \text{acos} \left(\frac{C_{l_{\text{act}}}^T C_{l_{\text{est}}}}{\|C_{l_{\text{act}}}\| \|C_{l_{\text{est}}}\|} \right) \quad (7)$$

3 Experiments

Here we first discuss the experimental setup and then present three different experiments and the results they produced: first the filter selection (Section 3.2), then the multispectral imaging (Section 3.3) and finally the illuminant estimation (Section 3.4).

3.1 Experimental Setup

The experimental setup comprises of a camera, filters, illuminants, reflectance data and test targets. A modern digital stereo camera from Fujifilm: the *Fujifilm FinePix REAL 3D W1* (in short, Fujifilm 3D) has been used. Fig. 1 illustrates a multispectral camera system constructed from this camera with a filter in front of one of its lenses. The sensitivities of this camera as measured and used by Shrestha et al. 16 has been used. The sensitivities of its left and right cameras are shown in Fig. 2.

To make the simulated multispectral system more realistic, as much as 4% normally distributed Gaussian noise is introduced as a random shot noise and 12-bit quantization noise is incorporated by directly quantizing the simulated camera responses after the application of the shot noise. Simulated D50 illuminant, and the CIE 1964 10° color matching functions are used for color computation as it is the logical choice for each color checker patches subtends more than 2° from the lens position. In order to evaluate the system, 63 patches of the Gretag Macbeth Color Checker DC has been used as the training target; and 122 patches remained after omitting the outer surrounding achromatic patches, multiple white patches at the center, and the glossy patches in the S-column of the DC chart have been used as the test target. The training patches have been selected using linear distance minimization method (LDMM) proposed by Pellegrini et al. [15]. A color whose associated system output vector has maximum norm among all the target colors is selected first. The method then chooses the colors of the training set iteratively based on their distance from those already chosen; the maximum absolute difference is used as the distance metric.

For the illuminant estimation, the 1995 Munsell surface reflectances (denoted as R) and the illuminants: the 87 measured training illuminants (L_{87}) and the 287 test illuminants (L_{287}), all the data from Barnard et al. [1] have been used. 265 optical filters of three different types: exciter, dichroic, and emitter from Omega are used. Transmittances of the filters available in the company website [14] have been used. One supplier has been chosen as a one point solution for the filters, and the Omega Optical Inc. has been chosen as they have a large selection of filters and data is available online.

3.2 Experiment I: Filter Selection

As discussed in Section 2, an optimal filter that produces the minimum estimation errors is selected. We have used the color estimation error (ΔE_{ab}^*) as the criteria, and the filter that produces the minimum ΔE_{ab}^* by the multispectral system and the minimum illuminant estimation (e_{ang}) error that lies within a acceptable error threshold values is selected from a given set of filters through exhaustive search.

For the illuminant estimation, the transformation matrices $M_i, i = 1 \dots 87$ are computed by imaging the whole surface reflectances, R under the training illuminants L_{87} , and they will be used to estimate the test illuminants. The test illuminants are estimated using the chromagenic algorithm.

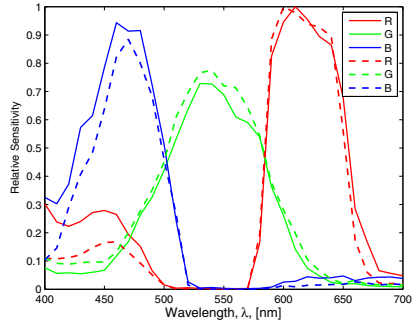


Fig. 2. Spectral sensitivities of the Fujifilm 3D camera (Left - solid, Right - dotted)

Depending on the application, an appropriate threshold values can be set for the color and the illuminant estimation errors. Here, as an illustration, we have chosen the $\Delta E_{ab}^* < 1.25$ and $e_{ang} < 2^\circ$ as the threshold values. Fig. 3 shows the XY plot of estimation errors with the 265 Omega filters, with the color estimation error along the X-axis and the angular error along the Y-axis. The filter selection algorithm chooses the XF1078 filter as shown in the figure. This filter has been selected with all the four spectral estimation methods discussed above. Fig. 4 shows the transmittance of the filter. This filter is then used in the next two experiments for the multispectral imaging and the illuminant estimation.

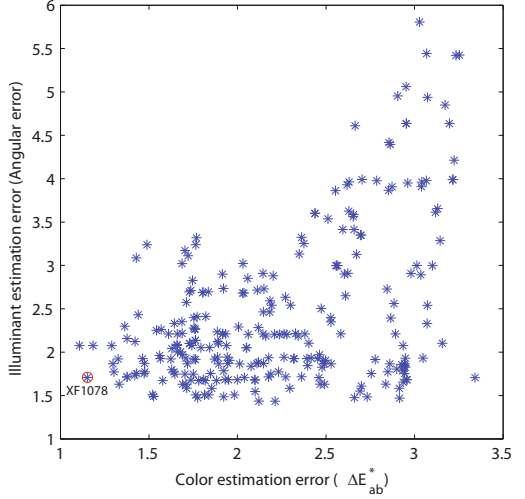


Fig. 3. Selection of a filter (red circled) from a set of 265 Omega filters

3.3 Experiment II: Multispectral Imaging

This experiment evaluates the proposed multispectral system constructed from the Fujifilm 3D and the selected filter, the Omega XF1078. Fig. 5 shows the normalized spectral sensitivities of the multispectral imaging system. The performance of the system has been investigated with all four spectral estimation methods (IM, LR, PN and NN) discussed in Section 2.1 and the results from three evaluation metrics (GFC, RMS and ΔE_{ab}^*) are reported. Table 1 shows the statistics (maximum/minimum, mean and standard deviation) of estimation errors side-by-side for the 3-channel and the proposed 6-channel systems.

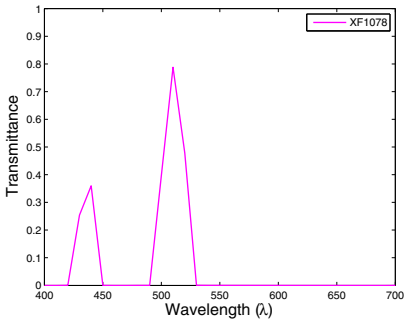


Fig. 4. Transmittance of the Omega XF1078 filter

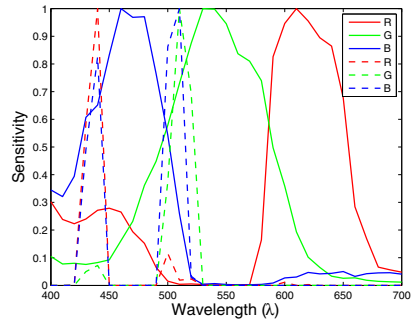


Fig. 5. Normalized spectral sensitivities of the 6-channel multispectral system

Table 1. Statistics of the estimation errors produced by the 6-channel system

Metric		3-Channel System				6-Channel System			
		IB	LR	PN	NN	IB	LR	PN	NN
GFC	Min	0.890	0.889	0.898	0.889	0.910	0.900	0.902	0.880
	Mean	0.990	0.990	0.990	0.990	0.996	0.996	0.996	0.996
	STD	0.017	0.017	0.015	0.017	0.009	0.010	0.010	0.012
RMS	Max	0.151	0.152	0.148	0.151	0.126	0.132	0.132	0.136
	Mean	0.031	0.031	0.029	0.031	0.021	0.020	0.020	0.021
	STD	0.023	0.023	0.021	0.023	0.018	0.019	0.018	0.020
ΔE_{ab}^*	Max	16.052	16.174	13.631	16.035	4.926	4.592	4.561	5.666
	Mean	3.486	3.542	3.552	3.484	1.153	1.130	1.196	1.050
	STD	3.320	3.359	2.766	3.316	0.958	0.816	0.825	0.872

The results show that the proposed 6-channel multispectral system outperforms the classic 3-channel system in terms of all the three metrics. We can also see that the performance of the four spectral estimation methods are comparable.

3.4 Experiment III: Illuminant Estimation

In this experiment, we use the real images generated from hyperspectral images of the eight natural scenes from Nascimento et al. [13]. The RGB images generated from the hyperspectral images using the Fujifilm3D camera and one of the 87 possible illuminant L_{87} are shown in Fig. 6. These hyperspectral images are available online in $820 \times 820 \times 33$ over 400-700nm bands in 10nm steps. However, the real image contents are less than 820×820 , but padded with zeros. Those padded empty data are removed and only real image contents are used. From these hyperspectral images, we obtain the unfiltered and the filtered versions of each image for every test illuminant L_{287} . The test illuminant is estimated in each case with the chromagenic algorithms using the transformation matrices $M_i, i = 1 \dots 87$ computed as discussed in Section 3.2. The median angular errors produced by the chromagenic algorithms along with the gray world [2] and the max-RGB [10] methods are given in Table 2. The results show significantly better estimation of the illuminant with the proposed system compared to the grayworld and max-RGB methods.

4 Discussion

The optimal filter used to construct the proposed system has been selected by setting acceptable error threshold values for the color and the illuminant estimation errors. As a further work, this selection could be made automatic with a single combined metric. Our experiments here are based on simulated images, and we assume that the two images are perfectly registered and there is no occlusion. It would be interesting to work further on experimental validation with real experiments taking into account the geometry of the stereo camera and two images from slightly different perspectives.



Fig. 6. The RGB images obtained from hyperspectral images of the 8 natural scenes from Nascimento et al. [13]

Table 2. Median angular errors for the 8 images generated from hyperspectral data of the scenes

Scene #	Gray World	Max RGB	Chromag.
1	5.50	4.75	3.52
2	9.86	21.85	6.19
3	9.45	3.20	4.62
4	5.50	4.75	3.52
5	7.32	11.04	2.05
6	2.83	6.94	2.21
7	0.99	2.12	1.64
8	2.87	3.10	3.49
Average	5.54	7.22	3.40

The experimental results show that the 6-channel multispectral system outperforms the 3-channel system significantly with all the four spectral estimation methods. As an example, the 6-channel system produces mean ΔE_{ab}^* of 1.05 with the NN method, while the 3-channel system produces 3.484. Moreover, the illuminant estimation with the chromagenic algorithm produces better results than the gray world and the max RGB methods. Finlayson et al. [4] have shown that the chromagenic based algorithms outperforms other color constancy algorithms like neural network, LP gamut mapping, Bayesian method, and color by correlation. The experimental results infer that simply selecting an optimal filter from a set of available filters produces promising results not only in the spectral and color reproduction from the multispectral imaging but also in the illuminant estimation. The performance could be improved further by using a set of a large number of filters, possibly from different manufacturers. The performance could also be improved significantly by using a custom designed filter [5].

The proposed system could be useful in digital color imaging where more accurate color imaging is required, and in color vision and robotics where color constant imaging is essential. Moreover, as the system is capable of acquiring multispectral and 3D images, it could also be used in multispectral imaging, for example, of culture and heritage.

5 Conclusion

This paper proposes a system capable of acquiring both a normal RGB image as well as a 6-channel multispectral image in a single shot, and at the same time capable of estimating the illuminant under which the image has been acquired. The system can be constructed from a off-the-shelf commercial stereo camera and

a filter. An optimal filter could either be selected from a set of available filters or custom designed. This allows a user flexibility to capture a color constant RGB image or a multispectral image or both. The system could be useful in many color imaging applications and computer vision.

References

1. Barnard, K., Cardei, V.C., Funt, B.: A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing* 11(9), 972–984 (2002)
2. Buchsbaum, G.: A spatial processor model for object colour perception. *Journal of the Franklin Institute* 310(1), 1–26 (1980)
3. Connah, D.R., Hardeberg, J.Y.: Spectral recovery using polynomial models. In: *Color Imaging X: Processing, Hardcopy, and Applications*. SPIE Proceedings, vol. 5667, pp. 65–75 (2005)
4. Finlayson, G.D., Hordley, S.D., Morovic, P.: Chromagenic colour constancy. In: 10th Congress of the International Colour Association (AIC), Granada, Spain, pp. 8–13 (May 2005)
5. Finlayson, G.D., Hordley, S.D., Morovic, P.: Chromagenic filter design. In: 10th Congress of the International Colour Association (AIC), Granada, Spain, pp. 1079–1083 (May 2005)
6. Hardeberg, J.Y., Schmitt, F., Brettel, H.: Multispectral color image capture using a liquid crystal tunable filter. *Optical Engineering* 41(10), 2532–2548 (2002)
7. Hordley, S.D., Finlayson, G.D.: Reevaluation of color constancy algorithm performance. *J. Opt. Soc. Am. A* 23(5), 1008–1020 (2006)
8. Huang, H.H.: Acquisition of multispectral images using digital cameras. In: *Asian Association on Remote Sensing, ACRS* (2004)
9. Imai, F.H., Berns, R.S.: Spectral estimation using trichromatic digital cameras. In: *International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, pp. 42–49 (1999)
10. Land, E.H.: The retinex theory of color vision. *Scientific American* 237(6), 108–128 (1977)
11. Maloney, L.T., Wandell, B.A.: Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A* 3(1), 29–33 (1986)
12. Mansouri, A., Marzani, F.S., Gouton, P.: Neural networks in two cascade algorithms for spectral reflectance reconstruction. In: *IEEE International Conference on Image Processing*, pp. 2053–2056 (2005)
13. Nascimento, S.M.C., Ferreira, F.P., Foster, D.H.: Statistics of spatial co-excitation ratios in natural scenes. *J. Opt. Soc. Am. A* 19(8), 1484–1490 (2002)
14. Omega: Omega filters. Omega Optical, Inc., <https://www.omegafilters.com/Products/Curvomatic> (last visited: February 2012)
15. Pellegrini, P., Novati, G., Schettini, R.: Selection of training sets for the characterisation of multispectral imaging systems. In: *PICS*, pp. 461–466 (2003)
16. Shrestha, R., Hardeberg, J.Y., Mansouri, A.: One-shot multispectral color imaging with a stereo camera. In: *Digital Photography VII, Electronic Imaging, Proceedings of SPIE/IS&T Electronic Imaging*, vol. 7876, p. 787609. SPIE, San Francisco (2011)

17. Shrestha, R., Mansouri, A., Hardeberg, J.Y.: Multispectral imaging using a stereo camera: Concept, design and assessment. *EURASIP Journal on Advances in Signal Processing* 2011(1) (September 2011)
18. Takita, K., Aoki, T., Sasaki, Y., Higuchi, T., Kobayashi, K.: High-accuracy subpixel image registration based on phase-only correlation. *IEICE Trans. Fundamentals* E86-A(8), 1925–1934 (2003)
19. Yamaguchi, M., Haneishi, H., Ohyama, N.: Beyond Red–Green–Blue (RGB): Spectrum-based color imaging technology. *Journal of Imaging Science and Technology* 52(1), 10201 (2008)

Multisource Fusion/Classification Using ICM and DSmT with New Decision Rule

Azeddine Elhassouny¹, Soufiane Idbraim¹, Aissam Bekkari¹,
Driss Mammass¹, and Danielle Ducrot²

¹IRF-SIC Laboratory, Faculty of Science, Agadir, Morocco
{info_azeddine,soufianeidbraim,a_bekkari}@yahoo.fr,
mammass@uiz.ac.ma

² CESBIO Laboratory Toulouse, France

Abstract. In this paper we introduce a new procedure for classification and change detection by the integration in a fusion process using hybrid DSmT model, both, the contextual information obtained from a supervised ICM classification with constraints and the temporal information with the use of two images taken at two different dates. Secondly, we have proposed a new decision rule based on the DSMP transformation, which is as an alternative and extension and overcoming the inherent limitations of the decision rules thus use the maximum of generalized belief functions.

The approach is evaluated on two LANDSAT ETM+ images, the results are promising.

Keywords: Detection of the changes, Image classification, Fusion, Hybrid DSmT model, Decision rule, DSMP, Satellite images, ICM.

1 Introduction

The management and the follow-up of the rural areas evolution are one of the major concerns for country planning. The satellite images offer a rapid and economic access to accurate homogeneous and updated information of studied territories. An example of application which results from this is related to the topic of the changes cartography, in this paper, we are interested to study the most subtle changes of the Argan land cover and other themes in the region of Agadir (Morocco) by contextual fusion /classification multidates based on hybrid DSmT model [1-3] and ICM with constraints [4, 5].

Our work environment, is the theory of Dezert-Smarandache [1-3] which is recent and very little implemented or used before the covered work of this paper, it was applied in multirate fusion for the short-term prediction of the winter land cover [6-9] and, recently, for the fusion and the multirate classification [10-12], although the theory of evidence, it is more exploited for fusion/classification [12-18] also, for classifier fusion [19-20].

Our methodology can be summarized as following, after preprocessing of the images, a supervised ICM classification with constraints [4, 5] is applied to the two

images, in order to recover the probabilities matrices for an after using in a step of fusion/classification basing on the theory of plausible and paradoxical reasoning known as Dezert-Smarandache Theory (DSMT) which allows to better assign the suitable pixels to the appropriate classes and also to detecte the changes.

In this paper, in section 2, we describe the mathematical basis of the recent theory of plausible and paradoxical reasoning (DSMT), and give a description of our decision rule. In section 3, we provide the results of our experimentation where the algorithm was applied to a LANDSAT ETM+ image, the classification results are discussed in the same section followed by conclusions in section 4.

2 Dezert - Smarandache Theory (DSMT)

2.1 Principles of the DSMT

The DSMT theory was conceived jointly by Jean Dezert and Florentin Smarandache [1-3], it is a new way of representing and fusing uncertain information. DSMT, considered as a generalization of the evidence theory of Dempster-Shafer [14], was developed to overcome the inherent limitations of DST (Dempster-Shafer Theory) [1-3]. The basic idea of DSMT rests on the definition of the hyper power set D^Θ from elements of Θ with \cup and \cap operators. From which the mass functions, the combination rules and the generalized belief functions are built.

We define a map as follows:

$$m_s(\cdot) : D^\Theta \rightarrow [0,1] \quad (1)$$

Associated to a given body of evidence s as

$$m_s(\phi) = 0 \quad (2)$$

and

$$\sum_{X \in D^\Theta} m_s(X) = 1 \quad (3)$$

with $m_s(X)$ is called the basic belief assignment/mass (bba) of X made by the source s .

The DSMT contains two models : the free model and the hybrid model [1-3], the first presents limits concerning the size of the hyper power set D^Θ , whereas the second has the advantage of minimizing this size, for this reason, it will be used in the continuation of our study.

Within the framework of DSMT, there are several rules combination [1-3], we have applied and implemented the majority of these rules in order to choose those which allow us to have good performances such as the Proportional Conflict Redistribution (PCR5).

Mainly, the PCR5 rule is based on the principle of the (total or partial) conflicting masses redistribution [1-3] to the non-empty sets involved in the conflicts proportionally with respect to their masses assigned by the sources (it can be also generalized for $N > 2$ sources) [2,3].

The generalized belief functions used in this study namely the Credibility (Cr), Plausibility (Pl) and DSmp Transformation [3] are defined for D^\ominus in $[0,1]$ and are given with more detail in [1-3].

2.2 Proposed Decision Rule

The decisions after combinations could be taken from the basic belief assignment/mass (bba) or the generalized belief functions (Credibility (Cr), Plausibility (Pl), (DSmp) transformation \dots etc), thus, to decide the belonging of a pixel to a given class, two cases are distinguished:

- The pixel belonging to a simple class (used to improve a classification): in this case we use one of the following decision criteria: maximum of bba, maximum of the Credibility (Cr) (with or without rejection), maximum of the Plausibility (Pl), Appriou criterium, DSmp criterium \dots etc) [6, 7, 11, 13, 15].
- The pixel belonging to a composed class(e.g. in the case of change detection), in this case we cannot use the functions quoted previously because they are increasing functions and unsuited to the decision for the elements of union and of intersection.

In an original step, we have proposed a new decision rule based on DSmp transformation and confidence interval to take in account the composed classes. In this decision rule we exploited the confidence interval $[Bel(X), Pl(X)]$ by the definition of a new measurement which we have named: Global uncertainty *IncG* which is the sum of the uncertainties (*Inc*) of the hyper power set elements:

$$\forall X \in D^\ominus \quad Inc(X) = Pl(X) - Bel(X) \quad (4)$$

This new decision rule described in Algorithm below is applied as follows:

For a given pixel x , we compare the Global uncertainty *IncG* of this pixel with a threshold (chosen experimentally). If it is lower than this threshold, the pixel is affected to the simple class which maximizes the DSmp transformation of all the simple classes, if not; it is affected to the composed class which maximizes the mass (bba) of all the composed classes.

The proposed decision rule

$$IncG = \sum_{x \in D^\ominus} (Pl(x) - Bel(x))$$

If ($IncG \leq threshold$) then

If $DSmp[x](\theta_i) = \max\{DSmp[x](\theta_i) \text{ with } 1 \leq i \leq n\}$ then $x \in \theta_i$

End if

Else

If $m[x](\cap \theta_i) = \max\{m[x](\cap \theta_i) \text{ with } 1 \leq i \leq |D^\ominus| - (n+1)\}$ then

$$x \in \cap \theta_i$$

End if

End if

3 Result and Discussion

3.1 Study Area and Used Data

The study area is located in the region of Souss, it is in southern Morocco. Geologically, it is the alluvial basin of the Oued Souss, separated from the Sahara by the Anti-Atlas mountains. The natural vegetation in the Souss is savanna dominated by the Argan (*Argania spinosa*), a local endemic tree found nowhere else, part of the area is now a UNESCO Biosphere reserve to protect this unique habitat.

The satellite images used in this study were taken respectively, in 19 March 2002 and 12 April 2005 by the Landsat ETM+, and are defined by the coordinates (Path 203, Row 39).

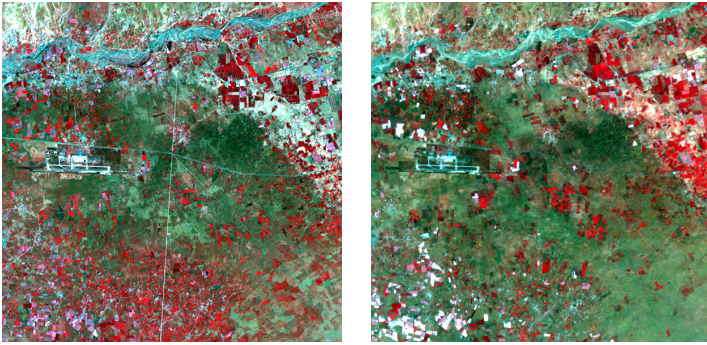


Fig. 1. RGB composite of the Landsat ETM+ images 2002 and 2005 (Agadir, Morocco)

3.2 ICM Classification with Constraints

After preprocessing of the two images and establishment of the samples trainings (Argan (A), Built/Oued (BO), Vegetation (V), Greenhouses (Gh) and Bareground (Bg)), a supervised contextual ICM classification with constraints [4, 5] is applied to the two images to have the probabilities of pixels belonging to classes. The classified image of 2002 is presented in figure 2 (a).

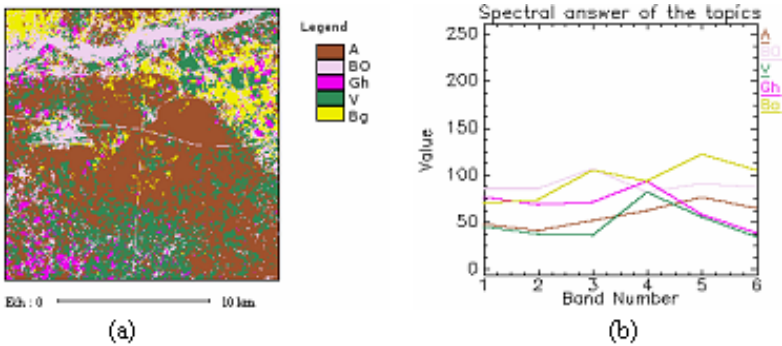


Fig. 2. ICM classification with constraints of the 2002 image (a) and Spectral response of the topics (b)

3.3 Multidates and Multi-source Fusion by the Hybrid DSMT Model

Our fusion process is composed of the following steps, first the definition of the framework Θ , then, the estimation of the mass functions of each focal element by the model of Appriou, finally, the application of the hybrid DSMT model.

3.3.1 Hyper Power Set

Taking in consideration the prior knowledge of the study area, we have identified 5 classes constituting the framework Θ section 3.2, Θ is defined as follows: $\Theta = \{A, BO, Gh, V, Bg\}$, some elements of the hyper power set D^Θ seem not being adjacents and exclusives. To realize a better adapted study to the real situations, some exclusivity constraints will be taken (hybrid DSMT model), for example $A \cap Gh$, which reduces the number of focal elements of the D^Θ .

3.3.2 Choice of Threshold

The decision is made for the simple classes and the classes of intersection by using our rule decision defined previously (section2.2), the threshold should be determined in advance by experimentation and analysis of total uncertainty distribution after standardization. We have tested our decision rule with various values of threshold, the following table 1 presents the occupancy rates of the simple classes (stable zones) and composed classes (change zones) according to the threshold.

Table 1. Occupancy rates of the simple classes and the composed classes according to the value of the threshold

Threshold	1.0e-26	1.0e-019	1.0e-016	1.0e-014	1.0e-012	1.0e-08
(%) of stable zones	0%	9.64%	32.12%	54.408%	76.86%	99.99%
(%)of change zones	100%	90.36%	67.88%	45.592%	23.32%	0.01%

The choice of threshold depends to the good results of detection with the use of the changed/ unchanged samples between the two dates (2002 and 2005), for this we have taken a suitable threshold that equal to 1.0e-014, which is coherent and appears to be close to reality basing on the test samples of the stable zones and the zones of changes between the two images.

3.3.3 Validation of the Results

3.3.3.1 *Prevalidation of the Results.* The fusion map obtained with an adequate threshold (1.0e-14) is presented in figure 3.

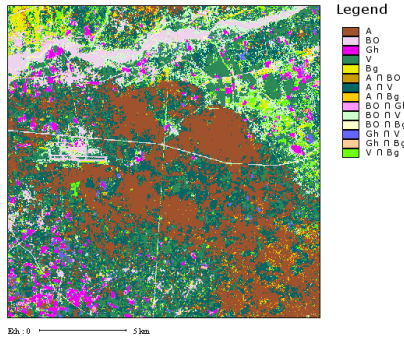


Fig. 3. Fusion map obtained with threshold of (1.0e-14)

From the fusion map, obtained with an adequate threshold we obtain the table 2 presenting the occupancy rate of the classes.

Table 2. Occupancy rate of the classes

Class	(%)	Class	(%)
A	29.04%	$A \cap Bg$	1.87%
BO	4.869%	$BO \cap Gh$	1.549%
Gh	3.18%	$BO \cap V$	7.25%
V	15.92%	$BO \cap Bg$	1.08%
Bg	1.399%	$Gh \cap V$	1.98%
$A \cap BO$	0.649%	$Gh \cap Bg$	0.5%
$A \cap V$	26.40%	$V \cap Bg$	3.30%

The table 2 illustrates the occupancy rates of the stable zones (simple classes) which reaches the rate of 54.40% and that of the zones of change (composed classes) which reaches 45.60%. From the table 2, we note that the Argan class (A) and Vegetation class (V) have known a great change compared to the other classes, indeed, we found for the composed class $A \cap V$ a rate of change of 26.40%.

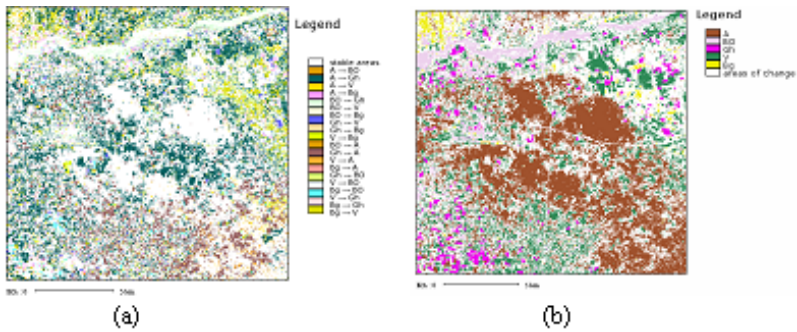


Fig. 4. Post-processed map of change zones (a) and stable zones map (b) obtained from fusion map

From the stable zones map figure 4-(a), we note that the zone of Oued Souss (BO), the zone of the urban areas along the road Agadir-Taroudant (BO), the international airport of Almassira (BO), the parcels of the Ouled teima region (V), the greenhouses (Gh) of the Biogra region and also some surface of argan (A) are all well detected as stable zones and are assigned to simple classes. This attribution is well justified because some zones are built and are unchangeable by nature.

In the post-processed fusion map figure 4-(b), areas that have known major changes are the Argan class which becomes bare ground or vegetation, which is well explained because of the deforestation. As result we found a rate of 18.95 of the Argan class changed to vegetation in 2005.

3.3.3.2 Validation per Spectral Signature of the Results. For this spectral evaluation, we have compared the spectral signatures of the some area of changes for the two images by taking in account the distributions of spectral signatures of different themes.

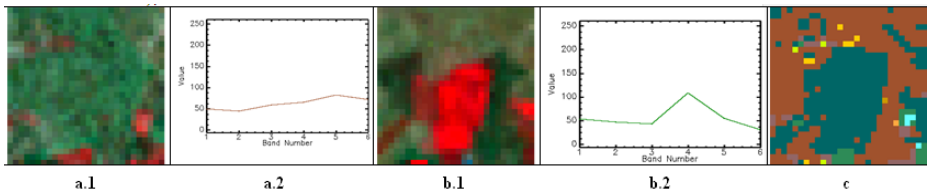


Fig. 5. Spectral signature of the Argan class (a.2) (ETM+2002 a.1) and of the Vegetation class (b.2) (ETM+ 2005 b.1), extract of the fusion map(c)

From the figure 5, we note that the extract area has known a change of Argan theme (Class A) to vegetation theme (class V), what is shown by the change of the pixels spectral signature of the extract area which had in 2002 a spectral signature of the Argan (Class A), and became in 2005 that of the vegetation (Class V). The table 5 summarizes all the changes detection and figure 6 shows the associated post-processed map of the fusion.

Table 3. Rate of change obtained by fusion between LANDSAT ETM+ 2002 and ETM+ 2005

Class	Number of pixels	Occupancy rate (%)	Class	Number of pixels	Occupancy rate (%)
A	104,542	29.0394	Gh→ V	3,112	0.8644
BO	17,538	4.8717	Gh→ Bg	471	0.1308
GH	11,458	3.1828	V→ Bg	2,662	0.7394
V	57,320	15.9222	V→ A	26,889	7.4692
Bg	5,048	1.4022	Bg→ A	2,439	0.6775
A→ BO	797	0.2214	Gh→ BO	915	0.2542
A→ V	68,202	18.9450	V→ BO	502	0.1394
A→Bg	4,295	1.1931	Bg→ BO	647	0.1797
BO→ Gh	4,650	1.2917	V→ Gh	4,005	1.1125
BO→ V	25,587	7.1075	Bg→ Gh	1,324	0.3678
BO Bg	3,227	0.8964	Bg→ V	12,831	3.5642

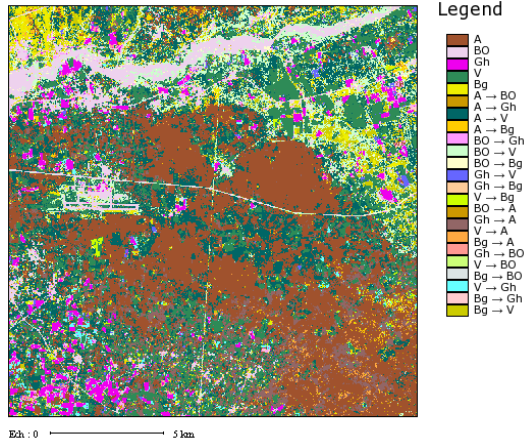


Fig. 6. Post-processed fusion map

4 Conclusion

In this paper we have proposed a new method for land cover changes detection. In first, we have included the spatial information in the process of fusion/classification by using a hybrid DSMT model with the introduction of contextual information using ICM classification with constraints, the use jointly of DSMT and ICM improves performance of the changes detection in terms of accuracy and exactitude. Secondly, we have proposed a new decision rule that has shown its performance and allowed us to overcome the limitations of rules decision based on the maximum generalized belief functions which are increasing and unsuited to the decision of the union and the intersection of elements. The application of this method for the cartography of the land cover changes is promising.

Acknowledgments. This work was funded by CNRST Morocco and CNRS France Grant under Convention CNRST CRNS program SPI09/11.

References

1. Smarandache, F., Dezert, J. (eds.): Advances and Applications of DSMT for Information Fusion (Collected works), vol. 1. American Research Press, Rehoboth (2004)
2. Smarandache, F., Dezert, J.: Advances and Applications of DSMT for Information Fusion (Collected works), vol. 2. American Research Press, Rehoboth (2006)
3. Smarandache, F., Dezert, J.: Applications and Advances of DSMT for Information Fusion (Collected works), vol. 3. American Research Press, Rehoboth (2009)
4. Idbraim, S., Ducrot, D., Mammass, D., Aboutajdine, D.: An unsupervised classification using a novel ICM method with constraints for land cover mapping from remote sensing imagery. International Review on Computers and Software (I.RE.CO.S.) 4(2) (2009)

5. Ducrot, D.: Méthodes d'analyse et d'interprétation d'images de télédétection multi-sources Extraction de caractéristiques du paysage, Habilitation thesis, France, Décembre 1 (2005)
6. Mercier, G.: Outils pour la télédétection opérationnelle, Habilitation thesis, Rennes I university, France, March 2 (2007)
7. Corgne, S., Hubert-Moy, L., Dezert, J., Mercier, G.: Land cover change prediction with a new theory of plausible and paradoxical reasoning. In: ISIF 2003, Colorado, USA (March 2003)
8. Moraa, B., Fourniera, R.A., Foucherb, S.: Application of evidential reasoning to improve the mapping of regenerating foreststands. *International Journal of Applied Earth Observation and Geoinformation* (2010)
9. Basse, R.M.: Université de Nice, La prise en compte de l'incertitude dans une démarche de modélisation prédictive. In: MoDyS, Lyon, France, Novembre 8-9 (2006)
10. Bouakache, A., Belhadj-Aissa, A.: Satellite image fusion using Dezert-Smarandache theory, DSMT-book3, Master Project Graduation, University Houari Boumediene (2009)
11. Khedam, R., Bouakache, A., Mercier, G., Belhadj-Aissa, A.: Fusion multidata à l'aide de la théorie de Dempster-Shafer pour la détection et la cartographie des changements: application aux milieux urbain et périurbain de la région d'alger. *Télédétection* 6(4), 359–404 (2006)
12. Djiknavorian, P.: Fusion d'informations dans un cadre de raisonnement de Dezert-Smarandache appliquée sur des rapports de capteurs ESM sous le STANAG 1241, Memory to obtain the degree (M.Se.), Laval University, Quebec (2008)
13. Anne-Laure, J., Martin, A., Maupin, P.: Gestion de l'information paradoxale contrainte par des requêtes pour la classification de cibles dans un réseau de capteurs multi-modalités. In: SCIGRAD 2008, Brest, France, Novembre 24-25 (2008)
14. Foucher, S., Germain, M., Boucher, J.M., Bénié, G.B.: Multisource Classification Using ICM and Dempster-Shafer Theory. *IEEE Transaction on Instrumentation and Measurement* 51(2) (April 2002)
15. Bloch, I.: Fusion d'informations en traitement du signal et des images. In: IC2, Hermès Science, Traité IC2, Paris, France (2003)
16. Lemeret, Y., Lefevre, E., Jolly, D.: Fusion de données provenant d'un laser et d'un radar en utilisant la théorie de Dempster-Shafer. In: MAJECSTIC 2004, France (2004)
17. Fiche, A., Martin, A.: Bayesian approach and continuous belief functions for classification. In: LFA, Annecy, France, November 5-6 (2009)
18. Germain, M., Boucher, J.M., Bénié, G.B., Beaudry, E.: Fusion évidentielle multi source basée sur une nouvelle approche statistique floue. In: ISIVC 2004, Brest, France (2004)
19. Martin, A.: Fusion de classifieurs pour la classification d'images sonar. In: RNTI-E 2005, pp. 259–268 (Novembre 2005)
20. Chitoub, S.: Combinaison de classifieurs: une approche pour l'amélioration de la classification d'images multisources multitudes de télédétection. *Télédétection* 4(3), 289–301 (2004)

Text Enhancement by PDE's Based Methods

Zouhir Mahani¹, Jalal Zahid¹, Sahar Saoud¹,
Mohammed El Rhabi², and Abdelilah Hakim²

¹ Université Ibn Zohr, ESTA, Laboratoire Matriaoux,
Systèmes et Technologies de l'information,
B.P: 33/S, 80000 Agadir - Maroc
{z.mahani,s.saoud}@uiz.ac.ma, j.zahid@gmail.com
<http://www.esta.ac.ma/>

² Université Cadi Ayyad, Faculté des Sciences et Techniques - Guéliz (FSTG),
Laboratoire de Mathématiques Appliquées et Informatique,
Bd. Abdelkrim El Khattabi , B.P. 618 Guéliz, 40000 Marrakech- Maroc
{elrhabi,abdelilah.hakim}@gmail.com
<http://www.fstg-marrakech.ac.ma/>

Abstract. In this work, we propose a new method to enhance text in document-image. Firstly, we introduce a classical model and a way to solve it by means of a non-convex optimization problem. So, a simultaneous estimation of the reflectance and the luminance is obtained when the non uniform illumination (also called luminance) is a smooth function and the reflectance is a function of bounded variation. We give an analyse of this problem and some conditions of existence and unicity. Then, we consider the “log” of the classical model. A new pde's model is proposed. This method is based on the resolution of an original partial differential equation (PDE) estimating the log of the luminance. We assume that the luminance is enough smooth and is the solution of a non classical second order's PDE. Then we deduce the reflectance from the estimated luminance and the acquired image. The effectiveness and the robustness of the proposed process are shown on numerical examples in real-world situation (images acquired from cameraphones). Then, we illustrate the ability of this method to improve an Optical Character Recognition (OCR) in text recognition.

1 Introduction

Computer vision based documents recognition could be an interesting way to dematerialize informations to manage clients and company's internal documents, offering enterprise wide fast access to business critical information while enhancing the achitecture in place. Typically the dematerialized document formats are PDF. Here, the problem could be separated as least in two steps: localization and recognition. Computer approach increases the performance of both steps, localization and recognition (see [1]). In the computer vision approaches, low cost cameras (webcam, cameraphone, ...) could introduce some distortions and noise artifacts, see [2] for an overview of document image degradation modeling.

The image restoration images by PDE's based models starts to produce significant results as images with the diffusion of documents to remove noise while preserving important information for readability ([3]) or to separate the back to front ink interferences ([4]). These works have shown that could be improved up to 30% recognition rate of OCR by restoring the document images.

Generally speaking, an image u could be seen as the product of a reflectance v and the illumination effect I (see [5]). This model has been used in [6,7] to restore blurred barcode signals under nonuniform illumination.

This paper provide a new method based on an anisotropic diffusion to estimate the luminance (or the log of the luminance) of a document-image in order to enhance the text. This method could be use for example in a mobile phone scanning solution or as a first step OCR processing.

We develop previously an isotropic method with some relative success(see [8]). This process allows us to estimate the luminance and to enhance texts then to share or print document images. By document images, we mean all kinds of images which include some handwritten or machine printed text. The basic camera phones often produce images of lower qualities. In addition, the conditions make it difficult shooting the readability and printing of document images. Regarding the text, these conditions make it difficult to interpret the information contained in this image without proper pretreatment. We have developped this method on a server side project, named "Qipit" : Qipit is a way to capture and share written documents with your camera phone or digital camera. Handwritten notes, signed contracts, whiteboards can be transformed into clean, crisp digital copies. After a specific treatment, this method has also directly embedded on photophones (Nokia, Samsung, Iphone . . .). As an example, on a Nokia N73 (3.15 megapixels, Symbian OS 9.1, S60 3rd edition), the time processing is less than 6 secondes (it depends on the choosen resolution). So, we have built an OEM business with the likes of Samsung, Sanyo and others with over 100 million copies sold and shipped factory-installed by mobile phone manufacturers to date. In this paper, since we present a preliminary result, the time processing is not discussed.

Our methods can produce three kinds of documents: black and white, grayscale or color ones. In all cases, input images are color pictures. Producing color documents should be seen as a different workflow, involving mostly the same algorithms as in the grayscale processing.

Document images are supposed to be obtained from a mobile device - a cameraphone more exactly, but could also come from any digital camera. In the following, we will only suppose the device to be a cameraphone as this is the case where most of the problems occur. Basic cameraphones often produce document images with poor quality - meaning it should be really tedious or even impossible to be read by someone or by an OCR (Optical Character Recognition). Because of constraints design, cameraphones have some limitations which we have to take into account, if we want to produce legible documents.

The paper is organized as follows : section [2] gives a brief description of a camera lens. Section [3] presents the non uniform illumination problem and defines a global model. Then two methods are given as solutions, namely a natural criterion

afterward a pde's based model. Section 4 describes the method and shows how to implement it. Finally, some numerical results illustrate this work in Section 5.

2 Description

Photographers control the camera and lens to expose the recording material (digital sensor or film) to the required amount of light. The controls include the focus of lens, the aperture of the lens (amount of light allowed to pass through the lens), the focal length and type of lens (macro, wide angle, or zoom), the duration of exposure (or shutter speed), the sensitivity of the medium to light intensity and color/wavelength, the nature of the light recording material, for example its resolution as measured in pixels or grains of silver halide.

Camera controls are inter-related, as the total amount of light reaching the image plane (the “*exposure*”) changes proportionately with the *duration of exposure*, *aperture* of the lens, and *focal length* of the lens (which changes as the lens is focused, or zoomed). Changing any of these controls alter the exposure. Many cameras automatically adjust the aperture of the lens to account for changes in focus, and some will accommodate changes in zoom as well.

The duration of an exposure is referred to as *shutter speed*, often even in cameras that don't have a physical shutter, and is typically measured in fractions of a second. *Aperture* is expressed by an *f-number* or *f-stop* (derived from focal ratio), which is proportional to the ratio of the focal length to the diameter of the *aperture*.

Exposures can be achieved through widely differing combinations of shutter speed and aperture. The chosen combination has an impact on the final result. In addition to the subject or camera movement that might vary depending on the shutter speed, the aperture (and focal length of the lens) determine the *depth of field*, which refers to the range of distances from the lens that will be in focus.

3 The Model

We suppose that the document-image u has been acquired by a cameraphone, we consider two sorts of problems due to this acquisition : image distortions and noise. Here, we adress a distortion, so-called “non uniform illumination” or variations of brihtness (see [9]). These variations render image processing difficults. One of these effects is : the shadows. In image processing, a shadow is considered as a region with low lightness and high gradients contours. So, we could separate two kinds of shadows, the own shadow and the shadows due to the acquisition. Own shadow occurs when the light hits a surface with a slope change. The brightness of pixels corresponding to the area decreases as the angle of incidence deviates from the normal of the surface. The brightness reaches its minimum when the incident light and the surface normal are orthogonal. Drop shadow occurs when the light source is obscured by an object before the light reflection on the surface. Others distortions problems like blur or warping are beyond of the scope of this paper.

In this paper, we consider an image $u \in \mathbb{R}^{N_1 \times N_2}$ where N_1 et N_2 are two integers. The non uniform illumination can be modeled as a multiplicative effect. This modelling combining the reflectance and the luminance of the image was proposed by Barrow and Tenenbaum in 1978 [10]. That said, due to various factors that may be involved in the construction of the image (the illumination of the object, the geometry of the scene acquired, the camera settings ...), such modelling is very difficult to tackle.

In 1999, Laszlo [11] has proposed a generative model of the image, based on a combinaison of Fredholm integral and a modelling of the settings of the camera. This model is very difficult to implement. Thus, it is the global illumination method [11] remains the most widely used :

$$u(x, y) = I(x, y) v(x, y) \cos_{\theta}(x, y) + b(x, y), \quad (1)$$

$u(x, y)$ is the grayscale of the pixel (x, y) , I is the luminance or the non uniform illumination, $v(x, y)$ the reflectance and $\cos_{\theta}(x, y)$ the cosine of the angle between the incident light ray and the surface normal at the point of the object. In image processing, this modelling is even more simplified by integrating $\cos_{\theta}(x, y)$ on the component $I(x, y)$. Thus, the final model becomes:

$$u(x, y) = I(x, y) v(x, y) + b(x, y). \quad (2)$$

This modelling of the image is far from perfect, because it does not take into account neither the problems of geometry of the object (the presence of surfaces which can create shadows on the object ...), or external factors in the formation of the image. The advantage of this simple model is to estimate the reflectance of an object from an approximation of its luminance.

The estimation of the reflectance v is crucial, because here we have an opportunity to characterize an object independently of illumination problems. Here, we suppose that the noise b is a gaussian random variable.

In the following, we suppose u, I and $v \in L^2(\Omega)$ (where $L^2(\Omega)$ is the square intergable space, Ω is the domain of the image) and the distortion is introduced by the measuring device and the conditions of acquisition.

3.1 Natural Criterion

In this part, we consider the model defined by the equation [1]. Thus and ideally, we would like to estimate both I and v up to the noise, a natural criterion could be:

$$J_1(I, v) = \frac{1}{2} \|Iv - u\|_{L^2(\Omega)}^2 + \lambda_1 R_1(v) + \lambda_2 R_2(I), \quad (3)$$

where λ_1, λ_2 are two positive hyperparameters, R_1, R_2 are two regularization functions allowing to control the noise and to provide some *a priori* on the pair of solution.

If we suppose that I is a smooth function and that the singularities are reported on v . A classical choice for the regularization function is :

$$R_1(v) = \int_{\Omega} |\nabla v| d\Omega, \quad R_2(I) = \int_{\Omega} |\nabla I|^2 d\Omega, \quad (4)$$

in other terms, $v \in BV(\Omega)$ (see [12] for more details on this space) and $I \in H^1(\Omega)$.

Then, we are looking for:

$$(I, v) = \arg \min_{l \in H^1(\Omega), w \in BV(\Omega)} J(l, w). \quad (5)$$

Proposition 1. *If we suppose that the noise $b = 0$ and if solutions of (5) exist where $I \in]0, 1]$ and v in $E = \{u \in L^2(\mathbb{R}^2), u(x, y) = 0 \text{ or } u(x, y) = 1\}$, let us consider two solutions (I_1, v_1) and (I_2, v_2) of (5) then $v_1 = v_2$ and when $u_1 = u_2 \neq 0$ then $I_1 = I_2$.*

Proposition 2. *If we suppose that I is a positive unknown constant in $]0, 1]$ there exist a solution (I, v) satisfying problem (5) in $E_A = \{v \in E, v(x, y) = 0 \forall x, y \in \mathbb{R}^2 \setminus A\}$ where $A = [0, 1]^2$.*

Proof —Under the assumption of the proposition, let I_j and v_j be a minimizing sequence of the problem (5), the sequence $(v_j)_j$ belongs to E_A and has bounded variations. Then by a classical compactness result (see [13]) for the functions with bounded variation and the fact that the all element of the sequence $(I_j)_j$ belongs to a compact $K \subset]0, 1]$. Then, we can extract a convergent subsequence, this subsequence converges toward (I_{∞}, v_{∞}) with $I_{\infty} \in K$. As E_A is a closed subset of $L^2(\mathbb{R}^2)$, we have $v_{\infty} \in E_A$.

Following these propositions, it could be interesting to add a constraint to project the solution v as closed as possible to E . A natural candidate function could be :

$$g : [m, M] \rightarrow \mathbb{R}^+ \\ x \mapsto \frac{(M-x)(x-m)}{M-m}$$

where m is the minimum value of the acquired image u and M is its maximum value.

So, we could penalized the criterion defined in (5) as follow:

$$J_2(I, v) = \frac{1}{2} \|Iv - u\|_{L^2(\Omega)}^2 + \lambda_1 R_1(v) + \lambda_2 R_2(I) + \lambda_3 G(v), \quad (6)$$

where $\lambda_3 > 0$ and $G(v) = \int_{\Omega} |g(v)|^2 d\Omega$.

Then, the researched solution is:

$$(I, v) = \arg \min_{l \in H^1(\Omega), w \in BV(\Omega)} J_2(l, w). \quad (7)$$

The Euler-Lagrange equations associated to the problem [6](#) are:

$$\begin{aligned} \frac{\partial J}{\partial v}(I, v) &= I(Iv - u) - \lambda_1 \operatorname{div} \left(\frac{\nabla v}{|\nabla v|} \right) + \lambda_3 (M + m - 2v) \frac{(M-v)(v-m)}{(M-m)^2} = 0, \\ \frac{\partial J}{\partial I}(I, v) &= v(Iv - u) - \lambda_2 (\Delta I) = 0. \end{aligned} \quad (8)$$

3.2 A Direct PDE Based Method

The objective of this approach is to extract the reflectance v from the non uniform illumination I . If we neglect for the moment the noise.

The model [2](#) becomes:

$$\begin{aligned} u(x, y) &= I(x, y)v(x, y), \\ \log(u(x, y)) &= \log(I(x, y)v(x, y)), \\ \log(u(x, y)) &= \log(I(x, y)) + \log(v(x, y)), \end{aligned} \quad (9)$$

Remark 1. *The grayscale image could be shifted by an offset to avoid the null values.*

We are thus reduced to find \tilde{I} and \tilde{v} satisfying:

$$\tilde{u} = \tilde{I} + \tilde{v} \quad (10)$$

So, as \tilde{I} is assumed smooth, we will find a way to approach \tilde{I} regardless of \tilde{v} , then deduce \tilde{v} by subtraction with \tilde{u} .

Here, we suppose that \tilde{I} is the solution of:

$$\begin{cases} w_t = s \max(0, s\Delta^A w), \\ \frac{\partial w}{\partial n} = 0 \text{ on } \partial\Omega, \\ w(t=0) = u, \end{cases} \quad (11)$$

where $s = 1$ if the grayscales of the text are on average “smaller” than the background ones and $s = -1$ else. Δ^A is defined as:

$$\Delta^A := \operatorname{div} \left(\varphi'(|\nabla u|) \frac{\nabla u}{|\nabla u|} \right). \quad (12)$$

Now, we introduce the orthonormal set (τ, n) , where n is defined by:

$$n(x) = \frac{\nabla u}{|\nabla u|}.$$

The vector fields τ et n are respectively tangent and normal to the level curves (isocontours) of u .

Then [\(12\)](#) becomes:

$$\Delta^A = \left(\frac{\varphi'(|\nabla u|)}{|\nabla u|} u_{tt} + \varphi''(|\nabla u|) u_{nn} \right), \quad (13)$$

where :

$$\begin{aligned} u_{tt} &= \tau^t \nabla^2 u \tau = \frac{1}{|\nabla u|^2} (u_{x_1}^2 u_{x_2 x_2} + u_{x_2}^2 u_{x_1 x_1} - 2u_{x_1} u_{x_2} u_{x_1 x_2}) \text{ and} \\ u_{nn} &= n^t \nabla^2 u n = \frac{1}{|\nabla u|^2} (u_{x_1}^2 u_{x_1 x_1} + u_{x_2}^2 u_{x_2 x_2} + 2u_{x_1} u_{x_2} u_{x_1 x_2}). \end{aligned}$$

u^t is the transpose of u , $\nabla^2 u$ is the hessian, u_{x_i} and $u_{x_i x_j}$, $i, j \in \{1, 2\}$ designe the first and second order partial derivative of u . This allows to separate both directions of diffusion: τ et n .

The term $\frac{\varphi'(|\nabla u|)}{|\nabla u|} u_{tt}$ defines the diffusion in direction τ , while $\varphi''(|\nabla u|) u_{nn}$ defines the diffusion in the direction n .

Thus, we have some qualitative requirements on the function φ : an isotropic diffusion on homogeneous region (part) of the image, $\varphi'(0) = 0$. We have also

$$\lim_{t \rightarrow 0} \frac{\varphi'(t)}{t} = \varphi''(0) > 0.$$

On the edge of the image, where we have a high gradient, we prefer a tangential diffusion and not in its transverse direction. So, we demand: $\lim_{t \rightarrow +\infty} \frac{t\varphi''(t)}{\varphi'(t)} = 0$.

An example of function satisfying these requirement is:

$$\varphi(t) = \sqrt{1 + t^2}, \quad (14)$$

If \tilde{I} is estimated then we can deduce \tilde{v} :

$$\tilde{v} = \frac{(1+s)}{2} \max(\tilde{u}) - s \left| \tilde{I} - \tilde{u} \right|, \quad (15)$$

4 Discretization

In the discrete form, an image is composed of a set of pixels indexed by (i, j) , $1 \leq i \leq N$, $1 \leq j \leq M$. $u = (u_{i,j})_{1 \leq i \leq N, 1 \leq j \leq M}$ belongs in X , where $X = \mathbb{R}^{N \times M}$. The space X is equipped with the euclidian inner scalar product:

$$\forall u, v \in X, \langle u, v \rangle_X = \sum_{i=1}^N \sum_{j=1}^M u_{i,j} v_{i,j}.$$

By a minor abuse of the notation, we state for X^m , where $m \geq 1$, the space $(\mathbb{R}^m)^{N \times M}$. The gradient of $u \in X$, written ∇u belongs to X^2 and could be defined by several manners. One of them consists to set $\nabla u = (g^{(1)}, g^{(2)})$ with:

$$g_{i,j}^{(1)} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N, \\ 0 & \text{si } i = N. \end{cases} \quad g_{i,j}^{(2)} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < M, \\ 0 & \text{si } j = M. \end{cases} \quad (16)$$

This div operator is defined in X^2 to X as the adjoint operator of $-\nabla$. So, for all $p = (p^{(1)}, p^{(2)}) \in X^2$, we have:

$$\forall z \in X, \langle \text{div} p, z \rangle = -\langle p, \nabla z \rangle.$$

In the case where the gradient is given by (16), one can prove that

$$(\operatorname{div}p)_{i,j} = (\operatorname{div}p)_{i,j}^{(1)} + (\operatorname{div}p)_{i,j}^{(2)}, \quad (17)$$

with

$$(\operatorname{div}p)_{i,j}^{(1)} = \begin{cases} p_{i,j}^{(1)} - p_{i-1,j}^{(1)} & \text{if } 1 < i < N, \\ p_{i,j}^{(1)} & \text{if } i = 1, \\ -p_{i-1,j}^{(1)} & \text{if } i = N. \end{cases}$$

$$(\operatorname{div}p)_{i,j}^{(2)} = \begin{cases} p_{i,j}^{(2)} - p_{i,j-1}^{(2)} & \text{if } 1 < j < M, \\ p_{i,j}^{(2)} & \text{if } j = 1, \\ -p_{i,j-1}^{(2)} & \text{if } j = N. \end{cases}$$

We state for all $u \in X$,

$$\Delta u = \operatorname{div}(\nabla u). \quad (18)$$

and

$$\Delta^A u = \operatorname{div} \left(\varphi'(|\nabla u|) \frac{\nabla u}{|\nabla u|} \right) \quad (19)$$

Then, from the definition of the divergence, we keep in this discretete form:

$$\forall u, v \in X, \langle \Delta u, v \rangle = -\langle \nabla u, \nabla v \rangle = \langle u, \Delta v \rangle. \quad (20)$$

Then, it follows the natural algorithm from [7, 17] and [18]:

Data: u the acquired image

Result: I the non uniform illumination and v the reflectance

initialization : Given $\varepsilon_1 > 0, \varepsilon_2 > 0, \lambda_1, \lambda_2$ and $\lambda_3 > 0, I^0 = \frac{u}{\max(u)},$

$v^0 = u$ and $\mu > 0$ adequately chosen

do

• update I and v :

$$v^{p+1} = v^p - \mu \left(I^p (I^p v^p - u) - \lambda_1 \operatorname{div} \left(\frac{\nabla v^p}{|\nabla v^p|} \right) + \lambda_3 (M + m - 2v^p) \frac{(M - v^p)(v^p - m)}{(M - m)^2} \right)$$

$$I^{p+1} = I^p - \mu \left(v^{p+1} (I^p v^{p+1} - u) - \lambda_2 \Delta I^p \right)$$

until $\|v^{p+1} - v^p\| < \varepsilon_1$ & $\|I^{p+1} - I^p\| < \varepsilon_2$

$v = v^{p+1}, I = I^{p+1}$

Algorithm 1. Model [7]

Now, if we use an explicit scheme in time and the discretization [19] for the anisotropic laplacian, we can derive the following algorithm from [11]:

Data: u the acquired image, $s = 1$ or -1
Result: v the reflectance
initialization : Given $\varepsilon > 0$, $\tilde{I}^0 = \tilde{u} := \log(u)$ and $dt > 0$ adequately chosen
do
 • update I :

$$\tilde{I}^{p+1} = \tilde{I}^p + dt \max(0, \Delta^A I^p)$$

until $\|\tilde{I}^{p+1} - \tilde{I}^p\| < \varepsilon$
 $\tilde{I} = \tilde{I}^{p+1}$ deduce \tilde{v} from [\[15\]](#) and $v = \exp(\tilde{v})$.

Algorithm 2. Model [\[11\]](#)

5 Numerical Results

In this section, we present some simulation results comparing both models. We show the ability of the proposed algorithms to successfully estimate the reflectance and the non uniform illumination and we compare also their performance by means of a text recognition software. As we don't control the recognition software errors, we present only how our process can improve this software. In other words, we show the recognition of the acquired image afterwards the recognition of the results obtained by the algorithms [\[1\]](#) and [\[2\]](#).

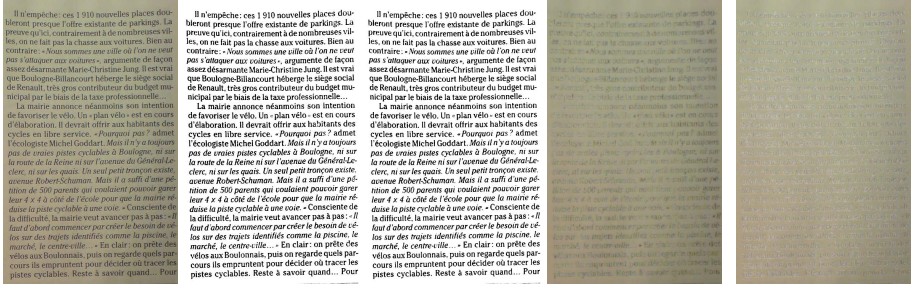
In the example 1, the original image is acquired from a Sony Ericksson K800i (3.2 megapixels) while images in Examples 2 and 3 are obtained from an Iphone 3GS (3 megapixels).

In these tests, the parameters of the algorithm [\[1\]](#)- criterion [\[7\]](#) are taken equal to $\lambda_1 = \lambda_2 = 0.1$, $\lambda_3 = 0.01$ and $\varepsilon = 0.00001$. For algorithm [\[2\]](#)- model [\[11\]](#), we take $s = 1$, $dt = 0.5$ and $\varepsilon = 0.00001$.

Figure [\[1\]](#) shows the estimated non uniform illumination and the estimated reflectance obtained from both algorithms. In this example, The image is affected by its own shadow. The resulting estimations are correct and quite similar.

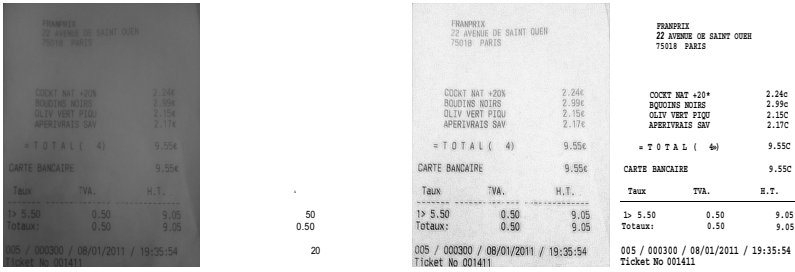
In figures [\[2\]](#) and [\[3\]](#), the images are acquired in a severe rough environment. These tests are very interesting because the images contain receipts. If we could recognize the contents of the receipts. The consumer can then follow his budget in a detailed way. Then, we can propose an alternative solution to barcode scanning. This solution is more tractable since we have only one shot and one scan when the barcode solution needs to decode the barcode for each product. Moreover, barcodes acquired from photophones have also distortion problems that we must take into account during the decoding process (see [\[7,6,14\]](#)).

In figure [\[2\]](#), the lightning conditions are very low and we deal with distortions which are beyond the scope of this work (blur and noise). Thus, here, we are on a "edge case" for both algorithms. As we can see, the text recognition software recognize some texts of the estimated reflectance for both algorithms whereas this software does not recognize any text from the original image. Both results are correct but the result of algorithm [\[2\]](#)- model [\[11\]](#) seems more suitable if we would like to read or to print the result. In figure [\[3\]](#), the image are affected by

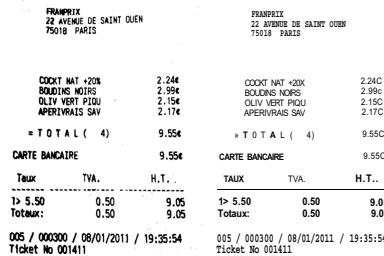


(a) Example 1 : original image ; (b) Text enhancement : estimated reflectance model [7] ; (c) Text enhancement : estimated reflectance model [11] ; (d) Text enhancement : estimated uniform illumination ; (e) Text enhancement : estimated uniform illumination ; model [7] ; (f) Text enhancement : estimated uniform illumination ; model [11]

Fig. 1. Original image acquired from a Sony Ericsson K800i



(a) Example 2, original image ; (b) Recognition by a classical OCR software ; (c) Text enhancement ; (d) Recognition by a classical OCR software ; (e) Text enhancement ; (f) Recognition by a classical OCR software ; (g) Text enhancement ; (h) Recognition by a classical OCR software ; (i) Text enhancement ; (j) Recognition by a classical OCR software



(e) Text enhancement ; (f) Recognition by a classical OCR software ; (g) Text enhancement ; (h) Recognition by a classical OCR software ; (i) Text enhancement ; (j) Recognition by a classical OCR software

Fig. 2. Original image acquired from an iPhone 3GS: first test

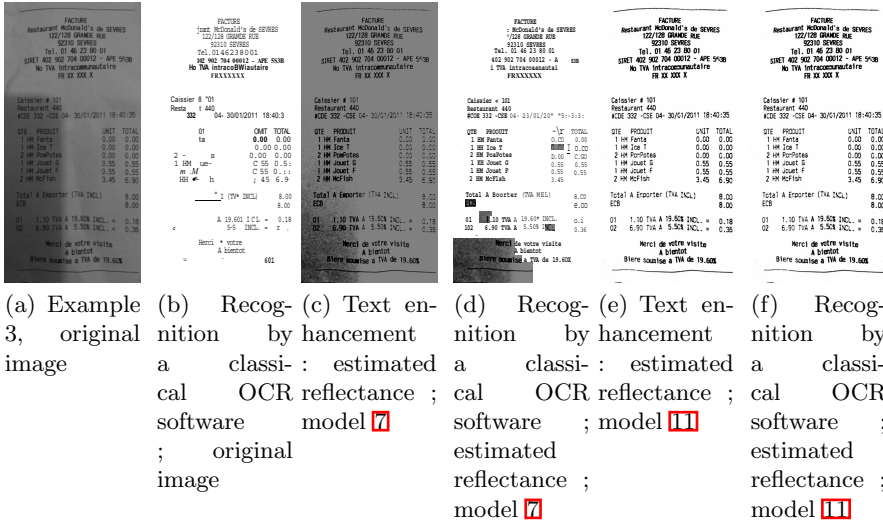


Fig. 3. Original image acquired from an iPhone 3GS: second test

two kinds of shadows : its own shadow and a drop shadow. In this test, the algorithm [2]-model [11] gives the best result.

6 Conclusion

Two methods for enhancing text in image-documents are proposed. The first is based on an highly non convex optimization problem while the second is based on a direct PDE's resolution. So far we have shown that the methods are robust to low lightning conditions. The second method based on an anisotropic pde's model provides the best result and is easier to implement. Moreover, the resulting algorithm depends only on few fixed parameters. However, this model needs some theoretical investigations and a rigorous analysis (conditions of existence, unicity, numerical analysis of the associated algorithm ...) that deserves further investigations. Based on the resolution of the associated global illumination model [2], both methods allow to estimate the luminance in a robust and reproducible way . In particular, this approach has been successfully applied to improve an OCR software to recognize texts from real images taken from photophones even in tough environment.

References

1. Wu, J., Caelli, T.: Model Based 3D object localization and recognition from a single intensity image. In: Adam, K., Tony, K., Cheng, S.Y. (eds.) Computer Vision and Shape Recognition, pp. 21–67 (1988)
2. Baird, H.S.: The State of the Art of Document Image Degradation Modeling. In: Proc. of 4th IAPR International Workshop on Document Analysis Systems, Rio de Janeiro, pp. 1–16 (2000)

3. Drira, F., Le Bourgeois, F., Emptoz, H.: Document images restoration by a new tensor based diffusion process: Application to the recognition of old printed documents. In: 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelone, pp. 321–325 (2009)
4. Moghaddam, R.F., Cheriet, M.: RSLDI: Restoration of single-sided low-quality document images. *Pattern Recognition, Special Issue on Handwriting Recognition* (42), 3355–3364 (2009)
5. Horn, B.K.: *Robot Vision*. MIT Press (1986)
6. Kim, J., Lee, H.: Joint nonuniform illumination estimation and deblurring for bar code signals. *Optic Express* 15(22), 14817–14837 (2007)
7. Dumas, L., El Rhabi, M., Rochefort, G.: An evolutionary approach for blind deconvolution of barcode images with nonuniform illumination. In: *IEEE Congress on Evolutionary Computation*, pp. 2423–2428 (2011)
8. Martin, A., Lefebure, M.: Realeyes3D SA, patent, <http://www.prior-ip.com/application/33411968/>
9. Gross, R., Brajovic, V.: An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In: Kittler, J., Nixon, M.S. (eds.) *AVBPA 2003. LNCS*, vol. 2688, pp. 10–18. Springer, Heidelberg (2003)
10. Barrow, H.G., Tenenbaum, J.M.: Recovering intrinsic scene characteristics from images. In: *CVS 1978*, p. 326 (1978)
11. Laszlo, S.-K.: Monte-Carlo Global Illumination Methods - State of the Art and New Developments. In: *SCCG 1999 (1999)* (invited talk)
12. Cohen, A., Wolfgang, D., Daubechies, I., DeVore, R.: Harmonic analysis of the space BV. *Rev. Mat. Iberoamericana* 19(1), 235–263 (2003)
13. Giusti, E.: *Minimal Surfaces and Functions of Bounded Total Variation*. Monographs in Mathematics, vol. 80. Birkhauser, Boston (1984)
14. El Rhabi, M., Rochefort, G.: Realeyes3D SA, patent, <http://www.wipo.int/patentscope/search/en/W02009112710>

Kernel-Based Laplacian Smoothing Method for 3D Mesh Denoising

Hicham Badri¹, Mohammed El Hassouni^{1,2}, and Driss Aboutajdine¹

¹ LRIT, Faculty of Science,

² DESTEC-FLSHR,

University Mohammed V -Agdal- Rabat, Morocco

{Hichambadri, Mohamed.Elhassouni}@gmail.com, aboutaj@fsr.ac.ma

Abstract. In this paper, we present an improved Laplacian smoothing technique for 3D mesh denoising. This method filters directly the vertices by updating their positions. Laplacian smoothing process is simple to implement and fast, but it tends to produce shrinking and oversmoothing effects. To remedy this problem, firstly, we introduce a kernel function in the Laplacian expression. Then, we propose to use a linear combination of denoised instances. This combination aims to reduce the number of iterations of the desired method by coupling it with a technique that leads to oversmoothing. Experiments are conducted on synthetic triangular meshes corrupted by Gaussian noise. Results show that we outperform some existing methods in terms of objective and visual quality.

1 Introduction

Denoising is one of the greatest challenges in image processing and computer graphics. Fast and efficient algorithms are needed to recover noised data (images and 3D models) while preserving their geometrical structure. Measurements are perturbed by noise in all real applications. Notably, 3D models acquisition using scanners. These scanners provide the real scanned object as a 3D digital mesh, usually represented as a triangular mesh, that can be manipulated by any 3D processing tool, for many purposes in various fields such as medical imaging, video games, etc ...

In recent years, various partial differential equations (PDE)-based techniques have been used for 2D image denoising such as the anisotropic diffusion proposed by P. Perona et al. [7]. 2D image denoising techniques were adapted for 3D mesh denoising. Taubin [8] proposed a Laplacian-based technique called Laplacian flow that repeatedly adjusts the location of each vertex to the geometric center of its vertex neighborhood. This technique is quite simple and fast but produces an oversmoothing result. Many anisotropic diffusion methods were proposed such as the diffusion and curvature flow technique presented in [9], this method gives better results than the Laplacian flow technique but takes more processing time.

Y. Zhang et al. proposed an efficient diffusion technique [1] based on solving a nonlinear discrete partial differential equation. M. El Hassouni et al. improved

this technique [2] by using other diffusion functions such as Laplacian, Reduced Centered Gaussian and Rayleigh function instead of the Cauchy function.

In this article, we propose a vertex-based method for 3D mesh denoising. The main idea is to reduce the smoothing effect by introducing a kernel function in the Laplacian flow expression. Then we present a linear combination of denoised instances in order to reduce the number iterations of the desired technique by combining it with a fast (oversmoothing) technique like the local averaging method. Experimental results show that the proposed work gives competitive results in comparison with existing methods while reducing the processing time.

This paper is organized as follows : Section 2 presents the problem formulation. Section 3 describes the Laplacian smoothing method. In section 4, we present the proposed work. Section 5 deals with experimental results. Finally, we give some concluding remarks in Section 6.

2 Problem Formulation

A 3D object is usually presented as polygonal or triangular mesh. A triangle mesh \mathbb{M} denotes a triple $\mathbb{M} = (\mathcal{V}, \mathcal{E}, \mathcal{T})$ where $\mathcal{V} = \{v_1, \dots, v_k\}$ represents the set of vertices, $\mathcal{E} = \{e_{ij}\}$ denotes the set of edges and $\mathcal{T} = \{t_1, \dots, t_n\}$ denotes the set of triangles. An edge e_{ij} consists of two vertices $\{v_i, v_j\}$ and we say that two vertices $v_i, v_j \in \mathcal{V}$ are adjacent if they are connected by an edge $e_{ij} \in \mathcal{E}$ (and we write $v_i \sim v_j$). We define the neighborhood of a vertex v_i , the set of adjacent vertices $v_i^* = \{v_j \in \mathcal{V} : v_i \sim v_j\}$. The degree of a vertex v_i is the number of the neighboring vertices $d(i) = |v_i^*|$. We denote by t_i^* the set of all triangles sharing a vertex or an edge with a triangle $t_i \in \mathcal{T}$ and by $\mathcal{T}(v_i^*)$ the set of triangles of the neighborhood v_i^* .

We denote by $n(t_j)$, with $t_j \in t_i^*$ a triangle, the unit normal of t_j and by n_i the normal at a vertex v_i given by the following formula :

$$n_i = \frac{1}{d_i} \sum_{t_j \in \mathcal{T}(v_i^*)} n(t_j) \quad (1)$$

We denote by $A(t_j)$ the area of the triangle t_j . The mean edge length \bar{l} of the mesh is given by :

$$\bar{l} = \frac{1}{|\mathcal{E}|} \sum_{e_{ij} \in \mathcal{E}} \|e_{ij}\| \quad (2)$$

Where

$$\begin{cases} \|e_{ij}\| = \|v_i - v_j\| & \text{if } v_i \sim v_j \\ \|e_{ij}\| = 0 & \text{otherwise} \end{cases}$$

Measurements are perturbed by noise (supposed additive in our case) in all real applications. This can be formulated by :

$$\hat{v} = v + \eta \quad (3)$$

Where \hat{v} is the observed vertex, v is the original one and η is a random noise process assumed to be Gaussian.

3 Laplacian Smoothing

Laplacian Smoothing is a very simple PDE-based smoothing approach formulated as follows [8]:

$$v_i \leftarrow v_i + \sum_{v_j \in v_i^*} \left(\frac{v_j - v_i}{d_i} \right) \quad (4)$$

This process can be done repeatedly to correct the location of each vertex to the geometric center of its neighboring vertices. This approach is simple and fast, however, it produces an oversmoothing result after few iterations.

Note that the *Laplacian Smoothing* is just a special case of the *Weighted Laplacian Filter* [11], also called *Local Averaging*:

$$v_i \leftarrow v_i + \frac{1}{\sum_{v_j \in v_i^*} w_{ij}} \sum_{v_j \in v_i^*} w_{ij} (v_j - v_i) \quad (5)$$

Where w_{ij} are the weights defined as :

$$w_{ij} = \begin{cases} > 0 & \text{if } v_j \in v_i^* \\ 0 & \text{otherwise} \end{cases}$$

For $w_{ij} = 1$ if $v_j \in v_i^*$ we retrieve the *Laplacian Smoothing* update rule, $\sum_{v_j \in v_i^*} w_{ij} = d_i$.

4 Proposed Method

In this section, we present at first an improved version of the Laplacian technique by introducing a kernel function. Then, we propose a linear combination of two denoised instances in order to reduce the number of iterations.

4.1 Kernel Based Laplacian Smoothing

The proposed approach consists in introducing a kernel function g to attenuate the added term, hence, overcome the oversmoothing problem.

The update rule is given by :

$$v_i \leftarrow v_i + \sum_{v_j \in v_i^*} \left(\frac{v_j - v_i}{d_i} \right) \left(\frac{g(|\nabla v_i|)}{\rho} \right) \quad (6)$$

Where

$$|\nabla v_i| = \left(\sum_{v_j \in v_i^*} \left\| \frac{v_i}{\sqrt{d_i}} - \frac{v_j}{\sqrt{d_j}} \right\|^2 \right)^{1/2} \quad (7)$$

g is a kernel function and ρ is a parameter to estimate, it can be either a scalar or a 3D vector (x,y,z). In this paper, we are going to use ρ as a scalar. Another update rule can also be used, inspired by [1]:

$$v_i \leftarrow v_i + \sum_{v_j \in v_i^*} \left(\frac{v_j - v_i}{d_i} \right) \left(\frac{g(|\nabla v_i|) + g(|\nabla v_j|)}{\rho} \right) \quad (8)$$

The same functions in [2] can be used as a kernel function. Note that for the following ρ value applied to the equation (8), we retrieve the formula presented in [1] :

$$\begin{cases} \rho = \frac{\sqrt{d_i d_j}}{d_i + \sqrt{d_i d_j} + \beta} \\ \beta = \frac{-v_j \sqrt{d_i d_j} + v_i d_i}{v_j - v_i} \end{cases}$$

The update rule (6) is the one that will be evaluated in this paper.

4.2 Linear Combination

The following method consists in combining 2 techniques using the following linear formula :

$$\widetilde{\mathbb{M}} = \alpha \widetilde{\mathbb{M}}_1 + (1 - \alpha) \widetilde{\mathbb{M}}_2, \quad \alpha \in [0, 1] \quad (9)$$

Where $\widetilde{\mathbb{M}}_1$ is the denoised mesh with the method 1 and $\widetilde{\mathbb{M}}_2$ and denoised mesh with the method 2.

This technique can be used to reduce the number of iterations of the desired technique by coupling it with a fast method that leads quickly to oversmoothing (like the Local Averaging).

The only combining approach that will be evaluated in this paper is the Local Averaging. The version used is the same one in [10]. Other combinations may give better results.

5 Experimental Results

This section presents simulation results where the proposed method is applied to 3D models contaminated by an additive zero-mean Gaussian noise.

For ease to use, we developed a simple Graphical User Interface using Matlab/Java and the toolbox graph [10]. The 3D objects used in this section are presented in Figure 1. To quantify the performance of the proposed method in comparison with other 3D mesh denoising techniques, we compute the face-normal error metric [4] given by :

$$E_{fne} = \frac{1}{A(\widetilde{\mathbb{M}})} \sum_{\hat{t}_i \in \hat{T}} A(\hat{t}_i) \| n(t_i) - n(\hat{t}_i) \|^2$$

Where $n(t_j)$ and $n(\hat{t}_i)$ are the unit normals, $A(\hat{t}_i)$ is the area of the triangle \hat{t}_i and $A(\hat{\mathbb{M}})$ is the total area of the mesh defined by the following formula :

$$A(\hat{\mathbb{M}}) = \sum_{\hat{t}_i \in \hat{\mathcal{T}}} A(\hat{t}_i)$$

We also use the visual error metric given by :

$$\begin{cases} E_{ve} = \frac{1}{2m} (\sum_{i=1}^m \|v_i - \hat{v}_i\|^2 + \sum_{i=1}^m \|I(v_i) - I(\hat{v}_i)\|^2) \\ I(v_i) = v_i - \frac{1}{d_i} \sum_{v_j \in v_i^*} v_j \end{cases}$$

This metric is computed on mesh vertices.

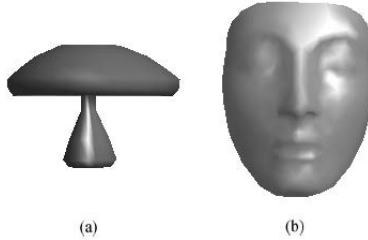


Fig. 1. 3D meshes used for experimentation : (a) mushroom (226 vertices), (b) nefertiti (299 vertices)

For all methods, parameters have been tuned experimentally in order to get the best trade-off between the visual error and the face-normal error for each technique. Also, We choose to report results only for one Kernel function and one noise level for each processed object. Here, use Laplacian as a Kernel function for all experiments with ($c = 2$).

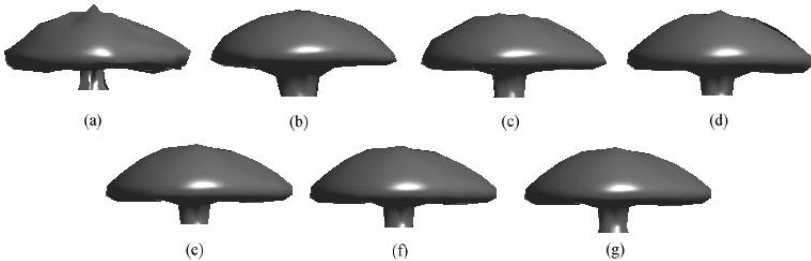


Fig. 2. Mesh denoising results for the *mushroom* instance. (a) Noisy 3D model $\sigma^2 = 0.0015$, (b) Laplacian smoothing (1 iteration), (c) Local Averaging (1 iteration), (d) Improved vertex-based diffusion (5 iterations), (e) Kernel-based Laplacian (1 iteration, $\rho = 0.65$), (f) Linear combination : Local Averaging (1 iteration) + improved vertex-based diffusion, (3 iterations, $\alpha = 0.53$), (g) Linear-combination : Local averaging (1 iteration) + kernel-based Laplacian (3 iterations, $\alpha = 0.58$, $\rho = 2$)

Figures 3 and 5 show the visual errors and Face-normal error for compared denoising methods. We remark that proposed method (kernel based Laplacian (e)) outperforms the standard Laplacian smoothing (b) and the improved vertex-based diffusion (d). Note that only one iteration is sufficient in this case, while 5 iterations were needed for the improved vertex-based diffusion. According to these graphs, linear combination gives better results for the improved vertex-based diffusion (f) while reducing the number of iterations. Combining proposed method and local averaging improves slightly the denoising performance.

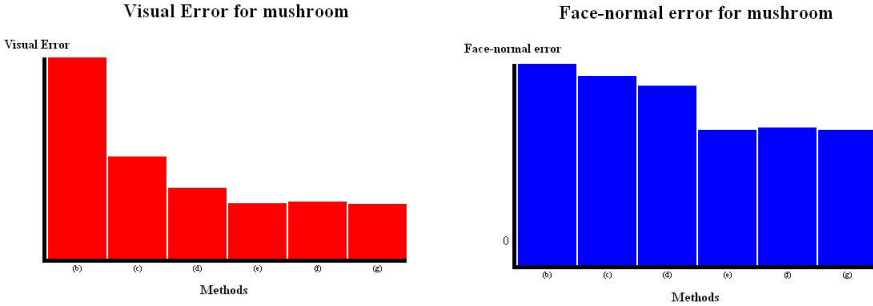


Fig. 3. Visual Error and Face-normal error for the *mushroom* object

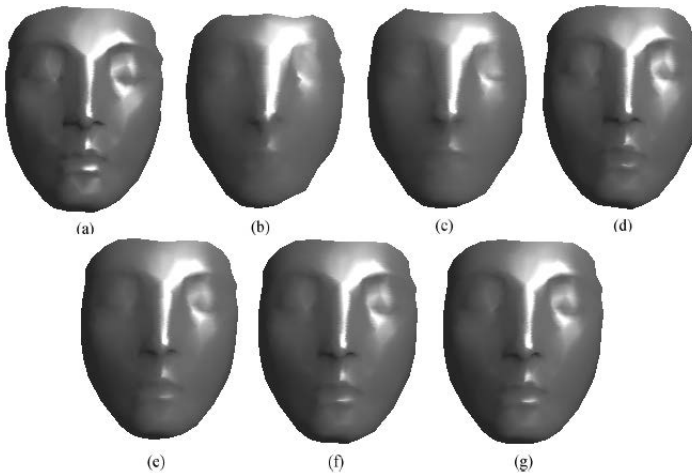


Fig. 4. Mesh denoising results for the *nefertiti* instance. (a) Noisy 3D model $\sigma^2 = 0.001$, (b) Laplacian smoothing (1 iteration), (c) Local Averaging (1 iteration), (d) Improved vertex-based diffusion (3 iterations), (e) Kernel-based Laplacian (1 iteration, $\rho = 0.72$), (f) Linear-combination : Local Averaging (1 iteration) + improved vertex-based diffusion (1 iteration, $\alpha = 0.3$), (g) Linear-combination : Local averaging (1 iteration) + kernel based Laplacian (1 iteration, $\alpha = 0.28$, $\rho = 3$)

Figure 2 and 4 are given to assess the visual impact of the denoising. We can see that for both case the Kernel Based Laplacian processed objects exhibit a more appealing visual appearance.

Note that Local Averaging using one iteration was the only technique used in the linear combination scheme. Other combinations may give better results.

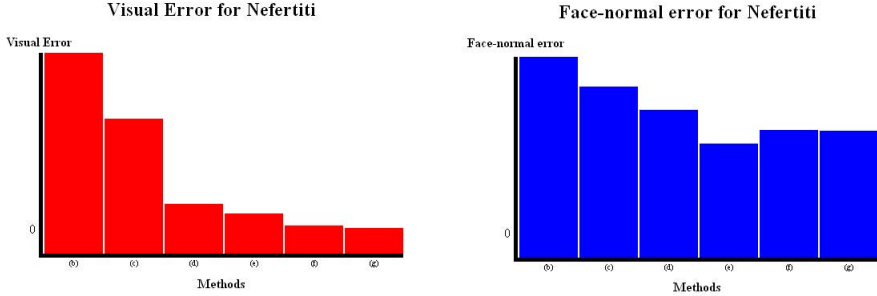


Fig. 5. Visual Error error and Face-normal error for the *nefertiti* object

6 Conclusion

We presented in this paper an kernel-based Laplacian smoothing method for 3D mesh denoising. The main idea is to reduce the smoothing effect by introducing a kernel function. Then we proposed a linear combination of denoised instances by different techniques. This method can be used to reduce the number of iterations in iterative denoising techniques of the desired method by combining it with a very fast (oversmoothing) technique like the local averaging method. Experimental results showed that the kernel-based Laplacian method proposed outperforms the standard Laplacian technique and the improved vertex-based diffusion [2] while using only 1 or 2 iterations. The linear-combination technique gave also good results while reducing the number of iterations. Using other techniques instead of the Local Averaging may give better results.

References

- [1] Zhang, Y., Ben Hamza, A.: Vertex-Based Diffusion for 3-D Mesh Denoising. IEEE Transactions on Image Processing 16(4) (April 2007)
- [2] El Hassouni, M., Aboutajdine, D.: 3D-Mesh denoising using an improved vertex based anisotropic diffusion. International Journal of Computer Science and Information Security (IJCSIS) 8(2) (2010)
- [3] Field, D.: Laplacian smoothing and Delauney triangulations. Comm. App. Num. Meth. 4, 709–712 (1988)
- [4] Taubin, G.: Linear Anisotropic Mesh Filtering. Res. Rep. RC2213 IBM (2001)
- [5] Karni, Z., Gotsman, C.: Spectral compression of mesh geometry. In: Proc. SIGGRAPH, pp. 279–286 (2000)

- [6] Mashiko, T., Yagou, H., Wei, D., Ding, Y., Wu, G.: 3D Triangle Mesh Smoothing via Adaptive MMSE Filtering. In: Proceedings of the Fourth International Conference on Computer and Information Technology (CIT 2004), pp. 734–740 (2004)
- [7] Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7), 629–639 (1990)
- [8] Taubin, G.: A signal processing approach to fair surface design. In: Proc. SIGGRAPH, pp. 351–358 (1995)
- [9] Desbrun, M., Meyer, M., Schroder, P., Barr, A.: Implicit fairing of irregular meshes using diffusion and curvature flow. In: Proc. SIGGRAPH, pp. 317–324 (1999)
- [10] http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/meshproc_3_denoising/
- [11] Desbrun, M., Meyer, M., Schröder, P., Barr, A.: Implicit fairing of irregular meshes using diffusion and curvature flow. In: Proc. SIGGRAPH, pp. 317–324 (1999)

Embedded Real-Time Video Processing System on FPGA

Yahia Said¹, Taoufik Saidani¹, Fethi Smach², Mohamed Atri¹, and Hichem Snoussi³

¹ Laboratory of Electronics and Microelectronics (EμE),
Faculty of Sciences Monastir, 5000 Tunisia
said.yahia1@gmail.com, saidani_toufik@yahoo.fr,
mohamed.atri@fsm.rnu.tn

² Active Networks
1 rue de Terre Neuve, BP 127 - 91944 Courtaboeuf Cedex - France
smach_fethi@yahoo.fr

³ Université de technologie de Troyes
Institut Charles Delaunay ICD, UMR STMR 6279
BP 2060 - 10010 TROYES Cedex
hichem.snoussi@utt.fr

Abstract. Image Processing algorithms implemented in hardware have emerged as the most viable solution for improving the performance of image processing systems. The introduction of reconfigurable devices and high level hardware programming languages has further accelerated the design of image processing in FPGA.

This paper briefly presents the design of Sobel edge detector system on FPGA. The design is developed in System Generator and integrated as a dedicated hardware peripheral to the Microblaze 32 bit soft RISC processor with the EDK embedded system. The input comes from a live video acquired from a CMOS camera and the detected edges are displayed on a DVI display screen.

Keywords: Sobel Edge detector, Real Time, Microblaze Processor, Field Programmable Gate Arrays (FPGA), Embedded Development Kit (EDK), System Generator (SysGen).

1 Introduction

Computationally Intensive DSP applications such as Image Processing is getting widely used in embedded systems for many applications, such as object detection, space exploration, security or video surveillance.

Reconfigurable hardware in the form of Field Programmable Gate Arrays (FPGAs) has been proposed as a way of obtaining high performance for Image Processing, even under real time requirements [1]. Implementing image processing algorithms on reconfigurable hardware minimizes the time-to-market cost, enables rapid prototyping of complex algorithms and simplifies debugging and verification. Therefore, FPGAs are an ideal choice for implementation of real time image processing algorithms [2].

Edge detection is a fundamental tool used in most image processing applications to obtain information from the frames as a precursor step to feature extraction and object segmentation. This process detects outlines of an object and boundaries between

objects and the background in the image. An edge-detection filter can also be used to improve the appearance of blurred or anti-aliased video streams [3].

FPGAs offer many performance benefits for executing image processing applications. The FPGA design can also reduce the system costs with various verification techniques such as behavioral simulation and post-route simulation. Moreover, Xilinx Embedded Development Kit (EDK) tools make it possible to implement a complete video processing system on a single FPGA using hardware/software codesign methods.

The objective of this work is to develop a real-time edge detection system with an input from a CMOS camera and output to a DVI display and verified the results video in real time.

This paper is organized as follows: Section 2 describes system architecture and functions of each block. Section 3 covers the architecture for edge detector core developed in System Generator. In section 4, experimental results of the proposed system are shown. Finally, a brief conclusion and directions for future work are given in Section 5.

2 System Architecture

The setup for implementation consists of the Spartan-3A DSP FPGA Video Starter Kit (VSK), a development platform consisting of the Spartan-3A DSP 3400A FPGA, the FMC-Video daughter card, and a VGA camera (“Fig. 1”).

The Spartan-3A DSP 3400A Development Platform is built around a Spartan-3A DSP XC3SD3400A device that provides significant resources (for example, 126 embedded DSP blocks) for implementing high performance video processing systems and co-processors.

The VSK includes a VGA camera based upon the Micron MT9V022 image sensor of resolution 720 x 480 pixels delivering serial frames at 60 fps through a FPGA Mezzanine Card (FMC) Daughter card which is an add-on card that augments the video capabilities of the Spartan-3A DSP 3400A Development Platform [4].

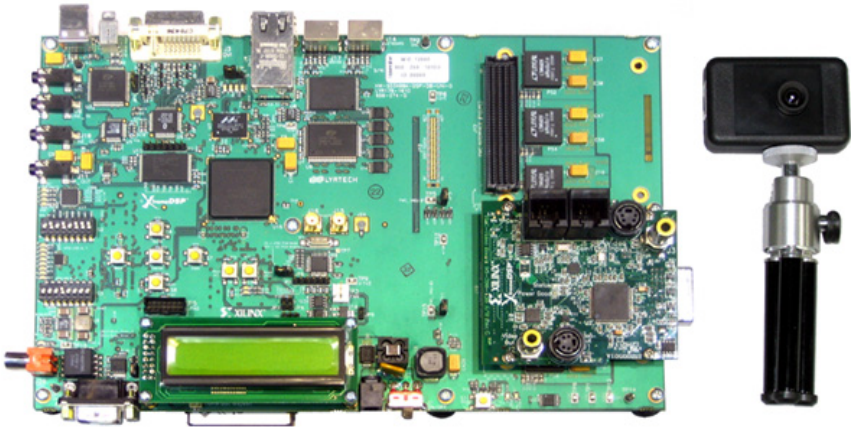


Fig. 1. Spartan-3A DSP 3400 Development Platform, FMC-Video, and Camera

Figure 2 shows a detailed system diagram of the implemented video filtering design. The complete streaming video application includes Video interfaces, a run-time configurable processing blocks, a real-time edge detection filter, and a MicroBlaze embedded processor for embedded control of the video subsystem.

The video processing application is designed as a system on a programmable chip with the help of Embedded Design Kit. The serial video is de-serialized on the FMC-Video card. The resulting parallel data stream is the input to the Camera In block. The Camera PCORE registers the signals, and groups the video signals into a unified bus that is connected to the Camera Processing block, which is included in the camera frame buffer reference designs shipped with the VSK [4], to control brightness, contrast and other parameters. The edge filter is applied on the input signal arriving from the Camera Processing block. The output signal is Gamma corrected for the output DVI monitor and is driven by Display controller to the DVI output monitor. The Video to VFBC core manages the storing of video into frame buffers in external memory. It writes the video data to the VFBC interface on the MPMC memory controller.

The Display Controller core reads video frames out of memory from a VFBC interface of MPMC and displays them to the output screen by applying the correct timing signals. The edge filter core developed in System Generator for DSP will be detailed in the next section.

The video pipeline demonstrated by our design is created using the Xilinx Embedded Development Kit (EDK) [5] and System Generator for DSP [6]. The Embedded Development Kit is a collection of Intellectual Property (IP) cores and tools for building FPGA-based embedded systems. System Generator for DSP enables the use of the Simulink/MATLAB modeling environment for FPGA design by providing a Simulink blockset of over 100 Xilinx optimized DSP building blocks.

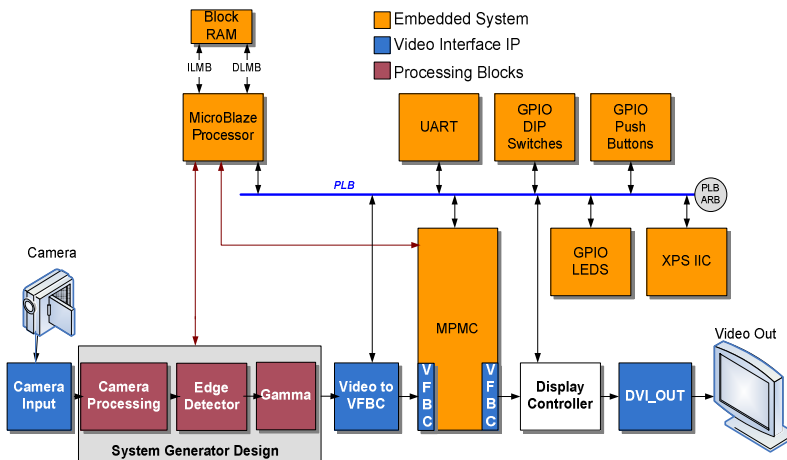


Fig. 2. Video Pipeline with MicroBlaze processor and peripheral

The architecture consists of a set of modules interconnected with buses, as seen in “Fig. 2”. The Camera Processing, Gamma and Edge Detection cores are connected to

the Embedded MicroBlaze processor through Processor Local Bus (PLB). The Processor is connected to dual-port SRAM, called Block RAM (BRAM), using a dedicated Local Memory Bus (LMB). This bus features separate 32-bit wide channels for program instructions and program data, using the dual-port feature of the BRAM. The LMB provides single-cycle access to on-chip dual-port Block RAM.

The MicroBlaze soft processor core [7] provided by Xilinx is central in the system and used as an embedded video controller. It is a reduced instruction set computer (RISC) optimized for implementation in the Xilinx Field Programmable Gate Arrays (FPGAs). Figure 3 shows the block diagram of MicroBlaze.

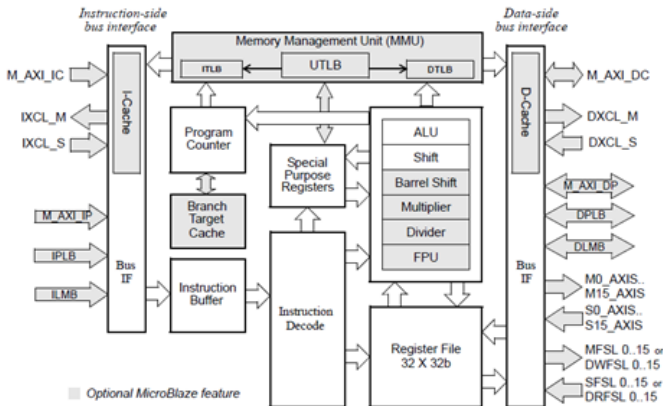


Fig. 3. MicroBlaze Core Block Diagram

3 Sobel Edge Detector

Edge detection is the process of localizing pixel intensity transitions. The edge detection has been used by segmentation, motion analysis, object recognition, target tracking, and many more [8]. Therefore, the edge detection is one of the significant techniques in the field of image processing.

The most well known technique for edge detection is gradient-based. The gradient method looks the edges by finding maximum and minimum in the first derivative of the image. Sobel is gradient based edge detection algorithm which performs a 2-D spatial gradient measurement on the video data. It uses a pair of 3X3 convolution masks, one estimating gradient in x-direction and other in y-direction. Then the value of the gradient magnitude is computed from the above two gradients.

First, RGB data are converted into grayscale to obtain image intensity, using the following equation:

$$I = (0.2989 \times R + 0.5870 \times V + 0.1140 \times B) \tag{1}$$

Then horizontal and vertical gradient are calculated as shown in “(2)”.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I \quad \text{and} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I \quad (2)$$

The magnitude and orientation are obtained as follow:

$$G = |G_x| + |G_y| \quad \text{and} \quad \theta = \text{Arctan} \left(\frac{G_y}{G_x} \right) \quad (3)$$

We build the sobel edge detector as a video processing accelerator, using System Generator for DSP and Simulink. The design of our filter is shown in “Fig. 4”.

System Generator supports hardware in-the-loop co-simulation using the Spartan-3A DSP 3400A development platform, which can accelerate the performance of Simulink simulations up to 100x. This acceleration enables video algorithm development and debug using real-time video streams read into Simulink using The Mathworks’ Data Acquisition Toolbox [9].

System Generator for DSP can automatically generate accelerator blocks in the form of a custom peripheral for the embedded video application that allows the MicroBlaze processor to read and write shared memories in the accelerator block. This includes an automatically generated hardware interface for the PLB bus, a software C driver file, and software documentation for using the DSP co-processor.

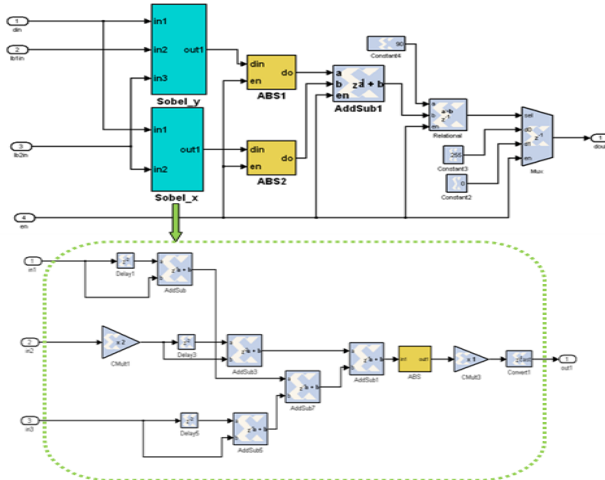


Fig. 4. Sobel architecture with system generator

The System Generator design contains an EDK Processor block that can be exported as an EDK pcore using the EDK Export Tool compilation target. The export process creates a PLB-based pcore, which is integrated to the Microblaze 32 bit soft RISC processor with the Xilinx Platform Studio (XPS) [6].

4 Experimental Results

In the system setup a DVI display shows the output edge from the camera. Experimental setup for implementation of sobel edge detection is presented in “Fig.5”.

The total resource usage for the system, including the MicroBlaze, bus structure, the sobel edge core and peripherals, is 9094 slices, equaling 38% of the Spartan 3A DSP 3400. The system was implemented to run at 62.5MHz. It is possible that higher frequencies are attainable, up to a limit of around 125MHz. The MicroBlaze has a maximum frequency of 125MHz on the Spartan 3A DSP 3400, and the sobel core has a post-synthesis maximum estimate of 68.432MHz.

Table 1 shows the amount of logic used for the sobel edge module. A maximum of 5% of the FPGA’s total resources are used by this module. The post-synthesis resource usage of the MicroBlaze processor is 1531 slices. The synthesis results of the overall system are given in Table 2.

From the synthesis results of our system, we can see that few resources of the FPGA are used; hence space is available for other complex image and video processing applications.

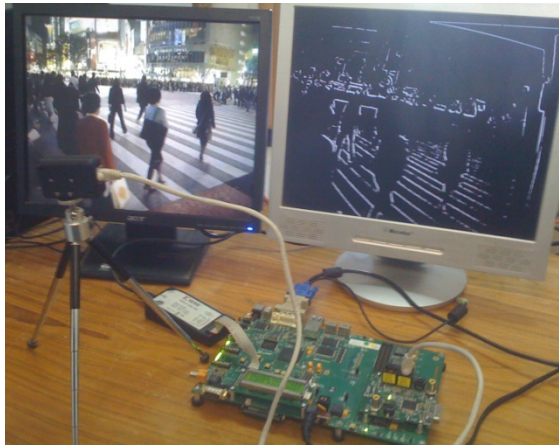


Fig. 5. Experimental setup for implementation of edge detection. Input is from CMOS camera and the output is on a DVI display.

Table 1. Post-Synthesis device utilization for the Sobel edge module implemented on the Spartan 3A DSP 3400

Resource Type	Used	Available	%
Slices	1284	23872	5%
Slice Flip Flops	1745	47744	3%
4 input LUTs	1713	47744	3%
bonded IOBs	0	469	0%
BRAMs	5	126	3%
DSP48s	4	126	3%
Maximum Frequency		68.432 MHz	

Table 2. The synthesis results of the overall system

Resource Type	Used	Available	%
Slices	9094	23872	38%
Slice Flip Flops	11451	47744	24%
4 input LUTs	12883	47744	27%
bonded IOBs	78	469	17%
BRAMs	69	126	55%
DSP48s	7	126	6%
Maximum Frequency	88.547MHz		

5 Conclusion

Continual growth in the size and functionality of FPGAs over recent years has resulted in an increasing interest in their use as implementation platforms for image processing applications, particularly real-time video processing [10].

In this paper, we propose a design for real-time video processing system on a Spartan 3A DSP FPGA. Sobel edge detector was implemented at a rate of 60 fps for an input image of resolution 720x480.

The implemented system architecture has 88.547MHz maximum frequency and uses 9094 CLB slices with 38% utilization, so there is possibility of implementing some more parallel processes with this architecture on the same FPGA.

System Generator for DSP enables the use of Simulink for Xilinx FPGA designs by providing a rich set of DSP building blocks, optimized for Xilinx devices. DSP designs captured in System Generator can be converted into custom peripherals for Platform Studio and connected to the embedded system using the processor local bus.

Future works include the use of the Xilinx System Generator and EDK development tools for the implementation of other blocks used in computer vision like feature extraction and object detection on Xilinx Programmable Gate Arrays (FPGA).

References

1. Crookes, D.: Design and implementation of a high level programming environment for FPGA-based image processing. *IEE Proceedings on Vision, Image, and Signal Processing* 147(4), 377 (2000)
2. Rao, D.V., Patil, S., Babu, N.A., Muthukuma, V.: Implementation and Evaluation of Image Processing Algorithms on Reconfigurable Architecture using C-based Hardware Descriptive Languages. *International Journal of Theoretical and Applied Computer Sciences* 1(1), 9–34 (2006)
3. Neoh, H., Hazanchuk, A.: Adaptive Edge Detection for Real-Time Video Processing using FPGAs. *Global Signal Processing* (2004)
4. Spartan-3A DSP FPGA Video Starter Kit user Guide, <http://www.xilinx.com>
5. Xilinx Inc. Embedded System Tools Reference Manual, <http://www.xilinx.com>

6. Xilinx System Generator user Guide, <http://www.xilinx.com>
7. MicroBlaze soft processor, <http://www.xilinx.com>
8. Senal, H.G.: Gradient Estimation Using Wide Support Operators. *IEEE Transaction on Image Processing* 18(4) (April 2009)
9. Mallet, J.: Updated Starter Kit Speeds Video Development. *Xell Journal* (67), 18–21 (2009)
10. Hutchings, B., Villasenor, J.: The Flexibility of Configurable Computing. *IEEE Signal Processing Magazine* 15, 67–84 (1998)

Edge Preserving Image Fusion Based on Contourlet Transform

Ashish Khare, Richa Srivastava, and Rajiv Singh

Department of Electronics & Communication,
University of Allahabad, Allahabad, India

ashishkhare@hotmail.com, {gaur.richa,jkrajivsingh}@gmail.com

Abstract. Image fusion is an emerging area of research having a number of applications in medical imaging, remote sensing, satellite imaging, target tracking, concealed weapon detection and biometrics. In the present work, we have proposed a new edge preserving image fusion method based on contourlet transform. As contourlet transform has high directionality and anisotropy, it gives better image representation than wavelet transforms. Also contourlet transform represents salient features of images such as edges, curves and contours in better way. So it is well suited for image fusion. We have performed experiments on several image data sets and results are shown for two datasets of multifocus images and one dataset of medical images. On the basis of experimental results, it was found that performance of proposed fusion method is better than wavelet transform (Discrete wavelet transform and Stationary wavelet transform) based image fusion methods. We have verified the goodness of the proposed fusion algorithm by well known image fusion measures (entropy, standard deviation, mutual information (MI) and Q_{AB}^F). The fusion evaluation parameters also imply that the proposed edge preserving image fusion method is better than wavelet transform (Discrete wavelet transform and Stationary wavelet transform) based image fusion methods.

Keywords: Image fusion, Contourlet transform, Wavelet transform, Edge preserving image fusion, Laplacian and directional filter banks.

1 Introduction

Image fusion [1, 2] is a technique of computer vision that integrates all relevant and complementary information from different source images into a single composite image without introducing any artifact or noise. The objective of image fusion [3] is to combine information from multiple source images which contains the best relevant information coming from source images and is more suitable for human or machine perception.

Image fusion is an interesting area of research having a number of applications in medical imaging [4], remote sensing [5], satellite imaging [6], target tracking [7], concealed weapon detection [8] and biometrics [9]. Most of the imaging sensors are capable to capture a single type of image that provides only a specific kind of

information. For example, in remote sensing, we have multispectral image with low spatial resolution and high spectral resolution and panchromatic image with high spatial resolution and low spectral resolution. Similarly in medical imaging, Positron Emission Tomography (PET) image provides functional information whereas the Magnetic Resonance Imaging (MRI) image gives anatomical information. These examples explain that a single imaging sensor is not able to provide all relevant information. Hence fusion of different source images is required in order to retrieve desired information in a single image.

Generally image fusion can be classified into three categories [10]: pixel level fusion, feature level fusion and decision level fusion. In the present work, we propose a pixel image fusion technique, as pixel level image fusion is computationally efficient and original quality of images are not lost during fusion process.

Image fusion techniques vary from spatial domain to transform domain. Limitations of spatial domain fusion techniques [11] lead to transform domain fusion techniques. Wavelet transform based fusion methods [3, 12, 13, 14] are widely found in literatures. Discrete Wavelet Transform (DWT) [15, 16] and Stationary Wavelet Transform (SWT) [17] are examples of wavelet transform based image fusion methods. But wavelet transform based fusion methods are limited as they can isolate the discontinuities at edges but failed to capture the smoothness along the contours and curves. Also wavelet transform has limited directionality (e.g. two dimensional DWT provides directional information in three orientations: horizontal, vertical and diagonal), that leads to loss of directional information in resulting fused image.

Due to these limitations of wavelet transform, we have used Contourlet Transform (CT) [18-23] for image fusion which provides more directional information and smoothness about contours. In the present work, we have proposed an edge preserving image fusion method using contourlet transform and compared the proposed method with wavelet transform based fusion techniques (DWT and SWT) using different fusion evaluation parameters.

The rest of the paper is organized as follows: Section 2 briefly describes the contourlet transform. In section 3, the proposed fusion method is elaborated. Results and discussions are given in section 4. Finally, conclusions of the work are given in section 5.

2 The Contourlet Transform

The Contourlet Transform [18-23] has a rich set of basis functions and it can represent salient features of images such as edges, curves and contours efficiently. It satisfies the basic requirement for image representation and provides a multiresolution representation of images with increased directionality and anisotropy. The construction of contourlet transform follows double filter banks approach, the laplacian pyramid (LP) and the directional filter banks (DFB). These filter banks jointly known as pyramidal directional filter banks (PDFB).

The contourlet transform is performed in two steps. First is subband decomposition and second is directional transform. The laplacian pyramid is used to capture the point discontinuities and the directional filter bank is used to link the point discontinuities

into linear structures. Thus the filter banks used in contourlet transform gives a detailed high directional image representation and captures intrinsic geometric structures as well. Hence it is well suited for image fusion with increased directionality and smooth representations along edges, curves and contours. The contourlet transform framework is shown in figure 1.

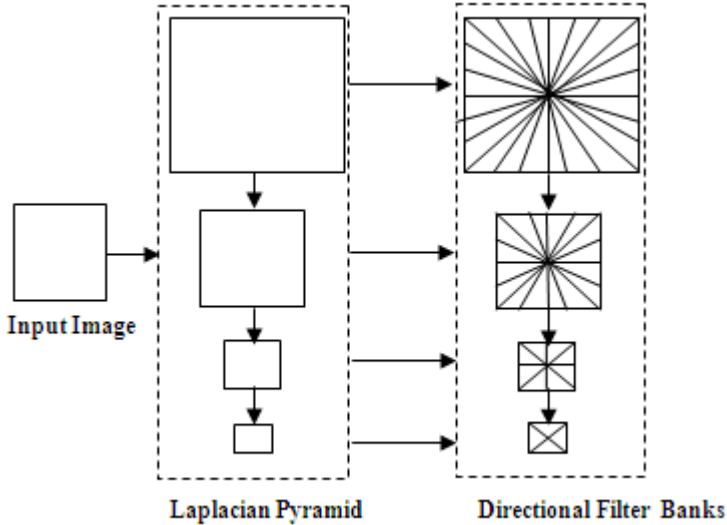


Fig. 1. The Contourlet Transform Framework

3 The Proposed Method

We know that edges and boundaries are salient features of images and required for better visual representation of images. Hence, we have proposed a new edge preserving image fusion method using contourlet transform. As contourlet transform has high degree of directionality and captures geometric structures in better way, it leads to a better detailed representation of images. The proposed method uses contourlet transform with edge preservation for fusion of different source images. In the proposed fusion method, first, the source images are decomposed using contourlet transform to obtain contourlet coefficients, followed by edge point computation using sobel gradient operator. Secondly, we preserve contourlet coefficients that correspond to edge points in the source images. Then we apply maximum fusion rule on remaining contourlet coefficients to obtain coefficients for fused image. Reconstruction of the obtained contourlet coefficients will result in fused image.

The steps of the proposed fusion algorithm can be summarized as follows:

1. Decompose source images $I_1(x, y)$ and $I_2(x, y)$ using Contourlet Transform (CT) to obtain contourlet coefficients $C_1(x, y)$ and $C_2(x, y)$ respectively.

$$C_1(x, y) = CT [I_1(x, y)]$$

$$\text{and } C_2(x, y) = CT [I_2(x, y)] \quad (1)$$

2. Use sobel operator to compute edge points $E_1(x, y)$ and $E_2(x, y)$ of contourlet coefficients $C_1(x, y)$ and $C_2(x, y)$ for source images $I_1(x, y)$ and $I_2(x, y)$ respectively.
3. Preserve edge points in the fused coefficient set $C(x, y)$ as following:

$$C(x, y) = \begin{cases} C_1(x, y), & \text{if } E_1(x, y)=1 \\ C_2(x, y), & \text{if } E_2(x, y)=1 \end{cases} \quad (2)$$

4. Apply maximum fusion rule on remaining contourlet coefficients to obtain coefficients for fused image i.e.

$$C(x, y) = \begin{cases} C_1(x, y), & \text{if } |C_1(x, y)| > |C_2(x, y)| \\ C_2(x, y), & \text{if } |C_2(x, y)| > |C_1(x, y)| \end{cases} \quad (3)$$

5. Reconstruction of $C(x, y)$ provides fused image $F(x, y)$.

$$F(x, y) = \text{Inverse CT } [C(x, y)] \quad (4)$$

4 Results and Discussions

This section gives visual and quantitative comparison of the proposed method with DWT and SWT based fusion techniques. We have performed the proposed fusion method over several image data sets and fusion results for representative multifocus and medical image datasets are shown in figure 2, figure 3 and figure 4. The proposed edge preserving fusion method is compared with DWT [15, 16] and SWT [17] based fusion techniques in which fusion is performed using average, average-maximum and maximum fusion rules. The proposed method preserves contourlet coefficients that correspond to edge points in the source images. This edge preserving step in the proposed method provides better edge information in the fused image that can be viewed from the visual representation of results shown in the figure 2, figure 3 and figure 4.

For objective evaluation of the proposed fusion method, we have used well known fusion measures [24, 25] - entropy, standard deviation, mutual information (MI) and Q_{AB}^F . Higher values of these fusion metrics will imply better fused image. To compare the proposed fusion method with DWT and SWT based fusion methods, we have computed above mentioned fusion measures and tabulated them in table 1, table 2 and table 3 for multifocus and medical image data sets. By observing table 1, one can easily found that fused image obtained by the proposed fusion method has higher values for three fusion metrics (entropy, mutual information and Q_{AB}^F) and very closed (62.1098) to maximum (62.4296) for standard deviation metric. By observing table 2, again we have highest values for entropy (7.1266) and mutual information (0.5643). Similarly in table 3, the proposed method has the highest values for three metrics

(entropy, mutual information and Q_{AB}^f) and closer to standard deviation metric for maximum fusion case. These comparisons show that the proposed edge preserving fusion method has better visual representation than DWT and SWT based fusion techniques as well as has higher values for different fusion metrics. Therefore, it can be concluded that the proposed edge preserving fusion using Contourlet Transform is able to preserve detail features of source images into resulting fused image and performs better than DWT and SWT based fusion methods.



Fig. 2. Fusion results for multifocus images. (a). Source image 1, (b). Source image 2, (c). Fused image by the proposed fusion method, (d). DWT maximum fused image, (e). DWT average fused image, (f). DWT average-maximum fused image, (g). SWT maximum fused image, (h). SWT average fused image and (i). SWT average-maximum fused image.

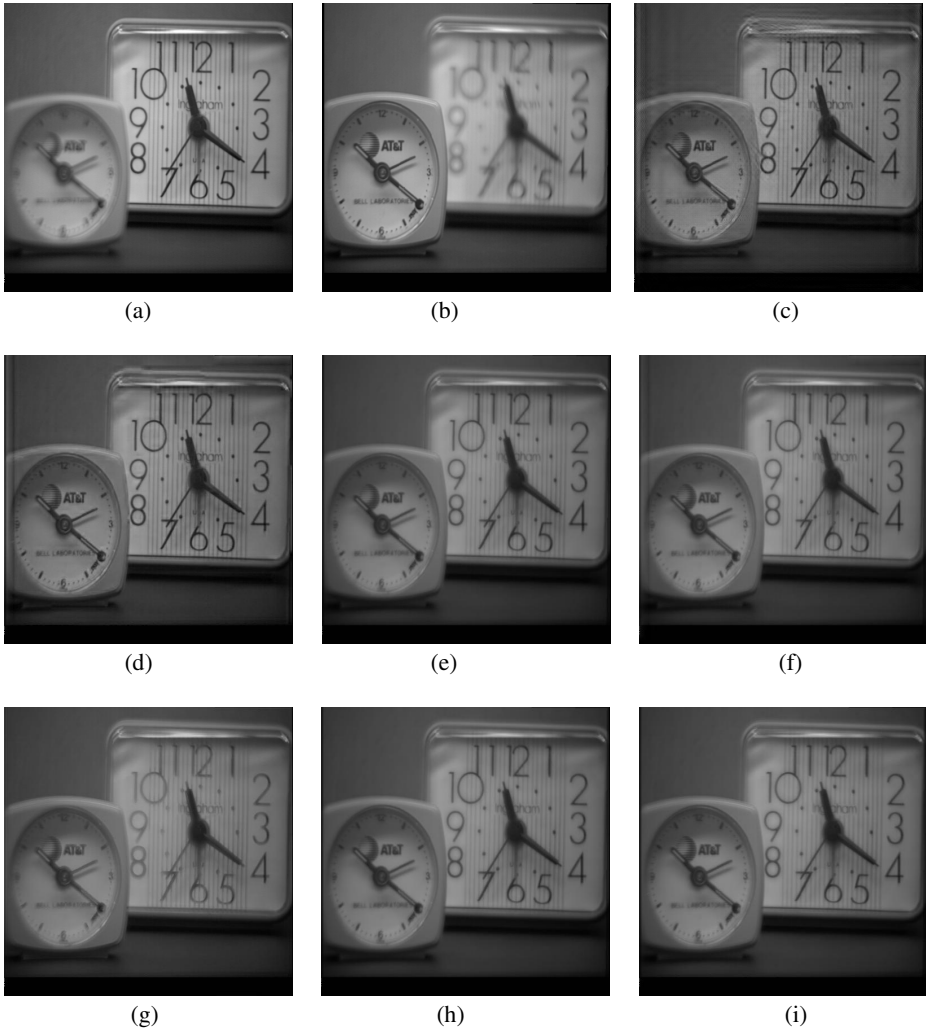


Fig. 3. Fusion results for Clock images. (a). Source image 1, (b). Source image 2, (c). Fused image by the proposed fusion method, (d). DWT maximum fused image, (e). DWT average fused image, (f). DWT average-maximum fused image, (g). SWT maximum fused image, (h). SWT average fused image and (i). SWT average-maximum fused image.

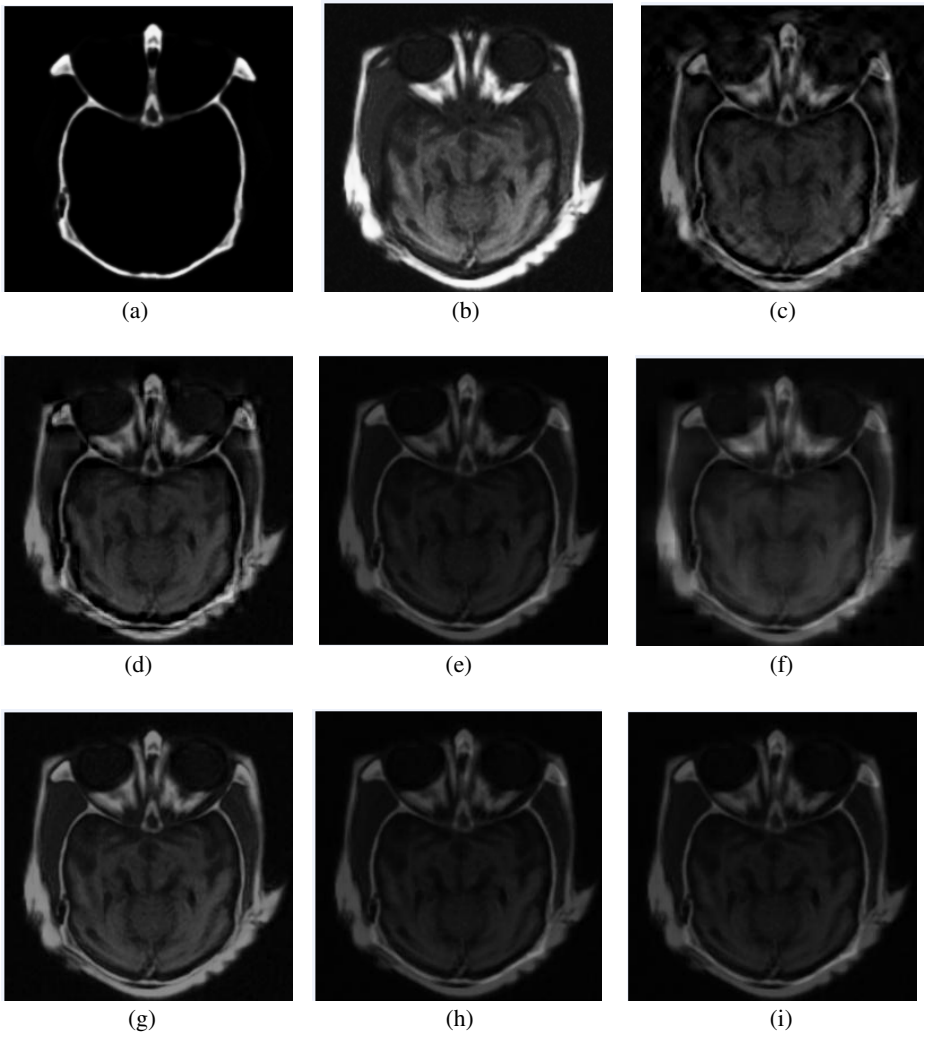


Fig. 4. Fusion results for medical images. (a). Source image 1, (b). Source image 2, (c). Fused image by the proposed fusion method, (d). DWT maximum fused image, (e). DWT average fused image, (f). DWT average-maximum fused image, (g). SWT maximum fused image, (h). SWT average fused image and (i). SWT average-maximum fused image.

Table 1. Fusion performance evaluation for Multifocus Image Data Set

Fusion Method	Entropy	Standard Deviation	MI	Q_{AB}^F
CT(EDGE)	7.7707	62.1098	4.5222	0.6697
DWT(MAX)	7.7688	62.4296	4.2728	0.6307
DWT(AVG)	7.6059	56.4075	4.5725	0.3330
DWT(AVGMAX)	7.6709	57.9649	4.5424	0.3385
SWT(MAX)	7.7701	60.9528	4.4907	0.4744
SWT(AVG)	7.6090	56.4817	4.5786	0.3302
SWT(AVGMAX)	7.6873	58.0404	4.3518	0.5000

Table 2. Fusion performance evaluation for Clock Image Data Set

Fusion Method	Entropy	Standard Deviation	MI	Q_{AB}^F
CT(EDGE)	7.1266	40.5562	5.5334	0.5643
DWT(MAX)	7.0328	40.6854	5.8496	0.5612
DWT(AVG)	6.9687	39.6026	6.5894	0.4980
DWT(AVGMAX)	6.9794	39.5327	6.2704	0.4979
SWT(MAX)	6.9967	40.3693	6.9971	0.4360
SWT(AVG)	6.9689	39.6728	6.5973	0.4946
SWT(AVGMAX)	6.9904	39.6661	6.5528	0.4899

Table 3. Fusion performance evaluation for Medical Image Data Set

Fusion Method	Entropy	Standard Deviation	MI	Q_{AB}^F
CT(EDGE)	6.3326	32.0629	3.6420	0.7787
DWT(MAX)	6.1270	27.3918	2.0188	0.3617
DWT(AVG)	5.1009	18.6230	2.9370	0.3325
DWT(AVGMAX)	6.1270	27.3918	2.0188	0.3617
SWT(MAX)	5.9902	32.9014	3.5055	0.6805
SWT(AVG)	5.1164	18.6744	2.9766	0.3292
SWT(AVGMAX)	5.1912	18.8342	2.9618	0.3737

5 Conclusions

In the present work, we have proposed a new edge preserving fusion method using Contourlet Transform. As Contourlet Transform has higher directionality and it captures smooth contours. Contourlet Transform based fusion using edge preservation is capable to preserve edge values in more efficient way. Experiments are performed over three set of image data sets (multifocus and medical). Visual representation of

experimental results indicate that the proposed fusion method is better than wavelet transform (DWT and SWT) based fusion methods and provides better visual representation of fused image. Also we have verified the goodness of the proposed fusion method with well known fusion measures (entropy, standard deviation, mutual information (MI) and Q_{AB}^F). This comparison also proved that the proposed edge preserving fusion method using Contourlet Transform is better than wavelet transform (DWT and SWT) based fusion methods.

Acknowledgements. This work was supported in part by the Department of Science and Technology, New Delhi, India, under grant no. SR/FTP/ETA-023/2009 and the University Grants Commission, New Delhi, India, under grant no. 36-246/2008(SR).

References

1. Goshtasby, A., Nikolov, S.G.: Image Fusion: Advances in the state of the art, Guest editorial. *Information Fusion* 8(2), 114–118 (2007)
2. Toet, A., Hogervorst, M.A., Nikolov, S.G., Lewis, J.J., Dixon, T.D., Bull, D.R., Canagarajah, C.N.: Towards cognitive image fusion. *Information Fusion* 11(2), 95–113 (2010)
3. Pajares, G., Cruz, J.M.: A wavelet-based image fusion tutorial. *Pattern Recognition* 37(9), 1855–1872 (2004)
4. Darasthy, B.V.: Information fusion in the realm of medical applications – A bibliographic glimpse at its growing appeal. *Information Fusion* 13(1), 1–9 (2012)
5. Simone, G., Farina, A., Morabito, F.C., Serpico, S.B., Bruzzone, L.: Image fusion techniques for remote sensing applications. *Information Fusion* 3(1), 3–15 (2002)
6. Ranchin, T., Aiazzi, B., Alparone, L., Baronti, S., Wald, L.: Image fusion- the ARSIS concepts and some successful implementation schemes. *ISPRS Journal of Photogrammetry & Remote Sensing* 58(1-2), 4–18 (2003)
7. Janczak, D., Sankowski, M.: Data fusion for ballistic targets tracking using least squares. *AEU- International Journal of Electronics and Communications* (2011) (article in press), <http://dx.doi.org/10.1016/j.aeue.2011.11.003>
8. Xue, Z., Blum, R.S., Li, Y.: Fusion of Visual and IR Images for Concealed Weapon Detection. In: *Proceedings of International Conference on Image Fusion (ISIF)*, vol. 2, pp. 1198–1205 (2002)
9. Ross, A., Jain, A.: Information Fusion in Biometrics. *Pattern Recognition Letters* 24(13), 2115–2125 (2003)
10. Pohl, C., Genderen, J.L.V.: Multisensor image fusion in remote sensing: concept, methods and applications. *International Journal of Remote Sensing* 19(5), 823–854 (1998)
11. Li, H., Manjunath, B.S., Mitra, S.K.: Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing* 57(3), 235–245 (1995)
12. Amolins, K., Zhang, Y., Dare, P.: Wavelet based image fusion techniques—an introduction, review and comparison. *ISPRS Journal of Photogrammetry & Remote Sensing* 62(4), 249–263 (2007)
13. Singh, R., Srivastava, R., Prakash, O., Khare, A.: DTCWT based Multimodal Medical Image Fusion. In: *Proceedings of International Conference on Signal, Image and Video Processing (ICSIVP 2012)*, IIT Patna, Patna, pp. 403–407 (2012)

14. Singh, R., Srivastava, R., Prakash, O., Khare, A.: Mixed scheme based multimodal medical image fusion using Daubechies Complex Wavelet Transform. Accepted to appear in Proceedings of International Conference on Informatics, Electronics & Vision (IEEE/IAPR ICIEV 2012), Dhaka, Bangladesh, May 18-19 (2012)
15. Shangli, C., Junmin, H.E., Zhongwei, L.: Medical Images of PET/CT Weighted Fusion Based on Wavelet Transform. *Bioinformatics and Biomedical Engineering*, 2523–2525 (2008)
16. Singh, R., Khare, A.: A Wavelet Based Multimodal Medical Image Fusion. In: Proceedings of International Symposium on Medical Imaging: Perspectives on Perception and Diagnostics, in Conjunction with Seventh Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP-2010), IIT- Delhi, New Delhi, December 9-10 (2010)
17. Singh, R., Vatsa, M., Noore, A.: Multimodal Medical Medical Image Fusion using Redundant Wavelet Transform. In: Proc. of Seventh International Conference on Advances in Pattern Recognition, pp. 232–235 (2009)
18. Do, M.N., Vetterli, M.: The Contourlet Transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing* 14(12), 2091–2106 (2005)
19. Do, M.N., Vetterli, M.: Contourlets: a directional multiresolution image representation. In: Proceedings of International Conference of Image Processing, pp. 357–360 (2002)
20. Do, M.N., Vetterli, M.: Contourlets. In: Stoeckler, J., Welland, G.V. (eds.) *Beyond Wavelets*, pp. 1–27. Academic Press, New York (2002)
21. Tang, L., Zhao, Z.: The Wavelet-based Contourlet Transform for Image Fusion. In: Proceedings of Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp. 59–64 (2007)
22. Asmare, M.H., Asirvadam, V.S., Iznita, L., Hani, A.F.M.: Image Enhancement by Fusion in Contourlet Transform. *International Journal on Electrical Engineering and Informatics* 2(1), 29–42 (2011)
23. Yang, L., Guo, B., Ni, W.: Multifocus Image Fusion Algorithm based on Contourlet Transform and Region Statistics. In: Fourth International Conference on Image and Graphics (IJIG), pp. 707–712 (2007)
24. Kotwal, K., Chaudhuri, S.: A novel approach to quantitative evaluation of hyperspectral image fusion techniques. *Information Fusion* (2011) (article in press), <http://dx.doi.org/10.1016/j.inffus.2011.03.008>
25. Xydeas, S., Petrovic, V.: Objective Image Fusion Performance Measure. *Electronics Letters* 36(4), 308–309 (2000)

Selecting Vision Operators and Fixing Their Optimal Parameters Values Using Reinforcement Learning

Issam Qaffou, Mohamed Sadgal, and Aziz Elfazziki

Département Informatique, FSSM,
Université Cadi Ayyad, Marrakech, Morocco
{i.qaffou, sadgal, elfazziki}@ucam.ac.ma

Abstract. Selecting the appropriate operators with the optimal values for their parameters represents a big challenge for users. In this paper we present a solution for this problem. This solution uses a multi-agent architecture based on reinforcement learning to automate the process of operator selection and parameter adjustment. The architecture consists of three types of agents: the User Agent, the Operator Agent and the Parameter Agent. The User Agent determines the phases of treatment, and for each phase it determines a library of possible operators and possible values of their parameters. The Operator Agent constructs all possible combinations of operators and decides for the best one. The Parameter Agent, the core of the architecture, adjusts the parameters of each combination of operators by processing a large number of images. Towards the end, the agents must offer the best combination of operators and the best values of their parameters.

Keywords: Computer Vision, Reinforcement Learning, Multi-Agent System, Parameter Adjustment, Operator Selection, Q-learning, Segmentation.

1 Introduction

To accomplish an image processing task (segmentation, detection, object recognition, etc.) the user finds him-self faced with a multitude of applicable operators averaging the fixation of values for several parameters. The quality of results depends essentially on the operator chosen and the values assigned to its parameters. The lack of a general rule that guides the user in his choices pushes him usually to use his experience and sometimes his intuition. He, generally, proceeds by trial and error until the identification of a satisfactory result. The problem is already remarkable when the task needs to apply just one operator but with several parameters to adjust. However, in the majority of vision tasks, the user is required and sometimes even is obliged to combine several operators whose each one has a multitude of parameters to adjust. Therefore the user must select the operators, adjust their parameters and then test them sequentially on the image. This process is repeated for a long time before deciding on the quality of the results. It's a tedious work with a great waste of time. To help the user to perform vision tasks, several solutions have been proposed as systems and GUI. For example, Pandore and Ariane [1]. These semi automatic

solutions provide a library of operators and a set of parameters for each operator, the selection of operators and the adjustment of their parameters are done manually by the user by using a GUI. Even though, the user finds always difficulty to choose the appropriate operators and adjust their parameters in order to find the best result.

Some authors searched to automate the operator selection process. Draper proposed ADORE in 2000. It is a system of object recognition based on MDP (Markov Decision Process) to choose, from a current situation, the operator to apply [2]. Draper used a library of ten operators to recognize duplexes in aerial images. ADORE is based on a method which is robust theoretically, but which cannot always ensure good results because, on one hand, Draper uses a predefined and limited library of operators, and on the other hand he didn't talk about the problem of parameter adjustment. Other authors proposed methods to automatically adjust parameters of vision operators. B.NICKOLAY et al. proposed a method to automatically optimize the parameters of a machine vision system for surface inspection by using specific Evolutionary Algorithms (EA) [3]. A few years later, Taylor [4] proposed a reinforcement learning framework which uses connectionist systems as function approximators to handle the problem of determining the optimal parameters for a computer vision application even in the case of a highly dimensional, continuous parameter space. More recently, Farhang et al. [5] introduced a new method for segmentation of the prostate in transrectal ultrasound images, using a reinforcement learning (RL) scheme. He divided the initial image into sub-images and works on each sub-image in order to reach a good result. In [6] and [7], we proposed a reinforcement learning method to adjust automatically the parameters of vision operators. Despite all these researches and their results, they stay limited to a predefined type of images or depend on some particular conditions. Until today, there is no method robust, sure and automatic which provides the user the appropriate operators and their optimal parameters values depending on the vision task and the class of images. Hence, we need systems that allow, generally and for any vision task, to automatically determine the best combination of operators and their optimal parameters values to apply.

In this paper we present a solution for this problem by proposing a multi-agents architecture based on reinforcement learning to select automatically the best operators to apply in a vision task, that's while adjusting their parameters values without the user intervention. In the second section we present an overview on reinforcement learning, multi-agent systems. The third section details the proposed approach. The forth section discusses the experience and its results. The last section concludes the paper.

2 Reinforcement Learning

According to the definition of S.Sutton and G.Barto [8], reinforcement learning defines a type of interaction between an agent and its environment. From a real situation « s » in the environment, the agent chooses and executes an action « a » which causes a transition to the state « s' ». It receives in return a reinforcement signal « r », which is a penalty if the action leads to a failure or a reward if the action is beneficial; a zero

signal means the inability to assign a penalty or a reward. The agent uses then this signal to improve its strategy, action sequence, in order to maximize the accumulation of its future rewards. For this purpose, it must balance exploration and exploitation. The exploration is to test new action, which could lead to higher earnings. Whereas the exploitation consists to apply the best strategy previously acquired. Watkins has developed Q-learning, a well-established on-line learning algorithm, as a practical RL method [9]. In this algorithm, the agent maintains a numerical value for each state-action, representing a prediction of the worthiness of taking an action in a state. Table 1 represents an iterative policy evaluation for updating the state-action values where r is the reward value received for taking action a in state s , s' is the next state, α is the learning rate, and γ is the discount factor. An ϵ -greedy policy is used to make a balance between exploration and exploitation. The ϵ -greedy selects the action with the highest Q-value in the given state with a probability of $1 - \epsilon$ and others with a probability of ϵ . The reward r is defined according to each state-action pair (s, a) . The goal is to find a policy to maximize the discounted sum of rewards received over time. The principal concerns in RL are the cases where the optimal solutions cannot be found, but can be approximated. Ideally, the RL agent does not require a set of training samples. Instead, it can continuously learn and adapt while performing the required task.

Table 1. Q-learning algorithm

Initialize $Q(s, a)$ arbitrary

Repeat (for each episode)

Initialize s

Repeat (for each step of episode)

Choose a from s using policy derived from Q

Take action a , observe r, s'

Take action a , observe r, s' $Q^x(s_t, a_t) = (1 - \alpha_t)Q^x(s, a) + \alpha_t(r + \gamma \max_{a \in A} Q^x(s', a'))$

$s \leftarrow s'$;

Until s is terminal

3 Multi-Agent Systems

The agent concept has been studied for a long time in various disciplines. Multiple definitions of agent have been given depending on the field of application. In our work, we use the definition adopted by Haroun [10] based on M. Wooldridge's works: "an agent is a computer system, situated in some environment, that acts autonomously and flexibly in order to achieve its delegated goals".

A multi-agents system consists of a set of multiple agents living at the same time, sharing common resources and communicating with each other. The key point of multi-agents systems is the formalization of coordination between agents. The agents are able to perceive and act on a common environment that they share. Perceptions allow agents to acquire information about their environment evolution, and their actions allow them to change it.

4 Proposed Approach

Generally, to accomplish a vision task we've to pass through some phases of processing. Each phase contains a set of operators, usually predefined in a system with their parameters whose some of their values are given by default. The users find themselves faced with a tedious work of choosing the best operator to apply and adjusting its parameters.

In our approach we propose a multi-agent architecture which helps automatically the user in his choices (operators and parameters values). The architecture is composed globally by three types of agents. Each one of them is charged to accomplish a specific task in the process. Fig. 1. shows how these agents are linked.

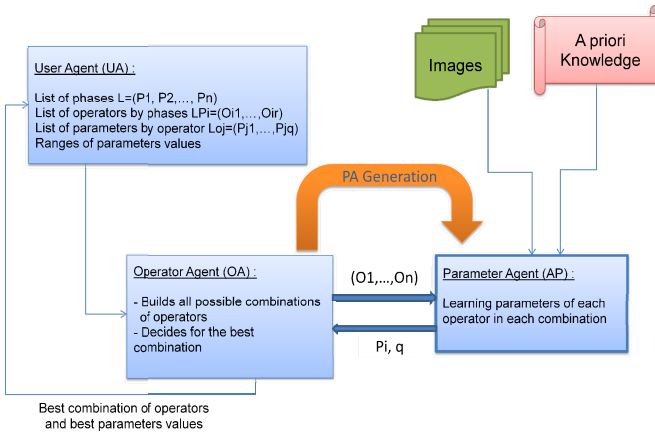


Fig. 1. Global schema of the proposed approach. OA proceeds in collaboration with PA.

4.1 User Agent (UA)

Depending on the vision task to accomplish, UA gives the list of processing phases. For each phase, it determines a set of possible operators. For each operator it defines parameters to adjust by specifying ranges of their possible values. It also proposes a class of images for learning, on which the system will run, as well as a ground truth for each image. The work of UA is necessary so that the operator agent and the parameter agent can proceed.

4.2 Operator Agent (OA)

Operator agent proceeds in two steps: the first one is to build, according to the phases determined by UA, all possible combinations of operators. Each combination contains a number of operators which is equal to the number of the phases determined by UA. For each combination, the agent OA generates an agent PA (Parameter Agent) specialized to adjust parameters of its operators. There are then so many agents PA as possible combinations. Each agent PA has its own combination of operators.

After adjusting parameters, according to the task at hand, each agent PA returns its combination of operators with the best parameters values. It also returns the result quality of this combination after applying it on the class of images determined by UA. The second step of the agent OA is to decide among all these combinations of operators which one is the best to apply. The best combination corresponds to this one having the higher result quality; it is then returned to the agent UA.

Each agent PA uses reinforcement learning to adjust the parameters of each operator. It applies actions on a set of images and receives a return which may be a punishment or a reward. This return is determined depending to a ground truth proposed by an expert (manual processing). More details about how the agents PA proceed are given hereafter.

4.3 Parameter Agent (PA)

For each combination of operators, there is an agent PA to adjust parameters of these operators. To do this, a range of values for each parameter and a set of images with their ground truth are given by the agent UA. The agent PA has no prior knowledge about the best parameters values. It proceeds by reinforcement learning to find values giving the best result. Fig. 2. presents a general schema about the functioning of each agent PA.

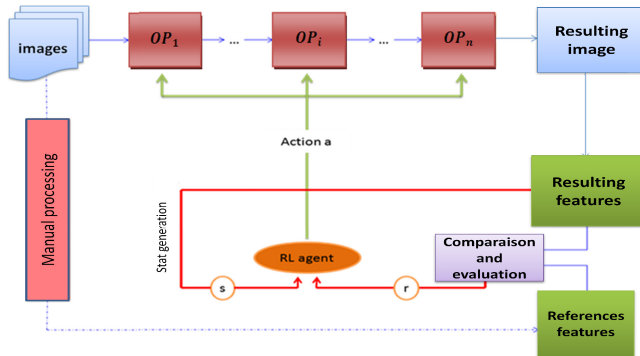


Fig. 2. General schema about the functioning of each agent PA for operator’s parameter learning

The input image is a processing subject of a series of operators. Each operator has a set of parameters to adjust, and each parameter has a range of possible values. The agent PA must find the best parameter values for each operator in order to get the best result. The agent PA uses reinforcement learning as an automatic method to explore all possible values and then exploit the best ones. The agent PA must then define the actions a , states s and reward r . We define actions as all possible combinations of parameters values. States are defined by features describing the image. These features are defined according to the task at hand. The agent PA chooses an action and applies the combination of operators on the input image and gets a resulting image. Each image has its ground-truth; it is a resulting image through a manual processing by an

expert. The ground-truth represents a reference for the agent PA. To assess the chosen action, the agent PA compares the resulting image with the reference one. It extracts some features from the resulting image and compares them with the same features extracted from the reference image. An evaluation metric is used to assess the result and produce a reward. Each action has its own reward. The best action is the most rewarded. The details about how the three components, namely state, action, and reward are defined in our proposed approach are described in the next subsections.

1) *Defining actions*: Generally, all possible combination of parameters values of operators is defined as an action for the agent PA. The set of the actions is then the set of all possible values combination, see fig. 2.

Each operator OP_k has a series of parameters:

$$(P_1^k, P_2^k, \dots, P_n^k)$$

Each parameter P_j^k has a range of values:

$$V_j^k = \{V_{j1}^k, V_{j2}^k, \dots, V_{jm}^k\}$$

An elementary action of the operator OP_k is:

$$a_k = (u_{j1}^k, \dots, u_{jr}^k) \text{ where } u_{j1}^k \in V_j^k$$

An action of the agent PA is defined by the combinations of the elementary actions of operators as it is defined above:

$$a = (a_1, a_2, \dots, a_n)$$

2) *Defining states*: A state is defined by a set of features extracted from the resulting image:

$$s = [\chi_1, \chi_2, \dots, \chi_n]$$

χ_i is a feature reflecting the state of the image after the processing. The type of the extracted features depends on the task at hand. Here we give a general definition, and in the experience we define them explicitly according to the application.

3) *Defining the reward*: The return is a reward if the agent PA chooses the right action, else it is a punishment. It is defined according to the quality of the processing result. This quality is assessed by using ground-truth models (manually processed images). To define the return we calculate the similarity between the resulting image and the ground truth image. That is depending on the task at hand. For example, if we use an edge detection approach for image segmentation we would calculate error measures which give global indices about the result quality: over-detection error, under-detection error, localization error [11]. But if we use a region approach we would calculate, for example, errors of Yasnoff [12] or the criterion of Vinet [12], etc. After measuring the similarity's criterions, we assess the result of our system using a weighted sum of the differences of these criterions' scalars:

$$D = \sum_i w_i D_i$$

The weights w_i are chosen according to the importance of each criterion D_i .

In our experiments, we've used three error measures: over-detection error, under-detection error and localization error [11] which are formally expressed in the next section.

A general form of the reward definition in the proposed approach is presented by:

```
Reward: r= -10, 0 or 10;
      if (D < ε) r = +10; f=true;
          elseif ( (D > ε) && (D < ε + δ) )
              r = 0;
          else r = -10;
      end
end
```

The values 10 and -10 represent respectively the reward and the punishment depending to a predefined threshold. Using the set of images determined by the agent UA, each agent PA returns to the agent OA its combination of operators with the best values of their parameters. It returns also the quality of the result corresponding to the highest reward. The agent OA retrieves then all the combinations it has built with the best parameters values of each operator and the qualities of their results. The agent OA returns to the agent UA the best combination of operators corresponding to the highest quality. Thus it decides which the best combination of operators to apply is.

In the following section we test this approach for segmentation tasks.

5 Results and Discussion

In this section we test practically the multi agent architecture to choose the right operators and their best parameters values for segmentation tasks. The operators used in image segmentation differ from a class of images to another, they are not necessary the same. Our main goal in this section is to show how the proposed approach can determine, for a class of images, the best combination of operators to apply for their segmentation. A dataset of 70 images of traffic signs with their ground truth are used in this experience.

5.1 The Agent UA

It defines three phases of processing: preprocessing, processing and post processing.

1) *Preprocessing phase*: three filters are defined: '**medfilt2**'; '**ordfilt2**'; '**wiener2**' as they are predefined in Matlab. The size of the used filter is the parameter to adjust. An operator is defined as:

Op= {operator name, number of parameters, List of possible Values}

2) *Processing phase*: one operator is defined: '**edge**', with two parameters to adjust: the filter to select and the threshold to remove edges with poor contrast. Contours are formed by pixels higher than a given threshold.

3) *Post processing phase*: one operator is defined: '**bwareaopen**', with two parameters to adjust: the connectivity and the maximal size of the objects to remove.

5.2 The Agent OA

It constructs all possible combinations of operators. As the agent UA determines three phases to accomplish the segmentation task, each one of these combinations will contain three operators.

For each combination, the agent OA generates an agent PA to adjust the parameters according to the dataset of images proposed by the agent UA. There are then three agents PA1, PA2 and PA3 which treat respectively the combinations: C1, C2 and C3.

5.3 The Agent PA

The agent PA is, generally, charged to adjust parameters of each operator in order that the segmentation result will be as close as possible to the segmentation done manually by an expert. To adjust parameters of each operator, the agent PA uses reinforcement learning. It must then define actions, states and reward.

Actions: Actions are all possible combinations of parameters values. We select another action by choosing other parameters values. An example of an action for the agent PA1: Action= [3, ('sobel', 0.02), (5,8)]

States: States are defined by some features extracted from the image. In this application, we define a state by three features:

$$s = [\chi_1, \chi_2, \chi_3]$$

χ_1 is the ratio between the number of contours in the resulting image and the number of those in the reference image (ground truth).

χ_2 is the ratio between the total of white pixels in the resulting image and those in the reference image.

χ_3 is the ratio between the length of the longest contour and the length of the longest contour in the reference image.

Reward: Reward is defined by a weighted sum of three error measures which give some global indices about the quality of boundary-based segmentation: over detection error D1, under-detection error D2 and localization error D3 [11]. These criteria evaluate a result of edge detection. The weights used in the definition of the reward are chosen according to the importance of each criterion. The reward is defined as:

$$D = w_1 D_1 + w_2 D_2 + w_3 D_3$$

where w_i is a weight for D_i

For each combination of operators, the agent PA finds the best parameters values which give the best segmentation. In this experience, we test the proposed architecture to segment two different types of images. Fig. 3. shows the result of segmentation for 5 images taken randomly from the processing of the dataset.

It is important to note that we evaluate an operator on the whole of the dataset of images and not one by one. The fixation of some parameters of the Q-learning algorithm affects largely the result. The results showed in fig. 3. are for: $\alpha=0.5$, $\gamma=0.8$, $\varepsilon=0.5$, number of episodes=200 and number of steps=80.

The combination of operators having the highest quality of segmentation is (wiener2, edge, bwareaopen) and the most rewarded action is (5; (prewitt, 3.000000e-002); (10, 8)).

It is the combination of operators decided by the agent OA and returned to the agent UA.

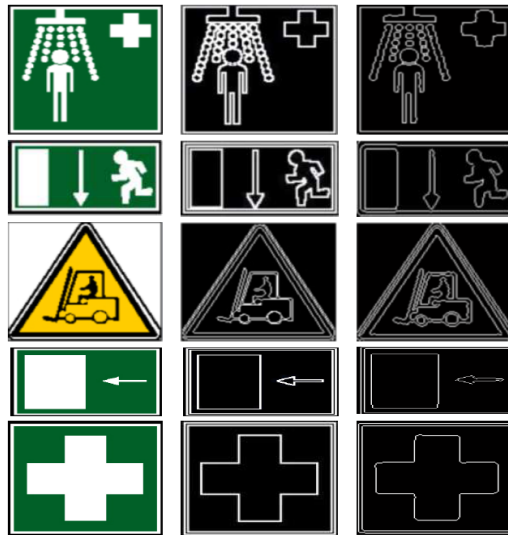


Fig. 3. From left to right: the initial image, the image segmented manually and the result of the proposed approach

6 Conclusion

Choosing the appropriate operators to apply and then adjusting their parameters values to accomplish a vision task represent a big challenge for users. In this paper we presented a multi-agents architecture based on reinforcement learning, which helps users by proposing them the optimal series of operators to apply and their best parameters values. Our system proceeds automatically to decide for his choices. Through the reinforcement

learning mechanism, our architecture does not consider only the system opportunities but also the user preferences. We intended to propose a general new way of thinking about the automatic selection of operators and the automatic adjustment of their parameters without the user intervention. The proposed approach constitutes then a theoretical robust basis for vision users and not just a solution for a particular problem. The experience we have done does not restrict the application of the approach to the image processing field, but its theoretical procedure shows that it can be applied to any decision process using parametric methods. Despite the theoretical strength of the idea and the obtained results, we acknowledge that we must improve the learning algorithm and study the reward expression using a function based on the similarity between the resulting images and the ground-truth.

References

1. Clouard, R., Elmoataz, A., Angot, F.: PANDORE: une bibliothèque et un environnement de programmation d'opérateurs de traitement d'images. Rapport interne du GREYC, Caen, France, Mars (1997)
2. Draper, B.A., Bins, J., Baek, K.: ADORE: Adaptive Object Recognition. *Videre* 1(4), 86–99 (2000)
3. Nickolay, B., Schneider, B., Jacob, S.: Parameter Optimization of an Image Processing System using Evolutionary Algorithms. In: Sommer, G., Daniilidis, K., Pauli, J. (eds.) CAIP 1997. LNCS, vol. 1296, pp. 637–644. Springer, Heidelberg (1997)
4. Taylor, G.W.: A Reinforcement Learning Framework for Parameter Control in Computer Vision Applications. In: Proceedings of the First Canadian Conference on Computer and Robot Vision (CRV 2004). IEEE (2004)
5. Sahba, F., Tizhoosh, H.R., Salama, M.: Application of reinforcement learning for segmentation of transrectal ultrasound images. *BMC Medical Imaging* 8, 8 (2008)
6. Qaffou, I., Sadgal, M., Elfazziki, A.: A Reinforcement Learning Method to adjust Parameters of Vision Operators. In: Sixth International Conference on Intelligent Systems: Theory and Application, Rabat, pp. 23–29 (2010)
7. Qaffou, I., Sadgal, M., Elfazziki, A.: A reinforcement learning method to adjust parameter of a texture segmentation. In: The 7th International Conference on Informatics and Systems (INFOS), Cairo, pp. 1–5 (2010)
8. Sutton, R.S., Barto, A.G.: Reinforcement Learning. MIT Press, Cambridge (1998)
9. Watkins, C.J.C.H., Dayan, P.: Q-Learning. *Machine Learning* 8, 279–292 (1992)
10. Haroun, R.: Segmentation des tissus cérébraux sur des images par résonance magnétique. Master's thesis, Université des sciences et de la technologie Houari Boumediène (2005)
11. Chabrier, S., Laurent, H., Rosenberger, C., Zhang, Y.J.: Supervised evaluation of synthetic and real contour segmentation results. In: European Signal Processing Conference, EUSIPCO (2006)
12. Do, M.C.: Évaluation de la segmentation d'images. Rapport final TIPE. Institut de la francophonie pour l'informatique. Nano

A Phase Congruency Based Document Binarization

Hossein Ziaei Nafchi¹ and Hamidreza Rashidy Kanan²

¹ Department of Electrical, Computer and IT Engineering, Qazvin Branch,
Islamic Azad University, Qazvin, Iran
h.ziaei@qiau.ac.ir

² Department of Electrical Engineering, Bu-Ali Sina University, Hamedan, Iran
h.rashidykanan@basu.ac.ir

Abstract. In this paper, three new methods proposed for binarization of degraded documents and manuscripts. Phase congruency used to select regions of interest (ROI) of document's foreground. The main idea is to achieve an optimal recall measure (recall~1), while the precision value is at an acceptable level. Further processing should be performed to focus on the ROI. We also used a modified adaptive thresholding method in the next step. This method uses a global variance, a global mean and local means for thresholding. Finally, a new method called early exclusion criterion (EEC) proposed for document enhancement. The experimental results on the datasets introduced in DIBCO 2009, H-DIBCO 2010 and DIBCO 2011 shows that near optimal recall value (recall~0.99) obtained, while precision measure's value is acceptable.

Keywords: Degraded document binarization, Phase congruency, Adaptive thresholding, Early exclusion criterion.

1 Introduction

The purpose of document binarization is to convert input gray-scale documents into binary form. Usually, the latter form will be used in most document analysis systems as the first step. The performance of document binarization step highly impacts the subsequent steps such as page segmentation and optical character recognition. A number of historical and badly degraded documents can be found in libraries and archives. Usually, reading or processing these documents is not easy. For converting such injured documents, adaptive thresholding techniques are the best choices. Adaptive thresholding technique is a robust method to handle strong illumination changes. A global binarization method such as Otsu's method [1], usually fail in such an environment conditions. Global binarization methods try to find a threshold and use it for whole document image.

In this paper, three methods proposed for document processing. Phase congruency [2] is a well-known edge detector. It widely used in the machine vision literature. Phase congruency shows advantages against gradient-based edge detectors such as Sobel and Canny due to their sensitivity to variations in image illumination, blurring and magnification [2]. In this paper, phase congruency is used to select edges of

foreground. Then, a morphological grey-scale image reconstruction [3] used to fill the obtained edges. After conversion to binary form, it contains all the foreground information. This means that recall~1.

A number of adaptive thresholding methods have been proposed [4, 5, 6, 7, 8, 9, 10, 11] and [12]. Sauvola et al [4] proposed an adaptive thresholding method for image binarization. They used local mean and local variance to compute a threshold for each pixel in the input image. Shafiat el al [5] improved the speed of the Sauvola's approach by using two integral images. An adaptive thresholding approach has been proposed by Wellner [6]. In this approach, each pixel is set to 0 (black) if the pixel intensity value is smaller than 85% of average of the intensity values of some surrounding pixels in the x-axis. Bradley and Roth [7] improved this approach by including surrounding pixels in both x and y axes and speeded-up their method by utilizing integral image and achieved real-time thresholding. We improved this approach by interfering the global mean and global standard deviation.

In the face detection literature, Liu [13] proposed a method called early exclusion criterion (EEC) to exclude windows which cannot be faces at all. This criterion used as first step (early exclusion) to speed up overall face detection speed. We used a different form of this approach for document enhancement.

The rest of the paper is organized as follows: In section 2, we discussed the related works. In section 3, proposed phase congruency based ROI selection is introduced. Section 4 elaborates our proposed modified adaptive thresholding method. In section 5, the proposed early exclusion (EEC) method for document enhancement is introduced. Section 6 deals with the experimental results and Section 7 draws a conclusion.

2 Related Works

Many of the thresholding methods have been surveyed in [14, 15] and [16]. Sauvola's method uses local mean and local variances to compute a threshold $t(x,y)$ for pixel $g(x,y)$ in the grey-scale image g . Each threshold is computed by the following equation:

$$t(x, y) = m(x, y) \left[1 + k \left(\frac{\sigma(x, y)}{R} \right) - 1 \right]. \quad (1)$$

where $m(x, y)$ is the local mean, $\sigma(x, y)$ is local standard deviation, R is the maximum value of the standard deviation and k is a parameter which takes the positive values in the range of [0.2,0.5], [4]. The value of k is 128 for grey-scale images. If the contrast of surrounding pixels of $g(x, y)$ is high, then $\sigma(x, y) \cong R \cong 128$ and then $t(x, y) = m(x, y)$. If the value of $s(x, y)$ becomes low then $t(x, y)$ becomes less than $m(x, y)$, and as a result the dark side of background will be removed. This approach shows acceptable results even for the severely degraded documents. Sauvola's method is the modification of Niblack's method [8]. A problem with Sauvola's approach is its slow process time. Computing local mean and local variance for each pixel is a slow process. Shafiat et al [5] improves the sauvola's method by using two integral images. Remember that integral image widely used in the machine vision literature, after utilizing by Viola and Jones [17]. By utilizing the integral image, we can compute sum of rectangles in the constant time independent of rectangle size.

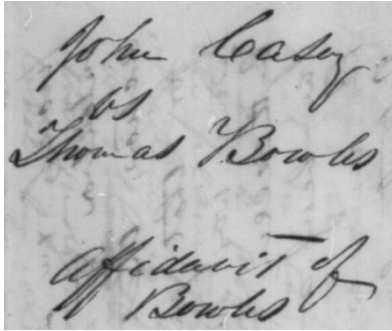
Bradley et al [7] used local mean of the surrounding pixels $S \times S$ for thresholding. $S \times S$ means that pixels of the both x and y axes used. They choose the surrounding pixels as $1/8$ of image length. A pixel is set to dark if its value is less than the product of a coefficient C and local mean. They chose the value of C as 0.85 like the Wellner [6].

Otsu's method [1] tries to choose a threshold to minimize the intraclass variance of the black and white pixels. Moghaddam et al [10] introduced AdOtsu, which is an adaptive and parameterless generalization of the Otsu's method. They used multiscale background estimation and a skeleton-based postprocessing to remove false positive sub-strokes. Moghaddam et al [11] proposed a multi-scale framework for adaptive binarization of degraded document images. This framework is based on the several binarizations on different scales and use of AdOtsu. Hedjam et al [12] proposed a spatially adaptive statistical method for image binarization. They used Sauvola's method to obtain an initial map and adapt a two-class maximum likelihood classifier to the pixels. The parameters of the class are computed locally from the grey-level distribution. Moghaddam et al [18] proposed a non-local patch means (NLPM) restoration and reconstruction method for preprocessing degraded document images. The image data is represented by a content-level descriptor based on patches. Then a modified genetic algorithm is used to correct the patched image based on the similar patches identified.

Finally, a number of hybrid methods used for document binarization. For example, Gatos et al [19] proposed a hybrid adaptive thresholding method based on combination of several methods. They also used edge information and enhancement step based on mathematical morphology operation.

3 Phase Congruency Based ROI Selection

A new procedure, based on phase congruency is proposed to perform document image binarization. Phase congruency is a robust method to detect edges and corners. Phase congruency's robustness to image variations stems from the multi-scale and multi-orientational approach to phase congruency calculation and from the fact that phase rather than magnitude information is considered for line or edge detection. We refer to reference [2] for more study about phase congruency. In this paper, phase congruency is used to detect edges of document's information. A number of parameters impact the phase congruency output. Specially, the number of $2D$ log-Gabor filters scales p and the number of orientations of $2D$ log-Gabor filters r should be set according to the application. As we see in Fig. 1 and Fig. 2 Phase congruency detects edges. As a result inner information of edges of document information will be lost. Therefore, we use grey-scale image reconstruction [3] to fill the inner parts of edges. Then a conversion to binary form must be performed. We choose a global threshold for this purpose. This global threshold can be either the mean of the filled image or even can be obtained threshold from Otsu global thresholding method. In our experiments, we used half of Otsu's global threshold. This threshold achieve near optimal recall measure while the precision value is acceptable for a ROI selector. The output of this step is the input for the proposed adaptive thresholding method. Fig. 1 and Fig. 2 shows above mentioned process.



a) A degraded document image (from DIBCO 2009 dataset)



b) The edge image obtained by using the phase congruency

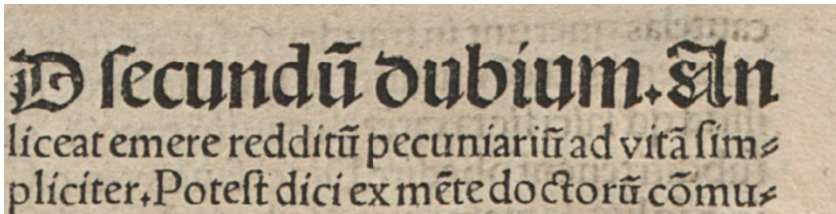


c) Filled image of (b)



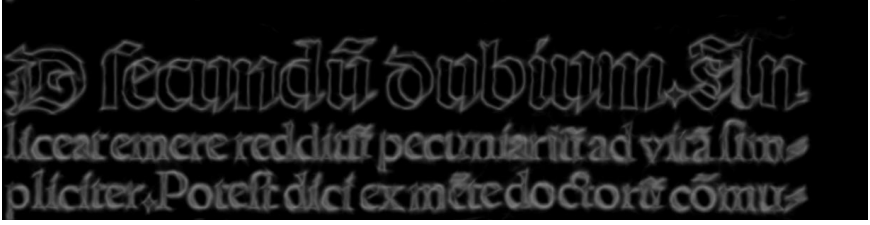
d) Binary conversion of (c).

Fig. 1. Phase congruency based ROI selection process. We used $p = 10$, $r = 10$ in our experiments. The phase congruency was able to reject a majority of the background pixels. ($Recall = 100$, $F - measure = 44.35$).



a) A degraded Input image (from DIBCO 2009 dataset)

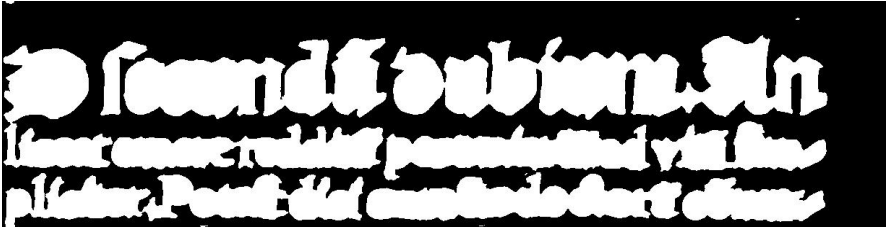
Fig. 2. Phase congruency based ROI selection process. ($Recall = 100$, $F - measure = 61.95$).



b) The edge image obtained by using the phase congruency



c) Filled image of (b)



d) Binary conversion of (c).

Fig. 2. (Continued)

4 The Proposed Adaptive Thresholding Method

While many of the adaptive thresholding methods use local mean and local variance to compute a threshold for each pixel, we use a different manner. The proposed method which is an improvement of the Bradley et al [7] approach is as follows. For each pixel, we compute the average of $S \times S$ surrounding pixels of that pixel and compare value of that pixel with product of obtained average with a coefficient. This coefficient is computed with the following equation:

$$C = 0.85 + \frac{|\mu + \sigma|}{1000}. \quad (2)$$

where, μ and σ are the average and standard deviation of intensities of input image. A pixel is set to 0 (dark) if the value of that pixel is smaller than the product of S mean values and C , otherwise pixel is set to 1 (white). Wellner [6] and Bradley et al [7] choose the value of the C as 0.85 and number of S as 1/8 of image length. Furthermore, Bradley & Roth suggested that for different applications one can use a

different C . Suppose that we have an image with high intensity pixels, Bradley et al adaptive thresholding method may fail because pixel value usually becomes more than surrounding pixels and maybe set to 1 erroneously. In images with low intensity values the same scenario repeated in setting pixels with 0 (dark). The proposed method interfere the mean and standard deviation of input image in the coefficient to overcome this problem. If input image has dark values and also low variations, then proposed approach considers low intensity values as background in according to its $S \times S$ mean. This is because the coefficient C becomes less than 1. We also choose the number of S as $1/16$ of image length.

In this paper, the output of the phase congruency ROI is the input for adaptive thresholding. In the input image those pixels in the ROI which classified as background, replaced by the mean of the input image. Instead of replacing by mean value some methods replaced the background values by *NaN* value. This approach also can be used as a preprocessing step without replacing by the mean value. This approach changes the F-measure value to some extent. Adaptive thresholding method converts this input image to binary form. At the end of this step, many of unnecessary holes will be removed while approximately all the sub-strokes remains. Results can be found in section 6.

5 Early Exclusion Criterion Enhancement

An early exclusion criterion has been proposed by Liu [13] for face detection preprocessing. A completely different manner of this criterion is used in this paper for document enhancement. While this criterion used as the first step in the face detection applications, we use it as final step. The reason is its slower process time in comparison with adaptive thresholding used in section 4. The purpose is to remove those background pixels which already classified as foreground. The proposed document enhancement method is as follow.

After removing isolated pixels from previous step, a 5×5 sliding window slides across whole input image. Average intensity of the sliding window m is computed from integral image. Then m_h is computed, where m_h is the average intensity of those pixels which has higher value than m . A pixel is set to background if it satisfies two conditions: $in(i, j) \times k^* > m_h$ and $m_h \neq 0$, where $k^* > 1$. We set $k^* = 1.05$ in our experiments. We observed that some inner parts of large connected foregrounds may set to background erroneously. Therefore, we use a third condition in addition to above two mentioned conditions. The value of $bw(i, j)$ and at least one of its 3×3 mask should be unequal. This condition solves the large connected foreground problem. It's clear that only foreground pixels from previous steps are evaluated by EEC. EEC removes those pixels in which they cannot be foreground. The experimental results show that outstanding results achieved for an enhancement procedure.

6 Experimental Results

The proposed methods have been tested on the standard datasets provided in DIBCO'09 [16], H-DIBCO'10 [20] and DIBCO'11 [21]. The measures recall, precision and F-measure used in our experiments:

$$Recall = \frac{TP}{TP+FN}. \quad (3)$$

$$Precision = \frac{TP}{TP+FP}. \quad (4)$$

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision}. \quad (5)$$

where, TP, FP, FN denote the true positive, false positive and false negative values, respectively. Table 1 provide the results of the proposed methods, Otsu [1], spatially adaptive [12], and AdOtsu methods [10]. To the best of our knowledge, this approach [10] has highest F-measure reported to date.

Table 1. Experimental results comparison between proposed methods, Otsu's method, spatially adaptive method and AdOtsu method for DIBCO'09 dataset

<i>Method</i>	<i>Measures</i>		
	Recall	Precision	F-measure
Phase congruency (1)	99.83	30.17	44.83
Phase congruency + Adaptive thresholding (2)	98.82	61.15	71.85
(1) + (2) + Early Exclusion Criterion	98.50	68.18	78.30
Otsu [1]	94.25	73.66	78.59
Spatially adaptive [12]	92.10	90.72	91.35
AdOtsu [10]	90.09	93.22	91.57

The proposed adaptive thresholding method improved the F-measure value from phase congruency by about 60% at the cost of 1% decrease of the recall value. The EEC improved F-measure by about 10% while reduction of recall value is only 0.3%. Recall and F-measure values for printed images in the DIBCO'09 dataset (P01-P05) are 98.21 and 89.30, respectively. That is 98.78 and 67.30 for handwritten images. Total running time to convert all 10 images in the DIBCO'09 dataset is about 31 seconds. The experiments performed on a Pentium 4, 3.4 GHz with 4 GB of RAM.

Although our proposed methods achieve smaller F-measure values, it should be noticed that a recall~99 achieved. An alternative method can further improve the performance by focusing on the ROI instead of dealing with all pixels in the input image. We also noticed that many of the methods evaluated in the DIBCO'09 contest [16] have higher F-measure value. We believe that a well combination of several methods can result in a near optimum F-measure. Each of these methods must have optimal or near-optimal recall value and acceptable precision value. For example, Gatos et al [19] combined several state-of-the-art methods to achieve high F-measure value. For this end, we proposed three mentioned methods.

To show the robustness of the proposed preprocessing step by using the phase congruency, results for the H-DIBCO'10 and DIBCO'11 are listed in Table 2. The experimental results of the DIBCO'10 and DIBCO'11 reports [20, 21] indicate that these datasets includes more problematic document images than DIBCO'09 dataset.

Table 2. Experimental results of preprocessing by using the proposed phase congruency based region of interest selection

<i>Dataset</i>	<i>Measures</i>		
	Recall	Precision	F-measure
H-DIBCO 2010	99.72	29.01	44.32
DIBCO 2011 (Handwritten)	99.09	32.25	48.21
DIBCO 2011 (Machine Print)	98.33	40.45	57.12

7 Conclusion

In this paper, three methods proposed to select regions of interest and binarize degraded document images. Phase congruency, image filling and Otsu's global threshold used to select regions of interest. Then, we used regions of interest information obtained from first method and a modified adaptive thresholding method to further improve the binarization result. Finally, an enhancement method called early exclusion criterion proposed and used for further document enhancement. Experimental results on the DIBCO'09, H-DIBCO'10 and DIBCO'11 datasets shows that near optimal recall value obtained, while precision value is acceptable. A subsequent method should be employed to achieve better results. In future work, we focus on the shape based document binarization based on phase congruency.

References

1. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man, and Cybernetics* 9(1), 62–66 (1979)
2. Kovese, P.: Image Features from Phase Congruency. *Videre: Journal of Computer Vision Research* 1(3) (1999)
3. Soille, P.: *Morphological Image Analysis, Principles and Applications*. Springer (2007)
4. Sauvola, J., Pietikainen, M.: Adaptive binary image binarization. *Pattern Recognition* 33(2), 225–236 (2000)
5. Shafiat, F., Keysers, D., Breuel, T.M.: Efficient Implementation of Local Adaptive Thresholding Techniques Using Integral Images. *Document Recognition and Retrieval XV* (2008)
6. Wellner, P.D.: Adaptive thresholding for the digitaldesk. *Tech. Rep. EPC-110* (1993)
7. Bradley, D., Roth, G.: Adaptive thresholding using the integral image. *Journal of Graphic Tools* 12(2), 13–21 (2007)
8. Niblack, W.: *An Introduction to Image Processing*. Prentice-Hall Press (1986)

9. Gatos, B., Pratikakis, I., Perantonis, S.J.: Adaptive degraded document image binarization. *Pattern Recognition* 39(3), 317–327 (2006)
10. Moghaddam, R.F., Cheriet, M.: AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization. *Pattern Recognition* 45, 2419–2431 (2012)
11. Moghaddam, R.F., Cheriet, M.: A multi-scale framework for adaptive binarization of degraded document images. *Pattern Recognition* 43(6), 2186–2198 (2010)
12. Hedjam, R., Moghaddam, R.F., Cheriet, M.: A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images. *Pattern Recognition* 44(9), 2184–2196 (2011)
13. Liu, C.: A Bayesian discriminating features method for face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 25(6), 725–740 (2003)
14. Trier, O., Taxt, T.: Evaluation of binarization methods for document images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 17, 312–315 (1995)
15. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging* 13(1), 146–165 (2004)
16. Gatos, B., Ntirogiannis, K.: ICDAR 2009 document image binarization contest (DIBCO 2009). In: *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1375–1382 (2009)
17. Viola, P., Jones, M.: Robust Real-Time Face Detection. *International Journal of Computer Vision* 57(2), 137–154 (2004)
18. Moghaddam, R.F., Cheriet, M.: Beyond pixels and regions: A non-local patch means (NLPM) method for content-level restoration, enhancement, and reconstruction of degraded document images. *Pattern Recognition* 44(2), 730–743 (2011)
19. Gatos, B., Pratikakis, I., Perantonis, S.J.: Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. In: *International Conference on Pattern Recognition*, pp. 1–4 (2008)
20. Pratikakis, I., Gatos, B., Ntirogiannis, K.: H-DIBCO 2010 – Handwritten Document Image Binarization Competition. In: *International Conference on Frontiers in Handwritten Recognition*, pp. 727–732 (2010)
21. Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICDAR 2011 Document Image Binarization Contest (DIBCO 2011). In: *International Conference on Document Analysis and Recognition*, pp. 1506–1510 (2011)

Porting a H264/AVC Adaptive in Loop Deblocking Filter to a TI DM6437EVM DSP

Abdellah Skoudarli¹, Mokhtar Nibouche², and Amina Serir¹

¹ USTHB-Faculty of Electronic and Informatic Laboratory of Image Processing and Radiation,
BP 32 El Alia Bab Ezzouar Alger, Algeria

² Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UWE, UK
askoudarli@usthb.dz, mokhtar.nibouche@uwe.ac.uk,
aserir@hotmail.com

Abstract. Complementary units in the form encoders and decoders are generally involved in video compression standards. Both the encoder and the decoder integrate an adaptive deblocking filter, which is very beneficial in preserving and enhancing the video quality. Deblocking filters are extremely popular in improving the visual quality of decoded frames in the H.264/AVC video coding standard. The prime goal of the current paper is to efficiently implement a H.264/AVC adaptive deblocking filter using the Texas Instruments DM6437EVM DSP processor. The adopted approach requires an initial identification of the portions of the algorithm wherein parallel processing can be exploited. The functions are then re-written and the instructions rearranged using the features of the targeted hardware architecture. The adaptive deblocking algorithm was optimised and ported to a DM6437EVM DSP platform. A quick comparison shows that the optimised code is a 32 % better, in terms of speed, than the non-optimised code.

Keywords: H.264/AVC, Filtering, Adaptive Deblocking Filter, DM6437EVM DSP, C/C++ optimization.

1 Introduction

H.264/AVC is the latest video compression standard jointly developed by the ISO and ITU [1][2]. The standard achieves the best encoding performance in terms of video quality and compression ratio than its predecessor by adopting a number of new techniques including, variable block size based motion estimation in inter mode prediction, multiple directions of intra prediction, quarter-pel accuracy in motion estimation, multiple reference frames, weighted prediction, rate distortion estimation and highly adaptive in loop deblocking filter. Both the encoder and the decoder must apply the normative deblocking filter at block boundaries. The standard specifies that the filter should be applied within the motion compensation loop, and as such, the filter is often referred to as a “loop filter”.

Deblocking filters are used to improve the visual quality of decoded frames in the H.264 video coding standard [1]. These filters attempt to remove the artifacts

produced by block-based operations, which consists of 4x4 DCT blocks and motion compensation prediction. Although these deblocking filters help tremendously in improving the subjective and objective quality of the output frames, they are generally computationally intensive. In fact, even after the tremendous efforts that have been made to optimise the speed of these filtering algorithms, unfortunately, they still easily account for one third of the computational complexity of a decoder [1]. This complexity is mainly due to the high adaptivity of the filter, which requires conditional processing on the block edge and sample levels. These are known to be very time consuming and present a real challenge for parallel processing in DSP hardware.

In embedded, real-time video applications, the implementation of the H.264/AVC requires high performance, low power consumption and low cost, as well as a level of flexibility, which can be very beneficial in relation to these requirements.

Complexity analysis shows that loop filtering uses 5% and 33% of the execution time of the encoder and of the decoder, respectively. Since the filtering process is normative, it can be accelerated by processor-dedicated parallel processing instructions. DSP processors, such as the TI DM6437EVM, are specialised platforms for fast execution of specific numerical operations like multiplications and additions and as such are excellent targets for implementing loop filtering algorithms.

The remainder of this paper is organised as follows: Section 2 presents an overview of the adaptive deblocking filter algorithm. Section 3 provides a brief description of the DM6437EVM DSP. Section 4 describes the optimisation approach and section 5 summarises the experimental results. Section 6 is dedicated to the conclusion.

2 Adaptive Deblocking Filter

2.1 Deblocking Filters

There are a number of deblocking algorithms that have been proposed for reducing the block artifacts in block DCT based image compression with minimal smoothing of true edges, as illustrated in Figure 1.b. Three of the most popular techniques include:

- Projection On Convex Sets (POCS)
- Weighted Sum of Symmetrically Aligned Pixels
- Adaptive Deblocking Filter.

The POCS based iterative algorithm [3] is implemented as a two stage process. The first stage involves the band limiting of the image by low pass filtering. Then, the image is transformed to obtain the transform coefficients, which are then subjected to quantisation constraints.

In the Weighted Sum of Symmetrically Aligned Pixels [4], the value of each pixel in the picture is recomputed with a weighted sum of itself and the other pixel values, which are symmetrically aligned with respect to block boundaries.

In case of the Adaptive Deblocking Filter algorithm [5], the deblocking process is separated into two stages. In the first stage, the edge is classified into different boundary strength with the pixels along the normal to an edge. In the second stage, different filtering scheme is applied according to the strengths obtained in stage one. The algorithm flow in each of these algorithms is highly iterative either at the pixel, block or edge level.

There are two main methods for implementing deblocking filters for video codecs. They can be implemented either as post filter or loop filter, with tradeoffs inherent in either implementation. The loop filter is normative in the H.264/AVC standard and provides better visual quality and rate distortion performance [5].

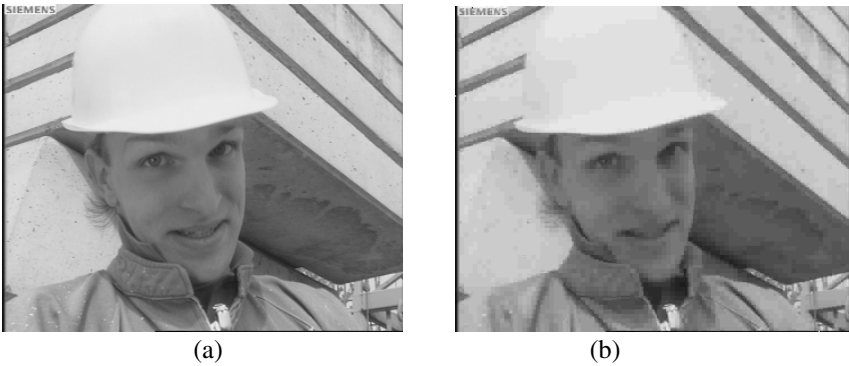


Fig. 1. (a) Original Foreman image (b) Reconstructed Foreman image without filtering

2.2 ADF Algorithm

H264/AVC uses an adaptive deblocking filter that operates on horizontal and vertical block edges within the prediction loop in order to remove artefacts caused by both the 4x4-block based transform and the coarse quantization of the transform coefficients. The filtering is based on 4x4 block boundaries, in which two pixels on either side of the boundary may be updated using different filter. The filter adjusts its filter strength adaptively according to the image local characteristics, leading to better image quality [5].

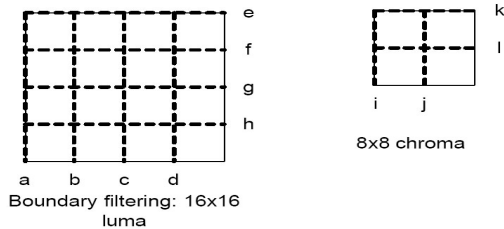


Fig. 2. Edges to be filtered in luma and chroma components

The deblocking filter is applied to both luma and chroma components separately. For each macroblock (MB), vertical edges are filtered from left to right; horizontal edges are processed from top to bottom, as illustrated in figure 2. The filtering is performed on MBs of a picture in a raster scan fashion. The filter should be applied to all 4x4 block edges of a picture, except the edges at the boundary frame, or any edges for which the filtering is disabled under certain conditions by a special flag, as described in the flowchart of figure 3.

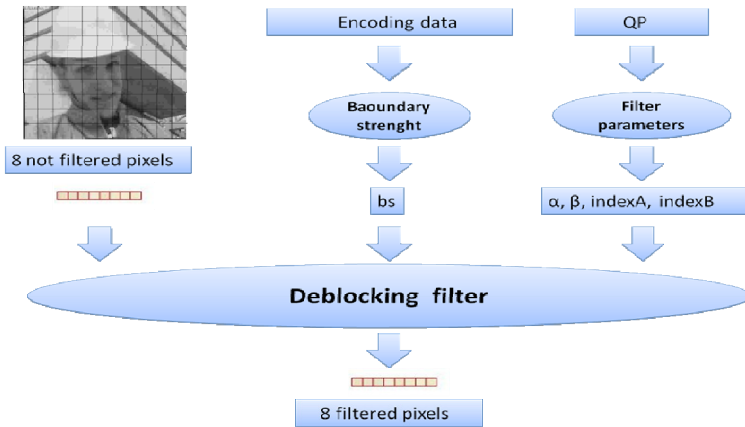


Fig. 3. Filtering Process

Depending on the coding type of the 4x4 blocks and their position within the array, a boundary strength (BS) is assigned to each edge [5]. This parameter determines the strength of filtering to be applied as shown in Table 1. ,

Table 1. Conditions for Determining a BS Value

BS	Rule
4	One of the blocks is intra and the edge is a macroblock a macroblock edge;
3	One of the blocks is intra;
2	One of the blocks has coded residuals;
1	Difference of block motion ≥ 1 luma sample distance;
1	Motion compensation from different reference frames;
0	Otherwise.

The filter is turned on or off, for each pixels across each line based on the values of p1, p0, q0 and q1 and two thresholds Alpha $\alpha(QP)$ and Beta $\beta(QP)$, which have values depending on the quantisation parameter QP of the current frame.

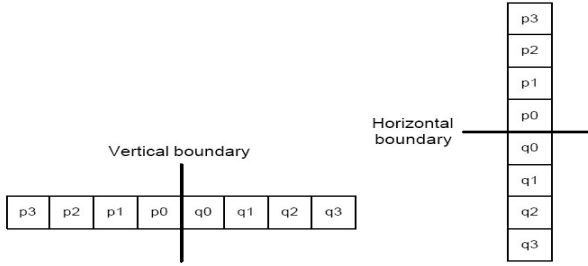


Fig. 4. Samples Across a 4x4 block Horizontal/Vertical Luma or Chroma edge

The filtering is applied to p0 and q0 only if these conditions are true:

$$\begin{aligned}
 &bs \neq 0 \\
 &abs(p_0 - q_0) < \alpha(QP) \\
 &abs(p_1 - p_0) < \beta(QP) \\
 &abs(q_1 - q_0) < \beta(QP) \quad \text{with } \beta(QP) < \alpha(QP)
 \end{aligned}$$

The filtering is extended to p1 and q1 respectively if the further conditions are also true:

$$abs(p_2 - p_0) < \beta(QP) \quad \text{or} \quad abs(q_2 - q_0) < \beta(QP)$$

The length of the filtering is also determined by the sample values over the edge, which determines the “activity parameters”. These parameters determine whether none, one or two pixels on either side of the edge are modified by the normative filter. Consequently, this analysis assesses the likelihood of an edge in the image being natural, or the result of a block based transform.

The equations calculating pixel values are defined in [2]. The equations can be classified into five categories, according to the BS values, as follows:

$$p_2 + p_1 + p_0 \tag{1}$$

$$p_2 + 2 \times p_1 + 2 \times p_0 \tag{2}$$

$$3 \times p_3 + 3 \times p_2 + p_1 + p_0 \tag{3}$$

$$2 \times p_1 + p_0 \tag{4}$$

$$(p_0 + q_0 + 1) \gg 1 \tag{5}$$

The filter is “stronger” where there is likely to be significant blocking distortion (high values of BS=4)

2.3 Complexity

The filtering algorithm is very complex. This is due to the highly adaptive nature of the algorithm of the deblocking filter, as well as to the huge quantity of pixel data to be read from memory and processed [7][8][9].

The filtering process consists of these two principal tasks:

First: **Get Strength**, which involves a large number of conditional branches. Filtering decision can be made from multiple data in parallel, in a way that pixels can be packed to operate simultaneously.

Second: **Loop Filtering** with multi-tap filter applied to the edge pixels in the decoded frame. Some optimisation techniques can be adopted conveniently to get a significant reduction.

Before performing the optimisation, the complexity of the filtering algorithm is analysed. The complexity can be summarised in the following four points:

The ADF is highly adaptive.

It is applied to each edge of all 4x4 luma and chroma blocks in a MB.

It can update three pixels in each direction where the filtering takes place.

In order to be applied to an edge, the related pixels in the current and neighboring 4x4 blocks must be read from memory and processed.

3 Overview of DM6437EVM DSP

The DaVinci™ TMS320DM6437 Digital Video Development Platform (DVDP) is a high performance video DSP processor from Texas Instruments. It is based on the third generation high performance, advanced Velocity™, TI's very long instruction word (VLIW) [10]. With performance of 4800 million of instructions per second at the clock rate 600 MHz, the DM6437EVM core offers solution to high performance DSP programming challenges. The DM6437EVM has an application-specific hardware logic, on chip memory, and additional on chip peripherals.

The DM6437EVM core uses a two level cache-based architecture [11]. The level 1 program memory cache (L1P) consists of a 32 Ko memory space and the level 1 data (L1D) consists of 80 Ko memory space. The level 2 memory cache (L2) consists of a 128 Ko memory space that is shared between program and data space. L2 memory can be configured as mapped memory, cache or combined memory.

Existing C6x DSPs support various instructions to execute packed operations between two registers. These operations are very useful for video processing [12]. Figure 5 illustrates the TMS320 DM6437EVM block diagram.

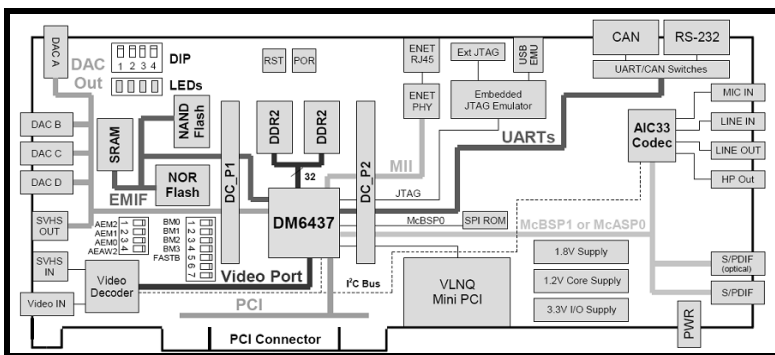


Fig. 5. TMS320 DM6437EVM block diagram

4 The Proposed Optimisation Approach

The proposed optimisation approach realises substantial gains from the performance of the C/C++ code by refining the code in terms of areas execution time, code size and memory access.

4.1 Using Ininsics to Replace Complicated C Code

The C6000 compiler provides intrinsic, special functions that map high-level operations directly to the inline C64xx instructions to speed up the C codes [12]. All instructions that are not easily expressed in C codes are supported as intrinsics. For example, the intrinsic operator “_abs” calculate the saturated absolute value.

4.2 Using Word Access to Operate on 16-Bit Data Stored in the High and Low Parts of a 32-Bit Register

In order to maximize data throughput, it is often desirable to use a single load or store instruction to access multiple data values consecutively located in memory. For example, C6x have instructions with associated intrinsics, such as “_add2()”, “_mpyh()”, “_mpylh()”, etc that operate on the 16-bit data stored in the high and low parts of 32-bit register [12]. When operating on a stream of 16-bit data, word accesses can be used to read two 16-bit values at a time, and then another C6x intrinsic is used to operate on the data. Ideally, we would like to get all units simultaneously operating on all individual instructions. This parallelism is still hard to achieve by the compiler and may still need hand coding in some cases.

4.3 Memory Management

The memory management becomes very important as the DSP has a small amount of fast internal memory. Using internal memory to store instructions and data helps in increasing the processing speed. Generally, each 4x4 block in a MB has 4 edges, then each pixel in 4x4 block may be read or updated four times before the 4x4 block is filtered completely. Since the pixels of a MB (256 luma and 128 chroma pixels) are accessed frequently during the filtering process, they are stored in the internal memory, leading thus to a reduction of memory access.

5 Experimental Results

To evaluate the effectiveness of the proposed optimised algorithm, the adaptive deblocking filter was implemented on DM6437EVM platform. A TI DM6437EVM development environment including target board and Code Composer Studio 3.3 profile tools was set up [13]. Furthermore, system level optimisation methods were adopted according to TI’s technical documentation [12].

The optimised ADF was ported to a DM6437EVM achieving a speed performance of 32% in comparison to a direct implementation (no optimisation). The performance of the optimizing approach has been measured on Foreman and Paris QCIF video sequences, using three different QP values. The performances, in terms of video quality, with DSP implementation are shown in the table 2.



Fig. 6. Performance of the deblocking filter for a highly compressed image (QP=38) (a) Reconstructed Foreman Image without Filtering. (b) Reconstructed Foreman Image with Filtering



Fig. 7. Performance of the deblocking filter for a highly compressed image (QP=38) (a) Reconstructed Paris Image without Filtering. (b) Reconstructed Paris Image with Filtering

Table 2. PSNR of non filtered and filtered Foreman and Paris images with QP=28,33,38

QP	Image	PSNR Image non filtered	PSNR Image Filtered
28	Foreman	33,92	34,78
	Paris	30,27	30,92
33	Foreman	32,44	33,07
	Paris	30,06	30,62
38	Foreman	31,69	32,21
	Paris	29,58	30,23

6 Conclusion

In this paper, a DSP based specific method to decrease the adaptive deblocking filter module complexity in the H.264/AVC encoder and decoder is proposed. The implementation of the adaptive deblocking filter on DM6437EVM DSP using code optimisation reduces the module cycles consumption by 32%. The losses, in terms of video quality, are minimal compared to the non-optimised implementation.

The results presented in this paper show that the same image quality, but with less computing time can be obtained through optimisation for a target specific implementation. As perspective, the specific optimisations of this module, exploiting the architecture features of the DM6437EVM will be carried out. This will allow a speeding up of the filtering process for real time applications.

References

1. Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology* 13(7), 560–576 (2003)
2. Draft ITU-T Recommendations and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC/14496-10 (E) AVC) (July 2004)
3. Zakhor, A.: Iterative procedures for reduction of blocking effects in transform image coding. *IEEE Transactions on Circuits and Systems for Video Technology* 2(1), 91–95 (1992)
4. Averbuch, A.Z., Schlar, A., Donoho, D.L.: Deblocking of Block-Transform Compressed Images Using Weighted Sums of Symmetrically Aligned Pixels. *IEEE Transactions on Image Processing* 14(2), 200–212 (2005)
5. List, A., Joch, A., Lainema, J., Bjontegaard, A., Karczewicz, M.: Adaptive Deblocking Filter. *IEEE Transactions on Circuits and Systems for Video Technology* 13(7), 614–619 (2003)
6. Lin, H.C., Wang, Y.J.: Cheng, K.T., Yeh, S.Y., Chen, W.N., Tsai, C.N., Chang, T.S., Hung, H.M.: Algorithm and DSP Implementation of H.264/AVC. In: ASPDAC, pp. 742–749 (2006)
7. Lin, H.C., Wang, Y.J., Cheng, K.T., Yeh, S.Y., Chen, W.N., Tsai, C.N., Chang, T.S., Hung, H.M.: SIP Approach for Implementation of H.264/AVC. *Journal of Signal Processing Systems* 50(1), 53–67 (2008)
8. Warrington, S., Shojania, H., Sudharsanan, S., Chan, W.Y.: Performance Improvement of the Deblocking Filter Using SIMD Instructions. In: ISCAS 2006 (2006)
9. Major, A., Nousias, I., Khawan, S., Milward, M., Yi, Y., Arslan, T.: H.264/AVC In Loop De-Blocking Filter Targeting a Dynamically Reconfigurable Instruction Cell Based Architecture. In: IEEE 2nd NASA/ESA Conference on Adaptive Hardware and Systems (2007)
10. Texas Instrument, The New TMS320C64x Architecture Enhancements over the TMS320C62x
11. Texas Instrument, TMS320DM6437 Evaluation Module Technical Reference (2006)
12. Texas Instrument, TMS320C6000 Programmer Guide (2001)
13. Texas Instrument, TMS320C6000 Code Composer Studio Tutorial (1999)

Methodology for Acoustic Characterization of a Labial Constraint in Speech Production

Leila Falek^{*}, Hocine Teffahi, and Amar Djeradi

USTHB, Electronics and Computer science Faculty, Algiers, Algeria
lfalek@hotmail.fr

Abstract. We propose in this study, a method allowing us to characterize a speech occurred in a stress situation. For this, we created an artificial disturbance (stress lip) and then, we analyzed the effects of stress on the acoustic parameters of the signals produced. We have developed a methodology allowing us to analyze the timing, the fundamental frequency, formants frequencies (calculated with wavelet transform methodology) and the coarticulation between consonant and vowel. These parameters are generally used to produce vector descriptors in communication systems. Articulatory interpretation of results typifies well the constraint used.

Keywords: speech disturbance, acoustics parameters, labial constraint, formants frequencies, fundamental frequency, coarticulation, locus equation.

1 Introduction

At present, in an effort to get as close as possible to a natural speech, recognition systems, coding or speech synthesis are in search of invariants in the acoustic cues of natural speech taking into account all the constraints under which useful information is produced. Also, much research has emerged in the study of speech artificially disturbed [1]. These studies provide a better understanding of motor control in a constraint situation and thus demonstrate the effect of stress on these acoustic parameters to better use in communication systems.

Moreover, it is possible to disturb speech production in several ways: either directly or indirectly by modifying the vocal tract geometry, or by interrupting some kind of feedback (tactile, auditory or other). We were interested here at the vocal tract geometry constraint [1]. (The geometric perturbation can mean an oral handicap, or when a person speaking with an object in the mouth).

A tube inserted between the lips of four speakers served as a perturbation of the lips protrusion, during the production of 7 French vowels. The aim is then to analyze the effects of this disturbance on the acoustic parameters of recorded sounds, by analyzing the compensatory strategies adopted by the speakers for better adaptation to the stress, for reach their target sound.

^{*} Corresponding author.

With Using the experiments of artificial perturbation of speech production, McFarland et al. (1996)[8] have revealed a significant difference : the vowels production is more affected by a change of oral function. The behavior of the vowels under stress

then is revealing information about motor control of speech. This justifies vowels analysis of in this study.

2 Experimental Setup for Labial Constraint Realization

The aim is to constrain the lip area, so we chose to use a lip ring diameter 3.4 cm and 1 cm in length (rigid), so as to compel the production of all vowels: it will allow for increase the lips area in the production of rounded french vowels [u] and [ou], and lower lip area in the production of [i], [a], [e] [é]. Although the variation in the lip area is the desired result, the insertion of a tube between the labial lips of the speaker also has the effect of causing a slight increase in the lips protrusion, as the speaker must use his lips to completely surround the tube, so that it can't fall and there is no air flow around the tube.

3 The Sentences Corpus

We conducted a corpus of french sentences composed of sequences consonant-vowel-consonant (CViC), which were incorporated into an interrogative carrier sentence for preserve the natural speech: nVina est la? With Vi = [a], [i], [ou], [o], [u], [e] and [é]. The consonant /n/ is chosen because it is a voiced consonant, to facilitate the acoustic analysis. The sentences of the corpus are then: “Nana est la ? “; “ Nina est la ?” ; “ Nouna est la ? “; “ Nena est la ?” ; “ Néna est la ?”; “ Nona est la ?”; “Nuna est la ?”. Each sentence is repeated seven times. These sentences were recorded initially without constraint and with constraint. The number of speakers is 4 (3 women and 1 man). What makes a corpus of: $7 \times 10 \times 4 \times 2 = 560$ sentences. The recording of the corpus was conducted in a calm, using a desktop computer with a sound card "Sound Blaster" , the Praat software and a microphone.

4 The Units Analyzed

The units in question are /CViC/, Vi is one of the seven french vowels (/a/, /i/, /ou/, /o/, /u/, /e/, /é/), in consonantic context /C/= /n/, for each speaker : /nan/, /nin/, /nen/, /nun/, /nén/, /non/, /noun/. to obtain these units, we manually segmented the corpus of sentences using Praat software. The analysis was done in the time domain and frequency domain.

In the time domain: we analyzed the durations (CViC) depending with the duration of the sentence, (d(CViC): average length of the unit CViC and d(sentence): average

duration of the sentence , corresponding to the unit: CViC). The aim is to study the constraint impact on the temporal distribution in the analyzed sentence.

In the frequency domain: we analyzed the formant frequencies F1 and F2 and the fundamental frequency F0 of the vowel "Vi" (taken in the vowels middle).

We then analyzed the influence of the constraint on vowel space and on the coarticulation (using the slope of the straight locus equations)

5 Analysis of Labial Constraint Influence on the Timing of "VCiV"

The comparison between the normal condition and in constraint condition is made on dispersions ellipses basis of average durations (d(CVC), d(sentence)) of each vowel , for each speaker, to take account of all samples. The abscissas of dispersion ellipses correspond to the average durations of sentences and the ordinates to length units VCiV. In blue color we have the situation without constraint and in red color with stress.

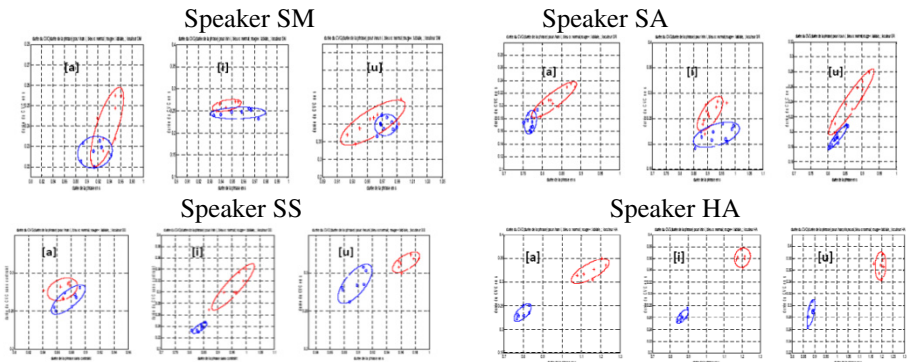


Fig. 1. Dispersion ellipses of average lengths (d(CViC), d(sentence)) of each vowel for each speaker

Figures 1 show an increase in the duration of VCiV in the sentence, for four speakers. So, the constraint has the effect of slowing the speech rate

6 Constraint Influence Analysis on the Vocalic Space

We calculated the formants frequencies F1 and F2 in the middle of the vowel, using a calculation method of formant trajectories, based on the complex continuous Morlet wavelet transform that we have developed.

6.1 Methodology for Calculating Formants Based on Complex Continuous Wavelets Transform

The wavelet transform is to decompose a signal $x(t)$ in a family of functions $\psi(t)$, localized in time and frequency called wavelets. It is defined by the relation [5 13 19]

$$(W_\psi x)(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \tag{1}$$

It is used to describe the frequencial content of a signal $x(t)$, locally near a point (a, b) , in the time-scale ("a" scale and "b" is the time). It indicates us the relative importance of frequency "1/a" around time "b" for the signal $x(t)$.

The complex continuous wavelet transform is used to define the notion of instantaneous frequency (and formants), when using an analytic wavelet [17]. The analytic signal is a complex signal associated with a real signal, which provides access to the signal phase, therefore, at its frequency.

It was proposed by VILLE in 1948 [20]. This latter has interesting properties, particularly in terms of its Fourier transform, which is zero for negative frequencies.

An analytic signal $z(t)$ can be calculated from the Hilbert transform [10] of a real signal $x(t)$ such that:

$$H(x(t)) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(s)}{t-s} ds \tag{2}$$

$$z_x(t) = x(t) + iH(x(t)) \tag{3}$$

$H(x(t))$ is the Hilbert transform of $x(t)$.

It is possible to define [6] from an analytic signal, an instantaneous frequency $f_x(t)$ [7], where $z_x(t)$ is an analytic signal

$$f_x(t) = \frac{1}{2\pi} \frac{d \arg(z)}{dt} (t) \tag{4}$$

6.1.1 The Morlet wavelet

It is an analyzing wavelet [15] for small oscillations (a center frequency f_c : around 1 Hz). It is very well localized in time. It is inspired by the Gabor elementary signal. It is obtained by modulation of a Gaussian. It is given by the following equation [3]:

$$\Psi(t) = \frac{1}{c} e^{j\omega_c t} \left[e^{-\frac{t^2}{2\sigma_t^2}} - \sqrt{2} e^{\frac{\omega_c^2 \sigma_t^2}{4}} e^{-\left(\frac{t^2}{\sigma_t^2}\right)} \right] \tag{5}$$

The product $\omega_c * \sigma_t$ fixed the link between the width of the Gaussian envelope of the wavelet and its oscillation frequency f_c . For the Morlet wavelet $\omega_c = 2\pi f_c$: $5 \leq \omega_c \leq 6$, so, $\omega_c * \sigma_t \geq 5$, $0.8 \leq F_c \leq 1$) Hz [3]

6.1.2 Presentation of the Method

The method we developed was made from some properties of time-frequency distribution and wavelet transform, namely: For a complex time-frequency distribution, there is information in the module and phase of the distribution.

Near the center frequency f_c of wavelets, whose cyclicity corresponds to the nature of the signal: the wavelet transform amplitude has a maximum [4]. The phase coefficients of the time-frequency distribution varies cyclically at a frequency close to the instantaneous frequency of the signal, therefore the derivative of the phase coefficient is close to the instantaneous frequency of the signal. So, we will say that for a given time, the instantaneous frequency can be estimated by: the maximum of the distribution module, the fixed point of the time derivative of the phase of time-frequency distribution [6] [19]

For formants calculation, it is important to add that in this case, the phase derivative alone is not sufficient since the formant frequencies correspond to the instantaneous maximum power. From this information we developed an algorithm for calculating the instantaneous frequencies of formants contained in a speech signal [10].

The following figures show the different steps of the method developed, applied to a signal $x(t)$ composed of a sum of four sinusoids:

$$x(t)=5\sin(2\pi*300t)+10\sin(2\pi*1000t)+ 15\sin(2\pi*3000t)+ 10\sin(2\pi*5000t)$$

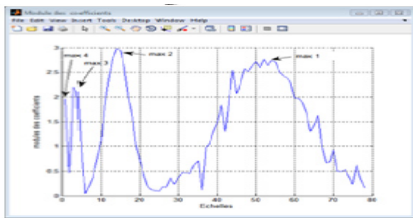


Fig. 2. modulate coefficients of the wavelet transform according to the scales

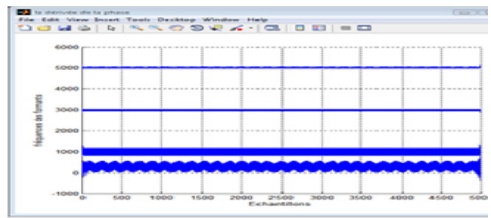


Fig. 3. Phase derivative of wavelet transform corresponding for each maximum

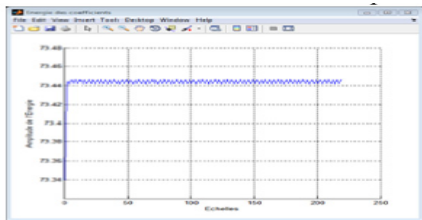


Fig. 4. Coefficients energy of the wavelet transform, for one frequency

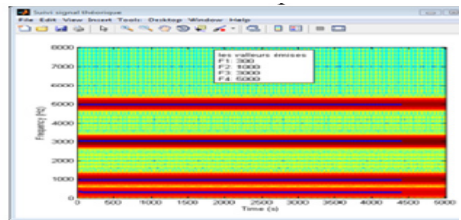


Fig. 5. Maximum energy for the 4 instantaneous frequencies

The results obtained in Figures 2, 3, 4 and 5 show that the method used to calculate the instantaneous frequencies allows to find the frequencies of the analyzed signal $x(t)$.

Using this method, we have determined the formant values F1 and F2 in order to trace the vocalic spaces corresponding, in normal condition and in labial constraint, for the 10 samples, for each vowel and for each speaker. We then represented each vowel by its dispersion ellipse in the vowel space. We then plot surfaces of the vowel triangles [7] in the vocalic space (F1, F2) for each speaker in two conditions: normal condition and in labial constraint. Figure 6 shows results obtained for each speaker, and table 1 shows effects of stress on areas of vocalic triangles.

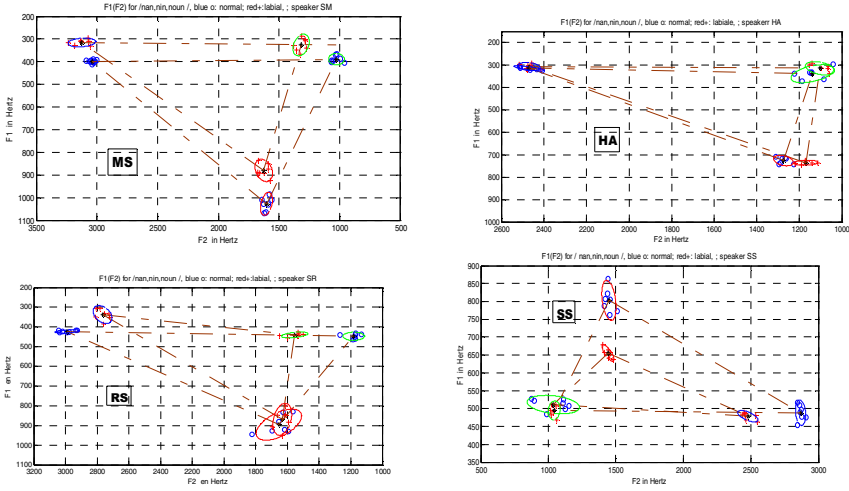


Fig. 6. Dispersion ellipses and vocalic triangle corresponding to vowels [a], [i] and [u] for each speaker HA, SS, RS and MS. In normal condition: in blue color and labial constraint: in red color.

Table 1. Areas of vocalic triangles by speaker (In normal condition and in labial constraint)

speaker	Area(Hz ²) Normal condition	area(Hz ²) With constraint
HA	259960	288350
SS	282060	110570
RS	411330	263430
MS	639750	502090

Results obtained in Table 1 show a clear vowel reduction caused by the stress for all speakers. This is consistent with results obtained by Lane et al. [13] and other authors who have noticed that the size of the vocalic space was reduced in case of disturbance of speech. However, according Gendrot et al. [12], in case of a good motor control, reduction of the vowel space is carried out according to a centralization of vowels in the vowel space. This would mean that speakers reach their acoustics targets (thus remains intelligible despite the constraint).

Figure 7 shows a centralization case. The vowels are represented by their dispersion ellipses (corresponding to several realizations of the same vowel, by the same speaker) in the vocalic space (F2, F1). Results show that in articulatory terms, the vowel [i] become less anterior and more open, while the vowel [a] would be slightly less posterior and a little less open. These changes would influence the formant values: for the [i] vowel, an increase in the formant F1 and a decrease in the formant F2; for the vowel [a], an increased of F2 and decreased of F1. This is called

centralization. The ellipses move toward to the center of the vocalic triangle in constraint situation. It would be interesting for us to see if our speakers centralize their vowels.

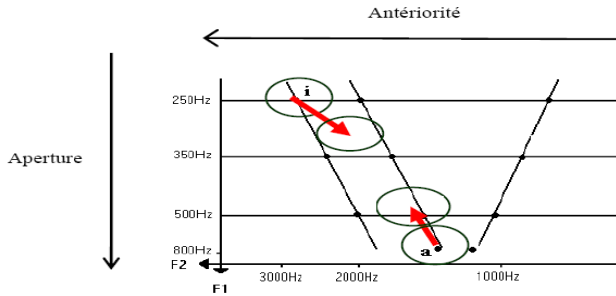


Fig. 7. Case of vowels [i] and centralization in vocalic space [21]

So, we studied the variation direction of formants F1 and F2 with the constraint for the three vowels [a], [i] [u], for four speakers. Results obtained showed that most speakers do not centralize. However, centralization is a sign of good motor control. This would mean that the stress is very important and the speakers have their motor control which has decreased. So the acoustic targets are not well met. (This means that the sounds are barely understandable).

7 Labial Constraint Influence Analysis on Intonation F0 and on Mouth Aperture

We have determinate the fundamental frequency and the formant F1 in the middle of the three vowels, in the normal case and in labial constraint, for the seven samples for each speaker. As F1 correspond to the mouth opening and F₀ to the intonation, we chose to represent the intonation according to the aperture to see the behavior of the speaker facing the stress in F₀ (F1). As we have 7 samples per vowel we have represented the F₀ and F1 by their ellipses of dispersion. The results obtained are regrouped in Figures 8.

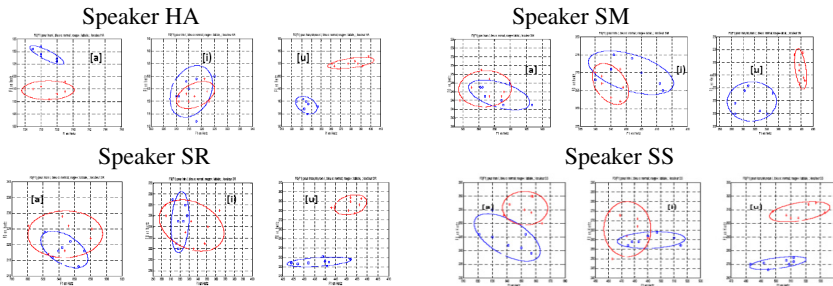


Fig. 8. F₀(F1) variation for each vowel and for each speaker

Figures 4 show in overall a small increase of F_0 for vowels [a] and [i] with very little variation of F_1 (since the lips are already stretched, the speaker has no difficulties to pronounce these two vowels). However, for the vowel [u], there is a considerable increase in F_0 and F_1 simultaneously with the constraint. These results appear right, as the formant frequency F_1 is related to the aperture and F_0 for the intonation: so, the speaker seems to make a major effort by increasing its F_0 for trying to reduce its aperture to say the vowel [u], since this one requires almost closing the mouth to be uttered (which justifies the increase in F_1 since the constraint prevents mouth from closing). Vowel [u] is then pronounced as a vowel [o], thus increasing the F_1).

8 Analysis Effects of the Constraint on the Coarticulation Degree

The coarticulation occurs when a vowel sees its characteristics change because of its phonetic environment. The experiences of disruption of speech have shown that in presence of constraint, the coarticulation increases. This is manifested by a vocalic reduction. An example of coarticulation can be observed when the consonant is followed by a rounded vowel: there is anticipation of the borough of the vowel during the consonant (see Figure 9). What changes acoustically consonant

Spectral characteristics of [k] vary depending on the nature of the following vowel. The main information of coarticulation is carried by the transition between the consonant and the vowel (or vis versa), surrounding the value of the form F_2 . The F_{2onset} variation (in the beginning of the vowel, in the CV transition) compared to the $F_{2stable}$ (middle of the vowel) is then a source of information for coarticulation. Thus, in type sequences VCV or CVC, for example, the vowel or consonant of the middle is imbued with the surrounding phonemes. The degree of coarticulation is usually estimated by the slope of the straight locus. The concept of the locus equation has been developed by Lindblom in 1963[15]. Lindblom has found that there is a linearity in the second formant frequency variation at limit time of the beginning transition of the consonant to the post-consonantal vowel (F_{2o} or F_{2onset}), according to the second formant frequency target variation of this vowel. Lindblom has established the locus equations from a study of sequences CV.

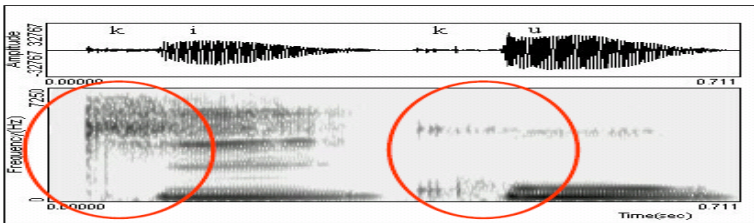


Fig. 9. Anticipation example in the production of CV /ki/ and /ku/

The locus equation is written as: $F2o = k * F2s + c$, where "c" and "k" are constants. In case $k = 0$, the equation is constant, $F2o = c$, which would imply that the production of sequence CV is a simple articulatory chaining where the coarticulation between the consonant and the vowel is virtually non-existent. In case where $k=1$ and $c = 0$, the coarticulation with the vowel context is considered maximal.

In this study, we have plotted the straight lines of locus equations: $F2_{onset}(F2_{satable})$, with considering seven French vowels, for all CVi (C: consonant [n] and Vi: one of the seven French vowels considered), for all speakers, in normal condition and with constraint, on the same figure. Each vowel is represented by its dispersion ellipse. Results obtained are illustrated by figures 10.

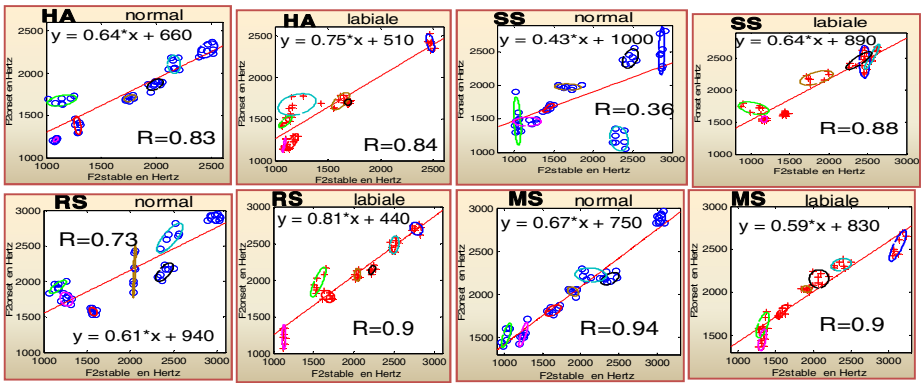


Fig. 10. Straight lines of locus by speaker (in normal condition and with labial constraint)

Locus equation is written: $F2_{onset} = m * F2_{stabil} + b$ where m and b are the slope and intercept respectively. R is the regression coefficient. The results obtained are summarized in the following table:

Table 2. Parameters of straight lines of locus equations obtained

speakers	Parameters of straight lines of locus equations					
	Without constraint			With constraint		
	m	b	R	m	b	R
HA	0,64	660	0,83	0,75	510	0,84
SS	0,43	1000	0,36	0,64	890	0,88
RS	0,61	940	0,73	0,81	440	0,9
MS	0,67	750	0,95	0,59	830	0,9

This table shows that in labial constraint, we have a net increase in slopes of the straight lines of locus, so of the coarticulation

9 Conclusion

We studied in this work the influence of a labial constraint in speech production. We acoustically analyzed the behavior of the vowels in constraint, then the influence of the constraint on the coarticulation. The results obtained show a great influence of the constraint on the acoustic parameters in particular, the frequential parameters. They can be summarized such as:

- In the temporal field: - temporal increase, therefore the speaker tends to slow down his speech flow for better controlling his elocution rate with the constraint.
- In the frequential field: vocalic spaces are reduced but without vowels centralization. Thus the speakers reach with difficulty their acoustic targets with the constraint or then of the whole.
- In variation of $F_0(F_1)$: There is a small increase of F_0 for vowels [a] and [i] with very little variation of F_1 but for the vowel [u], there is a considerable increase in F_0 and F_1 simultaneously with the constraint
- With straight lines of locus equations, results shows that we have a net increase of coarticulation with the constraint.

All these results show that, the evaluation methods of the acoustic parameters we propose, accurately characterize the constraint used.

References

- [1] Aubin, J.: Effets acoustiques et articulatoires des perturbations labiales sur la parole des enfants et des adultes. Université du Québec à Montréal (Décembre 2005)
- [2] Castellengo, M., Dubois, M.: Timbre ou timbres? Propriété du signal, de l'instrument ou construction cognitive (2005) 88
- [3] Chui, C.: An introduction to wavelets. Academic Press (1995)
- [4] Cnockaert, L.: Analyse du tremblement vocal et application à des locuteurs parkinsoniens. Thèse de doctorat en sciences de l'ingénieur. Université libre de Bruxelles ULB, Bruxelles (Décembre 2007)
- [5] Cohen, L., Lee, C.: Instantaneous frequency and time-frequency distributions. In: IEEE International Symposium on Circuits and Systems, pp. 1231–1234 (1989) 11
- [6] Emiya, V.: Spectrogramme d'amplitude et de fréquences instantanées (safi). Master's thesis, Aix-Marseille II (2004) 52
- [7] Hirsch, F., Ferbach-Hecker, V., Fauvet, F., Vaxelaire, B.: Étude de la structure formantique des voyelles produites par des locuteurs bègues en vitesses d'élocution normale et rapide. Jep (2006)
- [8] Mc Farland, D.H., Baum, S.R., Chabot, C.: Speech compensation to structural modifications of the oral cavity. Journal of the Acoustical Society of America 100(2), 1093–1104 (1996)
- [9] Ferbach-Hecker, V.: La résistivité de la qualité des voyelles orales du français. In: SCOLIA, vol. 20, pp. 115–134 (2005)
- [10] Flandrin, P.: Temps-fréquence. Hermès (1993) 11, 27, 29, 3.

- [11] Jones, J.A., Munhall, K.G.: Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America* 113(1), 532–543 (2003)
- [12] Gendrot, C., Adda-Deker, M.: Analyses formantiques automatiques de voyelles orales : évidences de la réduction vocalique en langues française et allemande. In: MIDL (2004)
- [13] Lane, H., Denny, M., Guenther, F.H., Matthies, M., Ménard, L., Perkell, J.S., Stockmann, E., Tiede, M., Vick, J., Zandipour, M.: Effects of bite blocks and hearing status on vowel production. *Journal of the Acoustical Society of America* 118(3), 1636–1646 (2005)
- [14] Lardiès, J., Ta, M.N., Berthillier, M.: Modal parameter estimation from output-only data using the wavelet transform. *Archive of Applied Mechanics* 73, 718–733 (2004)
- [15] Lindblom, B.: *Economy of Speech Gestures. The Production of Speech*. Springer, New York (1983)
- [16] Mallat, S.: *Une exploration des signaux en ondelettes*, Editions de l'école Polytechnique (2000)
- [17] Navarro, L.: Analyse temps-fréquence de signaux vibratoires issus d'un réacteur de culture osseuse. *Journée de la Recherche de l'EDSE* (2007a) 143
- [18] Rihaczek, A., Bedrosian, E.: Hilbert transforms and the complex representation of real signals. *Proceedings of the IEEE* 54(3), 434–435 (1966) 11
- [19] Torrèsani, B.: *Analyse continue par ondelettes*, Editions du CNRS, Paris (1995)
- [20] Ville, J.: *Théorie et applications de la notion de signal analytique*. Câbles et Transmissions 1, 61–74 (1948) 11,13
- [21] Calabrino, A.: Effets acoustiques du débit sur la production de la parole chez des locuteurs enfants et adultes. *Actes du Xe Colloque des étudiants en Sciences du Langage* 61–83 (2006)
- [22] Gendrot, C., Adda-Deker, M.: Analyses formantiques automatiques de voyelles orales: évidences de la réduction vocalique en langues française et allemande. In: MIDL (2004)

Performance of OFDM in Radio Mobile Channel

Mohamed Tayebi and Mrahi Bouziani

Tayebi_med@hotmail.com

Abstract. The OFDM has emerged in the second half of last century. However, it remains a technique used in broadband wireless communication systems. Which limits its use is its vulnerability to frequency shifts. The Doppler effect and imperfections of the local oscillators of the transmitter and receiver are at the origin of these shifts. The imperfections of local oscillators shift the spectrum of the OFDM signal, while the Doppler-effect expands it (or compresses it). In this paper, the impact of the Doppler-effect on the OFDM modulation is analyzed, we give a new expression of the contribution of each sub-carrier at the transmitter on each subcarrier demodulated at the receiver. Using the method of inter-carrier interference self-cancellation proposed by Zhao and Haggman, the system becomes more efficient.

Keywords: Orthogonal frequency division multiplexing (OFDM), carrier frequency offset (CFO), Doppler effect, Inter carriers Interferences (ICI), carrier to interferences ratio (CIR).

1 Introduction

The high spectral efficiency and robustness against multipath make OFDM one of the most promising techniques in wireless communications systems. Currently, OFDM has been adapted to the digital audio and video broadcasting (DAB/DVB) system, high-speed wireless local area networks (WLAN) such as IEEE802.11, HIPERLAN II, ADSL, and recently in the optical communication [1]. OFDM divides the available spectrum into N equal intervals, the symbols are sent in parallel channels with lower flow rates [2]. The high mobility is a major challenge for wireless communications. It creates frequency shifts in part caused by the Doppler effect [3]. These offsets destroy the orthogonality between subcarriers of the OFDM signal, and generates inter-carrier interference responsible for the degradation of system performance [4]. A number of studies has been developed to eliminate inter-carrier interference [5],[6], we will be interested by the method of Inter-carrier interferences self-cancellation for the first time proposed by Zhao and Haggman [6]. In this paper we study the superposition of the offset caused by the imperfections of local oscillators of the transmitter and receiver, and the Doppler-effect and its impact on all the subcarriers that constitute the OFDM signal. To improve system performance, we used the method of ICI self-interference [6]. This work was structured as follows, section 2 describes the nature of the frequency shift. The third topic focuses on the nature of the interference generated. The fourth section

gives the system performance in terms of CIR and we apply an existing method for eliminating interference[6]; and last we finalize with a conclusion.

2 Frequency Shifts in a Mobile Radio Channel

In this section we will introduce the impact of the Doppler-effect on all the subcarriers that constitute the OFDM signal. We will assume that the local oscillators of the transmitter and receiver do not wobble at the same frequency. At the transmitter, the frequencies are equal to:

$$f_k = \frac{k}{T} \quad k = 0, 1, \dots, N - 1 \quad (1)$$

Where k represent the index of subcarriers. Then at the receiver, frequencies are equal to:

$$f_k = \frac{k}{T} + \Delta f \quad k = 0, 1, \dots, N - 1 \quad (2)$$

Where N is the total number of subcarriers and Δf the frequency difference between transmitter and receiver. More if the receiver moves to the transmitter at a relative speed v , the Doppler-effect is manifested and the frequency received signal will be equal to:

$$f_k = \left(\frac{k}{T} + \Delta f \right) \left(1 \pm \frac{v}{c} \cos \alpha \right) \quad (3)$$

Where c is the speed of light, and α the angle between the velocity and direction of the electromagnetic wave. Then f_k can be written:

$$f_k = \frac{1}{T} (k + \varepsilon)(1 + \beta) \quad (4)$$

With

$$\beta = \pm \frac{v}{c} \cos \alpha \quad (5)$$

Where ε is the normalized frequency offset due to the imperfections of local oscillators. The total normalized offset δ will be equal to:

$$\delta(k) = \beta k + \varepsilon(1 + \beta) \quad (6)$$

δ is proportional to the index k of the subcarrier. In other words, the offset is different for each subcarrier, we are thus faced with a compression of the spectrum of OFDM signal if β is negative, or otherwise, with an expansion of the spectrum in if β is positive. For a given connection, the imperfections of the oscillators generate a constant normalized offset ε , while the Doppler effect produces a shift proportional to the index of the subcarrier k .

3 Inter-carrier Interferences

In the time domain, the OFDM signal is written as:

$$x(t) = \sum_{k=0}^{N-1} X(k) \exp(j2\pi f_k t) \quad (7)$$

Where $X(k)$ is the signal to be transmitted and f_k represents the N different frequencies of sub-carriers that form the OFDM signal. By discretizing the signal $x(t)$, the signal obtained at the receiver is then:

$$y(n) = \sum_{k=0}^{N-1} X(k) \exp\left(j2\pi \frac{n}{N} (k + \varepsilon)(1 + \beta)\right) \quad (8)$$

$$= \exp\left(j2\pi \frac{n\varepsilon}{N} (1 + \beta)\right) \times \sum_{k=0}^{N-1} X(k) \exp\left(j2\pi \frac{nk}{N} (1 + \beta)\right) \quad (9)$$

The signal $y(n)$ is out of phase with the original signal, this is due to imperfections of local oscillators. The signal undergoes an expansion (or compression) of the signal, since the Doppler-effect affects each subcarrier differently. At the end of the FFT, we get the symbols $Y(k)$, their expression is given by the equation:

$$Y(k) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} X(k) \exp\left(j2\pi \frac{n}{N} (k - m + k\beta + \varepsilon(1 + \beta))\right) \quad (10)$$

The complex coefficients will be expressed as:

$$S_{km}(k - m) = \frac{\sin\pi(k - m + k\beta + \varepsilon(1 + \beta))}{N \sin \frac{\pi}{N} (k - m + k\beta + \varepsilon(1 + \beta))} \times \exp j\pi \left(1 - \frac{1}{N}\right) (k - m + k\beta + \varepsilon(1 + \beta)) \quad (11)$$

The amplitudes of the complex coefficients have different values for positive and negative relative velocities. The influence of the subcarrier m on subcarrier k is different. Figure 1 shows the variation of the amplitudes of complex coefficients for a fixed value of the normalized frequency offset ε , but with three values of relative speed β , the first zero, the second positive and the last negative.

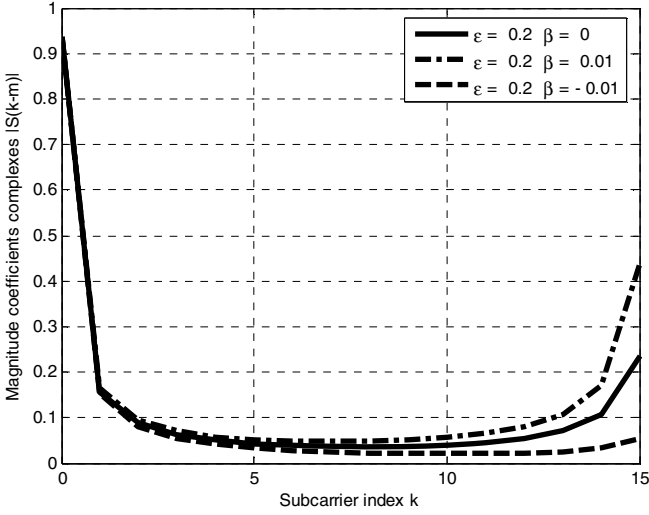


Fig. 1. Plot of the amplitudes of complex coefficients as a function of the index k of subcarriers for different values of relative speed

4 System Performances

The useful signal, is assigned the complex coefficient $S_{kk}(0)$ calculated for the indices $k = m$:

$$S_{kk}(0) = \frac{\sin \pi(k\beta + \varepsilon(1 + \beta))}{N \sin \frac{\pi}{N}(k\beta + \varepsilon(1 + \beta))} \exp j\pi \left(1 - \frac{1}{N}\right)(k\beta + \varepsilon(1 + \beta)) \quad (12)$$

The complex coefficients of the different sub-carriers are not the same magnitude, since they depend on the index of the subcarrier k . Figure 2 shows the degradation of the useful signal as a function of the index k . Similarly, we will study all the interference caused by the subcarriers index $k \neq m$. The power of interference σ_{ICI} is given by the relation:

$$\sigma_{ICI} = \sum_{\substack{k=2 \\ k \neq m}}^{N-1} |S_{km}(k - m)|^2 \quad (13)$$

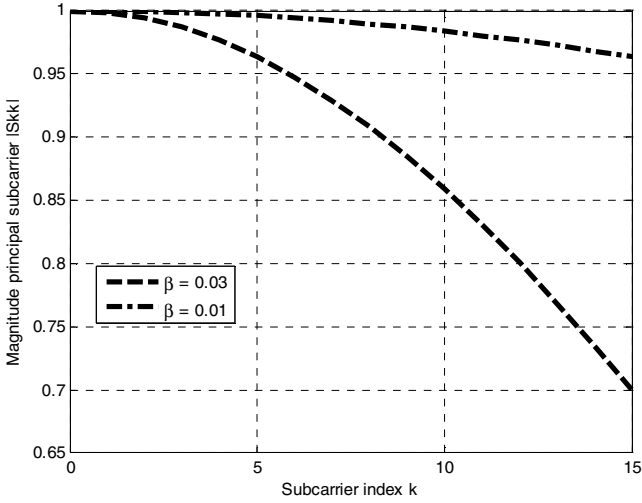


Fig. 2. Amplitude of the subcarrier based on the main index subcarriers k

This power is calculated for different values of m . To study the system performance, we calculated the ratio of carrier power on the power of all interference, it is noted by CIR. In the case studied, the frequency shift is proportional to the index of the subcarrier, which implies that the CIR is different for different subcarriers. Its expression is given by the relation:

$$CIR_{km}(\varepsilon, \beta) = \frac{|S_{km}(0)|^2}{\sum_{\substack{k=2 \\ k \neq m}}^{N-1} |S_{km}(k-m)|^2} \tag{14}$$

Figure 3 shows that for two different sub-carriers, the difference can reach 10dB. To improve system performance, we use the method of ICI self-cancellation. The parallel data streams are remapped as the form of :

$$X(1) = -X(0), X(3) = -X(2), \dots, X(N-1) = -X(N-2) \tag{15}$$

The CIR is given by the relation:

$$CIR_{km}(\varepsilon, \beta) = \frac{|2 * S_{kk}(0) - S_{kk}(1) - S_{kk}(-1)|^2}{\sum_{\substack{k=2 \\ k \neq m \\ k \text{ even}}}^{N-1} |2 * S_{km}(k-m) - S_{km}(k-m+1) - S_{km}(k-m-1)|^2} \tag{16}$$

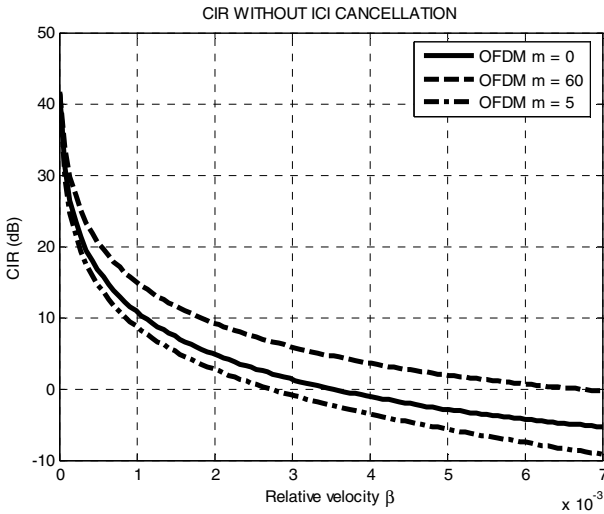


Fig. 3. Plot of the CIR as a function of relative velocity for different values of m

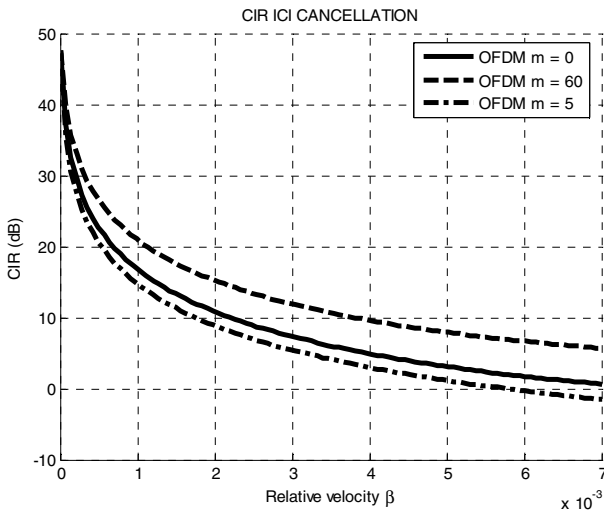


Fig. 4. Plot of CIR with ICI cancellation as a function of the speed for different values of m

Figure 3 plots the performance without using the method of ICI self-cancellation. However, Figure 4 plots the performance using the cited method. The gain varies from 7 to 10 dB for different subcarriers.

5 Conclusion

In this paper, we analyzed the performance of OFDM systems in a mobile radio channel. This analysis enabled us to lead to new mathematical expressions. Indeed, the Doppler effect present in the mobile radio channels, shifts not only the carrier frequency, but extended (or compress) the spectrum of OFDM signal because it shifts each subcarrier differently. So we calculated the performance for each subcarrier. Applying the method of ICI cancellation, we improved the system performance. This improvement has resulted in a gain ranging from 7 to 10 Db.

References

1. Armstrong, J.: OFDM for Optical communications. *Journal of Lightwave Technology* 27(3) (February 1, 2009)
2. Weinstein, S., Ebert, P.: Data transmission by frequency-division multiplexing using the discrete Fourier transform. *IEEE Trans. Commun.* 19, 628–634 (1971)
3. Russell, M., Stüber, G.L.S.: Interchannel interference analysis of OFDM in a mobile environment. In: *Proc. VTC 1995, Chicago, IL*, pp. 820–824 (July 1995)
4. Sathanathan, K., Tellambura, C.: Performance analysis of an OFDM system with carrier frequency offset and phase noise. In: *IEEE Vehicular Technology Conference, VTC 2001 Fall, Atlantic City, NJ, USA, October 7-11*, vol. 4, pp. 2329–2332 (2001)
5. Zhao, Y., Haggman, S.-G.: Intercarrier interference self-cancellation scheme for OFDM mobile communication systems. *IEEE Trans. Commun.* 49, 1185–1191 (2001)
6. Ryu, H., Li, Y., Park, J.: An Improved ICI Reduction Method in OFDM Communication System. *IEEE Transactions on Broadcasting* 51(3) (September 2005)

Spatial Correlation Characterization for UWB Indoor Channel Based on Measurements

H. Chaibi¹, R. Saadane², My A. Faqih¹, and M. Belkasm¹

¹ Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat,
Suissi, Rabat, Maroc

² SIR2C2S/LETI-EHTP, Ecole Hassania des Travaux Publiques, Casablanca Maroc
{has.chaabi, rachid.saadane}@gmail.com,
{faqih, belkasm}@ensias.ma

Abstract. The major aim of this paper is to present a statistical model for small scale fading and spatial correlation evaluation for indoor UWB channel. A some distributions are suggested for small scale fading statistics where the number of scatterers is unknown. Also, the impact of a number of system parameters on spatial correlation at the receiver is evaluated. The complex correlation coefficient decays slowly with distance under line-of-sight, but decreases quickly under non Line of sight.

Index Terms: UWB Channel Characterization Modeling, Measurements, Small Scale Statistics, Amplitude and Phase Distribution.

1 Introduction

Due to current developments in digital consumer electronics technology, Ultra Wide-band (UWB) is becoming extra attractive for low cost personal communication applications. UWB systems are now emerging across a diversity of commercial and military applications, including communications, radar, geolocation, and medical. First generation commercial wireless UWB creations are anticipated to be extensively deployed almost immediately. This has been fueled by a command for high frequency consumption and a large number of users requiring simultaneous multidimensional high data rate access for applications of wireless internet and e-commerce [2-6, 10, 11].

UWB systems are often defined as systems that have a relative bandwidth larger than 25 % and/or an absolute bandwidth of more than 500 MHz (FCC). The UWB systems using large absolute bandwidth, are robust to frequency-selective fading, which has significant implications on both design and implementation [10]. Among its significant characteristics, the UWB technology are low power devices, exact localization, high multi-path resistance, low complexity hardware structures and carrier-less architectures [8, 9, 14]. As well, the spreading of the information over a very huge frequency range decreases the spectral density and makes UWB technology deployment compatible with existing systems.

As first step in for designing and implementing any wireless communication system, channel measurements and modeling are a basic necessity. Several theoretical and

practical studies, have shown an extreme difference with respect to narrow-band channels [15]. In the area of UWB channel modeling, the researchers are interested to characterize the path loss law, shadowing, multi-path delay spread, coherence bandwidth, average multi-path intensity profile and received amplitude distribution of the multi-path components... But, there is no universal UWB channel model proposed.

The contribution of this paper is a simple modeling of small scale fading and simple empirical model relating correlation coefficient to distance of UWB indoor channel. The propagation scenarios deal with both Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) situations. We have assumed two hypotheses, the first one is that the indoor channel is considered to be time invariant because the transmitter and the receiver are static and no motions take place in the channel. The second one is that the signal excitation is assumed to be close to an ideal Dirac-Delta impulse which means that the received signal can be seen as a good approximation of the channel impulse response.

The rest of the paper is organized as follows. In Section 2 presents our measurement specification and set-up. Section 3 presents channel model and Statistical distributions. The Section 4 describes our spatial correlation channel analysis and results. Finally, conclusions are provided in Section 5.

2 Measurement Specification and Set-Up

Measurements are performed at spatially different locations under both Line of Sight (LOS) and Non Line of Sight (NLOS). These are carried out in Eurecom Mobile Communication Laboratory, which has a typical laboratory environment (radio frequency equipment, computers, tables, chairs, metallic cupboard, glass windows,...) with plenty of reflective and diffractive objects, rich in reflective and diffractive objects. For the NLOS case, a metallic plate is positioned between the transmitter and the receiver [6]. We have complete database of 4000 channel frequency responses corresponding to different scenarios with a transmitter-to-receiver distance varying distance varying from 1 meter to 14 meters. The Electrical specifications of the used antennas in this work are [12]:

- Frequency Range: 3.1 – 10.0 GHz.
- Gain: 4.4 dBi peak at 4.5 GHz.
- VSWR¹ < 2.0 : 1 : across 3.6 – 9.1 GHz.
- Polarization: Linear.
- Radiation Pattern: Azimuth Omni-directional.
- Feed Impedance: 50 Ω (Ohms) Unbalanced.

The Mechanical Specification [12]:

- Antenna Element: 0.54 × 0.63 × 0.12 in 13.6 × 16.0 × 3.0 mm
- Assembly PCB: 1.03 × 0.73 × .04 in 26.2 × 18.5 × 1.0 mm
- Area of PCB that is Ground: 0.48 × 0.73 in 12.2 × 18.5 mm
- Antenna Element Weight: 0.5 g

¹ Voltage Standing Wave Ratio.

The antennas presents a VSWR varying from 2 to 5 for example an efficiency about 82% at 5.2 GHz. From the Fig. 2 in [6] we see a plate response over frequency range, this a very important characteristics of UWB antenna, we recommend reference [13] to learn about analysis of the antennas impact on the channel measurements.

2.1 Configuration and Set-Up

In our spatial correlation analysis we have used a virtual antenna arrays at the receiver, with a fixed transmitter. A manual controlled positioning grid is used to scan the area by 1 cm. With respect of this scheme, receiving linear grid is synthesized in both broadside and end-fire configurations with minimum inter-element spacing $d = 1$ cm, maximum aperture width $D = 100$ cm, and up to $N_d = 40$ elements, as depicted in Fig. 1. Due to aperture synthesis and the calibration processes, this analysis does not include the effects of antenna coupling, and the scope of the present conversation is limited to spatial correlation analysis, for LOS and NLOS configurations.

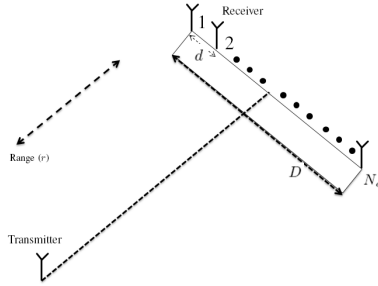


Fig. 1. General configuration of Eurecom UWB channel Measurement

2.2 Channel and Time Stationarity

The time stationarity of the channel is a obligation for the validity of the array synthesis approach, and is ensured by completely immobilizing the measurement environment, as established in [16] for example. This is in line with previous studies demonstrating the time stationarity of indoor office and residential environments [15, 16]. A different operating scenario, such as an outdoor UWB channel, may not reveal time stationarity, and the spatial correlation characteristics may then be different in time [16]. Several indoor small-office environments with size of the order of $10 \text{ m} \times 6 \text{ m}$ and strong number of multi-paths are detected. The settings within the laboratories are varied, and the transmitter receiver separation ranges from 1 to 6 m. Representing the receiving antenna position on the Cartesian measurement grid by the cross-range, x , and range, r , components, and the corresponding sets of locations by X and R , respectively, we can express the Complex Channel Transfer Function (CCTF) as [16]

$$T(x, r, f) = \sum_{n=1}^F A(x, r, n) \exp^{j\theta(x, r, n)} \delta(f - n\Delta f) \quad (1)$$

where $r \in R$ is the range location, $x \in X$ is the cross-range location, A is the amplitude and θ is the phase. Under this representation, a variety of values of $x \in X$, with r being constant, define a broadside array.

In this work, the spatial correlation coefficient, ρ , is calculated from the set of measured channel responses in the frequency domain. The receive correlation coefficient signifies the statistical correlation between the signals received at two different locations after being emitted by the same transmitter's position, at the same time as the transmit correlation is the converse quantity, i.e. the correlation with two transmitters and one receiver. Consider two CCTFs, $T_1 = T(r_1, x_1, f)$ and $T_2 = T(r_2, x_2, f)$, measured at locations (r_1, x_1) and (r_2, x_2) , respectively, separated by distance, d , given by

$$d = \sqrt{(r_1 - r_2)^2 + (x_1 - x_2)^2}. \quad (2)$$

The degree of likeness between these two CCTFs can be predicted in terms of their cross-correlation. We note that the CCTF is a random process as the frequency-domain fading coefficients are stochastic.

3 Proposed Statistical Distributions Description

The multipath indoor radio propagation channel impulse response on time domain is normally molded as a complex low-pass equivalent impulse response. To characterize the probability density function of the power variations in frequency domain ($H(f)$) we plot the histogram's measurement data [13]. The power variations are fitted with an analytical probability density function (pdf) approximation, namely a **Weibull** pdf. The general formula for the **Weibull** pdf is given by:

$$f(z) = \frac{\gamma}{\alpha} \left(\frac{z - \mu}{\alpha} \right)^{(\gamma-1)} \exp \left\{ - \left(\frac{z - \mu}{\alpha} \right)^\gamma \right\} \quad (3)$$

where $\alpha, \gamma, \mu \in R$, $\alpha, \gamma > 0$ and $z \geq \mu$, α is the scale parameter, γ is the shape parameter, and μ is the location parameter.

The Probability density function Student's t-distribution has the probability density function:

$$f(z) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{z^2}{\nu} \right)^{-(\nu+1)/2} \quad (4)$$

where ν is the number of degrees of freedom and Γ is the Gamma function. The Rice law, The probability density function is:

$$f(z|\nu, \sigma) = \frac{z}{\sigma^2} \exp \left[- \frac{(z^2 + \nu^2)}{2\sigma^2} \right] I_0 \left(\frac{x\nu}{\sigma^2} \right) \quad (5)$$

where $I_0(z)$ is the modified Bessel function of the first kind with order zero. When $\nu = 0$, the distribution reduces to a Rayleigh distribution.

3.1 Evaluation

About 2000 LOS measurements were used to qualifying the adequate distribution. Testing the results against the **Weibull** t distribution in LOS, the Rice and the Student's models, we conclude as follows: For the amplitude distribution we found the **Weibull** pdf to have the best (highest) value of log likelihood score in about 70% of the measurements. The same results are founded for NLOS. For the phases distribution we have founded that is very fitted by a non parametric distribution with normal kernel for both LOS and NLOS cases see Figures 2 and 3.

4 Spatial Correlation

Now, the correlation coefficient, ρ , between two complex random variables $u(\xi)$ and $v(\xi)$ can be evaluated using the general expression [16]

$$\rho(u, v) = \frac{E(uv^\dagger) - E(u)E(v^\dagger)}{\sigma_u \sigma_v} \quad (6)$$

where $E(\cdot)$ denotes expectation, \dagger denotes conjugation,

$$\sigma_u = \sqrt{E\{|u|^2\} + |E(u)|^2}. \quad (7)$$

and σ_v is defined in a similar manner. In order to calculate the complex correlation coefficient, ρ_a , for the UWB channel using the CTFs, we use (6) and (7) with $u = H_1$, $v = H_2$ and $\xi = f$. The envelope correlation, ρ_e , and power correlation, ρ_p , defined in [16], provide alternative definitions of the correlation coefficient. However, ρ_e and ρ_p do not make use of the phase information in the complex CTFs. To calculate ρ_e , we put $u = |H_1|$ and $v = |H_2|$ in (6) and (7), while for ρ_p , we use $u = |H_1|^2$ and $v = |H_2|^2$. The approximation

$$\rho_e \approx \rho_p \approx |\rho_a|^2. \quad (8)$$

holds for Rayleigh-faded narrowband wireless channels [17], but not necessarily for other fade distributions. We evaluate these three types of correlation for UWB channels, analyzing the effect of the spatial offset, d . If the channel is asymptotically isotropic in the horizontal plane, the correlation becomes a two dimensional Bessel function of d with contours circularly symmetric about the point of reference [16]. Fortunately, indoor UWB channels typically show large angular spreads [16], which is also reflected in our measurements.

4.1 Correlation Evaluation

This part will present the analysis to the spatial correlation characteristics as a function of range (distance), which signifies the performance of end fire arrays [16]. In order to achieve this, we calculate the expectation, over the measurement set, of the spatial correlation coefficient for a given range offset. The latter, in turn, is evaluated in terms of the correlation coefficient between the central element of the x^{th} column of the

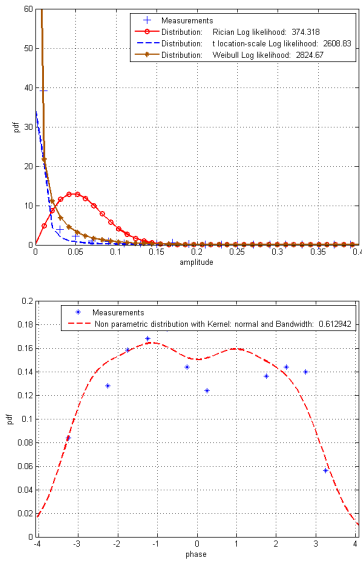


Fig. 2. Amplitude (left) and Phase (right) histogram together with the appropriate probability density functions for LOS case

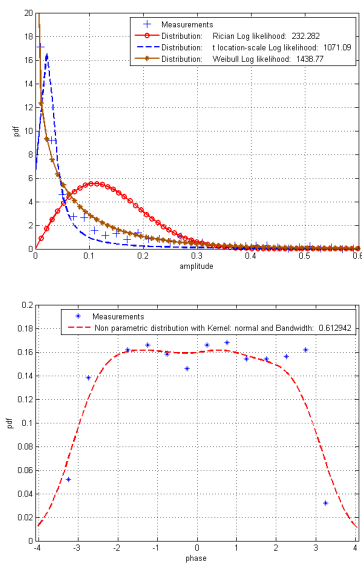


Fig. 3. Amplitude (left) and Phase (right) histogram together with the appropriate probability density functions for NLOS case

spatial measurement grid and the other elements in that column. The mean correlation coefficient magnitudes thus obtained are shown in Fig. 4. It is observed that $\rho_e \approx \rho_p$ for all considered offset d . The range correlation has a broader main lobe support compared with the cross-range correlation. The first null of ρ_e and ρ_p is obtained at $d = 20$ cm approx. ($d = 15$ cm approx. is reported in [16]), while the $\rho_e = 0.5$ threshold is crossed at $d = 9$ cm ($d = 7$ cm approx. is reported in [16]).

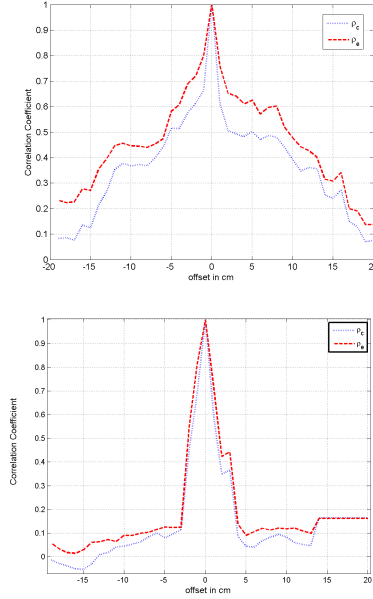


Fig. 4. The spatial correlation magnitude versus the spatial offset in LOS (left) and NLOS (right). The insets show the correlation for the corresponding LOS and NLOS channels. The complex (ρ_c) and envelope (ρ_e) correlation values are shown.

5 Conclusion

In this work, we have presented a simple empirical model for statistical small scale fading of UWB indoor channel with unknown number of scatters. The results shown that the channel small scale fading can be fitted very well by **Weibull** distribution for LOS and NLOS. Also, the analysis of correlation coefficient shows that the the complex correlation decays less rapidly with distance under LOS, but decreases rapidly under NLOS.

References

1. First report and order, revision of part 15 of the commission's rules regarding ultra-wideband transmission systems. FCC, ET Docket 98-153, February 14 (2002)
2. Ghassemzadeh, S.S., Jana, R., Rice, C.W., Turin, W., Tarokh, V.: Measurement and modeling of an ultra-wide bandwidth indoor channel. *IEEE Trans. Commun.* 52(10), 1786–1796 (2004)

3. Molisch, A.F.: Ultrawideband propagation channels-theory, measurement, and models. *IEEE Transaction Vehicular Technology* (2005) (invited paper)
4. Cassioli, D., Win, M.Z., Molisch, A.F.: The ultra-wide bandwidth indoor channel: From statistical study to simulations. *IEEE J. Select. Areas Commun.* 20(6), 1247–1257 (2002)
5. Kunisch, J., Pamp, J.: Measurement results and modeling aspects for the UWB radio channel. In: *Proc. UWBST*, pp. 19–23 (May 2002)
6. Saadane, R., Menouni, A., Knopp, R., Aboutajdine, D.: Empirical eigenanalysis of indoor ULB propagation channels. In: *IEEE Global Telecommunications Conference*, November 29–December 3 (2004)
7. Molisch, A.F., Kannan, B., Chnog, C.C., Emami, S., Karedal, A., Kunisch, J., Shantz, H., Schuster, U., Siwiak, K.: *IEEE 802.15.4a Channel model - final Report*. IEEE 802.15-04-0662-00-004a, San Antonio, TX, USA (November 2004)
8. Menouni Hayar, A., Knopp, R., Saadane, R.: Subspace analysis of indoor UWB channels. *EURASIP Journal on Applied Signal Processing, Special Issue on UWB - State of the Art* 2(3), 287–295 (2005)
9. Saadane, R., Menouni Hayar, A., Knopp, R., Aboutajdine, D.: On the estimation of the degrees of freedom of in-door UWB channel. In: *VTC Spring 2005*, May 29–June 1 (2005)
10. Saadane, R., Aboutajdine, D., Menouni Hayar, A., Knopp, R.: UWB Channel and Degrees of Freedom Evaluations. *International Journal on Wireless and Optical Communications Special Issue in Ultra Wide*, April 17 (2007)
11. Saadane, R., Aboutajdine, D., Menouni Hayar, A.: Ultra Widebandwidth Large and Small scale Characterization With Different Environments. In: *ICTIS 2007 Conference*, Fez, Morocco, April 3–5 (2007)
12. (2010), <http://www.skycross.com>
13. Saadane, R., El Aroussi, M., Hayar, A., Aboutajdine, D.: UWB Channel Modelling, Indoor Propagation: Statistical model Based on Large and Small Scales Analysis. *IJSC International Journal of Computational Science*, 1992–6669 (Print) 1992–6677 (2008)
14. Keignart, J., Pierrot, J.B., Danièle, N., Álvarez, Á., Lobeira, M., García, J.L., Valera, G., Torres, R.P.: Radio channel sounding results and model, Deliverable number: D31, IST-2001-32710-U.C.A.N
15. Chong, C.-C., Kim, Y., Lee, S.-S.: Statistical characterization of the UWB propagation channel in various types of high-rise apartments. In: *Wireless Communications and Networking Conference*, pp. 944–949 (March 2005)
16. Malik, W.Q.: Spatial correlation in ultrawideband channels. *IEEE Trans. Wireless Commun.* 7(2), 604–610 (2008)
17. LaMaire, R.O., Zorzi, M.: Effect of correlation in diversity systems with Rayleigh fading, shadowing, and power capture. *IEEE J. Select. Areas Commun.* 14(3) (April 1996)

Nonlinear Blind Source Separation Applied to a Simple Bijective Model

Shahram Hosseini¹, Yannick Deville¹, Sonia El Amine¹, and Hicham Saylani²

¹ Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse,
UPS-CNRS-OMP, 14 Av. Edouard Belin, 31400 Toulouse, France

² Laboratoire d'Electronique, de Traitement du Signal et de Modélisation Physique,
Faculté des Sciences, Université Ibnou Zohr, BP. 8061, 80000 Agadir, Maroc
{shosseini,ydeville}@irap.omp.eu, sonia_elamine@yahoo.fr,
h.saylani@uiz.ac.ma

Abstract. This paper deals with nonlinear Blind Source Separation (BSS) applied to a simple bijective “toy” model. Our objective is to better understand the difficulties encountered in nonlinear BSS, especially when estimating the parameters of mixing or separating structures. The results of this study and the proposed solutions may then be used by the BSS researchers dealing with actual nonlinear physical models. The simulation results confirm the usefulness of our proposed solutions.

1 Introduction

Blind Source Separation (BSS) aims at restoring source signals from their mixtures when the mixing parameters are unknown. While linear BSS has been widely studied, little work is available about nonlinear BSS. It is well known that the independence hypothesis is not sufficient for separating general nonlinear mixtures because of the very large indeterminacies which make the problem ill-posed [1]. A natural idea for reducing these indeterminacies is to constrain the structure of mixing and separating models to belong to a certain set of transformations [2]. Thus, the problem should be studied separately for each considered mixing structure. Even in this simplified case, nonlinear BSS is much more difficult than linear BSS because of the following problems:

1. most nonlinear models are not bijective so that even in the non-blind case when the mixing parameters are known, it is not possible to retrieve the sources in a unique manner without supplementary assumptions,
2. even when the mixing model is known, it is not always possible to find an analytical expression for its inverse,
3. the study of the identifiability and separability of nonlinear mixtures is a hard task and should be done model-by-model to determine which families of source distributions are not separable for each nonlinear model,
4. the blind estimation of the parameters in mixing (or separating) structure is another issue which is generally more difficult than in linear BSS. In particular, the matrix-based estimation algorithms can no longer be used.

The goal of our paper is the last issue, *i.e.* parameter estimation. The papers addressing this problem may be classified in the following categories¹:

- the papers considering the models which may be reduced to a linear model using some transformations (*e.g.* [1], [5]). The estimating methods proposed in these papers are especially developed for the particular considered model and cannot be generalized to other models,
- the papers studying non-bijective mixing models (*e.g.* [6], [7]). Since in this case there are several difficult problems to handle simultaneously, these papers do not focus especially on parameter estimation,
- the papers addressing this issue in general, without considering practical examples (*e.g.* [8]).

In this paper, we address the problem in the case of bijective models with known inverse and study in particular a simple “toy” model. Thus, we can focus our efforts on the parameter estimation. Although this model does not fit any known physical system, we believe this study will be useful for the BSS researchers dealing with other actual nonlinear physical models.

2 Problem Statement

Consider the mixing equation $\mathbf{x} = \mathcal{F}(\mathbf{s}, \boldsymbol{\theta}^*)$ where $\mathbf{s} = [s_1, \dots, s_K]^T$ is the vector of K independent unknown sources, $\mathbf{x} = [x_1, \dots, x_K]^T$ is the vector of K observations and \mathcal{F} is a bijective parametric function, defined by the unknown parameter vector $\boldsymbol{\theta}^*$. Denote \mathcal{G} the inverse of \mathcal{F} so that $\mathbf{s} = \mathcal{G}(\mathbf{x}, \boldsymbol{\theta}^*)$. BSS may possibly be achieved by constructing the separating model

$$\mathbf{y} = \mathcal{G}(\mathbf{x}, \boldsymbol{\theta}) \quad (1)$$

and looking for a parameter vector $\boldsymbol{\theta}$ which makes the components of $\mathbf{y} = [y_1, \dots, y_K]^T$ independent. It is clear that $\boldsymbol{\theta}^*$ is one of the solutions which provides the original sources. The other possible solutions depend on the indeterminacies involved in the problem. To make the components of \mathbf{y} independent, we can minimize the mutual information criterion defined as $I = E[\log f_{\mathbf{y}}(\mathbf{y})] - \sum_{k=1}^K E[\log f_{y_k}(y_k)]$ where $f_{\mathbf{y}}$ and f_{y_k} are respectively the joint and the marginal probability density functions (pdf) of the variables y_k . Since the model is supposed bijective, (1) yields $f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{x}}(\mathbf{x})/|J|$ where J is the Jacobian of the separating model. Then, we obtain

$$I = E[\log f_{\mathbf{x}}(\mathbf{x})] - E[\log(|J|)] - \sum_{k=1}^K E[\log f_{y_k}(y_k)] \quad (2)$$

To minimize this criterion using an optimization algorithm we need to compute its gradient and possibly its Hessian with respect to the parameter vector $\boldsymbol{\theta}$. As shown in [1], the gradient reads

¹ In this classification, we do not consider the non model-based papers like [3] and [4].

$$\frac{dI}{d\theta} = -E \left[\frac{1}{J} \frac{dJ}{d\theta} \right] + \sum_{k=1}^K E \left[\psi_{y_k}(y_k) \frac{dy_k}{d\theta} \right] \quad (3)$$

where $\psi_{y_k}(y_k) = -d \log f_{y_k}(y_k)/dy_k$ is the score function of y_k . Then, the element (i, j) of the Hessian matrix \mathbf{H} can be obtained as follows:

$$H_{ij} = \frac{d}{d\theta_j} \frac{dI}{d\theta_i} = -E \left[\frac{d}{d\theta_j} \left(\frac{1}{J} \frac{dJ}{d\theta_i} \right) \right] + \sum_{k=1}^K E \left[\frac{d\psi_{y_k}(y_k)}{dy_k} \frac{dy_k}{d\theta_j} \frac{dy_k}{d\theta_i} + \psi_{y_k}(y_k) \frac{d}{d\theta_j} \frac{dy_k}{d\theta_i} \right] \quad (4)$$

The mutual information may be minimized using e.g. the gradient descent algorithm $\theta_{new} = \theta_{old} - \mu \frac{dI}{d\theta}$ or the Newton algorithm $\theta_{new} = \theta_{old} - \mathbf{H}^{-1} \frac{dI}{d\theta}$. The online (stochastic) version of the gradient descent algorithm may be obtained by removing the expected values in (3).

The score functions required in the equations must be estimated from the outputs y_1 and y_2 and be updated at each iteration of the optimization algorithm. They may be estimated for example using the approach proposed in [9] which consists in writing $\psi_{y_k}(y_k) = \sum_{m=1}^M c_{km} \phi_m(y_k)$ where $\phi_m(y_k)$ are some basis functions and in computing the coefficients c_{km} by solving the following equation

$$\mathbf{G}_k [c_{k1}, \dots, c_{kM}]^T = \mathbf{g}_k \quad (5)$$

where $\mathbf{G}_k = E[\phi(y_k)\phi(y_k)^T]$, $\mathbf{g}_k = E[\phi'(y_k)]$ with $\phi(y_k) = [\phi_1(y_k), \dots, \phi_M(y_k)]^T$ and $\phi'(y_k)$ its derivative with respect to y_k . This derivative may also be used for estimating the score function derivatives required in (4).

An online estimate of the score functions may be obtained [10] at each time t by updating the matrices \mathbf{G}_k and the vectors \mathbf{g}_k using

$$\mathbf{G}_k(t) = \rho \mathbf{G}_k(t-1) + (1-\rho)\phi(y_k)\phi(y_k)^T, \quad \mathbf{g}_k(t) = \rho \mathbf{g}_k(t-1) + (1-\rho)\phi'(y_k) \quad (6)$$

where ρ is a “forgetting factor” (for example equal to $(t - 1)/t$), then solving $\mathbf{G}_k(t)[c_{k1}(t), \dots, c_{kM}(t)]^T = \mathbf{g}_k(t)$ to find the coefficients $c_{km}(t)$.

In the following sections, we study a simple toy problem to show the different practical aspects of nonlinear BSS.

3 A Simple Bijective Model

We consider the following inverse structure

$$s_1 = a^* x_1^3 + b^* x_2, \quad s_2 = -b^* x_1 + a^* x_2. \quad (7)$$

This model, which is defined by the parameter vector $\theta^* = [a^*, b^*]^T$ is bijective if $b^* \neq 0$ (and if $b^* = 0$ but $a^* x_1 \neq 0$): in this case its Jacobian $3a^{*2} x_1^2 + b^{*2}$ is always positive. The above equations yield $a^* x_1^3 + (b^{*2}/a^*)x_1 + (b^*/a^*)s_2 - s_1 = 0$

which can be solved using Cardan's formula with respect to x_1 to obtain one of the two mixing equations

$$x_1 = \left(\frac{-q}{2} + \sqrt{\Delta} \right)^{1/3} + \left(\frac{-q}{2} - \sqrt{\Delta} \right)^{1/3} \quad (8)$$

where $\Delta = \frac{q^2}{4} + \frac{p^3}{27}$, $q = ((b^*/a^*)s_2 - s_1)/a^*$ and $p = (b^*/a^*)^2$. The other mixture may then be obtained using

$$x_2 = (s_2 + b^*x_1)/a^*. \quad (9)$$

BSS may be achieved by constructing the separating structure

$$y_1 = ax_1^3 + bx_2 \quad , \quad y_2 = -bx_1 + ax_2 \quad (10)$$

and minimizing the mutual information of y_1 and y_2 with respect to $\boldsymbol{\theta} = [a, b]^T$. This model yields $J = 3a^2x_1^2 + b^2$, $dJ/d\boldsymbol{\theta} = [6ax_1^2, 2b]^T$, $dy_1/d\boldsymbol{\theta} = [x_1^3, x_2]^T$, $dy_2/d\boldsymbol{\theta} = [x_2, -x_1]^T$. Using (3) and (4), we obtain the following expressions for the gradient and the Hessian

$$\begin{aligned} \frac{dI}{d\boldsymbol{\theta}} &= E \left[\frac{-1}{3a^2x_1^2 + b^2} [6ax_1^2, 2b]^T + \psi_{y_1}(y_1)[x_1^3, x_2]^T + \psi_{y_2}(y_2)[x_2, -x_1]^T \right] \\ \mathbf{H} &= E \left[\frac{1}{J^2} \begin{pmatrix} -6x_1^2J + (6ax_1^2)^2 & 12abx_1^2 \\ 12abx_1^2 & -2J + (2b)^2 \end{pmatrix} \right] + E \left[\frac{d\psi_{y_1}(y_1)}{dy_1} \begin{pmatrix} x_1^6 & x_1^3x_2 \\ x_1^3x_2 & x_2^2 \end{pmatrix} \right] \\ &+ E \left[\frac{d\psi_{y_2}(y_2)}{dy_2} \begin{pmatrix} x_2^2 & -x_1x_2 \\ -x_1x_2 & x_1^2 \end{pmatrix} \right]. \quad (11) \end{aligned}$$

From (7) and (10), it is evident that if $\boldsymbol{\theta} = k[a^*, b^*]^T$, then $y_1 = ks_1$ and $y_2 = ks_2$ so that y_1 and y_2 are independent and minimize the criterion. One of the solutions to fix this indeterminacy consists in defining $s'_i = s_i/a^*$ and $c^* = b^*/a^*$ which yields the inverse model

$$s'_1 = x_1^3 + c^*x_2 \quad , \quad s'_2 = -c^*x_1 + x_2 \quad (12)$$

and the corresponding separating structure

$$y_1 = x_1^3 + cx_2 \quad , \quad y_2 = -cx_1 + x_2 \quad (13)$$

In this case, there is only one parameter to estimate so that the gradient and the Hessian are scalars: $\frac{dI}{dc} = E \left[\frac{-2c}{3x_1^2 + c^2} + \psi_{y_1}(y_1)x_2 - \psi_{y_2}(y_2)x_1 \right]$, $\mathbf{H} = \frac{dI^2}{dc^2} = E \left[\frac{-6x_1^2 + 2c^2}{(3x_1^2 + c^2)^2} + \psi'_{y_1}(y_1)x_2^2 + \psi'_{y_2}(y_2)x_1^2 \right]$.

4 Local Minima and Separability of the Model

In a first experiment, we mixed two 10000-sample independent, zero-mean and unit-variance, uniformly distributed sources s_1 and s_2 using the mixing model

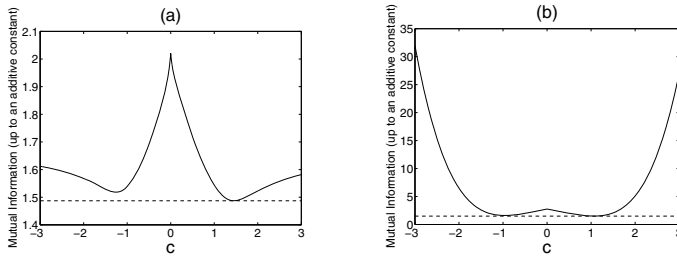


Fig. 1. Estimation of mutual information (up to an additive constant) (a) with pdf shape re-estimated for each value of c . (b) with pdf shape estimated for $c = 1$.

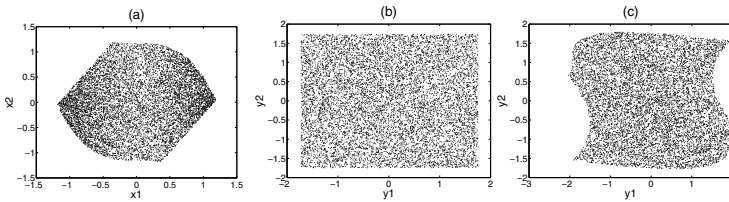


Fig. 2. scatter plots of (a) mixtures, (b) output components corresponding to the good local minimum, (c) output components corresponding to the bad local minimum

(8-9) and the parameters $a^* = 1$ and $b^* = 1.5$ which is equivalent to the inverse model (12) with $c^* = 1.5$ and $s'_i = s_i$. Then, we constructed the separating model (13), varied the parameter c between -3 and 3 , and for each value of c estimated the mutual information of the outputs (up to the additive constant $E[\log f_{\mathbf{x}}(x)]$). The result is shown in Fig. 1.a. As can be seen, this function has two local minima but only one of them (which is also the global minimum) corresponds to the actual value of the parameter and provides the independent components corresponding to the actual sources². When initialized with negative values of c , the optimization algorithms like gradient descent or Newton converge towards this “bad” local minimum. However, this value may be rejected a posteriori using an independence test. Fig. 2 shows the scatter plots of the mixtures and of the output components corresponding to these minima.

Note also that in practice, at each iteration of an optimization algorithm, one first estimates (using the current value of the parameter c) the coefficients c_{km} in (5) which determine the shape of score functions and related pdf, then freezes them and performs a minimization step for the mutual information related to these pdf with respect to c . Since the estimated pdf change during successive iterations, the shape of the function to be minimized changes too. For example, Fig. 1.b shows the mutual information as a function of c in the above example

² Note that replacing x_1^3 by x_1 in (12) and (13) yields a linear model for which the criterion has two good local minima at c^* and $-1/c^*$ leading respectively to $[y_1, y_2]^T = [s_1, s_2]^T$ and $[y_1, y_2]^T = [-s_2, s_1]^T/c^*$.

(with $c^* = 1.5$) corresponding to the coefficients c_{km} estimated using the value $c = 1$. As can be seen this function is not the same as in Fig. 1a. The practical optimization is then more difficult than what is suggested by Fig. 1a. This example also shows the sensitivity of the method to the estimation of score functions: if the estimated score functions are not updated in the following iterations, the optimization algorithm converges towards the minimum of Fig. 1b, i.e. $c = 1.07$.

The separability of our one-parameter model may be formulated as follows: are there a family of source distributions and a value of the parameter c in the separating model (13) for which y_1 and y_2 are independent but contain mixtures of s_1 and s_2 ? To answer this question, one has to solve an independence conservation functional equation [8]. Here, we only try to respond partially to this question by the following experiment: we consider the generalized Gaussian distribution family defined by the parameter α . For the values of α between 0.5 and 20 we generated the mixtures x_1 and x_2 for $a^* = 1$ and $b^* = 1.5$ (so that $c^* = 1.5$), then the outputs y_1 and y_2 using (13) for the values of c between -20 and 20. For each value of α , we estimated the mutual information I as a function of c (like in Fig. 1a) and verified if there was a value of c different from c^* for which $I \simeq 0$. Since we did not find such values, we can say “experimentally” that the model is separable for generalized Gaussian distributions.

5 Simulation Results

The first two lines of Table 1 compare the batch versions of the gradient descent (with a constant learning rate $\mu = 0.02$) and Newton methods applied to the mixtures generated as in Section 4. The algorithms were run 100 times corresponding to 100 different source signals and 100 different initial random values of the parameter c (uniformly distributed over $[0.1, 2.1]$). The score functions were estimated using the method described in Section 2 with $\phi_m(y_k) = y_k^{m-1}$ for $m = 1, \dots, 5$. In each simulation, the algorithm was stopped if $|c_{new} - c_{old}| < 10^{-6}$ and the performance was measured using the Signal to Interference Ratio (SIR) criterion defined by $SIR = \frac{1}{2} \sum_{i=1}^2 10 \log_{10} \frac{E[s_i^2]}{E[(s_i - \hat{s}_i)^2]}$ where \hat{s}_i is the estimate of s_i computed using the final estimate of the parameter c . We also tested the initial model with two parameters (Eq. 7-10) using $a^* = 4$, $b^* = 6$ (so that $b^*/a^* = 1.5$) and the parameters a and b initialized with positive random values. In this case, we estimate the two parameters simultaneously. The sources may be then estimated only up to a scaling factor. The last two lines of

Table 1. Comparing batch versions of gradient and Newton algorithms

	Mean(SIR)	Std(SIR)	Iterations per simulation	time per simulation
Gradient (1 param)	52.6 dB	9.6 dB	1208	17.87 sec
Newton (1 param)	52.6 dB	9.6 dB	111	1.7 sec
Gradient (2 param)	60.5 dB	7.8 dB	75	1.7 sec
Newton (2 param)	60.5 dB	7.8 dB	4	0.2 sec

Table 1 show the results. The SIR was computed after normalizing the estimated sources so that they had the same variances and signs as the actual sources. Note that in the Newton method, the Hessian \mathbf{H} may be badly conditioned and even negative-definite. To avoid this problem, after the eigenvalue decomposition of $\mathbf{H} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, the eigenvalues smaller than a positive value δ (chosen equal to 10^{-4} in our experiments) are replaced by δ [11].

As can be seen, this separating model leads to better results. This is probably because the parameters space defined by two parameters has a better shape so that the optimization algorithms converge better towards its minimum. In all the experiments, the Newton algorithm is much less time consuming than the gradient algorithm while providing the same performance.

Then, we tested the stochastic gradient algorithm. When using this algorithm, we have two principal problems: carefully choosing the learning rate (because the algorithm is extremely sensitive to this choice), and carefully estimating the score functions. The choice of learning rate μ is widely discussed in the neural networks literature [12]. We found that the following adaptation rule for updating the learning rate gives good results

$$\mu_i(t) = \mu_i(t-1) \cdot \max \left(0.5, 1 + q \nabla_i(t) \frac{\overline{\nabla}_i(t-1)}{\overline{\nabla}_i^2(t)} \right) \quad (14)$$

where $\nabla_i = \frac{dI}{d\theta_i}$, $\overline{\nabla}_i(t-1) = \rho \overline{\nabla}_i(t-2) + (1-\rho) \nabla_i(t-1)$ and $\overline{\nabla}_i^2(t) = \rho \overline{\nabla}_i^2(t-1) + (1-\rho) \nabla_i^2(t)$ with ρ a forgetting factor. The main idea is to increase the learning rate when the new gradient points in the same direction as the average past gradient $\overline{\nabla}_i(t-1)$ (normalized by the average of the squared gradient to make it better conditioned), and to decrease it otherwise. The multiplier is limited below by 0.5 to guard against very small (or even negative) factors. In our experiments, we used $\rho = \frac{t-1}{t}$, $q = 1.5$ and $\mu(0) = 0.02$.

We also need a new estimation of the score functions at each time t . Our experiments show that the approach proposed at the end of Section 2 and based on Eq. (6) does not give good results because at the first stages of the algorithm the estimation of c and consequently the estimation of the score functions are bad. Using (6), this bad estimation of the score functions does not change significantly afterwards. Hence, we propose another approach which consists in updating the score functions at each time t from all the past data, using $\mathbf{G}_k(t) = \frac{1}{t} \sum_{n=1}^t \phi(y_k(n)) \phi(y_k(n))^T$ and $\mathbf{g}_k(t) = \frac{1}{t} \sum_{n=1}^t \phi'(y_k(n))$.

We repeated the experiment with the one-parameter model using the stochastic gradient algorithm and the signals containing 10000 samples. The mean and the standard deviation of the SIR using 10 simulations were 42.0 dB and 15.2 dB with a runtime of about 90 seconds for each simulation. Figure 3 shows the evolution of the parameter c and the learning rate μ in one of the simulations. The same experiment using the two-parameters model led to an average SIR of 46.9 dB with a standard deviation of 6.3 dB and about 140 sec per simulation.

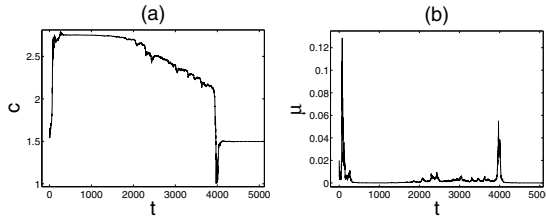


Fig. 3. Evolution of (a) the parameter c and (b) the learning rate μ in the stochastic algorithm versus sample index t

6 Conclusion

In this paper, we studied the BSS problem for one of the simplest bijective nonlinear models. Even for this simple model, the problem is much more difficult than linear BSS because of the existence of spurious local minima, the high sensitivity of the optimization algorithms to the estimation of score functions, the importance of parameter tuning in these algorithms, etc. We proposed solutions to cope with these problems which may be helpful for the future works using more realistic models. More experiments using constrained optimization algorithms are required for treating the case of non-bijective models.

References

1. Taleb, A., Jutten, C.: Source separation in post-nonlinear mixtures. *IEEE Trans. on Signal Processing* 47(10), 2807–2820 (1999)
2. Jutten, C., Babaie-Zadeh, M., Hosseini, S.: Three easy ways for separating nonlinear mixtures? *Signal Processing* 84(2), 217–229 (2004)
3. Almeida, L.B.: MISEP - linear and nonlinear ICA based on mutual information. *Journal of Machine Learning Research* 4, 1297–1318 (2003)
4. Zhang, K., Chan, L.: Kernel-Based Nonlinear Independent Component Analysis. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) *ICA 2007*. LNCS, vol. 4666, pp. 301–308. Springer, Heidelberg (2007)
5. Eriksson, J., Koivunen, V.: Blind identifiability of class of nonlinear instantaneous ICA models. In: *Proc. of EUSIPCO 2002, Toulouse*, vol. 2, pp. 7–10 (September 2002)
6. Hosseini, S., Deville, Y.: Blind Maximum Likelihood Separation of a Linear-Quadratic Mixture. In: Puntotnet, C.G., Prieto, A.G. (eds.) *ICA 2004*. LNCS, vol. 3195, pp. 694–701. Springer, Heidelberg (2004), ERRATUM: <http://arxiv.org/abs/1001.0863>
7. Duarte, L.T., Jutten, C.: Blind Source Separation of a Class of Nonlinear Mixtures. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) *ICA 2007*. LNCS, vol. 4666, pp. 41–48. Springer, Heidelberg (2007)
8. Taleb, A.: A generic framework for blind source separation in structured nonlinear models. *IEEE Trans. Sig. Proc.* 50(8), 1819–1830 (2002)

9. Pham, D.T., Garat, P.: Blind separation of mixtures of independent sources through a quasi maximum likelihood approach. *IEEE Trans. Sig. Proc.* 45(7), 1712–1725 (1997)
10. Pham, D.T.: Séparation aveugle de mélange instantanée de sources à l'aide de fonctions séparatrices ajustées. In: *Proc. GRETSI 1997*, pp. 969–972 (September 1997)
11. Nocedal, J., Wright, S.: *Numerical optimization*. Springer (2006)
12. Bishop, C.M.: *Neural networks for pattern recognition*, Oxford (1995)

Seismic Signal Discrimination between Earthquakes and Quarry Blasts Using Fuzzy Logic Approach

El Hassan Ait Laasri, Es-Saïd Akhouayri, Dris Agliz, and Abderrahman Atmani

Seismic Signal Processing Team, Electronic,
Signal Processing and Physical Modelling Laboratory,
Physics' Department, Faculty of Sciences,
IBN ZOHR University, B.P. 8106, Agadir, Morocco
{hassan.or, driss_agliz, atmani_abderrahman}@hotmail.com,
akhouayri@msn.com

Abstract. Seismic analysts identify earthquakes signals from those of explosions by visual inspection and calculating some characteristics of seismogram. Such work supposes a great deal of workload for seismic analysts. Therefore, an automatic classification tool reduces dramatically this arduous task, turns classification reliable, removes errors associated to tedious evaluations and changing of personnel. In this present paper we are interested in transforming the analysts' knowledge of classifying seismograms into an automated soft classification system based on fuzzy logic. This is primarily due to its capability of modelling human reasoning and decision-making, managing complexity and controlling computational cost. These capabilities are essential for manipulating high dimensionality and complexity of seismic signal. To conduct effective discrimination, relevant seismogram characteristics are extracted based on human experience. Using these characteristics, a fuzzy classifier is built and tested with real seismic data. The results are found to be encouraging.

Keywords: Seismic classification, Seismic signal processing, Fuzzy logic.

1 Introduction

The seismic database of Agadir was implemented in May 1998 by realization of an automatic station for detecting seismic events, and then in November 2001 by installation of the local seismic network [1]. The seismic events are detected by a power detector whereby the power over a short time-window (short-term average, STA) is compared with the power over a long time-window (long-term average, LTA). The basic idea of the algorithm is that an event is considered detected when the STA/LTA ratio exceeds a pre-determined threshold [2]. Because of several quarries located surrounding the Agadir city, many quarry explosions seismogram are recorded per day. As recorded explosions can mislead scientists interpreting the active tectonics and lead to erroneous results in the analysis of seismic hazards in the area, an event classification task is an important step in seismic signal processing. Such

task analyses data in order to find to which class each recorded event belongs. In this work, we focus our efforts on identifying earthquakes seismograms from those of explosions in order to construct an accurate and classified database. In view of very high volume of data, the Considerable operations conducted by humans can be tedious and stressful. Therefore, constructing a reliable automatic discrimination system is a crucial issue. Many different seismogram discrimination methods have been developed, where the event signal is reduced to a set of features. The latter can be extracted from time-domain representation of the signal, time frequency representation or spectral representation. Spectral ratios P/S and P/L are commonly presented as good discriminants between earthquakes and quarry blasts [3] [4] [5] [6]. Nevertheless, due to the low magnitude and overlapped P and S waves of quarry blast events, we cannot use the previous discriminant methods.

As the seismic analyst arrived to distinguish between earthquake and explosion by visual inspection and calculating some characteristics of event signal, we are interested in transforming the analyst knowledge of classifying seismograms into an automated soft classifier. Such classifier can imitate the reasoning processes of experts in solving classification problem. The discriminant method was developed using fuzzy logic, considering that the latter is shown to be very useful in acquiring knowledge from human and a powerful tool to incorporate imprecise knowledge.

The rest of this article is organized as follow. Data characteristics are discussed in Section 2. Section 3 is devoted to the proposed discrimination method of fuzzy logic approach. Experimental results are shown in Section 4. Finally, Section 5 outlines some conclusions.

2 Data

The term “seismic source” is a comprehensive name for all events or, more generally, for any radiator of seismic waves. A seismic source can be described by its strength and its spatial and temporal characteristics, i. e., by parameters such as source dimension, geometry and time function of the radiation. Explosion sources are instantaneous and produce a spherically expanding compressional first motion in all directions while the tectonic earthquake sources produce first motions of different amplitude and polarity in different directions. Due to these source characteristics, it is obvious that the forces acting at the source of an explosion are very different from those acting at the source of a natural earthquake. These different forces introduce diverse characteristics to the seismic signal that can be readily used to identify each type of source process and discriminate between explosions and tectonic earthquakes.

In this section we present a data set of quarry blast events recorded by the seismic local network of Agadir. Natural earthquake seismograms are also collected for comparison study between these two events. Such study is useful to extract some characteristics of the signal generated by each type of these events. Fig. 1 depicts the vertical components seismogram of two earthquakes (a, b) and two quarry blasts (c, d).

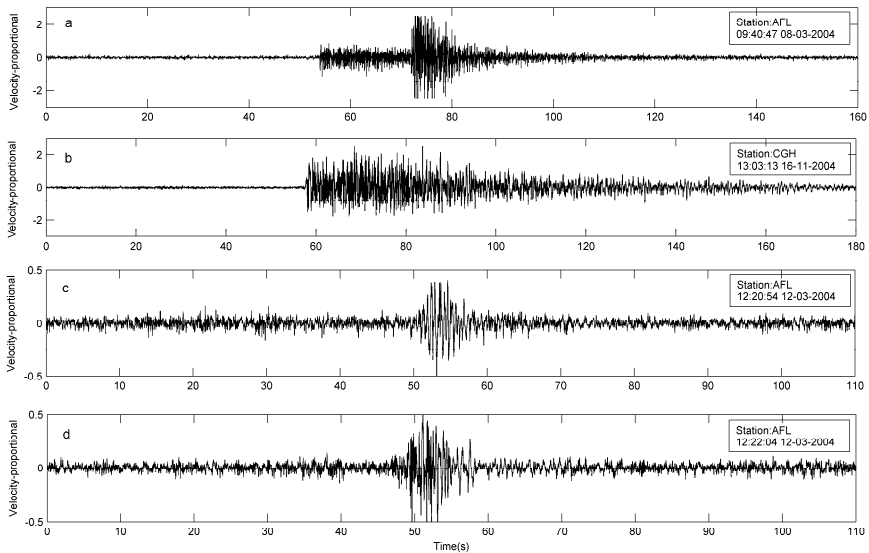


Fig. 1. Shows the vertical component seismogram of two earthquakes (a, b) and two quarry blasts (c, d) recorded by the local seismic network of Agadir

In a preliminary look at these seismograms, it seems that signal envelope is a promising discrimination parameter. The quarry blasts recording is characterized by overlapped P and S waves, less impulsive onset and short duration of coda waves. All these characteristics blend together to form a Gaussian envelope. Signal associated with an earthquake differs appreciably from that of explosion in that it involves large S waves, isolated or overlapped P and S waves (it depends on source-station distance) and exponential decay of coda amplitude with time.

Analysis of many explosion signals shows that all the events have almost the same envelope, and can be recognized using only the envelope. Unfortunately, not only the explosion signals show this feature; some earthquake signals have also the same envelope as explosion. In order to find another parameter which can differentiate these types of event, we analysed their signals in the frequency domain. In fig. 2, seismogram of an earthquake and a quarry blast with their associated FFT are compared. As can be seen in fig. 2, envelopes of these two events are quite similar, but the difference between their spectral content is clearly visible. It was observed that seismograms of quarry blast exhibit very low frequency amplitude below 1Hz. Contrary to earthquake seismograms, which often show very high frequency amplitude in the same band.

Another important parameter to be considered is time duration of the event. Time duration is influenced by many factors, mainly the characteristic dimensions of the source. Thus, it should be expected that time duration of natural earthquakes would be longer than the time duration of explosions.

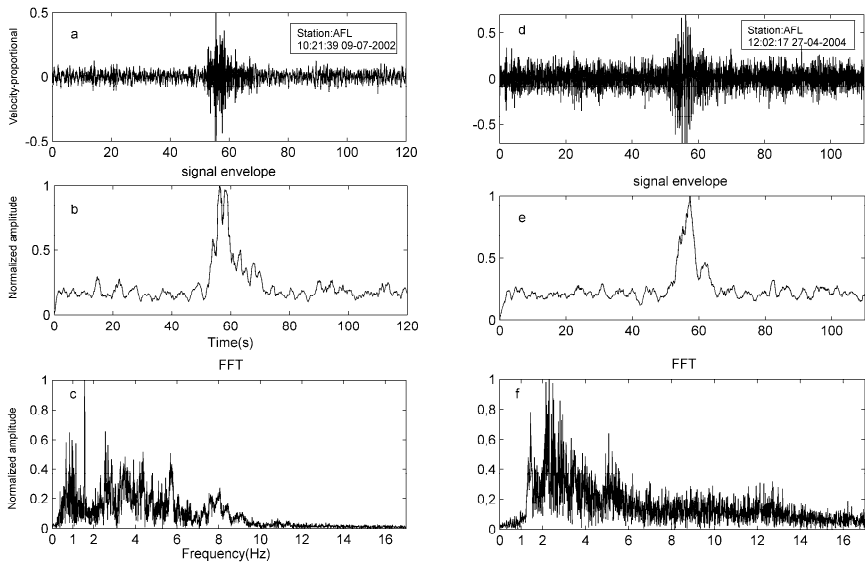


Fig. 2. Comparison between seismogram of an earthquake (a) and a quarry blast (d) and their corresponding envelope (b, e) and FFT (c, f)

Examination of the data displays that explosion records have durations of less than 40 seconds while tectonic earthquake records may last for several minutes.

The simplest variable which can be used for discrimination between earthquake and explosion is the hour of detection. This variable rely on the fact that explosion are exploded during specified hours and forbidden during the night. Therefore, we cannot record a quarry blast event during night time. Such variable is unreliable because it cannot separate seismograms recorded during day time, but can be reliable if it is used with other parameters.

3 Method

The problem of discrimination between earthquakes and quarry blasts is formulated as a problem in fuzzy logic. Such problem may be separated into two stages, feature extraction and classification.

3.1 Feature Extraction

The performance of classification method is affected by feature set used to reduce the high dimensionality of seismogram. Therefore, extract the most relevant feature from known seismograms contribute to more reliable discrimination of unknown seismograms. In this work the feature set defined by analysts corresponded to:

Envelope: To extract the signal envelope, we use the Hilbert Transform which is capable of tracking the amplitude envelope of the signal:

$$\begin{aligned} H[x(t)] &= x(t) * \frac{1}{\pi t} \\ &= \frac{1}{\pi} \left[\int_{-\infty}^{+\infty} x(\tau) \cdot \frac{1}{(t-\tau)} \cdot d\tau \right] \end{aligned} \quad (1)$$

From the given signal $x(t)$, a complex signal $A[x(t)]$ (also known as analytical signal) that is associated with the original signal can be constructed as :

$$A[x(t)] = x(t) + jH[x(t)] \quad (2)$$

The envelope of the signal is defined as:

$$E(t) = \sqrt{x(t)^2 + H[x(t)]^2} \quad (3)$$

A finite impulse response filter (FIR) is designed to minimize the rapid variation of the envelope. As quarry blasts have usually the same envelope, we have chosen an envelope of this event as template envelope, which will be compared with upcoming events envelope.

Time Duration: The time duration t_d is defined as the total duration in seconds of the event record from the P wave onset t_p to the end of the signal t_{end} . The latter is defined as the point where the signal is no longer seen above the noise.

$$t_d = t_{end} - t_p \quad (4)$$

Hour: Quarry blast events occur during the time day from 11:00 a.m. to 02:00 p.m. and from 05:00 p.m. to 06:00 p.m. Beyond this time intervals the explosion are absent, and the seismicity pattern during these hours should not be affected by these events.

$$Hour = hour + minute/60 + second/3600 \quad (5)$$

Frequency Content: The frequency amplitude of seismic event signal is calculated in the frequency band $[f_1 \ f_2] = [0.5 \ 1]$ Hz.

$$E_s = \int_{f_1}^{f_2} a(f) df \quad (6)$$

3.2 Fuzzy Classifier

Fuzzy logic theory, introduced by Zadeh [1965], deals with reasoning that is approximate rather than fixed and exact, which is the case in traditional logic theory. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value

may range between completely true and completely false. In classical concept of a set, the transition from one set to another is always abrupt; the membership of elements in a set is assessed in binary terms. So, an element either belongs or does not belong to the set. By contrast, the boundary of a fuzzy set is not precise. That is, the change from one class to its neighbours in a fuzzy set may be gradual rather than abrupt. This gradual change is expressed by a membership function valued in the real unit interval [0, 1]. Because the real world data, where information is often incomplete or sometimes unreliable, do not have sharply defined boundaries, fuzzy classifier provides the appropriate tool to manipulate the real data. What makes Fuzzy Logic so powerful is its ability of describing a system in linguistic terms rather than in terms of numbers. This enables the design of such system with more human-like reasoning using linguistic terms of natural languages.

Let $C=(C1, C2)$ be the two classes of seismic event, which can be described by a set of features or attributes $X_i, i=1, \dots, n$. i.e., a given event to classify is an element $x=(x_1, \dots, x_n)$ of $X_1 \times \dots \times X_n$, where x_i is the value taken by attribute i for this event. In the sequel, X_i will indicate either the attribute (or variable) itself or its set of values, while x_i indicate possible values of X_i . The problem of designing the classifier is to define a mapping F such as:

$$F: X_1 \times X_2 \times \dots \times X_n \rightarrow C \quad (7)$$

By employing fuzzy logic as a nonlinear mapping F , it is possible to describe the degree to which an event belongs to one class or the other.

In order to solve a fuzzy classification problem, it is necessary to fuzzify inputs (Fuzzification), determine all IF-THEN rules (rule base), process them (Interference engine) and provide result in a usable and understandable form (Defuzzification) [7][8].

Fuzzification: Fuzzification interface converts inputs of system into fuzzy variable. It contains predefined sets of linguistic terms and the degree to which inputs belong to each of the appropriate fuzzy sets is determined via membership functions. Qualitative interpretation of different available values of input variables is achieved by fuzzification.

Fuzzy Rules: The rules refer to variables and the adjectives that describe those variables. Rules are fuzzy conditional statements (implications) and usually expressed in the form:

$$IF \text{ variable } IS \text{ adjective } THEN \text{ class} \quad (8)$$

Fuzzy Inference: Fuzzy inference simulates human decision making to assign a class to each input based on his knowledge and interpretation of the input vector. So, the point of fuzzy inference is to map an input space to an output space, and the primary mechanism for doing this is the list of 'if-then' rules. All rules are evaluated in parallel, and the order of the rules is unimportant. The most commonly types of fuzzy inference systems that can be implemented in Fuzzy Logic are: Mamdani-type and Sugeno-type which is used in this paper [9]. In fact they are similar in many respects. The main difference between them is that the output membership functions are only

linear or constant for Sugeno-type fuzzy inference. A typical rule in a Sugeno fuzzy model has the following form:

If Input 1 = x and Input 2 = y , then Output z is:

$$z = ax + by + c \tag{9}$$

where a , b and c are constants.

Defuzzification: Defuzzification interface converts fuzzy outputs into numerical value, which corresponds to event class.

4 Results and Discussion

The dataset contains two classes of seismic signals: earthquake and quarry blast. Discriminative parameters are illustrated in Table 1, where their corresponding fundamental descriptive statistical information for each class is provided. Such information is useful in defining the membership function. To represent the variable Hour, the histogram was used (fig. 3).

Table 1. Discriminative parameters and their corresponding fundamental descriptive statistical information for the two classes

	<i>Duration</i>		<i>envelope</i>		<i>frequency</i>	
	earthquake	Quarry blast	earthquake	Quarry blast	earthquake	Quarry blast
minimum	13.69	9.06	0.47	0.09	0.08	0.22
maximum	736.80	32.03	0.98	0.55	190.12	7.44
mean	135.39	17.30	0.82	0.39	20.51	2.38
Standard deviation	150.45	4.83	0.12	0.09	39.51	1.58

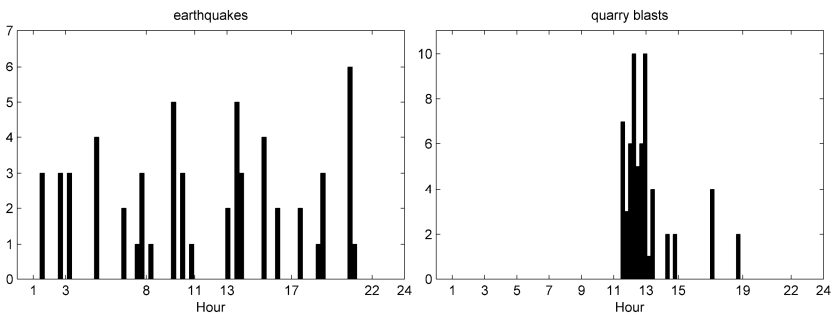


Fig. 3. Histogram of the parameter Hour for both earthquake class and quarry blast one

From these statistical data, it turns out that there is no single parameter with fixed classification threshold; as it can be seen in Table 1, similar values can be found for each parameter in the two classes. Therefore a single parameter cannot distinguish

between the two classes because they are overlapped. The most accepted solution for discrimination is the combination of these different parameters. At this point, the fuzzy inference system plays an important role. Fuzzy classes reflect reality better and allow decision makers to describe input attributes and output classes more intuitively using linguistic variables, overlapping classes and approximate reasoning. Events that belong to more than one class are treated in all classes where they have partial membership.

The classification system is made up using the first-order Sugeno fuzzy model with fore inputs and an output. Trapezoidal type membership function is chosen to be more suitable for the input variable because of its shape. Based on descriptions of the input (Hour is night, envelope is Gaussian, duration is short, frequency is low...etc) and output classes (earthquake, quarry blast), 09 rules are constructed. With 120 events (60 earthquakes, 60 quarry blasts), results were as following:

- correctly classified events: 120
- misclassified: 0
- accuracy: 100%

The results of this research have shown that the fuzzy approach is perfectly suitable for distinguishing earthquake events from quarry blast ones. Such technique, achieved a high discrimination performance with low complexity, could be employed in real time discrimination. Moreover, by using fuzzy logic rules, the maintenance of the structure of the algorithm decouples along fairly clean lines. The features characteristics of each class might change in the future, but the underlying fuzzy classifier will be the same. For example, exploded hour of quarry blasts can be changed in the future but, the system can be recalibrated quickly by simply shifting the fuzzy set that defines Hour or just rewriting the fuzzy rules without touching the complex programming code. Also, adding more rules to the bottom of the list to increment or expand the scope of the knowledge-base, as processes develop or new events are found, is relatively easy and without needing to undo what had already been done. In other words, the subsequent modification was pretty easy. The last statement is perhaps the most important one and deserves to be addressed here. Since fuzzy logic is built on top of linguistic terms used by ordinary people on a daily Basis, fuzzy logic allows anyone to edit and modify the rules without worrying about underlying code.

5 Conclusion

In this paper, an automated discriminant method between earthquakes and explosions in Agadir database is developed using fuzzy logic. Each event is represented by a set of features deduced from corresponding signal. The fuzzy system interprets the values in the input vector and, based on some set of rules, assigns each input to its class. Fuzzy logic is used due to its human-like-reasoning nature, its feasibility of implementation nonlinear problems and its capabilities to deal with uncertainties and imprecision, but otherwise the fuzzy should be considered in view of its simplicity and transparency. This simplicity however does not limit its effectiveness. In this work fuzzy classifier appear as a

powerful tool to deal with seismic signal, which is distorted, weakly, noisy and complex. Fuzzy classifier results show good performance with low complexity. However, fuzzy methods are still dependent on expert knowledge.

References

1. Akhouayri, E., Ait Laasri, H., Agliz, D., Atmani, A.: Agadir's seismic central acquisition management. In: 27th European Center for Geodynamics and Seismology, ECGS, Luxembourg, November 17-19 (2008)
2. Akhouayri, E., Agliz, D., Fadel, M., Ait Ouahman, A.: Automatic detection and indexation of seismic events. *AMSE Periodicals, Advances in Modeling and Analysis, série C* 56(3), 59–67 (2001)
3. Yıldırım, E., Gulbag, A., Horasana, G., Dogan, E.: Discrimination of quarry blasts and earthquakes in the vicinity of Istanbul using soft computing techniques. *Computers & Geosciences*, 01–09 (2010)
4. Allmann, P.B., Shearer, P.M., Hauksson, E.: Spectral discrimination between quarry blasts and earthquakes in Southern California. *Bulletin of the Seismological Society of America* 98(4), 2073–2079 (2008)
5. Ursino, A., Langer, H., Scarfi, L., Grazia, G.D., Gresta, S.: Discrimination of quarry blasts from tectonic microearthquakes in the Hyblean Plateau (south-eastern Sicily). *Annali di Geofisica* 44, 703–722 (2001)
6. Orlic, N., Loncaric, S.: Earthquake-explosion discrimination using genetic algorithm-based boosting approach. *Computers & Geoscience* 36, 179–185 (2010)
7. Ruiz Reyes, N., Vera Candeas, P., Garcia Galan, S., Munoz, J.E.: Two-stage cascaded classification approach based on genetic fuzzy learning for speech/music discrimination. *Engineering Applications of Artificial Intelligence* 23, 151–159 (2010)
8. Kovacic, Z., Bogdan, S.: *Fuzzy Controller Design -Theory and Applications*. Taylor & Francis Group, LLC, Boca Raton (2006)
9. Sivanandam, S.N., Sumathi, S., Deepa, S.N.: *Introduction to Fuzzy Logic using MATLAB*. Springer, Heidelberg (2007)

Ultra Wide-Band Channel Characterization Using Generalized Gamma Distributions

Zakaria Mohammadi¹, Rachid Saadane^{1,2}, and Driss Aboutajdine¹

¹ GSCM-LRIT Laboratory Associated with CNRST,
Faculty of Science, University Mohammed V- Agdal, Rabat.

² LETI laboratory, Ecole Hassania des Travaux Publiques,
Km 7 Route d' El Jadida, B.P 8108, Casa-Oasis, Casablanca
Zakariamhm@gmail.com, rachid.saadane@ehpt.ac.ma,
aboutaj@fsr.ac.ma

Abstract. In this paper, we present an experimental characterization of the ultra-wide bandwidth (UWB) indoor channel using Generalized Gamma distributions. This investigation is also based on the analysis of the statistical properties of the multipath profiles when thresholding the estimated Power Delay Profile PDP over a spaced measurement grid. This characterization procedure was applied to the ultra wideband channel measurements collected from a measurement campaign, which has been performed within the whyless.com project by the IMST group, over a bandwidth of 10 GHz.

Keywords: Ultra Wideband, Power Delay Profile, Generalized Gamma distribution, Small Scale Fading, channel Modeling.

1 Introduction

The challenge of the recent communication systems is to field the user's requirement on low power consumption, little interferences with other systems, and high rate transmission. Therefore, ultra wideband (UWB) radio is an emerging technology, and has received great interest from the research and industry community, notably for Wireless Personal Area Network WPAN. It consists generally to transmit very short pulses, over a large frequency bandwidth, in the order of 500 MHz to several GHz, according to the specification of the Federal communication commission (FCC) [1]. However, an appropriate knowledge of the UWB channel proves necessary for the design, simulation, and performance evaluation of such systems. Many characterizations have been performed for several environments: Indoor, Outdoor, Corridor, Office... [2][3][4]. All this contributions are based on experimental measurements of the UWB channel. Many measurements campaigns have been performed within the past few years, mainly due to emerging UWB standards (e.g. multiband OFDM-UWB, IEEE 802.15.4a, and IEEE 802.15.3c). It can be done either in frequency or time domain. Usually, time domain sounding consists to the transmission of a pseudo-noise sequence and to estimate the channel impulse response by correlating the emitted sequence with the received one [5]. While this technique is

more suited to outdoor context, the frequency domain sounding is more adapted to indoor measurements, by using a Vector Network Analyzer VNA, which emits a series of tones with frequency f at Port 1 and measures the relative amplitude and phase with respect to Port 2, providing automatic phase synchronization between the two ports. This technique was performed by the IMST group using a frequency VNA to acquire indoor UWB channel data. The rest of this paper is organized as follow: The second section describes briefly The IMST measurements campaign setup, scenarios and estimation of the channel response. The Generalized Gamma (GG) distribution is presented in the third section, while we describe the procedure by which we processed the measured data to extract a set of model parameters, before evaluating and comparing the model with the available Power Delay Profile PDP.

2 Measurement Campaign

To study the characterization of the UWB channel using GG distribution, we worked with the IMST Data, collected from a measurement campaign done within the Whyless.com project, in frequency domain using a VNA in indoor environment as shown in Fig .1. We are also interested in only two scenarios: the Line Of Sight (LOS) and Non-Line Of Sight (NLOS). All the radio measurements have been performed at the IMST premise within an office with dimension $5\text{m} \times 5\text{m} \times 2.6\text{m}$. The office has a single door, one wall with windows, and contains a metal cabinet. Both the transmitter and receiver deploy a bi-conical horn antenna with approx. 1dBi gain, positioned at a height of 1.5m . The attenuation and the phase of the channel response have been measured from 1 to 11 GHz with 6.25 MHz frequency spacing. The measurement process is described more in detail in [6]. We denote by local PDP the Power Delay profile measured at a fixed Receiver Rx and transmitter Tx distance, given as $PDP(\tau, d) = |h(\tau, d)|^2$. The small scale effects are then shown in the variation of the measured PDP caused by small change of the transmitter position.

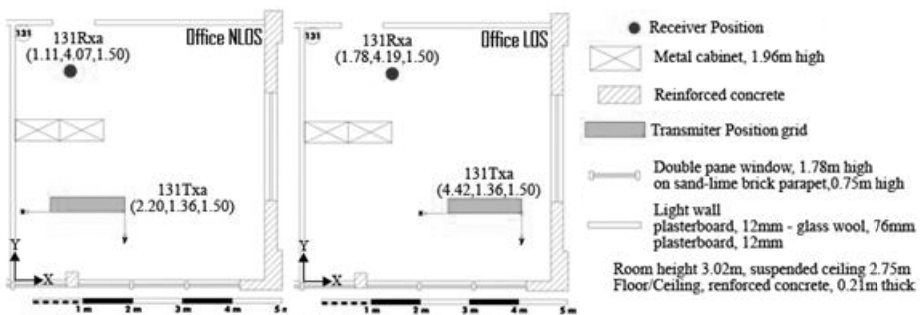


Fig. 1. Measurements Campaign Scenarios

Globally, the Tx position changes over a rectangular grid $150\text{cm} \times 30\text{cm}$ with a spatial resolution $\Delta d = 1\text{cm}$, resulting with a total of 4530 PDPs. Since the PDP amplitude vary when changing the Tx-Rx distance d , we proceed to a normalization of the PDPs with

respect to the first Multipath Component MPC arrival time, determined as $\tau_{\text{ref}} = d/c$, where c is the speed of light. Then it was identified as the origin of the delay axis. The goal of this normalization is to use the averaging technique safely, by anticipating the effect of smearing for an accurate extraction of statistical parameters.

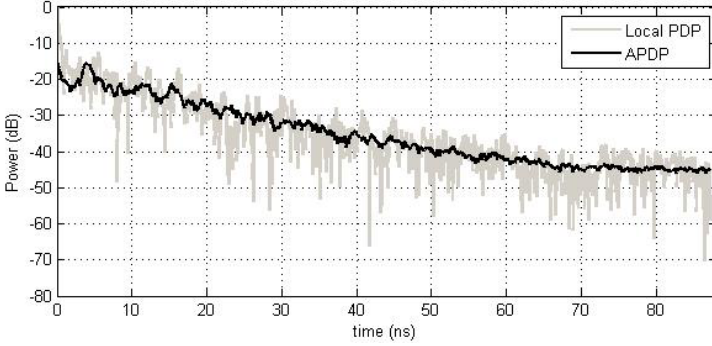


Fig. 2. Example of normalized estimated LOS-APDP in logarithmic scale

3 Generalized Gamma Distribution

Because of its flexibility and high quality adjustment, the Generalized Gamma Distribution was used in many fields like signal & image processing, mobile communication and many others. It was first introduced in [7]. The GGD Probability Density Function PDF is defined as follows:

$$f(x; \alpha, \beta, k, \gamma) = \frac{k(x-\gamma)^{k\alpha-1} \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^k\right)}{\beta^{k\alpha} \Gamma(\alpha)} \quad (1)$$

For $\gamma \leq x < +\infty$, where α and k are the positive real valued shape parameters, β the continuous scale parameter ($\beta > 0$), γ the location parameter ($\gamma = 0$ yields the three parameters GGD) and $\Gamma(\bullet)$ is the Gamma function. Among the interesting properties of this distribution, it has one more parameter than the most used distributions rendering it more flexible to the measurement data, moreover it contains a large variety of other distributions for different values of scale and shape parameters: Rayleigh ($k=2, \alpha=1$), exponential ($k=1, \alpha=1$), Nakagami ($k=2$), Gamma ($k=1$), log-normal ($\alpha \rightarrow \infty$), and Weibull ($\alpha=1$). For assessing the goodness of fit, we have to estimate the model-parameters. Since the APDPs values are all positive, the location parameter γ is assumed zero-valued ($\gamma=0$). To estimate the parameters α, β and k , we adopt the Maximum Likelihood ML method. Assuming $X = \{X_1, X_2, \dots, X_N\}$ a vector of mutually independent data, the Log likelihood is expressed as:

$$\begin{aligned} L(X; \alpha, \beta, k) &= \log f_x(X; \alpha, \beta, k) \\ &= N \log \left(\frac{k}{\beta^{k\alpha} \Gamma(\alpha)} \right) + (\alpha k - 1) \sum_{i=1}^N \log(x_i) - \sum_{i=1}^N \left(\frac{x_i}{\beta} \right)^k. \end{aligned} \quad (2)$$

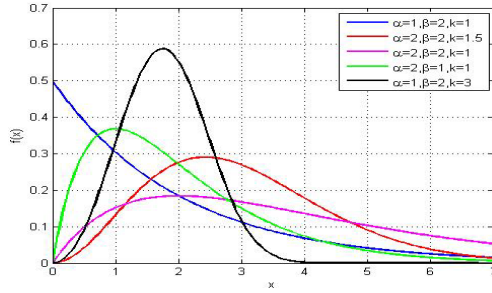


Fig. 3. PDF of the GID with $\gamma=0$ for different values of α , β and k

By settings the derivative of this function with respect to α , β and k to zero, we obtain the expression of the Maximum Likelihood ML estimators for the three parameters. Specifically, the expression of the estimators is:

$$\frac{\partial L(X; \alpha, \beta, k)}{\partial \alpha} = -N(k \log(\beta) - \psi(\alpha)) + \sum_{i=1}^N k \log(x_i) = 0, \psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}. \quad (3)$$

$$\frac{\partial L(X; \alpha, \beta, k)}{\partial k} = N \left(\frac{1}{k} - \alpha \log(\beta) \right) + \sum_{i=1}^N \alpha \log(x_i) - \left(\frac{x_i}{\beta} \right)^k \log \frac{x_i}{\beta} = 0. \quad (4)$$

$$\frac{\partial L(X; \alpha, \beta, k)}{\partial \beta} = -\frac{N\alpha k}{\beta} + \frac{k\beta^{-k}}{\beta} \sum_{i=1}^N x_i = 0. \quad (5)$$

Then the estimators can be straightforwardly determined by solving a three equations system [8]. After computing the estimated parameters, we use the test of Kolmogorov Smirnov [9] to evaluate the Goodness-Of-Fit of the fitted distribution regarding the data-set. It consists to test how well a hypothesized distribution function $F(x)$ fits an empirical distribution function $F_n(x)$, which equals the fraction of x_i that are less than or equal to x for each $-\infty < x < +\infty, i.e.$

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{x_i \leq x\}} \quad (6)$$

One of the simplest measures is the largest distance between the two functions $S(x)$ and $F(x)$, measured in a vertical direction. This statistic was suggested by Kolmogorov, and was used in our simulations for evaluating quality of adjustment.

4 Simulations and Results

In the majority of results, the power of APDP varies in logarithmic scale between 0db and -50db with respect to the strongest MPC arrived at τ_{ref} . The first simulation consists to apply several thresholds to the APDPs and computing the number of

MPCs arriving within the cut-off threshold. The usual thresholding processing comprises cutting off all data below a previously determined threshold and keep only the MPCs above it. In this simulation, the average MPCs number was calculated for several threshold values in both scenarios. We can observe an increase of the MPCs number when decreasing the threshold value. Then it is possible to estimate the total energy carried by the thresholded APDP. The large values of captured energy indicate that the temporal dispersion parameters can be estimated accurately, as denoted in [10]. Otherwise, this will enhance the estimation of temporal parameters when the total energy captured is more than 90%. Then the APDPs can be cutting-off using appropriate thresholds. The adopted threshold was -45dB and -35dB for LOS and NLOS respectively, corresponding to average MPCs number of 720 and 780 components. Then the 45 resulting APDPs are referred to our GFD-Model using the ML estimator to extract different values of shape and scale parameters. It was found that the Log-Logistic distribution gives the best agreement with the empirical distribution of the shape parameter α for different APDPs delay bin in both scenarios. Previous results showed that the 45 shape parameters α values are well modeled as log-logistic, denoted as $\text{LL}(\alpha, \beta, \gamma)$ with $(\alpha=92.869, \beta=0.68477, \gamma=0.32788)$ for LOS, while $(\alpha=39.336, \beta=0.24988, \gamma=0.75758)$ for NLOS. The table 1 shows the goodness of fit using the test of Kolmogorov-Smirnov for Log-Logistic, Normal, Log-Normal and Nakagami, which gives the best values of KS test. The histograms of the experimental α and the theoretically fitted distribution are shown in Fig.4. Applying the same procedure to characterize the second shape parameter k , we found that it is also log-normally and log-logistically distributed for LOS and NLOS respectively, i.e. $k_{\text{LOS}} \approx \text{LN}$ ($\sigma=1.6589, \mu=-6.8785, \gamma=8,0738 \text{ E-5}$), and $k_{\text{NLOS}} \approx \text{LL}$ ($\alpha=1,0781, \beta=0,00339, \gamma=4,9560 \text{ E-4}$), while assuming that the large values of parameter k correspond to the first multipath components, and the small k -values are associated with the later MPCs, with respect to linear fitting value for both scenarios.

Table 1. Goodness-Of-Fit for shape parameter α

		Log-Logistic (α, β, γ)	Normal (σ, μ)	Log-Normal (σ, μ, γ)	Nakagami (m, Ω)
LOS	Statistic	0.0541	0.09914	0.09785	0.09923
	Distribution Parameters	(92.869, 0.68477, 0.32788)	(0.01531, 1.0131)	(0.01512, 0.01293, 0.32168)	(1096.6, 1.0267)
NLOS	Statistic	0.02979	0.04217	0.0388	0.04159
	Distribution Parameters	(39.336, 0.24988, 0.75758)	(0.01164, 1.0079)	(0.02807, -0.8761, 0.5913)	(1874.1, 1.0159)

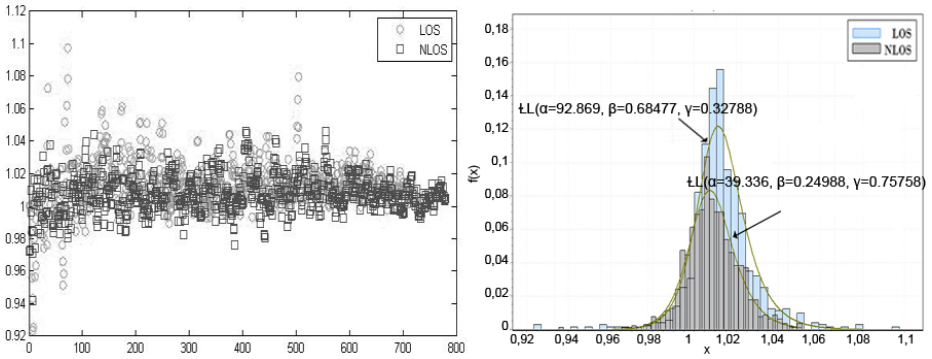


Fig. 4. Values and histograms of the shape parameter α and the theoretically fitted distribution

The previous shape parameters determine the shape of the distribution, while the scale parameter β determine the statistical dispersion of the probability distribution. If β presents large values, the distribution will be more spread out. Otherwise, the distribution will be more concentrated. The results indicate that the first LOS delay bins are more expanded than the others, whereas we notice that as far as the bin delay index increases, for LOS and NLOS scenarios, the bin values becomes more spread. This can be seen through FIG.5. Note that using the K-S test, we find that the scale parameter follows a Log-Logistic distribution for LOS scenario $\beta_{LOS} \approx \text{LL}(\alpha=2.829, \beta=15.91, \gamma=0.95684)$, while it is Log-Normally distributed for NLOS $\beta_{NLOS} \approx \text{LN}(\sigma=0.24244, \mu=3.2558, \gamma=-7.7394)$.

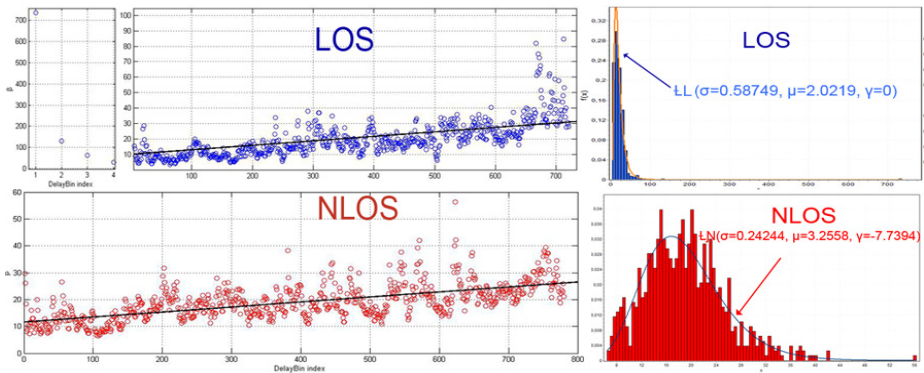


Fig. 5. Values and histograms of β with the theoretically fitted distributions

Then, it is possible to build simulated channel responses according to our simulation procedure, as described in FIG.6. In the model described above, the different shape and scale parameters are modeled with Log-Logistic or Log-Normal distributions, depending on the scenarios to simulate. The FIG.7 shows an example of comparison between a resulted APDP from our model with experimental one. It can be shown that the resulted Averaged Power Delay Profile reproduce roughly the same trend of our experimental data.

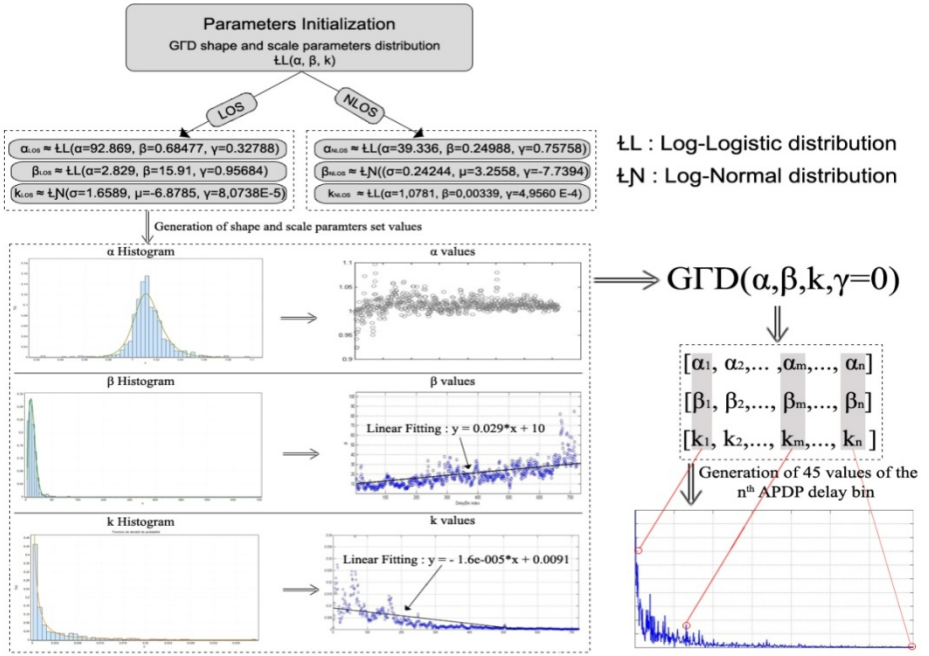


Fig. 6. Flowchart of the procedure for generating simulated APDP

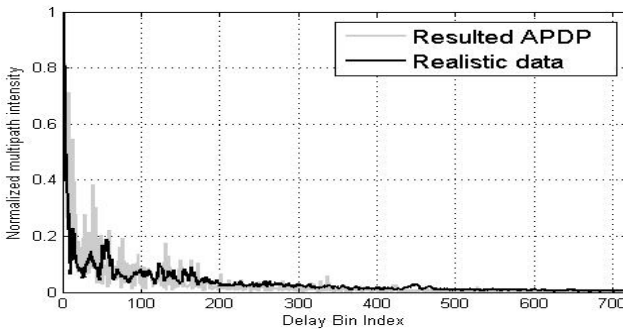


Fig. 7. Example of simulated VS Realistic APDP in LOS scenario

For the moment, the qualitative visual inspection remains the used method to evaluate the quality of modeling, while there are several other ways to assess the proposed model. Among them, quality of signal, calculated by integrating the theoretical and the modeled APDPs over all delays bin. This technique remains sometimes inaccurate, because of possibility of signal amplification which can affect its quality. However, the most used technique consists to compare some temporal parameters between the experimental response and the result built with the proposed model. More precisely, most authors use the Root mean square (RMS) delay spread

τ_{RMS} as a significant value which provide the time dispersion and the frequency selectively of the Power Delay Profile due to Multipath propagation. Moreover, to improve our model, we can enhance our flowchart by introducing some modification of our modeling process: for example, the linear fitting for the shape and scale parameters k and β proves be less suitable to describe the ascending and descending trend of this parameters. For this reason, we can use exponential or polynomial fitting.

5 Conclusion

In this paper, we performed a statistical analysis of UWB channel realizations, obtained from a measurement campaign in an indoor office environment using the parametric Generalized Gamma Distribution. Based on these results, we proposed a statistical model for generating UWB propagation channel. The experiment shows that the resulted channel Power Delay Profile can reproduce the same trend of the real channel response. In this work, we are limited to evaluate the quality of modeling with visual inspection, while we expected to improve our basic model and to assess our model using more developed tools are discussed before in future works.

References

1. Federal Communication Commission (FCC): First Report and Order in the Matter of Revision of Part 15 of the Commission's rule regarding Ultra-wideband transmission Systems. ET-Docket 98-153, FCC 02-48 (released, April 2002)
2. Udary, N., Kantz, K., Viswanathany, R., Cheung, D.: Characterization of Ultra Wideband Communications in Data Center Environments. In: IEEE International Conference on Ultra-wideband, Singapore (September 2007)
3. Alvarez, A., Valera, G., Lobeira, M., Toress, R., Garcia, J.L.: Ultra wideband channel characterization and modeling. In: Proc. Int. Workshop on Ultra Wideband Systems, Oulu, Finland (June 2003)
4. Rusch, L., Prettie, C., Cheung, D., Li, Q., Ho, M.: Characterization of UWB propagation from 2 to 8 GHz in a residential environment. IEEE Journal on Selected Areas in Communications (submitted for publication)
5. Rappaport, T.S.: Wireless Communications - Principles and Practice. Prentice Hall PTR, Upper Saddle River
6. Kunisch, J., Pamp, J.: Measurement results and modeling aspects for UWB radio channel. In: Proc. of IEEE Conference Ultra Wideband Systems and Technologies, pp. 19–23 (2002)
7. Stacy, E.W.: A generalization of the Gamma distribution. Ann. Math. Statist. 33(3), 1187–1192 (1962)
8. Chang, J.H., Shin, J.W., Kim, N.S., Mitra, S.K.: Image Probability Distribution Based on Generalized Gamma Function. IEEE Signal Process. Lett. 12(4), 325–328 (2005)
9. D'Agostino, R., Stephens, M.: Goodness-of-Fit Techniques (Marcel Dekker) (1986)
10. Molisch, A.F., Cassioli, D., Chong, C.C., Emami, S., Fort, A., Kannan, B., Karedal, J., Kunisch, J., Schantz, H.G., Siwiak, K., Win, M.Z.: A comprehensive standardized model for UWB propagation channels. IEEE Trans. Antennas and Propagation 54(11), 3151–3166 (2006)

Design of an Antenna Array for GNSS/GPS Network

Hocine Hamoudi^{1,2}, Boualem Haddad², and Philippe Lognonné³

¹ National Institute of the Post and I.C.T, Box 156, 16220 Eucalyptus, Algiers, Algeria
h_hamoudi@inptic.edu.dz

² University of Sciences and Technology Houari Boumedién, Laboratory of Image Processing
and Radiation, Box 32, 16111 El Alia, Bab Ezzouar, Algeria
bhaddad_57@yahoo.fr

³ Institut de Physique du Globe de Paris (IPGP), 94107 Saint Maure, France
lognonne@ipgp.fr

Abstract. This work focuses precisely on the design of a smart antenna printed on dielectric substrate operating at frequency L1 = 1575.42 MHz. This device consists of an antenna array to be integrated, in a GNSS/GPS network, with the aim of detecting ionosphere disturbances associated with land-based. To address such concerns, we studied an antenna array, consisting of four square elements, patch type, operating at the L1 frequency. As a first step, a simple square printed radiating structure was designed to test adaptation and radiation characteristics. In a second step, a square shape (2 * 2) antenna array has been studied. This type of sensor (networks) should respond no later than fifteen minutes after the main shock of an earthquake.

Keywords: Antenna array, GNSS antenna, ionosphere perturbations.

1 Introduction

At the end of the last decades, significant progress in detecting and modeling the ionosphere disturbances induced by seismic waves, were performed. This research is now an important part of the assignment and monitoring projects in the upper atmosphere [1] [2]. We are particularly interested in the behavior of the ionosphere, which is considered the seat of physical phenomena such as refraction and reflection of electromagnetic waves.

The extremely low power level delivered at the satellite makes the GPS prone to jamming and interference disturbances. Interference in question may be unintended equipment produced by other radio or from hostile interference. Studies by [3] and [4] show that an antenna array and adaptive directional beams are a promising method to overcome this problem. The use of high frequencies and microwave systems for microstrip structure has been responsible for the development of printed antennas. They are most often used in networks to improve performance and to enable the achievement of very specific functions, such as pointing and scanning electronic jammers rejection, and adaptive detection. There are two possible methods to dynamically change the radiation pattern and mitigate the effects of interference and multipath while increasing the coverage and distance [3]: (1) The switched beam and

(2) the adaptive beamforming. However, the first system provides limited performance compared to the second one. Therefore, we will focus our work on adaptive beam antennas which can react dynamically to changing RF environment. The unwanted signals (interferences) are deleted, while the main beam is directed to the signal. It is precisely to meet these requirements we are considering to use the antenna arrays controlled by servo. The servo leads the radiation beam to point to a desired mobile user and tracks the moving user, while minimizing interference from other users. Precisely, a part of this result we will try to achieve, in this study for a design in the industry, a network of smart antennas. The present work will be divided into two main parts. The first one will be devoted to the study of a single element (patch) square, while the second part, will address the computations and optimization of an antenna array.

2 Simple Element Design

The probably most two popular GNSS L1 antenna implementations are the patch and helix approaches. Nevertheless, others also exist [5]. This patch will be fed by a microstrip line with an impedance of 50Ω . The study of such a unit cell allows us to observe their behavior and frequency radiation at the working frequency L1.

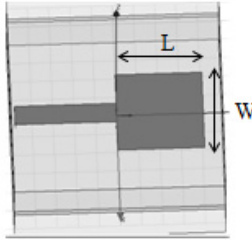


Fig. 1. Geometry of a single square patch

The described element will be design to operate at L1 frequency 1575.42 MHz. Our computations were based on transmission line theory. The width and length are given by the following equations [6]:

$$W = \frac{c}{2L_1 \sqrt{\frac{\epsilon_r + 1}{2}}} \quad (1)$$

Where the operating frequency $L_1=1575.42$ MHz, $c = 3 \cdot 10^8$ m/s and $\epsilon_r = 4.32$. Substituting above values, the width of the patch (W) becomes 61.1 mm. The effective length (L_{eff}) of the element can be calculated with equations (2) and (3).

$$L_{\text{eff}} = \frac{c}{2L_1 \sqrt{\epsilon_{\text{eff}}}} \quad (2)$$

And

$$\epsilon_{\text{eff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(\frac{1}{\sqrt{1 + 12h/W}} \right) \quad (3)$$

In this design, substrate has been used with parameters are: $\tan\delta = 0.001$, thickness of substrate (h) equal to 1.59 mm. Substituting $W = 61.1$ mm, $\epsilon_r = 4.32$ in equation (3), we get $\epsilon_{\text{eff}} = 4.10$. Hence $L_{\text{eff}} = 49.9$ mm. Due to the board effect, we derive the length extension (ΔL), using the following equation :

$$\Delta L = 0.412h \frac{(\epsilon_{\text{eff}} + 0.3) \left(\frac{W}{h} + 0.264 \right)}{(\epsilon_{\text{eff}} - 0.258) \left(\frac{W}{h} + 0.8 \right)} \quad (4)$$

Substituting $\epsilon_{\text{eff}} = 4.10$ and the values of W and h , we get $\Delta L = 0.741$ mm. In final, we obtain, using this equation: $L = L_{\text{eff}} - 2\Delta L = 49.9$ mm $-$ 1.482 mm $=$ 48.41 mm. However, some adjustments to the dimensions will be required in order to obtain a better adaptation and optimization of dimensions at the L1 frequency. The new dimensions of the patch, after optimization, are given in the following table.

Table 1. Dimensions of the patch

Dimensions (mm)	Theoretical values	After optimization
Length (L)	48.41 mm	43 mm
Width (W)	61.1 mm	43 mm
Thickness (h)	1.59 mm	1.59 mm

The major characteristics, at the desired frequency of a single element, of adaptation and radiation are shown, respectively, in the Fig. 2, Fig.3, Fig.4 and Fig.5.

A good agreement is obtained at 1575.42 MHz where the reflexion coefficient is about -14 dB. This value shows that the printed square patch antenna is better matched to its feeding strip line because $S_{11} \leq -10$ dB. In this investigation, VSWR is less than 2. Taking as a test transmission below -10 dB to define the frequency at which the patch works (Fig. 2), there is then a cut ranging from 1.55 GHz to 1.60 GHz - a bandwidth of about 4% compatible with the intended application. The antenna gain equal 4.91 dB in the direction of maximum radiation, that is perpendicular to the patch 0° . These results are satisfactory compared to the gain of patch antennas which seldom exceeds 6 dB [7].



Fig. 2. Reflexion coefficient (S_{11}) for single element

Taking as a test transmission below -10 dB to define the frequency at which the patch works (Fig. 2), there is then a cut ranging from 1.55 GHz to 1.60 GHz a bandwidth of about 4% compatible with the intended application.

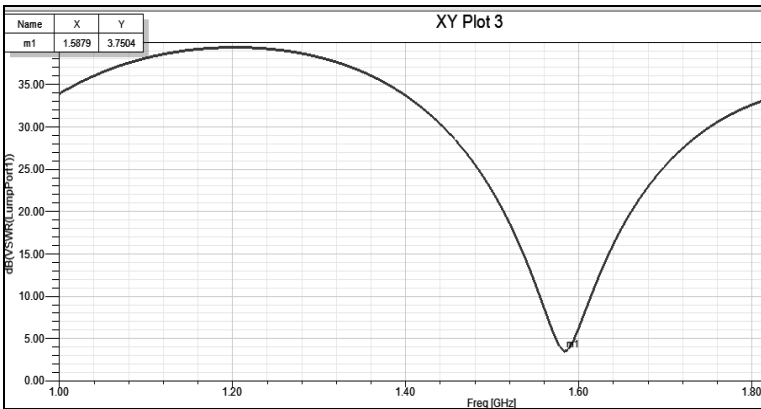


Fig. 3. VSWR for single element

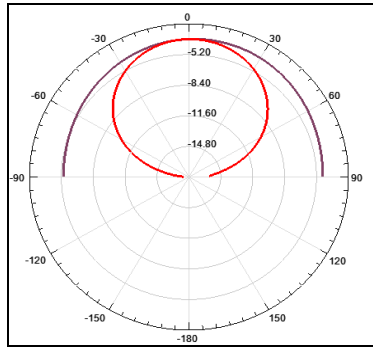


Fig. 4. Elevation pattern gain simulated for $\varphi=0^\circ$ (red) and $\varphi=90^\circ$ (purple)

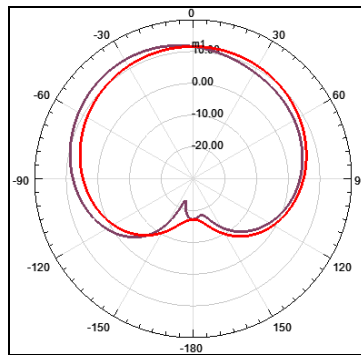


Fig. 5. Simulated radiation pattern at L1 frequency $\varphi=0^\circ$ (red) and $\varphi=90^\circ$ (purple)

3 Antenna Array Design

In the case of single square element, it has been observed that the antenna gain is quite low, not exceeding 6 dB. In order to increase the gain and improve the radiation characteristics, we use a network of antennas instead of a single antenna element. The major advantage of this array is its electronic scanning capability. Moreover, the major lobe can be steered toward any direction by changing the phase of the excitation current at each array element.

The most important point in the design of an antenna array is the feed network. In our case, we opted for a parallel feed. The parallel feed, also called the corporate feed, where the patch elements are fed in parallel by the power division transmission lines. The transmission line divides into two branches and each branch divides again until it reaches the patch elements. This is constructed by first connecting two adjacent elements together with a transmission line and this can be calculated from (5) and (6). Now, two separate groups, each containing two elements, need to be connected together. This is done with a transmission line drawn between the centers of the 0.35 mm wide transmission line. Figure 6 show the geometry of the proposed array.

$$Z_0 = \frac{\eta_0}{\sqrt{\epsilon_{\text{eff}}}} \left[\frac{W_e}{t} + 1.393 + 0.667 \ln\left(\frac{W_e}{t} + 1.444\right) \right]^{-1} \quad (5)$$

Where

$$\frac{W_e}{t} = \frac{W}{t} + \frac{1.25}{\pi} \frac{h}{t} \left[1 + \ln\left(\frac{4\pi W}{h}\right) \right] \quad (6)$$

Where W_e is the effective width of the patch, t is the thickness of the dielectric substrate, Z_0 is the impedance of the transmission line and η_0 is the free space intrinsic impedance. The transmission line is split using T-junction with equal power split. So both branches will receive input power, as is showing by the following equation:

$$P_{\text{in}} = \frac{1}{2} \frac{V_0^2}{2Z_0} \quad (7)$$

In that case where P_a and P_b be the output power, then

$$P_a = P_b = \frac{1}{2} P_{\text{in}} = \frac{1}{2} \frac{V_0^2}{2Z_{\text{out}}} \quad (8)$$

As equal power split is needed, the output impedance (Z_{out}) of the transmission line using (7) and (8) is obtained as $Z_{\text{out}} = 2Z_0$. In general, the impedance of a patch is between 100 and 400 Ω [8]. In our case, the transmission line is equal to 108 Ω and knowing that the impedance $Z_{\text{out}} = 2Z_0$. Hence $Z_0 = 54 \Omega$. We finally obtained a line width of 7 mm. The array is fed by a probe of diameter 1.2 mm in the middle of the thicker transmission line by using SMA of impedance 50 Ω . From equation (8) we get that the probe ideally should have an impedance of 48 Ω .

In this method, the inner conductor of the coax is connected to the patch through the substrate while the outer conductor is attached to the ground plane.

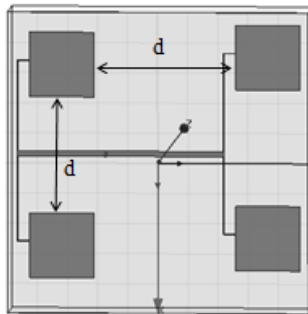


Fig. 6. Design of antenna array (2*2)

Particular interest should be worn to calculate the distance between element(s) in order to minimize the coupling between the elements. Several studies [8] and [9] showed that a distance of about $\lambda / 2$ reduces the effects of this phenomenon. Given that λ is the wavelength in the dielectric, equal to $\lambda_0 / \sqrt{\epsilon_{\text{eff}}}$ with λ_0 the wavelength in vacuum and ϵ_{eff} effective dielectric constant of the patch. In our case, the distance (d) between the elements of the antenna has been optimized and set at 8 cm. This parameter has a significant impact not only on minimizing coupling phenomena but also the shape of the radiation pattern [4].

In future work, we will focus on the inter-element distance patches to study the influence of this parameter and optimize the performance of our network. The antenna is made of right hand circular polarization (RHCP) and it is on this basis that the antenna was carried out. The choice of this polarization is defined and not arbitrary [10].

The microstrip antenna array radiates normal to its patch surface. The diagram consists essentially of a main lobe containing the maximum power, in the normal direction of the patch, suitable for our application. The simulated E-plane and H-plane pattern and the reflexion coefficient (S11) are illustrated in the figure 7 and 8. The reflexion coefficient of the antenna array is -14.82 dB at the L1 frequency for a peak gain at design frequency of 7 dB.

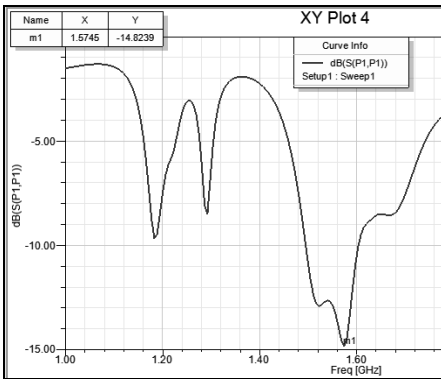


Fig. 7. Reflexion coefficient of the array

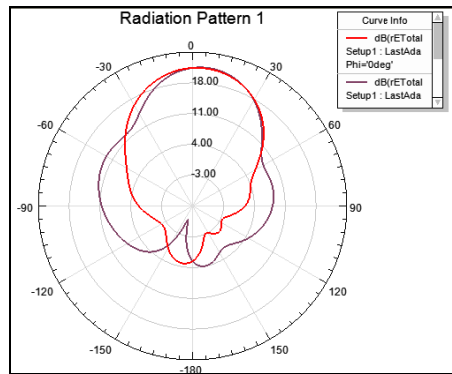


Fig. 8. Simulated E-plane and H-plane

4 Conclusion

The study of printed antennas shows that, we were able to design an antenna array consisting of four square elements, which will then be integrated into GPS/GNSS networks to monitor atmospheric phenomena. The results using simulation tools in the vicinity of 1547.42 MHz, showed satisfactory characteristics of adaptation and radiation. The appropriate approach to reach our goal is to study as a first step, and optimize the performance of a simple radiating component obtaining thus a reflexion coefficient close to -15 dB with a gain of 5 dB and a suitable radiation pattern. Indeed, the maximum radiation is obtained in the normal direction of the patch. The results are fully consistent with the results in the literature.

The second step consists in the optimization of antenna array by finding an optimal configuration of the network, addressing the key issue the distance between elements in order to obtain a small footprint while maintaining correct characteristics of adaptation and radiation. In this case, too, the radiation characteristics and adaptation remains satisfactory. We obtain a gain of about 7 dB and a radiation pattern consists mainly of a main lobe in the direction normal to the surface of the patch, suitable for our application. The inter-element distance was optimized and fixed to the half wavelength, about 8 cm

In the case of single element it has been observed that the antenna gain is quite low. But, while employing the array, gain increases significantly. This is one of the most advantages of the array structure. The results obtained at the L1 frequency, make it possible to design a network of printed antennas that can be used further for the design of an adaptive antenna.

The network thus proposed with high gain, low cost and small footprint meets our goal. As a perspective to our work, we propose the use of this network for the design of an adaptive antenna capable to modify the antenna pattern in order to have the benefits of the signal environment. A beam adaptive antenna arrays using controlled by a well-defined process. This control directs the radiation beam to a desired mobile user and tracks the moving user, while minimizing interference from other users by introducing nulls in the direction of interference.

References

1. Lognonne, P., Artru, J., Garcia, R., Crespon, F., Ducic, V., Jeansou, E., Occhipinti, G., Helbert, J., Moreaux, G., Godet, P.E.: Ground-based GPS imaging of ionospheric post-seismic signal. *Planetary and Space Science* 54, 528–540 (2006)
2. Blečki, J., Parrot, M., Wronowski, R.: Studies of the electromagnetic field variations in ELF frequency range registered by DEMETER over the Sichuan region prior to the earthquake. *Intern. J. of Remote Sensing* 31(13), 3615–3629 (2010)
3. Malmstrom, J.: *Robust Navigation with GPS/INS and Adaptive Beamforming*, Swedish Defence Research Agency (2003)
4. Fenn, A.J.: *Adaptive Antennas and Phased Arrays for Radar and Communications*. Artech House, Boston (2008)
5. Borre, K., Akos, D.M., Bertelsen, N., Rinder, P., Jensen, S.H.: *A software-defined GPS and GALILEO receiver: A single frequency approach*. Birkhauser, Boston (2007)
6. Sainati, R.A.: *CAD of micro strip antenna for wireless applications*. Artech House, Inc. (1996)
7. Gardiol, F.: Design and layout of micro strip structures. *Proceedings IEEE* 35(3), 145–157 (1988)
8. Visser, H.: *Array and phased array antenna*. Wiley, UK (2005)
9. Ghosh, C.K., Parui, S.K.: Design, Analysis and Optimization of A Slotted Microstrip Patch Antenna Array at Frequency 5.25 GHz for WLAN-SDMA System. *International Journal on Electrical Engineering and Informatics* 2(2), 106–110 (2010)
10. Rogstad, D.H., Mileant, A., Pham, T.T.: *Antenna arraying techniques in the deep space network*. JPL deep space communications and navigation series. Wiley (2003)

Blind Separation of Convolutive Mixtures of Non-stationary and Temporally Uncorrelated Sources Based on Joint Diagonalization

Hicham Saylani¹, Shahram Hosseini², and Yannick Deville²

¹ Laboratoire d'Electronique, de Traitement du Signal et de Modélisation Physique
Faculté des Sciences, Université Ibnou Zohr, BP. 8061, 80000 Agadir, Maroc

² Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse,
UPS-CNRS-OMP, 14 Av. Edouard Belin, 31400 Toulouse, France
h.saylani@uiz.ac.ma, {shosseini,ydeville}@irap.omp.eu

Abstract. In this paper, we propose a new method for blindly separating convolutive mixtures of non-stationary and temporally uncorrelated sources. It estimates each source and its delayed versions up to a scale factor by Jointly Diagonalizing a set of covariance matrices in the frequency domain, contrary to most existing second-order methods which require a Block Joint Diagonalization algorithm followed by a blind deconvolution to achieve the same result. Consequently, our method is much faster than these classical methods especially for higher-order mixing filters and may lead to better performance as confirmed by our simulation results.

1 Introduction

In this paper, we propose a new method for blindly separating convolutive mixtures of non-stationary and temporally uncorrelated signals. Consider M mixtures $x_i(n)$ of N discrete-time sources $s_j(n)$ and suppose the mixing filters are FIR (Finite Impulse Response). Denoting by $A_{ij}(z) = \sum_{k=0}^K a_{ij}(k)z^{-k}$ the transfer function of each mixing filter where K is the order of the longest filter, we can write

$$x_i(n) = \sum_{j=1}^N \sum_{k=0}^K a_{ij}(k)s_j(n-k), \quad i = 1, \dots, M. \quad (1)$$

This convolutive mixture may be rewritten as an instantaneous mixture [1–4] in the following manner. Considering delayed versions of the mixtures, i.e. $x_i(n-l)$ ($l = 0, 1, \dots, L-1$), Eq. (1) reads

$$x_i(n-l) = \sum_{j=1}^N \sum_{k=0}^K a_{ij}(k)s_j(n-(k+l)), \quad (i, l) \in [1, M] \times [0, L-1]. \quad (2)$$

These ML **generalized observations** $x_{il}(n) = x_i(n-l)$, $(i, l) \in [1, M] \times [0, L-1]$ can be then considered as instantaneous mixtures of $N(K+L)$ **generalized sources** $s_{jr}(n) = s_j(n-r) = s_j(n-(k+l))$, $(j, r) \in [1, N] \times [0, K+L-1]$.

This mixture is (over-)determined if $ML \geq N(K + L)$. It is clear that this condition may be satisfied only if $M > N$ i.e. if the original convolutive mixture is strictly over-determined. In this case, by choosing the integer number L so that $L \geq \frac{NK}{M-N}$, the reformulated instantaneous mixture (2) is (over-)determined. To represent the reformulated mixture in vector form, we define

$$\begin{aligned}\tilde{\mathbf{x}}(n) &= [x_{10}(n), x_{11}(n), \dots, x_{1(L-1)}(n), \dots, x_{M0}(n), x_{M1}(n), \dots, x_{M(L-1)}(n)]^T, \\ \tilde{\mathbf{s}}(n) &= [s_{10}(n), s_{11}(n), \dots, s_{1(K+L-1)}(n), \dots, s_{N0}(n), s_{N1}(n), \dots, s_{N(K+L-1)}(n)]^T,\end{aligned}$$

which yield using (2) :

$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(n), \quad (3)$$

$$\text{where } \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{M1} & \dots & \mathbf{A}_{MN} \end{pmatrix} \text{ and } \mathbf{A}_{ij} = \begin{pmatrix} a_{ij}(0) \dots a_{ij}(K) & 0 & \dots & 0 \\ & \ddots & & \ddots \\ & & \ddots & \\ 0 & \dots & 0 & a_{ij}(0) \dots a_{ij}(K) \end{pmatrix},$$

each block \mathbf{A}_{ij} being a matrix of dimension $L \times (K + L)$.

Then, Eq. (3) models an (over-)determined instantaneous mixture with $M' = ML$ observations $x_{il}(n)$ and $N' = N(K + L)$ sources $s_{jr}(n)$. The $M' \times N'$ mixing matrix $\tilde{\mathbf{A}}$ is supposed to admit a *pseudo-inverse* $\tilde{\mathbf{A}}^+$, called the separating matrix, that we want to estimate for retrieving the **generalized source vector** $\tilde{\mathbf{s}}(n)$.

Several second-order methods, initially developed for separating Linear Instantaneous Mixtures (LIM), have been reformulated in this manner and used to separate convolutive mixtures. For example, *SOBI* [5], *BGML* [6], and *TFBSS* [7] are three well-known methods proposed for separating LIM of mutually uncorrelated sources. Since the covariance matrix of the source vector, $\mathcal{R}_{\mathbf{s}}(n, \tau)$, is diagonal $\forall n, \tau$ for mutually uncorrelated sources, these methods jointly diagonalize a set of such matrices to achieve source separation. The approaches proposed in [1], [2] and [3], called respectively *SOBI-C*, *BGML-C* and *TFBSS-C* in the following, result from the generalization of these three methods to convolutive mixtures using the above reformulation. However, after reformulation the diagonality property of the covariance matrix of the generalized source vector, $\mathcal{R}_{\tilde{\mathbf{s}}}(n, \tau)$, is no longer met $\forall n, \tau$, but $\mathcal{R}_{\tilde{\mathbf{s}}}(n, \tau)$ is **block-diagonal**, whatever the nature of the original sources $s_j(n)$. As a result, the convolutive methods *SOBI-C*, *BGML-C* and *TFBSS-C* are based on **Joint Block-Diagonalization (JBD)** of a set of covariance matrices. The JBD algorithm provides several filtered versions of each initial source. Then, a blind deconvolution algorithm [3] may be used to estimate each of the generalized sources $s_{jr}(n)$, and in particular each of the initial sources $s_j(n)$, up to a scale factor.

In [4], we recently proposed a frequency-domain second-order approach for separating convolutive mixtures of non-stationary sources, also based on the reformulation of the mixture as an LIM and on the JBD. Contrary to the three methods mentioned above [1-3], our approach [4] requires neither global stationarity (supposed in [1]) nor piecewise stationarity (supposed in [2]) of the

sources nor their sparseness (supposed in [3]). Our simulation results in [4] using speech sources (i.e. non-stationary and temporally correlated signals) confirmed the better performance of this approach [4] compared to the other methods [1–3]. Nevertheless, the main drawback of all above four methods [1–4] is the high computational cost, especially for high-order mixing filters. This cost is mainly due to the JBD algorithm. That’s why we propose in this paper another algorithm which avoids JBD and blind deconvolution. We show that when the sources are non-stationary and temporally uncorrelated, it is possible to directly estimate each of the generalized sources $s_{j_r}(n)$ up to a scale factor just by jointly diagonalizing a set of covariance matrices in the frequency domain.

2 Proposed Approach

In [8], we proposed a new method for separating LIM of non-stationary and temporally uncorrelated signals based on the joint diagonalization of covariance matrices in the frequency domain. The approach proposed in the current paper is an extension of that method to convolutional mixtures and uses the same joint diagonalization algorithm as in [8]. Like in the initial method [8], we suppose that the initial sources $s_j(n)$ are

- (H1) : real and non-stationary,
- (H2) : zero-mean and temporally uncorrelated, i.e. $\forall j, \forall n \neq m, E[s_j(n)s_j(m)] = 0$,
- (H3) : mutually uncorrelated, i.e. $\forall j \neq k, \forall n, m, E[s_j(n)s_k(m)] = 0$.

Our spectral decorrelation method proposed in [8], which deals with frequency-domain sources $S_j(\omega)$ (the Fourier transforms of temporal sources $s_j(n)$), is based on the following principal properties¹:

- (P1) : *Uncorrelatedness* and *non-stationarity* in the time domain are transformed respectively into *wide-sense stationarity* and *autocorrelation* in the frequency domain. The **frequency-domain sources** $S_j(\omega)$ are then **wide-sense stationary** and **autocorrelated**.
- (P2) : Since the temporal sources $s_j(n)$ are mutually uncorrelated, their Fourier transforms $S_j(\omega)$ are **mutually uncorrelated** too.

Thanks to the linearity of the Fourier transform, by mapping the initial time-domain LIM into the frequency domain, we obtain another LIM with the same mixing matrix, but with respect to the frequency-domain sources $S_j(\omega)$ which are wide-sense stationary and autocorrelated. Then, we can separate them using the classical BSS algorithms initially developed for separating mixtures of time-domain wide-sense stationary, time correlated signals like *SOBI* [5]. The main advantage of our approach [8] is that thanks to the wide-sense stationarity in the frequency domain, the expected values involved in the computation of covariance matrices can be rigorously estimated by frequency averages. In the following, we denote by *SOBI-F* the frequency-domain version of the *SOBI* algorithm.

¹ See [8], and in particular Theorem 4, for more details.

Note finally that the separating matrix may be estimated by jointly diagonalizing covariance matrices if and only if the following two conditions are satisfied [8]:

- (C1) : the covariance matrix of the source vector² $\mathbf{s}(n)$, $\mathcal{R}_{\mathbf{s}}(n, \tau)$, is diagonal $\forall n, \tau$ (this condition is guaranteed by Hypothesis (H3)),
 (C2) : the sources $s_j(n)$ have *different normalized variance profiles*.

In [8], we showed that for a given frequency shift ν_1 , Condition (C2) is equivalent to the following identifiability condition:

$$\frac{E[S_i(\omega)S_i^*(\omega - \nu_1)]}{E[|S_i(\omega)|^2]} \neq \frac{E[S_j(\omega)S_j^*(\omega - \nu_1)]}{E[|S_j(\omega)|^2]}, \quad \forall i \neq j. \quad (4)$$

In the following, we present our extension of the above method to convolutive mixtures. As mentioned in Section 1, a convolutive mixture can be reformulated as an LIM mixture $\tilde{\mathbf{x}}(n) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(n)$. If we want to apply the above spectral decorrelation method (using a Joint Diagonalization algorithm) to this reformulated LIM for estimating the separating matrix $\tilde{\mathbf{A}}^+$, we must at first check the above two conditions (C1) and (C2). Nevertheless, we know that the matrix $\mathcal{R}_{\tilde{\mathbf{s}}}(n, \tau)$ is not diagonal $\forall n, \tau$, but only block-diagonal, whatever the nature of sources $s_j(n)$. However, using Hypothesis (H2) on the initial sources $s_j(n)$, we now show that this matrix is diagonal for $\tau = 0$. In fact, according to Hypothesis (H2), the generalized sources $s_{jr}(n)$ satisfy the following equation:

$$\forall j = k, \quad \forall r \neq d, \quad \forall n, \quad E[s_{jr}(n)s_{kd}(n)] = E[s_j(n-r)s_j(n-d)] = 0, \quad (5)$$

and using Hypothesis (H3) we can write:

$$\forall j \neq k, \quad \forall r, d, \quad \forall n, \quad E[s_{jr}(n)s_{kd}(n)] = E[s_j(n-r)s_k(n-d)] = 0. \quad (6)$$

Equations (5) and (6), together yield

$$\forall n, \quad E[s_{jr}(n)s_{kd}(n)] = \begin{cases} 0 & \forall j \neq k \text{ or } r \neq d \\ E[s_{jr}(n)^2] & \text{for } j = k \text{ and } r = d \end{cases} \quad (7)$$

Thus, the generalized sources $s_{jr}(n)$ are *instantaneously* mutually uncorrelated, so that the matrix $\mathcal{R}_{\tilde{\mathbf{s}}}(n, \tau)$ is diagonal for $\tau = 0$. We now propose a trick to transform these generalized sources $s_{jr}(n)$ into new sources which are mutually uncorrelated for every time lag τ so as to satisfy Condition (C1) and to apply our spectral decorrelation method for LIM. This trick is based on the following theorem.

Theorem 1. *Let $u_p(n)$ ($p = 1, \dots, \mathcal{N}$) be \mathcal{N} real, zero-mean and instantaneously mutually uncorrelated random signals i.e.*

$$\forall (p, q) \in [1, \mathcal{N}]^2, \quad p \neq q, \quad \forall n, \quad E[u_p(n)u_q(n)] = 0. \quad (8)$$

² In LIM, the considered source vector^T is defined as $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$.

Suppose $g(n)$ is a real, zero-mean, stationary, temporally uncorrelated random signal, independent from all signals $u_p(n)$. Then, the signals $u'_p(n)$ defined by $u'_p(n) = g(n)u_p(n)$ are real, zero-mean, **temporally uncorrelated** and **mutually uncorrelated**. Moreover, each new signal $u'_p(n)$ has the same normalized variance profile as the original signal $u_p(n)$.

Proof: See Appendix.

Multiplying each generalized observation $x_{il}(n)$ by a random signal $g(n)$ satisfying the conditions of the above theorem, we obtain new observations denoted by $x'_{il}(n) = g(n)x_{il}(n)$. These new observations are LIM of the new sources $s'_{jr}(n) = g(n)s_{jr}(n)$ with the same mixing matrix $\tilde{\mathbf{A}}$, because denoting $\tilde{\mathbf{x}}'(n) = g(n)\tilde{\mathbf{x}}(n)$ and $\tilde{\mathbf{s}}'(n) = g(n)\tilde{\mathbf{s}}(n)$ and using (3) we obtain

$$\tilde{\mathbf{x}}'(n) = g(n)\tilde{\mathbf{x}}(n) = g(n)(\tilde{\mathbf{A}}\tilde{\mathbf{s}}(n)) = \tilde{\mathbf{A}}(g(n)\tilde{\mathbf{s}}(n)) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}'(n). \quad (9)$$

Moreover, thanks to the above theorem (applied to signals $u_p(n) = s_{jr}(n)$), the new sources $s'_{jr}(n) = g(n)s_{jr}(n)$ are:

1. real and non-stationary with the same normalized variance profiles as the sources $s_{jr}(n)$,
2. zero-mean and temporally uncorrelated,
3. mutually uncorrelated for each time lag, i.e. $\mathcal{R}_{\tilde{\mathbf{s}}'}(n, \tau)$ is **diagonal** $\forall n, \tau$.

Thus, the first condition (C1) for applying our spectral decorrelation method for LIM is now satisfied because $\mathcal{R}_{\tilde{\mathbf{s}}'}(n, \tau)$ is diagonal $\forall n, \tau$. Besides, if the sources $s_{jr}(n)$ have different normalized variance profiles, then the new sources $s'_{jr}(n)$ have too so that the second condition (C2) is also verified. In this case, the new frequency-domain sources $S'_{jr}(\omega)$, which are the Fourier transforms of $s'_{jr}(n)$, satisfy the following identifiability condition

$$\forall j \neq k \text{ or } r \neq d, \quad \frac{E [S'_{jr}(\omega)S'^*_{jr}(\omega - \nu_q)]}{E [|S'_{jr}(\omega)|^2]} \neq \frac{E [S'_{kd}(\omega)S'^*_{kd}(\omega - \nu_q)]}{E [|S'_{kd}(\omega)|^2]}, \quad (10)$$

so that our spectral decorrelation method for LIM can be used to compute an estimate of the separating matrix $\tilde{\mathbf{A}}^+$, denoted $\tilde{\mathbf{A}}^+_{est}$. To this end, we start by computing the Fourier transform of the new observation vector $\tilde{\mathbf{x}}'(n) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}'(n)$ which yields:

$$\tilde{\mathbf{X}}'(\omega) = \tilde{\mathbf{A}}\tilde{\mathbf{S}}'(\omega), \quad (11)$$

where $\tilde{\mathbf{S}}'(\omega) = [S'_{10}(\omega), \dots, S'_{1(K+L-1)}(\omega), \dots, S'_{N0}(\omega), \dots, S'_{N(K+L-1)}(\omega)]^T$ and $\tilde{\mathbf{X}}'(\omega) = [X'_{10}(\omega), \dots, X'_{1(L-1)}(\omega), \dots, X'_{M0}(\omega), \dots, X'_{M(L-1)}(\omega)]^T$, with $X'_{il}(\omega)$ the Fourier transform of $x'_{il}(n)$. The modified generalized sources $s'_{jr}(n)$ being zero-mean, non-stationary, temporally uncorrelated and mutually uncorrelated, their Fourier transforms $S'_{jr}(\omega)$ are wide-sense stationary, autocorrelated and mutually uncorrelated, thanks to Properties (P1) and (P2). Therefore, we can apply the *SOBI-F* algorithm to compute $\tilde{\mathbf{A}}^+_{est}$ as follows:

1. we compute the $N' \times M'$ whitening matrix \mathbf{W} which yields a new observation vector $\tilde{\mathbf{Z}}'(\omega) = \mathbf{W}\tilde{\mathbf{X}}'(\omega)$ so that $E[\tilde{\mathbf{Z}}'(\omega)\tilde{\mathbf{Z}}'^H(\omega)] = \mathbf{I}_{N'}$, by diagonalizing the matrix $\mathcal{R}_{\tilde{\mathbf{X}}'}(0) = E[\tilde{\mathbf{X}}'(\omega)\tilde{\mathbf{X}}'^H(\omega)]$,
2. we compute the rotation matrix \mathbf{U} by Jointly Diagonalizing (JD) several covariance matrices $\mathcal{R}_{\tilde{\mathbf{Z}}'}(\nu_q) = E[\tilde{\mathbf{Z}}'(\omega)\tilde{\mathbf{Z}}'^H(\omega - \nu_q)]$ ($q = 1, 2, \dots$),
3. an estimate of the separating matrix $\tilde{\mathbf{A}}^+$ is given by:

$$\tilde{\mathbf{A}}_{est}^+ = \Re\{\mathbf{U}^H\mathbf{W}\} \simeq \mathbf{PD}\tilde{\mathbf{A}}^+, \tag{12}$$

where \mathbf{P} is a permutation matrix and \mathbf{D} is a real diagonal matrix [5, 8].

Once $\tilde{\mathbf{A}}_{est}^+$ has been computed by this method, we can directly find an estimate of the generalized source vector $\tilde{\mathbf{s}}(n)$, denoted by $\tilde{\mathbf{s}}_{est}(n)$, using (3) as follows:

$$\tilde{\mathbf{s}}_{est}(n) = \tilde{\mathbf{A}}_{est}^+ \tilde{\mathbf{x}}(n) \simeq (\mathbf{PD}\tilde{\mathbf{A}}^+)(\tilde{\mathbf{A}}\tilde{\mathbf{s}}(n)) \simeq \mathbf{PD}\tilde{\mathbf{s}}(n). \tag{13}$$

Thus, each generalized source $s_{jr}(n)$, and in particular each initial source $s_j(n)$ ($= s_{j0}(n)$), can be estimated up to a scale factor (and a permutation). In the following, we call our method **SOBI-F-C_{JD}**.

3 Simulation Results

In this section, we present our simulation results using $M = 3$ artificial FIR convolutive mixtures of $N = 2$ artificial sources containing $N_s = 65536$ samples. The sources are generated using $s_j(n) = r_j(n)\mu_j(n)$, where $r_j(n)$ are mutually uncorrelated, zero-mean i.i.d. (independent and identically distributed) Gaussian signals, $\mu_1(n) = \cos(\omega_0 n)$ and $\mu_2(n) = \sin(\omega_0 n)$ with $\omega_0 = \pi/7$. This choice allows us to generate two non-stationary and temporally uncorrelated initial sources $s_1(n)$ and $s_2(n)$ with different normalized variance profiles. The mixtures are generated using FIR filters of order $K \in \{1, 3, 5\}$. The coefficients $a_{ij}(k)$ of each transfer function $A_{ij}(z) = \sum_{k=0}^K a_{ij}(k)z^{-k}$ are generated randomly. For each value of K we choose in the model (2) the integer L equal to $2K$. This choice provides $M' = 6K$ generalized observations $x_{il}(n)$ and $N' = 6K$ ($\in \{6, 18, 30\}$) generalized sources $s_{jr}(n)$ so that the matrix $\tilde{\mathbf{A}}$ is square⁴.

To apply our **SOBI-F-C_{JD}** method, we first multiply all generalized observations $x_{il}(n)$ by an i.i.d., real, zero-mean and uniformly distributed signal $g(n)$, independent from the generalized sources. After whitening data as explained in the previous section, we jointly diagonalize 4 covariance matrices corresponding to 4 different frequency shifts, yielding an estimate of each of the generalized sources $s_{jr}(n)$ up to a scale factor.

We compare our results with those obtained using the time-domain method **BGML-C** [2] which exploits the non-stationarity of signals without requiring

³ ‘C’ for Convolutive and ‘JD’ for Joint Diagonalization.

⁴ Having originally 3 FIR mixtures of 2 sources, i.e. $M = 3$ et $N = 2$, we obtain $M' = ML = 6K$ and $N' = N(K + L) = 6K$ after reformulation as in (2) .

them to be temporally autocorrelated. To apply *BGML-C*, we consider 128 covariance matrices computed over 128 adjacent frames of 512 samples. To block-diagonalize these matrices, we use the orthogonal algorithm proposed by Févotte et al. in [3]. After the JBD stage, we obtain $K + L$ filtered versions of each initial source $s_j(n)$. Then, we use a blind deconvolution method proposed in [3] which allows us to estimate each of the generalized sources $s_{jr}(n)$ up to a scale factor. Performance is measured using the Signal to Interference Ratio (SIR) defined as $SIR = \frac{1}{2}(SIR_1 + SIR_2)$ where:

$$SIR_j = \max_r \left\{ 10 \log_{10} \left[\frac{E\{s_{jr}(n)^2\}}{E\{(\hat{s}_{jr}(n) - s_{jr}(n))^2\}} \right] \right\}, (j, r) \in [1, 2] \times [0, K + L - 1],$$

after normalizing the estimated generalized sources $\hat{s}_{jr}(n)$ so that they have the same variances and signs as the original generalized sources $s_{jr}(n)$. The SIR as well as the computation time⁵ are given in Table 1 for our method and the *BGML-C* method. We also repeat our simulations by varying the number of samples N_s . The results for $N_s \in \{2^{17}, 2^{18}, 2^{19}\}$ and $K = 1$ are shown in Table 2.

Table 1. SIR (in dB) and computation time T_j (in minutes) versus filter order K for $N_s = 2^{16} = 65536$

	$K = 1 (N' = 6)$			$K = 3 (N' = 18)$			$K = 5 (N' = 30)$		
Method	SIR	T_j (mn)	T_2/T_1	SIR	T_j (mn)	T_2/T_1	SIR	T_j (mn)	T_2/T_1
<i>SOBI-F-C_{JD}</i>	40.43	$T_1 = 0.04$		32.84	$T_1 = 0.20$		26.28	$T_1 = 0.50$	
<i>BGML-C</i>	28.07	$T_2 = 0.06$	1.50	10.18	$T_2 = 1.54$	7.70	7.14	$T_2 = 13.31$	26.62

Table 2. SIR (in dB) and computation time T_j (in minutes) versus number of samples N_s for $K = 1$

	$N_s = 131072$			$N_s = 262144$			$N_s = 524288$		
Method	SIR	T_j (mn)	T_2/T_1	SIR	T_j (mn)	T_2/T_1	SIR	T_j (mn)	T_2/T_1
<i>SOBI-F-C_{JD}</i>	42.90	$T_1 = 0.08$		46.22	$T_1 = 0.15$		53.33	$T_1 = 0.31$	
<i>BGML-C</i>	32.94	$T_2 = 0.09$	1.13	33.09	$T_2 = 0.18$	1.20	37.86	$T_2 = 0.36$	1.16

As can be seen:

- our method outperforms *BGML-C* in all of the tested configurations, especially for higher-order filters. For $K = 5$, it is about 26 times faster and leads to an SIR about 20 dB higher than *BGML-C*. This can be justified considering that *BGML-C* supposes the non-stationary signals to be piecewise stationary while this condition is not satisfied by our test signals, and

⁵ The algorithms were implemented on a 2.10 GHz Dual-Core Pentium processor with 3GB memory.

- it uses a JBD algorithm which is more time consuming than a JD algorithm,
- not surprisingly, for both methods the SIR increases with N_s and decreases with K , while the computation time increases with N_s and K .

4 Conclusion and Perspectives

In this paper, we proposed an extension of our spectral decorrelation method, initially developed for LIM, to convolutive mixtures. The proposed method, called *SOBI-F-C_{JD}*, may be used for separating convolutive mixtures of non-stationary and temporally uncorrelated sources. Just by using a joint diagonalization algorithm, it provides an estimate of each generalized source up to a scale factor, contrary to the existing approaches [1–4] which need a block-joint diagonalization algorithm followed by a blind deconvolution to achieve the same result. The first simulations confirmed the better performance of our method in terms of both separation quality and rapidity compared to the *BGML-C* method. Nevertheless, it would be interesting to confirm these results using more statistical tests. For example, increasing the number of covariance matrices used in JD algorithm would improve the performance.

References

1. Bousbia-Salah, H., Belouchrani, A., Abed-Meraim, K.: Blind separation of convolutive mixtures using joint block diagonalization. In: ISSPA, Kuala Lumpur, Malaysia, vol. 1, pp. 13–16 (August 2001)
2. Bousbia-Salah, H., Belouchrani, A., Abed-Meraim, K.: Blind separation of non stationary sources using joint block diagonalization. In: SPWSSP, Singapore, pp. 448–451 (August 2001)
3. Févotte, C., Doncarli, C.: A unified presentation of blind source separation methods for convolutive mixtures using block-diagonalization. In: ICA 2003, Nara, Japan (April 2003)
4. Saylani, H., Hosseini, S., Deville, Y.: Blind Separation of Convolutive Mixtures of Non-stationary Sources Using Joint Block Diagonalization in the Frequency Domain. In: Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R., Vincent, E. (eds.) LVA/ICA 2010. LNCS, vol. 6365, pp. 97–105. Springer, Heidelberg (2010)
5. Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E.: A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing* 45(2), 434–444 (1997)
6. Pham, D.-T., Cardoso, J.-F.: Blind separation of instantaneous mixtures of non stationary sources. *IEEE Trans. on Signal Processing* 49(9), 1837–1848 (2001)
7. Févotte, C., Doncarli, C.: Two contributions to blind source separation using time-frequency distributions. *IEEE Signal Processing Letters* 11(3), 386–389 (2004)
8. Hosseini, S., Deville, Y., Saylani, H.: Blind separation of linear instantaneous mixtures of non-stationary signals in the frequency domain. *Signal Processing* 89(5), 819–830 (2009)

Appendix: Proof of Theorem 1

Denote $\mathbf{u}(n) = [u_1(n), u_2(n), \dots, u_{\mathcal{N}}(n)]^T$ and $\mathbf{u}'(n) = g(n)\mathbf{u}(n)$.

1. Since $g(n)$ is independent from all the zero-mean signals $u_p(n)$, we can write

$$\forall p \in [1, \mathcal{N}], \quad E [u'_p(n)] = E [g(n)u_p(n)] = E [g(n)] E [u_p(n)] = 0. \quad (14)$$

Hence, the new signals $u'_p(n)$ ($p = 1, \dots, \mathcal{N}$) are also zero-mean.

2. Whatever the times n_1 and n_2 , we have

$$E [\mathbf{u}'(n_1)\mathbf{u}'(n_2)^T] = E [g(n_1)g(n_2)\mathbf{u}(n_1)\mathbf{u}(n_2)^T]. \quad (15)$$

The independence of $g(n)$ from all the signals $u_p(n)$ yields

$$E [\mathbf{u}'(n_1)\mathbf{u}'(n_2)^T] = E [g(n_1)g(n_2)] E [\mathbf{u}(n_1)\mathbf{u}(n_2)^T], \quad (16)$$

and since $g(n)$ is zero-mean, stationary and temporally uncorrelated

$$E [\mathbf{u}'(n_1)\mathbf{u}'(n_2)^T] = \sigma_g^2 \delta(n_1 - n_2) E [\mathbf{u}(n_1)\mathbf{u}(n_1)^T] \quad (17)$$

where σ_g^2 is the variance of $g(n)$. The signals $u_p(n)$ being zero-mean and instantaneously mutually uncorrelated, the matrices $E [\mathbf{u}(n_1)\mathbf{u}(n_1)^T]$ and so $E [\mathbf{u}'(n_1)\mathbf{u}'(n_2)^T]$ are diagonal. As a result, the new zero-mean signals $u'_p(n)$ are mutually uncorrelated. Moreover, according to Eq. (17), the diagonal entries of the matrix $E [\mathbf{u}'(n_1)\mathbf{u}'(n_2)^T]$ can be written as

$$E [u'_p(n_1)u'_p(n_2)] = \sigma_g^2 \delta(n_1 - n_2) E [u_p(n_1)u_p(n_1)] = \sigma_g^2 \delta(n_1 - n_2) E [u_p^2(n_1)]. \quad (18)$$

Hence, the new signals $u'_p(n)$ are temporally uncorrelated. Furthermore, by choosing $n_1 = n_2 = n$, Eq. (18) becomes $E [u_p'^2(n)] = \sigma_g^2 E [u_p^2(n)]$ which means that the new signals $u'_p(n)$ have the same normalized variance profiles as the original signals $u_p(n)$.

Maximizing Network Lifetime through Optimal Power Consumption in Wireless Sensor Networks

El Abdellaoui Saïd¹, Fakhri Youssef^{1,2}, Debbah Merouane³, and Aboutajdine Driss¹

¹ LRIT, Unité Associée au CNRST (URAC 29), Faculty of Sciences,
University Mohammed V- Agdal, Rabat, Morocco

² LARIT, équipe Réseaux et Télécommunication, Faculty of Sciences,
University Ibn Tofail, Kenitra, Morocco

³ Alcatel-Lucent Chair on Flexible Radio, Supelec, Gif-sur-Yvette Cedex, France
elabdellaoui.said@yahoo.fr, fakhri-youssef@univ-ibntofail.ac.ma,
merouane.debbah@supelec.fr, aboutaj@fsr.ac.ma

Abstract. Energy efficiency is a foremost concern in Wireless Sensor Networks (WSNs). It aims to maximize the network lifetime which is defined as the time duration until the battery depletion of the first node. The aim of our approach is to provide the optimal transmission power taking into account the signal to noise ratio (SNR) constraint at the Fusion Center (FC) while guaranteeing the required performance. In this article, we address the lifetime maximization problem under non-orthogonal channels assuming two cases. In the first case, the nodes have the perfect knowledge of all channel gains. While in the second case, we propose several extensions to the unacknowledged channel gains by the nodes. In both cases, we consider that the nodes transmit their data to the FC over Quasi-Static Rayleigh fading Channel (QSRC). Simulation results show that the proposed optimal power allocation method maximizes the network lifetime better than the EP method.

Keywords: Energy-Efficiency, WSNs, MIMO Cooperative, Cooperation Communication, Optimal Power Allocation.

1 Introduction

Wireless Sensor Networks (WSNs) represent a technological revolution resulting from convergence of electronic and wireless communication systems. A WSN is a special network composed from a large number of nodes equipped with an embedded processor, sensors and a radio. These nodes have very limited resources which should be wisely used while trying to provide an acceptable QoS. Since nodes in WSNs are battery powered and changing batteries is a very difficult operation due to highly varying topology and deployment characteristics, the energy consumption should be taken into account in order to maximize nodes lifetime. Then, the most important objective for the WSN is to maximize the network lifetime by making the nodes run for a long time. The network lifetime has been defined in various ways. It may be defined as the time until the first sensor runs out of energy as in [1], others have defined it as the time until the last sensor runs out of energy[2][3]. In this work, we

consider the first assumption. In literature, there are several works that have treated the same issue and which will be quoted in brief in this article.

We begin by mentioning the work of Belmega et al. in [4] which conclude that MIMO systems are more energy efficient than SISO systems if only the transmitted power consumption is taken into account. However, when the circuitry energy consumption is considered, this conclusion is no longer true.

In the WSN, however, the node cannot carry multiple antennas at the same time due to his limited physical size. Therefore, a new transmission technique called ‘‘Cooperative MIMO’’ has been proposed in [5] [6] for a better diversity reception. This technique is based on the cooperation principle where the participating nodes (relays) form a distributed antenna array to achieve the diversity gain of the MIMO system, in other words, the MIMO technology is virtually introduced.

Several studies have addressed the problem of maximizing the network lifetime using various methods for minimizing energy consumption. In [7] [8] optimal solutions are presented for maximizing a static network lifetime through a graph theoretic approach using static broadcast tree. In [8] Thomas et al. have presented an optimal solution for maximizing the network lifetime through a graph theoretic approach using a static multicast tree. Bhardwaj et al. [9], [10] have explored the fundamental limits of energy-efficient collaborative data-gathering by deriving upper bounds on the lifetime of increasingly sophisticated sensor networks assuming that sensor nodes only consume energy when they process, send or receive data. In [7], the authors have studied the node density vs. network lifetime tradeoff for a cell-based energy conservation technique.

In this paper, we introduce a novel method for maximizing the network lifetime under the non-orthogonal channels taking into account the total SNR constraint at the FC. The next part of this paper is organized as follows: Section 2 explains our method applied to the non-orthogonal channels considering that the nodes have direct access to the FC and they transmit their data over a QSRC. We assume two cases; in the first case, we consider that the nodes have the perfect knowledge of all channel gains. In the second, we consider that the nodes do not have knowledge of all channel gains. Section 3, presents the conducted experiments and the last section concludes the paper.

2 Background and Definitions

In this section, we give a background and precisely define the terms used throughout this paper. We assume a Fusion Centre (FC) and M sensors randomly distributed in the area of interest and these sensors have a direct access to the FC (see figure 1). We consider that the nodes transmit their data over quasi-static Rayleigh fading channels and each sensor has an initial energy noted by E_{int} .

We assume that the sensed observation, when a monitored event occurs, is contaminated with Additive White Gaussian Noise (AWGN) noted by n_i . The noisy observation from the i^{th} sensor can be written as:

$$x_i = \theta + n_i \quad (1)$$

Where θ is the actual parameter being measured and n_i is the additive complex Gaussian noise with $n_i \sim \mathcal{CN}(0; \sigma_{it}^2)$. To ensure that this noisy observation x_i to be transmitted to the FC, it must be multiplied by the transmitter gain w_i . We note that the transmission power is written as $p_i = w_i^2 (1 + \sigma_{it}^2)$ considering that $E[\theta^2] = 1$, where, $E[\cdot]$ is the mathematical expectation operator [14].

We assume that the noisy observations transmitted to the FC have another noise noted n_r with an additive complex Gaussian distribution $n_r \sim \mathcal{CN}(0; \sigma_r^2)$ and h_i is the i^{th} channel coefficient from the sensor i to the FC. We consider that $|h_i|$ has a Rayleigh distribution where σ_{hi}^2 represents the well known variance, where,

$$f(|h_i|) = \frac{|h_i| e^{\left(\frac{-|h_i|^2}{2\sigma_{hi}^2}\right)}}{\sigma_{hi}^2}$$

In this article, our goal is to maximize the network lifetime that is written as follows:

$$L = N * T \tag{2}$$

Where the T is the period measurement of channel condition (we consider that $T=1$ to simplify), N is the number of transmissions before the network misses energy. Consequently, to maximize the networks lifetime it is sufficient to maximize the number of transmission for each sensor, taking into account the estimation of overall SNR constraint at the FC, then the general formulation of our problem as:

$$\begin{cases} \text{Max } E[N] \\ \text{SNR} \geq \gamma \\ 0 \leq P_i \leq E_{int} \end{cases}$$

In our work, we focus on the optimal power allocation problem for WSNs under Non-Orthogonal channels assuming two cases quoted previously. In addition, we suppose a linear minimum mean square-error (LMMSE) detector is used at the receiver.

2.1 Non-orthogonal Channel (Known Channel States)

We assume M sensors randomly distributed in the area of interest using non-orthogonal channels between the FC and each sensor (Figure 1). We consider that the nodes have Channel State Information (CSI).

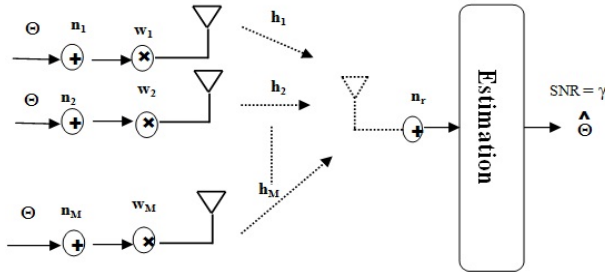


Fig. 1. System model

The received signal at the FC is defined by:

$$y = \sum_{i=1}^M h_i w_i (\theta + n_i) + n_r \quad (3)$$

Assuming that we use real channels, the SNR at the FC corresponding to M sensors using the MMSE detector is given by:

$$SNR = \frac{(\sum_{i=1}^M |h_i| w_i)^2}{\sum_{i=1}^M w_i^2 |h_i|^2 \sigma_{it}^2 + \sigma_r^2} \quad (4)$$

Our aim is to maximize the batteries lifetime duration while keeping the expected value of SNR greater than or equal to a target value γ .

At the l^{th} instant, maximizing the lifetime relies on minimizing the power consumption; therefore, the problem formulation is given as follows:

$$\begin{cases} \text{Min} \sum_{i=1}^M w_i^{(l)2} (1 + \sigma_{it}^2) \\ SNR^{(l)} \geq \gamma \quad l = 1, \dots, N \\ p_i \geq 0 \end{cases} \quad (5)$$

To find the optimal points, we use the Lagrange method while satisfying the constraints quoted before. The Lagrangian \mathcal{L} can be written as follows:

$$\begin{aligned} \mathcal{L}(w_i^{(l)}, \lambda, v) = & \sum_{i=1}^M w_i^{(l)2} (1 + \sigma_{it}^2) - \sum_{i=1}^M \lambda_i w_i^{(l)} \\ & + v_l \left[\gamma \left[\sum_{i=1}^M w_i^{(l)2} |h_i^{(l)}|^2 \sigma_{it}^2 + \sigma_r^2 \right] - \left[\sum_{i=1}^M w_i^{(l)} |h_i^{(l)}| \right]^2 \right] \end{aligned} \quad (6)$$

Let consider the Karush-Kuhn-Tucker (KKT) [14] conditions for the problem:

$$\begin{cases} \lambda_i \geq 0, v_l \geq 0, \lambda_i w_i^{(l)} = 0 \quad \forall i, l = 1, 2 \\ v_l \left[\gamma \left[\sum_{i=1}^M w_i^{(l)2} |h_i^{(l)}|^2 \sigma_{it}^2 + \sigma_r^2 \right] - \left[\sum_{i=1}^M w_i^{(l)} |h_i^{(l)}| \right]^2 \right] = 0 \\ \frac{\partial \mathcal{L}}{\partial w_k^{(l)}} = 0 \end{cases} \quad (7)$$

Then, the partial derivative of \mathcal{L} with respect to w_k is:

$$\frac{\partial \mathcal{L}}{\partial w_k^{(l)}} = 2 w_k^{(l)} (1 + \sigma_{kt}^2) - \lambda_k + 2 v_l \gamma |h_k^{(l)}|^2 \sigma_{kt}^2 w_k^{(l)} - 2 v_l |h_k^{(l)}| \left[\sum_{i=1}^M w_i^{(l)} |h_i^{(l)}| \right]$$

Taking into account the KKT conditions, we find that $v_1 > 0$ and $\lambda_k = 0$. Thus,

$$w_k^{(1)} = \frac{v_1 |h_k^{(1)}| (\sum_{i=1}^M w_i^{(1)} |h_i^{(1)}|)}{(1 + \sigma_{kt}^2) + v_1 \gamma |h_k^{(1)}|^2 \sigma_{kt}^2} \tag{8}$$

To find the value of $\sum_{i=1}^M w_i^{(1)} |h_i^{(1)}|$ we replace (8) in (7), and it becomes:

$$\left[\sum_{i=1}^M w_i^{(1)} |h_i^{(1)}| \right]^2 = \frac{\gamma \sigma_r^2}{1 - \gamma v_1^2 \left[\sum_{i=1}^M \frac{|h_i^{(1)}|^4}{\left[(1 + \sigma_{it}^2) + v_1 \gamma |h_i^{(1)}|^2 \sigma_{it}^2 \right]^2 \sigma_{it}^2} \right]} \tag{9}$$

Finally, equation (8) becomes:

$$w_k^{(1)} = \frac{v_1 |h_k^{(1)}| \sigma_r \sqrt{\gamma}}{\left[(1 + \sigma_{kt}^2) + v_1 \gamma |h_k^{(1)}|^2 \sigma_{kt}^2 \right] \sqrt{1 - \gamma v_1^2 \left[\sum_{i=1}^M \frac{|h_i^{(1)}|^4}{\left[(1 + \sigma_{it}^2) + v_1 \gamma |h_i^{(1)}|^2 \sigma_{it}^2 \right]^2 \sigma_{it}^2} \right]}}$$

Now, the challenge is to find the value of v_1 . Therefore, we multiply equation (8) by $|h_k^{(1)}|$, After that, we compute the sum of all the resulting equations, we obtain:

$$\sum_{i=1}^M |h_k^{(1)}| w_k^{(1)} \left[1 - \sum_{i=1}^M \frac{v_1 |h_k^{(1)}|^2}{(1 + \sigma_{kt}^2) + v_1 \gamma |h_k^{(1)}|^2 \sigma_{kt}^2} \right] = 0$$

Since $\sum_{i=1}^M h_k w_k^{(1)} \neq 0$, we obtain:

$$\sum_{i=1}^M \frac{|h_k^{(1)}|^2}{1 + \sigma_{kt}^2 \left[1 + v_1 \gamma |h_k^{(1)}|^2 \right]} = \frac{1}{v_1} \tag{10}$$

This equation is not written in a closed-form solution. So, it can be solved numerically using the function "Fminsearch" [13].

2.2 Non-orthogonal Channel (Unknown Channel States)

Since the previous assumption is not actually valid for some practical systems, then, in this section, we consider the same assumptions of the previous section except that in which the nodes do not have a Channel State Information (CSI). The received signal at the FC from i^{th} sensor is defined by:

$$y_i = h_i w_i (\theta + n_i) + n_r$$

The average SNR at the FC in this case is not similar to that in the previous section, it can be written as follows:

$$SNR = \frac{\sum_{i=1}^M w_i^2 |h_i^{(l)}|^2}{\sum_{i=1}^M \sigma_{it}^2 w_i^2 |h_i^{(l)}|^2 + \sigma_r^2} \quad (11)$$

Our aim is to maximize the lifetime of our network by taking into account the estimation of overall SNR at FC. To maximize the networks lifetime it is adequate to minimize the transmission power for each sensor, then, our problem formulation as:

$$\left\{ \begin{array}{l} \text{Min} \sum_{i=1}^M w_i^{(l)2} (1 + \sigma_{it}^2) \\ SNR \geq \gamma \\ P_i \geq 0 \end{array} \right. \quad (12)$$

To find the optimum power, we will use the Lagrange method as an optimization method, while satisfying the constraints quoted before. Using the equation (12), the Lagrangian \mathcal{L} can be written as follows:

$$\begin{aligned} \mathcal{L}(w_i^{(l)}, \lambda, \nu) = & \sum_{i=1}^M w_i^{(l)2} (1 + \sigma_{it}^2) - \sum_{i=1}^M \lambda_i w_i^{(l)2} \\ & - \nu \left[\sum_{i=1}^M w_i^2 |h_i^{(l)}|^2 - \gamma \sum_{i=1}^M \sigma_{it}^2 w_i^2 |h_i^{(l)}|^2 - \gamma \sigma_r^2 \right] \end{aligned}$$

With the Karush-Kuhn-Tucker (KKT) Conditions for the problem are given by:

$$\left\{ \begin{array}{l} \lambda_i \geq 0, \nu \geq 0, \lambda_i w_i^{(l)} = 0 \\ \left[\sum_{i=1}^M w_i^2 |h_i^{(l)}|^2 - \gamma \sum_{i=1}^M \sigma_{it}^2 w_i^2 |h_i^{(l)}|^2 - \gamma \sigma_r^2 \right] = 0 \\ \frac{\partial \mathcal{L}}{\partial w_k^{(l)}} = 0 \end{array} \right.$$

The partial derivative of \mathcal{L} with respect to $w_k^{(l)}$ is:

$$\frac{\partial \mathcal{L}}{\partial w_k^{(l)}} = 1 + \sigma_{kt}^2 - \lambda_k - \nu |h_k^{(l)}|^2 [1 - \gamma \sigma_{kt}^2] = 0$$

In that case, the Lagrangian method does not lead us to solve our problem. According to the equation (12), and knowing that the denominator is positive, we have:

$$\sum_{i=1}^M w_i^2 |h_i^{(l)}|^2 - \gamma \sum_{i=1}^M \sigma_{it}^2 w_i^2 |h_i^{(l)}|^2 \geq \gamma \sigma_r^2$$

Then,

$$\sum_{i=1}^M w_i^2 |h_i^{(l)}|^2 (1 - \gamma \sigma_{it}^2) \geq \gamma \sigma_r^2$$

We can observe that our equation is written as: $a^T X \geq b$ that is called a linear matrix inequality (LMI) [17] in x : $a^T X = x_1 a_1 + \dots + x_n a_n \geq b$, with $X = [w_1^2, w_2^2, \dots, w_M^2]$, $b = \gamma \sigma_r^2$ and $a = [|h_1^{(l)}|^2 (1 - \sigma_{1t}^2), |h_2^{(l)}|^2 (1 - \sigma_{2t}^2), \dots, |h_M^{(l)}|^2 (1 - \sigma_{Mt}^2)]$. Then, the problem can be solved numerically.

3 Simulation

Several simulations have been conducted using MATLAB in order to compare and evaluate the behavior of both the Equal Power (EP) method [16] and our novel approach. For each simulation, we study the network lifetime while increasing the number of nodes. The simulations parameters are generated randomly such that each parameter p belongs to a uniform distribution between ψ and φ , $p \in U[\psi; \varphi]$.

3.1 Non-orthogonal Channel (Known Channel States)

Figure 2 shows the lifetime network while increasing the number of nodes using non-orthogonal channels where the channel coefficients are known. As it can be seen, the proposed approach improves EP method concerning the network lifetime. Actually, the network lifetime is extended by an average of 82, 80%. Table 1 shows the parameters used for simulations.

Table 1. Simulations parameters

Estimate	Parameters
U[0.1, 0.2]	σ_{hi}^2 : The variances of channel estimation
0.5	σ_r^2 : The noise variance at the FC
U[0.02, 0.1]	σ_{it}^2 : The observation noise variances
U[200, 500]	ϵ_i : The initial energy

3.2 Non-orthogonal Channel (Unknown Channel States)

In Figure 3, we can observe that our new method is more effective than the EP method concerning network lifetime. The batteries lifetime duration is extended by an average of 79, 98%. Table 1 shows the parameters used for simulations.

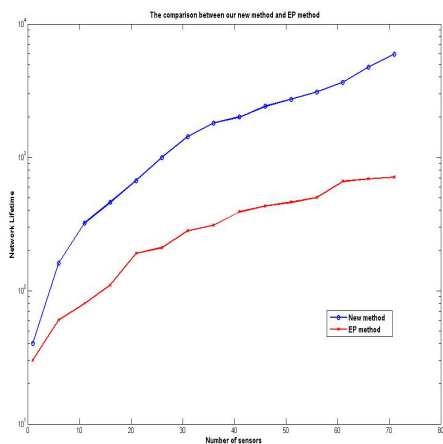


Fig. 2. Non-Orthogonal Channel (Known Channel States)

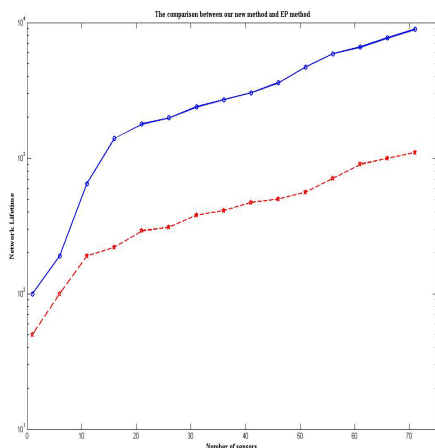


Fig. 3. Non-Orthogonal Channel (Unknown Channel States)

4 Conclusion

This paper presented a new algorithm which aims to maximize the network lifetime under Non-Orthogonal channel configuration. This method takes into consideration the estimation of the overall SNR at the FC. We showed that our new method in the both cases consumes less energy than the EP method. The future work is to adapt our new method to the wireless communication using energy harvesting transmitters.

References

1. Ratnasamy, S., Estrin, D., Govindan, R., Karp, B., Yin, L., Shenker, S., Yu, F.: Data-centric storage in sensor networks. In: Proceeding of the ACM SIGCOMM Computer Communication Review, Pittsburgh, PA, USA, pp. 137–142 (2002)
2. Choi, W., Shah, P., Das, S.K.: A Framework for Energy-Saving Data Gathering Using Two-Phase Clustering in Wireless Sensor Networks. In: Proceeding of the Mobile and Ubiquitous Systems, Boston, pp. 203–212 (2004)
3. Ephremides, A.: Energy concerns in wireless networks. *IEEE Wireless Communications* 9(4) (2002)
4. Belmega, E.V., Lasaulce, S., Debbah, M.: A survey on energy-efficient communications. In: *IEEE Intl. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC 2010)*, Istanbul, Turkey, p. 289 (2010) (invited paper)
5. Cui, S., Goldsmith, A.J., Bahai, A.: Energy efficiency of MIMO and cooperative MIMO techniques in sensor networks. *IEEE Jour. on Selected Areas in Comm.* 22(6), 1089–1098 (2004)
6. Nguyen, T., Berder, O., Sentieys, O.: Cooperative MIMO schemes optimal selection for wireless sensor networks. In: *IEEE 65th Vehicular Technology Conference*, pp. 85–89 (2007)

7. Winters, J.: The diversity gain of transmit diversity in wireless systems with Rayleigh fading. *IEEE Transactions on Vehicular Technology* 47(1), 119–123 (1998)
8. Thomas Hou, Y., Shi, Y., Hanif Sherali, D., Jeffrey Wieselthier, E.: Multicast Communications in Ad Hoc Networks Using Directional Antennas: A Lifetime-Centric Approach. *IEEE Transactions on Vehicular Technology* 56(3) (2007)
9. Chandrakasan, A.: Design considerations for distributed micro-sensor systems. In: *Custom Integrated Circuits Conference (CICC)*, pp. 279–286 (1999)
10. Shih, E.: Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. In: *Proc. of the Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking*, pp. 272–286 (2001)
11. Bian, F., Goel, A., Raghavendra, C.S., Li, X.: Energy-efficient broadcasting in wireless ad hoc networks: lower bounds and algorithms. *JINet* 3, 149–166 (2002)
12. Wei, H., Sasaki, H., Kubokawa, J.: A decoupled solution of hydro-thermal optimal power flow problem by means of interior point method and network programming. *IEEE Transactions on Power Systems* 13(2), 286–293 (1998)
13. <http://www.mathworks.com/help/techdoc/ref/fminsearch.html#fminsearch>
14. Namin, F., Nosratinia, A.: Pragmatic Lifetime maximization of cooperative Sensor Networks via a decomposition approach. In: *Acoustics, Speech and Signal Processing- Las Vegas, Nevada, U.S.A., vol. (4)*, p. 3017 (2008)
15. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press (2004)
16. Goudarzi, H., Pakravan, M.R.: Equal Power Allocation scheme for cooperative diversity. In: *4th IEEE/IFIP International Conference, Tashkent-uzbekistan*, pp. 1–5 (2008)

Evolutionary Spectrum for Random Field and Missing Observations

Rachid Sabre

AgroSup/Laboratoire Le2i Université de Bourgogne, 26,
bd Docteur Petitjean 21000 Dijon
r.sabre@agrosupdijon.fr

Abstract. There are innumerable situations where the data observed from a non-stationary random field are collected with missing values. In this work a consistent estimate of the evolutionary spectral density is given where some observations are randomly missing.

Keywords: spectral density, non-stationary processes, periodogram, smoothing estimate, oscillatory process.

1 Introduction

Spectral analysis for stationary processes has been extensively studied in recent years. However, in many applications the signals must be modeled as non-stationary processes. This has motivated several authors to study non-stationary processes assuming that they are locally stationary. Priestley ([14], [15]) established the theory of the evolutionary spectrum generalizing spectral analysis for stationary processes. The evolutionary spectrum is time-dependent and describe the local power-frequency distribution at each instant of time. Other studies based on the Wold-Cramér decomposition have contributed to the development of the evolutionary spectrum [10], [17], [16],[18]. The applications of the evolutionary spectrum cover various scientific fields: signal and image processing [3], [1], seismic [20], oceanography, music [4]. The estimation of the evolutionary spectral density is studied in [15], [10], [8], [19], [9].

On the other hand, Jones [6] is the first to consider the missing data problems in spectral analysis. More precisely he studied the case where a block of observations is periodically unobtainable. In parallel, the theory of amplitude-modulated stationary processes was developed by Parzen [12], he applied this theory to solve periodic missing data problems. Bloomfield [2] has considered stationary processes with randomly missing data. He gives an asymptotically unbiased estimator of the spectral density and shows under suitable conditions that its variance converges to zero. We cite in this paper a few works that have contributed to find solutions to problems of missing observations: [21], [13],[7].

The aim of the present paper is to consider the problem of the randomly missing data for the class of non-stationary oscillatory random fields. Using the same techniques introduced by Bloomfield [2] for stationary processes, we give a

consistent estimate of the evolutionary spectral density. The paper is organized as follows. In section 2, we give some notations, assumptions and the amplitude modulating function Y_{t_1, t_2} . In section 3, we construct a periodogram and we show that it is an asymptotically unbiased estimator. Since, we smooth the periodogram in the neighborhood of the time-instant t via a weight function and we show that it is a consistent estimate of the (weighted) average value of $h_{t_1, t_2}(\omega_{01}, \omega_{02})$ in the neighborhood of the time-instant (t_1, t_2) . Section 4 is reserved to prove the theorems. In section 5, we study numerical results and simulation. Concluding comments are given in section 6.

2 The Amplitude Modulating Function, Y_{t_1, t_2}

As in Priestley ([14], [15]), we consider a non-stationary centred oscillatory random field X_{t_1, t_2} , $t_1, t_2 \in \mathbb{Z}$ i.e.

$$X_{t_1, t_2} = \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} e^{i(t_1\omega_1, t_2\omega_2)} A_{t_1, t_2}(\omega_1, \omega_2) dZ_1(\omega_1, \omega_2); \quad t_1, t_2 \in \mathbb{Z}, \quad (1)$$

where the function $A_{t_1, t_2}(\omega_1, \omega_2)$ is given by

$$A_{t_1, t_2}(\omega_1, \omega_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i(\theta_1 t_1, \theta_2 t_2)} dF_{\omega_1, \omega_2}(\theta_1, \theta_2),$$

$t_1, t_2 \in \mathbb{Z}$ and $\omega_1, \omega_2 \in [-\pi, \pi]$,

where F_{ω_1, ω_2} is a measure satisfying: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |dF_{\omega_1, \omega_2}(\theta_1, \theta_2)| = 1$ and Z_1 is a processus with orthogonal increments defined on the interval $[-\pi, +\pi]^2$ and $E |dZ_1(\omega_1, \omega_2)|^2 = d\mu_1(\omega_1, \omega_2)$ where μ_1 is a positive measure. The evolutionary spectral measure is defined by Priestley ([14], [15]) at each (t_1, t_2) by

$$dH_{t_1, t_2}(\omega_1, \omega_2) = |A_{t_1, t_2}(\omega_1, \omega_2)|^2 d\mu(\omega). \quad (2)$$

Our choice of oscillatory random field is motivated by the fact that it has a physical interpretation and the variance of the process is interpreted as a measure of the total power of the process at time t , because $Var(X(t_1, t_2)) = \int_{-\infty}^{+\infty} dH_{t_1, t_2}(\omega_1, \omega_2)$. The evolutionary spectral density of the process $\{X(t_1, t_2)\}$ is given by $h_{t_1, t_2}(\omega_1, \omega_2)$ and defined as follows:

$$h_{t_1, t_2}(\omega_1, \omega_2) = \frac{dH_{t_1, t_2}(\omega_1, \omega_2)}{d\omega_1 d\omega_2}, \quad \omega_1, \omega_2 \in \mathbb{R}. \quad (3)$$

Assume that the process $\{X_{t_1, t_2}\}$ is observed with randomly missing observations. As Bloomfield [2], we consider the process L_{t_1, t_2} defined as the product of the process $\{X_{t_1, t_2}\}$ and an other process $\{Y_{t_1, t_2}\}$ defined as follows:

$$L_{t_1, t_2} = X_{t_1, t_2} Y_{t_1, t_2} \quad \text{where} \quad Y_{t_1, t_2} = \begin{cases} 1 & \text{if } X_{t_1, t_2} \text{ is observed} \\ 0 & \text{otherwise.} \end{cases}$$

The process L_{t_1, t_2} is equal to a modified version of the original process $\{X_{t_1, t_2}\}$ by replacing the missing observations by $E(X_{t_1, t_2})$ their mean value, which is zero because $\{X_{t_1, t_2}\}$ is centred.

To simplify, we suppose, as Bloomfield [2], that $\{Y_{t_1, t_2}\}$ is stationary, independent of X_{t_1, t_2} and satisfying:

$$P \{Y_{t_1, t_2} = 1\} = p > \frac{1}{2},$$

$$P \{Y_{t_1, t_2} = 0\} = 1 - p,$$

The assumption of stationarity means that the statistical properties of the process Y does not depend on time. This case is often encountered in practice especially when collecting data provided by devices partially defective. Set

$$\xi_{r_1, r_2} = \frac{1}{p} E \{Y_{t_1, t_2} Y_{t_1+r_1, t_2+r_2}\} \tag{4}$$

$$\nu_{q, r, s} = \frac{1}{p^2} E \{Y_{t_1, t_2} Y_{t_1+q_1, t_2+q_2} Y_{t_1+r_1, t_2+r_2} Y_{t_1+s_1, t_2+s_2}\}; \quad q_i, r_i, s_i \in \mathbb{Z} \tag{5}$$

Since $E(Y_{t_1, t_2}) = p$, we obtain

$$Cov \{Y_{t_1, t_2}, Y_{t_1+r_1, t_2+r_2}\} = E \{Y_{t_1, t_2} Y_{t_1+r_1, t_2+r_2}\} - E \{Y_{t_1, t_2}\} E \{Y_{t_1+r_1, t_2+r_2}\}$$

$$= p\xi_{r_1, r_2} - p^2 = p(\xi_{r_1, r_2} - p).$$

This implies that ξ_{r_1, r_2} is symmetric in (r_1, r_2) . In the remainder of this paper, we assume the following hypotheses:

$\mathcal{H}_1)$ There exists a real number $V > 0$ such that

$$\sum_{q=-\infty}^{\infty} |\nu_{r, q, q+s} - \xi_{r_1, r_2} \xi_{s_1, s_2}| \leq V (|(r_1, r_2)| + |(s_1, s_2)| + 1) < \infty, \tag{6}$$

$\mathcal{H}_2)$ $\xi > 0$ and $p\xi_{r_1, r_2} \geq 2p - 1 > 0 \quad r_1, r_2 \in \mathbb{Z}$ (7)

Remark 1. – The first hypothesis $\mathcal{H}_1)$ means that the sum,

$$\sum_{q=-\infty}^{\infty} Cov(Y_{t_1, t_2} Y_{t_1+r_1, t_2+r_2}, Y_{t_1+q_1, t_2+q_2} Y_{t_1+q_1+s_1, t_2+q_2+s_2})$$

is bounded by a function proportional to $p^2 (|(r_1, r_2)| + |(s_1, s_2)| + 1)$.

– The second hypothesis $\mathcal{H}_2)$ implies for each (t_1, t_2) , the probability that X_{t_1, t_2} is observed (not missing) is greater than $\frac{1}{2}$.

3 Estimation of the Evolutionary Spectral Density

We begin by given some definitions introduced by Priestley ([14], [15]). Let \mathcal{F} the family of oscillatory functions $\{A_{t_1, t_2}(\omega_1, \omega_2) e^{i(t_1 \omega_1 + t_2 \omega_2)}\}$. For each family \mathcal{F} , we

define the function $\mathcal{B}_{\mathcal{F}}(\omega_1, \omega_2) = \int ||(\theta_1, \theta_2)|| |dF_{\omega_1, \omega_2}(\theta_1, \theta_2)|$. Let \mathcal{C} in the class of families \mathcal{F} such that $\mathcal{B}_{\mathcal{F}}(\omega_1, \omega_2)$ is bounded for all (ω_1, ω_2) . For each family \mathcal{F} we define the following constant $\mathcal{B}_{\mathcal{F}}$ termed the characteristic width of \mathcal{F} :

$$\mathcal{B}_{\mathcal{F}} = \left[\sup_{(\omega_1, \omega_2)} \mathcal{B}_{\mathcal{F}}(\omega_1, \omega_2) \right]^{-1}$$

The characteristic width of the process X_{t_1, t_2} is defined by $\mathcal{B}_X = \sup_{\mathcal{F} \in \mathcal{C}} \mathcal{B}_{\mathcal{F}}$. For more details about definitions see Priestley ([14], [15]).

In this section, we propose a periodogram constructed as follows:

$$I_{t,T}(\omega_{01}, \omega_{02}) = \left| \sum_{u=t-T}^{t+T} g_u \frac{L_{t_1-u_1, t_2-u_2}}{S} e^{-i(\omega_{01}(t_1-u_1) + (t_2-u_2)\omega_{02})} \right|^2, \tag{8}$$

where $S = \left(2\pi \sum_{u_1, u_2} p \xi_{0,0} |g_{u_1, u_2}|^2 \right)^{\frac{1}{2}}$, and $\{g_{u_1, u_2}\}$, is a filter satisfying the following conditions:

- C_1 : $g_{u_1, u_2} \geq 0$; $g_{u_1, u_2} = g_{-u_1, -u_2}$,
- C_2 : $\sum_{u_1, u_2, v_1, v_2} p \xi_{u_1-v_1, u_2-v_2} g_{u_1, u_2} g_{v_1, v_2}^* < \infty$, where ξ is defined in (4)
- C_3 :) g_{u_1, u_2} has finite ‘‘width’’, defined by:

$$\mathcal{B}_g \simeq \sum_{u_1, u_2, v_1, v_2 = -\infty}^{+\infty} p |\xi_{u_1-v_1, u_2-v_2}| ||(u_1, u_2)|| |g_{u_1, u_2}| |g_{v_1, v_2}^*| < \infty, \tag{9}$$

- C_4 : $\mathcal{B}_g \ll \mathcal{B}_{\mathcal{F}}$,
- C_5 : For any real numbers k_1, k_2 , we have

$$\left| \int_{-\infty}^{\infty} \Gamma(s, s) h_{t_1, t_2}(s_1 + k_1, s_2 + k_2) ds_1 ds_2 - h_{t_1, t_2}(k_1, k_2) \int_{-\infty}^{\infty} \Gamma(s, s) ds_1 ds_2 \right| < \frac{\mathcal{B}_g}{\mathcal{B}_{\mathcal{F}}},$$

where the function Γ is defined by:

$$\Gamma(s, s') = \sum_{u_1, u_2, v_1, v_2} p \xi_{u_1-v_1, u_2-v_2} g_{u_1, u_2} g_{v_1, v_2}^* e^{-i(u_1 s_1 - v_1 s'_1 + u_2 s_2 - v_2 s'_2)}.$$

The function Γ_1 is highly concentrated relative to the function h_{t_1, t_2} .

When this condition is satisfied, we say as Priestley ([15] page 829) that the function Γ_1 is δ -function with respect to h_{t_1, t_2} in order $\left(\frac{\mathcal{B}_g}{\mathcal{B}_{\mathcal{F}}} \right)$.

$$C_6$$
 : $g_{u_1, u_2} = O(e^{-||(\omega_1, \omega_2)||})$

The following theorem shows that the periodogram $I_{t,T}(\omega_{01}, \omega_{02})$ is an asymptotically unbiased estimator of the evolutionary spectral density $h_{t_1, t_2}(\omega_{01}, \omega_{02})$.

Theorem 1. *Let t_1, t_2 be an integer numbers and ω_{01}, ω_{02} are real numbers, suppose that $\frac{\mathcal{B}_g}{\mathcal{B}_X} < \epsilon$, then*

$$E [I_{t,T}(\omega_{01}, \omega_{02})] = h_{t_1, t_2}(\omega_{01}, \omega_{02}) + O(\epsilon).$$

To prove the theorem 1, we have need the two following lemmas

Lemma 1. For any $t_1, t_2, t'_1, t'_2, \lambda_1, \lambda_2$ real numbers, we have

$$\left| \int e^{i(t_1 s_1 + t_2 s_2)} e^{-i(t'_1 s'_1 + t'_2 s'_2)} \Gamma(s + k, s' + k') dF_{\lambda_1, \lambda_2}(s_1, s_2) dF_{\lambda_1, \lambda_2}(s'_1, s'_2) - \Gamma(k, k') \int e^{i(t_1 s_1 + t_2 s_2)} e^{-i(t'_1 s'_1 + t'_2 s'_2)} dF_{\lambda_1, \lambda_2}(s_1, s_2) dF_{\lambda_1, \lambda_2}(s'_1, s'_2) \right| < 2 \frac{\mathcal{B}_g}{\mathcal{B}_F}$$

Lemma 2. Let $\theta_1, \theta_2, \lambda_1, \lambda_2, t_1, t_2$ and t'_1, t'_2 be real numbers, we have

$$\left| A_{t_1, t_2}(\lambda_1, \lambda_2) A_{t'_1, t'_2}^*(\lambda_1, \lambda_2) \right| \left| \Gamma_{t_1, t_2, t'_1, t'_2, \lambda_1, \lambda_2}(\theta_1, \theta_2) - \Gamma(\theta, \theta) \right| \leq 2 \frac{\mathcal{B}_g}{\mathcal{B}_F}, \text{ where}$$

$$\Gamma_{t_1, t_2, s_1, s_2, \lambda_1, \lambda_2}(\theta_1, \theta_2) = \sum_{u_1, u_2, v_1, v_2} p \xi_{u_1 - v_1, u_2 - v_2} g_{u_1, u_2} g_{v_1, v_2}^* \beta(u, v, \theta) \quad (10)$$

where

$$\beta(u, v, \theta) = \frac{A_{t_1 - u_1, t_2 - u_2}(\lambda_1, \lambda_2) A_{s_1 - v_1, s_2 - v_2}^*(\lambda_1, \lambda_2)}{A_{t_1, t_2}(\lambda_1, \lambda_2) A_{s_1, s_2}^*(\lambda_1, \lambda_2)} e^{-i((u_1 - v_1)\theta_1 + (u_2 - v_2)\theta_2)}.$$

In order to obtain a consistent estimate of $\{h_{t_1, t_2}(\omega_{01}, \omega_{02})\}$, we smooth the periodogram in the neighborhood of the time-instant (t_1, t_2) via a weight function:

$$\widehat{h}_{t_1, t_2}(\omega_{01}, \omega_{02}) = \sum_{v_1, v_2 \in M} w_{T'_1, T'_2, v_1, v_2} \widehat{I}_{t_1 - v_1, t_2 - v_2}(\omega_{01}, \omega_{02}). \quad (11)$$

where $w_{T'_1, T'_2, v_1, v_2}$ is a weight-function depending on the parameters T'_1, T'_2 and satisfying

- a) $w_{T'_1, T'_2, v_1, v_2} \geq 0$, for all v_1, v_2, T'_1, T'_2
- b) $w_{T'_1, T'_2, v_1, v_2} = 0, v_1, v_2 \notin M$, where M is a set of integers surrounding zero.
- c) $w_{T'_1, T'_2, v_1, v_2} = w_{T'_1, T'_2, -v_1, -v_2}$,
- d) $\sum_{v_1, v_2 \in M} w_{T'_1, T'_2, v_1, v_2} = 1$,
- e) $\sum_{v_1, v_2 \in M} w_{T'_1, T'_2, v_1, v_2}^2 < \infty$.
- f) We assume that there exists a constant C such that

$$\lim_{T'_1, T'_2 \rightarrow \infty} T'_1, T'_2 \sum_{u_1, u_2 \in M} |W_{T'_1, T'_2, u_1, u_2}|^2 = C, \text{ where}$$

$$W_{T'_1, T'_2, u_1, u_2} = \sum_{v_1, v_2 \in M} e^{-i(u_1 v_1 + u_2 v_2)} w_{T'_1, T'_2, v_1, v_2}.$$

The following theorem show that the estimator $\widehat{h}_{t_1, t_2}(\omega_{01}, \omega_{02})$ is an asymptotically unbiased of the (weighted) average value of $h_{t_1, t_2}(\omega_{01}, \omega_{02})$ in the neighborhood of (t_1, t_2) .

Theorem 2. *Let $-\pi \leq \omega_{01}, \omega_{02} \leq \pi$, suppose that $\frac{\mathcal{B}_a}{\mathcal{B}_X} < \epsilon$, then*

$$E \left[\widehat{h}_{t_1, t_2}(\omega_{01}, \omega_{02}) \right] = \bar{h}_{t_1, t_2}(\omega_{01}, \omega_{02}) + O(\epsilon)$$

where $\bar{h}_{t_1, t_2}(\omega_{01}, \omega_{02}) = \sum_{v_1, v_2 \in M} w_{T'_1, T'_2, v_1, v_2} h_{t_1 - v_1, t_2 - v_2}(\omega_{01}, \omega_{02})$

To show that the variance converges to zero, as Priestley ([14]) and Melard [10], we assume that the process L_{t_1, t_2} is Gaussian.

Theorem 3. *Let $-\pi \leq \omega_1, \omega_2 \leq \pi$ and suppose that the process L_{t_1, t_2} is Gaussian, then we have*

$$Var \left[\widehat{h}_{t_1, t_2}(\omega_{01}, \omega_{02}) \right] = O\left(\frac{1}{T'_1, T'_2}\right).$$

4 Numerical Studies

As in Bloomfield [2], we suppose that our process $\{X_{t,s}\}_{t,s \in \mathbb{Z}}$ is observed at the successively instants $(t_1, s_1), (t_2, s_2), \dots, (t_n, s_n)$ where $\tau_i = |t_{i+1} - t_i|$ $\tau'_i = |s_{i+1} - s_i|$ are independent random variables, each with the probability distribution $\{f_{r_1, r_2} = P[(\tau, \tau') = (r_1, r_2)]\}$, and finite mean p^{-1} . As in Feller ([5], pp 282-283), we define a process $\{Y'_{t,s}\}$ which coincides with $\{Y_{t,s}\}$ except at origin $Y'_{0,0} = 1$. the event " $Y' = 1$ " is termed persistent and recurrent event. Using (6) we obtain

$$\begin{aligned} \xi_{r_1, r_2} &= p^{-1} E \{Y_{t_1, t_2} Y_{t_1+r_1, t_2+r_2}\} = P \{Y_{t_1+r_1, t_2+r_2} = 1 / Y_{t_1, t_2} = 1\} \\ &= P \{Y'_{r_1, r_2} = 1\} \end{aligned}$$

Feller ([5], pp 282-283) has shown that

$$\xi_{r_1, r_2} = \sum_{s=1}^{r_1, r_2} f_{s_1, s_2} \xi_{r_1 - s_1, r_2 - s_2}, \quad r_i, s_i = 1, 2, \dots$$

The processus L_{t_1, t_2} was obtained from $X_{t,s}$ by omitting certain observations with a renewal-type mechanism defined above with $f_{1,1} = \frac{8}{9}, f_{2,2} = \frac{1}{9}, f_{r_1, r_2} = 0$ otherwise.

The simulation of the process X :

Using the same method in [11] for the simulation of Markov Gauss random field, we simulate the Gaussian random field $Y = \{Y(n_1, n_2)\}_{n_1, n_2 \in \mathbb{Z}}$ such that $R_Y(n_1, n_2)$ the covariance function is given by $R_Y(n_1, n_2) = e^{-\sqrt{(n_1+n_2)}}$, and its spectral density is $f_Y(\lambda_1, \lambda_2) = \frac{1}{\pi(1+\lambda_1^2+\lambda_2^2)}$.

the random field $X_{t,s}, t, s \in \mathbb{Z}$ is given by the following model

$$X_{t,s} = c_{t,s} Y_{t,s}, \quad t, s \in \mathbb{Z}..$$

where $c_{t,s} = e^{-\frac{(t+s-500)^2}{2 \cdot 200^2}}$ $A_{t,s}(\omega_1, \omega_2) = c_{t,s}$ is independent of ω . With respect to the family $\mathcal{F} = \{c_{t,s} e^{i(\omega_1 t + \omega_2 s)}\}$, $X(t, s)$ has evolutionary spectral density function $h_{t_1, t_2}(\omega_1, \omega_2) = c_{t_1, t_2}^2 f_Y(\omega_1, \omega_2)$.

The curve of the estimator with 5000 observations (Fig. 2) and that of the spectral density (Fig. 1) are very similar. So the estimator is quite satisfactory. If we take more observations (around 10000), the estimator becomes more smoother and the curve approaches the density much.

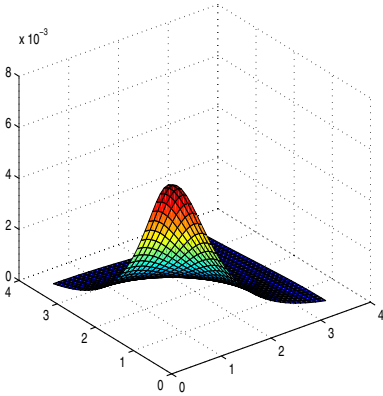


Fig. 1. Density $h_{100,12}$

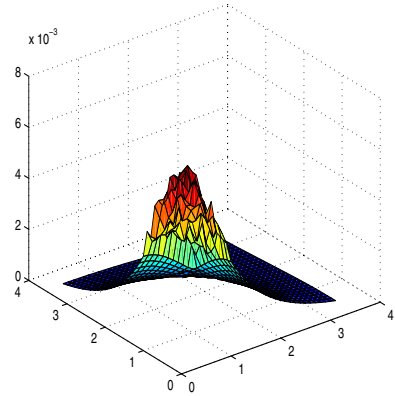


Fig.2. Estimator $\hat{h}_{100,12}$

5 Conclusion

We have proposed in this paper some results about the estimation of the evolutionary spectral density for non-stationary random fields where the data observed are collected with missing values. The approach is based on the technique used by Bloomfield [2] for stationary processes combining estimates of evolutionary spectrum introduced by Priestley ([14]). This work could be applied to several cases when the process is non-stationary as for example for:

- the segmentation of a sequence of images of a dynamic scene, detecting weeds in a farm field.
- the study of geostatistical mapping of certain chemical factors in agricultural soil.

This work could be supplemented by the study of optimal smoothing parameters using cross validation methods that have proven in the field. It will also be extended to non-Gaussian process by assuming some hypotheses as for example the cumulants are finite.

Acknowledgments. I would like to thank the anonymous referees for their interest in this paper and their valuable comments and suggestions.

References

1. Al-Shoshan, A.I., Chaparro, L.F.: Identification of non-minimum phase systems using evolutionary spectral theory. *Signal Processing* 55(1), 79–92 (1996)
2. Bloomfield, P.: Spectral analysis with randomly missing observations. *J. Roy. Statistical Society* 32(3), 369–380 (1970)
3. Cabrera, S.D., Flores, B.C., Thomas, G., Vega-Pineda, J.: Evolutionary spectral estimation based on adaptive use of weighted norms. In: Luk, F.T. (ed.) *Advanced Signal Processing Algorithms, Architectures, and Implementations IV*. SPIE, vol. 2027, pp. 168–179 (1993)
4. Caetano, M., Rodet, X.: Evolutionary spectral envelope morphing by spectral shape descriptors. In: *Proceeding of International Computer Music Conferences (ICMC)* (August 2009)
5. Feller, W.: *An Introduction to Probability Theory and its Applications*, 2nd edn., vol. I. Wiley, New York (1957)
6. Jones, R.H.: Spectrum estimation with missing observations. *Annals of the Institute of Statistical Mathematics* 23(1), 387–398 (1971)
7. Hung, J.-C.: A genetic algorithm approach to the spectral estimation of time series with noise and missed observations. *Information Sciences* 178(24), 4632–4643 (2008)
8. Kayhan, A.S., El-Jaroudi, A., Chaparro, L.F.: Data-Adaptive Evolutionary Spectral Estimation. *IEEE Transactions on Signal Processing* 43(1), 204–213 (1995)
9. Kayhan, A.S., Amine, M.G.: Spatial evolutionary spectrum for DOA estimation and blind signal separation. *IEEE Transactions on Signal Processing* 48(3), 791–798 (2000)
10. Mélard, G., de Schutter, A.H.: Contributions to evolutionary spectral theory. *Journal of Time Series Analysis* 10(1), 41–63 (1988)
11. Messaci, F.: Estimation de la densité spectrale d'un processus en temps continu par échantillonnage poissonnien, Univ de Rouen (Thèse de 3^{ème} cycle) (1986)
12. Parzen, E.: On spectral analysis with missing observations and amplitude modulation. *Sankhya, Series A* 25, 383–392 (1963)
13. Broersen, P.M.T.: Automatic spectral analysis with missing data. *Digital Signal Processing* 16(6), 754–766 (2006)
14. Priestley, M.B.: Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser., B* 27, 204–237 (1965)
15. Priestley, M.B.: *Spectral analysis and time series*. Probability and Mathematical Statistics. Academic Press (1981)
16. Rachdi, M., Sabre, R.: Mixed-spectra analysis for stationary random fields. *Statistical Methods and Applications* 18, 333–358 (2009)
17. Sabre, R.: Spectral density estimation for stationary stable random fields. *Journal Applications Mathematicae* 23(2), 107–133 (1995)
18. Sabre, R.: Discrete estimation of spectral density for symmetric stable process. *Statistica* 2, 1–26 (2000)
19. Shah, S.I., Chaparro, L.F., El-Jaroudi, A., Furman, J.M.: Evolutionary Maximum Entropy Spectral Estimation and HeartRate Variability Analysis. *Multidimensional Systems and Signal Processing* 9(4), 453–458 (1998)
20. Tezcan, J.: Evolutionary Power Spectrum Estimation Using Harmonic Wavelets. *Seismic Design and Analysis of Structures*, 37–41 (2003)
21. Wang, Y., Stoica, P., Li, J., Marzetta, T.L.: Nonparametric spectral analysis with missing data via the EM algorithm. *Digital Signal Processing* 15(2) (2005)

Iris-Biometric Fuzzy Commitment Schemes under Signal Degradation*

C. Rathgeb and A. Uhl

Multimedia Signal Processing and Security Lab.
Department of Computer Sciences
University of Salzburg, A-5020 Salzburg, Austria
{crathgeb,uhl}@cosy.sbg.ac.at

Abstract. Low intra-class variability at high inter-class variability is considered a fundamental premise of biometric template protection, i.e. it is believed that biometric traits need to be captured under favorable conditions in order to provide practical recognition rates. In this work the impact of blur and noise to fuzzy commitment schemes is investigated and is compared to the impact observed on the accuracy of the underlying recognition scheme. Iris textures are successively blurred and noised in order to measure the robustness of iris-biometric fuzzy commitment schemes.

1 Introduction

Biometric template protection schemes are designed to meet major requirements of biometric information protection (ISO/IEC FCD 24745), i.e. irreversibility (infeasibility of reconstructing original biometric templates from the stored reference data) and unlinkability (infeasibility of cross-matching different versions of protected templates). In addition, template protection schemes, which are commonly categorized as biometric cryptosystems and cancelable biometrics, are desired to maintain recognition accuracy [1]. Due to the sensitivity of template protection schemes it is generally conceded that deployments of biometric cryptosystems as well as cancelable biometrics require a constraint acquisition of biometric traits, in order to minimize any sort of signal degradation.

Biometric fuzzy commitment schemes (FCSs) [2], biometric cryptosystems which represent instances of biometric key-binding, have been proposed for several modalities (e.g. fingerprints, iris) achieving practical key retrieval rates at sufficient key sizes. While it is generally considered that template protection schemes, such as the FCS, are restricted to be operated under constraint environment detailed performance analysis in the presence of signal degradation have remained elusive. The contribution of this work is the investigation of the impact of signal degradation on the performance of FCSs. Two types of conditions, blur and noise, applied in the order illustrated in Fig. 1, are investigated:

* This work has been supported by the Austrian Science Fund, project no. L554-N15.

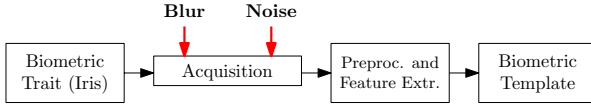


Fig. 1. Supposed blur and noise occurrence within a biometric recognition system

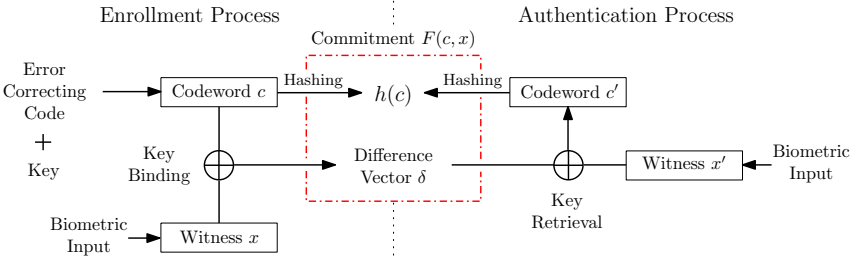


Fig. 2. Basic operation mode of the Fuzzy Commitment Scheme

- **Blur:** focusing on image acquisition out of focus blur represents a frequent distortion.
- **Noise:** noise represents an undesirable but inevitable product of any electronic device.

Experimental studies are carried out on iris-biometric data employing different feature extraction algorithms to construct FCSs. Various combinations of different intensities of blur and noise are applied to simulate signal degradation. It is demonstrated that, opposed to current opinions, signal degradation, within a restricted extent, does not necessarily effect the key retrieval performance of a template protection scheme, even if this is the case for original recognition algorithms.

This paper is organized as follows: in Section 2 related work regarding biometric cryptosystems and FCSs is reviewed. Subsequently, a comprehensive case study on iris-biometric FCS is presented in Section 3. Finally, a conclusion is given in Section 4.

2 Fuzzy Commitment Schemes

In past year numerous template protection schemes have been proposed [1]. In 1999, Juels and Wattenberg [2] proposed the FCS, a bit commitment scheme resilient to noise. A FCS is formally defined as a function F , applied to commit a codeword $c \in C$ with a witness $x \in \{0, 1\}^n$ where C is a set of error correcting codewords of length n . The witness x represents a binary biometric feature vector which can be uniquely expressed in terms of the codeword c along with an offset $\delta \in \{0, 1\}^n$, where $\delta = x - c$. Given a biometric feature vector x expressed in this

Table 1. Experimental results of proposed Fuzzy Commitment Scheme

Author(s)	Modality	FRR/ FAR	Key Bits	Remarks
Hao <i>et al.</i> [3]	Iris	0.47/ 0	140	ideal images
Bringer <i>et al.</i> [4]		5.62/ 0	42	short key
Rathgeb and Uhl [5]		4.64/ 0	128	–
Teoh <i>et al.</i> [6]	Fingerprint	0.9/ 0	296	user-specific tokens
Nandakumar [7]		12.6/ 0	327	–
Van der Veen <i>et al.</i> [8]	Face	3.5/ 0.11	58	>1 enroll. sam.
Ao and Li [9]		7.99/ 0.11	>4000	user-specific tokens

way, c is concealed applying a conventional hash function (e.g. SHA-3), while leaving δ as it is. The stored helper data is defined as,

$$F(c, x) = (h(x), x - c). \quad (1)$$

In order to achieve resilience to small corruptions in x , any x' sufficiently “close” to x according to an appropriate metric (e.g. Hamming distance), should be able to reconstruct c using the difference vector δ to translate x' in the direction of x . In case $\|x - x'\| \leq t$, where t is a defined threshold lower bounded by the according error correction capacity, x' yields a successful decommitment of $F(c, x)$ for any c . Otherwise, $h(c) \neq h(c')$ for the decoded codeword c' and a failure message is returned. In Fig. 2 the basic operation mode of the FCS is illustrated.

Key approaches to FCSs with respect to applied biometric modalities, performance rates in terms of false rejection rate (FRR) and false acceptance rate (FAR), extracted key sizes, and applied data sets are summarized in Table 1. The FCS was applied to iris-codes in [3]. In this scheme 2048-bit iris-codes are applied to bind and retrieve 140-bit cryptographic keys prepared with Hadamard and Reed-Solomon error correction codes. Hadamard codes are applied to eliminate bit errors originating from the natural biometric variance and Reed-Solomon codes are applied to correct burst errors resulting from distortions. In order to provide an error correction decoding in an iris-based FCS, which gets close to a theoretical bound, two-dimensional iterative min-sum decoding is introduced in [4]. A matrix formed by two different binary Reed-Muller codes enables a more efficient decoding. Different techniques to improve the accuracy of iris-based FCSs have been proposed in [5,10]. In [7] a binary fixed-length minutiae representation obtained by quantizing the Fourier phase spectrum of a minutiae set is applied in a FCS where alignment is achieved through focal points of high curvature regions. In [6] a randomized dynamic quantization transformation is applied to binarize fingerprint features extracted from a multichannel Gabor filter. Subsequently, Reed-Solomon codes are applied to construct the FCS incorporating a non-invertible projection based on a user-specific token. A similar FCS based on a face features is presented in [9]. A FCS based on face biometrics is presented in [8] in which real-valued face features are binarized by simple thresholding followed by a reliable bit selection to detect most discriminative

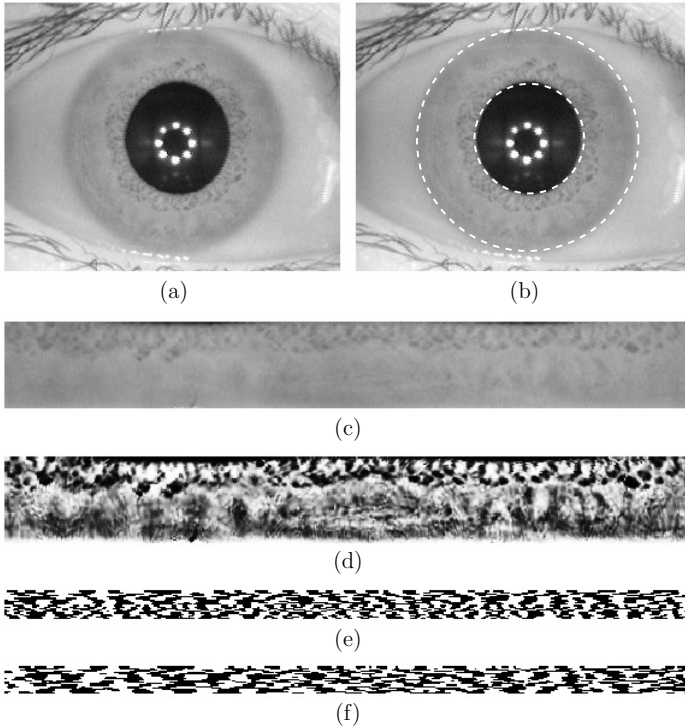


Fig. 3. Preprocessing and feature extraction: (a) eye (b) detection of pupil and iris (c) unwrapped and (d) preprocessed iris texture, iris-code of (e) Masek and (f) Ma *et al.*

features. It has been found that FCSs (template protection schemes in general) reveal worse performance on non-ideal data sets (e.g. in [4]), however, this is the case for underlying recognition algorithms, too. To our knowledge, so far, no detailed investigations about the impact of signal degradation based on a certain ground truth have been proposed.

3 A Case Study on Iris-Biometric FCSs

3.1 Experimental Setup

Experiments are carried out using the CASIA-v3-Interval iris database¹. In experiments only left-eye images (1332 instances) are evaluated. At preprocessing the iris of a given sample image is detected, un-wrapped to a rectangular texture of 512×64 pixel, and lighting across the texture is normalized as shown in Fig. 3 (a)-(d).

In the feature extraction stage custom implementations of two different iris recognition algorithms are employed. The first one was proposed by Ma *et al.* [11]. Within this algorithm a dyadic wavelet transform is performed based on

¹ The Center of Biometrics and Security Research, <http://www.idealtest.org>

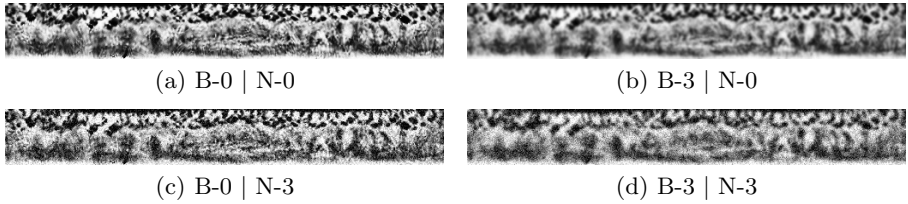


Fig. 4. Signal degradation: (a)-(d) different intensities of blur and noise applied to a sample iris texture.

which two fixed subbands are selected. Local minima and maxima above a adequate threshold are located an encoded extracting 10240 bit. The second feature extraction method follows an implementation by Masek² in which filters obtained from a one-dimensional Log-Gabor function are utilized to generate iris-codes of 10240 bit. Sample iris-codes of both algorithms are shown in Fig. 3 (e)-(f).

3.2 Iris-Biometric FCSs

The applied fuzzy commitment scheme follows the approach in [12]. For the applied algorithm of Ma et al. and the Log-Gabor feature extraction we found that the application of Hadamard codewords of 128-bit and a Reed-Solomon code $RS(16, 80)$ reveals the best experimental results for the binding of 128-bit cryptographic keys. At key-binding, a $16 \cdot 8 = 128$ bit cryptographic key R is first prepared with a $RS(16, 80)$ Reed-Solomon code. The Reed-Solomon error correction code operates on block level and is capable of correcting $(80 - 16)/2 = 32$ block errors. Then the 80 8-bit blocks are Hadamard encoded. In a Hadamard code codewords of length n are mapped to codewords of length 2^{n-1} in which up to 25% of bit errors can be corrected. Hence, 80 8-bit codewords are mapped to 80 128-bit codewords resulting in a 10240-bit bitstream which is bound with the iris-code by XORing both. Additionally, a hash of the original key $h(R)$ is stored as second part of the commitment. At authentication key retrieval is performed by XORing an extracted iris-code with the first part of the commitment. The resulting bitstream is decoded applying Hadamard decoding and Reed-Solomon decoding afterwards. The resulting key R' is then hashed and if $h(R') = h(R)$ the correct key R is released. Otherwise an error message is returned.

3.3 Signal Degradation

Signal degradation is simulated by means of blur and noise where blur is applied prior to noise (out of focus blur is caused before noise occurs). Different intensities (including absence) of blur and noise, which are summarized in Table 2, are considered, and combinations of these. In order to avoid segmentation errors blur and noise is incorporated after preprocessing (deformation of blur and noise

² L. Masek: Recognition of Human Iris Patterns for Biometric Identification, Master's thesis, University of Western Australia, 2003.

Table 2. Blur and noise conditions considered for signal degradation (different denotations of σ are defined in Eq. 2 and Eq. 3)

Blur		Noise	
Abbrev.	Description	Abbrev.	Description
B-0	no blur	N-0	no noise
B-1	$\sigma = 0.6$	N-1	$\sigma = 10$
B-2	$\sigma = 1.0$	N-2	$\sigma = 20$
B-3	$\sigma = 1.2$	N-3	$\sigma = 30$

caused by an unwrapping of the iris is ignored, however, signal degradation still decreases recognition accuracy of the applied algorithms). Examples of adding according signal degradation to a sample iris texture are shown in Fig. 4 (a)-(d). Out of focus blur represents a frequent distortion in image acquisition mainly caused by an inappropriate distance of the camera to the acquired eye (another type of blur is motion blur caused by rapid movement which is not considered in this work). We simulate the point spread function of the blur as a Gaussian

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (2)$$

which is then convoluted with the specific image. Amplifier noise is primarily caused by thermal noise. Due to signal amplification in dark (or underexposed) areas of an image, thermal noise has a high impact on these areas. Additional sources contribute to the noise in a digital image such as shot noise, quantization noise and others. These additional noise sources however, only make up a negligible part of the noise and are therefore ignored during this work.

Let P be the set of all pixels in image $I \in \mathbb{N}^2$, $\omega = (\omega_p)_{p \in P}$, be a collection of independent identically distributed real-valued random variables following a Gaussian distribution with mean m and variance σ^2 . We simulate thermal noise as additive Gaussian noise with $m = 0$, variance σ^2 for pixel p at position x, y with N being the noisy image, for an original image I as

$$N(x, y) = I(x, y) + \omega_p, \quad p \in P. \quad (3)$$

3.4 Performance Evaluation

Experimental results for both feature extraction methods and FCSs according to different intensities of blur and noise are summarized in Table 3, including average peak signal-to-noise ratios (PSNRs) caused by signal degradation and the number of corrected block errors after Hadamard decoding. Obtained performance rates for FCSs under various forms of signal degradation for the feature extraction of Ma *et al.* are plotted in Fig. 5 (a)-(d). For the recognition algorithm of Ma *et al.* and Masek in verification mode (columns “HD” in Table 3), FRRs of 2.54% and 6.59% are obtained at a FAR of 0.01% where the Hamming distance is applied as dis-similarity metric. Focusing on the feature extraction of Ma *et al.* FCSs provide a FRR of 5.90%. With respect to the feature extraction of Masek a FRR of 8.01% is obtained.

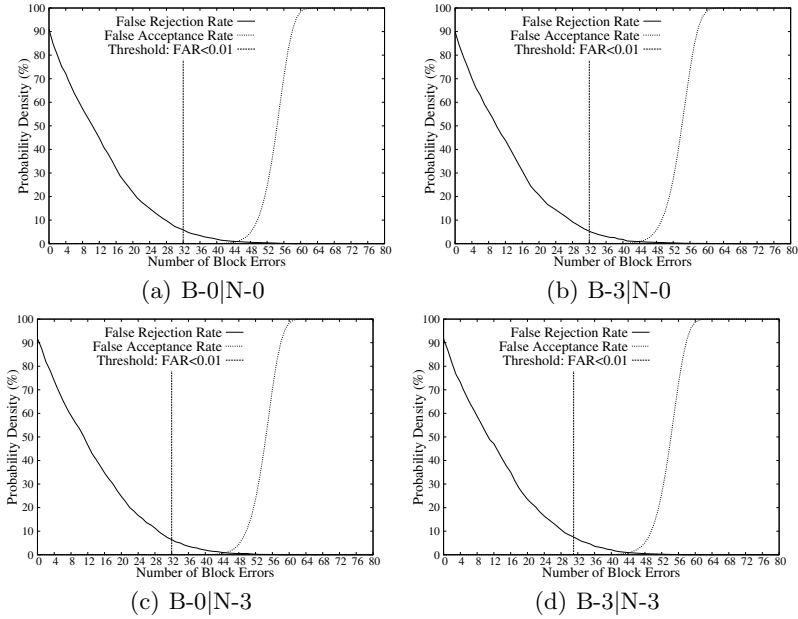


Fig. 5. Performance rates: (a)-(d) FCSs based on the algorithm of Ma *et al.* under various signal degradation conditions.

Simulating signal degradation, recognition accuracy is significantly effected for both recognition algorithms leading to FRRs above 4% and 10% at a FAR of 0.01%, respectively. In contrast, FCSs based on both feature extraction methods appear rather robust to signal degradation. Focusing on FCSs based on the algorithm of Ma *et al.* FRRs do not significantly increase, for drastic signal degradation FRRs of $\sim 6.50\%$ are obtained compared to a FRR of 5.90% without signal degradation. It is found that incorporating a certain amount of blur even improves key retrieval rates obtaining FRRs of $\sim 5.00\%$, since, on average, extracted iris-codes are even more alike (iris-codes extracted from blurred textures do not encode detailed features), i.e. slight blurring is equivalent to denoising. Focusing on the algorithm of Masek a more predominant decrease in key retrieval rates is observed, however, results are still comparable to those obtained in the absence of blur and noise. In case of drastic signal degradation FRRs of $\sim 10.00\%$ are obtained (partially outperforming the original recognition algorithm), compared to 8.01% without signal degradation. Again, in case of a slight blur performance is improved or retained. For both feature extraction methods, characteristics of FCS's FRRs and FARs remain almost unaltered in presence of signal degradation, i.e. all types of investigated FCSs appear rather robust to a certain extent of signal degradation based on blur and noise.

Table 3. Results for both FCSs under various signal degradation conditions

			Ma <i>et al.</i>			Masek		
			HD	FCS		HD	FCS	
Blur	Noise	\emptyset PSNR	FRR at FAR \leq 0.01	FRR at FAR \leq 0.01	Corr. Blocks	FRR at FAR \leq 0.01	FRR at FAR \leq 0.01	Corr. Blocks
B-0	N-0	–	2.54 %	5.90 %	32	6.59 %	8.01 %	28
B-1	N-0	26.47 dB	3.82 %	5.69 %	32	9.92 %	7.86 %	28
B-2	N-0	21.04 dB	3.75 %	4.88 %	32	10.62 %	7.59 %	26
B-3	N-0	19.62 dB	4.36 %	5.22 %	32	10.94 %	8.61 %	27
B-0	N-1	28.32 dB	4.25 %	5.94 %	32	9.51 %	8.75 %	27
B-1	N-1	24.27 dB	3.36 %	5.76 %	32	10.15 %	9.02 %	27
B-2	N-1	20.21 dB	3.84 %	5.56 %	32	10.80 %	8.95 %	27
B-3	N-1	19.07 dB	4.15 %	6.30 %	31	10.69 %	8.88 %	27
B-0	N-2	22.54 dB	4.88 %	6.51 %	32	9.92 %	9.22 %	27
B-1	N-2	20.99 dB	4.09 %	5.76 %	32	10.62 %	9.17 %	28
B-2	N-2	18.58 dB	3.86 %	5.76 %	32	9.97 %	9.02 %	27
B-3	N-2	17.70 dB	4.27 %	5.83 %	32	10.69 %	10.44 %	26
B-0	N-3	19.14 dB	4.36 %	6.44 %	32	10.33 %	9.86 %	28
B-1	N-3	18.28 dB	4.43 %	6.37 %	32	10.49 %	10.37 %	26
B-2	N-3	16.82 dB	4.56 %	6.24 %	32	10.96 %	9.43 %	27
B-3	N-3	16.19 dB	4.27 %	6.58 %	32	9.54 %	9.29 %	27

4 Conclusion

In this paper we investigate the impact of signal degradation on the performance of template protection schemes, in particular, the effect of blur and noise to FCSs based on iris. Based on different feature extraction methods FCSs are constructed and a significant amount of blur and noise is added successively to iris biometric data to simulate out of focus blur and thermal noise. It is found that, opposed to current opinions, FCSs appear rather resilient to a certain amount of signal degradation within biometric data obtaining key retrieval rates comparable to those achieved in the absence of signal degradation, even if this is not the case for underlying recognition algorithms.

References

1. Rathgeb, C., Uhl, A.: A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security* 2011 (2011)
2. Juels, A., Wattenberg, M.: A fuzzy commitment scheme. In: *Sixth ACM Conference on Computer and Communications Security*, pp. 28–36 (1999)
3. Hao, F., Anderson, R., Daugman, J.: Combining Cryptography with Biometrics Effectively. *IEEE Trans. on Computers* 55, 1081–1088 (2006)
4. Bringer, J., Chabanne, H., Cohen, G., Kindarji, B., Zémor, G.: Theoretical and practical boundaries of binary secure sketches. *IEEE Trans. on Information Forensics and Security* 3, 673–683 (2008)

5. Rathgeb, C., Uhl, A.: Adaptive fuzzy commitment scheme based on iris-code error analysis. In: Proc. of the 2nd European Workshop on Visual Information Processing (EUVIP 2010), pp. 41–44 (2010)
6. Teoh, A., Kim, J.: Secure biometric template protection in fuzzy commitment scheme. *IEICE Electron. Express* 4, 724–730 (2007)
7. Nandakumar, K.: A fingerprint cryptosystem based on minutiae phase spectrum. In: Proc. of IEEE Workshop on Information Forensics and Security, WIFS (2010)
8. Van der Veen, M., Kevenaar, T., Schrijen, G.J., Akkermans, T.H., Zuo, F.: Face biometrics with renewable templates. In: SPIE Proc. on Security, Steganography, and Watermarking of Multimedia Contents, vol. 6072, pp. 205–216 (2006)
9. Ao, M., Li, S.Z.: Near Infrared Face Based Biometric Key Binding. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 376–385. Springer, Heidelberg (2009)
10. Zhang, L., Sun, Z., Tan, T., Hu, S.: Robust Biometric Key Extraction Based on Iris Cryptosystem. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 1060–1069. Springer, Heidelberg (2009)
11. Ma, L., Tan, T., Wang, Y., Zhang, D.: Efficient Iris Recognition by Characterizing Key Local Variations. *IEEE Trans. on Image Processing* 13, 739–750 (2004)
12. Rathgeb, C., Uhl, A.: Systematic Construction of Iris-Based Fuzzy Commitment Schemes. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 940–949. Springer, Heidelberg (2009)

Sfax-Miracl Hand Database for Contactless Hand Biometrics Applications

Salma Ben Jemaa¹, Mayssa Frikha², Imen Moalla²,
Mohamed Hammami², and Hanene Ben-Abdallah¹

¹ MIRACL-FSEG, Sfax University, Road Aeroport Km 4, 3018 Sfax, Tunisia

² MIRACL-FS, Sfax University, Road Sokra Km 3 BP 802, 3018 Sfax, Tunisia
benjemaa.salma@hotmail.com, {frikha.mayssa, imen.moalla}@hotmail.fr,
mohamed.hammami@fss.rnu.tn, hanene.benabdallah@fsegs.rnu.tn

Abstract. A new branch of biometrics, hand recognition, has attracted increasing amount of attention in recent years. In this paper, we propose an approach of hand detection based skin color pixel for biometric applications using multi layer perceptron (MLP) neural network. This later has the ability to classify skin pixels belonging to people with different skin tones and captured in different lighting conditions and complex background environments. To improve the achieved results, a succession of post-processing was proposed. The choice of the database is an important step in testing a biometric process. For this, we build a database named “Sfax-Miracl hand database”. This database contains a total of 1080 images having the advantage of being captured from freely posed hands in contact free settings. Various conducted experiments on this database show promising results and demonstrate the effectiveness of the proposed approach.

Keywords: Hand biometric, Contactless hand detection, Skin color pixel, Multi layer perceptron (MLP).

1 Introduction

Automatic recognition using biometric characteristics is becoming more and more popular in the current e-world. As an important member of the biometric family, hand has merits such as robustness, user-friendliness, high accuracy and cost-effectiveness. Hand biometric has many modality characteristics such as palmprint, fingerprint, hand shape, etc. All these modalities need a detection step of the hand that represents the first stage of biometric recognition process. In the literature, three approaches have been proposed to solve the problem of skin detection. The color based approach [1-11], the shape based approach [12, 13] and the hybrid approach [14]. Color is a powerful fundamental feature for skin detection. In fact, it is invariant to the change of position or scale of a given hand and it had shown satisfactory results in the literature. For these reasons, we interested in the color based approach. Regarding this approach, we find three categories of skin modeling named parametric [2, 6, 7], non-parametric [1, 4, 5, 9-11] and explicit models [3, 8]. The importance of non-parametric skin modeling

methods is to estimate skin color distribution from the training data without deriving an explicit model of the skin color. In addition, they are characterized by their rapidity in both training and classification which is very interesting in real-time applications. Therefore, they could be used efficiently for skin detection. The most known methods in the non-parametric skin modeling methods are the normalized lookup table [4, 5], the histogram based segmentation [1], the induction graph [11] and the neural network [9, 10]. The neural network has the ability of a machine learning which produce a generic prediction model from a simple input data. Thus, we opted for the neural network to construct our prediction model.

In the most existing hand biometric systems in the literature, the hand is placed on a glass plate which implies the presence of noise that may provide diseases transfer. Our contribution consists mainly to establish a robust approach for contactless hand detection. A wide variety of color spaces have been applied to the problem of skin color modeling. In the most previous works, researchers opted for a specific color space to construct their skin color model. Our work consists of studying the most widely used color spaces in the problem of skin color modeling in order to determine the adequate color space for our prediction model. In our knowledge, the most existing hand databases in the literature are linked to several constraints like grayscale images with a simple and familiar environment or in other case, markers are placed on the fingertips which disturb the users. For these reasons, another novelty of our work is to build a database which contains hand images detected from freely posed hands in contact free settings that we have called “Sfax-Miracl hand database”.

The remainder of this paper is organized as follows: Section 2 presents the construction details of our database “Sfax-Miracl hand database”. Section 3 describes our proposed hand detection approach. Section 4 discusses the experimental results. Finally, section 5 draws conclusions and futures work.

2 Data Collection: “Sfax-Miracl Hand Database”

Data acquisition is the first stage of every biometric system. Between February 2011 and April 2011, we collected a database of over 1080 hand images captured from 54 peoples where 29 of them are woman and 25 of them are men. This contactless hand images, named “Sfax-Miracl hand database”, consists mainly of hand images collected from volunteers include students, graduate students, workers, retired, etc. 48 of them are less than 30 years old, 3 of them are between 30 and 60 and 2 of them are more than 60 years old. Each one of them is asked to provide about 10 images for their left hand and 10 images for their right hand in one session. In total, each subject provides about 20 images. So that, our database contains a total of 1080 images captured from 108 different hands. This database is built by using Samsung ES75 digital camera. All the images are available in JPEG format with a resolution of 1024*768 pixels. In the process of acquisition, the user places his hand opened in front of the perpendicular axe of the camera with a distance of about 30 cm. The hand images are

captured by taking into account several problems such as the variation of the background (complex, simple, etc.), the variation of the skin tones according to the person and to the races (white, black, brown, red, etc.), the variation of the position and the orientation of the hand, the variation of the hand size which vary according to the age and to the sex, the presence of hand's accessories (ring, bracelet, watch, etc.) and the variation of the lighting conditions (indoor and outdoor environments). We note the presence of parts or the totality of the face for the concerned person or others for some images. The main objective in building this database is to have the unsupervised conditions for the imaging which attempts to represent more realistic application environment. This database is publically available in the contact of the authors. Fig. 1 shows an example of typical acquisitions.

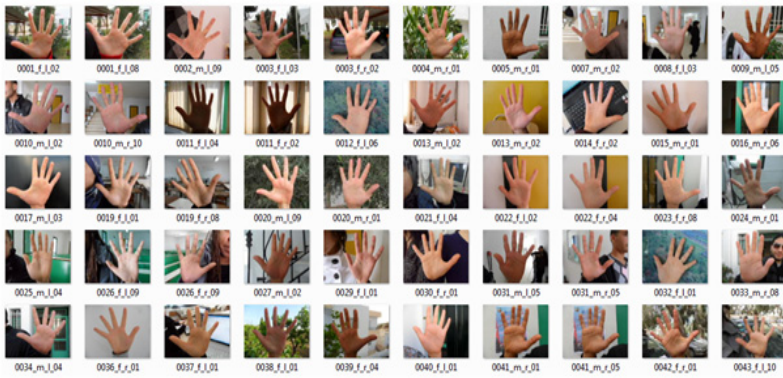


Fig. 1. Extract of images from the database “Sfax-Miracl hand database”

Besides, our database has the advantage to contain masks in which the skin regions were preserved and the background was replaced by black color. These masks are created manually for each images of our database using Adobe Photoshop 7 platform and it were stored in independent files with a name corresponding to its origin images. The creation of these masks served us to extract automatically the skin color and the non-skin color pixels from the original image to be used for the preparation of the training data set and test data set. Fig. 2 shows the mask corresponding to the acquisitions of Fig. 1.

3 The Proposed Hand Detection Approach

Our proposed approach involves two steps: (1) the learning step to construct the adequate prediction model to discriminate the pixels of skin from those of non-skin and (2) the detection step to only conserve the hand region. The total process of our proposed approach is shown in Fig. 3.

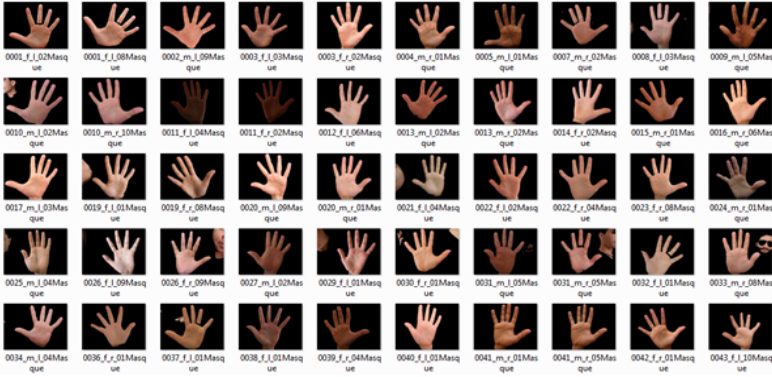


Fig. 2. Corresponding masks to images acquisition for “Sfax-Miracl hand database”

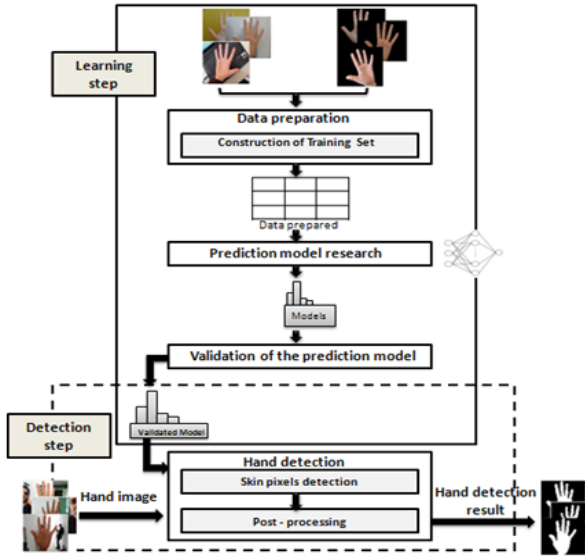


Fig. 3. The total process of our proposed approach

3.1 Learning Step

The learning step is composed of three major phases: (1) a data preparation phase for the construction of the training data set, (2) a prediction model research phase which looks for a generalizable prediction model and (3) a validation phase which consists on assessing the quality of the learned prediction model.

Data Preparation. In this phase, our purpose is to build a two-dimensional table from our training corpus. Each column in the table represents a color space axes and each rows represents a skin pixel value in each color space axes. In the

most previous works, researchers have proposed many prediction models based on skin color, but everyone has opted for a specific color space to present his model. In our work, we try to fill this vacuum by studying the most widely used color spaces in the problem of skin color modeling in order to determinate the adequate color space for our prediction model. The set of training pixels are extracted automatically from the training images and their corresponding binary masks. Thereafter, for each pixel we computed its representation in the used color spaces: RGB, YCrCb, and HSV. This leads to a features vector composed of 9 color space axes. With each pixel features vector is associated its class label denoted as 1 for skin color and 0 for non-skin color.

Prediction Model Research. In this phase, we propose to build a prediction model based on a supervised learning approach. In the literature, there are several techniques of supervised learning, each having its advantages and disadvantages. In our approach, we opted for the use of the MLP as a prediction model method. The MLP [15] is characterized by their capacity of memorization and generalization of the data in addition to their ability to solve not linearly separable problem. A typical MLP network consists of a set of source nodes forming the input layer, one or more hidden layers of computation nodes, and an output layer of nodes. The input signal propagates through the network layer-by-layer. The computations performed by such a feedforward network with a single hidden layer with nonlinear activation functions and a linear output layer can be written as:

$$y = f(x) = B\alpha(Ax + a) + b. \quad (1)$$

where x and y are respectively the vector of inputs and outputs and α is the activation function. A is the matrix of weights of the first layer and a is the bias vector of the first layer. B and b are, respectively, the weight matrix and the bias vector of the second layer.

One of the important aspects in designing an MLP neural network is how to determine the network topology. The input size is dictated by the number of available inputs features. Each color component of the used color space is treated as an input neuron; in our case we need three neurons. The output layer will have one neuron. Thus, the two decisions that must be made regarding the hidden units are to determine the number of hidden layers and the number of neurons in each hidden layer. Fu in 1994 [16] stated that using only one hidden layer is sufficient to solve many practical problems, so in our work we used one hidden layer MLP neural network. The determination of the number of neurons in the hidden layer will be discussed in section 4. The network is trained using error back propagation training algorithm.

Validation of the Prediction Model. Several possible metrics have been proposed in the literature for assessing the quality of a prediction model for skin color pixel detection. Among this metrics, four types of rates were used in our work: (1) The Correct Detection Rate (CDR): represent the probability of the correct classified pixels, (2) The False Detection Rate (FDR): represent the

probability of the wrongly classified pixels, (3) The A Priori Error Rate (APER): represent the probability of the skin pixels which are not assigned to this class, (4) The A POsteriori Error Rate (APOER): represent the probability of pixels assigned to the skin class and which do not belong to this class.

3.2 Detection Step

The detection procedure is done by scanning all the pixels in the image and for each pixel, the MLP neural network takes as input three neurons, each one takes the value of each color component of the chosen color space. Then the MLP neural network computes the probability that each pixel is a skin pixel. So, the pixels whose probability is higher to the threshold value will considered as skin color and takes the value of “1” otherwise it will be considered as background tone and will be set to “0”. This process produces a binary image highlighting the skin color pixels with white color and the others with the black color as shown in Fig. 4(b).

The result of detection produced by the MLP neural network may contain hand region that is corrupted by false detection (Fig. 4(b)). Therefore, a succession of post-processing was proposed allowing us to remove this false detection. We started by applying morphological operator: opening with a circular structuring element. The basic effect for this operation is to remove some false skin pixels detected in the background. Given that in some hand images, the presence of parts or the totality of the face for the concerned person or others, so finally, we eliminate the regions having a surface lower than 500 to preserve only the larger white objects in the image e.g. the hand. The result of post-processing phase is shown in Fig. 4(c).

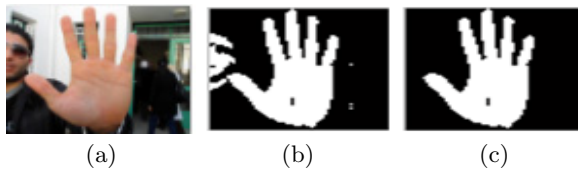


Fig. 4. Hand detection results: (a) original image (b) detection results of skin color (c) hand image after post-processing phase

4 Experimental Results and Evaluation

In the herein reported experiments, we try to validate our contribution and to evaluate the effectiveness and robustness of our approach. All the images used in the following experiments are taken from our database and resized with a size of 64*64 pixels. For the construction of our training data set, we try to choose images from different individuals and having different variations. This training data set is composed of 245760 pixels from which 83349 are skin color pixels and 162411 are non-skin color pixels. Independently of the training data set, we

construct our test data set. This later is composed of 4177920 pixels from which 1376772 are skin color pixels and 2801148 are non-skin color pixels.

Since the transfer function used is a sigmoid function, the MLP neural network produces an output between 0 and 1. Therefore, the output of the neural network needs to be modified so that it is either 0 or 1. To achieve this, we need a threshold value. To choose the threshold value, we calculate the CDR on the training data set using a single neuron in the hidden layer (Table 1).

Table 1. Correct detection rate for different thresholds value

Correct Detection Rate(%)	0.5	0.6	0.7	0.8	0.9
RGB	33.9148	89.7587	89.1207	88.3309	87.168
YCrCb	33.9148	89.9512	89.4775	88.798	86.8542
HSV	33.9148	85.4822	84.8039	80.3259	66.0852

As we have seen from these results, our approach achieved its best performance with the threshold 0.6 for the various used color spaces.

The choice of the adequate neurons number is determined by the calculation of the CDR on the training data set by modifying the neuron number in the hidden layer from 1 to 25 with a step of 2 neurons (Table 2).

Table 2. Correct detection rate for the different neurons number in the hidden layer for the three color spaces: RGB, YCrCb and HSV

Correct Detection Rate(%)	1	3	5	7	9	11	13	15	17	19	21	23	25
RGB	89.76	90.52	66.09	90.35	90.5	90.53	90.41	90.46	90.5	90.4	90.37	90.5	66.09
YCrCb	89.95	90.49	90.48	90.63	90.69	66.08	90.67	90.78	90.66	90.58	90.74	90.73	90.76
HSV	85.48	90.13	89.9	90.11	90.31	90.26	90.37	90.34	90.31	90.38	66.28	90.46	66.08

From the results of previous experiment, we can observe that the best result of correct detection rate is achieved using 15 neurons in the hidden layer.

In order to determinate the adequate color space for our prediction model, we conducted a comparison between the obtained results by the RGB, the YCrCb, and the HSV color space. This comparison does not only concern the obtained CDR but also the FDR, the APER, and the APOER (Table 3).

From this experiment, since a compromise between the CDR, the FDR, the APER, and the APOER, the best result is obtained with the RGB color space. The execution time is another interesting factor in a detection skin process. Table 4 illustrates the average computing time of our proposed approach.

As illustrated from the previous result, our approach is not only effective but also very fast. Therefore, we have once again shown the performance of our proposed approach.

Table 3. Comparison between the CDR, the FDR, the APER, and the APOER using the RGB, the YCrCb, and the HSV color space

(%)	CDR	FDR	APER	APOER
RGB	94.53	5.46	15.57	2.15
YCrCb	94.52	5.47	15.13	2.02
HSV	94.41	5.58	15.67	2.12

Table 4. Average computing time of our proposed hand detection approach

Color space	Average computing time (s)
RGB	0.0625 s

5 Conclusions and Futures Works

We have introduced in this paper a method for hand detection captured without contact and without constraints on the capture environment for hand biometric applications. This method is based on color based approach adopting the MLP for the skin color modeling. First of all, we begin with presenting all the details of the construction process of our database “Sfax-Miracl hand database”. Then, we describe the various steps of our approach. This later involves two steps: the first is the learning step to construct the adequate prediction model and the second is the detection step to only conserve the hand region. Finally, we conducted several experiments in our database. These experiments showed that our approach provide an efficient detection results with a CDR of 94.53% in the RGB color space.

Our future orientations consist of studying the possibility of generating an hybrid color space to better discriminate the skin pixels. Given that the run-time is important for such a biometric application, it would be interesting to explore other learning techniques such as the induction graphs.

References

1. Saxe, D., Foulds, R.: Toward robust skin identification in video images. In: 2nd International Face and Gesture Recognition Conference, pp. 379–384. IEEE Computer Society Press, USA (1996)
2. Yang, M., Ahuja, N.: Detecting human faces in color images. In: IEEE International Conference on Image Processing, Chicago, pp. 127–130 (1998)
3. Garsia, C., Tziritas, G.: Face detection using quantized skin color region merging and wavelet packet analysis. J. IEEE Trans. on Multimedia 1, 264–277 (1999)
4. Berard, F.: Vision par ordinateur pour l’interaction homme-machine fortement couplee. PhD thesis, Joseph Fourier University, France (2000)

5. Wu, A., Shah, M., da Vitoria Lobo, N.: A virtual 3d blackboard: 3d finger tracking using a single camera. In: Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 536–543. IEEE Press, France (2000)
6. Greenspan, H., Goldberger, J., Eshet, I.: Mixture model for face-color modeling and segmentation. *J. Patt. Recogn. Lett.* 22, 1525–1535 (2001)
7. Lee, J.Y., Yoo, S.I.: An elliptical boundary model for skin color detection. In: International Conference on Imaging Science Systems and Technology, Las Vegas, pp. 579–584 (2002)
8. Chiang, C.C., Tai, W.K., Yang, M.T., Huang, Y.T., Huang, C.J.: A novel method for detecting lips, eyes and faces in real time. *J. Real-Time Imaging* 9, 277–287 (2003)
9. Bhoyar, K.K., Kakde, O.G.: Skin color detection model using neural networks and its performance evaluation. *J. of Computer Science* 6, 963–968 (2010)
10. Doukim, C.A., Dargham, J.A., Chekima, A., Omatu, S.: Combining neural networks for skin detection. *J. Signal and Image Processing* 1, 1–11 (2010)
11. Hammami, M., Tsishkou, D., Chen, L.: Data-mining based Skin-color Modeling and Applications. In: Third International Workshop on Content-Based Multimedia Indexing, pp. 157–162. SuviSoft Oy Ltd., France (2003)
12. Cootes, T.F., Taylor, C.J.: Statistical models of appearance for computer vision, Technical report, University of Manchester, United Kingdom (1998)
13. Liu, N., Lovell, B.C.: Hand gesture extraction by active shape models. In: Digital Imaging Computing: Techniques and Applications Conference, pp. 59–64. IEEE CS Press, Australia (2005)
14. Doublet, J., Lepetit, O., Revenu, M.: Hand Detection for Contact less Biometrics Identification. In: International Conference Cognitive System with Interactive Sensors, France (2006)
15. Krose, B., Smagt, P.V.D.: An introduction to neural networks, Amsterdam (1996)
16. Fu, L.M.: Neural networks in computer intelligence. McGraw-Hill, India (1994)

Spiral Cube for Biometric Template Protection

Chouaib Moujahdi¹, Sanaa Ghouzali^{1,2}, Mounia Mikram^{1,3},
Mohammed Rziza¹, and George Bebis⁴

¹ LRIT, Faculty of Sciences, Mohammed V-Agdal University, Rabat, Morocco

² Information Technology Department, CCIS, King Saud University, Saudi Arabia

³ The School of Information Sciences, Rabat, Morocco

⁴ Dept of Computer Science and Engineering, University of Nevada, Reno

moujahdi_chouaib@yahoo.fr

Abstract. In this paper we present a new approach for biometric template protection. Our objective is to build a preliminary non-invertible transformation approach, based on random projection, which meets the requirements of revocability, diversity, security and performance. We use the chaotic behavior of logistic map to build the projection vectors using a new technique that makes the construction of the projection matrix depend on the biometric template and its identity. Experimental results conducted on several face databases show the ability of our technique to preserve and increase the performance of protected systems. Moreover, we demonstrate that the security of our approach is sufficiently robust to possible attacks.

Keywords: Template protection, random projection, logistic map, revocability, security.

1 Introduction

The growing concern for the problem of identity theft and the urgent need for individual privacy make the conception of personal authentication / identification systems increasingly important. These systems must authenticate users respecting several requirements, like speed, reliability, accurately and protection of user's privacy. Traditional systems of personal authentication which use passwords or ID cards are not able to meet all these requirements. For against, authentication systems based on biometrics, which use physiological (face, iris, etc.) and behavioral (signature, etc.) modalities, have proven a priority over traditional systems. But while biometrics ensure uniqueness, they do not provide the secrecy. For example, a person let his fingerprints on every touched surface and face images can be seen everywhere. Consequently, many attacks can be launched against the biometric systems, which reduce the credibility of these systems. Therefore, although biometric technologies have inherent advantages over traditional methods of personal authentication / identification, the problem of ensuring the security of biometric data is critical.

In practice, opponents exploit the structure of biometric systems to launch their attacks. All biometric systems consist of four main modules (Fig. 1): the sensor module.

The feature extraction module that selects the most significant characteristics in an image sent by the sensor and builds a biometric template test. The module of the database containing the biometric templates of legitimate users, and the module of comparison or classification is responsible for comparing the test templates with the templates stored in the database to make a final decision. *Ratha et al* have identified eight points or levels of attack in a biometric system [1] (Fig. 1), but since the principle of some attacks is repeated, *Jain et al* include them into four categories [2]. Firstly, the attacks on the user interface (sensor), mainly due to the presentation of falsified biometric data, for example *spoofing / mimicry* attacks [4] (Level 1). Secondly, the attacks on the interface between modules, an adversary can either destroy or interfere communication interfaces between modules, for example *replay* attacks [3] and *hill climbing* attacks [4] (Levels 2, 4, 7 and 8). Thirdly, the attacks on software modules, the executable program on a module can be modified so that it always returns the desired values by the opponent. It is the *Trojan-horse* attacks (Levels 3 and 5). Finally, the attacks on database (Level 6), one of the most damaging attacks on a biometric system is against the biometric templates stored in the database system. For example, a biometric template can be replaced by an impostor template to obtain unauthorized access to the system. In addition, a physis parody (spoof) can be created from a stolen template [5] to obtain unauthorized access to the system. The irrevocability of biometric templates makes this attack very dangerous, because, unlike a stolen credit card or password, if a template is stolen it is not possible for a legitimate user to revoke their biometric templates and replace them with another set of identifiers.

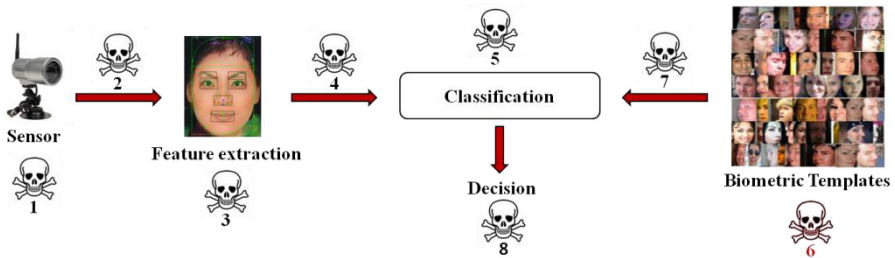


Fig. 1. The eight levels of attack in a biometric system

Because of these security issues, several schemes and methods have been proposed for biometric template protection. The concept of *revocable* (also called *cancelable*) biometric has been proposed, for the first time, as a template security solution by *Ratha et al* [6]. Revocability means that we can revoke a compromised template and replace it with another in the same way as a stolen password is. All the proposed approaches are based on this concept. In this work, we propose a new approach for the protection of biometric template based on the random projection and the phenomenon of chaos, that meets several requirements including the revocability.

The rest of the paper is organized as follows. Section 2 presents an overview of template protection approaches. Section 3 describes the technique of random projection, the phenomenon of chaos, the proposed approach and a security analysis. Experimental results are discussed in Section 4, conclusions and perspectives are drawn in Section 5.

2 Overview of Template Protection Approaches

Personal authentication systems based on biometrics have shown new problems and challenges related to the protection of personal data, inexistent in traditional authentication systems. Because of these problems of security and privacy, there are currently many research efforts to protect biometric systems against the possible attacks. We can divide the proposed solutions into two main categories: *preventive* solutions and *palliative* solutions. Each category can be divided into two main types: *hardware* approaches and *software* approaches.

The objective of *palliative* solutions is, once the attack has been made, to minimize the probability of rupture in the system. The hardware approaches of these solutions try to add specific devices (smell, blood pressure, etc.) in biometric sensors to detect the liveliness / fraud of presented features. Among software approaches of these solutions, one has received more attention from researchers and industry, called *liveliness detection*. The design of these solutions depends on the used biometric trait and there is no one standard approach for all biometric systems.

Preventive solutions are designed to prevent the commission of an attack. In general, these solutions are trying to protect biometric templates. The hardware approaches of these solutions try to put all the modules and interfaces of biometric system on a chip card or a secure processor in general. The software approaches of these solutions are designed to protect the stored biometric templates. The idea is, instead of storing the templates themselves, to store a function of each template used directly in the task of classification. This work is primarily concerned with these solutions of template protection.

An ideal approach of biometric template protection must meet four requirements [7]:

- *Revocability*: it should be possible to revoke a template and put a new template based on the same biometric data.
- *Diversity*: if a revoked template is replaced by a new model, it should not correspond with the former. This property ensures the privacy of the user.
- *Security*: it must be difficult, computationally, to obtain the original template from the protected template.
- *Performance*: The protection approach should not degrade the recognition performance of system.

The major challenge to design an approach of template protection, which meets all requirements, is the presence of intra-subject variations, because multiple acquisitions of the same biometric trait do not lead an identical set of features.

Jain *et al* have classified these approaches into three main categories [2]: *feature transformation* approaches, *biometric cryptosystem* approaches and *hybrid* approaches. The basic idea of feature transformation approaches is to apply a transformation function F to the original biometric template T using a key K , and the transformed template $F(T, K)$ is stored in the database. The function F is also used to transform the test template Q , and we can directly compare the transformed templates $F(T, K)$ and $F(Q, K)$ in the transformation domain to determine whether the user is accepted or not. Depending on the transformation function F , feature transformation schemes can be divided into two classes: *biohashing* and *non-invertible transformation*. For biohashing [4][8], F is

invertible; if an opponent has the key and the transformed template, he/she can recover the original biometric template (or an approximation of it). Therefore, the *biohashing* scheme security is based on the security of the key. For non-invertible transformation [9][10][11], the function F is not invertible. The main property of this approach is that even if the key and / or the transformed template are known, it is difficult for an adversary to recover the original template (in terms of computational complexity). In biometric cryptosystems, the principle of classical cryptosystems is combined with the principle of biometrics to improve security of personal authentication systems based on biometrics. The main objective of these schemes is to minimize the amount of biometric data stored in the database. In these approaches, an error correcting code on the original template B and the key K are applied to extract the helper data H . At the time of authentication, an error correcting code on the helper data H and test template Q are applied to recover the key K and make a decision.

Each of these approaches has its own advantages and limitations [2]. They do not meet, contemporaneously, the requirements of revocability, diversity, security and high performance recognition. Thus, there is no best approach for protecting biometric data and available protection schemes are not yet mature enough for widespread deployment.

In this paper, we propose a new non-invertible transformation approach that allows diversity, revocability, security and performance with no need for a user's key. Our method is based on random projection, a technique that has been applied on various types of problems. We also use the chaotic behavior of logistic map to build the projection vectors which makes the construction of the matrix depend on the biometric template and its identity. The next section describes the proposed approach.

3 Proposed Approach

In this Section, we present a non-invertible transformation approach for biometric template protection, based on the principle of random projection and use the chaotic behavior of logistic map to build the projection vectors.

3.1 Random Projection

Random projection has been applied on various types of problems [12] including the biometric template protection. It uses orthogonal random matrices to project the biometric templates in a space where distances are preserved. To make the projection non-invertible, a quantization step was included in [13].

Stages of the non-invertible random projection are (Fig. 2):

- Generate m random vectors from user key.
- Apply the Gram-Schmidt orthogonalization algorithm on the m random vectors to compute an orthogonal matrix A ($AA^t = I$).
- Transform the original template z using the matrix A :

$$y = Az$$

- y is the transformed template.

- Apply quantization on the transformed template y .

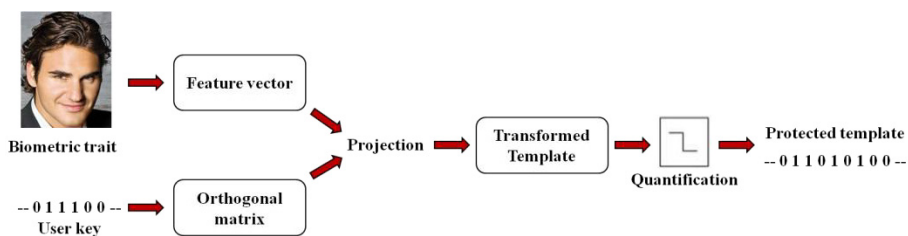


Fig. 2. The non-invertible random projection

The generation of projection matrices using Gram-Schmidt orthogonalization algorithm is time consuming, but there are less expensive methods which do not require this algorithm. For example, *Achlioptas* [14] has proposed a new approach which yields significant computational savings during the computation of the matrix A and the projection Az . Testing this algorithm is one of our objectives in the future work.

In the other hand, the Gram-Schmidt orthogonalization algorithm returns a set of orthogonal vectors if and only if the input vectors are linearly independent. Therefore, the generation of random vectors from the used key will be relatively limited by this requirement. This has motivated us to use the chaotic behavior of logistic map to generate linearly independent vectors that will be used to construct the projection matrices.

3.2 Logistic Map

Logistic map is a sequence whose recurrence is not linear. Its recurrence relation is:

$$x_{n+1} = \mu x_n (1 - x_n)$$

According to the values of μ , we can observe a chaotic behavior in the val $[3.5699456, 4]$. Thus, logistic map is very sensitive to initial conditions. According to this feature, logistic map was used in several applications, such as the protection of data content. For example, random sequences of the chaotic zone can be used to cryptographically secure the transmission channels in several telecommunications systems.

In our work, we use logistic map to generate multiple random vectors. These vectors will be stored in a 4D matrix, called *spiral cube*. Spiral cube will be used to construct the projection matrices. The construction of cubic spiral depends on the size of the original template. Suppose that the feature vector contains n values, the cube will consist of n spiral cells (3D matrix), each cell being of size $m \times m$ (m is the nearest integer greater than or equal to \sqrt{n}) and each cell corresponds to a specific value of μ . Therefore, each cell contains an $m \times m$ box, and each box contains a vector generated using the values μ in the interval $[3.5699456, 4]$ (Fig. 3).

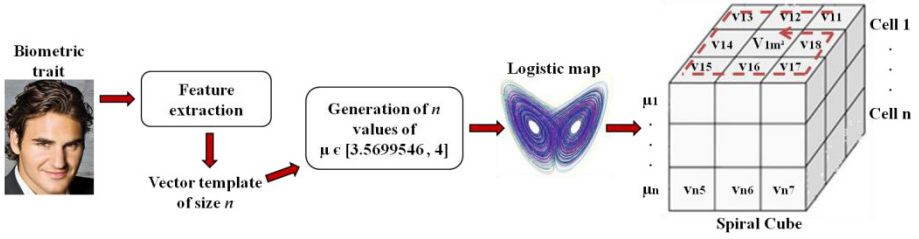


Fig. 3. Spiral cube construction

3.3 Proposed Approach

We propose a non-linear mechanism of random projection. Our objective is to build a non-invertible method for biometric template protection that meets all the requirements of security and performance with no need for a user’s key. It should be noted that the proposed approach is applicable to any biometric system that uses *feature vectors* for classification.

During enrollment, after the extraction of the features from the training templates, (i.e., we assume that the training database contains x templates of size n , each identity is presented by y templates, assuming z identities: $x = y \times z$). Our approach and the mechanism of protection start with the following steps:

- For each training template T , we calculate ∂ :

$$\partial = \frac{|\max(T) - \min(T)|}{m^2} \tag{1}$$

- m is the nearest integer greater than or equal to \sqrt{n} .

- Then, we calculate the quantized vector Q of the template T :

$$Q_i = \begin{cases} 1 & \text{if } T_i = \min(T) \\ m^2 & \text{if } T_i = \max(T) \\ \text{ceil}\left(\frac{|T_i - \min(T)|}{\partial}\right) & \text{else} \end{cases} \quad i \in [1, n] \text{ and } Q_i \in [1, m^2] \tag{2}$$

- $\text{ceil}(a)$ calculates the nearest integer greater than or equal to a .

- At the end of the previous step, we have x quantized vectors. For each identity, we keep a single quantized vector (randomly chosen among the y vectors). Finally, we obtain a matrix \emptyset which contains z quantized vectors. We call it the *map cube*.
- We construct the projection matrices for each identity using the *spiral cube* and the *map cube* (Fig. 4). Assuming that we calculate the projection matrix of identity 1 (first vector of the map cube), the first value of the quantified

vector (size n) of this identity corresponds to the first cell in the spiral cube, and so on for the other values. For example, if the first value is 3, we extract the vector number 3 of the first cell of the spiral cube. Finally, we obtain n vectors and we apply the Gram Schmidt algorithm to construct the projection matrix of the identity 1. The principle is similar for the other identities.

- Finally, we store the protected templates, spiral cube and cube map in the system database (storage of spiral cube and cube map is public).

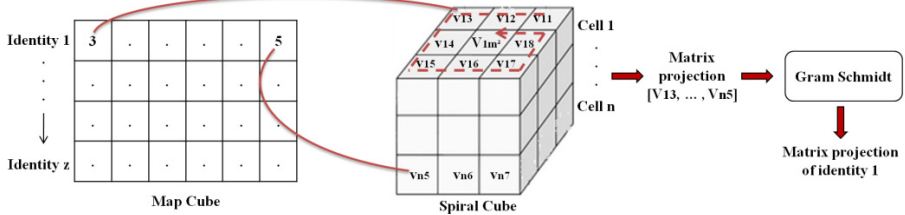


Fig. 4. Mechanism of projection matrix construction

During authentication, after the extraction of the features from the test template, the protected system works as follows:

- We apply quantization on the template test which is similar to that applied on the training templates during enrollment stage (equations 1 and 2).
- We use a KNN classifier to find the nearest vector in the *map cube* to the quantized test template.
- The closest vector is used to find the projection matrix corresponding to the test vector (Fig. 4).
- The protected template is compared directly with the protected training templates; the comparison will be carried out according to the type of classifier used by the system.

3.4 Analysis Study

The proposed technique meets the requirements of revocability, diversity and security. Knowing that multiple acquisitions of the same biometric trait do not yield an identical set of features, the dynamics of our approach allow us to create different templates for the same identity in the presence of these variations. In addition, we can protect a compromised template by changing partially the map cube. We change, specifically, the quantized vector corresponding to the identity of the compromised template, either by redoing the quantization or by changing partially this quantized vector. Thus, the revocability and diversity are assured. It should be noted that the change of the spiral cube, or the order of cells in the cube, require to redo the training task again for all templates in the database; this is a weak point in our approach which we plan to address in our future work.

The security analysis of existing methods is mainly based on the complexity of brute force attacks which assume that biometric data are uniform. We assume that the

size of the original template is n . We analyze the scenario where the adversary has access to the protected template and the spiral cube. To find the original template we need to find the used projection matrix. In this scenario, we have $(m^2 \times n)^n$ (m is the nearest integer greater than or equal to \sqrt{n}) possible projection matrices. For example, if $n=100$ the number of possible matrices is 100^{200} which provides high robustness against brute force attacks. Let us assume now that the opponent has access to the protected template, the spiral cube and the map cube (all public data). If he/she does not know the role of the map cube, the number of possibilities is similar to the previous scenario. Otherwise, if he/she knows the role of a map cube, the number of possibilities is $(m^2 \times z)^n$ (i.e., z is the number of identities), since he/she does not know that the projection vectors are stored spirally in the cube and there is no evidence in the database or public data to determine the storage manner. We have found that even in the worst case scenario where the adversary has all the public data and the template protected, the security is enough to be robust to attacks.

In practice, however, an adversary can exploit the non-uniform structure of data to launch an attack that may require far fewer attempts to reach the security of the system. A rigorous analysis of security of these methods, like [17], is necessary, and will be the objective of future work.

4 Experimental Results

In this Section, we present our experimental results using the YALE and UMIST face databases (Fig. 5). The YALE face database consists of 165 face images of 15 distinct persons. Images are characterized by variations in facial expressions and lighting conditions. The UMIST face database consists of 550 face images of 20 distinct persons. Faces in the database cover a wide range of poses from profile (90°) to frontal (0°) views.



Fig. 5. Examples from UMIST database (top) and YALE database (down)

The biometric system used in our experimentation is based on an efficient feature extraction method LST [15] followed by a multi-class dimensionality reduction approach SVDA [16] for feature selection, and a KNN classifier for classification [18]. For the YALE database, each person is presented with five images. Thus, the training database contains 75 templates while the test database contains 90 images. For the UMIST database, each person is presented with six images and the *leave-one-out*

approach is used for testing on 120 training images [18]. According to the leave-one-out approach, the algorithms are run N times. In each round, $N-1$ samples are used for training and the remaining sample is used for testing. If the test sample is correctly predicted, the test accuracy of the round is 100%, otherwise it is 0%. The overall test accuracy is the mean accuracy of all the N predictions. Figure 6 shows a comparison between the unprotected system and the protected system using the two databases.

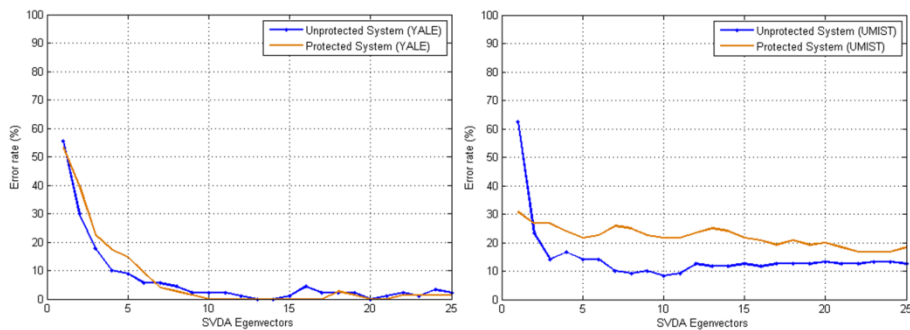


Fig. 6. Comparison of performance between protected systems and unprotected systems

In the case of the YALE database, the performance of the unprotected system is very high because the feature extraction (i.e., LST) and selection (i.e., SVDA) techniques are very efficient. On the other hand, the variation in facial expressions and lighting conditions do not degrade the performance significantly. In the case of the protected system, we can notice a small improvement compared to the unprotected system.

In the case of the UMIST database, the performance of the unprotected system is average because each person is presented with six images (one image per view class). Knowing the principle of the leave-one-out, in each test round, the view class of the test image is not present in the training database, which presents a very complex test situation. Knowing also that the major challenge in designing a template protection approach is the presence of intra-subject variations (the multi-view in our case), we believe that the observed decrease in performance by 3.34% is quite acceptable

Our results show the ability of our technique to increase the performance of protected system in ideal test conditions and to preserve it in non-ideal test conditions.

5 Conclusions and Perspectives

In this paper, we proposed a new approach for biometric template protection. We used the logistic map vector to generate the vectors of projection. We have stored these vectors in a spiral cube, which is used to generate the matrix of protections and depends on the template to be protected. Our approach meets revocability, diversity and security, required in an ideal method for template protection. In addition, it does not only preserve recognition performance but increases it; due to using a dynamic projection matrix for each identity. Thus, it manages better the intra-subject variations. In future work, we will test other biometric modalities such as fingerprints. As for the security analysis, we plan to use the analytical equations presented in [17].

Acknowledgments. Dr. George Bebis is a Visiting Professor in the Department of Computer Science at King Saud University, Riyadh, Saudi Arabia. The first author would like to thank Dr. Abdul Wadood (Computer Engineering Department, CCIS, King Saud University, Riyadh, Saudi Arabia) for his pedagogic assistance. This work has been supported by the Fulbright joint supervision program partially with the project N° PMI 16/09.

References

1. Ratha, N.K., Connell, J.H., Bolle, R.M.: An Analysis of Minutiae Matching Strength. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 223–228. Springer, Heidelberg (2001)
2. Jain, A.K., Nandakumar, K., Nagar, A.: Biometric Template Security. *EURASIP Journal on Advances in Signal Processing* (2008)
3. Syverson, P.: A taxonomy of replay attacks. In: *The Computer Security Foundations Workshop* (1994)
4. Adler, A.: Vulnerabilities in Biometric Encryption Systems. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 1100–1109. Springer, Heidelberg (2005)
5. Adler, A.: Images can be regenerated from quantized biometric match score data. In: *The Canadian Conference on Electrical and Computer Engineering* (2004)
6. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication system. *IBM Systems Journal* (2004)
7. Breebaart, J., Yang, B., Buhan-Dulman, I., Busch, C.: Biometric Template Protection: The need for open standards. *Privacy and Data Security Journal* (2009)
8. Lam, K., Beth, T.: Timely authentication in distributed systems. In: *The European Symposium on Research in Computer Security* (1992)
9. Bolle, R.M., Connell, J.H., Ratha, N.K.: Biometric perils and patches. *Pattern Recognition* (2002)
10. Teoh, A.B.J., Toh, K.-A., Yip, W.K.: 2^N Discretisation of BioPhasor in Cancellable Biometrics. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 435–444. Springer, Heidelberg (2007)
11. Yang, B., Hartung, D., Simoens, K., Busch, C.: Dynamic Random Projection for Biometric Template Protection. In: *The 7th Framework Programme of the European Union, Project TURBINE, ICT-2007-216339* (2010)
12. Goel, N., Bebis, G., Nefian, A.: Face Recognition Experiments with Random Projection. In: *SPIE Defense and Security Symposium (Biometric Technology for Human Identification)*, Orlando, FL, March 28–April 1 (2005)
13. Wang, Y., Plataniotis, K.N.: Face Based Biometric Authentication with Changeable and Privacy Preservable templates. In: *Biometrics Symposium* (2007)
14. Achlioptas, D.: Database-friendly random projections. In: *ACM Symposium on the Principles of Database Systems*, pp. 274–281 (2001)
15. Gu, S., Tan, Y., He, X.: Laplacian Smoothing Transform for Face Recognition. *Science in China Series F-Information Sciences* (2009)
16. Gu, S., Tan, Y., He, X.: Discriminant Analysis via Support Vectors. *Neurocomputing* (2009)
17. Nagar, A., Nandakumar, K., Jain, A.K.: Biometric Template Transformation: A Security Analysis. *SPIE Digital Library* (2010)
18. Moujahdi, C., Ghouzali, S., Mikram, M., Abdul, W., Rziza, M.: Inter-communication classification for multi-view face recognition. In: *International Conference on Multimedia Computing and Systems (ICMCS)*, Tangier, Morocco (2012)

Sparse Representation Based Classification for Face Recognition by k -LiMapS Algorithm

Alessandro Adamo¹, Giuliano Grossi², and Raffaella Lanzarotti²

¹ Dipartimento di Matematica, Università degli Studi di Milano
Via Saldini 50, 20133 Milano, Italy
alessandro.adamo@unimi.it

² Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano
Via Comelico 39, 20135 Milano, Italy
{grossi, lanzarotti}@dsi.unimi.it

Abstract. In this paper, we present a new approach for face recognition that is robust against both poorly defined and poorly aligned training and testing data even with few training samples. Working in the conventional feature space yielded by the Fisher's Linear Discriminant analysis, it uses a recent algorithm for sparse representation, namely k -LiMAPS, as general classification criterion. Such a technique performs a local ℓ_0 pseudo-norm minimization by iterating suitable parametric nonlinear mappings. Thanks to its particular search strategy, it is very fast and able to discriminate among separated classes lying in the low-dimension Fisherspace. Experiments are carried out on the FRGC version 2.0 database showing good classification capability even when compared with the state-of-the-art ℓ_1 norm-based sparse representation classifier (SRC).

1 Introduction

The face recognition problem has been widely studied in several fields, involving biological researchers, psychologists, and computer scientists. This interest is motivated by the still big disparity between the performances achieved by existing automatic face recognition systems (FRSs) [1,2] and human ability in solving this task. FRSs can be classified in local-based or holistic. The first extract local features either on the whole face [3] or in correspondence to peculiar fiducial points [4]. By construction such methods are more robust to variations caused by either illumination or pose changes. Their main disadvantages are the computational cost and the fact that they require a certain image resolution and quality, which cannot be guaranteed in real world applications. The holistic approaches are more suitable in case of low quality images considering they do not require to design and extract explicit features. The most popular are Eigenface [5], Fisherface [6] and Laplacianface [7]. More recently a new approach [8] based on the sparse representation theory [9,10] has been proposed, proving its effectiveness. This method aims to recognize a test image as a sparse representation of the training set, assuming that each object covers a certain subspace. The main disadvantage of this method, and of all the holistic approaches in general, is that

it requires a very precise (quasi-perfect) alignment of all the images both in the training and in the test sets: even small errors affect heavily the performances [11]. Besides, they require to have numerous images per subject for training and it is computationally expensive. All these characteristics are not conceivable for real world applications.

The misalignment problem has been tackled in [12,13], showing a robustness increment of these systems.

In this paper we propose a completely automatic and fast FRS based on the sparse representation (SR) method. Both the training and the test sets are pre-processed with the off-the-shelf face detector presented in [14] plus the eyes and mouth locator presented in [15]. The obtained face sub-images are projected in the Fisher space and then sparsity is accomplished applying the recently proposed algorithm k -LiMAPS [16]. Such method is based on suitable Lipschitzian type mappings providing an easy and fast iterative schema which leads to capture sparsity in the face subspace spanned by the training set.

We tested our method on the FRGC version 2.0 database [17], and compared it with the SRC method. These experiments prove that, despite the system is completely automatic, it is robust with respect to mis-alignments and variations in expression or illumination.

2 Sparse Recovery by k -LiMapS

In this section, we first briefly recall the general sparse recovery (or sparse representation, SR) framework, then we detail our proposed k -LiMAPS algorithm.

2.1 Sparse Recovery

The mathematical problem statement of SR consists in finding the sparsest representation of a vector $x \in \mathbb{R}^n$ given an overcomplete dictionary $\Phi = [\phi_1, \dots, \phi_m]$ assumed to be a collection of $m > n$ atoms or vectors in \mathbb{R}^n . A sparse representation for x can be expressed as a linear combination of atoms, i.e., $x = \sum_i \alpha_i \phi_i$, or equivalently in matricial form

$$\Phi \alpha = x, \quad (1)$$

and is measured in terms of ℓ_0 -norm $\|\alpha\|_0$, simply representing the number of non-zero elements in α . More generally, it is not sensible to assume that the available data x obeys precise equality (1) with a sparse representation $\|\alpha\|_0 = k \ll n$. A more plausible scenario assumes sparse approximate representation in which there is an ideal noiseless signal x (admitting a sparse representation) corrupted by noise, leading to the model $x = \Phi \alpha + \varepsilon$, in which error or noise $\varepsilon \in \mathbb{R}^n$ gives rise, for instances, to measurements or estimates. Adopting this noisy setting, the general goal of finding the sparsest decomposition of the signal x can be rephrased as the constrained minimization problem in ℓ_2 -norm

$$\min_{\alpha \in \mathbb{R}^m} \|x - \Phi \alpha\|^2 \quad \text{subject to} \quad \|\alpha\|_0 \leq k. \quad (\text{P}_0)$$

The optimization problem (P₀) is generally NP-hard. Therefore the objective becomes to find computationally efficient algorithms that can approximately solve (P₀), keeping the purpose of recovering as sparse as possible coefficient vectors α .

2.2 k -LiMapS Algorithm

To promote sparsest solutions to the underdetermined inhomogeneous system (1) we proposed the k -LiMAPS algorithm (k -COEFFICIENTS LIPSCHITZIAN MAPPINGS FOR SPARSITY) [16]. For a desired sparsity level $k > 0$ fixed a priori, the method iterates a parametric family of nonlinear mappings along the affine space associated to the system favoring sparse near-feasible solutions. To recover in turn admissible solutions, an alternating stage envisages the use of an orthogonal projector onto the feasible space. The process yields a Cauchy sequence in the Hilbert space ℓ_2^m for which limit point exists regardless of the initial guess. At the end of the process, depending on whether the signal under exam x admits or not a k -sparse representation, an hard thresholding operation is applied to solution $\alpha \in \mathbb{R}^m$ so that $\|\alpha\|_0 = k$.

More specifically, let $\mathcal{F} = \{F_\lambda : \mathbb{R}^m \rightarrow \mathbb{R}^m \mid \lambda \in \mathbb{R}^+\}$ the one-parameter family of nonlinear mappings promoting sparsity via near-feasible points defined by

$$F_\lambda(x) = x \odot \left(1 - e^{-\lambda|x|}\right), \tag{2}$$

where \odot denotes the Hadamard product. The orthogonal projector aimed to map every point falling in the range of (2) into the nearest point in the affine space $\mathcal{A}_{\Phi,x} = \{\alpha \in \mathbb{R}^m : \Phi\alpha = x\}$, supposed not empty, is the usual projector $P = I - \Phi^\dagger\Phi$, where I is the identity operator and $\Phi^\dagger = (\Phi^T\Phi)^{-1}\Phi^T$ the Moore-Penrose pseudo-inverse of the (assumed) full-rank matrix Φ . Moreover, the point $\nu = \Phi^\dagger x$ represents the closed-form least-squares solution of system (1), frequently used as starting point of the trajectory iteratively generated.

As a consequence, by composing mapping (2) and projector P , we are given with the new mapping $G_\lambda : \mathbb{R}^m \rightarrow \mathcal{A}_{\Phi,x}$, defined as

$$\begin{aligned} G_\lambda(\alpha) &= PF_\lambda(\alpha) + \nu \\ &= \alpha - P\alpha \odot e^{-\lambda|\alpha|}, \end{aligned} \tag{3}$$

whose iterations will help in finding sparse solutions of the system. In fact, the search is accomplished by the sequence of successive approximations inductively defined by

$$\begin{cases} \alpha_0 = \alpha \in \mathbb{R}^m \\ \alpha_n = \alpha_{n-1} - P\alpha_{n-1} \odot e^{-\lambda_n|\alpha_{n-1}|}, \quad n \geq 1 \end{cases} \tag{4}$$

where the sequence of positive parameters $\{\lambda_n\}_{n \geq 1}$ is suitably defined as follows. In order to meet the severe constraint of choosing k coefficients not null and discarding the remaining $m - k$, by denoting with $\hat{\alpha}_n$ the absolute values of α_n rearranged in descending order and $\hat{\alpha}_N(k+1)$ its $(k+1)$ -th element, the sequence

of parameters $\{\lambda_n\}$ is adaptively assumed to be $\lambda_n = 1/\hat{\alpha}_n(k+1)$. In this way the algorithm preserves the k most significant coefficients annihilating little by little the remaining $m-k$.

The overall algorithmic schema explained above is summarized by the pseudocode in Algorithm 1.

Algorithm 1. k -LIMAPS

Require: Projector $P = I - \Phi^\dagger \Phi$, sparsity level k , initial guess $\nu = \Phi^\dagger x$

```

1:  $\alpha \leftarrow \nu$ 
2: while [ $\|P\alpha \odot e^{-\lambda|\alpha|}\| > \epsilon$ ] do
3:    $\sigma \leftarrow \text{sort}(|\alpha|)$                                 <descending order coefficients>
4:    $\lambda \leftarrow 1/\sigma_k$                                 <sparsity ratio update>
5:    $\alpha \leftarrow \alpha - P\alpha \odot e^{-\lambda|\alpha|}$           <orthogonal projection>
6: end while
7:  $\alpha_j \leftarrow 0 \quad \forall j \text{ s.t. } |\alpha_j| \leq \sigma_k$    <thresholding>

```

Ensure: an approx. solution of $s = \Phi\alpha$ s.t. $\|\alpha\|_0 \leq k$

Regarding the convergence analysis, it can be stated that the sequence of iterates $\{\alpha_n\}_{n \geq 0}$ in k -LIMAPS converges to a fixed point of (3), by locally minimizing problem (P₀).

3 Face Recognition via k -LiMapS

Here we show that k -LIMAPS can be used in a FRS, enforcing the belief that sparsest representation is naturally discriminative. Our algorithm achieves small classification error by selecting the most representative subset of atoms belonging to the same target subject, while rejecting all other possible but less compact representations. To achieve good performances both in time and in representation quality, we choose to work into the feature space yielded by the LDA transform, briefly sketched in the next sub-section.

3.1 LDA Subspace Analysis and Image Embeddings

In a typical setting for face recognition, we face the problem of correctly determining to which class belongs a given test image among c distinct classes or subjects. Arranging data in a matrix structure, the training samples from the i -th class are represented as column vectors of the matrix $A_i = [x_1, \dots, x_{n_i}] \in \mathbb{R}^{n \times n_i}$. The training set collecting all subjects is then obtained by stacking all matrices A_i into matrix $A = [A_1, \dots, A_c]$.

In order to explore the feature structure of A_i , one of the most popular discriminative tool is the so called Linear Discriminant Analysis (LDA) [6] or Fishfaces. Defining the between-class scatter matrix S_B and the within-class scatter

matrix S_W for the training set A as in [6], LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Such a projection is chosen to maximize the Fisher Discriminant Criterion, i.e.,

$$W_{\text{OPT}} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|}.$$

The main drawback of LDA is related to the potential singularities of the matrix S_W because the dimension of the sample space is typically larger than the number of samples in the training set. To overcome this problem, the Fisherfaces method foresees a Principal Components Analysis (PCA) [5] stage in order to project the original data on a lower dimensional space where the within-class scatter matrix may be nonsingular. Formally, the complete transform is given by

$$W_{\text{LDA}}^T = W_{\text{OPT}}^T W_{\text{PCA}}^T, \quad (5)$$

where the space is reduced to the dimension $c - 1$, which is in general much smaller than n .

3.2 FRS Based on k -LiMapS

Our FRS requires to construct a dictionary Φ on the basis of the training images (k images per subject). At first all the training samples corresponding to the c subjects to recognize are collected in a matrix A (as described in subsection 3.1). Then, applying the LDA projection in (5), the dictionary $\Phi = W_{\text{LDA}} A$ is created, being each atom a $(c - 1)$ -dimensional vector. Successively, in the classification stage, given a test image x , we calculate the projected sample $y = W_{\text{LDA}} x$ and then we perform k -LiMAPS to find sparse vector α such that $\Phi \alpha \approx y$.

In the purpose of solving the membership i of the test image x , we look for the linear span (i.e., LDA subspace) of the training samples associated with the subject i that better approximates the feature vector y . In other words, by denoting with $\hat{\alpha}_i$ the coefficient vector whose only nonzero entries are the ones in α associated to class i , we classify y minimizing its residual with the linear combination $\Phi \hat{\alpha}_i$, i.e., applying the following rule:

$$j = \min_{i \in [1, \dots, c]} \|y - \Phi \hat{\alpha}_i\|.$$

The class assigned to x will be the j so found.

4 Experimental Results

We tested the proposed technique on the FRGC version 2.0 database [17]. The dataset reports images of 466 people acquired in several sessions (from 1 to 22, varying from person to person), over two periods (Fall 2003 and Spring 2004). A session consists of six images: four *controlled* and two *uncontrolled*, both acquired with either neutral or smiling face expression. Controlled images are

Table 1. The face recognition rate (%) on the FRGC 2.0 controlled, varying the cardinality. In brackets we report the number of features which brought to such percentage.

# Subj	50	100	150	200	239
$k = 3$	97.6 (100)	96.4 (180)	95.6 (200)	94.9 (340)	93.9 (360)
$k = 4$	98.4 (100)	98.3 (200)	97.0 (250)	96.9 (390)	95.4 (490)
$k = 5$	98.8 (160)	98.2 (230)	98.2 (280)	97.2 (340)	97.2 (390)

acquired in frontal pose, with homogeneous illumination, while the uncontrolled ones represent smaller faces, often blurred and acquired in several illumination conditions. For our experiments we considered only the subjects with at least three sessions per period. This brought us to 239 subjects.

All the experiments have been carried out on images automatically localized with the face detector proposed in [14] followed by the eyes and mouth locator presented in [15]. *No human intervention* is required. The misalignment we deal with is exemplified in Fig. 1.

**Fig. 1.** Examples of automatic cropping on uncontrolled images of two subjects

Furthermore the number of images in the training set has been deliberately kept low (k varying between 3 and 5) even if the database would allow a richer subject representation. This has been done in order to emulate real world settings. The results we report have been obtained mediating over 20 experiments; at each iteration, k images are randomly selecting for training and the remaining are used to construct the test set. Comparisons have been carried out with the state-of-the-art SRC [8], with a feature space dimension equal to 100, which is a good compromise between the performances and the computational costs.

We set up several experiments¹: first, we explored the *system scalability*: considering only the controlled images of people with neutral expressions, we tested the system performances incrementing the subjects cardinality. As shown in Table 1, the decrease of performances is more important for small values of k .

Second, we investigated how the *expression variation* influences the performances. In the first two columns of Table 2 we report the results obtained by both our algorithm and the SRC, varying k and the pool of images: either neutral or neutral and smiling. In all these experiments we considered 239 subjects. As we can see, the expression variation causes a loss of less than one percentage point for both our method and the SRC, showing a desirable invariance to the expressions.

¹ Matlab code of k -LiMAPS used in the tests is available on the website <http://dalab.dsi.unimi.it/software/klimaps-face-recognition.tgz>.

As last case, we explored the system behavior on *uncontrolled* images reporting the results in the last column of Table 2. This is the more realistic and challenging scenario, where the subjects are non-collaborative and the acquisition conditions non-optimal. In this case the performances are poorer, reflecting the challenge of the task. The low quality of these images affects the recognition percentage in two ways: first the face locator is less precise, resulting in more mis-aligned faces (see Fig. 1). Second, the feature extractor itself has to deal with less discriminative information deleted by blurring, and even misleading information caused by shadows or glasses. What we highlight however is the large gap between the performance we achieve and the SRC ones. This confirms that our method is more robust in presence of misalignment and unfavorable conditions.

Table 2. The face recognition rate (%) on 239 subjects of the FRGC 2.0 controlled, neutral versus neutral and smiling and FRGC 2.0 uncontrolled

	NEUTRAL		NEUTRAL AND SMILING		UNCONTROLLED	
	k-LiMapS	SRC	k-LiMapS	SRC	k-LiMapS	SRC
$k = 3$	93.9 (360)	92.8	93.2 (380)	91.8	77.1 (390)	68.4
$k = 4$	95.4 (490)	95.3	94.6 (500)	94.7	82.8 (360)	74.7
$k = 5$	97.2 (390)	96.6	96.3 (460)	96.2	87.2 (380)	79.1

Finally, we make two general considerations. First, we remark that the feature space dimension is not a critical parameter for our method: the reported performances are achieved for a wide range of feature dimensions (± 200 features from the optimal); here we report the best one to give an idea of the corresponding order of magnitude. Second, to hint at computational time, it turns out that in relation to the largest feature spaces considered in the experiments (equal to 1000), k -LiMAPS processes a test image in about 0.05 seconds, making it a very fast heuristic for face recognition.

5 Conclusions

In this work, we outline a new approach for face recognition based on the paradigm of sparsity recovery. We have experimentally shown that it produces well separated classes in low-dimensional subspaces (e.g., Fisherspace) exhibiting good performances also in case of misalignments and large variations in lighting and facial expressions. In particular it should be stressed that such a technique results particularly suitable for realistic world applications where one has to deal with not only uncontrolled conditions but also very few examples available for training purpose. Future work suggests to face up to more challenging problems like recognition of noisy or partially occluded images as well as to show the independence of any database. This would prove the applicability of our method in a wide variety of real world applications of FRSs.

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: a literature survey. *ACM Computing Surveys* 35, 399–458 (2003)
2. Rabia, J., Hamid, R.: A survey of face recognition techniques. *Journal of Information Processing Systems* 5 (2009)
3. Perez, C., Cament, L., Castillo, L.E.: Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count. *Pattern Recognition* 44, 951–963 (2011)
4. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19, 775–779 (1997)
5. Turker, M., Pentland, A.: Face recognition using Eigenfaces. *Journal of Cognitive Neuroscience* 3, 71–86 (1991)
6. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 711–720 (1997)
7. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face recognition using laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27, 328–340 (2005)
8. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 31, 210–227 (2008)
9. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution. *Comm. Pure Appl. Math.* 59, 797–829 (2004)
10. Candes, E., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* 59, 1207–1223 (2005)
11. Delac, K., Grgic, M. (eds.): *Face Recognition*. I-Tech Education and Publishing (2007)
12. Yan, S., Wang, H., Liu, J., Tang, X., Huang, T.: Misalignment-robust face recognition. *IEEE Transactions on Image Processing* 19, 1087–1096 (2010)
13. Wagner, A., Wright, J.: Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 34, 372–386 (2012)
14. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518 (2001)
15. Campadelli, P., Lanzarotti, R., Lipori, G.: Precise eye and mouth localization. *International Journal of Pattern Recognition and Artificial Intelligence* 23 (2009)
16. Adamo, A., Grossi, G.: A fixed-point iterative schema for error minimization in k -sparse decomposition. In: *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2011)*, pp. 167–172 (2011)
17. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K.: Overview of the face recognition grand challenge. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2005)

3D Face Recognition Using an Expression Insensitive Dynamic Mask

Sadegh Salahshoor and Karim Faez

E.E. Dept. Electrical Engineering
Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran

{S.Salahshoor, KFaez}@AUT.ac.ir

Abstract. Human face recognition is one of the most popular biometric approaches. In last decade 3D face recognition attracted much attention. In this paper, we present an automatic face recognition algorithm and demonstrate its performance on the Bosphorus 3D face database. A novel Dynamic mask is used to segment automatically the regions of face which are less sensitive to expressions. We applied a multilayer perceptron (MLP) to compute maskable region (MR). MR shows which percentage of face image pixels must be masked to produce the expression insensitive binary mask for 3D faces. We applied a modified nearest neighbor classifier for identification. We only used one neutral frontal face of each subject as gallery images and tested our algorithm with emotional expression images. The identification rate obtained is 85.36 percent in non-neutral expression.

Keywords: Biometrics, 3D Face Recognition, Facial Expression.

1 Introduction

Face recognition is a great challenge for computer vision and is one of the most attractive biometric approaches. In last decade, great advantages occurred in face recognition, but the corresponding methods are not robust and reliable enough. [1]

The performance of face recognition systems can be affected by five key factors: 1) Illumination variations; 2) Pose changes; 3) Expression variation; 4) Time delay; 5) Occlusions. [1]

Recent researches on 2D and 3D face recognition systems show that 3D face recognition is more robust to illumination variations and pose changes. [2]

Chua et al [3] used iterative closest point (ICP) for 3D face recognition and they applied a Gaussian distribution to separate the rigid and non-rigid parts of the face. Zhong et al [4] used 3D depth images. Xu et al [10] applied the principal component analysis (PCA) to construct the 3D eigenfaces and used a nearest neighbor classifier to recognition. Queirolo et al [5] applied the simulated annealing (SA) for registration with surface interpenetration measure (SIM) as similarity measure. They combined SIM values of four different face regions: circular and elliptical areas around the nose,

forehead, and the entire face region to achieve recognition. Xu et al [6] applied Gabor wavelet filter on the depth and intensity images; they used linear discriminate analysis (LDA) and adaboost learning to extract the features. Mian et al [7] used the spherical face representation (SFR) and scale-invariant feature transform (SIFT) to reduce the number of candidate faces. Since the eye-forehead and nose regions are less sensitive to expressions, they used a modified ICP on these regions to achieve recognition.

In this paper, we propose an automatic 3D face recognition system robust to expression variations. (see fig. 1) we applied a multilayer perceptron (MLP) to compute the maskable region (MR) for each probe image. This value is used to produce a dynamic mask which separates the least sensitive region of 3D face Image for recognition.

The rest of this paper is organized as follows. Section 2 introduces the preprocessing procedure including 3D face denoising and normalization which are very important to robust face recognition. Section 3 describes the proposed Artificial Neural Network dynamic mask structures. Section 4 reports our experimental result. Finally, in section 5, a conclusion summarizes this paper.

2 Preprocessing

Since most of acquired 3D face samples are noisy and without alignment, a preprocessing approach is needed. Preprocessing includes three steps.

At the first step, 3D face images must be denoised, three Gaussian filters are applied to remove the spikes and reduce the noises. These filters smooth the data with different variations. But smoothing may cause to lose some sharp details of images, so it is a trade-off between the denoising and keeping the important information of images.

At the second step, the region of interest (ROI) in the face is selected. For this purpose, we detect the nose tip at first. The point with highest z value in the middle of face region is selected as approximate nose tip. Then spherical region around the nose tip is cropped from the denoising data.

At the last step, the 3D faces have been aligned to a standard posture by ICP algorithm [8]. ICP is used to align the forehead and nose regions of a 3D face (see fig. 2b), which is less sensitive to facial expression, with a reference image (see fig. 2a). The reference image is a mean image of the neutral 3D faces of database. Finally the aligned face is normalized to a 160x135 depth image (see fig. 2c). The block diagram of the preprocessing procedure is shown in fig. 3.

3 3D Face Recognition Using Dynamic Mask

The algorithm starts by finding the maskable region (MR) for each probe image. An Artificial Neural Network (ANN) is applied to calculate a list of MRs for each gallery images. The minimum size MR in the list is selected as desired MR. An MR shows which percentage of face image pixels must be omitted by making a binary mask.

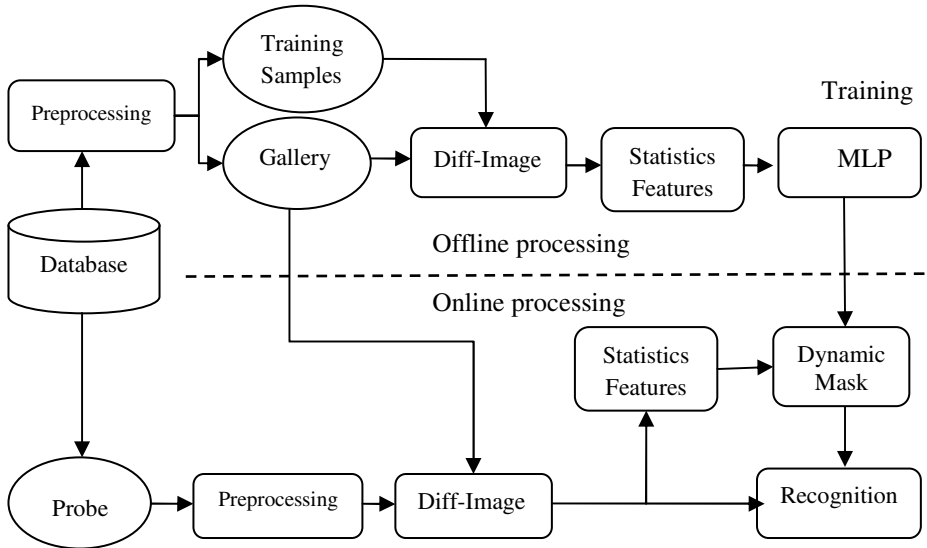


Fig. 1. Block diagram of our recognition algorithm. The dashed line separates the online and offline phases.

3.1 Artificial Neural Network (ANN)

A Multilayer perceptron (MLP) is applied to compute MR. The MLP is a feed forward ANN which comprises of multiple layers of nodes. Each node is fully connected to nodes of the previous and next layers. The nodes in the same layer do not connect together. The nodes are called neurons and each of them has a nonlinear activation function. Hyperbolic tangent and logistic function are the most well-known activation functions which are sigmoid, and described by:

$$f(x) = \tanh(x). \quad (1)$$

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (2)$$

where the former ranges from -1 to 1 and the latter ranges from 0 to 1.

Back propagation (BP) method is used to train an MLP. BP is a kind of supervised learning algorithm. In this algorithm, weights of the nodes are updated by back propagating the errors between desired and actual outputs.

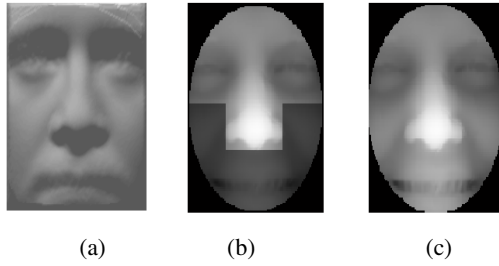


Fig. 2. (a) Reference image (mean face). (b) Forehead region which is used to align the expression faces. (c) A normalized depth image.

We used a simple three layers, which has 11 nodes in the input layer, nine neurons in the hidden layer and only one node in the output layer. Inputs are statistical features of the Differential-Image which is described by:

$$Diff_{Image} = |probe - gallery|. \tag{3}$$

where $|x|$ denotes the absolute number of x 's. Then we reshaped this $M \times N$ matrix to a $NM \times 1$ Differential-vector and sorted it into ascending order.

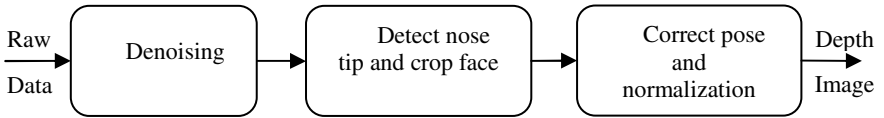


Fig. 3. Block diagram of the preprocessing Procedure

The statistical inputs are minimum, maximum, median, 1st and 3rd quartile and 9th decile of Differential-vector. The others Inputs are:

a) arithmetic mean:

$$mean = \frac{1}{n} \sum_{i=1}^n x(i). \tag{4}$$

b) interquartile mean:

$$IQmean = \frac{2}{n} \sum_{i=\frac{n}{4}+1}^{\frac{3n}{4}} x(i). \tag{5}$$

where x is an ascend vector.

c) Geometric mean:

$$Gmean = (\prod_{i=1}^n x(i))^{\frac{1}{n}}. \tag{6}$$

d) harmonic mean:

$$Hmean = \frac{n}{\sum_{i=1}^n \frac{1}{x(i)}}. \quad (7)$$

and e) the standard deviation:

$$std = \sqrt{\frac{1}{n} \sum_{i=1}^n (x(i) - mean)^2}. \quad (8)$$

3.2 Dynamic Mask

Differential-Images (Eq. 3) are used to make binary masks for each pair of probe-gallery faces. These masks separate the regions of face which are the least affected by facial expression. A dynamic threshold for each pair of probe-gallery faces is selected to make the dynamic mask.

$$DTh = Diff_{vector}(N * (1 - MR)). \quad (9)$$

where DTh denotes the dynamic threshold, $Diff_{vector}$ is a vector corresponding to the facial regions of differential- image, which is sorted into ascending order. N denotes the length of $Diff_{vector}$ and MR is the maskable region.

In order to make a dynamic mask, the pixels of differential-image whose magnitudes are less than the threshold are set to 1 and other pixels are set to zero. Some examples of dynamic mask are shown in fig. 4.

3.3 Classification

The classification is performed using a modified nearest neighbor algorithm (NN). NN computes the distances from a probe face to all gallery faces. The number of pixels equal to one in the dynamic masks, might be different for each pair of probe-gallery images; however, MRs are unique for each probe images. This can affect the accuracy of our algorithm, so we applied a modified mean squared error (MMSE) to measure the distance which is described by:

$$MMSE = \frac{1}{K} \sum_{i=1}^N \sum_{j=1}^M mask(i, j) * (P(i, j) - G(i, j))^2. \quad (9)$$

where P is a probe and G is a gallery face and the mask is a dynamic mask which is made for this pair of probe-gallery faces and all of them are $N \times M$ matrices. K is the number of pixels in dynamic mask which are equal to unity.

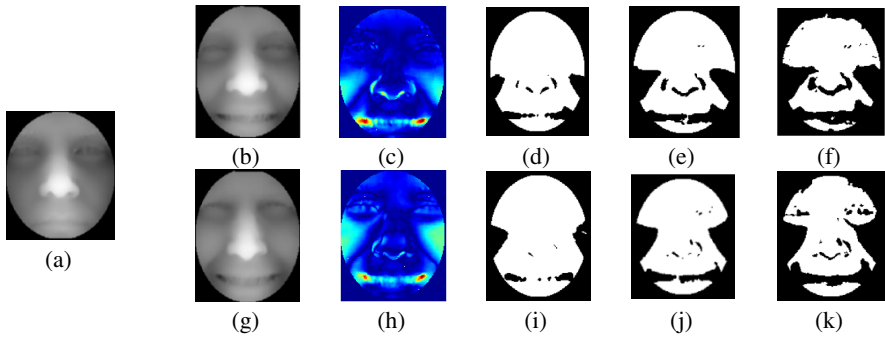


Fig. 4. some examples of dynamic mask. (a) neutral expression. (b, g) happy expressions. (c, h) the differential Images. (d, i) dynamic masks (IP=0.1). (e, j) dynamic masks (IP=0.2). (f, k) dynamic masks (IP=0.3). first row (b-f) and (a) are belonging to the same person.

4 Experimental Result and Discussions

We tested our 3D face recognition method on the emotional expression faces of Bosphorus database [9].

4.1 Bosphorus 3D Face Database

Bosphorus 3D face database comprises 4666, 3D faces (shape and texture). It consists of 105 subjects (61 men and 44 women) in various expressions, pose and occlusion.

Bosphorus database comprises two set of expressions, the first set consists of the expressions which are known as action units (AUs) [10] and the second set comprise emotional expressions, which are anger, disgust, fear, happiness, sadness, and surprise, but all subjects do not have all the emotional expressions. In table 1, the frequency of each emotional expression in the database is given; also fig. 5 shows some 3D faces of Bosphorus database.

Table 1. Frequency of each of emotional expressions in Bosphorus database

anger	disgust	fear	happiness	sadness	surprise
71	69	70	105	66	70

4.2 Identification Result

In the experiments, the neutral expression face is used as the single gallery of a person, and the emotional expression faces are used as probes.

At offline processing approach, the best ignored rates (MR) for some samples of database are calculated, and then these samples are used to train the Neural Network. Since the Maskable Region range from 0 to 1, the logistic function (Eq.2) is used as activation function of MLP neurons. We used this trained MLP to make the dynamic mask at online processing.

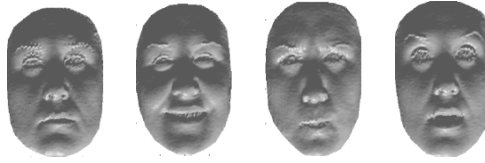


Fig. 5. Some samples of Bosphorus 3D face database

We compare our proposed dynamic mask with three different face regions: circular and elliptical areas around the nose and the entire face, and applied some fundamental methods such as, 3D eigenfaces [11], and Gabor features for this comparison. Gabor filters of five scales and eight orientations were used to extract the Gabor features, and the PCA was applied to reduce the dimensionality of the feature space. In table 2, the comparing results are given. It shows that the proposed method achieve an average accuracy of 85.36 percent and the recognition range for different expressions range from 74.29 to 95.43 percent.

Table 2. Comparing rank-one identification rates of the proposed method with other methods

	Circular areas around the nose		Elliptical areas around the nose		Entire face		
	3D eigenface [11]	Gabor + PCA	3D eigenface [11]	Gabor + PCA	3D eigenface [11]	Gabor + PCA	The proposed method
anger	80.28	71.83	81.69	70.42	83.10	36.62	87.32
disgust	68.12	63.77	78.26	57.97	34.78	26.09	78.26
fear	80	72.86	72.86	68.57	35.71	32.87	92.86
happiness	58.10	75.24	54.29	59.05	12.38	23.81	74.29
sadness	84.85	83.33	81.82	80.30	84.85	53.03	95.45
surprise	82.86	75.71	64.29	71.43	24.28	32.86	91.43
overall	74.28	73.84	70.73	67.18	43.02	33.26	85.36

5 Conclusions

In this paper, we proposed a new 3D face recognition system uses an expression insensitive dynamic mask for identification. In this new proposed method, a MLP is applied to estimate the Maskable Region (MR) which is used to make the dynamic mask. This mask separates the less sensitive region of face, and finally a modified nearest neighbor classification algorithm is used to recognition. Experiments on the Bosphorus 3D face database show that our method is more effective than the region based method. On the Bosphorus database with one neutral face per individual in the gallery, the proposed method achieved the rank-one identification rate of 85.36 percent.

References

1. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D Face Recognition: A survey. *Pattern Recognition Letters*, 1885–1906 (2007)
2. Bowyer, K., Chang, K., Flynn, P.: A Survey of approaches and challenges in 3D and multi-model 3D + 2D face recognition. In: *CVIU*, vol. 101, pp. 1–15 (2006)
3. Chau, C.S., Han, F., Ho, Y.K.: 3D human face recognition using point signature. *Automatic Face and Gesture Recognition*, 233–238 (2000)
4. Zhong, C., Sun, Z., Tan, T.: Robust 3D face recognition using learned visual codebook. *Pattern Recognition*, 1–6 (2007)
5. Queirolo, C.C., Silva, L., Bellon, O.R.P., Segundo, M.P.: 3D face recognition simulated annealing and the surface interpenetration measure. *IEEE Transaction, Pattern Analysis and Machine Intelligence* 32(2), 206–219 (2010)
6. Xu, C., Li, S., Tan, T., Quan, L.: Automatic 3D face recognition from depth and intensity Gabor features. *Pattern Recognition* 42, 1895–1905 (2009)
7. Mian, A.S., Bennamoun, M., Owens, R.: An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transaction, Pattern Analysis and Machine Intelligence* 29(11), 1927–1943 (2007)
8. Besl, P.J., McKay, N.D.: A method for registration of 3D shapes. *IEEE Transaction, Pattern Analysis and Machine Intelligence* 14, 39–256 (1992)
9. Savran, A., Alyüz, N., Dibeklioglu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus Database for 3D Face Analysis. In: Schouten, B., Juul, N.C., Drygajlo, A., Tistarelli, M. (eds.) *BIOID 2008*. LNCS, vol. 5372, pp. 47–56. Springer, Heidelberg (2008)
10. Ekman, P., Friesen, W.V.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto (1978)
11. Xu, C., Wang, Y., Tan, T., Quan, L.: A new attempt to face recognition using 3D-eigenfaces. In: *The 6th Asian Conference on Computer Vision (ACCV)*, vol. 2, pp. 884–889 (2004)

Score Fusion in Multibiometric Identification Based on Fuzzy Set Theory

Khalid Fakhar¹, Mohammed El Aroussi^{1,2},
Mohamed Nabil Saidi^{1,3}, and Driss Aboutajdine¹

¹ GSCM-LRIT Research Laboratory (associated to CNRST, URAC 29),
Mohammed V University – Agdal, Rabat, Morocco

kld.fakhar@gmail.com, aboutaj@fsr.ac.ma

² LETI, EHTP, Casablanca, Morocco

moha387@yahoo.fr

³ INSEA, Rabat, Morocco

msaidi@insea.ac.ma

Abstract. Multimodal biometric systems consolidate or fuse information from multiple biometric sources. They have been developed to overcome several limitations of each individual biometric system, such as sensitivity to noise, intra class invariability, data quality, non-universality and other factors. In this paper, we propose a general framework of multibiometric identification system based on fusion at matching score level using fuzzy set theory. The motivation for using fuzzy set theory is that it offers methods suited to treat (modeling, fusion,...) and take into account the information inherently uncertain and ambiguous. We note that our fusion system is based on face and iris modalities. Experimental results exhibit that the proposed method performance bring obvious improvement compared to unimodal biometric identification methods and classical combination approaches at score level fusion.

Keywords: Multimodal biometrics, identification, score level fusion, fuzzy set theory.

1 Introduction

Biometrics is the science of establishing human identity based on their physical or behavioral characteristics [1]. Biometric systems can operate in two modes [2]: verification and identification. In the verification mode, the system validates a query biometric by comparing only the captured biometric data with the biometric template of a specific identity stored in the database. There is one-to-one comparison in this case. In the identification mode, the user's biometric input is compared with the templates of all the persons enrolled in the database. Therefore, the system conducts a one-to-many comparisons to establish individual's identity (closed-set) or fails if the input is not enrolled in the database (open-set). The identification task is significantly more challenging than verification [3].

Biometric systems based on a single biometric trait suffer from limitations such as sensitivity to noise, data quality, non-universality and spoof attacks [4]. Multimodal biometric systems which combine multiple sources of biometric information have been developed in order to overcome those problems and to achieve better recognition performance [5]. The technique implementation of this kind of systems requires the use of information fusion theory. There are three main fusion strategies [6]: fusion at the feature extraction level, matching score level and decision level.

In our work, we focused on biometric fusion at score level because it offers the best trade-off in terms of the information content and the ease in fusion [5]. Unlike most studies that have been used in this fusion level and examined for the verification mode [7], we aim, in this paper, to introduce a general framework for the closed-set identification based on fuzzy set theory [8] where each biometric matcher output will be modeled as a fuzzy set.

The rest of this paper is organized as follow. In section 2, we introduce the proposed system, then we describe the multimodal biometric system used in our work in section 3. The experimental results are discussed in section 4. Finally some conclusions end this paper.

2 Fuzzy Set Theory and Multibiometric Identification System

2.1 Fuzzy Set Theory

Fuzzy set theory was introduced by Zadeh in 1965 as a means of representing and manipulating imprecise or uncertain information [8]. It plays a significant role in many complex systems because of their capability to model the vagueness and ambiguity data. A fuzzy set F is a subset of the universe of discourse U , represented as:

$$F = \{ (x, \mu_F(x)), x \in U \},$$

where $\mu_F(\cdot)$ is a membership function which gives to each $x \in U$ a degree of belongingness from $[0, 1]$. When $\mu_F(\cdot)$ takes a value only in $\{0, 1\}$, F reduces to a crisp set and $\mu_F(\cdot)$ represents the characteristic function of the set F .

Measure of fuzziness is an intrinsic property that estimates the average ambiguity in a fuzzy set. In the literature, several definitions have been proposed. In general, a measure of fuzziness is a function

$$h : \mathfrak{F}(U) \longrightarrow \mathbb{R}^+,$$

where $\mathfrak{F}(U)$ denotes the set of all fuzzy subsets of U . This function is required to satisfy the following conditions [9]:

1. $\forall F \subset U$, if $h(\mu_F) = 0$ then F is a crisp set in U .
2. $h(\mu_F)$ assumes a unique maximum if $\forall x \subset U, \mu_F(x) = 0.5$.

3. $\forall(\mu_F, \mu_{F'}) \in U^2, f(\mu_F) \geq f(\mu_{F'})$ if F' is "crisper" than F ,
 i.e., $\forall x \in U \begin{cases} \mu_G(x) \geq \mu_F(x) & \text{if } \mu_F(x) \geq 0.5 \\ \mu_G(x) \leq \mu_F(x) & \text{if } \mu_F(x) \leq 0.5 \end{cases}$
4. $\forall F \subset U, h(\mu_F) = h(\mu_{\bar{F}})$ where \bar{F} is the complement of F .

For each fuzzy set F , this function assigns a nonnegative real number $h(F)$ that expresses the degree to which the boundary of F is not sharp. De Luca and Termini [9] borrowed the concept of Shannon's [10] information theoretic entropy to define a fuzzy entropy satisfying the above properties. This fuzzy entropy is given by:

$$H_{DTE}(\mu_F) = K \sum_{i=1}^n S(\mu_F(x_i)), \tag{1}$$

where $S(x) = -x \log(x) - (1 - x) \log(1 - x)$ and K is a normalizing constant. Many other measures of fuzziness have been suggested, with similar properties, but fuzzy entropy is still one of the most interesting and important measures of fuzziness in a fuzzy set.

2.2 Score Fusion in Multibiometric Identification System

When a biometric system operates in the closed-set identification, the output of the system is always a non-empty candidate list. Therefore, our goal is to determine the true identity of the given query.

Let M denote the number of biometric modalities in the multibiometric system and N be the number of users enrolled in the system. Each of these users has his own reference model. To simplify the notation, we assume that there is only a single reference model associated with each user for each biometric modality. A way to work at the score level is the use a score matrix which containing the match scores output by each biometric matchers. Let s_m^n denote the generic match score output by the m^{th} biometric matcher for n^{th} user in the database, $m = 1, 2, \dots, M; n = 1, 2, \dots, N$. For a given query, we can get a $M \times N$ score matrix \mathbf{S} defined as:

$$\mathbf{S} = \begin{pmatrix} s_1^1 & \dots & s_1^N \\ \vdots & \ddots & \vdots \\ s_M^1 & \dots & s_M^N \end{pmatrix} = \begin{pmatrix} S_1 \\ \vdots \\ S_M \end{pmatrix},$$

where $S_m = \{s_m^1, s_m^2, \dots, s_m^N\}$, for $m = 1, 2, \dots, M$.

At the score level fusion, a normalization step is generally necessary before the match scores from different matchers can be combined, because the matching scores output by various modalities are heterogeneous. In general, the normalized score is obtained by using a normalization function that maps S_m to \bar{S}_m defined as:

$$\begin{aligned} \mu : S_m &\longrightarrow \bar{S}_m \subset [0, 1] \\ s_m^n &\longmapsto \mu(s_m^n) \end{aligned},$$

where S_m is the set of all the raw output values of the corresponding matcher m and \bar{S}_m is the set of normalized matching scores of the matcher m . Therefore, the normalized score matrix $\bar{\mathbf{S}}$ is defined as:

$$\bar{\mathbf{S}} = (\mu(s_m^n))_{M \times N} = (\bar{S}_1, \bar{S}_2, \dots, \bar{S}_M)^\top.$$

If we replace the normalization function by the membership function using the fuzzy set theory, we can consider the set $\bar{S}_m = \{\mu(s_m^1), \mu(s_m^2), \dots, \mu(s_m^N)\}$ as a fuzzy set in S_m , and we can write it in the following way:

$$\bar{S}_m = \{(s_m^n, \mu_{S_m}(s_m^n)), s_m^n \in S_m\},$$

where $\mu_{S_m}(s_m^n) \in [0, 1]$ is the grade of the membership function of s_m^n in \bar{S}_m . Therefore, for a given query, the output of each biometric matcher is then modeled as a fuzzy set. In other words, the outcome of the system can be viewed as the fusion of the different membership functions provided by the different biometric matchers. In this study each biometric matcher m is represented by piecewise-linear membership function defined as:

$$\mu_{S_m}(s_m^n) = \begin{cases} 1 & \text{if } s_m^n < \alpha_1, \\ (\alpha_2 - s_m^n)/(\alpha_2 - \alpha_1) & \text{if } \alpha_1 < s_m^n < \alpha_2, \\ 0 & \text{if } s_m^n > \alpha_2, \end{cases}$$

where α_1 and α_2 are the minimum value of the impostor and the maximum value of the genuine score distributions, respectively.

2.3 Proposed Fusion Approach

All matching scores provided by a biometric matcher is represented as a fuzzy set. We assume that when a biometric matcher provides a reliable result, this set will be close to a binary set. On the contrary, when the biometric matcher is unreliable no opportunity should be significantly higher than the others. In term of fuzziness degree, the set constructed with a reliable biometric matcher will have a low degree contrary to a set constructed with an unreliable biometric matcher. Based on measure of fuzziness we define the weights

$$w_i = \frac{\sum_{k=0, k \neq i}^m H_{DTE}(\mu_{S_k})}{(m - 1) \sum_{k=0}^m H_{DTE}(\mu_{S_k})} \tag{2}$$

where $H_{DTE}(\mu_{S_k})$ is the fuzziness degree of biometric matcher k defined in (II), and m is the number of biometric modalities. When a source has a low fuzziness degree, w_i is close to 1 and it only slightly affects corresponding fuzzy set.

For every person (enrolled in the database) submits a biometric query samples, m fuzzy sets are computed, one for each modality. Then each fuzzy set will be

weighted by w_i defined in (2). This set of fuzzy sets constitutes the input for the fusion process:

$$\mathbf{D} = \{w_1\bar{S}_1, w_2\bar{S}_2, \dots, w_M\bar{S}_M\}.$$

Based on fuzzy set theory, the query should be assigned to the identity I_{n_0} if

$$n_0 = \arg \max_{n=1}^N \bigcup_{m=1}^M \{w_m \mu_{S_m}\},$$

where $\bigcup_{m=1}^M \{w_m \mu_{S_m}\} = \max_{m=1}^M \{w_m \mu_{S_m}\}$.

3 Multimodal Biometric System

In this paper, the face and iris biometric traits are selected to construct multimodal biometric system, because face recognition is most natural and acceptable in identity authentication whereas iris recognition is one of the most accurate biometrics.

3.1 Face Recognition

Among various face recognition algorithms, the most popular are appearance based approaches. Here we use the Steerable Pyramid (S-P) wavelet transform [11]. The first task in face recognition is an initial alignment, all images are aligned with respect to the manually detected eye coordinates, scaled and histogram equalized. After this stage, each sample face image preprocessed is described by a subset of band filtered images containing steerable pyramid coefficients which characterize the face textures. We divide the S-P sub-bands into small sub-blocks, from which we extract compact and meaningful feature vectors using simple statistical measures. For recognition, the city-block distance is used to measure the dissimilarity of different feature vectors.

3.2 Iris Recognition

Iris recognition is receiving increased attention due to its high reliability [12]. The iris recognition system employed in this paper is mainly based on the open source code provided by Libor Masek [13]. The human iris is an annular region between the black pupil and the white sclera. The first stage of iris recognition is to isolate the boundaries between the pupil and iris region and between sclera and iris region using a method that based on canny edge detection and circular Hough Transform [14]. The iris region not only contains the region of interest, but also some unuseful parts such as eyelid, eyelash and pupil. The eyelashes and eyelids occlude the upper and lower parts of the iris region are marked so that they will not be taken into account in the comparison stage. A noise mask

was generated to record the position of the occluded region. After localizing the region of interest, the Rubber Sheet Model [12] is used for unwrapped the iris to a rectangular region with prefixed size where the iris texture is analyzed. Then, a template representing iris pattern information is created using a 1D log-Gabor wavelets. The created templates were compared using the Hamming distance [12].

4 Experimental Results

To evaluate the robustness and the performance of our approach, we used a multimodal database containing face and iris samples. Considering the independence between two biometric traits, our database is constructed by concatenating FERET face database [15] and CASIA iris database (version 1) [16]. Finally, The obtained virtual multimodal database contains 756 records corresponding to 108 subjects (7 records for each subject), and each record represents a face sample and an iris sample.

In the following experiments, the whole database is divided randomly into two data sets, 128 records to construct the training data and the remaining records are used for testing. The training dataset was used to get the distributions of the unimodal matching scores in order to estimate the parameters of the fuzzy membership function. For the testing dataset, we use it to evaluate the performance of multimodal system. In the closed-set identification system the cumulative match characteristic (CMC) is widely used as an accuracy performance measure. The CMC shows the ratio of correct identification versus the matching rank.

The goal of the multimodal biometric is to increase precision and enhance reliability than unimodal biometrics. In order to demonstrate the effectiveness of our proposed method, figure 1 shows the CMC curves of the individual face and iris matchers and for the proposed method. These results clearly show that our multimodal system provides a significantly better rate than the individual monomodal system.

The second experiment consist to evaluate our fusion system with different score fusion methods such as sum, product, min and max rules. The recognition accuracy with min-max normalization are presented in table 1. According to table 1, we note that the proposed method outperforms the sum rule, which is considered to be one of the best combination approaches at score level fusion [5].

Table 1. Recognition accuracy: proposed method vs. combination approaches

Sum rule	Product rule	Max rule	Min rule	Proposed method
96.76%	96.30%	92.13%	91.67%	97.69%

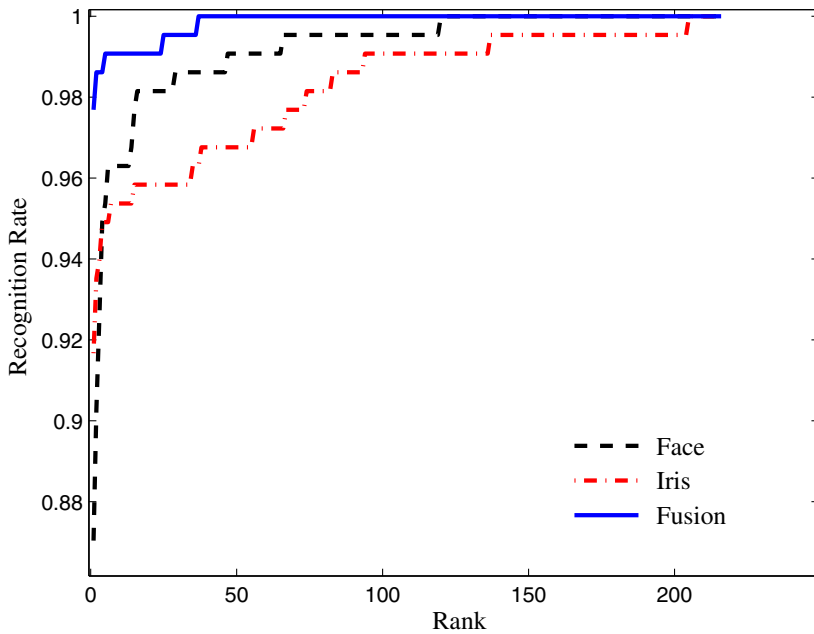


Fig. 1. Cumulative Match Characteristic (CMC) curves: proposed method vs. unimodal systems

5 Conclusion

In this paper, we proposed a general framework of multibiometric identification system based on fusion at matching score level using fuzzy set theory. In this approach, the output of each biometric matcher is modeled as a fuzzy set, and the corresponding degree of fuzziness estimates the reliability of the information provided by each biometric matcher. Then, the results are aggregated with a fuzzy combination rule. In order to prove the efficiency of our fusion approach, we have used a virtual multimodal database which integrates face and iris biometric modalities. Experimental results exhibit that our fusion approach achieves better performance than the best unimodal system and the classical combination approaches at score level fusion. However, the system needs to be tested on a large database.

References

1. Jain, A.K., Flynn, P., Ross, A.A.: Hand book of Biometrics. Springer (2008)
2. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(1), 4–20 (2004)

3. Nandakumar, K., Jain, A.K., Ross, A.: Fusion in Multibiometric Identification Systems: What about the Missing Data? In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 743–752. Springer, Heidelberg (2009)
4. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. *Pattern Recognition* 38(12), 2270–2285 (2005)
5. Ross, A.A., Nandakumar, K., Jain, A.K.: *Handbook of Multibiometrics* (International Series on Biometrics). Springer-Verlag New York, Inc., Secaucus (2006)
6. Ross, A., Jain, A.: Information fusion in biometrics. *Pattern Recognition Letters* 24, 2115–2125 (2003)
7. Nandakumar, K., Chen, Y., Dass, S.C., Jain, A.K.: Likelihood Ratio Based Biometric Score Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), 342–347 (2008)
8. Zadeh, L.A.: Fuzzy sets. *Information and Control* 8(3), 338–353 (1965)
9. De Luca, A., Termini, S.: A definition of a nonprobabilistic entropy in the setting of fuzzy entropy. *Inform. Control* 20, 301–312 (1972)
10. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. Jr.* 379–423, 621–659 (1946)
11. El Aroussi, M., El Hassouni, M., Ghouzali, S., Rziza, M., Aboutajdine, D.: Local appearance based face recognition method using block based steerable pyramid transform. *Signal Processing* (2010)
12. Daugman, J.G.: How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology* 14(1), 21–30 (2002)
13. Masek, L.: Recognition of human iris patterns for biometric identification. Bachelor of Engineering Degree Thesis, The University of Western Australia, Australia (2003)
14. Wildes, R.P.: Iris recognition: an emerging biometrics technology. *Proc. IEEE* 85, 1348–1363 (1997)
15. Phillips, P.J., nad Moon, H., Rauss, P.J., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22, 891–906 (2000)
16. Institute of Automation, Chinese academy of Science, CASIA Iris Image Database, <http://www.sinobiometrics.com/chinese/chinese.htm> (retrieved on April 2010)

Security Analysis of Key Binding Biometric Cryptosystems

Maryam Lafkih¹, Mounia Mikram^{1,2}, Sanaa Ghouzali^{1,3}, and Mohamed El Haziti⁴

¹ LRIT, Faculty of Sciences, Mohammed V University, Rabat, Morocco

² The School of Information Sciences, Rabat, Morocco

³ College of Computer and Information Sciences,
King Saud University, Riyadh, Saudi Arabia

⁴ Higher School of Technology, Sale, Morocco

maryam.lafkih@gmail.com

Abstract. The use of biometric systems is becoming an important solution to replace traditional authentication. However, biometric systems are vulnerable to attacks. When biometric data is compromised, unlike a password, it can't be changed. Therefore, the security of biometrics models is essential in designing an authentication system. To achieve this protection of biometric models, two categories of approaches are proposed in the literature, namely, methods based on transformation of characteristics and biometric cryptosystems. For the first type of approaches, a study is made to assess the security of biometric systems. In biometric cryptosystems the realized works are hampered by the lack of formal security analysis. Hence the purpose of this paper is to propose standard criteria for a formal security analysis of biometric cryptosystems. The proposed measures take into account the specific effect of key binding cryptosystems. The security analysis is illustrated by experiments on the techniques of *Fuzzy Commitment* and *Fuzzy Vault* which we use in this work for the protection of biometric face recognition system. Our analysis indicates that both techniques are vulnerable to intrusion and binding attacks because of the ease of obtaining the user's model using the elements known to the attacker.

Keywords: Security analysis, Biometric cryptosystems, Performance evaluation, Models transformation.

1 Introduction

Today, the need for security systems is becoming a necessity in the world. To better meet this need, biometrics is presented as a real alternative to passwords and other identifiers. It ensures that the user is who he claims to be, thereby reducing the risk of theft, loss or forgetfulness. However, biometric systems are not protected against attacks and a template stored in a database can be stolen by an attacker for an illegitimate access. This would mean that legitimate users should not be able to use the compromised model to authenticate [1]. To overcome this problem, one idea would be to secure biometric authentication scheme.

In the literature there are two types of methods to protect biometric templates: Methods based on the transformation of biometric features and biometric cryptosystems [2]. The first type of methods consists on applying a transform function on the biometric characteristics to build a model that will be stored in the database (enrollment phase). During authentication, the same function is applied to the biometric characteristics of the query template to obtain a model that is then compared to the stored reference model to allow or deny the access [2]. Biometric cryptosystems use a secret key to wrap the biometric characteristics and generate an auxiliary data that will be stored in the database (enrollment phase). In the authentication phase the secret key must be extracted from the biometric characteristics of the query and the auxiliary data stored [3].

However, these biometric technologies include several components that have weaknesses and limitations such as high cost, risk of tampering and poor performance. To this end a performance evaluation is a necessity for comparing different systems. In the case of characteristics transformation methods, a study is made by Nagar et al. [4] for the security evaluation of biometrics systems. Although cryptosystems are used in the real world (e.g. smart cards) [5], their practical applicability is hampered by the lack of a formal security analysis. Thus, the objective of this work is to propose a set of standard criteria to evaluate the overall security of biometric cryptosystems.

In the rest of this article, biometric cryptosystems are described in Section 2, and then in Section 3 the analysis of the security of cryptosystems is detailed. In Section 4, experimental results illustrate how the proposed measures can be used to evaluate biometric cryptosystems. Conclusion and perspectives are drawn in Section 5.

2 Biometric Cryptosystems

Biometric cryptosystems are techniques that aim to integrate the benefits of using a secret key (encryption) and biometric features in a security system [6] [7]. A several approaches developed in the field of biometric cryptosystems are based on two modes of generation of the secret key. Thus, we can distinguish between two types of cryptosystems along the two modes (1) *Key binding biometric cryptosystems* [5]: where the biometric template is linked with a secret key in a single entity to build an auxiliary data. This data reveal no information on the key or the biometric template and (2) *Key generation biometric cryptosystems* [8]: where the key is derived directly from the biometric data. Authentication is successful if the key is retrieved.

In the literature there are two main approaches to perform key binding biometric cryptosystems: *Fuzzy Commitment* and *Fuzzy Vault* [9]. The first approach, proposed by Juels and Wattenberg [10], consists on using biometric features and secret key to generate a helper data. The pair that contains the helper data and the secret key encrypted is then stored in the database. In the authentication phase the key must be regenerated from the helper data stored in the database and biometric features of the query template. The second approach, proposed by Juels and Sudan [11], aims to generate a polynomial p from a secret code and biometric characteristic and then add false points to construct a Vault V that will be stored in the database. During

authentication it must find the secret code from the Vault stored and the biometric features of the query in order to succeed the access.

However, these biometric cryptosystems include several components that have gaps and limitations such as the high cost, risk of falsification and poor performance. To this end, a performance evaluation is a prerequisite for comparison between different biometric systems. To ensure the security and protection against the risks associated with these systems, the security analysis of biometric cryptosystems consists of measuring these risks according to the probability or frequency of their appearance and their possible effects. Nagar and al [4] have made a study for security analysis of biometric systems. In the case of biometric cryptosystems the studies are made specifically for each approach; there is no formal analysis to analyze the security of all biometric cryptosystems. In next section we propose a set of generalized criteria to evaluate the overall security of biometric cryptosystems.

3 Security Analysis of Biometrics Cryptosystems

The security analysis plays an important role to evaluate the performance of biometric systems; it can test several components such as the ease of the system, the security ... etc. To analyze the security of biometric cryptosystems, we focused on vulnerability to intrusion attacks and binding attacks. The term “*Intrusion*” is the access to a biometric system by submitting fake authentication data for the system. “*Binding*” attacks involve the mapping of multiple biometric models generated from different encryption parameters to find the original model. To cope with these attacks, it is important to analyze the probability of their success in a cryptosystem.

To describe the security measures of biometric cryptosystems, we used the following notation: X_U and X_U' represent the model of the user and biometric characteristics of the request for the same user, X_A the biometric characteristics of the attacker, H is the auxiliary data, K_U and K_U' are two different keys of the user and K_A is the key of the attacker. D_O (respectively D_E) is a function of distance between the original models (respectively auxiliary data in encrypted domain). The user will be accepted by the system if the distance between the model and biometric characteristics of the query is below a threshold ϵ .

We have proposed criteria for (1) measure of the usability of the system, (2) measure of the security for intrusion threats evaluation and (3) measure of the security for binding threats evaluation.

3.1 Measure of the Usability of a System

Measuring the usability of a system is made in terms of False Rejection Rate “*FRR*”. The *FRR* is the percentage of the users rejected by the system out of the total number of users in the database [12]. Therefore we distinguish two cases; before encryption and after encryption.

The False Rejection Rate of the biometric system in Original domain i.e. before the encryption, “*FRR_O*” is expressed by the probability that the distance between the biometric characteristics of the user X_U and the biometric characteristics of the request X_U' is greater than or equal to the threshold ϵ .

$$FRR_O(\epsilon) = P(D_O(X_U, X_U') \geq \epsilon) \quad (1)$$

The *False Rejection Rate of the biometric system after the application of encryption*, i.e. False Reject Rate in encrypted domain, " FRR_E " is expressed by the probability that the distance between the helper data of the user and the helper data of the request is greater than or equal to the threshold ϵ as given by the following equation:

$$FRR_E(\epsilon) = P(DE(H_U(X_U, K_U), H_U(X_U', K_U)) \geq \epsilon) \quad (2)$$

3.2 Measure of the Security of Intrusion Threats

The measure of the security of intrusion threats is defined as the probability of a successful attack, assuming that the model stored in the database and encryption parameters are available to the attacker attempting to usurp the identity of a trusted user. The probability of successful intrusion threats is given by the *False Acceptance Rate "FAR"*. The *FAR* gives the percentage of accepted attackers among the number of attackers who come to the system [12]. We have proposed criteria for both cases; before encryption and after the encryption.

False Acceptance Rate of original biometric system before encryption "FAR_O" is given by the probability that the distance between the biometric characteristics of the user X_U and the biometric characteristics of the attacker X_A is lower than the threshold ϵ as it is illustrated in the following equation:

$$FAR_O(\epsilon) = P(D_O(X_U, X_A) < \epsilon) \quad (3)$$

For the case after encryption, the attacker is required to submit biometric characteristics with a set of encryption parameters for authentication. Therefore, there are two possibilities; the case where the parameters are '*Unknown*' to the attacker and the case where the parameters are '*known*' to the attacker. Suppose that the attacker doesn't know the encryption parameters for the specific user. We calculate in this case the *False Acceptance Rate with Unknown encryption parameters "FAR_{UP}"* given by the following equation which expresses the probability that the distance between the helper data of the user and the helper data of the attacker generated by its own key K_A is lower than the threshold ϵ .

$$FAR_{UP}(\epsilon) = P(DE(H_U(X_U, K_U), H_A(X_A, K_A)) < \epsilon) \quad (4)$$

If the attacker knows the encryption parameters of the user, the *False Acceptance Rate with Known encryption parameters "FAR_{KP}"* is defined by the probability that the distance between the helper data of the user and the helper data of the attacker generated by the same key of the user K_U is lower than the threshold ϵ , as indicated in the following equation:

$$FAR_{KP}(\epsilon) = P(DE(H_U(X_U, K_U), H_A(X_A, K_U)) < \epsilon) \quad (5)$$

In addition to the False Acceptance Rate, we considered other probabilities of intrusion after encryption where the stored model and the encryption parameters are

available to the attacker to gain an illegitimate access to a ‘*Different*’ biometric system which uses the same biometric characteristics. Suppose that the attacker knows also the encryption parameters of the second system. In this case, he will try to retrieve the biometric model using the model encrypted and the encryption parameters of the second system. The probability of success of such attack is called the *Cryptosystem Intrusion Rate of Different system with known Parameters* “ $CIRD_{KP}$ ” and is defined by the Equation 7 that expresses the probability that the distance between the helper data of the user stored in the second system H_U^{S2} and the helper data of the attacker H_A (generated by the feature X'_U estimated using the two keys of the user (the key of the first system and the key of the second system) and the helper data of the user stored in the first system) is lower than the threshold ϵ :

$$CIRD_{KP}(\epsilon) = P(D_E (H_U^{S2}(X_U, K_U), H_A (X'_U, K'_U)) < \epsilon) \tag{6}$$

If the attacker knows the helper data and the encryption parameters of the user in the first system without knowing the encryption parameters of the second system, the attack performed in this case is called the *Cryptosystem Intrusion Rate of Different system with Unknown Parameters* ‘ $CIRD_{UP}$ ’. The success of this attack can be expressed by the probability that the distance between the helper data of the user stored in the second system H_U^{S2} and the helper data of the attacker H_A (generated by the feature X'_U , estimated using the key of the user in the first system and both helper data of the user stored in the first and the second systems, and his key K_A) is lower than the threshold ϵ as specified by

$$CIRD_{UP}(\epsilon) = P(D_E (H_U^{S2}(X_U, K_U), H_A(X'_U, K_A)) < \epsilon) \tag{7}$$

3.3 Measure of the Security of Binding Threats

The measure of the security for the evaluation of binding attacks is defined as the probability of a successful attack to link different models of the same biometric trait of the user and different parameters encryption. Suppose that the two sets of encryption parameters are known to the attacker. The *Cross Rate in the Encrypted fields* “ CR_E ” can be defined by the probability that the distance between the helper data of the user in the first system H_U^{S1} and the helper data of the user in the second system H_U^{S2} is lower than the threshold (ϵ equation 8):

$$CR_E(\epsilon) = P(D_E (H_U^{S2}(X_U, K_U), (H_U^{S1}(X'_U, K'_U)) < \epsilon) \tag{8}$$

Besides these attacks, we assume the case where the attacker will attempt to combine the helper data of the trusted user and his own helper data H_A (generated from his biometric data and his own key K_A) which we named the *Combination Attack*; ‘ CA ’. To illustrate this scenario we consider the following criterion which consists of the probability that the distance between the result of combination and the helper data of the user is lower than the threshold ϵ .

$$CA(\epsilon) = P(D_E (H_U(X_U, K_U), (H_U(X'_U, K_U) + H_A(X_A, K_A)) < \epsilon) \tag{9}$$

We also proposed another criterion, *Combination Attack in a ‘different’ system* “ CA_{diff} ”, in which we assume that the attacker has the encryption parameters and the helper data of the user in the first system and tries to have access to a second system. We expressed this criterion by the probability that the distance between the helper data of the user stored in the first system H_U^{S1} and the result of combination (of the helper data of the user in the first system and the helper data of the attacker generated by the key of the user K'_U) is lower than the threshold ϵ by

$$CA_{diff}(\epsilon) = P(D_E(H_U^{S2}(X_U, K_U), (H_U^{S1}(X'_U, K'_U) + H_A(X_A, K'_U))) < \epsilon) \quad (10)$$

4 Experiments

In order to evaluate the proposed security analysis framework of biometric cryptosystems, we considered the example of biometric systems based on face recognition. Thus, we need two biometric systems using two different methods for extracting features of the face images and a technique to protect the authentication scheme.

4.1 Experimental Settings

At first we created two biometric systems, the first biometric system is based on “*Laplacian Smoothing Transform, LST*” [13] method used for feature extraction followed by “*Linear Discriminant Analysis, LDA*” [14]. The second biometric system [15] uses *LST* for feature extraction followed by Support Vector-Discriminant Analysis (*SVDA*) technique [16] for dimensionality reduction. We evaluated the performance of biometric systems using the *YALE face* database [17] separated into training and test subsets. Then we calculated the *Hamming distance* between the user of the test and the reference for matching. In a second step we used the *Key binding* biometric cryptosystems (*Fuzzy Commitment* and *Fuzzy Vault*) to secure the two biometric systems. We used *Reed Solomon error correcting code* that allows recovering the data even in the case of error transmission [18]. The hash function *SHA-1* [19] has been used in this work to encrypt the secret key in *Fuzzy Commitment* scheme. In a third step we applied the criteria proposed in Section 3 to analyze the security of these systems.

To evaluate the performance of biometric systems, there are several important components to test such as the system reliability and performance. We measured the performance of biometric systems using a false acceptance rate set correspondence with a false rejection rate. To view the performances of biometric systems when the threshold varies, we used the *ROC* (Receiver Operating Characteristics) [20] curves representing the *FAR* from *1-FRR*. The “ ROC_{orig} ” (FAR_o from $1-FRR_o$) curve presents the original system i.e. before encryption, the system after encryption in case the attacker knows the encryption parameters and the case where the attacker does not know the encryption parameters is presented by “ $ROC_{Unknown}$ ” (FAR_{UP} from $1-FRR_E$).

4.2 Security Analysis Results of Fuzzy Commitment Technique

Figure 1 (a) shows the different ROC curves of the first system. We notice performance degradation compared to the original model in case the encryption parameters are unknown and increased degradation in the case that the attacker knows the encryption parameters as indicated by the curve ROC_{known} . As it is shown, the original system ROC_{orig} is better than the system after encryption, in case where the attacker has just his biometric traits and attempts to gain illegitimate access to the system as expressed in $ROC_{Unknown}$ curve, we notice that the system accepts up to 47% of the attackers and in the case where the attacker knows also the encryption parameters of the system we notice less performance compared to the previous scenario (as indicated by the ROC_{known} curve).

In the second system represented by Figure 1 (b), we note that there is always a degradation of performance compared to the original model in the case of intrusion with unknown parameters as shown by the $ROC_{Unknown}$ curve, and the degradation of the performance increases if the attacker knows the encryption parameters as it is indicated by the curve ROC_{known} . ROC_{orig} curve shows also that the original system is better than the system after encryption; we note that the system accepts a maximum of 11.36% of the attackers in the case of unknown parameters ($ROC_{Unknown}$).

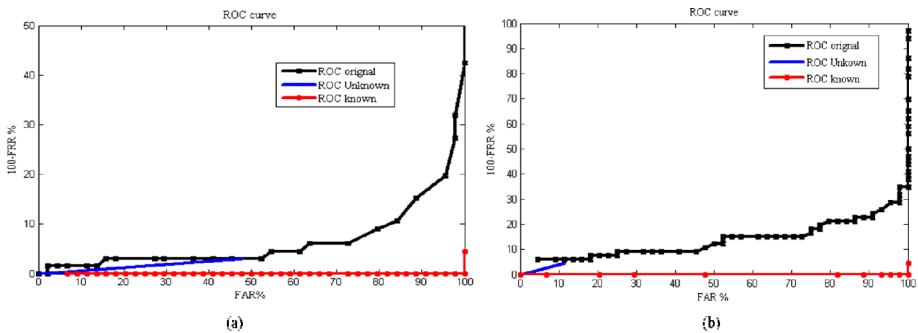


Fig. 1. ROC_{orig} , $ROC_{Unknown}$, ROC_{known} curve of the first system (a) and the second system (b)

As comparison of the two systems, the second system is more efficient than the first; this performance can be explained by the use of the $SVDA$ method which gives better results than LDA [14].

For intrusion attacks we also evaluated the measures of the criteria, $CIRD_{KP}$, CR_E and $CIRD_{UP}$.

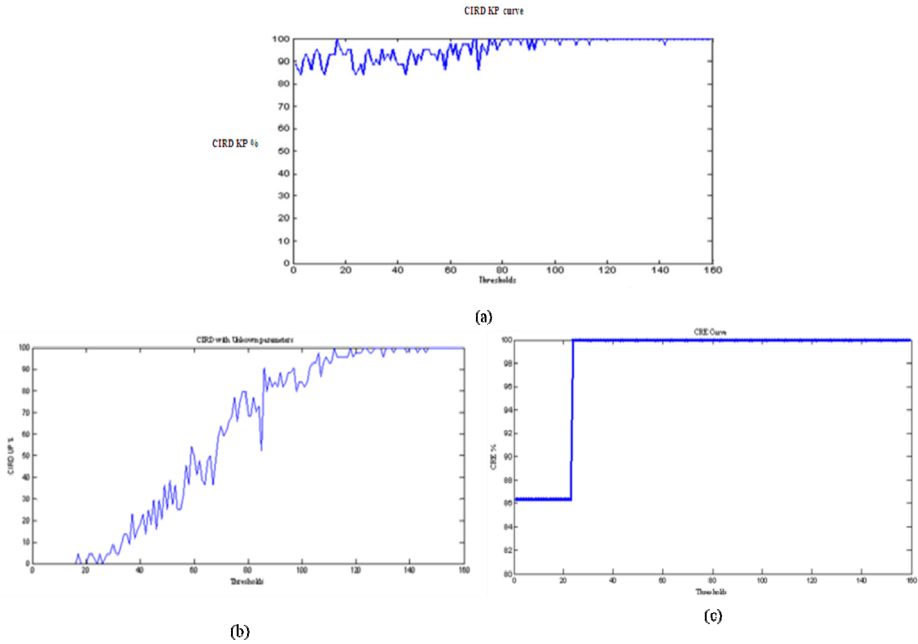


Fig. 2. $CIRD_{KP}$, CR_E , $CIRD_{UP}$ curves of *Fuzzy Commitment*

Figure 2 (a) shows the representation of $CIRD_{KP}$ for different thresholds. This figure shows the possibility of success of the attacker to access a ‘different’ system that uses the same biometric traits of the user. We note that the probability of success of the attacker is changed if the threshold is less than or equal to 80, due to the intra-class variation between the user and the attacker. This variation prevents the attacker to gain 100% access to the system. As it is shown in the curve, the minimum probability that an attacker can access to the system is equal to 86%. So even with the intra-class variation, the probability that an attacker succeeds to access the system remains high. For a threshold above 80, the value of $CIRD_{KP}$ increases up to 100% which means more vulnerability of the system. The probability of success of such attack is higher if the attacker knows the helper data stored in the database of the first system and the encryption parameters on both systems. The attacker tries to generate a helper data $H_U'(X_U, K_U)$ from the data of the first system $H_U^{S1}(X_U, K_U)$ and the two code words c_U^{S1} and c_U^{S2} of the two systems as given by the following equation.

$$H_U'(X_U, K_U) = (H_U^{S1}(X_U, K_U) + c_U^{S1}) - c_U^{S2} = X_U - c_U^{S2} \tag{11}$$

We can conclude that the method of *Fuzzy Commitment* is vulnerable to intrusion attacks. If the attacker knows the encryption parameters and the model stored in the system (represented by $CIRS$) then the probability that he may have access to the system is of 100%.

In the case where the attacker wants to access to another system that uses the same biometric features and has the helper data of the first system and the encryption parameters of the first and second systems (represented by $CIRD_{KP}$), protection with *Fuzzy Commitment* is not guaranteed against this type of attacks. Only the intra-class variation can decrease the access probability of the attacker, but from a certain threshold the attacker can access with a probability of 100%.

The Figure 2 (b) shows the *CIRD* with *Unknown Parameters*. We note that the attacker cannot access to the system if the threshold is below 17; the rate of intrusion increases with variation according to thresholds and equal to 100% when the threshold is greater than 140.

Figure 2 (c) shows the representation of ‘*Cross Rate in Encrypted domain*’ CR_E according to thresholds. For threshold values less than or equal to 38, the cross rate is equal to 86.36%. For other values of the threshold (above 38) the success rate of this attack is increased to 100%. In this type of attacks, the rate of vulnerability is due to the knowledge of two helper data by the attacker, which makes easy the connection in encrypted domains by just matching the different helper data. The vulnerability of *Fuzzy Commitment* according to the proposed scenario can be explained by the ease of obtaining the original model from the elements known by the attacker namely the helper data and the encryption parameters.

In ‘*Combination Attack*’ *CA* as shown in Figure 3 (a), the attacker and the user use the same model to authenticate. The acceptance rate of the attacker may be 0% for certain thresholds such as the range of thresholds [0, 20]. The rate of access to the system by the attacker is increased with variation because he uses the same record as the user. The maximum value of the vulnerability of the system is reached in the 119 threshold for a rate of attack equal 25%.

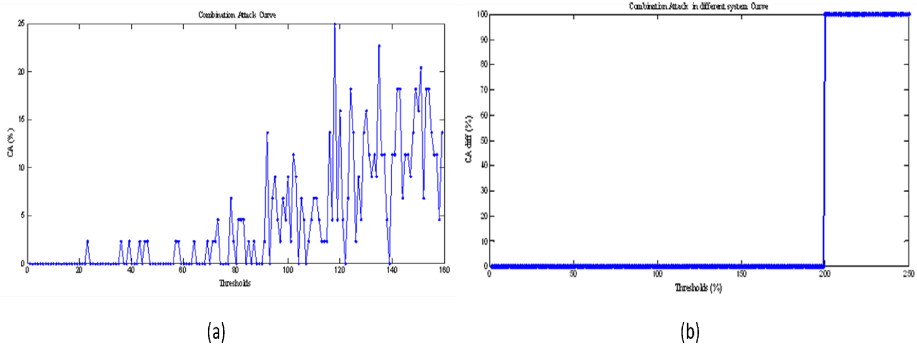


Fig. 3. CA , CA_{diff} curves of *Fuzzy Commitment*

In the case of ‘*Combination Attack in Different system*’ illustrated by Figure 3(b), the attacker has a helper data generated by these biometric data and the encryption parameters of the user, then he makes a combination with the auxiliary data of the user and tries to attack a second system that uses the same biometric trait of the user. We notice that the attacker does not have access to the system for thresholds below 199, after

this threshold the value of vulnerability is increased to 100% because the attacker uses the key of the user. *Fuzzy Commitment* is more vulnerable to this attack, where several helper data generated from the same biometric trait can be adapted by the attacker to extract the original biometric model, and thus the ability of the revocation is affected.

4.3 Security Analysis Results of Fuzzy Vault Technique

After applying the proposed criteria on the method of “*Fuzzy commitment*”, we analyzed the security of the second method i.e. “*Fuzzy Vault*” using same criteria.

Figure 4 shows the ROC curves before encryption ROC_{orig} , after encryption $ROC_{Unknown}$ curve and ROC_{Known} where the attacker knows the encryption parameters. We notice a less performance than the original system ROC_{orig} and degradation of performances if the attacker knows the encryption parameters. In case where the attacker does not have the encryption parameters, the possibility to be accepted is varied. If the encryption parameters are known to the attacker, the possibility of acceptance is 100% (ROC_{known}) while the acceptance in case of unknown parameters $ROC_{Unknown}$ is less than 100%. This vulnerability is due to the knowledge of the polynomial p by the attacker where the possibility of having an illegitimate access to the system of 100%.

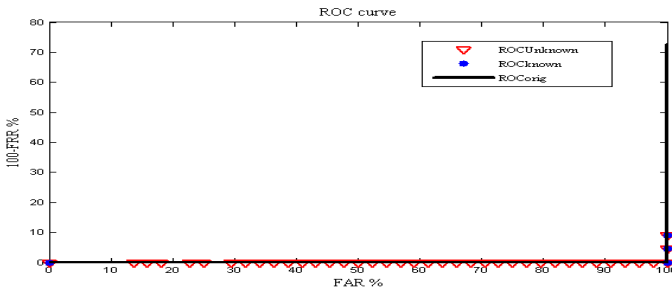


Fig. 4. ROC_{orig} , $ROC_{Unknown}$, ROC_{known} curves of *Fuzzy Vault*

Figure 5 (a) shows the ‘*Cryptosystem Intrusion Rate in a Different system with Known Parameters*’ $CIRD_{KP}$. In this scenario, the system is vulnerable after the threshold 3070 because the attacker knows the polynomial $p1$ and $p2$ of two biometric systems and also knows the Vault $V1$ stored in the first system. He has all the elements allowing to find the model X_U used to estimate the Vault $V2$ of the second system and hence has an illegitimate access to the system using the Equations 12 and 13.

$$X'_U = \text{Racine} (V_U^{S1} - FP) = \text{Racine} (\text{Projection} (p1, X_U)) \tag{12}$$

$$V'_U^{S2} = \text{Projection} (p_2, X'_U) + FP' \tag{13}$$

The attacker is rejected by the system up to the threshold 3070 due to the intra-class variation and the false points as well.

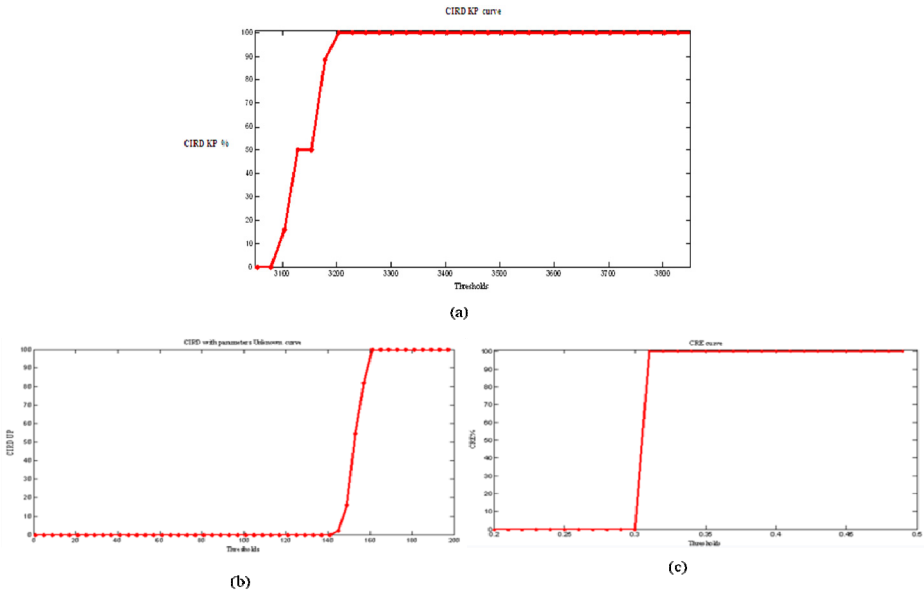


Fig. 5. $CIRD_{KP}$, $CIRD_{UP}$, CR_E curves of *Fuzzy Vault*

Figure 5 (b) shows the ‘*Cryptosystem Intrusion Rate in a ‘Different’ system with Unknown parameters*’ $CIRD_{UP}$. We note an increase in vulnerability of the system after the threshold 140. This vulnerability is due to the knowledge of two Vaults of the two systems (the first system and the second system) and the encryption parameters of the first system, the attacker tries to find the original model using known elements according to Equations 14.

$$X'_U = \text{Racine}(V_U^{S1} - FP'_A) = \text{Racine}(\text{Projection}(p1, X_U^{S1})) = \text{Racine}(V_U^{S2} - FP'_A) \quad (14)$$

Figure 5 (c) shows the ‘*Cross Rate in Encrypted domain*’ CR_E according to the thresholds. For threshold values less than or equal to 0.3, the attacker cannot link the two Vaults (the cross rate is 0%). This result can be explained by the false point that can make a difference between the two Vaults. For other threshold values (greater than 0.3) the success rate of this attack is increased to 100% because the attacker knows the two polynomials and also the two Vaults. Then to make the correspondence in encrypted domain, the attacker can simply match the two Vaults following thresholds higher than 0.3.

We note that the method of *Fuzzy Vault* is vulnerable to attack from a certain threshold depending on the proposed scenarios; this vulnerability is due to the possibility of obtaining the original model from the information known to the attacker i.e. encryption parameters and stored Vault in the database.

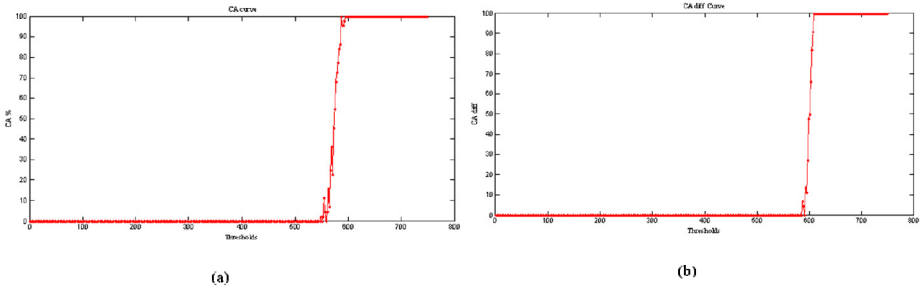


Fig. 6. CA ‘combination Attack’ and CA_{diff} ‘combination Attack in different system’ curve of the Fuzzy Vault

In CA attack (Figure 6 (a)), the attacker adds his Vault V_A generated with his key K_A to the Vault of the user V_U^{S1} stored in the first system S_1 (Equation 15) which can disrupt the system after certain thresholds. We notice that the attacker can access to the system after the threshold 550 and then the value of this attack increases according to the threshold up to 100 after the threshold 600.

$$V_A^{S1} = V_A(K_A) + V_U^{S1}(K_U^{S1}) \tag{15}$$

In CA_{diff} attack (Figure 6 (b)), the attacker adds the Vault of the user V_U^{S1} to his Vault V_A generated with the key of the user K_U^{S1} to attack a second system (Equation 16). The attacker can access the system after the threshold 585. The vulnerability of the attack increases up to 100% if the threshold exceeds 600.

$$V_A^{S2} = V_A(K_U^{S1}) + V_U^{S1}(K_U^{S1}) \tag{16}$$

In combination attack, the attacker has difficulty of access to the same system as illustrated by CA curve. This difficulty can be higher in case of attack in a second system that uses the same biometric features of the user knowing the key to the first system as shown in CA_{diff}.

5 Conclusion

Biometric cryptosystems are developed to protect the biometric models; however no study is conducted in this domain for a formal security analysis. In this paper, we have proposed different measures to assess the security strength of key binding biometric cryptosystems. We applied these criteria for the protection of a biometric facial recognition system. The emphasis here was on the security analysis, which was tested on *Fuzzy Commitment* and *Fuzzy Vault* techniques showing the interest of the proposed measures. Our analysis shows that both methods are vulnerable to ‘intrusion’ and ‘binding’ attacks especially if the attacker knows the encryption parameters in intrusion attacks and the helper data along with the encryption parameters in the cross attacks. Our experiments expressed that the method of *Fuzzy Commitment* is more vulnerable to proposed scenarios than *Fuzzy Vault*. This vulnerability can be explained by the ease of obtaining the original model from the auxiliary data and the encryption parameters. The proposed criteria allow evaluating the robustness of the biometric

cryptosystems (as shown for both techniques *Fuzzy Commitment* and *Fuzzy Vault*) and also make the difference between security and usability.

The experimental field in the future will be extended to include different parameters for the protection of biometric systems. As a future work, we plan to offer other attack scenarios.

References

1. Ratha, N.K., Connell, J.H., Bolle, R.M.: An Analysis of Minutiae Matching Strength. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 223–228. Springer, Heidelberg (2001)
2. Nagar, A.: Secure Biometric Recognition. In: PRIP Seminar (2008)
3. Nagar, A., Nandakumar, K., Jain, A.K.: A hybrid biometric cryptosystem for securing fingerprint minutiae models. Elsevier Pattern Recognition Letters (2010)
4. Nagar, A., Nandakumar, K., Jain, A.K.: Biometric Model Transformation: A Security analysis. SPIE (2010)
5. Jain, A.K., Nandakumar, K., Nagar, A.: Biometric Model Security. Eurasip Journal (2008)
6. Uludag, U., Pankanti, S., Prabhakar, S., Jain A.: Biometric cryptosystems: Issues and challenges, pp. 948–960. IEEE (2004)
7. Hao, F., Anderson, R., Daugman, J.: Combining crypto with biometrics effectively. IEEE Trans. Comput., 1081–1088 (2006)
8. Li, Q., Sutcu, Y., Memon, N.: Secure Sketch for Biometric Templates. In: Lai, X., Chen, K. (eds.) ASIACRYPT 2006. LNCS, vol. 4284, pp. 99–113. Springer, Heidelberg (2006)
9. Dodis, Y., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, pp. 523–540. Springer (2004)
10. Juels, A., Wattenberg, M.: A Fuzzy Commitment Scheme. In: Sixth ACM Conference on Computer and Communications Security, Singapore, pp. 28–36 (1999)
11. Juels, A., Sudan, M.: A Fuzzy Vault Scheme. In: IEEE International Symposium on Information Theory, Lausanne, Switzerland (2002)
12. Adair, K.L., Parthasaradhi, S.T.V., Kennedy, J.: Real World Evaluation: Avoiding Pitfalls of Fingerprint System Deployments. BiometricsIndia Expo. (2008)
13. Gu, S., Tan, Y., He, X.: Laplacian Smoothing Transform for Face Recognition, pp. 2415–2428. Springer (2010)
14. Khan, A., Farooq, H.: Principal Component Analysis-Linear Discriminant Analysis Feature Extractor for Pattern Recognition. International Journal of Computer Science Issues (IJCSI) 8(6) (2011)
15. Moujahdi, C., Ghouzali, S., Mikram, M., Abdul, W., Rziza, M.: Inter-communication classification for Multi-view Face Recognition. In: The 4th International Conference on Multimedia Computing and Systems (ICMCS), Tangier, Morocco (2012)
16. Gu, S., Tan, Y., He, X.: Discriminant Analysis via Support Vectors. Neurocomputing (2010)
17. Bellhumer, P.N., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Patt. Anal. and Mach. Intel. Special Issue on Face Recognition, 711–720 (1997)
18. MacWilliams, F.J., Sloane, N.J.A.: The Theory of Error-Correcting Codes. North Holland (1977)
19. Schneider, B.: Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2nd edn. Wiley, New York (1996)
20. Fawcet, T.: ROC Graphs: Notes and Practical Considerations for Researchers. HP Laboratories, 1143–1501 (2004)

Texture Analysis for Trabecular Bone X-Ray Images Using Anisotropic Morlet Wavelet and Rényi Entropy

Ahmed Salmi EL Boumnini El Hassani¹, Mohammed El Hassouni²,
Rachid Jennane³, Mohammed Rziza¹, and Eric Lespessailles⁴

¹ LRIT, Faculty of Science, University Mohammed V - Agdal - Rabat, Morocco

² DESTEC-FLSHR University Mohammed V - Agdal - Rabat, Morocco

³ PRISME Laboratory, University of Orléans, 12 rue de Blois, 45067 Orléans, France

⁴ Hospital of Orleans, IPROS, 1, rue Porte Madeleine, 45032 Orléans, France

Abstract. In this paper, we propose a new method based on texture analysis for the early diagnosis of bone disease such as osteoporosis. Our proposed method is based on a combination of four methods. First, bone X-ray images are enhanced using the algorithm of Retinex. Then, the enhanced images are analyzed using the fully anisotropic Morlet wavelet. This step is followed by the quantification of the anisotropy of the images using the Rényi entropy. Finally, the Rényi entropies are used as entries for a neural network. Applied on two different populations composed of osteoporotic (OP) patients and control (CT) subjects, a classification rate of 95% is achieved which provides a good discrimination between OP patients and CT subjects.

1 Introduction

Osteoporosis is considered as a major public health issue [1] due to an increase frequency of fractures of the hip, spine, and wrist. Osteoporosis is characterized by a severe degradation of the bone mass and an alteration of the bone microarchitecture. This problem is currently affecting more than 200 million people worldwide. Epidemiological studies provide a very significant increase in the number of osteoporotic fractures in the coming years [2]. Osteoporosis is clinically assessed by using BMD (Bone Mass Density). Despite the effectiveness of this technique, it does not give information about the microarchitecture of the bone tissue. If BMD is combined to an independent technique that describes the microarchitecture, this might enable a better and precise diagnosis [3] for the prediction of fracture risk. Obviously, it has to be non invasive for the patient, not expensive, reproducible and efficient. The calcaneus (the bone of the heel) is subject to forces of compression and tension produced by the gravity of the human being, making it very suitable for the characterization of the bone microarchitecture (Fig. 1). For a normal subject, the compression and tensile trabeculae are uniformly distributed. For an osteoporotic subject, the tensile trabeculae may disappear making the structure anisotropic. The modifications

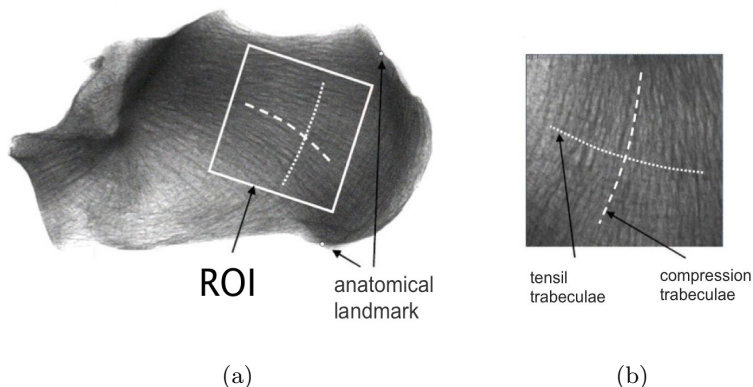


Fig. 1. (a) A typical radiography of the calcaneus with the Region Of Interest (ROI). (b) 256×256 extracted ROI.

of the organization of the trabeculae and their thickness can help evaluating the damage of the bone.

Several studies have attempted to evaluate osteoporosis to characterize the anisotropy of textured images. Sevestre *et al.* [4] developed a morphological study to establish a skeleton of the trabecular bone microarchitecture. Despite their quite interesting results, the tool is very complex to produce.

Many methods of texture analysis have been proposed over the last three decades [5,6]. These methods are evaluated over natural and textured surfaces which are quite distinctive for the human vision system. The texture present in osteoporotic and healthy bone radiographs, however, are visually close to each other, making the discrimination task very challenging. Other methods using fractal analysis for bone texture have been explored [7,8,9]. These methods gave interesting results but are still under investigation for an efficient characterization of bone texture organization. More recently, some of the authors of the present study [10] proposed a new descriptor called 1D LBP (One Dimensional Local Binary Pattern) for bone texture characterization. Results of this study demonstrated the importance of preprocessing the data to improve the classification rates to distinguish between CT and OP subjects. In the same way, Pramudito *et al.* [11] combines the coefficients of the wavelet and the fractal dimension to identify the disease. Their method offers a new perspective to analyze such kind of images.

In this work, we propose a method which enables characterizing the anisotropy of an image using the entropy of Rényi and a fully anisotropic Morlets. The Rényi entropy has shown its effectiveness especially to quantify the anisotropy [12]. The use of a fully anisotropic Morlet enables settling the problem of non-uniform changes.

This paper is organized as follows. Section 2, describes the methods used to characterize trabecular bone data on radiographs. Section 3 presents the experimental results obtained on two different populations composed of osteoporotic patients and control subjects. Finally, some concluding remarks are discussed in section 4.

2 Methods and Materials

Our goal is to study the effect of preprocessing the data of bone radiograph images for the diagnosis of osteoporosis. Different methods are considered. First, images are enhanced. Then, the fully anisotropic Morlet wavelet is used to analyze the images. After computing the two-dimensional histogram, the features of the Rényi entropy are used to distinguish between the two populations(OP and CT). Fig. 2 shows our studied Cases.

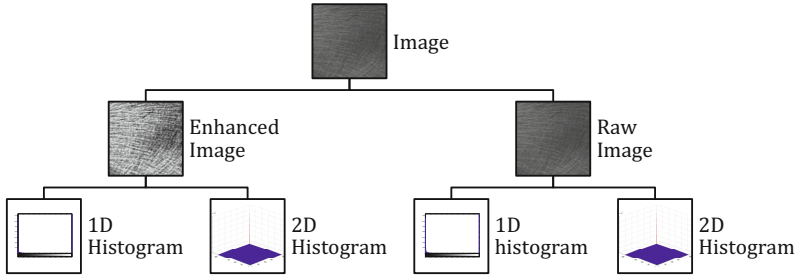


Fig. 2. Global chart of the proposed method

2.1 Preprocessing

To process the different data, first, we have used the Retinex algorithm introduced by Land *et al.* [13]. This method improves the contrast of the images using the reflection of light. There exist several versions of this algorithm and we have used the one defined by Funt *et al.* [14]. To keep the significant information of the trabecular bone patterns, a quantization over fewer gray levels was performed. Only 8 gray levels were kept to provide better and more easily exploitable images that are better suited for bone texture characterization. Figure 3 shows a sample of an enhanced and quantized image of a bone X-ray image.

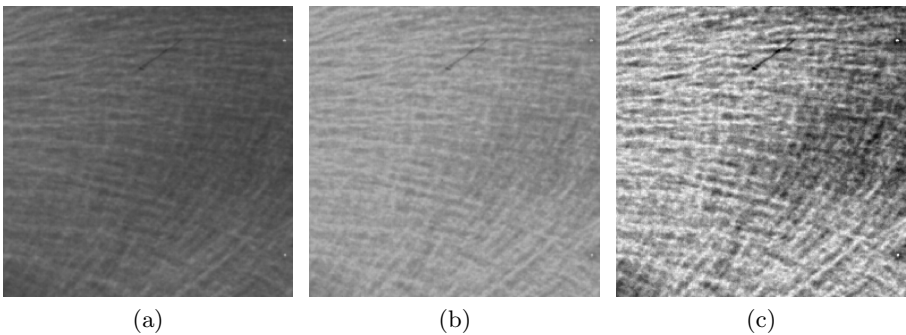


Fig. 3. (a) Original image of a calcaneus radiograph, (b) filtered image by the Retinex algorithm, (c) and quantized image over 8 gray levels

2.2 A Fully-Anisotropic Morlet Wavelet

The Morlet wavelet, was formulated by Goupillaud *et al.* [15]. Then, Antoine *et al.* [16] proposed anisotropic Morlet which is given by:

$$\psi(x) = e^{ik_0 \cdot x} e^{-1/2(x \cdot A^T A x)} \quad (1)$$

where $\mathbf{k}_0 = (0, k_0) \geq 5.5$ is a wave vector and $A = \text{diag}(L, 1)$ is an anisotropic matrix, "diag" denotes the diagonal matrix and L is the ratio of anisotropy. Kumar *et al.* [17] have controlled the orientation by defining $k_0 = (k_0 \cos\theta, k_0 \sin\theta)$ where θ is the parameter of orientation. The combination of the methods proposed by Kumar *et al.* [17] and Antoine *et al.* [16] produces an anisotropic and directional wavelet. This wavelet is not fully anisotropic. To solve this problem, Roseanna *et al.* [18] proposed a fully anisotropic Morlet where both the elliptical envelope and the wave vector are rotated through an angle defined by the orientation parameter θ . This wavelet is given by:

$$\psi(x, \theta) = e^{ik_0 \cdot Cx} e^{-1/2(Cx \cdot A^T A Cx)} \quad (2)$$

with $k_0 = (0, k_0)$, $k_0 \geq 5.5$, $A = \text{diag}(L, 1)$ and C is a linear transformation defined by:

$$C = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad (3)$$

So, the Wavelet coefficients are given by the following convolution:

$$W_\psi f(\mathbf{b}, a, \theta) = \frac{\sqrt{L}}{a} \int_{-\infty}^{\infty} f(x) \bar{\psi}\left(\frac{x - \mathbf{b}}{a}, \theta\right) dx = \frac{\sqrt{L}}{a} f(\mathbf{b}) * \bar{\psi}(-\mathbf{b}/a, \theta) \quad (4)$$

The exploitation of the fully anisotropic Morlet, enabled us solving the problem of orientation which is caused by the non-uniform changes. Fig. 4 shows a representative example of subband of an image from the database in different orientations.

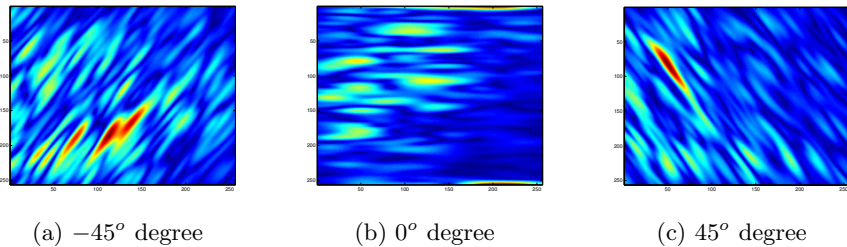


Fig. 4. A representative example of a sub-band for a Xray bone image in different orientations with $L = 0.2$ and $a = 4$

2.3 2D Histogram of Fully-Anisotropic Morlet Wavelet

To characterize the Morlet coefficients we use a two-dimensional (2D) histogram proposed by Sahoo *et al.* [19]. To Compute the 2D histogram for each sub-band, we proceed as follows. First, we calculate the average of the neighborhood for each coefficient. Let $g(x, y)$ be the average value of the neighborhood for the coefficient $f(x, y)$. Thus for a 3×3 neighborhood $g(x, y)$ is calculated as:

$$g(x, y) = \left\lfloor \frac{1}{9} \sum_{a=-1}^1 \sum_{b=-1}^1 f(x+a, y+b) \right\rfloor \quad (5)$$

where $\lfloor A \rfloor$ denotes the integer part of A . The average value is used for the construction of the normalized 2D histogram as:

$$Hist2D(k, l) = \frac{Prob(g(x, y) = k \cap f(x, y) = l)}{Number\ of\ Pixel} \quad (6)$$

Note that $(\sum_{i=1}^N \sum_{j=1}^M Hist2D = 1)$ where (M, N) is the size of the 2D histogram. The 2D histogram is used to compute the entropy as explained in the next section.

2.4 Rényi Entropy for 2D Histogram

The Rényi entropy [20] is widely used for the description of anisotropic textures. The Rényi entropy results from the generalization of the Entropy of Shannon. It is an efficient tool which has shown good performances [12]. The Rényi entropy, H_α , of order α ($\alpha \geq 0$, $\alpha \neq 1$) is defined as:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^n p_i^\alpha \right) \quad (7)$$

where p_i represents the probability density of $X = \{x_1 \cdots x_n\}$. In the literature and for most cases, the Rényi entropy refers to case $\alpha = 2$. In our case, the Rényi entropy was used as a feature for the description of each image. To this end, we have used the 2D histogram and the Rényi entropy, $Entro_{2D}$, as follows:

$$Entro_{2D} = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^N \sum_{j=1}^M Hist2D^\alpha(x, y) \right) \quad (8)$$

where $Hist2D$ is the Histogram 2D of each subband and M is the maximum gray level for the Histogram of sub-band and N is the maximum gray level for the Histogram of average of the same sub-band.

3 Experimental Results

For this study, we considered a population composed of 77 postmenopausal women suffering from osteoporotic vertebral crush fractures and control subjects. Among these subjects, there were 38 control (CT) cases and 39 patients

with osteoporotic (OP) fractures. As age has an influence on bone density and on trabecular bone microarchitecture, the control cases were age-matched with the vertebral crush fracture cases.

To realize calcaneus X-ray images a standardized procedure was followed. An X-ray clinical equipment was used. Focal-calcaneus distance was set at 1 m. The region of interest (ROI; Fig. 1(a)) was defined by a physician who marked anatomical markers on the calcaneus images. This way, we ensure that the ROI be acquired in the same area as well as in the same orientation from each bone radiography, since the effect of the orientation on the analysis is part of this study. This ROI of $2.7 \times 2.7 \text{ cm}^2$ was located in a region that contains only trabecular bone. The pixel size was $105 \mu\text{m}$.

The preprocessing as well as the orientation of analysis were evaluated. We have also compared the results obtained using either the two-dimensional or the one-dimensional histogram in the Rényi entropy.

Our method is based on a 4-step algorithm. First, the image content is enhanced. Then, each image is analyzed using the fully anisotropic Morlet wavelet in different orientations. Follows, the computation of the 2D histogram on each sub-band. Finally, the entropy from Rényi is estimated using each two-dimensional histogram. Namely, for the orientations, we used a range of $\theta = [-180, -135, -90, -45, 0, 45, 90, 135, 180]$. Thus, for each image, we choose $1 \leq N \leq 9$ for this range of orientations.

For the parameters of the Morlet wavelet, we chose $L = 0.2$ to take advantage of fully-anisotropic anisotropy [18]. For the scale, we use $a = 4$. Since, the purpose of this paper is take advantage for fully-anisotropic wavelet, the influence of the scales will be considered in a future work.

As a classifier, we used the neural networks with N as the size of the input vector with 30 nodes for the hidden layer and output. For the distribution of the data, we used 50% for learning, 25% for the test and 25% for the validation. Moreover, the Receiver Operating Characteristics curves (ROC) [21] were used

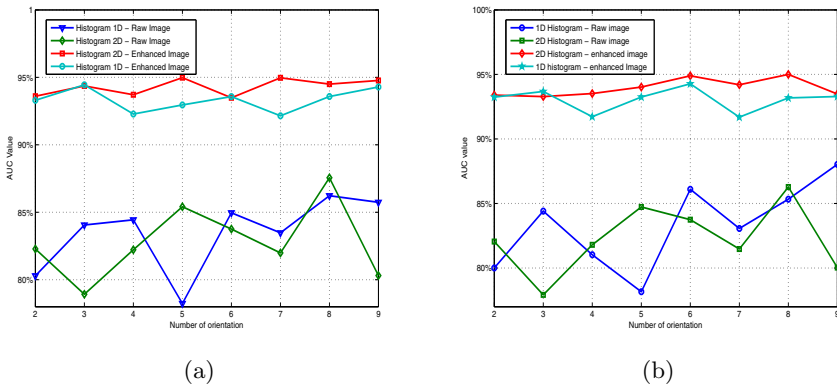


Fig. 5. AUC values depending on Number of Orientation using the Rényi entropy: (a) OP Class, (b) CT Class for 4 studied Case (Fig. 2)

to measure the influence of N on the rate of the classification. All the procedures were executed 100 times and the mean value was retained as a representative result.

Figure 5 shows the evolution of the Area Under Curve (AUC) of the ROC curves while varying the parameter N . Each graph corresponds to one of the proceeded data, raw or enhanced images using either the 1D or 2D histogram. As can be seen on figure 5, the enhancement improves the classification rates. The best classification rate is obtained for the enhanced image using the 2D histogram. $N = 4$ gives a good classification rate and seems to be a good trade off between efficiency and computation time.

4 Conclusion

In this work, we have proposed an original approach based on a fully-anisotropic Morelet and Rényi entropy for texture characterization with an application to bone X-ray images for the diagnosis of bone disease such as osteoporosis. Our technique combining image preprocessing and the entropy shows that it is possible to achieve better classification rates to distinguish between two different populations composed of osteoporotic patients and control subjects. The fully-anisotropic Morlet, helped us estimating non-uniform changes due to anisotropy variations induced by osteoporosis. The Neural Network classifier and the Receiver Operating Characteristics curves were used to distinguish between osteoporotic and control subjects. Combining our technique to Bone Mineral Density we can offer a new perspective for precise studies of bone disease such as osteoporosis.

References

1. Johnell, O.: The socioeconomic burden of fractures: today and in the 21st century. *Am. J. Med.* 103(2A), 20–25 (1997)
2. Cooper, C., Campion, G., Melton 3rd, L.J.: Hip fractures in the elderly: a world-wide projection. *Osteoporos Int.* 2(6), 285–289 (1992)
3. Dempster, D.W.: The contribution of trabecular architecture to cancellous bone quality. *J. Bone Miner. Res.* 15(1), 20–23 (2000)
4. Sevestre-Ghalila, S., Benazza-Benyahia, A., Ricordeau, A., Mellouli, N., Chappard, C., Benhamou, C.L.: Texture Image Analysis for Osteoporosis Detection with Morphological Tools. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3216, pp. 87–94. Springer, Heidelberg (2004)
5. Tuceryan, M., Jain, A.K.: Texture analysis, in the handbook of pattern recognition and computer vision, 2nd edn. (1998)
6. Petrou, M., Sevilla, P.G.: *Image processing: Dealing with texture* (2006)
7. Benhamou, C.L., Poupon, S., Lespessailles, E., Loiseau, S., Jennane, R., Siroux, V., Ohley, W., Pothuaud, L.: Fractal analysis of radiographic trabecular bone texture and bone mineral density: two complementary parameters related to osteoporotic fractures. *J. Bone Miner. Res.* 16(4), 697–704 (2001)

8. Jennane, R., Ohley, W.J., Majumdar, S., Lemineur, G.: Fractal analysis of bone x-ray tomographic microscopy projections. *IEEE Trans. Med. Imaging* 20(5), 443–449 (2001)
9. Pothuaud, L., Lespessailles, E., Harba, R., Jennane, R., Royant, V., Eynard, E., Benhamou, C.L.: Fractal Analysis of Trabecular Bone Texture on Radiographs: Discriminant Value in Postmenopausal Osteoporosis. *Osteoporosis International* 8, 618–625 (1998)
10. Houam, L., Hafiane, A., Jennane, R., Boukrouche, A., Lespessailles, E.: Trabecular Bone Anisotropy Characterization Using 1D Local Binary Patterns. In: Blanc-Talon, J., Bone, D., Philips, W., Popescu, D., Scheunders, P. (eds.) *ACIVS 2010, Part I. LNCS*, vol. 6474, pp. 105–113. Springer, Heidelberg (2010)
11. Pramudito, J.T., Soegijoko, S., Mengko, T.R., Muchtadi, F.I., Wachjudi, R.G.: Trabecular pattern analysis of proximal femur radiographs for osteoporosis detection. *Journal of Biomedical and Pharmaceutical Engineering* 1(1), 45–51 (2007)
12. Gabarda, S., Cristóbal, G., Rodríguez, P., Miravet, C., Del Cura, J.M.: A new Rényi entropy-based local image descriptor for object recognition. *Society of Photo-Optical Instrumentation Engineers*, vol. 7723 (2010)
13. Land, E.H., McCann, J.J.: Lightness and Retinex Theory. *Journal of the Optical Society of America* 61, 1–11 (1971)
14. Funt, B., Ciurea, F., McCann, J.: Retinex in matlab. *Journal of Electronic Imaging* 13(1), 48 (2004)
15. Goupillaud, P., Grossmann, A., Morlet, J.: Cycle-octave and related transforms in seismic signal analysis. *Geoexploration (former title)* 23(1), 85–102 (1984)
16. Antoine, J.P., Carrette, P., Murenzi, R., Piette, B.: Image analysis with two-dimensional continuous wavelet transform. *Signal Processing* 31(3), 241–272 (1993)
17. Kumar, P.: A wavelet based methodology for scale-space anisotropic analysis. *Geophysical Research Letters* 22(20), 2777–2780 (1995)
18. Neupauer, R., Powell, K.: A fully-anisotropic morlet wavelet to identify dominant orientations in a porous medium. *Computers and Geosciences* 31(4), 465–471 (2005)
19. Sahoo, P.: A thresholding method based on two-dimensional renyis entropy. *Pattern Recognition* 37(6), 1149–1161 (2004)
20. Rényi, A.: On Measures Of Entropy And Information. In: *Proc of the 4th Berkeley Symp. on Math., Stat. and Prob*, pp. 547–561 (1960)
21. Fawcett, T.: An introduction to roc analysis. *Pattern Recogn. Lett.* 27, 861–874 (2006)

Improving of Gesture Recognition Using Multi-hypotheses Object Association

Sebastian Handrich*, Ayoub Al-Hamadi, and Omer Rashid

Institute for Electronics, Signal Processing and Communications (IESK),
Otto-von-Guericke-University Magdeburg, Germany
{sebastian.handrich, ayoub.al-hamadi, omer.ahmad}@ovgu.de

Abstract. Gesture recognition plays an important role in Human Computer Interaction (HCI) but in most HCI systems, the user is limited to use only one hand or two hands under optimal conditions. Challenges are for instance non-homogeneous backgrounds, hand-hand or hand-face overlapping and brightness modifications. In this research, we have proposed a novel approach that solves the ambiguities occurred due to the hand overlapping robustly based on multi-hypotheses object association. This multi-hypotheses object association builds the basis for the tracking in which the hand trajectories are computed and this leads us to extract the features. The gesture recognition phase takes the extracted features and classifies them through Hidden Markov Model (HMM).

Keywords: hand tracking, multi hypotheses, HCI, gesture recognition.

1 Introduction

Multimodal human behavior analysis in modern human computer interaction (HCI) systems is becoming increasingly important, and it is supposed to outperform the single modality analysis. Like other modalities, for instance facial expression [1] and prosody, gestures play an important role, since they are very intuitive and close to natural human-human interaction [2]. The analysis of gestures in HCI systems requires a robust and realtime-capable hand tracking system. In our work we provide such a system. A lot of work has been done on hand tracking and gesture recognition. An overview can be found in [3]. However, hand tracking is due to the high number of freedoms, self occlusions and possible overlappings a difficult task. Many HCI applications are therefore limited to only one hand [4] or require other constraints, e.g. that the hand is the most foreground object [5], there are no overlappings [6] or that the user wears colored gloves [7]. In [8] a system was proposed that can handle hand-hand-overlappings if the appearance of both hands do not change during the overlapping. In [9] the authors developed a multi hypotheses based tracking approach. Such an approach provides the possibility to automatically correct false tracking results, which is important in a HCI environment.

* This work was supported by Transregional Collaborative Research Centre SFB/TRR 62 (“Companion-Technology for Cognitive Technical Systems”) funded by the German Research Foundation (DFG).

2 System Architecture

Figure 1 presents the architecture of the proposed approach which comprises of two main modules namely 1) hand detection and tracking, and 2) feature extraction and classification. Moreover, the hand detection and tracking problem is divided in two phases. In the first phase, skin colored objects are segmented and then clustered using 3D-data (i.e. image and depth information) of these objects by utilizing the distance from the camera. Further, the location and orientation of each object is re-estimated using Expectation Maximization (EM) algorithm at each frame. The second step contains a multi-hypotheses based tracking approach in which the hypotheses are generated based on the detected objects and fitted to the model of human body. The features are extracted in the second module which are derived from the hand trajectories and are then classified using discrete Hidden-Markov models (DHMM).

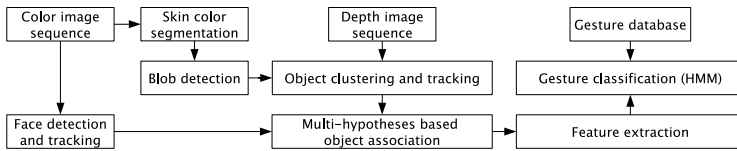


Fig. 1. System structure for hand gesture recognition

2.1 Hand Detection and Tracking

In the proposed approach, the data is acquired from Bumblebee2 camera which gives us 2D image and depth sequences. From these 2D images, the detection of skin colored objects starts with the classification of pixels as skin and non-skin pixels. For this purpose, a Gaussian mixture model (GMM) is trained using YCrCb color space. Moreover, a pixel is classified as skin color if $Y > 80$ and its probability $p(x)$ is above a threshold $p(skin) = 0.1$. $p(x)$ is determined as:

$$p(x) = \sum_{i=1}^N \pi_i \frac{e^{-0.5(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}}{2\pi \sqrt{|\Sigma_i|}}, \tag{1}$$

where $x = [Cr\ Cb]^T$ represents the color value of the pixel, $N = 4$ is the number of mixtures, and μ_i, Σ_i are the mean and covariance of each mixture. The trained parameters are shown in table 1. The skin color classification results in a mask image I_{skin} . Every skin-colored region (blob) in I_{skin} is detected using a blob-detection-algorithm. Since each blob B_i does not necessarily contain only one object, we create a histogram of depth values within each region. Finally, the initial number of objects and their positions are determined by peak-detection within each blob (see Fig 2). Each skin colored object O_k is described by (μ_k, Σ_k) , with $\mu_k = [x_k\ y_k\ z_k]^T$ as 3D object position and Σ_k as 3x3 covariance matrix describing the spatial distribution of the 3D-points q_i that are assigned to the object (Fig. 3). Here, tracking means to re-estimate both μ_k and Σ_k in each

Table 1. Trained parameters of the Gaussian mixture model for skin-detection. Samples were taken from a self created database with 8 different persons and 50 sample images per person (400 samples in total). The lighting conditions have remained stable due to the LED panels.

weight π_i	0.46	0.16	0.14	0.24
mean μ_i	(131.1 141.2)	(100.1 178.2)	(110.6 158.8)	(106.7 166.7)
covariance Σ_i	$\begin{pmatrix} 24.6 & 2.4 \\ 2.4 & 42.3 \end{pmatrix}$	$\begin{pmatrix} 10.3 & -13.1 \\ -13.1 & 30.6 \end{pmatrix}$	$\begin{pmatrix} 20.4 & -33.6 \\ -33.6 & 72.2 \end{pmatrix}$	$\begin{pmatrix} 17.0 & -20.6 \\ -20.6 & 40.2 \end{pmatrix}$

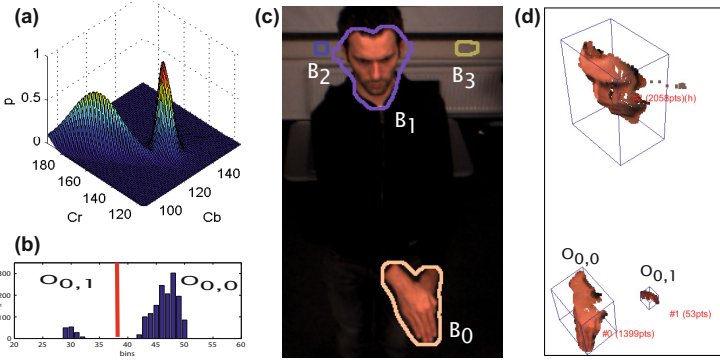


Fig. 2. Depth based object clustering: (a) Both hands share the same blob (B_0), (b) Depth-histogram of B_0 , (c) Hands were detected as two separate objects

frame and is done by EM-algorithm. In the E-step, we determine the probability $p_{k,j}$ for each skin-colored 3D-point $\mathbf{q}_j = [x_j \ y_j \ z_j]^T$ that it belongs to object $O_k : p = N(\mu_k, \Sigma_k)$. In the M-step, we then update μ_k and Σ_k of every object according to $p_{k,j}$. Moreover, to prevent the close objects to merge, only the M_k 3D-points with maximum probability of belonging to object O_k are used.

$$p_{k,j} = \frac{e^{-0.5(\mathbf{q}_j - \mu_k)^T \Sigma_k^{-1} (\mathbf{q}_j - \mu_k)}}{2\pi \sqrt{|\Sigma_k|}} \tag{2}$$

$$\mu_k = \frac{1}{|M_k|} \frac{\sum_j^{M_k} p_{k,j} \cdot \mathbf{q}_j}{\sum_j^{M_k} p_{k,j}}, \quad \Sigma_k = \frac{1}{|M_k|} \frac{\sum_j^{M_k} p_{k,j} \cdot (\mathbf{q}_j - \mu_k)^2}{\sum_j^{M_k} p_{k,j}} \tag{3}$$

The E- and M-step are repeated until convergence. To avoid numerical instabilities, the diagonal elements of Σ_k are set to $\sigma_k^2 = \max(\sigma_k^2, 10^{-4})$.

The second step in the proposed approach is the hand tracking. In this association problem, user’s hands are identified from the observed objects O_k detected at frame t . It is a difficult task for several reasons:

- Usually three objects are to be expected (head and hands). However, there can be more objects, e.g. skin-colored clothes.
- There can be less than the three expected objects (e.g. hand is hidden).
- Hands are hard to distinguish after an overlap.

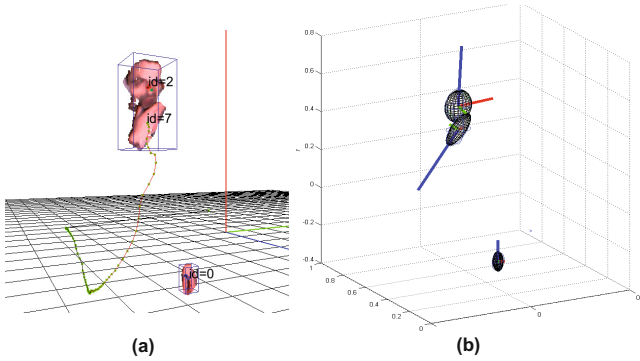


Fig. 3. Object tracking. (a) The user touches his head, but both are still separated objects. (b) the error ellipsoids of Σ_k and the corresponding eigen-vectors.

To tackle these issues, our approach contains the following steps:

- Head detection, based on face recognition. It is used for hypotheses creation.
- Based on the list of observed objects $\{O_k\}$, create a list $\{H_i\}$ of all possible hypotheses about the position of the hands at each time step t .
- Each hypothesis H_i is measured with a scoring function $S_p(H_i)$ which gives the probability of estimated pose according to model of the human body.
- Determine how each hypothesis matches the predictions of the N best hypotheses \widehat{H}_j of the previous time step ($S_M(H_i, \widehat{H}_j)$).
- Calculate the total score of all combinations and select the N best hypotheses. Discard all other hypotheses. In our system N was limited to 50 for computational reasons.

In HCI environments, where the user mostly faces the camera, it is unlikely that both elbows cross. So, each hypothesis contains the following assumptions: $H_i = \{x_h, x_{lh}, x_{rh}, x_{le}, x_{re}\}$, where x_h is the position of the head, determined by the face detector, x_{lh}, x_{rh} are the assumed positions of both hands and x_{le}, x_{re} are the estimated positions of the elbows. To estimate x_{le}, x_{re} , we assume that the elbow is in the direction of the largest hand extension. This is not always correct, however, sufficient for the validation of the hypotheses. So, for each skin-colored object $O_k = (\mu_k, \Sigma_k)$, we calculate the normalized largest eigenvector v_k of Σ_k and assume that the elbow is at $x_{e1,k1} = \mu_k \pm 0.3 \cdot v_k$, with 0.3 the estimated length of an underarm in meters. Figure 4 shows the generation of hypotheses in which three skin colored objects are observed with one object (H) recognized as head. This leads to a total of eight hypotheses. The next step is to calculate the score for each hypothesis and to discard the impossible hypotheses. So, first, we remove all the hypotheses where either the euclidian distance between the head and a hand is above 1.2meters or the hands are far behind the user ($x_{hand.z} - x_h.z > 0.5m$). The score S_p of the pose is determined by assuming a ground position of both elbows relative to the head ($z_{le/re} = x_h + [\pm 0.2 \ - 0.4 \ 0.1]^T$) and calculating:

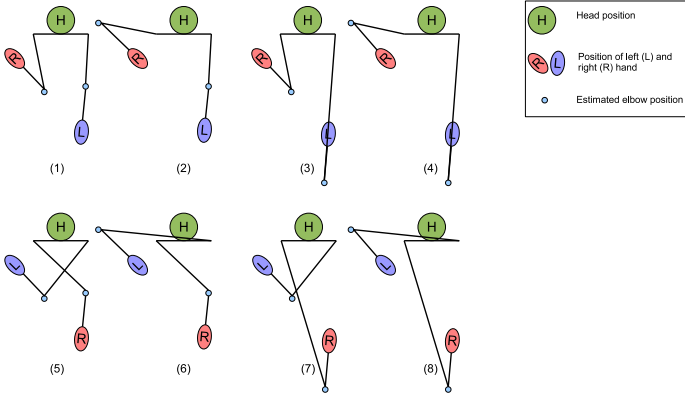


Fig. 4. Hand tracking: Based on the observed skin-colored objects, a list of all possible hypotheses about the body pose is created

$$S_p = e^{-|x_{te}-z_{ie}|/\sigma_e} \cdot e^{-|x_{re}-z_{re}|/\sigma_e}. \tag{4}$$

The parameter $\sigma_e = 1$ was chosen empirically. So, the pose is less likely to be higher than the distances between the estimated and assumed elbow positions. In the final step, for each hypotheses H_i of the current timestep, the score $S_M(H_i, \widehat{H}_j)$ is calculated which determines how well it matches N best hypotheses \widehat{H}_j of the previous timestep. This score is based on the euclidian distances between the predicted positions of the last hypothesis and the current one. This prediction is done by assuming that the user is moving his head and hands with the same velocities as in the previous timestep. So, the velocities are the euclidian distances between each hypotheses \widehat{H}_j and its parent hypothesis H_j . $S(H_i, \widehat{H}_j)$ is given by:

$$S_M(H_i, \widehat{H}_j) = 1 - \frac{1}{d_{max}} \sum_{k=1}^5 w_k \cdot d_k \tag{5}$$

with weights $w_k = [0.3 \ 0.25 \ 0.25 \ 0.1 \ 0.1]^T$ and d_k the distances between the prediction of H_j and H_i for all five hypothesis-elements, limited to d_{max} . So, the total score $S(H_i, \widehat{H}_j)$ is calculated as: $S(H_i, \widehat{H}_j) = S_M(H_i, \widehat{H}_j) \cdot S_p(H_i) \ \forall i, j = 1 \dots M, N$ with M the number of valid hypotheses in the current timestep and N the number of the best selected hypotheses of time step $t - 1$. Finally the hypotheses of $\{H_i\}$ with the N highest scores are selected: $\widehat{H} \leftarrow H(S = max_N(S))$.

2.2 Feature Extraction and Classification

The features are extracted from the detected hand trajectories at each frame. There are three features used in the proposed approach as:

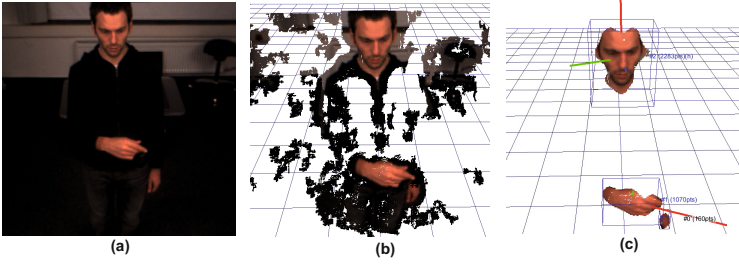


Fig. 5. Preprocessing: (a) A typical scene in the HCI environment. (b) Partially incomplete 3D data of the scene. (c) Extracted skin-colored objects (head and hands).

- Cylindrical coordinates of both hands relative to the head: $F_c = [h, r, \phi]$
- Velocities of the hands in m/s: $F_v = [v_x, v_y, v_z]$.
- Spatial orientation $F_o = \{diag(\Sigma_k)\}$ of the hands. Herefore, we used the diagonal elements of the covariance matrix Σ_k of objects, representing hands.

All the detected features are normalized to $[-1 \ 1]$ and are combined to one single feature vector at every frame F_t . A complete gesture path is then described as a temporal sequence of feature vectors: $G = \{F_{t-0} \dots F_{t-T}\}$. Since, we use discrete HMM for gesture classification, the feature vectors F_t are quantized to obtain discrete symbols z_t (vector quantization). It is done by using k-means clustering algorithm. The cluster index is then used as input to the DHMM. We have used a DHMM with LRB-architecture (left-right-banded) where Baum-Welch algorithm is used for training and Viterbi algorithm for evaluation.

3 Experimental Results

We tested our tracking and gesture recognition system on some videos taken from a database, which has been created at our university in the context of a research project that focuses on the development of companion-technology. In Fig. 5(a), a typical scene in an HCI environment is shown, in which the user is performing a gesture in front of the system. Fig. 5(b) shows the corresponding 3D-data captured with the Bumblebee2 stereo camera. Within the regions of user’s cloth, 3D-data is partially incomplete. In Fig.5(c), the results of skin-segmentation and object-clustering are shown. Our proposed hand tracking system is tested on app. 120,000 frames (ca. 90min) and provided very good results. One of the complicated cases is to track hands when the user crosses arms after a previous hand-hand-overlapping as shown in Fig. 6. The red (purple) circle represents the currently best assumed position of the right (left) hand. Starting from an initial position ($t=0$), there is a hand-hand contact ($t=23$). After that, the user crosses his arms ($t=26$) and during these time steps, the assignment of the hands was correct. However, a false assignment can occur during the hand-hand contact when the positions of the elbows are incorrectly estimated (Fig 7, $t=5$). However, due to the multi-hypotheses approach, the system is able to correct this false

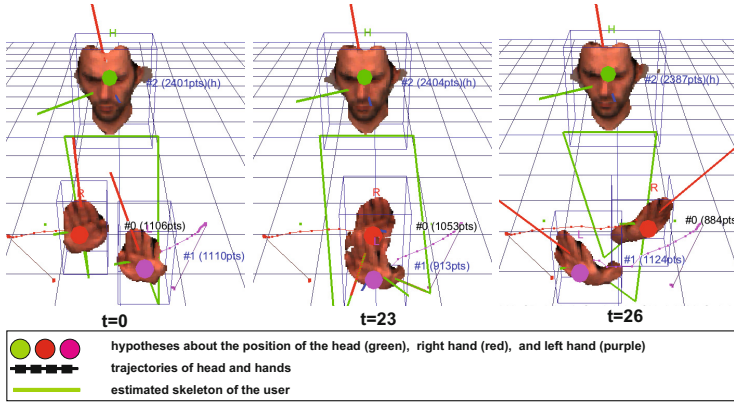


Fig. 6. Correct hand tracking in the case of hand-hand overlappings followed by a crossing of both arms

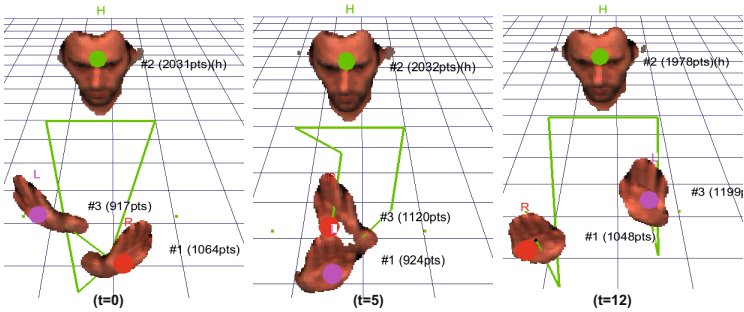


Fig. 7. Correction of false hypotheses during hand-tracking. In time step $t=5$ the assumed hand positions are incorrect (red/purple circles refer to left/right hand rather than vice versa). This is corrected in time step $t=12$.

assumption, since they become very unlikely ($t=12$). Since the hand detection is only based on skin-color, the system is only able to handle persons wearing long sleeves. Here, additional work to separate arms and hands has to be done. We combined our tracking module with a basic gesture recognition system. In Fig.8, row 1-3, time course of features (Section 2.2) for two consecutive gestures and the corresponding cluster results (row 4) are shown. Best results are achieved with $K=5$ clusters. Although, in Fig.8, the user performed an identical gesture twice, the features and so the cluster results ($t=0$) ($t=5$) ($t=12$) differ. However, this problem is handled by HMM. In the gesture recognition, HMM was trained on a dataset (40 sequences) and tested on a different set (40 sequences, different users). Best results are achieved with 5-state HMM, where 95% of all performed gestures were correctly recognized. An exemplary sequence is shown in Fig. 9.

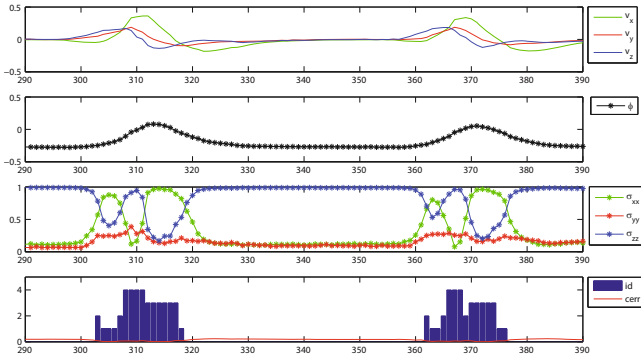


Fig. 8. Features for gesture recognition. Row 1-3 show the time course of features for two performed gestures. Bottom row: Result of k-Means algorithm used to obtain discrete observation symbols as input for the HMM.

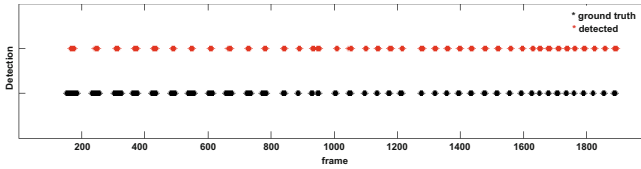


Fig. 9. Gesture recognition. The lower line (black stars) shows points in time at which the user performs a gesture (manually labeled). The upper line (red stars) shows the result of the automatic gesture recognition. All performed gestures were recognized.

4 Conclusions

Multimodal human behavior analysis in HCI environments is becoming increasingly important. Different modalities are involved in such analysis for instance facial expression, gestures and prosody. A gesture recognition system requires a robust hand tracking system. In our work, which is part of a HCI companion system, we provided such a system, which is able of tracking hands in non-trivial situations, for instance during hand-hand overlappings or hand-face-contacts. As an application we combined our tracking module with a basic gesture recognition system. The system was able to work in realtime (20 fps, 400 by 300px).

References

1. Niese, R., Al-Hamadi, A., Panning, A., Michaelis, B.: Emotion recognition based on 2d-3d facial feature extraction from color image sequences. *JMM* 5 (2010)
2. Hassanpour, R., Wong, S., Shahbahrami, A.: Vision based hand gesture recognition for human computer interaction: A review. In: *Int. Conference Interfaces and Human Computer Interaction*, pp. 125–134 (2008)

3. Shan, C., Tan, T., Wei, Y.: Real-time hand tracking using a mean shift embedded particle filter. *Pattern Recognition* 40, 1958–1970 (2007)
4. Suk, H.I., Sin, B.K., Lee, S.W.: Hand gesture recognition based on dynamic bayesian network framework. *Pattern Recognition* 43, 3059–3072 (2010)
5. Van den Bergh, M., Van Gool, L.: Combining rgb and tof cameras for real-time 3d hand gesture interaction. In: *Applications of Computer Vision, WACV* (2011)
6. El-Sawah, A., Joslin, C., Georganas, N., Petriu, E.: A framework for 3d hand tracking and gesture recognition using elements of genetic programming. In: *Canadian Conference on CRV*, pp. 495–502 (2007)
7. Keskin, C., Erkan, A., Akarun, L.: Real time hand tracking and 3d gesture recognition for interactive interfaces using hmm. In: *ICANN*, pp. 3–6 (2003)
8. Saeed, A., Niese, R., Al-Hamadi, A., Michaelis, B.: Solving the Hand-Hand Overlapping for Gesture Application. In: Choraś, R.S. (ed.) *Image Processing and Communications Challenges 3. AISC*, vol. 102, pp. 343–350. Springer, Heidelberg (2011)
9. Nickel, K., Stiefelhagen, R.: Visual recognition of pointing gestures for human-robot interaction. *Image and Vision Computing* 25, 1875–1884 (2007)

An Improved Images Watermarking Scheme Using FABEMD Decomposition and DCT

Noura Aherrahrou and Hamid Tairi

University Sidi Mohamed Ben Abdellah,
Faculty of Sciences, Dhar El mahraz,
LIIAN, Department of Informatics,
Fez, Morocco

noura.ah@hotmail.fr, htairi@yahoo.fr

Abstract. In this paper, we propose a new robust digital image watermarking scheme which integrates the Discrete Cosine Transform (DCT) and the Fast and Adaptive Bidimensional Empirical Mode Decomposition (FABEMD). The use of the FABEMD decomposition is motivated by the fact that FABEMD has better quality than any other decomposition technique in extracting intrinsic components known as Bidimensional Intrinsic Mode Functions (BIMFs) and residue. In the proposed approach, the watermark is embedded in the DCT coefficients of the residue, in order to achieve better performance in terms of perceptually invisibility and the robustness of the watermark. Experimental results and comparison analysis demonstrate that our method has better performance than the traditional watermarking method operating in the DCT domain.

Keywords: DCT, FABEMD, BIMFs, Watermarking.

1 Introduction

The fast development of the Internet in recent years has made it possible to easily create, copy, transmit, and distribute digital data. Consequently, this has led to a strong demand for reliable and secure copyright protection techniques for digital data. Digital watermarking has been proposed as valid solution for this problem.

In order to be successful, the watermark should be invisible and robust to premeditate any spontaneous modification of the image. It is very important to be robust against common image processing operations such as filtering, additive noise, resizing, cropping, and common image compression techniques.

Watermarking techniques can be categorized in different ways. They can be classified based on 1) the type of watermark being used (the watermark can be a visually recognizable logo or a sequence of random numbers) or 2) domain where the watermark is applied (spacial domain or the frequency-domain).

The most popular approaches for the image watermarking are the frequency-domain approaches. In these schemes, the image is being transformed via some common frequency transform and watermarking is achieved by altering the transform coefficients of the image. The transforms that are usually used are the DCT, DFT and the DWT. A question that raises in such approaches is the number and the position of the altered coefficients in the frequency representation of the image. Many different ideas have been presented, most of them originating from Cox's et al. system [5].

The method proposed by Cox's et al. computes the $N \times N$ DCT coefficients for an $N \times N$ image. The watermark of length n is placed into the n highest magnitude coefficients of the transform matrix. The motivation for choosing the higher value coefficients is that they represent the low frequency regions of an image, which contain most of the perceptually significant image information. Also, the human visual system attaches more resolution to the low frequency spectral components. Furthermore, it has been observed that common signal processing operations and distortions affect the perceptually insignificant regions of an image, which correspond to high-frequency components. So, the watermark has to be inserted in the low-frequency components.

Further performance improvements in the Cox's proposed method may be achieved by using the FABEMD decomposition. The use of FABEMD in Cox's method and the embedding of the watermark in the DCT coefficients of the residue is supported by the fact that FABEMD has better quality than any other decomposition technique in extracting intrinsic components (BIMFs) which contain the different frequency parts of the image from high to low frequency (from BIMFs to residue).

2 FABEMD (Fast and Adaptive Bidimensional Empirical Mode Decomposition)

2.2 FABEMD Overview

Empirical Mode Decomposition (EMD) is first developed by Huang et al. [1] and has shown to be a powerful tool for decomposing nonlinear and nonstationary signals. The concept of EMD is to decompose the signal into a set of zero mean functions called Intrinsic Mode Functions (IMF) and a residue. This decomposition technique has also been extended to analyze two-dimensional (2D) data/images, which is known as bidimensional EMD (BEMD). In EMD or BEMD, extraction of each IMF or BIMF requires several iterations. Hence, the extreme detection and interpolation at each iteration make the process complicated and time consuming. The situation is more difficult for the case of BEMD, which requires 2D scattered data interpolation at each iteration. For some images it may take hours or days for decomposition.

To overcome these limitations of BEMD, a novel approach called Fast and Adaptive BEMD (FABEMD) was proposed recently by Bhuiyan et al. [2, 3]. It substitutes the 2D scattered data interpolation step of BEMD by a direct envelop estimation method and limits the number of iterations per BIMF to one. In this technique, spatial

domain sliding order-statistics filters, namely, MAX and MIN filters, are utilized to obtain the running maxima and running minima of the data. Application of smoothing operation to the running maxima and minima results in the desired upper and lower envelopes respectively. The size of the order-statistics filters is derived from the available information of maxima and minima maps.

2.3 FABEMD Algorithm

Figure 1 illustrates the different steps of the FABEMD. Let the original image be denoted as I , a BIMF as $BIMF_i$, and the residue as R . In the decomposition process i th BIMF $BIMF_i$ is obtained from its source image S_i , where S_i is a residue image obtained as $S_i = S_{i-1} - BIMF_{i-1}$ and $S_1 = I$. It requires one or more iterations to obtain $BIMF_i$, where the intermediate temporary state of BIMF in j th iteration can be denoted as F_{Tj} . With the definition of the variables, the steps of the FABEMD process can be summarized as follows:

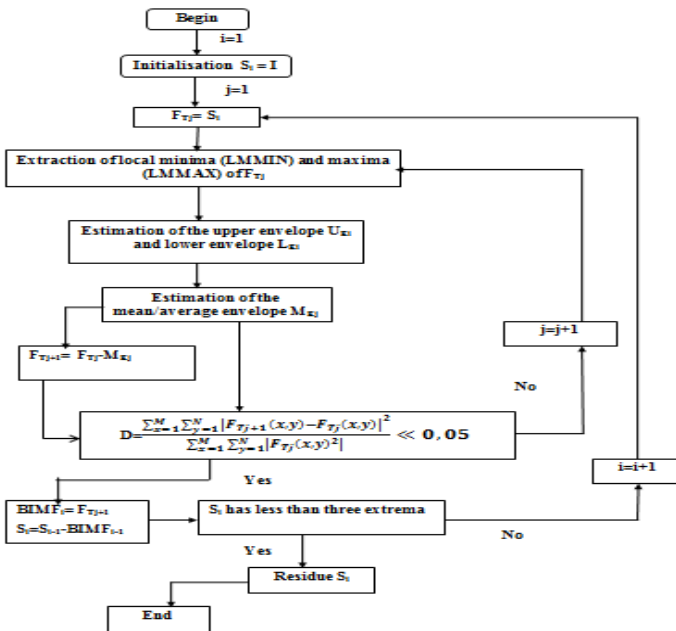


Fig. 1. FABEMD algorithm

Figure 2 shows an example of FABEMD decomposition.

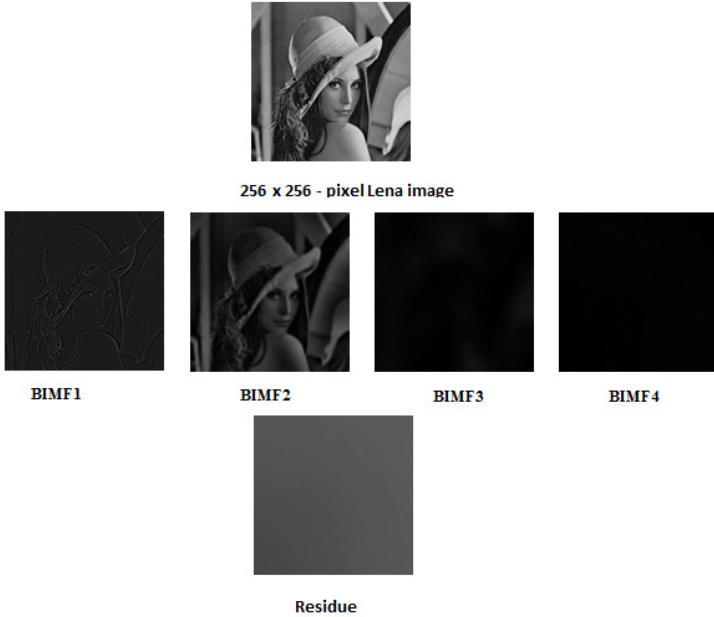


Fig. 2. Result of the FABEMD decomposition of Lena image

3 Proposed Approach

3.2 Watermark Embedding

The watermark $X = \{x_1, x_2, \dots, x_n\}$ consists of a pseudo-random sequence of M length.

To insert the watermark in the host image, the first step is to decompose the image into BIMFs and Residue, then calculated DCT, followed by introduced the watermark on the selected coefficients of the Residue.

The watermark will be introduced from the $L+1$ coefficient to the $M + L$ coefficient of the DCT coefficients range.

These coefficients generate the vector $T = \{t_{L+1} \dots \dots t_{L+M}\}$ and will be modified according to:

$$t'_{L+i} = t_{L+i} + \alpha |t_{L+i}| x_i \tag{1}$$

where α is the watermark strength, which determine the invisibility of the watermark.

These coefficients will be the elements of the vector $T' = \{t'_{L+1}, t'_{L+2} \dots \dots t'_{L+M}\}$ the modified coefficients.

Finally, the vector T' is reinserted into the original DCT coefficients according with the original order, and with the IDCT it is obtained the watermarked Residue which will be added to BIMFs to produce the watermarked host image.

Therefore, the watermark embedding steps are represented in figure 3 followed by a detailed explanation.

1. Decompose image into BIMFs and residue.
2. Generate the watermark X which has to be inserted.
3. Calculate the DCT of the Residue.
4. Generate the coefficient vector T.
5. Modify the coefficient vector T according to the X values to obtain T'.
6. Swap the original coefficients (T) by the modified in the vector T'.
7. Calculate the IDCT of the Residue to obtain the watermarked residue.
8. Build the watermarked image by adding all BIMFs and the watermarked residue

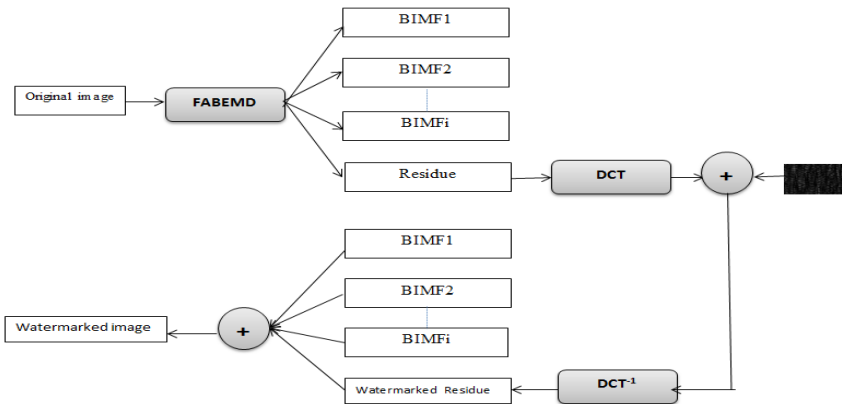


Fig. 3. Watermark embedding scheme

3.3 Watermark Detection

The first step to detect the watermark of a given image I^* consists of calculating the DCT transform. The DCT coefficients of I^* from $L + 1$ to $L + M$ are selected to form the vector of marked coefficients and perhaps modified $T^* = \{t_{L+1}^*, t_{L+2}^*, \dots, t_{L+M}^*\}$

To detect the mark it's necessary to correlate the marked coefficients and perhaps modified T^* with the watermark Y . The correlation formula to be used is defined by:

$$z = \frac{YT^*}{M} = \frac{1}{M} \sum_{i=1}^M y_i t_{L+i}^* \tag{2}$$

Where y_i is the watermark to be verified and t_{L+i}^* are the coefficients of the marked DCT and perhaps modified.

The DCT coefficients which are inside the vector T are always the coefficients which were obtained ignoring the first L elements and taking the next M elements, then the previous formula will be simplified as:

$$z = \frac{YT^*}{M} = \frac{1}{M} \sum_{i=1}^M y_i t_i^* \tag{3}$$

The correlation z can be used to determine whether a given mark is present or not, z is simply compared to a predefined threshold Tz[5], whereas z is computed for each of the watermarks and that with the largest correlation is assumed to be the one really present in the image.

To detect the watermark into a given image the following steps are shown in figure 4.

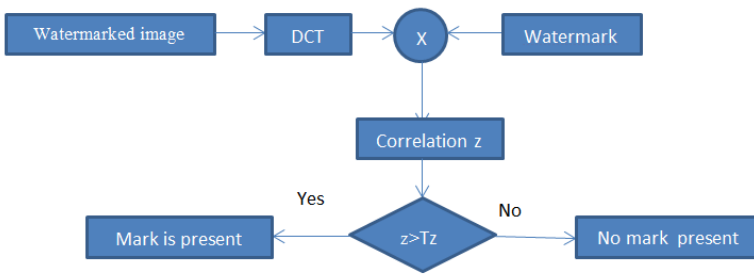


Fig. 4. Watermark detection scheme

4 Experimental Results

In this section the results of our study are shown. Several experiments are done to evaluate the effectiveness of the presented watermarking algorithm.

4.1 Invisibility of the Watermark

In this section the invisibility of the watermark is evaluated.

The PSNR is popularly used to measure the similarity between the original image and the watermarked image. While higher PSNR usually implies higher fidelity of the watermarked image.







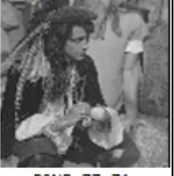








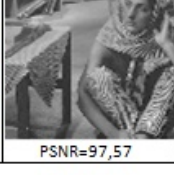


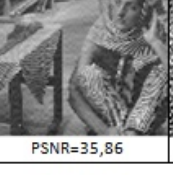
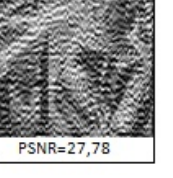
As can be seen in table1, for traditional method, we observe that if we fix the watermarking parameters L and M, then smaller watermarking strength α Results in higher robustness of watermark process, while the unreasonably big watermarking strength α Results in the watermark perceptually visible in the watermarked image. Here the invisibility of the watermark is demonstrated to show the successful use of the FABEMD in the traditional scheme.

4.2 Robustness of the Method against Attacks

To check if the proposed scheme is robust it has been implemented some attack functions to compare the expected results with the obtained results.

The mark will be specified by $\alpha = 0.2$ and $L = 16000$, $M = 16000$. Furthermore, to check the detection process behavior will be used a group of 100 watermarks, only one will be the correct (Only watermark number 40 is correct) while the others will be randomly generated.

Table 1. Comparison between our method and Cox’s method

	$\alpha=0.001$	$\alpha=0.01$	$\alpha=0.1$	$\alpha=1$	$\alpha=10$
Our method	 PSNR=93,50	 PSNR=93,50	 PSNR=93,50	 PSNR=93,42	 PSNR=85,23
Traditional method	 PSNR=97,57	 PSNR=77,71	 PSNR=53,81	 PSNR=35,46	 PSNR=27,72
Our method	 PSNR=89,54	 PSNR=89,54	 PSNR=89,54	 PSNR=89,52	 PSNR=86,45
Traditional method	 PSNR=97,57	 PSNR=78,35	 PSNR=54,20	 PSNR=35,86	 PSNR=27,78

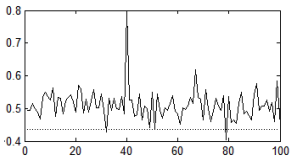
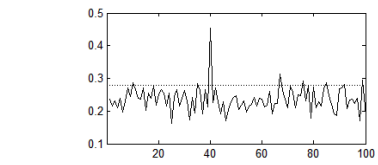
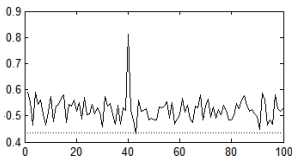
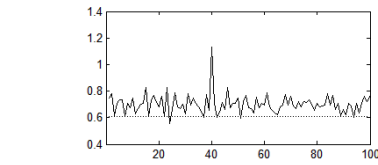
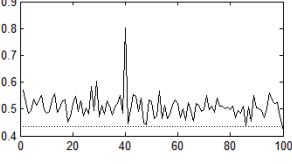
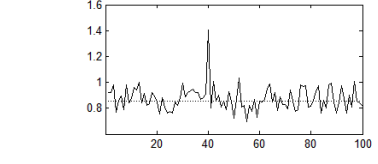
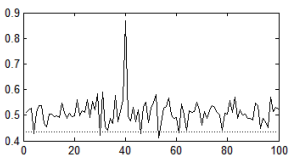
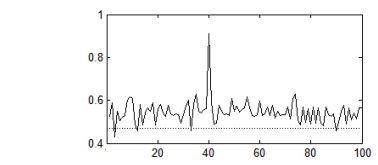
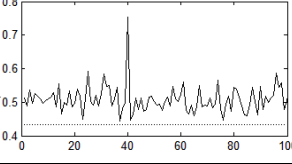
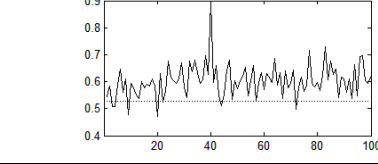
	Our method		Traditional method	
Median filter				
	Threshold	Response	Threshold	Response
	0.43	0.79	0.28	0.45
JPEG compression				
	Threshold	Response	Threshold	Response
	0.43	0.81	0.61	1.13
Gaussian noise				
	Threshold	Response	Threshold	Response
	0.43	0.80	0.85	1.40
Cropping				
	Threshold	Response	Threshold	Response
	0.43	0.86	0.46	0.91
Dithering distortion				
	Threshold	Response	Threshold	Response
	0.43	0.75	0.52	0.89

Fig. 5. Detector responses on 100 randomly generated watermarks, after 512X512 pixel Lena watermarked image been attacked with gaussian noise, dithering distortion, cropping , JPEG compression and median filter using Cox’s method and the proposed method. Only watermark number 40 is correct

5 Conclusion

We have presented a new robust digital image watermarking scheme based on joint FABEMD-DCT. Our scheme is shown to be resistant against several signal processing techniques, including dithering distortion, median filtering, Gaussian noise, cropping, and JPEG compression. Furthermore, we show that the implementation of the FABEMD algorithm leads to better performance in terms of invisibility of the watermark compared to traditional method.

References

1. Huang, N.E., Shen, Z., Long, S.R., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A* 454(1971), 903–995 (1998)
2. Bhuiyan, S.M.A., Adhami, R.R., Khan, J.F.: Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation. *EURASIP J. Adv. Signal Process* 2008, 1–18 (2008)
3. Bhuiyan, S.M.A., Adhami, R.R., Khan, J.F.: A novel approach of fast and adaptive bidimensional empirical mode decomposition. In: *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, March 31–April 4, pp. 1313–1316 (2008)
4. Cox, I., Kilian, J., Thompson Leighton, F., Shamon, T.: Secure Spread Spectrum Watermarking for Multimedia. *IEEE Trans. on Image Processing* 6(12) (December 1997)
5. Barni, M., Bartolini, F., Cappellini, V., Piva, A.: A DCT Domain System for Robust Image Watermarking. *Signal Processing* 66, 357–372 (1998)

A Fragile Watermarking Scheme Based CRC Checksum and Public Key Cryptosystem for RGB Color Image Authentication

Nour El-Houda Golea

Department of Computer Science, University of BATNA, Algeria
golea.nour@gmail.com

Abstract. The increased use of multimedia applications pose more problems concerning the preservation of confidentiality and authenticity of the transmission of digital data. These data, in particular the images should be protected from tampering. The solution is the use of fragile watermarking. Fragile watermarking can be modeled as a problem of communication of a signal over a noisy and hostile channel, where the attack takes place. Indeed, the use of error checking algorithms appear natural. . Cyclic redundancy check (CRC) code provides a simple, yet powerful, method for the detection of burst errors during digital data transmission and storage. CRC is one of the most versatile error checking algorithm used in various digital communication systems. In this paper, we propose a novel fragile watermarking scheme based CRC checksum and public key cryptosystem for RGB color image authentication.

Keywords: Fragile watermarking, image authentication, Cyclic redundancy check (CRC), public key cryptosystem, RGB color image watermarking.

1 Introduction

Digital watermarking technology is the process of embedding information into digital data in such a way that it is imperceptible to a human observer but easily detected by computer algorithm. A digital watermark is a invisible information pattern that is embedded into a suitable component of the data source by using a specific computer algorithm. Digital watermarks are signals added to digital data (audio, video, or still images) that can be detected or extracted later to make an assertion about the data [1, 2, 3].

Digital watermarking schemes can be classified as either *robust* or *fragile* . Robust watermarking schemes can be used to authenticate ownership [4, 5], whereas fragile watermarking schemes are commonly used for image authentication to verify whether the received image was modified during transmission or not. One may hide the watermark imperceptibly in the image before transmission and detect it after receiving to make sure that the received image is original or corrupted. Many watermarking schemes for image authentication have been proposed [6, 7, 8].

The watermarks used for the authentication must contain information which determines the integrity of the image. The watermark must invisible and fragile

(so, any modification in the watermarked image also in the signature must be detected and it is very desirable that it can detect the corrupted region.

In general, fragile watermarking schemes divide an original image into non-overlapping blocks, embed a signature, and detect the modified location for every block. Memon and Wong [6] proposed a method in which an image is divided into blocks and each block contains the hash value calculated from the MSB¹'s of the pixels forming that block. Fridrich [7] also proposed that the authentication watermark should exclusive-OR (XOR) the hash value of the block with more block information. Lin and al. [8] proposed a fragile block-wise, and content-based watermarking for image authentication and recovery. In this scheme, the watermark of each block is an encrypted form of its signature, which includes the block location, a content-feature of another block, and a CRC checksum. While the CRC checksum is to authenticating the signature. With block-based fragile watermarking we cannot detect exactly the modified pixels, but we can detect the corrupted block.

Fragile watermarking schemes are classified into those using the public key cryptosystem [7,8] and those using the private key cryptosystem as the tool for making the signature [6].

In this paper, we propose fragile watermarking based pixel detection approach using CRC and a public key cryptosystem. This approach is proposed to authenticate the RGB color image using the CRC checksum to detect the corrupted pixels.

The proposed method is decomposed from four functions: the first one generate a generator polynomial P_X with same size as the host image $n \times m$, each element $P_X(i,j)$ is used to create the watermark $W(i,j)$. The second function, use P_X to generate the watermark W of size $n \times m$, each element $W(i,j)$ is a binary sequence of six bits. The watermark $W(i,j)$ is the remainder of the division of the 18 MSB bits of the three colored pixels $R(i,j)$, $G(i,j)$ and $B(i,j)$ by $P_X(i,j)$. After this step, the P_X is encrypted using a secret key K_S , and performing the RSA encryption algorithm. The third function embed each two bits of $W(i,j)$ in the two LSB² of the corresponding three colored pixels. At the reception, the encrypted P_X is decrypted using a public key K_P . The last function, extract the watermark (CRC checksum) from the two LSB of the three colored pixels. The extracted watermark is appended at the end of the 18 MSB of the colored pixels. Then, divide this new sequence by $P_X(i,j)$, if the remainder is zero then the pixel $f(i,j)$ is authentic else it is corrupted.

The remainder of this paper is organized as follows: Section 2 gives a brief description of the CRC principle. Our proposed scheme is presented in Section 3. In Section 4, the experimental results are described and analyzed. Finally, we draw the conclusions of our work in Section 5.

2 Cyclic Redundancy Check

Normally, for the error detection in digital communication systems, a certain number of check bits, often called a *checksum*, is computed on the message that needs to be transmitted. The computed checksum is then appended at the end of the message

¹ Most Significant Bit.

² Least Significant Bit.

stream and is transmitted. At the receiving end, the message stream’s checksum is computed and compared with the transmitted checksum. If both are equal, then the message received is treated as error free. Cyclic Redundancy Code Check, or CRC works in a similar way, but it has greater capabilities for error detection than the conventional forms. CRC is one of the most versatile error checking technique used in various digital communication systems. Different CRC polynomials are employed for error detection. The size of CRC depends upon the polynomial chosen.

The message to be transmitted is treated as a polynomial and divided by an irreducible polynomial known as the *generator polynomial*. The degree of the generator polynomial should be less than that of the message polynomial. For a $n + 1$ bits generator polynomial, the remainder will not be greater than n bits. The CRC checksum of the data is the binary equivalent of the remainder after the division.

In general, an n -bit CRC is calculated by representing the data stream as a polynomial $M(x)$, multiplying $M(x)$ by x^n (where n is the degree of the polynomial P_X), and dividing the result by P_X . The rest of the division is the CRC checksum which is appended to the polynomial $M(x)$ and transmitted. The complete transmitted polynomial is then divided by the same P_X at the receiver end. If the result of this division has no remainder, there are no transmission errors [9].

3 Proposed Method

The proposed method is decomposed from four algorithms: P_X (generator polynomial) generation algorithm, W (watermark) generation algorithm, embedding and extraction algorithms. This method is modeled in Fig. 1.

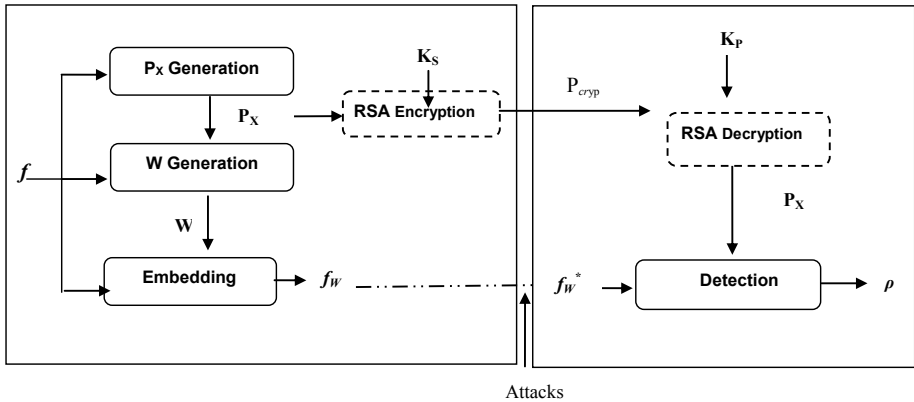


Fig. 1. Proposed model

The P_X generation algorithm create the matrix P_X of degree d taking the size of the original host image f and d . This function is described as $Generator_p$:

$$P_X = Generator_p(f, d). \tag{1}$$

The watermark generator algorithm generates a signature that contain the watermarking information, by taking the original host image f and a generator polynomial P_X , that is described as a function $Generator_W$:

$$W=Generator_W(f, P_X). \tag{2}$$

The embedding algorithm takes the signature and the host image, and generates the watermarked image f_w , that is described as a function E :

$$f_w= E(f, W). \tag{3}$$

The detection algorithm loads the watermarked, normal or corrupted image f_w^* and P_X , and calculate the measure ρ . The process can be described as function D :

$$\rho =D(f_w^*, P_X) . \tag{4}$$

if $\rho=0$ then the pixel is not corrupted, else it is corrupted.

The generator polynomial P_X must be encrypted using a secret key K_S and performing the asymmetric key *Encryption* algorithm. At the reception, the receptor must decrypt the generator polynomial P_{cryp} using the public key K_p and performing the *Decrypted* algorithm.

3.1 Generator Polynomial Generation Algorithm

This algorithm allows to create a matrix P_X of the same size of the host image f , each element $P_X(i, j)$ of this matrix is a generator polynomial used to calculate the CRC checksum corresponding to the pixel $f(i,j)$.

Input:

- $n \times m$: size of the host image f ;
- d : the maximal degree of $G(x)$, i.e., the maximal number of bits used to insert the watermark (in this case $d=6$).

Output:

P_X : the generator polynomial matrix of size $n \times m$.

Steps:

- for $i=1$ to n do
 - for $j=1$ to m do
 - Randomly generate a binary sequence g of size $d+1$;
 - $P_X(i,j)$ is calculated as:

$$P_X(i,j)=g_1 X^6 + g_2 X^5 + g_3 X^4 + g_4 X^3 + g_5 X^2 + g_6 X^1 + g_7 X^0. \tag{6}$$

To encrypt and decrypt the P_X , we propose to use the RSA algorithm (named for its inventors, Ron Rivest, Adi Shamir, and Leonard Adleman). The RSA cryptosystem is the most widely-used public key cryptography algorithm. It can be used to encrypt a message without the need to exchange a secret key separately. The RSA algorithm

can be used for both public key encryption and digital signatures. Its security is based on the difficulty of factoring large integers [10, 11].

3.2 Watermark Generation Algorithm

This algorithm generate a watermark W of size $n \times m$, where each element is presented by 6 bits. The watermark is generated depending on 18 MSB bits of the colored pixels (R, G and B). The following algorithm summarizes the way the watermark W is generated using the host image f and the generator polynomial P_X .

Input:

- f : RGB color host image of size $n \times m$;
- P_X : generator polynomial matrix of size $n \times m$.

Output:

- W : watermark of size $n \times m$, each element $W(i,j)$ is binary sequence of 6 bits $W = \{ W_1, W_2, \dots, W_6 \}$.

Steps:

- For each colored pixels $R(i,j)$, $G(i,j)$ and $B(i,j)$ do:
 1. Construct the message M by concatenating the 6 MSB bits of each pixel.
 2. Perform the CRC encoding to calculate the checksum:
 - Calculate $M' = M(x) \times x^d$;
 - The watermark $W(i,j)$ is the remainder of division of M' by $P_X(i,j)$.

Fig. 2 illustrates the block diagram of the watermark generation algorithm.

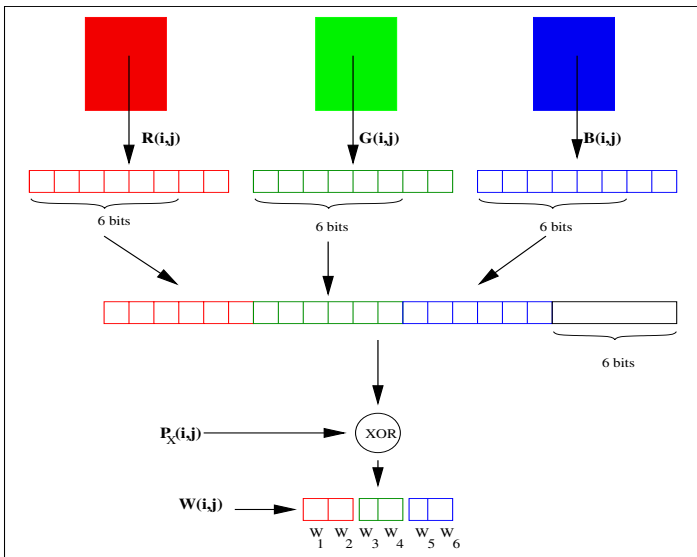


Fig. 2. Block diagram of the watermark generation algorithm

3.3 Embedding Algorithm

The following algorithm describes the way the watermark W is inserted in the host image f .

Input:

- f : RGB color host image of size $n \times m$;
- W : generated watermark of size $n \times m$.

Output:

- f_w : watermarked image of size $n \times m$.

Steps:

- For each colored pixels $R(i,j)$, $G(i,j)$ and $B(i,j)$ do :
 1. Replace the two LSB bits of $R(i,j)$ by $W_1(i,j)$ and $W_2(i,j)$;
 2. Replace the two LSB bits of $G(i,j)$ by $W_3(i,j)$ and $W_4(i,j)$;
 3. Replace the two LSB bits of $R(i,j)$ by $W_5(i,j)$ and $W_6(i,j)$.

Fig. 3 presents the block diagram of the watermark embedding algorithm.

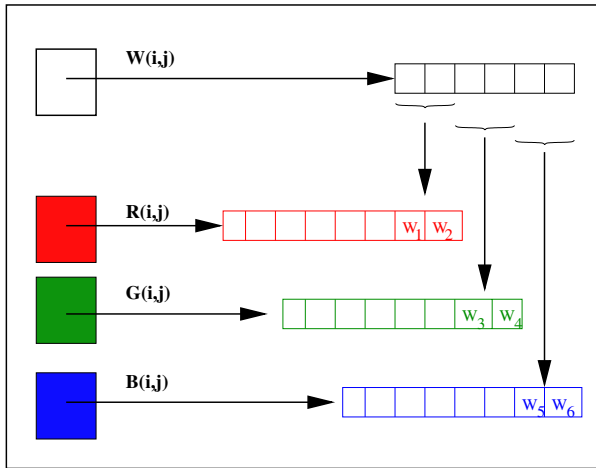


Fig. 3. Block diagram of the embedding algorithm

3.4 Detection Algorithm

The following algorithm summarizes the way the watermark is extracted from watermarked image f_w and $P(x)$.

Input:

- f_w : RGB color watermarked received image of size $n \times m$;
- P_X : generator polynomial matrix of size $n \times m$.

Output:

- ρ : decision parameter (confidentiality measure).

Steps:

- For each pixel $R_w(i,j)$, $G_w(i,j)$ and $B_w(i,j)$ do :
 1. Construct the message M' by concatenating the 6 MSB bits of each pixel.
 2. Extract the watermark W from the two LSB of each pixel.
 3. Perform the CRC decoding:
 - The watermark W is appended at the end of M' ;
 - Calculate the remainder ρ of the devising of M' by $P_X(i,j)$.
 if $\rho=0$ then the watermarked pixel $f_w(i,j)$ is not corrupted
 else $f_w(i,j)$ is corrupted.

Fig. 4 shows the block diagram of the detection algorithm.

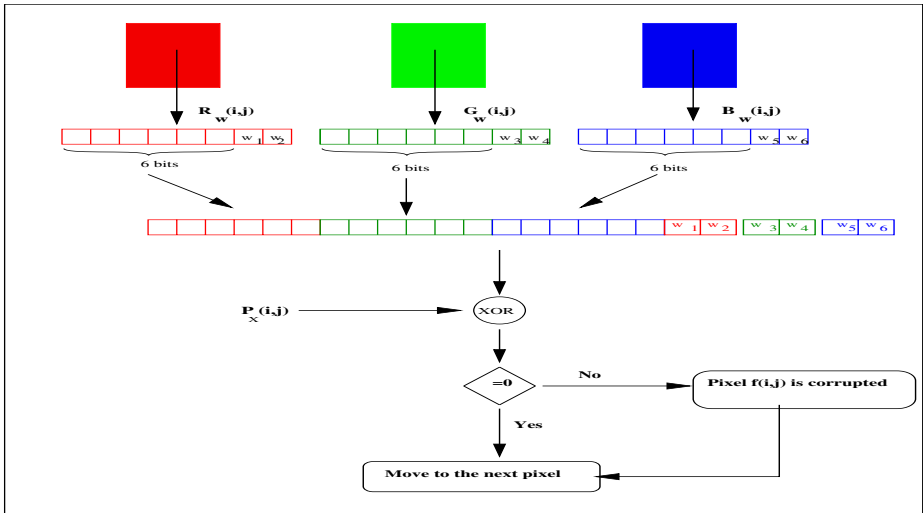


Fig. 4. Block diagram of the detection algorithm

4 Simulation and Experimental Results

In this section, we mainly demonstrate the imperceptibility and the fragility of our watermarking method. The experimental results reported here have been separated into two parts: the first one is for testing the imperceptibility property and the other one is for evaluating the fragility to malicious manipulations.

Imperceptibility Property

In order to test the imperceptibility property of our watermarking method, several typical RGB color images with size 128x128 such as *Baboon*, *Sailboat*, *Lena* and *house* have been watermarked. These original host images with their watermarked images have, respectively, been shown in Fig. 5.

From these result images, we could see that the differences between the original images and their corresponding watermarked images are hard to be perceived by human eyes.



Fig. 5. Host images with their corresponding watermarked images

To concretely estimate the quality of our method, we employed the *Peak Signal to Noise Ratio (PSNR)* to evaluate the distortion of the watermarked images. Table 2 depicts the name of some of the experimental images and their PSNR values of RGB components.

Table 1. Quality of the watermarked images

<i>Host image</i>	Baboon	Sailboat	Lena	House
<i>PSNR</i>	47.2633	47.2407	47.2578	47.4048

It can be seen from the PSNR values that the distortion between the watermarked images and the original ones is imperceptible. One can also conclude that the watermarked images have a good quality when the block size increases.

Fragility Property


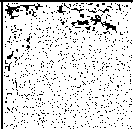


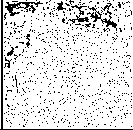
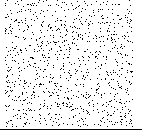
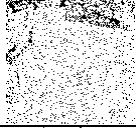

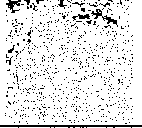


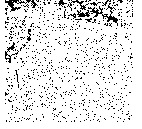
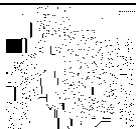
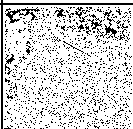
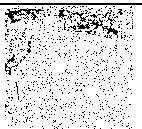
In order to evaluate the fragility to attacks, the CRC image is the matrix of the remainders of division of each pixel $f(i,j)$ by $P_x(i,j)$, if this image is black then the image is not corrupted, else it is corrupted. Fig. 6 presents the CRC images calculated from the watermarked images.



Fig. 6. CRC images extracted respectively from *Baboon*, *Sailboat*, *Lena* and *house* watermarked images

To highlight the fragility of our method, we have also taken into account many kinds of image watermarking attacks. Table 2 shows the extracted CRC images after different typical and standard attacks.

Table 2. Performances against several typical and standard attacks

Attack	CRC image	Attack	CRC image	Attack	CRC image
Rotation <i>Angle= 0.01°</i>		Average filter <i>3x3</i>		Salt & Pepper noise <i>D=0.002</i>	
Rotation <i>Angle=1°</i>		Gaussian filter <i>3x3</i>		Gaussian noise <i>M = 0.0, V = 0.001</i>	
Resize <i>128 ==>256</i>		Laplacian filter <i>Default parameters</i>		Winner filter <i>3x3</i>	
JPEG <i>Q=90</i>		Median filter <i>3x3</i>		Blurring <i>radius = 0.1</i>	
JPEG <i>Q=30</i>		Sharpen filter <i>1.0</i>		Blurring <i>radius = 1.0</i>	

Form Table 2, one can see that our watermark embedding method has strong fragile against many image attacks.

5 Conclusion

In this paper, we propose a fragile watermarking for RGB color images. The proposed method based pixel detection use CRC checksum and public key cryptosystem. The watermark is generated depending to the three colored pixels and using the generator polynomial. The generated watermark is inserted in the 2 LSB of each corresponding colored pixels. At the detection, the extracted watermark from the LSB of each colored pixels is appended at the end of the 18 MSB of the three colored pixels. This sequence is divided by the corresponding generator polynomial, if the remainder is null, then the pixel is not corrupted. The generator polynomial is encrypted using RSA public key cryptosystem.

The simulation results show that the proposed system performs fairly well when it is required to detect all kind of alteration, indicating precisely its altered region.

Acknowledgments. We are grateful to Dr Redha BENZID and Dr Rachid SEGHIR for early stimulating discussion and for their valuable comments and suggestions which lead to substantial improvements of this paper.

References

1. Voyatzis, G., Nikolaidis, N., Pitas, I.: Digital Image Watermarking: An Overview. In: 9th IEEE European Signal Processing Conference, vol. 1, pp. 9–12 (1998)
2. Barni, M., Cox, I., Kalker, T., Kim, H.J.: Digital Watermarking. In: 4th International Workshop, IWDW, Siena, Italy, September 15-17, Proceedings Series. Lecture Notes in Computer Science, vol. 3710 (2005)
3. Mitthelholzer, T.: An Information- Theoric Approach to Steganography and Watermarking. In: Pfitzmann, A. (ed.) IH 1999. LNCS, vol. 1768, pp. 1–17. Springer, Heidelberg (2000)
4. Golea, N.E.H., Seghir, R., Benzid, R.: A Bind RGB Color Image Watermarking Based on Singular Value Decomposition. In: IEEE/ACS International Conference on Computer Systems and Applications - AICCSA 2010, pp. 1–5 (2010)
5. Zhao, X., Ho, A.T.S.: An Introduction to Robust Transform Based Image Watermarking Techniques. In: Collection of Intelligent Multimedia Analysis for Security Applications, pp. 337–364 (2010)
6. Wong, P.W., Memon, N.: Secret and public key image-watermarking schemes for image authentication and ownership verification. IEEE Transactions on Image Processing 10(10) (2001)
7. Fridrich, J.: Security of Fragile Authentication Watermarks with Localization. In: Proc. SPIE, vol. 4675, pp. 691–700 (January 2002)
8. Lin, P., Huang, P., Peng, A.: A Fragile Watermarking Scheme for Image Authentication with Localization and Recovery. In: IEEE Sixth International Symposium on Multimedia Software Engineering (MSE 2004), Florida, USA (2004)
9. Tanenbaum, A.S.: Computer Networks, 4th edn. Pearson Education International, The Netherlands (2003)
10. Rivest, R.L., Shamir, A., Adleman, L.M.: A method for obtaining digital signatures and public-key cryptosystem. Communications of the ACM 21, 120–126 (1978)
11. Bellare, M., Rogaway, P.: Optimal Asymmetric Encryption – How to Encrypt with RSA. In: De Santis, A. (ed.) EUROCRYPT 1994. LNCS, vol. 950, pp. 92–111. Springer, Heidelberg (1995)

Maximum Likelihood Estimation, Interpolation and Prediction for Fractional Brownian Motion

Rachid Harba¹, Hassan Douzi², and Mohamed El Hajji²

¹ Laboratoire PRISME, Polytech'Orléans, Université d'Orléans 45067 Orléans, France
Rachid.Harba@univ-orleans.fr

² Laboratoire IRF-SIC, Faculté des sciences, Université Ibn Zohr, BP8106 Agadir, Maroc
{douzi_h, hajjimohmed}@yahoo.fr

Abstract. The maximum likelihood (ML) estimation approach for fractional Brownian motion (fBm) is explored in this communication. First, a ML based estimation of the H parameter is implemented on the signal itself. This approach on the signal itself can easily be applied on non-uniformly sampled data or directly useful in the case of incomplete data. Secondly, the method is extended to provide a ML prediction and a ML interpolation for fBm which could be of interest in many domains. Results also help to explain errors in other interpolating methods such as the midpoint displacement algorithm used to synthesize fBm data.

1 Introduction

Fractional Brownian motion (fBm) of H parameter in the range $]0 ; 1[$ is defined as an extension of Brownian motion [1]. One of the main issues when dealing with such data is to estimate the H parameter [2-3]. Among the numerous methods to achieve such a goal, the maximum likelihood (ML) approach proposed by Lundahl *et al.* [4] is often used due to its asymptotical efficiency [5]. It is also efficient in noisy environments [6]. But the ML based estimation of the H parameter is performed on the fBm increments which may be a limiting factor in some cases.

Here, we propose an ML estimate of the H parameter processed on the fBm itself. This allows direct extension of the method to include cases where there may be irregular sampling or incomplete data. Moreover, an ML based prediction and interpolation technique for a fBm signal easily result.

This communication is organized as follow. In the next section, fBm is defined and its main properties are derived. Then, the ML based estimation of the H parameter is achieved and is tested on exact fBm data. Finally, ML interpolation and prediction are presented and a real data example illustrates the methods.

2 FBM Properties

Continuous fBm of H parameter in $]0 ; 1[$, denoted $B_H(t)$, is defined as an extension of Brownian motion $B(t)$ [1]:

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H+\frac{1}{2})} \left\{ \int_{-\infty}^t (t-s)^{H-1/2} dB(s) - \int_{-\infty}^0 (-s)^{H-1/2} dB(s) \right\}. \tag{1}$$

Γ is the gamma function and when $H=1/2$, fBm reduces to Brownian motion.

From now on, we will focus on properties of discrete processes denoted $B_H[i]$ where i is a discrete time index. With a starting value $B_H[0]=0$, fBm is zero mean, Gaussian and second order non stationary as attested by its variance law deduced from (1):

$$\text{Var}(B_H[i]) = \sigma^2 i^{2H}. \tag{2}$$

Var is the variance operator, and σ^2 is the variance of fBm for the time index $i=1$. From (2) the autocorrelation function of the process follows [1]:

$$r_{B_H}[i, j] = E(B_H[i]B_H[j]) = \frac{\sigma^2}{2} (|i|^{2H} + |j|^{2H} - |i - j|^{2H}). \tag{3}$$

E is the expectation operator. Using time-frequency tools, it was shown that the averaged power spectral density of fBm is proportional to $|\omega|^{-1-2H}$ [7]. When considering discrete signals, there always will be aliasing problems.

fBm has no derivative, and thus its increments for a time lag m are of interest. They are named fractional Gaussian noises (fGn), denoted G_m , and defined as:

$$G_m[i] = B_H[i] - B_H[i - m]. \tag{4}$$

They are zero mean, Gaussian and stationary processes since their autocorrelation can be written as:

$$r_{G_m}[k] = E(G_m[i]G_m[i + k]) = \frac{\sigma^2}{2m^{2H}} (|k + m|^{2H} - 2|k|^{2H} + |k - m|^{2H}). \tag{5}$$

Without lost of generality, the case $m=1$ will be considered in the following. fGn for $m=1$ will be noted G_1 and its autocorrelation function derived from (5) becomes:

$$r_{G_1}[k] = \frac{\sigma^2}{2} (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}). \tag{6}$$

The following remarks regarding this equation will be useful in section 4 to explain some results.

For $H=0.5$, increments are uncorrelated and fGn is the white Gaussian noise process. For $H<0.5$, increments are negatively correlated. For $H>0.5$, they are positively correlated and the process is said to have long term memory since $r_{G_m}[k]$ decays hyperbolically with the lag k . It should be noticed that for $H\geq 1/2$ the function

$r_{G1}[k]$ is always nonnegative. This sequence is also decreasing and convex, *i. e.* second differences are positive. Finally, for $H < 1/2$, one gets $r_{G1}[k] < 0$ for any integer $k \neq 0$ [8].

3 ML H Parameter Estimation

There exist a lot of estimators of the H parameter [2-3]. Among all of these, the ML is of interest because of its asymptotical efficiency [5]. A ML estimation of the H parameter is developed in [4] based on the fGn. Here, we propose to perform it directly on the fBm data. First, let us define \mathbf{fBm} , the fBm vector composed of N samples. Since all the samples of \mathbf{fBm} are jointly Gaussian distributed, their likelihood function LF parameterised by H and σ^2 is:

$$\text{LF}(\mathbf{fBm}; H, \sigma^2) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{fBm}^T \mathbf{R}^{-1} \mathbf{fBm}\right). \quad (7)$$

N is the size of \mathbf{fBm} and \mathbf{R} is its $N \times N$ covariance matrix where each element $[\mathbf{R}]_{i,j}$ depends on the autocorrelation function as $[\mathbf{R}]_{i,j} = r_{BH}[i,j]$ as defined in equation (3).

The maximum of the log-likelihood function (LLF, the logarithm of equation 7) is to be found where constant terms are neglected:

$$\text{LLF}(\mathbf{fBm}; H, \sigma^2) = -\text{Ln} |\mathbf{R}| - \mathbf{fBm}^T \mathbf{R}^{-1} \mathbf{fBm}. \quad (8)$$

\mathbf{R} is first decomposed as $\sigma^2 \mathbf{R}'$ and the derivative of the LLF with respect to σ^2 is calculated. The value found by letting the derivative go to zero is inserted in the LLF and gives the final function to maximise with respect to H. The H estimator noted \hat{H} is:

$$\hat{H} = \underset{0 < H < 1}{\text{Max}} \left\{ -\text{Ln} |\mathbf{R}'| - N \text{Ln} \left(\frac{\mathbf{fBm}^T \mathbf{R}'^{-1} \mathbf{fBm}}{N} \right) \right\}. \quad (9)$$

As $\mathbf{fBm}[0]$ is zero by convention, the first row and column of \mathbf{R}' are all zeros and must not be considered, otherwise \mathbf{R}' is singular. The Gauss-Jordan elimination algorithm is used to compute the inverse and determinant of \mathbf{R}' .

This ML based estimator is tested on synthetic signals. In 1D, there exist two methods theoretically exact to synthesis fractional Brownian motion. The first one is the method based on the Choleski decomposition of the covariance function [4]. It requires high computational resources due to its complexity of $O(N^2)$. The second one is the circulant embedding method (CEM) [9]. Since based on the fast Fourier transform (FFT) algorithm, its complexity is only $O(N \log N)$. CEM is used in our experimental tests. 100 signals of 100 samples each were synthesised for three typical H values: $H=0.2$, $H=0.5$, and finally $H=0.8$. Mean \hat{H} values are compared to the true

H given during the synthesis of the reference signals. The standard deviations are also estimated. Results are shown in table 1.

Table 1. Mean \pm standard deviations (std) of the ML H estimators based on fBm in the first column. Computing are based on 100 synthetic signals of 100 samples each for H=0.2, 0.5 and 0.8. Second column shows results when a block of 100 unknown samples (indexed from 50 to 149) is added in the middle of each signal.

True H	a) \hat{H}	b) \hat{H}
	Mean \pm std	missing data Mean \pm std
0.2	0.197 \pm 0.048	0.205 \pm 0.048
0.5	0.496 \pm 0.060	0.501 \pm 0.059
0.8	0.796 \pm 0.058	0.797 \pm 0.058

The quality of this estimator can be studied. First, the bias of the estimates is low. A bilateral Student t test with a level of significance of 0.01 shows that these estimates are unbiased. In identical conditions, the bias could be as high as 0.3 for some other analysis methods [2]. The standard deviations of the estimates are close to the square root of the Cramer-Rao lower bound which is equal to 0.046, 0.059 and 0.057 for respectively H = 0.2, 0.5 and 0.8 for 100 samples [6]. An unilateral hypothesis test with a significance level of 0.01 shows that the variances of the estimates are equal to the respective Cramer-Rao lower bounds. These results show that in this case, the ML approach is efficient for data length as short as 100 samples.

This ML estimator can be easily used for non uniform sampling periods or when some samples are unknown. As an example, two blocks of 50 samples each are separated by 100 unknown samples for a fBm signal. The size of covariance matrix is 100 \times 100. Each element is computed using (3) where i and j are the position indexes of the known samples. Table 1 (b) shows the results. It can be noticed that the bias is still low and that the variance is nearly unchanged. Thus, an efficient ML estimate of the H parameter can be achieved for particular signals. Such cases arise when studying incomplete time series or for irregularly sampled 1D data.

4 ML Prediction and Interpolation

Two direct extensions of the above method can be derived, namely ML prediction and interpolation processed on the fBm signal itself. The prediction problem has been theoretically treated in [10] while the interpolation has not been considered. Here a practical study on true fBm data is carried out for prediction as well as for interpolation.

There are now three parameters to estimate: H, σ^2 and the value of the data to be found. The problem can be split into two parts: a ML H estimation is first carried out on the signal, then the value of the prediction or interpolation is computed with known H.

4.1 Prediction

Let $\mathbf{fBm}[N+x]$ for $x>0$ be a sample to predict given the H parameter and the first N samples of the vector. The theoretical correlation $r[i,N+x]$ between $\mathbf{fBm}[i]$ and $\mathbf{fBm}[N+x]$ with $1 \leq i \leq N$ is deduced from (3) with j being replaced by $N+x$. The covariance matrix \mathbf{R} of \mathbf{fBm} is now an $(N+1) \times (N+1)$ square matrix. It is identical to the one for the ML H estimation problem except that there are a row and a column added after respectively the last row and column to take into account the correlation between $\mathbf{fBm}[i]$ and $\mathbf{fBm}[N+x]$. \mathbf{R} can be decomposed as previously in $\sigma^2 \mathbf{R}'$. The final function is maximised with respect to $\mathbf{fBm}[N+x]$. This result can be seen as the mean prediction. The standard deviation easily follows based on the knowledge of the LLF.

We have tested this method on the same synthetic signals as previously described. Five samples of a typical realisation for $H = 0.2, 0.5$ and 0.8 are represented as shown in figure 1. Ten regularly spaced predictions are estimated after the last sample.

Remarks regarding equation (6) stated in section 2 are necessary to explain the results. For $H=0.5$, the ML mean estimate is equal to the last value of the signal. Indeed, as its increments are uncorrelated, the probability of an increase is equal to the probability of a decrease. For $H=0.8$, the ML estimate follows the trend of the past signal because increments are positively correlated. The shape looks similar to a polynomial prediction. For $H=0.2$, the estimate goes in the opposite direction because increments are negatively correlated. The standard deviation of the estimates follows a power law due to the fact that increments are zero mean with standard deviation proportional to the lag at the power H . This dependence on H is clearly seen in figure 1.

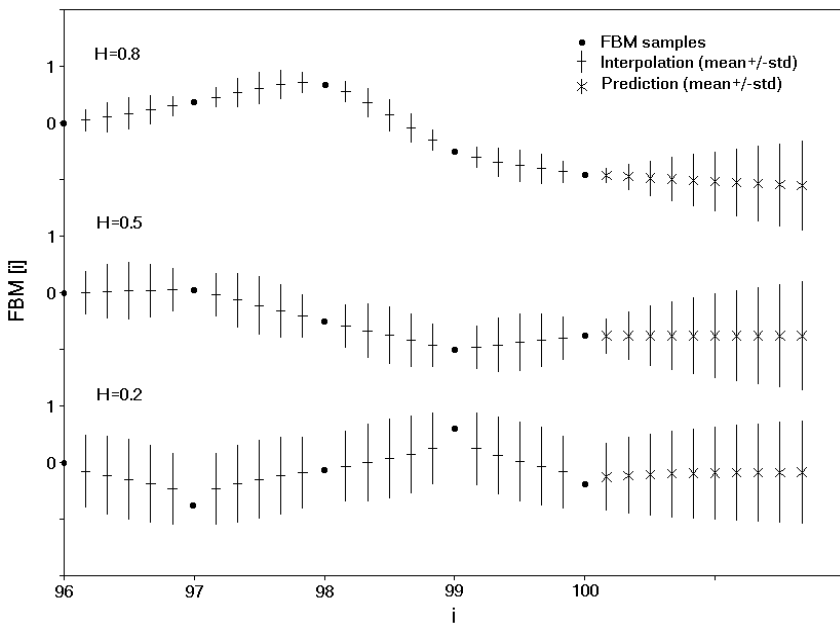


Fig. 1. Mean predictions and interpolations for a typical realisation of fBm for $H = 0.2, 0.5$ and 0.8 . The standard deviations of the estimates are also represented.

4.2 Interpolation

Given the H parameter and N samples of the observation, the value to estimate is now $fBm[k+x]$ for $1 \leq k < N$ and for $0 < x < 1$. The correlation between $fBm[i]$ and $fBm[k+x]$ is deduced from (3) with j being replaced by $k+x$. The covariance matrix \mathbf{R} is a $(N+1) \times (N+1)$ square matrix with a row and a column added between respectively the k and $k+1$ rows and columns. The same scheme as for the prediction process is applied. Figure 1 shows the results of five interpolations regularly spaced between each of the last five samples of the data.

For $H=0.5$, the ML mean interpolation is a linear interpolation. This can be explained from the prediction results. Indeed, interpolation can be seen as a weighted combination of a forward prediction (knowing the k first samples) and of a backward prediction (knowing the samples from the $k+1$ to the last one). Results for $H=0.2$ and $H=0.8$ can be identically explained. The standard deviations of the estimates are depending on H and on the distance from the nearest known sample.

4.3 Discussion

Prediction and interpolation for fractal signals can be applied to many real cases. One can mention financial domain to predict stock exchange or sub-pixel interpolation for fractal images. But, the above results enable a better understanding of conventional fBm synthesis techniques as the random midpoint displacement (MID) [11]. This iterative technique can be seen as a stochastic interpolation process. It consists in adding new points whose position along the horizontal axis is the middle of two adjacent points. The position on the vertical axis is given by a Gaussian random variable with mean equal to the average of the two adjacent points and with variance depending on H . It is known that it fails to provide true fBm signals when $H \neq 0.5$ [12]. Our results confirm this fact. For $H=0.5$, the mean ML position (a linear interpolation) is identical to the one given during the MID synthesis (an average). But, for other H values, it is not true. A solution to improve the MID synthesis method keeping the same scheme would be as follows: replace the random variable of the MID generating process by a new one with mean value and variance given by the ML interpolation for fBm as described above.

5 Conclusion

In this communication, we have presented ML approaches performed on the fBm signal itself. On reference fractal signals, it has been shown that the ML method gave efficient results. It also allows to measure the H parameter even when data are missing or when irregular sampling is present. Two direct extensions were derived concerning ML prediction and interpolation for fBm signals which could be of interest in many real cases. Results can be analysed taking into account the behaviour of the process for the various H values that were studied. They also explain

approximations of fractal synthesis methods such as the midpoint displacement method.

Future work will concern the synthesis of true 2D fBm images by using the new interpolation that is presented here. In addition the prediction of 1D signal as for stock exchange data will be a new and interesting application.

References

1. Mandelbrot, B.B., Van Ness, J.W.: Fractional Brownian Motion, Fractional Noises and Applications. *SIAM* 10(4), 422–438 (1968)
2. Gache, N., Flandrin, P., Garreau, D.: Fractal Dimension Estimators for Fractional Brownian Motion. In: *Proceedings of the ICASSP*, vol. 5, pp. 3557–3560 (1991)
3. Jennane, R., Harba, R., Jacquet, G.: Quality of Synthesis and Analysis Methods for Fractional Brownian Motion. In: *Proceedings of the IEEE Workshop on Digital Signal Processing*, pp. 307–310 (1996)
4. Lundahl, T., Ohley, W.J., Kay, S.M., Siffert, R.: Fractional Brownian Motion: A Maximum Likelihood Estimator and its Application to Image Texture. *IEEE Transactions on Medical Imaging* 5(3), 152–161 (1986)
5. Dahlhaus, R.: Efficient parameter estimation for self-similar processes. *The Annals of Statistics* 17, 1749–1766 (1989)
6. Hoeffler, S., Kumaresan, R., Pandit, M., Ohley, W.J.: Estimation of the Fractal Dimension of a Stochastic Fractal from Noise Corrupted Data. *Archiv fuer Electronic und Übertragungstechnik* 46(1), 13–21 (1992)
7. Flandrin, P.: On the Spectrum of Fractional Brownian Motions. *IEEE Trans. on Info. Theory* 35, 197–199 (1989)
8. Beran, J.: *Statistics for Long-Memory Processes*. Chapman & Hall (1994)
9. Perrin, E., Harba, R., Jennane, R., Iribaren, I.: Fast and exact synthesis for 1D fractional Brownian motion and fractional Gaussian noises. *IEEE Signal Processing Letters* 9(11), 382–384 (2002)
10. Gripenberg, G., Norros, I.: On the prediction for fractional brownian motion. *Journal of Applied Probabilities* 33, 400–410 (1996)
11. Peitgen, H.O., Saupe, D. (eds.): *The Science of Fractal Images*. Springer, New York (1988)
12. Mandelbrot, B.B.: Comment on Computer Rendering of Fractal Stochastic Models. *Communications of the ACM* 25, 581–583 (1982)

Gabor Filter-Based Texture Features to Archaeological Ceramic Materials Characterization

Mohamed Abadi¹, Majdi Khoudeir¹, and Sylvie Marchand²

¹ XLIM-SIC Department, UMR CNRS 6172, Chasseneuil-Futuroscope, France
{abadi, khoudeir}@sic.sp2mi.univ-poitiers.fr

² Institut français d'archéologie orientale, Cairo, Egypt
smarchand@ifao.egnet.net

Abstract. This paper presents a self-learning system for automatic texture characterization and classification on ceramic pastes or fabrics and surfaces. The system uses Gabor filter as pre-processing methods with feature extraction possibilities. On these features it applies a linear discriminant analysis (LDA) and k-nearest neighbor classifiers (k-NN) with its best parameters. Experimental results of the recognition ceramic materials, deals on the field and in the laboratory, for different ceramic pastes and surfaces show a good accuracy and applicability of the process on this type of data.

Keywords: Egyptian ceramic materials, ceramic fabrics and surface, texture characterization, feature extraction, classification algorithms.

1 Introduction

The history of archaeological Egyptian ceramics is an evolving discipline with new grid interpretations and pottery analysis, that archaeologists explore jointly taking two fundamental elements of ceramics study.

- The first element is the cultural aspect of the pottery vessel (archaeological context, chronology, shape, coating, decor, manufacturing techniques, function, etc.)
- The second element is the technical aspect of ceramic materials (characterization of the fabric: group designation for all the properties of the clay. The paste of the potter is a term for the plastic material from which the pots were made with the non plastic inclusions of mineral or organic origin [1])

The technical aspect constitutes a significant part of the pottery discovered at archaeological sites. It is considered as an identity card [1, 2, 3]. Now, the importance of material characterization to ceramics study is well established [1, 4]. The progress realized in this field during the last thirty years places the archaeological finds in several levels of analysis (production, consumption and distribution). For distribution levels, ceramic materials recognition is crucial because it shows the reconstruction of the traffic of archaeological objects in the inter-regional or international trading

routes, to allow us to distinguish between chronological and ethnic groups and to give some information on cultural relationships [5].

Traditionally, archaeologists examine the ceramic sherds on the field and thereafter in the laboratory [1, 5]. Generally, they use a binocular microscope to describe the fabric using a fresh break cut parallel to the vessel rim, and also the inner and outer surfaces of the sherd. This step represents the basis for any ceramic production classification to create groups of different fabrics [5, 6]. It is based on visual criteria (nature, size, shape, repartition of mineral and/or organic inclusions contained in the paste made by the potter, shaping methods, texture, color, methods of surface treatments, the firing of pottery). When it is necessary, the last levels for recording the properties of a pottery fabric are two main laboratory methods: the petrographic analysis (using thin sections) to identify the inclusions and chemical analysis to measure the chemical constituents of ceramic [2, 5]. Generally, these techniques are complementary and have two main goals:

- Establish and validate the classifications obtained on visual criteria with the microscope
- Characterize the ceramic fabric of the ceramic sherd in order to define the productions and to seek - where possible - the geographic origin

In some cases, however, the field and laboratory methods can provide contradictory conclusions. They therefore remain complex to interpret and there are uncertainties because every characterization process is done manually, by different archaeologists and under varying environments [6, 7]. In this context, archaeologists need to use machine vision to archive their ceramic materials [8, 9]. In this paper, we propose a solution based on texture analysis. Firstly, the ceramic paste or fabric on fresh break sherd and surface textures are described by texture characterization methods and secondly, obtained feature vectors are used to compare textures using several classification algorithms to allow the classification of ceramic materials.

2 Related Work

Texture analysis is the process to characterize and to classify different textures from the given images. It is considered as a key problem in a large variety of pattern recognition application areas, such as object recognition [10], industrial inspection [11], wood species recognition [12], rock classification [13] and so on. This kind of process requires the identification of proper features that differentiate the textures in the image for classification and recognition. In the real world, the images are often not uniform (changes in orientation, scale or other visual appearance) [14] and the extracted features are assumed to be uniform within the regions containing the same textures [14]. Several methods have been proposed in literature. Surveys of existing and comparative texture analysis may be found in Refs. [14-17].

The most important part of the historical ceramic materials classification is to define invariant features characterizing ceramic paste and surface textures, and which make possible a distinction between different kinds of ceramic materials. This

application is similar to rock texture analysis [13] but is more difficult. Unlike rock textures, ceramic paste and surface texture analysis is quite demanding. They are non-homogenous and not clear directional (when a vessel is shaped on a wheel, the inclusions like rod-shaped particules and straws follow the orientation and are parallel in the fresh break section to the rilling lines of the pot [1]). Also different granular size and other very small objects and straws can be integrated in some ceramic materials. [18-20] use texture analysis for quality control in ceramic tile production. Lindqvist and Akesson [21] present a literature review of image analysis applied to characterize rock structure and rock texture analysis. In Singh et al. [22], texture features for rock image classification are compared. The best performance was obtained by using co-occurrence matrices. Few research works on this few topic have been published. Smith et al. [23] use color and texture features based on well-known Scale Invariant Features Transform and we formulate a new feature based on total variation geometry for the reconstruction of archaeologically excavated ceramic fragments. Kampel and Sablatnig [7, 9] are developing an automated classification and reconstruction system for archaeological fragments based on shape and color information as a pre-classification process.

3 Archaeological Site

A set of macroscopic photographs made in IFAO laboratory by a device composed of :

- Reflex photo Kodak DCS-14 N camera.
- This camera is mounted on a Zeiss stemi 2000-C stereomicroscope.
- Schott KL 1500 LCD cold light source at a temperature of 3200 K.

They were obtained from ceramic materials classification of different fabrics representatives of Egyptian/local pottery production of three different sites. The samples cover a wide geographical area of Egypt (Marsa Matrouh, Abu Roach, different sites of Kharga oasis) and a large chronological period from the end of the Neolithic period (around 4800 BC) to the medieval period. The selected ceramic samples for this paper are derived from the Abu-Roach archaeological site which is located in Egypt in the Memphite area near the modern Cairo [24].

Abu-Roach, is an archaeological site¹ which is mainly known for Old Kingdom period (around 2500 BC) to be the pyramid complex of the king Djedefrê. The samples choose are the more representative from the classification of fabrics pottery link to the repertoire of pottery of this period found during the excavations. The vast majority of pottery vessels served domestic purposes (storage, preparation and consumption of food) and other types served ritual purposes (miniatures). But the site continues to be occupied after this period until the medieval period [3]. Few sherds from New Kingdom pottery (around 1300 BC), beginning of Ptolemaic pottery (around IVe century BC), and beginning of the Arabic period in the VIIe century AD have been also selected for this paper.

¹ <http://www.ifao.egnet.net/archeologie/abou-roach/>

The ceramic paste used in the manufacture of these objects includes the Nile fabrics or alluvial paste, the marl fabrics or marl clays (calcareous), and mixed clays fabrics (combination of marl and alluvial clays), and the foreign fabrics (Palestinian fabrics, Nubian fabrics, etc.). For this paper only Egyptian pottery sherds are presented here and the majority of these pots are locally produced in the Memphite area. If the fabrics are important to characterize and identify the pottery production, the ceramic surface too. For the selected samples from Abu Roach presented in this paper we found two main treatments of the ceramic surfaces (slip or not slip, and not slip with diffuse surface). For the not slip pots, the surface could be diffuse and present variation of colors on the surface, all this depends on the firing process.

4 Ceramic Pastes and Surfaces Characterization

Archaeologists consider that visual inspection is an important part of ceramic materials examination. However, this step remains complex because the ceramic pastes used in different manufacturing processes are extremely diverse and have heterogeneous composition. Similarly, ceramic surfaces of objects are often non-uniform. Thus automated visual inspection based on the machine vision system to ceramic materials classification requires the use of features and should describe the desired properties of ceramic pastes and surfaces.

4.1 Ceramic Texture Characterization Based Gabor Filter

In order to make an automated classification between different ceramic materials, some features have to be extracted, from ceramic pastes and surface textures. In this paper we try to apply a Gabor filters-based texture features. It has been successfully and widely applied to image processing, computer vision and pattern recognition [25, 26]. Its use is motivated by various factors. The characteristics of the Gabor filter, especially the frequency and orientation representations, are similar to those of the human visual system [27]. The statistics of these micro-features in a given region are often used to characterize the underlying texture information. In addition to accurate time-frequency location, they also provide robustness against varying brightness and contrast of images. A circular 2D Gabor filter in the spatial domain has the following general form [28]

$$G(x, y, \theta, u, \sigma) = \frac{1}{2\pi\sigma^2} e^{\left\{-\frac{x^2+y^2}{2\sigma^2}\right\}} \times e^{\{2\pi i(ux \cos \theta + uy \sin \theta)\}}$$

where $i = \sqrt{-1}$; u is the frequency of the sinusoidal wave; θ controls the orientation of the function and σ is the standard deviation of the Gaussian envelope. A filter bank of Gabor filters with various scales and rotations is created. In this work we have considered scales of 0, 2, 4, 6, 8, 10 and orientations of 0° , 45° , 90° and 135° . For each obtained response image we extract first three moments as features.

4.2 Ceramic Texture Classification

In image classification, the objects which are characterized by a feature vectors should describe the visual appearance of the texture or other attributes as accurately as possible. Generally, the features extracted are overlapping in the feature space and this makes the classification challenging. In this work, the classification algorithms are a supervised approach and a non-parametric discriminant analysis. In order to classify new samples we need a training set representing each category of ceramic pastes and/or surfaces. After wards, the linear discriminant analysis [29, 30] and the k-nearest neighbour, classifier [31] are applied on training set to estimate the optimal models. Finally, the tests set are assigned using these models. The selection of these classifiers is due to their robustness with homogenous and non-homogenous feature distributions of the ceramic paste and surface textures.

Linear Discriminant Analysis (LDA)

The linear discriminant analysis [29, 30] finds a transform matrix W , such that

$$W = \underset{W}{\operatorname{argmax}} \frac{W^T S_B W}{W^T S_W W}$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix, defined as

$$S_B = \sum_{i=1}^c N_i (x_i - \mu)(x_i - \mu)^T$$

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

In these expressions, N_i is the number of training samples in class i , c is the number of distinct classes, μ_i is the mean vector of samples belonging to class i and X_i represents the set of samples belonging to class i .

k-Nearest Neighbour Classifier (k-NN)

The k-NN classifier is very simple to understand and easy to implement. It is based on a distance function such as the Euclidean, City block, Cosine distance, Pearson's correlation or so on. These functions are computed for pairs of samples in N -dimensional space (number of features). Each sample is classified according to the class memberships of its k nearest neighbours, as determined by the distance function. k-NN has the advantages of simple calculation and the ability to perform well on data sets that are not linearly separable, often giving better performance than more complex methods in many applications [31]. Given the training feature dataset $X = \{x_1, x_2, \dots, x_n\}$, and a test feature vector x , we will find the distance function and the k nearest neighbors to the test feature vector, where each nearest neighbor has one vote for the class label c . The test label will base on the majority of the votes according cross validation accuracy (10-fold) to select the best parameters for k-nearest neighbor classifier.

5 Results

In this paper, we experiment with the texture analysis process to classify Egyptian ceramic materials. These materials are described by images representing fresh break section of the sherd, inner and outer surface view (figure 1).

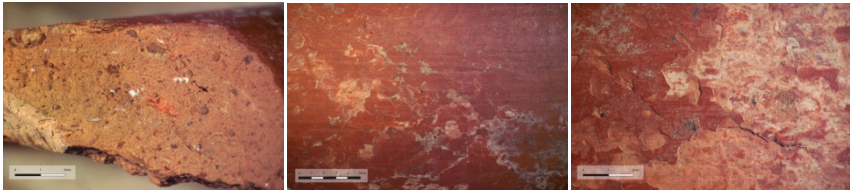


Fig. 1. Images representing fresh break section of the sherd, inner and outer surface (Abu-Roach site: sample n°1762)

In fact, the samples are manually classified by archeological experts based on the kind of ceramic fabrics and treatments of the surfaces. Table 1 shows the different categories and the sample's number in each category. Each sample is labeled by a value coded on three digits. All digits are ranging from 1 to 3). The First digits indicate the nature of samples (rupture, inner or outer view respectively 1, 2, 3). The second digits show the used characteristics (ceramic pastes or ceramic surface. The digit is 1 or 2 respectively. Finally, the third digits describe the nature of categories (1 for marl clay or Slip surface, 2 for alluvial clays or diffuse/not slip surface and 3 for mixed-clays or not-slip surface). Table 2 represents an example of three labeled samples. In this example, we can easily observe that the samples 1790 and 1771 have the same ceramic paste and their ceramic surface is different.

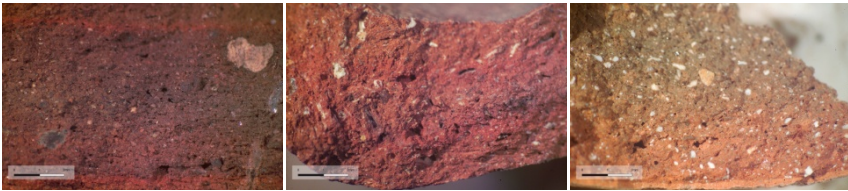
Table 1. Numbering of samples for different Abu-Roach ceramic categories

	Categories	Numbering of samples
Ceramic pastes fabrics	Marl (Ma)	1769, 1785, 1788, 1800, 1764, 1765, 1767, 1784 1791, 1795, 1798, 1777, 1796, 1797, 1761, 1762, 1772, 1779, 1780, 1781, 1782, 1783
	Alluvial (Al)	1763, 1790, 1774, 1775, 1776, 1778, 1787, 1789, 1771, 1792, 1794, 1768, 1793, 1766
	Mixed-clay (Mi)	1770, 1786, 1799, 1773
	Slip (S)	1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768, 1769, 1770, 1773, 1775, 1776, 1777, 1787, 1790, 1793, 1794, 1795, 1798, 1799, 1800
Ceramic surfaces	Diffuse/not slip (D)	1796, 1797
	Not-Slip (NS)	1774, 1771, 1772, 1778, 1779, 1780, 1781, 1782, 1783, 1784, 1785, 1786, 1788, 1789, 1791, 1792

Table 2. Samples labeled according ceramic pastes and surfaces

Samples	Ceramic paste	Ceramic surface
1790	1-1-2	1-2-1
1771	2-1-2	2-2-3
1796	3-1-1	3-2-2

Now, in each sample we have several images which represent both ceramic paste (fresh break section) and surfaces. Each image is labeled according to its belonging sample defined by archaeologist experts. Therefore, we have a database composed of 599 images with resolution 4500×3000 pixels. From each of these images, we have extracted four representative sub-images of size 512×512 pixels. Thus a new database is formed and it contains 2396 sub-images. The sub-images are distinguished by their origin, ceramic pastes or surfaces properties and other archaeological criteria. Fig. 2 shows an example of homogenous, non-homogenous and non-directionality textures.

**Fig. 2.** Textures corresponding to the samples in table 2

Once ceramic pastes and surfaces databases are defined, the Gabor filter-base texture features are computed in each sub-image. The obtained feature vectors are divided in training and test set using K-cross-validation method ($K = 10$). In order to obtain significant and correct statistical values, this operation is repeated 100 times. To study the effects of the feature extraction method on ceramic materials recognition by applying LDA and k-NN classifiers, we computed a better model for the training set. The influence of changing parameters can be assessed through examining the classification accuracy. The implementation of this procedure was performed in a batch mode. The best models or parameters retained are then used to predict the association of each pixel to adequate class. Accuracy assessment of four classification figures was performed by computing overall and category by category user's and producer's (Figures 3) accuracy using the validation dataset [32]. In fact, for k-NN classifier, we choose $k \in \{1; 2; \dots; 15\}$ and we have used four different distance measures: Euclidean, City block, Cosine distance and Pearson's correlation to study the effect on classification accuracy.

Figures 3 shows the performance of Gabor filter-based texture features to characterize ceramic pastes and surfaces separately. Regarding the overall classifications accuracies results (first tow bars), we can conclude that k-NN classifier have produced a best results and they show the quite similar overall accuracy

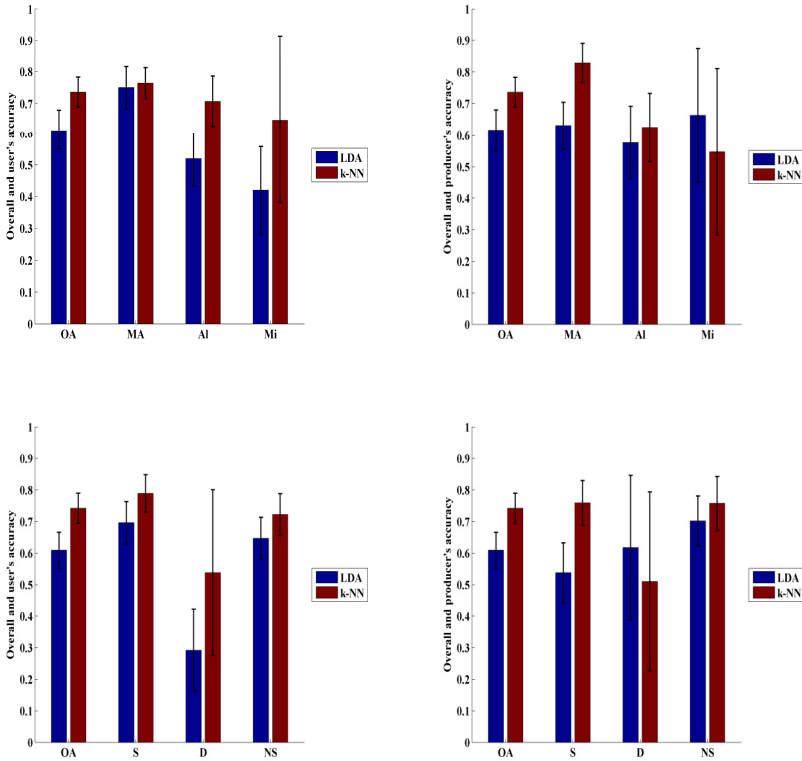


Fig. 3. Overall accuracy. By column: user's and producer's accuracy. By row: ceramic pastes and surfaces

between ceramic pastes and ceramic surfaces. Indeed, the best values obtained for LDA and k-NN are respectively (0.615 ± 0.065) and (0.736 ± 0.047) for ceramic pastes and (0.609 ± 0.057) and (0.742 ± 0.048) for ceramic surfaces. At Single category level, we observe a same results, k-NN classifier performed are better than LDA, specifically in user's accuracy. In producer's accuracy, the LDA return best results for mixed paste (Mi) and diffuse surface (D). For the classification of our complex data sets and when we choose good parameters, Gabor filter-based texture features and k-NN classifier appear to be the best classifier and texture features options because they can be used with any data. When we compare the accuracy classifications between ceramic pastes and ceramic surface separately, we observe that accuracy is differing from one category to another. For example, see results returned by LDA in figure 3. The LAD returns a lower value of user's accuracy (0.292 ± 0.129) from diffuse ceramic surface (D). To improve final classification results and take advantage of each single category best classification through combination between ceramic pastes and surfaces results could be used to improve results. Therefore, the results have shown that machine vision based on image texture analysis for Egyptian ceramic

materials classification can constitute a cost-effective approach for a characterization, assessment and archiving archaeological finds. Due to the good recognition and the existing complementarities between ceramic pastes and surfaces, using Gabor filter-based texture features and k-NN classifier should be confirmed by applying this process in other Egyptian archaeological sites.

6 Conclusion

In this paper, a texture analysis process based on Gabor filter-to texture feature extraction and classification algorithms (LDA, k-NN), to classify and analysis a non-homogenous Egyptian archaeological ceramic textures were proposed. In general, this is a difficult classification task due to strong homogeneities within samples in the same category. In fact, to classify the Abu-roach archaeological database, the ceramic materials are characterized separately by their ceramic paste and surface textures. The results are shown that texture analysis yields promising accuracy. It leads to an effective process for Egyptian ceramic materials recognition.

References

1. Arnold, D., Bourriau, J.: *An Introduction to Ancient Egyptian Pottery*, Mayence, p. 162 (1993)
2. Picon, M.: L'apport du laboratoire dans les identifications de céramiques. In: dans Lévêque, P., Morel, J.-P. (éd.), *Céramiques Hellénistiques et Romaines III*, Paris, pp. 9–30, (2001)
3. Marchand, S.: Abou Rawash à la IV^e dynastie. Les vases en céramique de la pyramide satellite de Rêdjedef. In: dans Rzeuska, T.I., Wodinska, A. (éd.), *Studies on Old Kingdom Pottery*, Varsovie, pp. 71–94 (2009)
4. Sinopoli, C.M.: *Approaches to Archaeological Ceramics*, New York (1991)
5. Bourriau, J., Nicholson, P.T., Rose, P.A.J.: Pottery. In: Nicholson, P.T., Shaw, I. (eds.) *Ancient Egyptian Materials and Technology*, pp. 121–147. Cambridge University Press (2000)
6. Orton, C., Tyers, P., Vince, A.: *Pottery in archaeology*. Cambridge University Press
7. Adler, K., Kampel, M., Kastler, R., Penz, M., Sablatnig, R., Schindler, K., Tosovic, S.: Computer Aided Classification of Ceramics - Achievements and Problems. In: Börner, W., Dollhofer, L. (eds.) *Proc. of 6th Intl. Workshop on Archaeology and Computers*, Vienna, Austria, pp. 3–12 (2001)
8. Sablatnig, R., Menard, C.: Computer based Acquisition of Archaeological Finds: The First Step towards Automatic Classification. In: Moscati, P., Mariotti, S. (eds.) *3rd International Symposium on Computing and Archaeology*, Rome, vol. 1, pp. 429–446 (1996)
9. Kampel, M., Sablatnig, R.: An Automated Pottery Archival and Reconstruction System. *Journal of Visualization and Computer Animation* 14(3), 111–120 (2003)
10. Bileschi, S., Wolf, L.: A Unified System for Object Detection, Texture Recognition and Context Analysis Based on the Standard Model Feature Set. In: *British Machine Vision Conference, BMVC* (2006)
11. Mitchell, T.A., Bowden, R., Sarhadi, M.: Efficient texture analysis for industrial inspection. *International Journal of Production Research* 38(4), 967–984 (2000)

12. Tou, J.Y., Tay, Y.H., Lau, P.Y.: A Comparative Study for Texture Classification Techniques on Wood Species Recognition Problem. In: ICNC, vol. 5, pp. 8–12 (2009)
13. Lepisto, L., Kunttu, L., Autio, J., Visa, A.: Rock Image Classification Using Non-Homogenous Textures and Spectral Imaging. In: WSCG (2003)
14. Tuceryan, M., Jain, A.K.: Texture Analysis, 2nd edn. The Handbook of Pattern Recognition and Computer Vision, pp. 207–248. World Scientific Publishing, Singapore (1998)
15. Yap, W.H., Khalid, M., Yusof, R.: Face Verification with Gabor Representation and Support Vector Machines. In: AMS, pp. 451–459 (2007)
16. Salem, Y.B., Nasri, S.: Texture Classification of Woven Fabric Based on a GLCM Method and using Multiclass Support Vector Machine. In: SSD (2009)
17. Chen, C.C., Chen, C.C.: Filtering methods for texture discrimination. *Pattern Recognition Lett.* 20, 783–790 (1999)
18. Tsai, D.M., Huang, T.Y.: Automated surface inspection for statistical textures. *Image and Vision Computing* 4(21), 307–323 (2003)
19. Jahabin, S., Bobik, A.C., Perez, E., Nair, D.: Automatic inspection of textured surfaces by support vector machines. In: International Congress SPIE - Proceedings (2009)
20. Lebrun, V., Macaire, L.: Aspect inspection of marble tiles by colour line scan cameras. In: Proc. of the Int. Conf. on Quality Control by Artificial Vision, vol. 2, pp. 403–408 (2001)
21. Lindqvist, J.E., Akesson, U.: Image analysis applied to engineering geology, a literature review. *Bulletin of Engineering Geology and the Environment* 60(2), 117–122 (2001)
22. Singh, M., Javadi, A., Singh, S.: A Comparison of Texture Features for the Classification of Rock Images. In: Yang, Z.R., Yin, H., Everson, R.M. (eds.) IDEAL 2004. LNCS, vol. 3177, pp. 179–184. Springer, Heidelberg (2004)
23. Smith, P., Bepalov, D., Shokoufandeh, A., Jeppson, P.: Classification of archaeological ceramic fragments using texture and color descriptors. In: Computer Vision and Pattern Recognition Workshops, pp. 49–54 (2010)
24. Lehner, M.: *The Complete Pyramids*, p. 120. Thames and Hudson, London (1997)
25. Adini, G.Y., Moses, Y., Ullman, S.: Face recognition: the problem of compensation for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* 19(7), 721–732 (1997)
26. Dunn, D., Higgins, W.E.: Optimal Gabor filters for texture segmentation. *IEEE Trans. Image Process.* 4(4), 947–964 (1995)
27. Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2(7), 1160–1169 (1985)
28. Jain, A., Healey, G.: A multiscale representation including opponent color features for texture recognition. *IEEE Trans. Image Process.* 7(1), 124–128 (1998)
29. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigen faces vs. Fisher faces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19(7), 711–720 (1997)
30. Swets, D.L., Weng, J.: Using discriminant eigen features for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18(8), 831–836 (1996)
31. Dudoit, S., Fridlyand, J.: Speed TP: Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data. Technical report 576, Mathematical Sciences Research Institute, Berkeley, CA (2000)
32. Congalton, R., Green, K.: Assessing the Accuracy of remotely Sensed Data: Principles and Practices, p. 137. CRC/Lewis Press, Boca Roton

RGB Color Distribution Analysis Using Volumetric Fractal Dimension

Dalcimar Casanova* and Odemir Martinez Bruno

USP - Universidade de São Paulo
IFSC - Instituto de Física de São Carlos, São Carlos, Brasil
dalcimar@gmail.com, bruno@ifsc.usp.br

Abstract. Over the years many approaches for texture analysis have been proposed. Most of these methods use, directly or indirectly, the spatial information to build the features. Although the spatial distribution of gray levels is a property *a priori* of the texture, some methods do not use this propriety to characterize it. The problem is that this class of methods has, generally, worst results than first one. Thus, in this work we propose a new method to classify color textures that does not use any type of spatial distribution information and still achieves high classification rates, comparable, if not better, than traditional texture analysis methods. The method is based on analysis of RGB color distribution using volumetric fractal dimension.

1 Introduction

The identification of visual patterns in images or objects is a key process in computer vision area. And, among the set of possible patterns, the texture is one of the most useful for experiments of image classification and identification.

Although there is no precise definition of texture, this attribute is easily perceived by humans being a rich source of visual information. However, while the ability of a human to distinguish different textures is apparent, the automated description and recognition of these same patterns has proved to be quite complex.

Many methods of texture analysis have been proposed recently, most of these methods use, directly or indirectly, the spatial distribution of gray levels to build the features. In statistical approaches, the greatest number of methods uses any information of the likelihood of neighboring pixel values (e.g. GLDM [18], GLCM [4]). The geometrical-based methods have an desirable property in defining local spatial neighborhoods (e.g. Voronoi tessellation features [15]). In model-based methods, such MRF [2], assume that the intensity at each pixel in the image depends on the intensities of only the neighboring pixels. And in signal processing methods, the frequency of an texture is determined by spatial distribution of the pixels [13]).

Although the spatial distribution of gray levels is a property *a priori* of the texture, some methods do not use this propriety to characterize it. First-order

* Dalcimar Casanova gratefully acknowledges the financial support FAPESP (São Paulo Research Foundation, Brazil) (2008/57313-2) for his PhD grant.

statistics, for example, measure the likelihood of observing a gray value at a randomly chosen location in the image. First-order statistics can be computed from the histogram of pixel intensities in the image. These depend only on individual pixel values and not on the interaction or co-occurrence of neighboring pixel values. The problem is that this class of methods are generally, not competitive against the methods that use some spatial distribution information.

So, in this work, we propose a new method that does not use any type of spatial distribution information to classify color textures. They are based on the color distribution analysis over RGB color model. This analysis is made with volumetric fractal dimension and posterior classification with LDA and Bayesian classifier. The results are best than methods of same class and very competitive against another methods with neighborhood relationship.

The rest of the paper is organized as follows. Section 2 presents the general methodology for RGB color cube transform, the volumetric fractal dimension and the classification procedure. Section 3 describes the experimental results and the comparison with others methods, while conclusions are presented in Section 4.

2 Materials and Methods

2.1 RGB Color Cube Transform

The first step of the proposed method is mapping the existing colors of texture in a cube represented by RGB coordinates (i.e. red, green and blue colors). Given an texture image $I(x, y)$, an cube $C(r, g, b)$ (that have the r-axis representing red values, g-axis as green values and b-axis as blue values), and the colors in I defined by three components, the $C(r, g, b)$ will be an function as follows:

$$C(r, g, b) = \begin{cases} 1, & \text{if } \exists I(x, y) = (r, g, b) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

In proposed method we use each axis in the range 0 to 255, representing 8-bit per channel. The basic idea of this transformation is summarize all existent colors in texture and map that in a cube, as is done in 3D color histogram representation, but without counting the number of image pixels in each bin. In Fig. 1 we show 3 different textures that represent this transformation. Note that the spacial distribution of the colors is different for each class. In the next session we will explore and quantify this propriety by use of volumetric fractal dimension.

2.2 Volumetric Fractal Dimension

Benoit Mandelbrot, in 1970s, introduced a new field of mathematics, named Fractal Geometry. He said that complex objects are generated by the interaction of simple rules and has non-integer dimension, which is related to its complexity. Since then many methods have been developed to estimate the fractal dimension of an given object, one of the most accurate is the Bouligand-Minkowski [14].

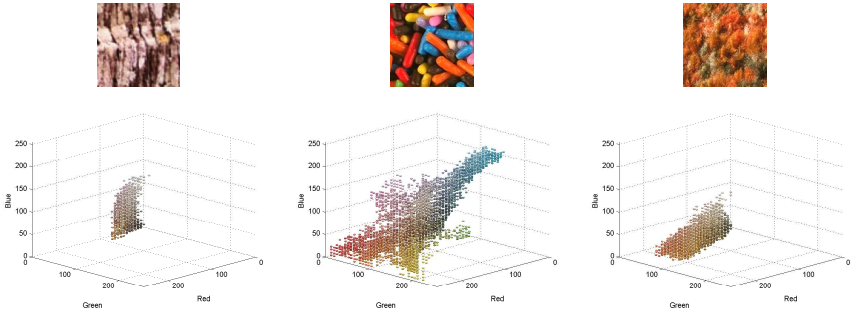


Fig. 1. Textures examples (above) and your respective RGB cube transforms (below). The main idea is summarize all existent colors in texture in a cube that represent RGB coordinates.

Since then some works of texture analysis has been made with fractals [7]. Recently [1] proposed a new texture descriptor based on fractals. In this, the texture is mapped to a cube, and the descriptor explores the differences in influence area of the 3D object formed. The same process will be used here to characterize the spacial distribution of the colors in our RGB color cube transform.

Given our 3D cube $C(r, g, b)$, we can obtain $V(r)$ through dilation of each point p of C using a sphere of radius r . This $V(r)$ is the influence volume of the object for a given radius r , and is very sensitive to structural changes of the object. As we have different objects, with different structures in your RGB color cube transform, this methodology is suitable to characterize it. The dilation curve expressed as volume $V(r)$ as a function of the dilation radius r_{max} is given by:

$$V(r) = \{p \in R^3 | \exists p' \in S : |p - p'| \leq r\} \tag{2}$$

In this method the arrangement of points in C alters the process of dilation. As the value of r grows, the spheres produced by the different points of object begin to interact. This interaction causes effects in $V(r)$, thus each object produces an characteristic growth of $V(r)$ and this makes possible the use of the values of $V(r)$ as descriptors. Thus, the feature vector \mathbf{x} is defined as the set of logarithm of influence volumes $V(r)$ calculated for all values of $r \in E$, where E is the set of possible Euclidean distances for a radius r_{max} :

$$E = \{1, \sqrt{2}, \sqrt{3}, \dots, r_{max}\} \tag{3}$$

$$\mathbf{x} = [\log V(1), \log V(\sqrt{2}), \dots, \log V(r_{max})] \tag{4}$$

The Fig. 2 exemplifies this process of dilation for different values of r . In order to complete the estimation of Bouligand-Minkowski fractal dimension we need plot the $\log V(r)$ versus $\log r$. The value of inclination of the straight line obtained gives us an estimative of the fractal dimension of the respective object (Equation 5).

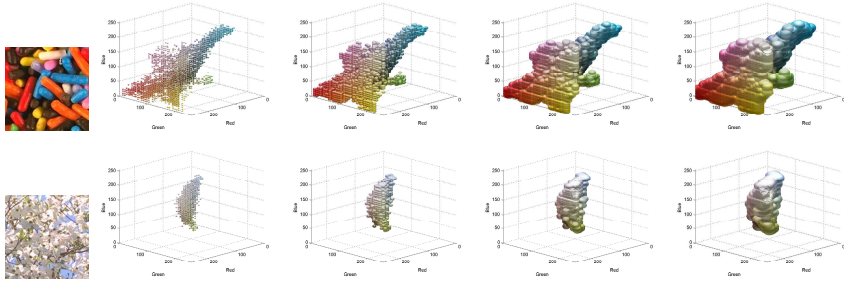


Fig. 2. Dilation process for two different textures. This allows an analysis of color distribution by the volume calculated.

$$D = 3 - \lim_{r \rightarrow 0} \frac{\log V(r)}{\log(r)} \tag{5}$$

The $V(r)$ can be calculated by using some fast Exact Distance Transform (EDT) algorithms [5, 9, 3]. A important characteristic is that only one parameter need be chosen, the r_{max} .

2.3 Data Analysis

A widely type of techniques are used for data analysis in a supervised multiclass classification task. Due the values of $V(r)$ are naturally dependent and highly correlated we opt by use the Linear Discriminant Analysis (LDA)+Bayesian classifier. This supervised task is performed under a 10-fold cross-validation scheme.

Linear Discriminant Analysis. Basically, LDA applies an geometric transformation (rotations) to the feature space with the purpose of generating new uncorrelated features based on linear combinations of the original ones, aiming seek a projection that best separates the data. Given the matrix S , indicating the total dispersion among the feature vectors, defined as:

$$S = \sum_{i=1}^N (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)' \tag{6}$$

and the matrix S_i indicating the dispersion of objects of C_i :

$$S_i = \sum_{i \in C_i} (\mathbf{x}_i - \mu_i)(\mathbf{x}_i - \mu_i)' \tag{7}$$

we can define the intra-class variability S_{intra} (indicating the combined dispersion within each class) and interclass variability S_{inter} (indicating the dispersion of the classes in terms of their centroids) as:

$$S_{intra} = \sum_{i=1}^K S_i \tag{8}$$

$$S_{inter} = \sum_{i=1}^K N_i(\mu_i - \mu)(\mu_i - \mu)' \tag{9}$$

where K is the number of classes, N , the number of samples, N_i , the number of objects in class i , C_i , the set of samples of class i , μ , the global average, and μ_i , the average of objects in class i . For these measures of dispersion we have necessarily:

$$S = S_{intra} + S_{inter} \tag{10}$$

Thus, the i -th canonical discriminant function is given by:

$$Z_i = a_{i1}\mathbf{X}_1 + a_{i2}\mathbf{X}_2 + \dots + a_{ip}\mathbf{X}_p \tag{11}$$

where p is the number of features of the model and a_{ij} are the elements of the eigenvector $a_i = (a_{i1}, a_{i2}, \dots, a_{ip})$ of matrix C given by:

$$C = S_{inter} * S_{intra}^{-1} \tag{12}$$

This formulation leads to a condition where there is no correlation between Z_i and Z_1, Z_2, \dots , within the classes. From p -original variables the p -principal components can be obtained. However, in general, a reduction in the number of variables to be assessed is desired, i.e., the information contained in the p -original variables be replaced by the information contained in $k(k < p)$ uncorrelated principal components. Thus, the system of random variability of the original vector with p -original variables is approximated by the variability of the random vector containing the k -principal components.

Bayesian classifier. The Bayesian classifier is based on the Bayesian decision theory and combines class conditional probability densities (likelihood), and prior probabilities (prior knowledge), to perform classification by assigning each object to the class with the maximum a posteriori probability. For g groups, the Bayes rule assigns an object to the group i when:

$$P(i|\mathbf{x}) > P(j|\mathbf{x}), \text{ for } \forall j \neq i \tag{13}$$

In this case, assuming the hypotheses of independence, we have for the random variables:

$$P(i|\mathbf{x}) = \frac{P(i) \prod_{k=1}^n P(x_k|i)}{\prod_{k=1}^n P(x_k)} \tag{14}$$

where:

$$P(x_k|i) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{(x_i - \mu_{ik})^2}{2\sigma_{ik}^2}} \tag{15}$$

being $P(\mathbf{x}|i)$ the probability of obtaining a particular set of features \mathbf{x} , given that the object belongs to the group i and $P(i)$ is the *a priori* probability, i.e. the probability of choosing the group i without known any feature of the object.

2.4 Database

The experiments are performed over VisTex color textures database [17]. This database is maintained by the Vision and Modeling group at the MIT Media Lab. The full database contains images representative of real-world textures under practical conditions (lighting, perspective, etc.). In this work the 54 images of resolution 512×512 were split into 16 non-overlapping sub-images of 128×128 . These images are available on de Outex site as test suite Contrib_TC_00006 [10].

3 Results and Discussion

In order to evaluate the quality of proposed method we set, based on work of [1], the $r_{\max} = 20$, totaling 335 successive dilations. Additionally we make a uniform quantization of the image I using a color map with 65536 colors. The source image I is quantized by matching colors with the nearest color in the color map. This procedure aims decrease the number of color in source images.

The Table 1 shows the result for the proposed method in Vistex database. The 95.25% of accuracy demonstrates the high quality of the proposed method. This results use all features between $r = 5$ and $r = 20$, totaling $313 \log V(r)$ features. We do not use the first's radius because they not contain relevant features. It is due the quantization used, that separates the possible colors points on RGB color cube transform (i.e. the initial dilation of the points does not have any interaction with other points due the distance). More studies about the ideal quantization and ideal r_{\max} parameter need be made in other databases. The high accuracy obtained impedes this research here, since several parameters will reach a high accuracy.

Obviously, this method needs be tested in more hard conditions, such different illuminations conditions and different acquisition devices. The RGB is a device-dependent color model, i.e. different devices detect or reproduce a given RGB value differently, since the color elements (such as phosphors or dyes) and their response to the individual R, G, and B levels vary from manufacturer to manufacturer, or even in the same device over time. Thus an RGB value does not define the same color across devices.

However, if this approach does not work with these difficulties, many alternatives to solve these problems are known. The use of color management systems are an alternative, but not always available. Apply this same methodology over other color spaces, such HSV, are another possibility.

3.1 Comparison with Methods That Do Not Use Spatial Information

In order to evaluate the quality of proposed method against another approaches, we compare it with 3 another methods of same class, i.e. who do not use information about spatial distribution of pixels to build the features. They are:

Table 1. Comparison between methods that do not use the spatial distribution of gray levels as features

Method	No. of descriptors	No. of images correctly classified	Success rate %
VFD RGB cube	313	823	95.25
Histogram ratio	uncertain	484	56.02
Chromaticity	25	599	69.32
First-order	18	777	89.93

- Histogram ratio features [12]: this method utilizes an the 3-D xyY color histogram of a given image to calculate the self-relative histogram ratio features. The number of features varies from class to class since it depends on how many common histogram bins exist among each class.
- Chromaticity moments [11]: The method uses the CIE xy chromaticity diagram of an image and a corresponding set of two-dimensional and three-dimensional moments to characterize a given color texture. We used the 5T-type + 5D-type moments (CM55), totaling 25 features.
- First-order statistics of RGB channels [16]: Given the image, simple statistics as mean, variance, skewness, kurtosis, energy and entropy are calculated of each RGB channels, totaling 18 features.

The table 1 show the results. We can see the superior quality of the proposed approach, since the closest result is the first-order method, with 89.93 of accuracy. Is important to say that the work of [12] show an accuracy of 96.36% in Vistex database. However the confection of the database is another. He perform the experiments in a set of 164 color textures images of size 128×128 , where he draws randomly from each image a subsample of 100×100 . This result in a database where all samples of same class are very similar, unlike of de Vistex database used here. Due this, very bad results are reached by this method. The same problem occurs with [11] work.

3.2 Comparison with Methods That Use Spatial Information

The most methods of texture analysis use the spatial information directly (e.g. GLCM) or indirectly (e.g. Gabor filters) to build they features. We will compare our methodology with them too. The configuration used in these methods is presented below.

- Gabor filters [6]: is, basically, a bi-dimensional Gaussian function modulated with an oriented sinusoid. The convolution of the image with the family of Gabor filters in different scales and rotations produce the features. We use 64 filters (8 rotations and 8 scale filters) with lower and upper frequency equal to 0.01 and 0.4, respectively. The individual parameter of each filter is defined by [8].

- Gray Level Co-occurrence Matrix (GLCM) [4]: they are the joint probability distributions between pairs of pixels at a determined distance and direction. For this comparison, distances of 1 and 2 pixels with angles of -45° , -90° , 45° , 90° were used. Contrast, correlation, energy and homogeneity measures are computed from resulting matrices, totalizing a set of 32 descriptors. A non-symmetric version has been adopted in experiments.

The Table 2 shows the results. Despite the difference in methodologies, the 95.25% of accuracy, against 94.44% of Gabor filters and 92.47% of GLDM, is a very impressive result for a method that use only color information to build their characteristics. Moreover, since each method explores different texture characteristics, the use of our approach in conjunction with traditional methods is quite possible. The Gabor filters or GLDM, for example, uses the spatial distribution of pixels to characterize the texture and, the proposed methodology explores, basically, the color distribution, they are complementary information.

For all results (except histogram ratio that use classification scheme of [12]), we use LDA analysis in a 10-fold cross-validation. We summarize the original features into principal components that represent 99.99% of total variance explained. The new features, also called canonical features, are then used in the Bayesian classifier. Thus, although our approach has the largest number of feature between all tested methods, the number of canonical features obtained after dimensionality reduction with LDA is very similar.

Table 2. Comparison between proposed method and traditionally methods that use spatial information

Method	No. of descriptors	No. of images correctly classified	Success rate %
VFD RGB cube	313	823	95.25
GLCM	32	799	92.47
Gabor filter	64	816	94.44

4 Conclusion

A simple and efficient method for color texture classification has been presented. The called RGB color cube transform map the existing colors of texture in a cube, and the volumetric fractal dimension uses these color distribution information to build the features.

A comparison with several methods are performed and, although do not use any type spatial information the proposed method achieves high classification rates, comparable, if not better, with traditional texture analysis methods. Further research will investigate the ideal quantization and the optimal r_{\max} parameter, and will also examine the performance in other databases with different illumination sources and acquisition devices.

References

1. Backes, A.R., Casanova, D., Bruno, O.M.: Plant leaf identification based on volumetric fractal dimension. *International Journal of Pattern Recognition and Artificial Intelligence* 23(6), 1145–1160 (2009)
2. Chellappa, R., Chatterjee, S.: Classification of textures using gaussian markov random fields. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33(4), 959–963 (1985)
3. Fabbri, R., da F. Costa, L., Torelli, J.C., Bruno, O.M.: 2D euclidean distance transform algorithms: A comparative survey. *ACM Computing Surveys* 40(1), 1–44 (2008)
4. Haralick, R.M.: Statistical and structural approaches to texture. *Proceedings of IEEE* 67(5), 786–804 (1979)
5. Hirata, T.: A unified linear-time algorithm for computing distance maps. *Information Processing Letters* 58, 129–133 (1996)
6. Jain, A.K., Farrokhnia, F.: Unsupervised texture segmentation using gabor filters. *Pattern Recognition* 24(12), 1167–1186 (1991)
7. Keller, J.M., Chen, S., Crownover, R.M.: Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and Image Processing* 45(2), 150–166 (1989)
8. Manjunath, B.S., Ma, W.-Y.: Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8), 837–842 (1996)
9. Meijster, A., Roerdink, J.B.T.M., Hesselink, W.H.: A general algorithm for computing distance transforms in linear time. In: *Proceedings of the 5th International Conference on Mathematical Morphology and its Applications to Image and Signal Processing*, pp. 331–340 (2000)
10. Ojala, T., Mäenpää, T., Pietikäinen, M., Viertola, J., Kyllönen, J., Huovinen, S.: Outex - new framework for empirical evaluation of texture analysis algorithms. In: *Proceedings 16th International Conference on Pattern Recognition*, pp. 701–706 (2002)
11. Paschos, G.: Fast color texture recognition using chromaticity moments. *Pattern Recognition Letters* 21(9), 837–841 (2000)
12. Paschos, G., Petrou, M.: Histogram ratio features for color texture classification. *Pattern Recognition Letters* 24(1), 309–314 (2003)
13. Randen, T., Husøy, J.H.: Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(4), 291–310 (1999)
14. Tricot, C.: *Curves and Fractal Dimension*. Springer (1995)
15. Tuceryan, M., Jain, A.K.: Texture segmentation using voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(2), 211–216 (1990)
16. Tuceryan, M., Jain, A.K.: Texture analysis. In: Chen, C.H., Pau, L.F., Wang, P.S.P. (eds.) *The Handbook of Pattern Recognition and Computer Vision*, pp. 207–248. World Scientific (1998)
17. VisTex. Vision texture database (2009), <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>
18. Weszka, J.S., Dyer, C.R., Rosenfeld, A.: A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics* 6(4), 269–285 (1976)

Multiobjective Genetic Algorithm for Image Thresholding

Layla Tahri and Mohamed Wakrim

IbnZohr University, Faculty of Sciences, EMMS, Agadir, Morocco
layla.tahri@gmail.com, m.wakrim@uiz.ac.ma

Abstract. In this paper we present a new image thresholding method based on a multiobjective Genetic Algorithm using the Pareto optimality approach. We aim to optimize multiple criteria in order to increase the segmentation quality. Thus, we've adapted the well-known Non Domination Sorting Genetic Algorithm [1] for this purpose so that it takes into consideration the contribution of the objective functions in improving the reproduction step and then improving the optimal Pareto front of solutions. Our method was tested against NSGAI algorithm and has shown effectiveness and convergence speed.

Keywords: Evolutionary approach, Genetic algorithms, Image segmentation, Image thresholding, Multiobjective optimization, Pareto optimization.

1 Introduction

Image segmentation has been approached from different ways of knowledge such as graph theory, statistics (multivariate analysis), artificial neural networks, fuzzy set theory and other areas. One of the most popular methods in this field is to perceive image segmentation as an optimization problem, where the best segmentation of a given image is achieved by optimizing one or more objective functions. In this sense, evolutionary algorithms have captivated since their apparition in 1995, the attention of the practitioners of optimization all over the world, due to their elevated degree of robustness, convergence speed and accuracy.

Moreover, they've shown effectiveness in solving complicated optimization problems and overcame the well-known problem of being trapped in local optima for the classical approaches.

To date, many attempts of applying evolutionary algorithms to find the "best" clustering of an image have been achieved. However, most of them present the disadvantage of being dependent on many parameters making their adaptation to different problems harder. This remains an important issue when it comes to apply the algorithm to different datasets of images.

To overcome this problem we propose a new algorithm based on an adapted version of the Non Sorted Genetic Algorithm II (NSGAI), using the multilevel thresholding extended from Otsu's method and the Shannon Entropy as objective functions.

This work is organized as follows: in the next section we present the extension of the classical Otsu’s thresholding method to the multilevel thresholding. In the section 3, we introduce a modified Shannon entropy measure. Then in the 4th section we present the formulation of image segmentation problem as a multiobjective optimization problem. The non-dominated sorting Genetic Algorithm is then introduced in section 5. And finally we present our new method.

2 Otsu’s Thresholding algorithm

The Otsu’s method aims to automatically segment an image into two classes based on the shape of image histogram[2]. The extension of this method to the multilevel thresholding is given by the following equations:

Let $\{t_1, \dots, t_N\}$ be the set of thresholds to evaluate. Finding the best set of thresholds is performed by maximizing the inter-classes variance:

$$\{t_1^*, \dots, t_N^*\} = \text{Maximize}\{\sigma_B^2(t_1, t_2, \dots, t_N)\}, 1 \leq t_1 < t_2 < \dots < t_N \leq \text{max}L \quad (1)$$

where $\text{max}L$ is the maximum value in the range of the thresholds.

and

$$\sigma_B^2 = \sum_{k=1}^{N+1} \omega_k (\mu_k - \mu_T)^2 \quad (2)$$

is the inter-classes variance,

with ω_k and μ_k are computed as follows:

$$\omega_k = \sum_{t_{k-1}+1}^{t_k} p_i, k \in [1, N] \quad (3)$$

and

$$\mu_k = \sum_{t_{k-1}+1}^{t_k} i * p_i, k \in [1, N] \quad (4)$$

then

$$\mu_T = \sum_{k=1}^N \omega_k \mu_k \quad (5)$$

This algorithm is wildly used, easy to implement and gives satisfactory results in terms of clustering quality. Its main disadvantage is the important computational time it needs for the execution. In fact, its exhaustive search involves $(L - N)^N$ possible combinations. This increases considerably with the number of classes in real life images.

Thus, our method has been developed to overcome this problem by combining a modified Genetic algorithm based on NSGA II and the multi-level thresholding method shown above.

3 Shannon Entropy

3.1 Bi-level Thresholding

Shannon entropy [3],[4] and [5] is a statistical criterion used to find the best threshold for segmenting an image into two classes and is defined as follows:

$$H_T(t) = -\sum_{i=1}^t \frac{p_i}{P_t} e^{1-\frac{p_i}{P_t}} - \sum_{i=t+1}^{maxL} \frac{p_i}{P_t} e^{1-\frac{p_i}{P_t}} \tag{6}$$

where t is the gray level, p_i is the frequency of t in the image, P_t is the prior probability of t and $maxL$ is the highest gray level in the image. Maximizing this entropy gives the best threshold for the image segmentation into two classes.

$$t^* = \text{ArgMax}\{H_T(t), t \in [1..maxL]\} \tag{7}$$

3.2 Multilevel Thresholding

The Shannon entropy can be extended to multilevel thresholding as follows:

$$H_T(t_1, \dots, t_N) = -\sum_{k=1}^N \sum_{i=t_{k-1}+1}^{t_k} \frac{p_i}{P_{t_k}} e^{1-\frac{p_i}{P_{t_k}}} \tag{8}$$

where $1 \leq t_1 < t_2 < \dots < t_{k-1} < t_k < \dots < t_N \leq maxL$
 and the best set of thresholds is obtained by maximizing this entropy:

$$\{t_1^*, \dots, t_N^*\} = \text{ArgMax}\{H_T(t_1, \dots, t_N)\} \tag{9}$$

4 Formulation of Image Segmentation Problem as a Multiobjective Optimization Problem:

In a problem where there are many (possibly conflicting) objectives to be optimized, there is no accepted notion as optimal solution. Hence, comparing between solutions becomes difficult. Generally, the optimizing methods produce a set of solutions of equivalent quality. The “best” solution is therefore subjective and depends on the decision maker.

The multi-objective optimization problem can be formally stated as:

Find the vector $t^* = (t_1^*, t_2^*, \dots, t_N^*)$ which will satisfy the following properties:

Optimizes the vector function $[f_1(t), f_2(t), \dots, f_{M-1}(t), f_M(t)]$

and satisfies the inequality and inequality constraints:

$$g_i(t) \leq 0, \quad i = 1, 2, \dots, k \tag{10}$$

and

$$h_i(t) = 0, \quad i = 1, 2, \dots, p \tag{11}$$

The constraints of equations (10) and (11) define the region Ω of admissible solutions and the vector t^* represents one of the optimal solutions in the set Ω .

In general, the uniqueness of the optimal solution of a multi-objective optimization problem is not satisfied, and poorly defined. The concept of Pareto optimality is as an appropriate response to this case.

For a minimization problem, the formal definition of the Pareto optimality can be given as follows:

A decision vector t^* is said to be pareto optimal if and only if there is no other vector t that dominates t^* . i.e there is no t such that :

$$\forall i \in \{1, 2, \dots, M\} : f_i(t) \leq f_i(t^*) \quad (12)$$

and

$$\exists j \in \{1, 2, \dots, M\} : f_j(t) < f_j(t^*) \quad (13)$$

this property (*i.e t* dominates t*) is written as: $t^* > t$

5 Non Domination Sorting Algorithm II (NSGAI)

There are different approaches to solve multiobjective optimization problems such as aggregating, population based non-pareto and pareto based techniques. The NSGAI belongs to the last category. Its main advantages of this algorithm are convergence speed due to its non-domination sorting applied to the generations and the elitism.

The pseudo code of the NSGAI is given as follows:

Input : the image to be clustered and the number of clusters.

Output: Pareto Front of optimal solutions

Initialization of the population

Evaluation of the objective functions

Non-dominated sort of the population

Selection of the parents

Crossover or mutation

While not (stopping criteria is reached)

 Evaluation of the objective functions

 Merge (population and offspring)

 Fronts \leftarrow Non dominated sort of the merged set

 Parents \leftarrow ensemble vide ; FrontL \leftarrow ensemble vide

 For each Front do the follow

 Compute and assign crowding distance to each individual in the ith Front

 If (size(parents)+ size(Fronti))>size(population)

 FrontL \leftarrow I; Break();

 Else

 Parents \leftarrow merge(parents, Fronti);

 End

 End

 If (size(parents)<size(population))

 FrontL \leftarrow SortByRankAndDistance(FrontL);

 For (P1 to size(population) - size(FrontL))

 Parents \leftarrow Pi

 End

 End

 Selected \leftarrow SelectParentsByRankAndDistance(Parents, size(population))

 Population \leftarrow offspring

 Offspring \leftarrow CrossoverOrMutation(Selected)

End

Return Children

6 The Proposed Multiobjective Method

In order to reach a rapid and non-parametric algorithm, we adapted the NSGAI algorithm at the following levels:

6.1 Initialization

Most of NSGA algorithms initialize the population randomly. They might have a set of individuals in the same range, which might lead to a non-well distributed population and therefore affects the next iterations. i.e a threshold variable t_i is initialized as $t_i \in [1, \dots, maxL]$, where $maxL$ is the highest level in the image.

In our method, we specify a range for each decision variable depending of the number of clusters. i.e knowing that the number of clusters is N , we divide the range $[1, \dots, maxL]$ to $N+1$ equal intervals. Then, each decision variable t_i is initialized within the appropriate interval such as: $t_i \in [t_{k-1} + 1, t_k]$, $k \in [1, N + 1]$

6.2 Selection

In most of NSGA algorithms, the selection process chooses randomly from the sorted population, the parents susceptible of performing the reproduction process. It might lead to redundancy of the parent sets in the reproduction pool.

For this reason, we've added a constraint which implies that the parents should be from different Pareto fronts, but having the best scores in terms of objective functions regarding other individuals from the same front. This can be written as follows:

Choose $Parent_i$ and $Parent_j$ such as:

$$\begin{aligned}
 &Parent_i \in Front_i \quad \text{and} \quad Parent_j \in Front_j \quad \text{for each } i \neq j \\
 &\text{with} \quad f_j(Parent_i) = ArgMax \{f_j(Parent_{i,k})\}, \\
 &\text{where} \quad Parent_{i,k} \in Front_i \quad \text{and} \quad j \in [1, \dots, M]
 \end{aligned} \tag{14}$$

6.3 Crossover

Our third intervention was at the crossover step. In fact, the NSGA algorithms use the SBX (Simulated Binary Crossover) β [6] in the process of generating offspring chromosome as described below:

$$child_1 = 1/2 \left(((1 + \beta) * Parent_1) + ((1 - \beta) * Parent_2) \right) \tag{15}$$

$$child_2 = 1/2 \left(((1 - \beta) * Parent_1) + ((1 + \beta) * Parent_2) \right) \tag{16}$$

Therefore, in order to take into consideration the weight of the parent's objective functions in improving the Pareto solutions, we use the following equations:

$$child_1 = 1/2 * \left(\left((1 + \omega_{11}) + (1 - \omega_{12}) \right) * Parent_1 \right) + \left(\left((1 + \omega_{21}) + (1 - \omega_{22}) \right) * Parent_2 \right) \quad (17)$$

$$child_2 = 1/2 * \left(\left((1 - \omega_{11}) + (1 + \omega_{12}) \right) * Parent_1 \right) + \left(\left((1 - \omega_{21}) + (1 + \omega_{22}) \right) * Parent_2 \right) \quad (18)$$

where

$$\omega_{11} = \frac{f_{11}(t)}{f_{11}(t)+f_{12}(t)} \text{ and } \omega_{12} = \frac{f_{12}(t)}{f_{11}(t)+f_{12}(t)} \quad (19)$$

represents the contribution weights of objective functions for the first parent $Parent_1$ and

$$\omega_{21} = \frac{f_{21}(t)}{f_{21}(t)+f_{22}(t)} \text{ and } \omega_{22} = \frac{f_{22}(t)}{f_{21}(t)+f_{22}(t)} \quad (20)$$

are the contribution weights of objective functions for the second parent $Parent_2$

Noting also that f_{ij} is the j^{th} objective function for the i^{th} individual.

6.4 Mutation

The NSGA uses a constant η_m called the polynomial mutation spread factor that needs to be tuned. Usually it takes 20 as value. Thus, in order to overcome this parameter-dependence, we mutate the parent as described below:

For each component p_j of the parent do

Convert p_j to the binary format.

Randomly select the i^{th} bit to change.

Mutate if as follows

If $p_j(i) = 0$ then

$p_j(i) = 1$

Else

$p_j(i) = 0$

Re-convert the component p_j to the original number format.

7 Results and Discussion

We've chosen a set of test images from the Berkeley Database [7] in order to evaluate and compare the performance of both NSGAI and the Adapted NSGAI algorithms. This choice is based on the nature of the images: the goat and elephant images present a weak contrast between the foreground objects and the background. In the opposite, the peppers and the water skier have an important contrast regarding the background image.

Table 1. Original images from Berkeley database



Fig. 1. Goat image



Fig. 2. Elephant image



Fig. 3. Water skier image



Fig. 4. Peppers image

Table 2. Comparison of the number of input and output parameters in NSGAII and our proposed method

Algorithm	Input parameters	Output
NSGA II	Image - Number of clusters Crossover spread factor Mutation spread factor Pool size - Tournament size	Pareto front
The adapted NSGA II	Image Number of clusters Pool size Tournament size	Pareto front

Table 3. Comparison between the Otsu’s variance, the modified Shannon entropy and the CPU time in NSGAI and our proposed method

Algorithm	Image	Number of clusters	Otsu’s Variance	Modified Shannon Entropy	CPU time (in seconds)
NSGAI	Goat	3	2445.8986	-10.1730	6.4176
	Elephant	3	1337.7215	-8.9751	6.2460
	Peppers	6	3243.9823	-22.6925	8.8013
	Water skier	4	1224.2536	-16.4316	7.8684
The adapted NSGAI	Goat	3	2338.6183	-9.7920	1.9344
	Elephant	3	1424.8782	-10.3751	1.3884
	Peppers	6	3535.4301	-18.3782	2.9812
	Water skier	4	1312.6678	-13.0296	2.0280

In table 2, we can see that the NSGA II requires more input parameters than the adapted version we proposed. Moreover, those parameters need tuning in order to be adjusted for different sets of images.

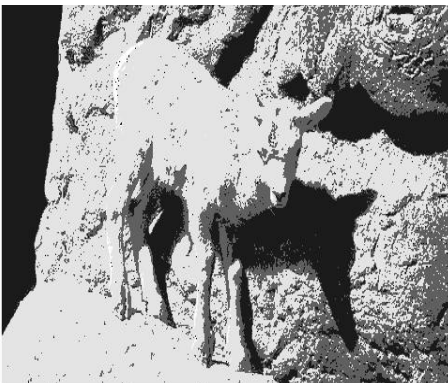
In table 3, we present the mean values of the Otsu’s Variance, the modified Shannon entropy and the CPU execution time for each algorithm over 50 runs and 100 generation.

The overall performance is close with better results for our algorithm for the peppers and Water skier images which can be explained by the introduction of the objective function’s weights in the crossover process.

In terms of execution time, our algorithms performs much better than the NSGAI, this might be owed to the elimination of the spread crossover and mutation factors which has led to less calculations and therefore required less computational time.

Table 4. Results of image segmentation performed by NSGAI and Adapted NSGAI

NSGAI results



The Adapted NSGAI results

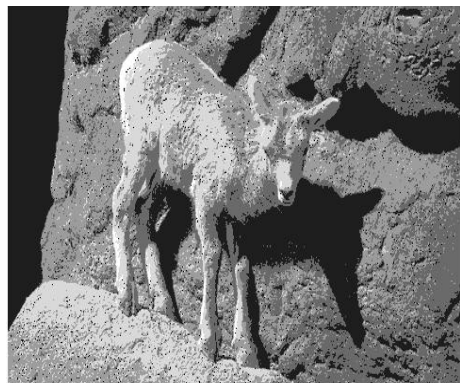


Fig. 5. Segmented goat image using the NSGAI

Fig. 6. Segmented goat image using the Adapted NSGAI

Table 4. (Continued)

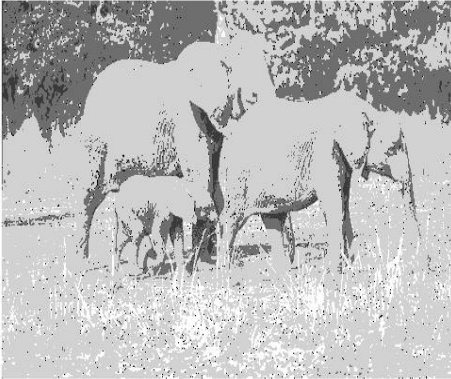


Fig. 7. Segmented elephant image using the NSGAI

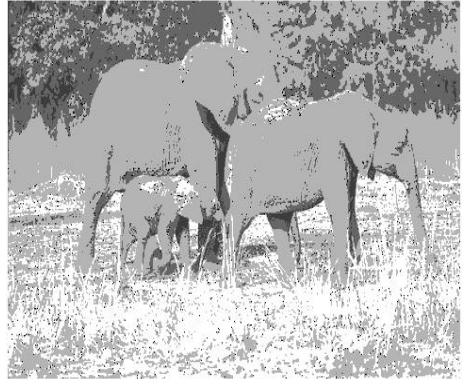


Fig. 8. Segmented elephant image using the Adapted NSGAI



Fig. 9. Segmented peppers image using the NSGAI

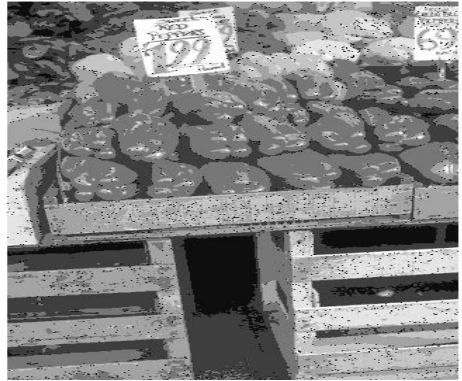


Fig. 10. Segmented peppers image using the Adapted NSGAI



Fig. 11. Segmented water skier image using the NSGAI



Fig. 12. Segmented water skier image using the Adapted NSGAI

Although the goat and elephant image present less contrast, the Adapted NSGAI performed well and gave a good segmentation result.

Furthermore, the proposed adapted algorithm shows better performance for the images with important contrast compared to the segmentation results of the NSGAI. This is consistent in general with the Otsu's variance and the modified Shannon entropy results in table 2.

8 Conclusion

In this article, a new image segmentation based on multiobjective genetic algorithm has been proposed. The question has been formulated as a multiobjective optimization problem.

The objective functions used were the Otsu's variance and a modified Shannon entropy measure, both extended to multilevel thresholding.

The algorithm has been applied to several images presenting different characteristics; the results were presented and compared to the NSGAI ones.

The good performance of our method for different types of images at the level of objective functions, convergence speed and independence of parameter tuning; shows that it might be motivating to use this algorithm in the field of image segmentation.

References

1. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6(2), 182–197 (2002)
2. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. on Syst., Man and Cyb.* 9(1), 62–66 (1979)
3. Kapur, J.N., Sahoo, P.K., Wong, A.C.K.: A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics and Image Processing* 29, 273–285 (1985)
4. Pal, N.K., Pal, S.K.: Entropy: A new definition and its applications. *IEEE Trans. Syst. Man. Cybern.* 21, 1260–1270 (1991)
5. Nakib, A.: Conception de métaheuristiques d'optimisation pour la segmentation d'images. Application à des images biomédicales, pp. 9–10 (2008)
6. Deb, K., Kumar, A.: Real-coded Genetic Algorithms with Simulated Binary Crossover: Studies on Multimodal and Multiobjective Problems. *Complex Systems*, 431–454 (1995)
7. The Berkeley database site,
<http://www.oracle.com/technetwork/database/berkeleydb/overview/index.html>

Dual-Resolution Active Contours Segmentation of Vickers Indentation Images with Shape Prior Initialization

Michael Gadermayr and Andreas Uhl*

Department of Computer Sciences, Salzburg University, Austria
uhl@cosy.sbg.ac.at

Abstract. Vickers microindentation imagery is segmented using the Chan-Vese level-set approach. In order to find a suitable initialization, we propose to apply a Shape-Prior gradient descent approach to a significantly resolution-reduced image. Subsequent local Hough transform leads to a very high accuracy of the overall approach.

1 Introduction

The Vickers hardness test uses a pyramidal diamond as indenter which is applied to a flat surface using a known force. The resulting indentation is captured using a microscope (see Fig. 1 for example images) and the diagonals are measured to determine the Vickers hardness of the material.

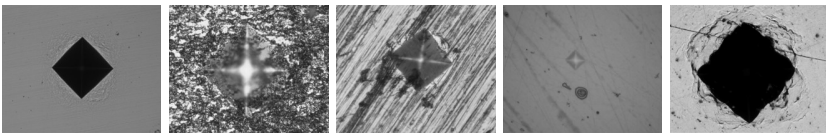


Fig. 1. Example images

There are several proposals for automated image segmentation of Vickers indentations, among them techniques based on template matching [1,2] and applying a local Hough transform to predefined vertex candidate regions [3,4]. The method introduced in [5] applies thresholding followed by a Hough transform. Other suggested methods also binarize the image using thresholding [6,7].

As can be seen in the figure, the leftmost example exhibits perfect properties for segmentation, while other images suffer from noise and/or low contrast. Therefore, Vickers indentation segmentation remains challenging. Active contours have not been investigated with respect to Vickers images so far, although the method is known to be an accurate state of the art segmentation tool. The contribution of this work is the proposal of a dual-resolution active contours algorithm (i.e. level-set approach) where the initial level set is found by a Shape-Prior gradient descent algorithm applied to a resolution-reduced image.

* Corresponding author.

This paper is structured as follows. In Sect. 2 we review existing level-set algorithms and discuss their limitations when being applied to indentation images. In Sect. 3 we introduce a Shape-Prior gradient descent method, that produces robust approximative segmentation results which serve as initialization for subsequent level-set techniques. Unlike existing Shape-Prior approaches, we restrict the evolution to the exact prior shape. In Sect. 4 we fuse the components into the final algorithm: computation of an initial level set by applying the Shape-Prior gradient descent to resolution-reduced images, indentation segmentation with the highly accurate Chan-Vese region based level-set technique, and application of a local Hough transform to vertex candidate regions to optimize the accuracy of the corner detection. Section 5 compares the results to a template matching based state of the art indentation segmentation [1] with respect to accuracy and computational effort. Section 6 concludes this paper.

2 The Level-Set Approach

Active contours [8] are closed curves, which iteratively converge at object boundaries. The curve is forced by an energy criterion, which is based on the homogeneity of the contour on the one hand and on the image information on the other hand.

In the level-set formulation [9] an explicit parametrization by frontier points is circumvented by using an intrinsic formulation. The evolving contour is given by its level set Γ :

$$\Gamma = \{(x, y) | \phi(x, y) = 0\} . \quad (1)$$

$\phi(x, y)$ is a function which is 1 inside, -1 outside of the region and exactly 0 at the frontier of the evolved shape. The evolution of the frontier happens by moving the initial level set in normal direction to the contour with a specified speed v . There exist various different ways of calculating the speed function v , which influences the behavior of the evolving level set.

One quite common approach is based on the gradient information of the image [10]. The speed v is adjusted in order to reduce occurrences of the contour in image regions without image gradients. The edge based level-set algorithm requires the propagation of edges to increase the capture range of single edge pixels, otherwise, an exact initialization is required. Another problem is that image gradients might be weak or blurred.

To bypass these issues, a region based approach has been introduced [11]. This method is based on the assumption that the object's surface and the surface outside of the object are homogeneous as far as its gray value is concerned. The following region based energy criterion has been introduced:

$$E_{CV} = \int_{\Gamma_{in}} (I(v) - \bar{I}_{in})^2 dv + \int_{\Gamma_{out}} (I(v) - \bar{I}_{out})^2 dv + \lambda \int_{\Gamma} \|\nabla\phi(v)\| dv . \quad (2)$$

I is the image gray value, \bar{I}_{in} and \bar{I}_{out} are the average values inside and outside of the contour, ∇ is the gradient operator and λ is the curvature weighting term.

Intuitively, energy is low if the gray values inside the contour are equal, the gray values outside the contour are equal and the contour is smooth.

Although the region based Chan-Vese approach is known to be less vulnerable to a poor initialization, we have observed that final segmentation results are accurate only in case of close initialization. Fig. 2 illustrates the dependence of the level-set approach on the initialization. An inappropriate initialization (left rectangle) causes a wrong segmentation (crosses).

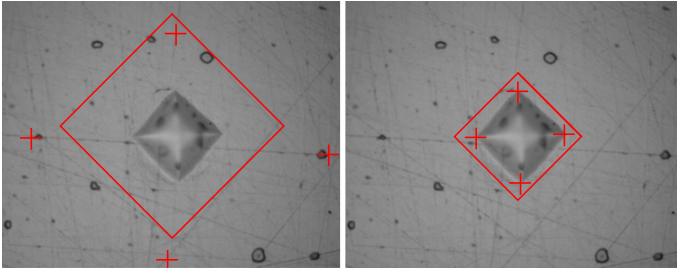


Fig. 2. The influence of the initialization (right: appropriate initialization, left: inappropriate initialization)

An additional issue is computational cost, which is tremendous in case the initialization is far away from the final contour. Having observed the importance of an appropriate initialization, we introduce a corresponding strategy in the following.

3 The Shape-Prior Gradient Descent Method

The approach is based on the fact, that the indentations approximately have a square shape which serves as the prior shape. Whereas traditional active contours and level-set methods evolve arbitrary curves, we consequently only evolve the parameters of a strict square template, which is represented by the following four parameters: x_0 (horizontal translation), y_0 (vertical translation), r_0 (scaling) and α (rotation).

The contour of the square is given by the points (x, y) with the distance $d = r_0$ to a center (x_0, y_0) . d is calculated in the following way, to ensure that the evolved contour has a square shape:

$$d = \begin{aligned} & |(x - x_0) \cdot \cos(\alpha) + (y - y_0) \cdot \sin(\alpha)| + \\ & |(x - x_0) \cdot \sin(\alpha) - (y - y_0) \cdot \cos(\alpha)| . \end{aligned} \quad (3)$$

Of course, this algorithm will not be able to segment Vickers images perfectly, as Vickers' shape often cannot be described by a perfect square. However, this is not our objective, but the found results can serve as good initialization for a subsequent more accurate strategy.

The regions in- and outside of the square are given by Γ_{in} and Γ_{out} :

$$\Gamma_{in} = \{(x, y) : d < r_0\} . \quad (4)$$

$$\Gamma_{out} = \{(x, y) : d > r_0\} . \quad (5)$$

As with the level-set approach, we define an energy criterion which is minimized by gradient descent. Different energy functions (edge based, region based) have been investigated. Tests showed that the following statistical criterion, which is derived from the approach proposed in [12], is the best choice:

$$E = - \int_{\Gamma_{in}} \log(p_{\Gamma_{in}}(f(v)))dv - \int_{\Gamma_{out}} \log(p_{\Gamma_{out}}(f(v)))dv . \quad (6)$$

$f(v)$ is an arbitrary feature of the point v . The higher the dimensionality of the feature vector f , the higher are the computational costs, as for each step of the iterative gradient descent, the n dimensional probability densities $p_{\Gamma_{in}}$ and $p_{\Gamma_{out}}$ have to be calculated. Moreover, a higher dimensionality causes the empirical probability function (which is a matrix of n dimensions) to become a sparse matrix, as the number of matrix elements is exponentially increasing whereas the number of features stays the same. When the elements of the matrix are sparse, the empirical distribution (gathered from the pixels inside or outside the contour) cannot be utilized straightforward. Consequently, it is necessary to estimate the real probability density function which is done by applying a Gaussian Parcen window in different sizes.

Empirical tests showed, that the following feature vector produces the best results:

$$f(v) = (I(v), \|\nabla I(v)\|) . \quad (7)$$

We compute $p_{\Gamma_{in}}$ and $p_{\Gamma_{out}}$ by applying a Gaussian Parcen window to the second dimension (edge information) with variance $\sigma = 2$. For the first dimension (gray value) the empirical density function is being used.

The evolved parameters are collected in the vectors $s_i = (x_0, y_0, r_0, \alpha)$. The vector s_0 is the initialization. s_{n+1} is defined recursively:

$$s_{n+1} = s_n + \lambda(\nabla E) . \quad (8)$$

λ which usually is a multiplicative component, is called step size. To allow lambda to act as a signum function (one pixel left, stay, one pixel right), which can deal with numerical issues, it is more generally defined as function. We use the following definition:

$$\lambda((x_1, \dots, x_n)^T) = (\text{sign}(x_1), \dots, \text{sign}(x_n))^T . \quad (9)$$

$$\nabla E = \left(\frac{dE}{dx}, \frac{dE}{dy}, \frac{dE}{dr}, \frac{dE}{d\alpha} \right)^T . \quad (10)$$

e.g. the partial derivative of the x dimension is calculated as:

$$\frac{dE}{dx}((x, y, r, \alpha)^T) = \frac{1}{2} \cdot [E((x + 1, y, r, \alpha)^T) - E((x - 1, y, r, \alpha)^T)] . \quad (11)$$

Although the introduced approach is already able to deal with local minima caused by noise, we still have not achieved a total invariance to the initialization s_0 . Local minima still prevent from a proper localization of the indentation in several cases. The balloon approach [13] introduced for active contours, deals with this problem by adding an energy term, forcing the contour to become smaller or larger. Our approach allows to apply a kind of balloon force in an easy but effective way: Instead of calculating the radius r_0 by gradient descent, r_0 is simply decreased by one in each iteration of the gradient descent.

If the contour starts outside the image boundaries, it necessarily has to cross the object's boundaries, when getting smaller and smaller.

Unlike unforced gradient descent, the proposed balloon-method does not stop before r_0 becomes zero (or a defined minimum). In a second step the history of the gradient descent has to be analyzed, to get the best fitting vector s_{res} from a set of several local minima. In our case the best results are achieved when using the vector s_i with the highest response (achieved by convolution) of the image information to the template (parametrized by s_i) shown in Fig. 3 with a thickness of 3 pixels.

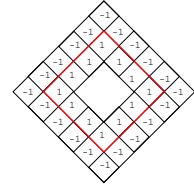


Fig. 3. Directed edge template (thickness 1)

4 Dual-Resolution Level-Set with Hough Postprocessing

Since the main aim of the Shape-Prior gradient descent is to provide a robust initialization for subsequent more accurate techniques allowing for more adapted contours, we apply the approach to downscaled versions of the original images. This limits the computational demand of the initialization. The calculated results are upscaled and form an initial level set in the full resolution image. Subsequently, the Chan-Vese level set approach with the given initialization is conducted.

After applying the level set algorithm a decision must be made in order to identify the corners of the indentation. Unfortunately, depending on the configuration, indentations are not always segmented perfectly, but often incomplete or ragged. Even small artefacts or cut corners considerably affect the accuracy of the segmentation process if simply the topmost, the rightmost, the bottommost and the leftmost pixels of the evolved level set are identified as the corners of the indentations. Consequently, we apply the local Hough transform [34] as a post-processing strategy which approximates lines in a defined distance from the precalculated corner points of the level set technique. In the first step, the Hough transform is computed separately for each corner in a surrounding region. In the second step, the correct two lines of the Hough transform have to be selected. The first chosen line is the line with the best Hough rating and the second one is the line with the best Hough rating which is almost orthogonal to the first one. The actual corner is assumed to be the intersection of the selected two Hough lines.

5 Experiments

For experimentation, 150 test images (1280x1024 pixels) acquired with EM-COTest Durascan hardware were used. In order to compare the calculated results with the ground truth, these images were manually evaluated with respect to the correct indentation vertex position by four experts independently. The ground truth was determined by taking the mean of all four measurements. For the first step of the dual-resolution approach, these images are downscaled by factor 10 (averaged with a Lanczos filter). The local Hough transform is applied in a circular area around the calculated vertex position with a radius of 60 pixels.

The results of the different strategies of the proposed dual-resolution algorithm are shown in Fig. 4. For each deviation of the calculated vertex to the ground truth in pixels as shown on the x-axis, the number of vertices with the respective distance to the real points are shown on the y-axis (bars). The right-most bar collects the outliers (vertices with distances ≥ 20 or ≥ 50 pixels) while the line graph represents the relative cumulative distribution.

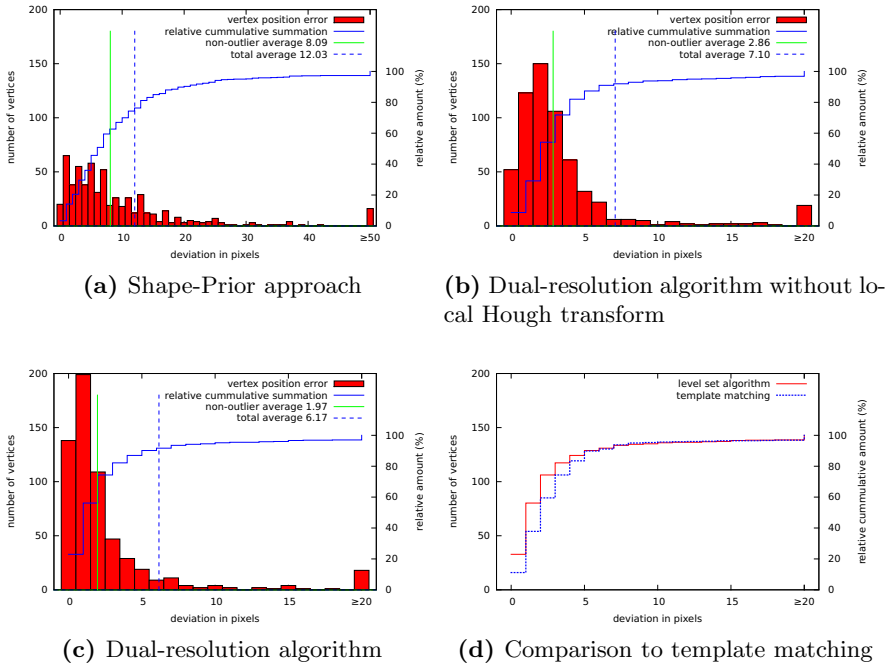


Fig. 4. Accuracy of indentation segmentation

In Fig. 4a, the provisional results (i.e. initializations for the level-set method) of the approximative Shape-Prior gradient descent approach are shown. Only vertices with a deviation of at least 50 pixels in the original image (i.e. 5 pixels in the downscaled image) are defined to be outliers. The number of outliers is already quite low, but the accuracy can be considerably increased by applying

the level-set segmentation method as shown in Fig. 4b. For example, the ratio of vertices with a deviation of maximal 2 pixels can be increased from 21% to 53% (note that for these results, the extremal points of the segmented areas are assumed to be the indentations' vertices).

Figure 4c shows the final results achieved with the proposed dual-resolution approach plus additional local Hough transform post-processing. The accuracy is again considerably increased (e.g., 74% of the vertices have a deviation of maximal 2 pixels as compared to 53% without the local Hough transform).

We also compare the performance of the proposed dual-resolution algorithm with an alternative robust and accurate template matching approach as introduced earlier [1]. Fig. 4d compares the cumulative distribution of the proposed technique and the referenced template matching method. Especially the probability of a very exact segmentation of the indentation corner points (Euclidean distances of 0 to 5 pixels) is considerably higher with the proposed method. The number of outliers is the same.

In Fig. 5, example segmentation results of our proposed approach are shown.

Finally, we evaluate runtime performance. Shape-Prior gradient descent and the Hough transform are implemented in Java and are not optimized for execution speed. For the exact Chan-Vese segmentation stage, the high performance Ofeli C++ level-set library is used. In Table 1, average runtimes per image of the proposed method and the template matching approach [1] (implemented in Java) are shown. The tests were executed on a notebook with an Intel Core 2 Duo T5500 1.66 GHz processor.

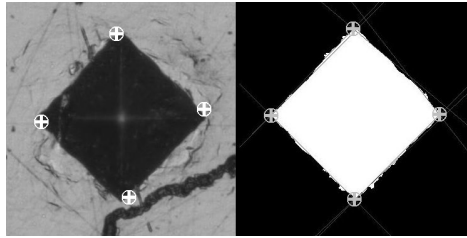


Fig. 5. Approximative results after stage 1 (left) and final results (right)

Table 1. Runtime comparison

Method	avg. runtime per image
Total costs dual-resolution method	4.3 s
Proposed statistical Shape-Prior algorithm	2.2 s
Proposed region based level-set method	1.3 s
Local Hough transform	0.7 s
Referenced template matching [1]	3.1 s

We see that although composed of three distinct procedures, the proposed technique is competitive to template matching in terms of runtime. Especially in the first stage, we still see room for runtime optimizations.

6 Conclusion

The introduced dual-resolution approach consisting of a robust gradient descent localization method and an exact level-set segmentation method is highly competitive. Especially, a very exact segmentation of high quality images can be achieved.

Acknowledgment. This work has been partially supported by the Austrian Federal Ministry for Transport, Innovation and Technology (FFG Bridge 2 project no. 822682).

References

1. Gadermayr, M., Maier, A., Uhl, A.: Algorithms for microindentation measurement in automated Vickers hardness testing. In: Pinoli, J.C., Debayle, J., Gavet, Y., Cruy, F., Lambert, C. (eds.) Tenth International Conference on Quality Control for Artificial Vision (QCAV 2011). Proceedings of SPIE, vol. 8000, pp. 8000M-1-8000M-10. SPIE, St. Etienne (2011)
2. Maier, A., Uhl, A.: Robust automatic indentation localisation and size approximation for vickers microindentation hardness indentations. In: Proceedings of the 7th International Symposium on Image and Signal Processing (ISPA 2011), Dubrovnik, Croatia, pp. 295-300 (September 2011)
3. Ji, Y., Xu, A.: A new method for automatically measurement of vickers hardness using thick line hough transform and least square method. In: Proceedings of the 2nd International Congress on Image and Signal Processing (CISP 2009), pp. 1-4 (2009)
4. Yao, L., Fang, C.-H.: A hardness measuring method based on hough fuzzy vertex detection algorithm. *IEEE Trans. on Industrial Electronics* 53(3), 963-973 (2006)
5. Macedo, M., Mendes, V.B., Conci, A., Leta, F.R.: Using hough transform as an auxiliary technique for vickers hardness measurement. In: Proceedings of the 13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006), pp. 287-290 (2006)
6. Mendes, V., Leta, F.: Automatic measurement of Brinell and Vickers hardness using computer vision techniques. In: Proceedings of the XVII IMEKO World Congress, Dubrovnik, Croatia, pp. 992-995 (June 2003)
7. Sugimoto, T., Kawaguchi, T.: Development of an automatic Vickers hardness testing system using image processing technology. *IEEE Transactions on Industrial Electronics* 44(5), 696-702 (1997)
8. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321-331 (1988)
9. Osher, S., Sethian, J.A.: Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics* 79(1), 12-49 (1988)
10. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *International Journal of Computer Vision* 22(1), 61-79 (1997)
11. Chan, T., Vese, A.: Active contours without edges. *IEEE Transactions on Image Processing* 10(2), 266-277 (2001)
12. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International Journal of Computer Vision* 72(2), 195-215 (2006)
13. Cohen, L.: On active contour models and balloons. *CVGIP: Graphical Models and Image Processing* 53(2), 211-218 (1991)

Matching Noisy Outline Contours Using a Descriptor Reduction Approach

Saliha Aouat and Slimane Larabi

¹ LRIA Laboratory, Computer Science Department, USTHB University, Algiers, Algeria
{saouat,slarabi}@usthb.dz

Abstract. Shape Matching is an important area in computer vision researches. We propose in this paper a method to match two outline shapes. Assuming that shapes are stored in the database using their textual descriptors, an iterative process is used to reduce descriptors. After the reduction process, the textual descriptors can be compared in order to perform the matching process. The Textual smoothing is done by applying transformations and reductions of the textual descriptors of shapes to be matched.

Keywords: Descriptors, matching, smoothing, reduction, Textual descriptors.

1 Introduction

Different representations of shapes have been proposed these last years and used in the recognition process. The most known representations are based onto appearance[1], outline contour [2, 3, 4, 5, 6, 7, 8], aspect-graph[9, 10], set of characteristic outline points [11], medial axis of silhouettes [12, 13], shock graph[14, 15], and shape axis trees (S-A-trees)[16]. A review of shape representation methods may be found in [17, 18]. In [19], authors propose a part-based method for silhouettes representation. Silhouettes are partitioned into parts, junction and disjunction lines. Each element is then geometrically described. The obtained description is written following an XML language noted **XLWDOS** (XML Language for Writing Descriptors of Outline Shapes). Since real images are noisy, there XLWDOS descriptors may be very different. This sensitiveness to noise does not facilitate their use in the matching and recognition processes. A notion of multi-scale descriptors of silhouettes is introduced and applied to match silhouettes. A Gaussian convolution of silhouettes is done in order to smooth outline shapes and eliminate noise depending on the value of the Gaussian scale. Also, noisy XLWDOS descriptors of silhouettes may be matched using a reduction technique that eliminates noisy elements from the descriptors. This paper is structured as follows:

We present in the second section an overview of the part-based method for describing outline shapes. In the third section we show the sensitiveness to noise of XLWDOS descriptors. In the fourth section, we explain our strategy based on the matching of XLWDOS descriptors using the reduction technique. The proposed method is validated using real images and the obtained results are presented and discussed in the fifth section.

2 Silhouettes Description

Concavity points for which direction of outer contour changes following top-bottom-top or bottom-top-bottom are considered as partition points (see figure 1.a). A silhouette is partitioned at these points onto parts, junction and disjunction lines: either, two parts or more are joined with a third part through a junction line, or a part is joined with two parts or more through a disjunction line. This process applied to the left silhouette in figure 1. produces seven parts, two junction lines and one disjunction line. The silhouette descriptor is the grouping of descriptors of its elements. A part is defined by its two boundaries (left and right) which begin at the highest left point and terminate at their lowest points (see figure 1.b). Using the inflection points, these boundaries are segmented into a set of primitives (line, convex and concave contours) and described by the parameters: type (line, convex or concave curve), degree of concavity or convexity, angle of inclination and length. Junction and disjunction lines are decomposed onto segments. Each segment is described with three parameters: type, the reference numbers of parts where it appertains and its length. Types of segment are: **Junction** if it is common for two parts, **Free-High** if it belongs only to the high part or **Free-Low** if it belongs to the low part. Applying this description to write the descriptor of part P₂ of the shape in the left of figure 1, we obtain:

$$P_2 \rightarrow \langle P_2 \rangle \langle L \rangle cv 6\% 56 76 \langle /L \rangle \langle R \rangle cv 16\% 107 77 \langle /R \rangle \langle /P_2 \rangle$$

This descriptor is read as follow: The left boundary of part 2 is composed by a convex contour with 0.06 as degree of convexity, 56° of inclination and 76 pixels as length. The right boundary is composed by a convex contour with 0.16 as degree of convexity, 107° of inclination and 77 pixels as length.

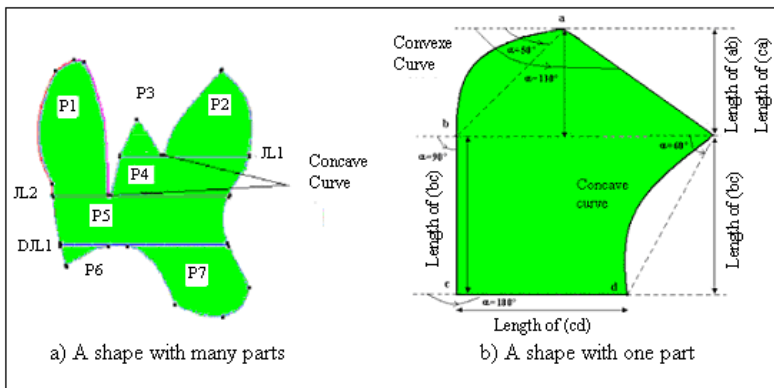


Fig. 1. Example of a silhouette, concave points, and parts

The notion of **composed part** is defined as a set of two (or more) parts joined to another part using a junction (or a disjunction) line and written as follows:

Composed Part $\rightarrow \langle CP \rangle P_1 P_2 \dots P_{n-1} \langle J \rangle$ Junction line $\langle J \rangle P_n \langle CP \rangle / \langle CP \rangle P_1 \langle D \rangle$
 Disjunction line $\langle D \rangle P_2 P_3 \dots P_n \langle CP \rangle$

Recursively, a composed part is considered as a part and may constitute (with other elements) other composed part. There are three composed parts in the XLWDOS descriptor of the left silhouette in figure 1:

- $\langle CP \rangle P_2 P_3 \langle J \rangle JL_1 \langle J \rangle P_4 \langle CP \rangle$
- $\langle CP \rangle P_1 \langle CP \rangle P_2 P_3 \langle J \rangle JL_1 \langle J \rangle P_4 \langle CP \rangle \langle J \rangle JL_2 \langle J \rangle P_5 \langle CP \rangle$
- $\langle CP \rangle \langle CP \rangle P_1 \langle CP \rangle P_2 P_3 \langle J \rangle JL_1 \langle J \rangle P_4 \langle CP \rangle \langle J \rangle JL_2 \langle J \rangle P_5 \langle CP \rangle \langle D \rangle DJL1 \langle D \rangle P_5 P_6 \langle CP \rangle$.

To write descriptors of silhouettes we use the following syntax:

Silhouette $\rightarrow \langle DXLWDOS \rangle \langle Name \rangle Objectname \langle /Name \rangle$ Composed Part $\langle /DXLWDOS \rangle$
 (DXLWDOS means description according to XLWDOS description).

Finally, the descriptor of the silhouette in figure 1 is:

```

<DXLWDOS><Name>Silhouette 1</Name>
<CP><CP><P1><L>r 32 10 cv 14% 88 102 r 90 10 </L><R> r 165 8 cv 6% 100
120</R></P1>
<CP><P2><L>cv 6% 56 76</L><R>cv 16% 107 77</R></P2>
<P3><L> r 64 32</L><R>r 123 32</R></P3> <J>j P3 P4 36 w P4 2 j P2 P4 75</J>
<P4><L> r 76 34 </L><R> r 63 34</R></P4><CP> <J>h P1 1 j P1 P5 48 w P5 2 j P4 P5
103 h P4 1</J>
<P5><L> r 98 41</L><R>cc 11% 88 41</R></P5><CP> <D>j P5 P6 42 h P5 17 j P5 P7
87 w P7 1</D>
<P6><L> r 105 19 </L><R>cc 6% 27 19</R></P6>
<P7><L>cc 10% 127 53 cv 7% 165 11</L><R> r 115 38 cv 17% 48 26</R></P7></CP>
</DXLWDOS>
    
```

3 Sensitiveness to Noise of XLWDOS Descriptors

XLWDOS descriptors are sensitive to noise which may produce additional parts, junction and disjunction lines. Noise may then change the global structure of XLWDOS descriptor, but in other cases it changes only the geometric description of its elements. Figure 2.b illustrates an example where noise produces in addition, two parts, one junction line, two other parts, and one disjunction line relatively to the shape 2.a. Also, noise may correspond to concave points that the change of position creates new parts and lines. The silhouette of figure 2.d illustrates an example where a concavity point change of position produces additional parts and lines relatively to the shape 2.c. Finally, the XLWDOS descriptors are robust to noise when it changes only the geometric description of contours; it's the case of the silhouette 2.e. Depending on the direction of computation of the XLWDOS descriptor; noise may change the global structure of XLWDOS descriptors.

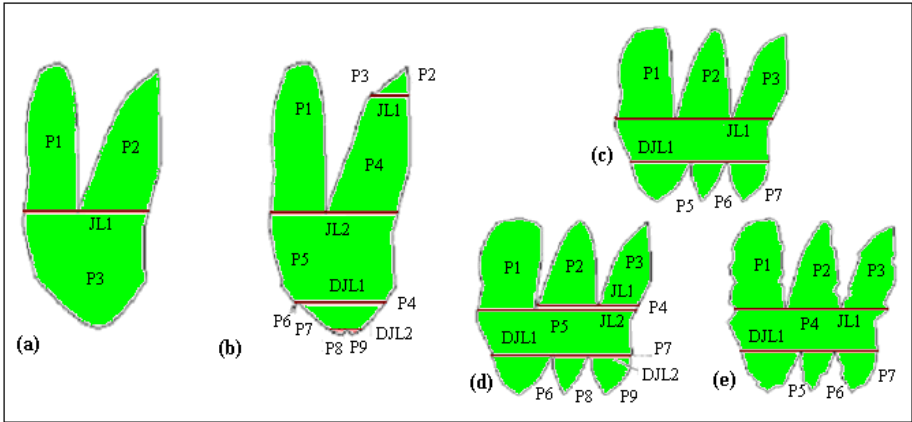


Fig. 2. Noisy outline shapes

4 Matching Descriptors of Silhouettes

A. Matching descriptors in multi scale space

As it is described above, XLWDOS descriptors are sensitive to noise. Therefore, it is necessary to take into account this noise in the matching process. In a previous work we have developed a method comparing silhouettes at different scales [20] to eliminate noise from outline shape applying a convolution with a Gaussian filter. We obtain for each value of σ a smoothed outline shape. We define the notion of multi-scale XLWDOS descriptors as the set of descriptors computed using these smoothed outline shapes obtained at different scales (values of σ) from the initial outline shape. More the value of σ increases, more the XLWDOS descriptor contains less elements whose number becomes steady from a certain value of σ . [20]. In this present work, we will see that it is possible to found same results with textual transformation of descriptors without Gaussian Smoothing.

B. Reduction method

We define a noisy part, as a part whose length is 1, 2, 3 or n pixels. Each noisy part will be marked in the written descriptor using the two tags $\langle N \rangle$ $\langle /N \rangle$. For example, the XML structure of XLWDOS descriptors of silhouette 2.a and 2.b are:

```

<DXLWDOS><Name>Silhouette2a</Name>
  <CP><P1></P1><P2></P2><J></J><P3></P3></CP></DXLWDOS>
<DXLWDOS><Name>Silhouette2b</Name>
  <CP><CP><P1></P1><CP><P2></P2><N><P3></P3></N><J></J><P4></
P4></CP><J></J><P5></P5></CP>
  <D></D><N><P6></P6></N><CP><P7></P7><D></D><N><P8></P8></N
><N><P9></P9></N></CP></CP>
</DXLWDOS>
    
```

These two descriptors cannot be matched because they have different structures. Therefore, the problem is to verify if we can reduce the two descriptors in order to

maintain in their XML structures only their main parts. For this, we propose a reduction method that takes into account all possible positions of noisy parts in the XML structure.

Also, the XML structure of XLWDOS descriptors of silhouettes 2.c and 2.d are:

```
<DXLWDOS><Name>Silhouette2c</Name>
  <CP><CP><P1></P1><P2></P2><P3></P3><J></J><P4></P4></C
  P>
  <D></D><P5></P5><P6></P6><P7></P7></C>
</DXLWDOS>
<DXLWDOS><Name>Silhouette2d</Name>
<CP><CP><P1></P1><CP><P2></P2><P3></>
<J></J><N><P4></P4></N></CP>
  <J></J><P5></P5></CP>
  <D></D><P6></P6><CP><N><P7></P7></N>
  <D></D><P8></P8><P9></P9></CP></CP>
</DXLWDOS>
```

The principle of our reduction method is as follows:

When the size of a part (Psi) is negligible in relation with sizes of main parts, the (Psi) will be considered as noisy and will be suppressed from the initial descriptor. The noisy part size is, therefore, less than a fixed threshold.

Let be:

$\langle CP \rangle Ps_1 Ps_2 \dots Ps(i-1) Psi Ps(i+1) \dots Psn Ji Pm \langle /CP \rangle$, a composed part according to XLWDOS Descriptor.

The same descriptor, after removing the noisy part (Psi), becomes:

```
<CP>Ps1 Ps2 ... P's(i-1) Ps(i+1)... Psn J'i Pm</CP>.
```

P's(i-1) is generated from Ps(i-1) where the description of left boundary of P's(i-1) is the same than the left boundary of Ps(i-1). The right boundary of P's(i-1), is the right boundary of Ps(i-1) for which we add the noisy part (Psi) and the segments of separating lines adjacent to (Psi).

The separating line descriptor will also be modified according to the new part obtained after the reduction process. Indeed all segments that were adjacent to (Psi), will be assigned to the new part P's(i-1).

The graphic illustration of this technique is given in Figure 3.

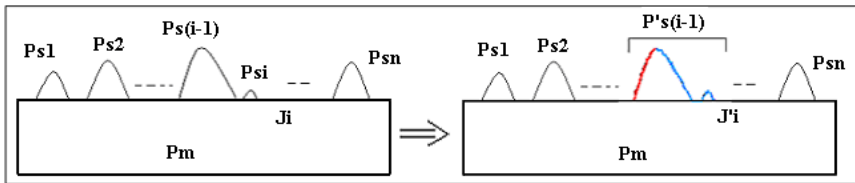


Fig. 3. Elimination of a noisy part in the textual description

This is a recursive process; it is applied to remove all noisy parts from the descriptor while parts sizes are less than the fixed threshold and while the two descriptors to be matched remain different during the reduction process.

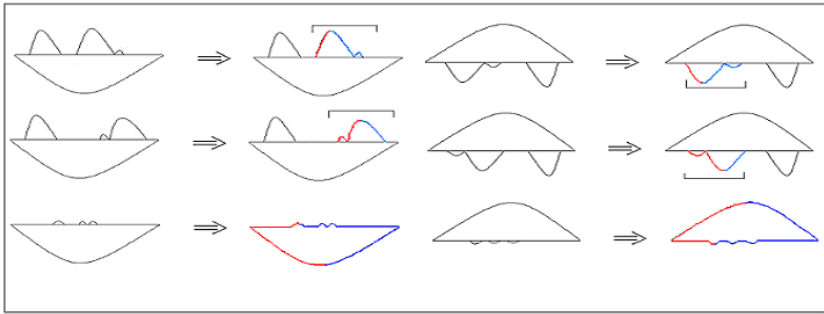


Fig. 4. Examples of descriptors reduction

For example, using the reduction method, the XML structure of XLWDOS descriptor of silhouette 2.d becomes:

```
<DXLWDOS><Name>Silhouette2d</Name>
<CP><CP><P1></P1> <P'2></P'2><P'3></P'3><J></J><P5></P5></CP>
<D></D><P6></P6> <P'8></P'8><P'9></P'9></CP>
</DXLWDOS>
```

Now, the XML structure of both descriptors (Silhouette2c and Silhouette2d) are similar, therefore the matching of their elements may be done.

Other examples of transformations are given by graphic illustrations in Figure 4.

5 Experimentation

In figure 5, we show two images of the same object (cup). There are parts and junction lines in the descriptor of the first image which could not be matched with other parts and lines in the descriptor of the second image.

After applying a convolution with a Gaussian filter for the two outline shapes, smoothed outlines shapes are obtained and the obtained XLWDOS descriptors of both two images become similar.

The matching problem has also been solved using our reduction method. The approach is been applied for the two noisy XLWDOS descriptors and gave same results. Indeed, the first noisy descriptor computed (of the first image) has the following structure:

```
<DXLWDOS><Name>Cup1</Name>
<CP><CP><CP><CP><N><P1></P1></N><N>
<P2></P2></N><J></J><P3></P3></CP><P4>
</P4><J></J><P5></P5>
</CP><N><P6></P6></N><J></J><P7></P7></CP>
<D></D><CP><N><P8></P8></N><D></D>
<N><P10></P10></N><N><P11></P11></N></CP><N><
P9></P9></N></CP>
</DXLWDOS>
```

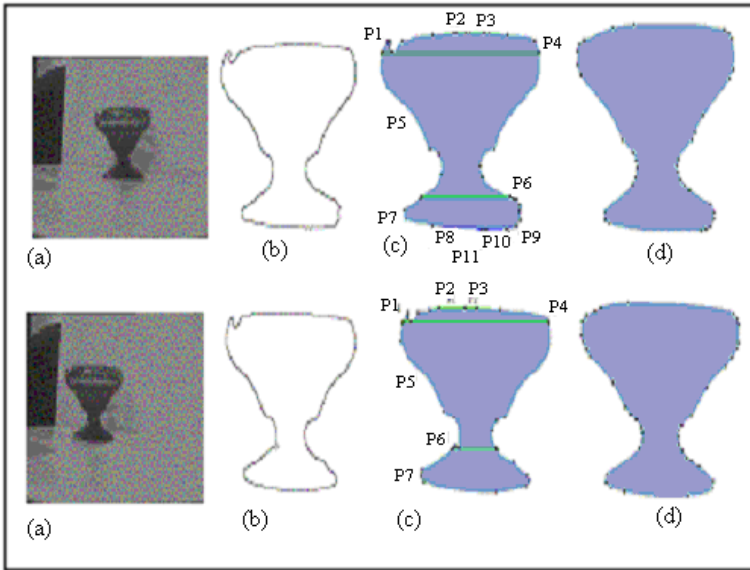



Fig. 5. For each image (a), the extracted outline shape of a cup (b), the result of applying XLWDOS description for outline shape (c) and the smoothed outline shape with $\sigma=10$ (d)

This descriptor is reduced as follows:

```
<DXLWDOS><Name>Cup1</Name>
<CP><P1∪P2∪P3></P1∪P2∪P3>
  <P4></P4><J></J><P5∪P6∪P7∪P8∪P10∪P11∪P9></P5∪P6∪P7∪P8∪P10∪
  11∪P9> </CP>
</DXLWDOS>
```

where: $\langle P1 \cup P2 \cup P3 \rangle \langle /P1 \cup P2 \cup P3 \rangle$ designates the part obtained as the union of the three parts P1, P2 and P3. Therefore the obtained descriptor may be written:

```
<DXLWDOS><Name>Cup1</Name>
  <CP><P'1></P'1><P4></P4><J></J><P'5></P'5></CP> </DXLWDOS>
```

The second noisy descriptor computed (of the second image) has the following structure:

```
<DXLWDOS><Name>Cup2</Name>
<CP><CP>
  <CP><N></N><P1></P1></N><N></N><P2></P2></N><J></J><P3></P3></CP>
<P4></P4><J></J><P5></P5></CP>
  <N></N><P6></P6></N><J></J><P7></P7></CP>
</DXLWDOS>
```

This descriptor is reduced as follows:

```
<DXLWDOS><Name>Cup2</Name>
<CP><P1∪P2∪P3></P1∪P2∪P3>
<P4></P4><J></J><P5∪P6∪P7></P5∪P6∪P7></CP>
</DXLWDOS>
```

Therefore the obtained descriptor may be written:

```
<DXLWDOS><Name>Cup2</Name>
<CP><P''1></P''1><P4></P4><J></J><P''5></P''5> </CP>
</DXLWDOS>
```

Now both the two images descriptors (of cup1 and cup2) can be matched and their similarity is obtained by comparing the elementary contours of the obtained parts.

Two other examples of experiments are given in Figure 6, the same process is applied and similar results are obtained; the descriptors of smoothed shapes become similar.

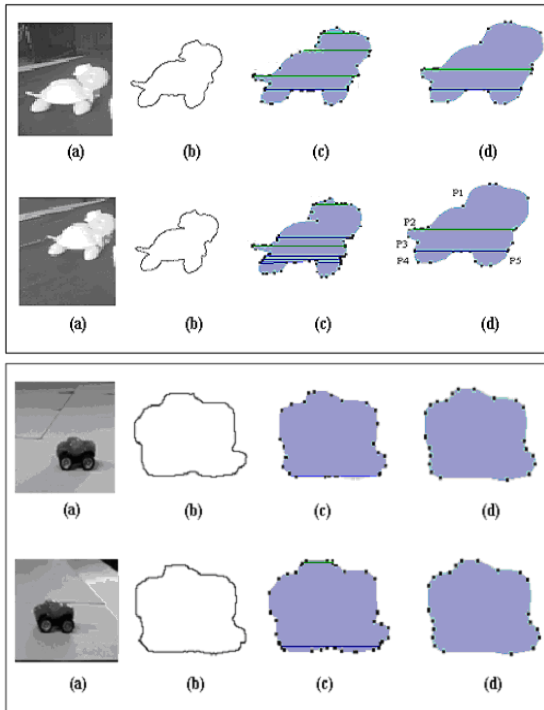


Fig. 6. For each image (a), the extracted outline shape (b), the result of applying XLWDOS description for outline shape (c) and the smoothed outline shape with $\sigma=10$ (d)

6 Conclusion

In this paper we proposed an efficient method for silhouettes matching despite the presence of noise. We have seen the possibility to match noisy silhouettes using their descriptors.

The proposed solution was illustrated by the writing of XLWDOS descriptors including the information of noisy parts. A reduction method has been developed in order to reduce their XML structures and let only main non-noisy parts.

The conducted experiments show the usefulness of the proposed approach and demonstrate the possibility to use XLWDOS descriptors for real images applications.

References

1. Trinh, N.H., Kimia, B.B.: Skeleton Search: Category-Specific Object Recognition and Segmentation Using a Skeletal Shape Model. *IJCV* 94(2), 215–240 (2011)
2. Sclaroff, S.: Deformable prototypes for encoding shapes categories in image databases. *Pattern Recognition* 30(4) (1997)
3. Mokhtarian, F., Mackworth, A.K.: A theory of multiscale, curvature-based shape representation for planar curves. *IEEE PAMI* 14(8) (August 1992)
4. Mokhtarian, F.: Silhouette-Based isolated object recognition through curvature scale space. *IEEE PAMI* 17(5) (1995)
5. Ma, T., Latecki, L.J.: From partial matching through local deformation to robust global shape similarity for object detection. In: *CVPR*, pp. 1441–1448 (2011)
6. Sethi, A., Renaudie, D., Kriegman, D., Ponce, J.: Curve and surface duals and the recognition of curved 3D objects from their silhouettes. *I.J.C.V.* 58(1), 73–86 (2004)
7. Wang, X., Bai, X., Liu, W., Latecki, L.J.: Feature context for image classification and object detection. In: *CVPR 2011*, pp. 961–968 (2011)
8. Orrite, C., Herreo, J.E.: Shape matching of partially occluded curves invariant under projective transformation. *Computer Vision and Image Understanding* 93(1) (2004)
9. Koenderink, J.J., Doorn, V.: The internal representation of solid shape with respect to vision. *Biol. Cyber.* 32 (1976)
10. Cyr, C.M., Kimia, B.B.: A similarity-based aspect-graph approach to 3D object recognition. *International Journal of Computer Vision* 57(1), 5–22 (2004)
11. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on P.A.M.I.* 24(24) (2002)
12. Ruberto, C.D.: Recognition of shapes by attributed skeletal graphs. *Pattern Recognition* 37(1), 21–31 (2004)
13. Zhu, S.C., Yuille, A.L.: FORMS: A flexible object recognition and modeling system. In: *Fifth International Conference on Computer Vision*, June 20–23. M.I.T., Cambridge (1995)
14. Siddiqi, K., Kimia, B.B.: A shock grammar for recognition. In: *Conference of Computer Vision and Pattern Recognition* (1996)
15. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(5) (May 2004)
16. Geiger, D., Liu, L., Kohn, R.V.: Representation and self-similarity of shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(1) (January 2003)

17. Zhang, D., Lu, G.: Review of shape representation and description techniques. *Pattern Recognition* 37(1), 1–19 (2004)
18. Campbell, R.J., Flynn, P.J.: A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding* 81, 166–210 (2001)
19. Larabi, S., Bouagar, S., Trespaderne, F.M., de la Fuente Lopez, E.: LWDOS: Language for Writing Descriptors of Outline Shapes. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 1014–1021. Springer, Heidelberg (2003)
20. Aouat, S., Larabi, S.: Matching Descriptors of Noisy Outline Shapes. *Int. Journal of Image and Graphics* (2010)

Brain MRI Image Segmentation in View of Tumor Detection: Application to Multiple Sclerosis

Rabeb Mezgar¹, Mohamed Ali Mahjoub², Randa Salem³, and Abdellatif Mtibaa¹

¹ Laboratory of Electronics and Microelectronics, Faculty of Sciences of Monastir,
University of Monastir, Tunisia

rabebmezgar@gmail.com, abdellatif.mtibaa@enim.rnu.tn

² Research Unit SAGE (Advanced Systems in Electrical Engineering) Eniso,
University of Sousse

medali.mahjoub@ipeim.rnu.tn

³ Laboratory of interventional Radiology,
University of Monastir, Tunisia

krmranda@yahoo.fr

Abstract. Multiple Sclerosis (MS) is an inflammatory and demyelization disease that causes the disorder of the central nervous system. Magnetic resonance imaging (MRI) becomes the most important means for a better understanding of the disease. A variety of methods to segment these lesions are available to make the lesions detection less fastidious. So, we use a robust algorithm on EM algorithm that proposes an original detection scheme for outliers. The results obtained are very satisfactory.

Keywords: EM algorithm, Multiple sclerosis, Magnetic Resonance Imaging, Levels-Sets.

1 Introduction

Multiple sclerosis is in fact a disease of the central nervous system characterized by demyelization process located in the white matter resulting in the formation of plaques in relapsing multiple and successive occurring at irregular intervals whose duration is unpredictable. The cause of MS is still partially unknown. It would be rather a multifactorial disease involving primarily environmental and genetic factors. The symptoms are varied: poor balance and vision, abnormal fatigue, tremors, etc.. The major impact on knowledge of the disease and the diagnosis is magnetic resonance imaging (MRI). This also allows to monitors lesions (figure 1) over time. Faced with the increase in the amount of data for each patient, a processing system dedicated to medical image analysis of pathology must be adaptable to a changing camera settings and acquisition robust to a change the quality of images. To this end, there are several known methods: The method used to distinguish T1 brain tissue (white matter (MB), gray matter (GM), the cerebrospinal fluid (CSF)) and T2 modality highlights the lesions and CSF. In this context, the segmentation of brain tissue is a preliminary step in the process of detection of lesions.

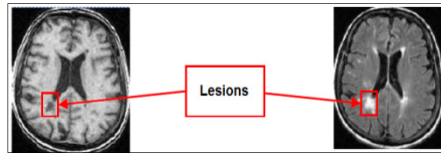


Fig. 1. MS lesions

One approach proposed for the segmentation of MRI brain is the representation of brain structures (white matter (MB), Grey Matter (GM), cerebrospinal fluid (CSF)) as components of mixture distributions for modeling the intensity as a Gaussian mixture. This model allows the modeling of the image intensities by a number of parameters. The parameter estimation is performed by the estimator of maximum likelihood (ML) using optimization methods. This classical approach is based on the Expectation-Maximization algorithm (EM) [1] and, unfortunately, is unsuitable for pathological cases characterized by atypical intensity in MRI lesions caused by MS. Changes are made to obtain the best detection results [2, 3,4].

Schroeter et al. [2] have also used a Gaussian mixture model but adding a uniform component model for the intensities of the lesions. While Van Leemput et al. [3] added a weighting that reflects the degree of "typicality" of each pixel. The data are atypical voxels (3D pixels) satisfying the following condition: their Mahalanobis distance from each component of mixture is greater than a predefined threshold. Dugas-Phocion et al. [4] introduced the calculation of the Mahalanobis distance directly in the iterations of the EM. But still, the results of these methods are still insufficient.

Our goal is therefore to provide a method then the segmentation of lesions for a given time from 3D T2-weighted FLAIR.L proposed algorithm (EM) algorithm is robust against noise and can be applied to many weights and different acquisition times. Our first contribution is the modeling of the exact multi-dimensional problem (the intensities in the image) and the robust estimation of parameters of the blend components. The second stage of our work is the extraction of Multiple Sclerosis lesions in the preamble as "outliers" of the established model. Finally, the proposed improvement is the calculation of the Mahalanobis distance of each pixel from the component mixture, which makes this data atypical of the model, after applying the optimization with the EM algorithm. The segmentation of healthy tissue and damage is then obtained.

The proposed processing chain is explained in Section 2. The following section presents the results and we conclude in Section 4.

2 The Processing Chain

MS lesions can be detected as voxels atypical ("Outliers") in relation to a statistical model (statistical atlas) brain images 'Normal'. In this regard, work has been performed to obtain a probabilistic atlas is used to provide information on statistical parameters of each class of brain tissue. In addition, some researchers [4] propose

methods for obtaining a probabilistic atlas which gives the probability for each pixel to belong to a class segment. Thus, we chose to initialize the statistical parameters with values close to the values of normal patients that the algorithm is applicable to all patient image as follows: Image acquisition, extraction of brain parenchyma, parameter initialization (values of normal patient), tissue segmentation by EM algorithm and lesion extraction.

2.1 The Brain Extraction

In brain MRI, there is the image of the entire head. However, our goal is to focus on the part of brain tissue. So we separate them from the fat, skin... We have therefore used the geometric deformable models (Level Sets). In our case, we applied the method of implicit active contours Caselles et al. [5] with a manual initialization to segment the brain. A Level-Set is the set of points having the same function, so it's an iso contour of a function (x, y, t), defined on the domain of the image. The principle a method of Level-Sets is to evolve a curve by updating a level-set function has fixed coordinates over time. This method is a method based on the contour: the gradient of the image is used to calculate the force function and the curve will be directed towards areas of high gradient. The algorithm has one parameter which is the specific term of propagation c, whose role is to push the contour to be inward or outward. Processing takes an execution time for the important initial contour being married. This treatment also requires several iterations.

2.2 The Segmentation Method: The Algorithm Expectation-Maximization (EM)

The algorithm (Expectation-Maximization) we present in this part was presented by Wells [6] and further developed by Van Leemput in [7, 8], and more particularly relates to the segmentation of brain MRI images. The data of the algorithm are:

X_i : The intensity of the images for each voxel i.

π_i^k : The a priori probability for each voxel i belonging to class k.

The results are to seek: $(\mu_k \Sigma_k)$: Parameters of Gaussians associated with each class

Y_i^k : The posterior probability of belonging to class k.

The a priori probability π_i^k to belong to a class for each voxel k i can be available as an atlas or statistical probabilities. The different parameters can be estimated as follows :

$$\mu_k = \frac{\sum_{i=1}^N x_i Y_i^k}{\sum_{i=1}^N Y_i^k} \tag{1}$$

$$\Sigma_k = \frac{\sum_{i=1}^N (x_i - \mu_k)(x_i - \mu_k)^T Y_i^k}{\sum_{i=1}^N Y_i^k} \tag{2}$$

If, however, the labelings are known, the parameters of classes are directly calculable from the labeling. The updating of posterior probabilities is done by the following formula:

$$\gamma_i^k = \frac{G_{\mu_k \Sigma_k}(x_i) \pi_i^k}{\sum_{l=1}^K \pi_l^k G_{\mu_l \Sigma_l}(x_i)} \quad (3)$$

The intensity distribution of each class of tissue is approximated by a Gaussian $G(\mu_k, \Sigma_k)$ average μ_k and covariance matrix Σ_k . And the reason for the use of mixtures of Gaussian modeling is the presence of different component in each voxel (idea proposed by Schroeter and Al) [9]. The probability distribution in a mixture model [10], $X | \theta$ is defined as a weighted sum of parametric functions K , whose set of parameters θ :

$$P(X|\theta) = \sum_{k=1}^K p((X, Z = k)|\theta) = \sum_{k=1}^K p(Z = k)p(X|(Z = k, \theta)) = \sum_{k=1}^K \pi_k p((X, Z = k)|\theta)$$

Indeed, $p(X | (Z=K, \theta))$ is a combination of K Gaussian parameters (μ_k, Σ_k) . In our model, the parameters are:

$$\theta = \{\pi_1, \pi_2, \dots, \pi_k, \mu_1, \mu_2, \dots, \mu_k, \Sigma_1, \Sigma_2, \dots, \Sigma_k\}$$

The EM algorithm can estimate the parameters θ based on the data X . In this model (model mixture of Gaussian), the membership of each sample to a class k is represented by the hidden variable Z . Estimate would be easy with this information on data labeled. The posterior probability of obtaining a labeling $Z_i = k$ according to the parameters θ_k class k , and x_i is given:

$$\gamma_i^k = p(Z_i = k | X = x_i, \theta_k)$$

The γ_i^k are labeling a posteriori data. The step of expectation is done using Bayes law:

$$\gamma_i^k = \frac{p(Z_i = k)p(X = x_i | Z_i = k, \theta_k)}{p(X = x_i, \theta_k)} = \frac{\pi_k G_{\mu(k), \Sigma(k)}(x)}{\sum_{l=1}^K \pi_l G_{\mu(l), \Sigma(l)}(x)} \quad (4)$$

The iteration of steps 2 and 3 ends when we reach the convergence criterion that is related to the change in the log-likelihood:

$$(L(X|\theta^{(t+1)}) - L(X|\theta^{(t)})) \leq \varepsilon \quad (5)$$

- The space of intensities

The joint histogram is a handy tool to see the space of currents: It represents the distribution of intensities based on grayscale and Gaussian distribution. Representation is possible: the idea is to isolate the "Outliers" by the CSF to put them in a class that decomposes the image into four classes instead of three. Histogram for the difference lies in the histograms (green) of the Gaussian distribution that become four (fig. 2).

- The images space

The result of the algorithm is a mask that represents the posteriori labeling of voxels. The EM algorithm takes as input image to segment (stage supratentorial) and the number of classes. In case $k = 3$, the mask has a good classification of regions and MS lesions are well defined in a whole class (fig. 3). The ownership of the T2 FLAIR provides good contrast for the detection of lesions of Multiple Sclerosis because it distinguishes the cerebra-spinal plates: only take the white lesions (hyper signal).

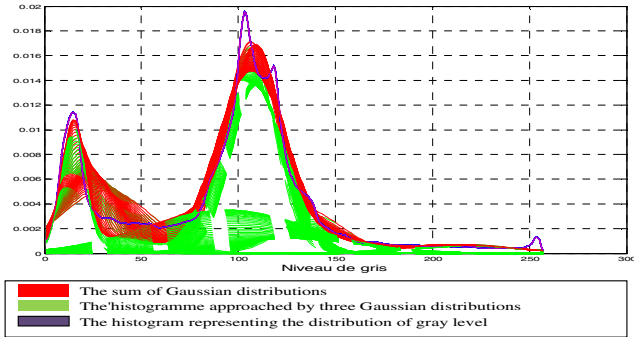


Fig. 2. Histograms representing the distributions of the intensity in the image (Case classification into 4 classes (white matter, gray matter, CSF, "Outliers"))

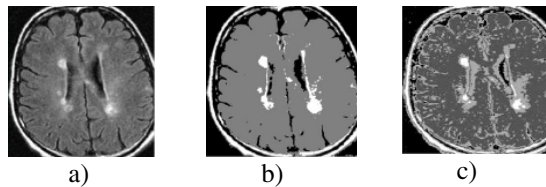


Fig. 3. Segmentation results by EM algorithm a) original image b) segmentation with 3 classes c) 4 classes

- The space of statistical parameters

Table 1 shows the statistical parameters by considering an EM algorithm for 4 classes:

Table 1. Statistical parameters for 4 classes

	Mean	Variance	Probability
Class 1	14	38,7	0,16
Class 2	108,2	156,7	0,42
Class 3	93,26	1776,8	0,37
Class 4	218,17	696	0,03

We have differentiated the CSF of "outliers" where each tissue represents a whole class. In this case to distinguish between the two classes, it should be noted that the class "outliers" is often a very significant variance compared to other classes. So the class concerned in this case is the third class.

2.3 Extracting the Lesions: The Mahalanobis Distance

Once the data model is established by MS, we then seek to extract the "Outliers" (atypical data). For each voxel i of the image, the Mahalanobis distance between the vector y_i and each class k is:

$$D_M(x) = \text{sqrt}((y_i - \mu_k)^T \Sigma_k^{-1} (y_i - \mu_k))$$

DM is the "Mahalanobis distance" from the point x to the mean μ of the distribution. In the implementation algorithms, we sought to determine the distance of the voxels from the voxels belonging to the third class (which contains outliers) so we looked for the difference in values of the voxels relative to the mean and variance of third Class. Voxels representing the lesions are those values above the mean and variance of the third class.

3 Experiments

In case the number of classes is three (MB, MG, LCR + 'Outliers') detection is not complete (Fig. 4). Indeed, there is elimination of noise but also removal of some MS lesions. The execution time is of the order of seconds. The EM algorithm takes 16.72 seconds, while the calculation of the Mahalanobis distance is 20.41 seconds.

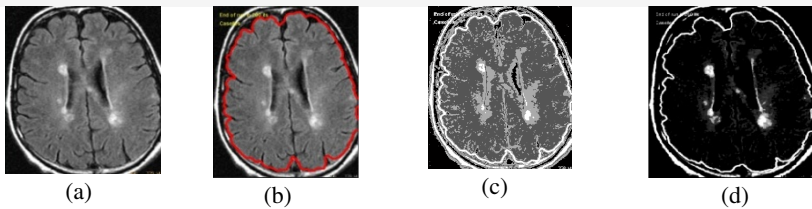


Fig. 4. Results of treatment ($k = 3$) original image b) parameters estimation c) determination of brain parenchyma mask d) lesion detection

If we differentiate CSF and "outliers" we obtain more accurate results (Fig. 5). The execution time of the EM algorithm for four classes is higher 29.55 seconds, because of increasing of iterations number for parameter estimation. The time for calculating the Mahalanobis distance is almost identical 20.99 seconds.

Our main objective is to obtain a better estimation of the parameters of classes, not just to get a good labeling for the detection of lesions. Our processing chain has been very acceptable detection of MS in a short execution time for $k = 4$ (Fig. 5). We managed to obtain clinical validation by a doctor, we have to validate our work by a quantitative assessment.

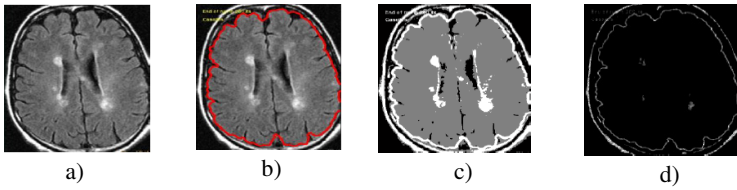


Fig. 5. Results of treatment (k = 4) Original image b) parameters estimation c) determination of brain parenchyma mask d) lesion detection

3.1 Evaluation and Validation of Results

Our algorithm is characterized by a very short execution time and the response is almost instantaneous. The evaluation parameters are:

- Sensitivity (Se): if the area is detected a lesion, the probability of being detected as lesions.
- Specificity (Sp): If the detected area is healthy, it is likely detected as sound.
- Positive predictive value (PPV): If the test is positive probability that the area is actually detected a lesion.
- Negative predictive value (NPV): If the test is negative, the probability that the area detected is actually healthy.
- TP(True Positive):detected as hyperintense pathological lesions.
- TN(True Negative):non pathological hyperintensities detected as such.
- FP(False Positive):non pathological hyperintensities(artifacts,noise..) detected as lesions.
- FN(False Negative):hyperintense pathological ignored by the processing chain.

Table 2. Quality assessment of a drug test

	Lesion	healthy	quantifier
positive test	TP	FP	V_{PP}
negative test	FN	TN	V_{PN}
quantifier	Se	Sp	----

The test is applied on subjects including sick or not lesion volume (column), it can be positive or negative (line). Good sensitivity is a sign that a large majority of lesions are detected as good damage-detection - while a good specificity suggests that few healthy areas are detected as injury - no false alarms. When:

$$Se = \frac{VP}{VP+FN} \quad ; \quad Sp = \frac{VN}{VN+FP} \quad ; \quad V_{PP} = \frac{VP}{VP+FP} \quad ; \quad V_{PN} = \frac{VN}{VN+FN} \quad ;$$

If the test is perfect, there will be no false positive or false negative. In practice, when a doctor sees the results of detection system, positive or negative, the question implies, what is the probability that the detected area is really an injury, knowing that

the review gave a positive (or negative) ? What are the predictive values that correspond to the concerns of doctors, and they might seem evaluation parameters the most significant. But really, it is the specificity and sensitivity that are most often used to assess the additional tests.

3.2 Quantitative Evaluation of Segmentation of MS Lesions

We validated our processing chain on T2 FLAIR MRI images of 15 patients by Multiple Sclerosis. The results of the chain of detection vary from case to case. Several factors affect the detection of lesions: the type of lesions, the presence of flow artifacts (false positives) and the presence of "black holes" (false negatives). The result of detection chain of MS lesions is presented in Fig. 6 concerning the first patient.

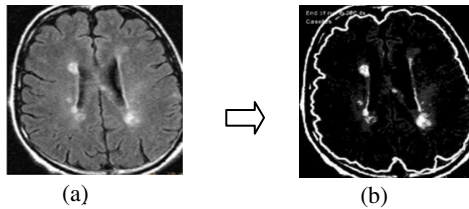


Fig. 6. Results of treatment of first patient original image b) lesion detection

The results of detection of per ventricular lesions are acceptable in terms of sensitivity and specificity (Table 3). The majority of lesions detected are well except that the horns of the ventricles and the space between the two ventricles often suffer from artifacts in T2 FLAIR flow, causing false positives.

Table 3. Evaluation of quality detection

	Lesion	healthy	quantifier
positive test	5	2	0.714
negative test	0	0	0
quantifier	1	0	---

3.3 The ROC (Receiver Operating Characteristics)

When a review provides the results of the continuous type, it determines the best detection limit among the pathological values. The ideal would be to obtain a sensitivity and specificity equal to 1. This is usually not possible, and we must try to obtain the highest values for these parameters, knowing that they vary in opposite directions. For a good choice, we make a graphical tool, the ROC curve (Fig. 7). It is the plot of the values of the sensitivity versus $1 - Sp$. Just then look for the point of the curve that is closest to the point coordinate ($Se = 1; 1 - Sp = 0$).

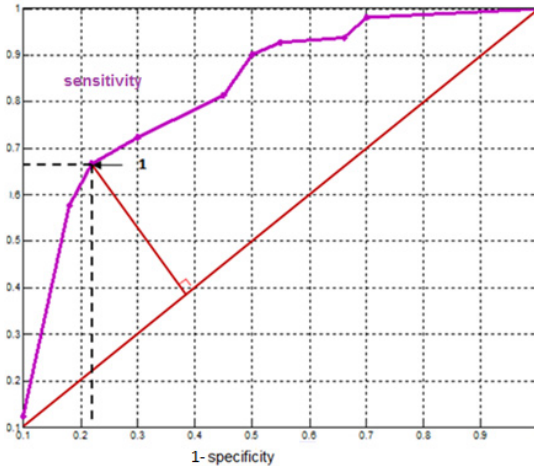


Fig. 7. ROC curve for l'examen of 15 patients

In this figure, the curve is a good diagnostic test for which we can simultaneously obtain high values of sensitivity and specificity. The threshold is the desired coordinates satisfying a maximum value of sensitivity and specificity (in the ideal case: 1, $1 - Se = Sp = 0$). Example of choice of an optimal threshold value with a ROC curve: a threshold value (sensitivity and specificity = 0.723 = 0.8) is obtained by finding the point on the curve farthest from the diagonal. The area under the curve associated with greater than 0.5 (≈ 0.75) so this is a test more or less perfect (a perfect test, the area under the curve is associated with 1). There are on average $VP = 7.6$ true positives and $FP = 1.46$ 8.92 false positives among the lesions seen by physicians. Although the number of false positives is not important compared to true positives but there is always a degradation factor because in the clinical examination should distinguish the lesions in order to properly assess the influence of treatment over several flares.

Table 4. Detection: Calculation of quantifiers

	Labeled as lesions	Labeled as healthy	Quantifiers
detected as lesions	7,6	1,46	$V_{pp}(\text{vox})=0,83$
Detected as healthy	1,32	----	----
Quantifiers	$Se(\text{vox})=0,85$		

In view of results, the segmentation system can be described as sensitive: $Se = 85\%$ of lesions are detected. And we see that 27% ($1 - V_{pp}$) lesions machines are false positives. For the detection of lesions of multiple sclerosis, we achieve good sensitivity. As part of a diagnostic aid, for example, a false negative lesion that is not detected by the chain, must be sought by the doctor, which costs a lot of time. On the other hand, a false positive can be corrected afterwards: a manual correction by an expert may be added a priori by the introduction of a probabilistic atlas.

4 Conclusion

This work is included in the development of approaches to help specialists to make accurate decisions in monitoring patients with lesions in the white matter. Our contribution is the development of a processing chain which aims to detect lesions as "Multiple Sclerosis". To achieve this, the segmentation of MRI brain is the essential step in the detection process. Thus, we proposed an efficient algorithm for segmentation of MRI T2-FLAIR based on EM algorithm. The convergence of such iterative scheme has been demonstrated. The initialization of parameters is done randomly; in order to improve our results we initialized them using a brain atlas flunked our data to rigidly. To improve the detection results of the processing chain, it is necessary to eliminate these false positives. In the literature, approaches are proposed to address this problem that are moving towards the introduction of a priori knowledge by using the statistical atlas, the integration of certain processes to improve the visual quality of the segmentation. We can think about different methods of segmentation such as hidden Markov chain prior to our detection to include constraints on the neighborhood. We can also think about separating the outliers by searching for a method to determine a threshold that can differentiate the "Outliers" in cerebrospinal fluid.

References

1. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39(1), 1–38 (1977)
2. Schroeter, P., Vesin, J.M., Langenberger, T., Meuli, R.: Robust Parameter Estimation of Intensity Distributions for Brain Magnetic Resonance Images. *IEEE Transactions on Medical Imaging* 17(2), 172–186 (1998)
3. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging* 20(8), 677–688 (2001)
4. Dugas-Phocion, G., González, M.A., Lebrun, C., Chanalet, S., Bensa, C., Malandain, G., Ayache, N.: Hierarchical Segmentation of Multiple Sclerosis Lesions in Multi-sequence MRI. In: *ISBI 2004, Arlington, USA (April 2004)*
5. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *Int. J. of Computer Vision* 22, 61–79 (1997)
6. Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive Segmentation of MRI Data. In: Ayache, N. (ed.) *CVRMed 1995. LNCS, vol. 905, pp. 59–69. Springer, Heidelberg (1995)*
7. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging*, 885–896 (October 1999)
8. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging* 897–908 (October 1999)

9. Dugas-Phocion, G.: Segmentation d'IRM Cérébrales Multi-Séquences et Application à la Sclérose en Plaques, Thèse de doctorat (2006)
10. McLachlan, G., Peel, D.: Finite mixture models. Wiley Series in Probability and Statistics (2000)
11. Flandin, G.: Utilisation d'informations géométriques pour l'analyse statistique des données d'IRM fonctionnelle. PhD thesis, Université de Nice-Sophia Antipolis (April 2004)
12. Mahjoub, M.A., Kalti, K.: Image Segmentation by Adaptive Distance Based on EM Algorithm. In: IJACSA, pp. 19–24 (2010)

3D shape Retrieval Using Bag-of-Feature Method Basing on Local Codebooks^{*}

El Wardani Dadi¹, El Mostafa Daoudi¹, and Claude Tadonki²

¹University of Mohammed First, Faculty of Sciences, LaRi Laboratory, Oujda, Morocco
wrd.dadi@gmail.com, m.daoudi@fso.ump.ma

²Mines ParisTech, Laboratory of Research in Computer, Math & System,
Fontainebleau, France
claude.tadonki@mines-paristech.fr

Abstract. Recent investigations illustrate that view-based methods, with pose normalization pre-processing get better performances in retrieving rigid models than other approaches and still the most popular and practical methods in the field of 3D shape retrieval [9,10,11,12]. In this paper we present an improvement of the BF-SIFT method proposed by Ohbuchi et al [1]. This method is based on bag-of-features to integrate a set of features extracted from 2D views of the 3D objects using the SIFT (Scale Invariant Feature Transform [2]) algorithm into a histogram using vector quantization which is based on a global visual codebook. In order to improve the retrieval performances, we propose to associate to each 3D object its local visual codebook instead of a unique global codebook. The experimental results obtained on the Princeton Shape Benchmark database [3] show that the proposed method performs better than the original method.

Keywords: 3D Content-based Shape Retrieval, bag-of-features, SIFT, vector quantization, codebook.

1 Introduction

Currently, there are an increasing number of 3D objects on the web, including large databases, thanks to recent digitizing and modeling technologies. The need of efficient methods for 3D shape-content based retrieval, in order to ease navigation into related large databases, and also to structure, organize and manage this new multimedia type of data, has become an active topic in various research communities such as *computer vision*, *computer graphics*, *mechanical CAD*, and *pattern recognition*.

One major challenge in 3D objects indexation is to design an efficient canonical characterization of the objects. In the literature, this characterization is referred to as a

^{*} E.W. DADI is supported by the "Excellence Grant of Moroccan Ministry of Higher Education. Grant No. G 08/004".

descriptor or a signature. Since the descriptor serves as a key in the search process, it is a critical kernel with a strong influence on the searching performances (i.e. computational efficiency and relevance of the results).

Various 3D shape description methods have been proposed in the literature. The reader may refer to a very good survey in [4] and a comparative study of 3D retrieval algorithms [5, 6, 7, 8]). Those algorithms can be clustered into two main families: 2D/3D approaches and 3D/3D approaches. For 2D/3D approaches, the description model is obtained through different 2D projections of the 3D shape, whereas for the 3D/3D approaches, the description model is obtained from the 3D information directly extracted from the 3D shape. Recent investigations illustrate that view-based methods with pose normalization pre-processing get better performance in retrieving rigid models than other approaches and still the most popular and practical methods in the field of 3D shape retrieval [9, 10, 11, 12].

Our work presented in this paper is inspired by the BF-SIFT method (Ohbuchi et al [13]), which is based on a global codebook (visual dictionary) used to describe each 3D objects in the database. We propose an improvement of the method by using local codebooks, since we think that using a unique global codebook badly influences the retrieval performance. On the Princeton Shape Benchmark (PSB) [3], that contains various shapes with more geometric details, experimental results show that our variant performs better and provides more accurate results.

The rest of the paper is organized as follows. Section 2 presents a description of the BF-SIFT algorithm. In section 3, we present our variant of the BF-SIFT method. Experimental results are provided and analyzed in Section 4. Section 5 opens some perspectives and concludes the paper.

2 The BF-SIFT Method

The BF-SIFT (Bag-of-Features - Scale Invariant Feature Transform) method proposed by Ohbuchi et al [13] compares 3D shapes using thousands of local visual features per model. A 3D model is rendered into a set of depth images, and from each image, local visual features are extracted by using the Scale Invariant Feature Transform (SIFT) algorithm of Lowe [2]. To efficiently compare among a large set of local visual features, the algorithm uses bag-of-features (BoF) approach in order to integrate, for each model, the local features into a vector of features. The BoF approach vector quantifies (or encodes) the SIFT features into a representative vector (or “visual word”), using a global codebook. The global codebook is generated with thousands of features extracted from a set of models in the retrieval database. In the following, we present an overview of the BF-SIFT algorithm:

- *Pose normalization (position and scale)*: The BF-SIFT performs pose normalization only for position and scale, so that the model is rendered with an appropriate size in each of the multiple-view images. Pose normalization is not performed for rotation.
- *Multi-view rendering*: in this step, a set of depth-buffer views of a 3D object are captured uniformly in all directions in order to catch up all symmetries.

- *SIFT feature extraction*: a 3D object can be approximately represented by a set of depth-buffers from which salient SIFT descriptors are extracted using the SIFT algorithm [2].
- *Vector quantization*: each 3D model is associated with thousands of local features. Each SIFT feature extracted from 3D models is quantified as a *vector* or *visual word* by using a global visual codebook. The vector quantification is to find frequencies of visual words (local features) generated from a model in the visual codebook which is learned, by using a clustering algorithm type (e.g. *k-means*, *kd-tree*, *ERC-tree*, and *Locality sensitive hashing*).
- *Histogram generation*: quantified local features or “visual words” are accumulated into a histogram with N_v bins (N_v is considered as the size of the codebook). The histogram becomes the feature vector of the corresponding 3D model.
- *Distance computation*: Dissimilarity among pairs of feature vectors (the histograms) is computed by using Kullback-Leibler Divergence (KLD).

$$D(x, y) = \sum_{j=1}^{N_v} (y_j - x_j) \ln \frac{y_j}{x_j}$$

Where $x = (x_i)$ and $y = (y_i)$ are the features vectors and N_v the dimension of the vectors.

3 Proposed Improvements Based on Local Codebooks

The construction of the visual codebook is one of the sensitive stages. Indeed, the descriptor of each object in the database will be calculated using the visual words in the codebook. For that, it is important to generate codebooks as representative as possible.

We propose to construct a visual codebook for each 3D object in the database. In this case, the vector quantification is based on local codebooks instead of a unique global codebook as the BF-SIFT method. We associate each 3D object with its codebook, which is learned from the features extracted from the 3D model. This is the fundamental difference between the two methods, but all other steps are similar.

After the construction of the visual codebook of each 3D object in the 3D models database, we calculate the corresponding descriptors using the codebook of each object its nearest neighbor in the codebook. This consists in finding the visual words appearance frequencies in the codebook for each 3D object. Thus, each object is associated with a histogram bins. By searching for the nearest neighbor in the codebook, a local descriptor is assigned to each 3D object.

After the vector quantification, the code vectors frequencies, also called “visual words”, are counted to create a histogram for each 3D model, whose number of bins equals the size of the codebook. The histogram is the features vector of the 3D model.

Finally, to compare two 3D objects, using our method, we calculate the descriptor of the 3D object query following the same steps described above and illustrated in Figure 1. We use the codebook of the 3D model with the query well be compared, to vector quantize, the feature vector of 3D object-query.

Using the descriptor, the 3D object query is compared with features vector of a 3D model in the database. The feature vector of the query is calculated every time an occurrence is found in order to be compared with a 3D model in the database.

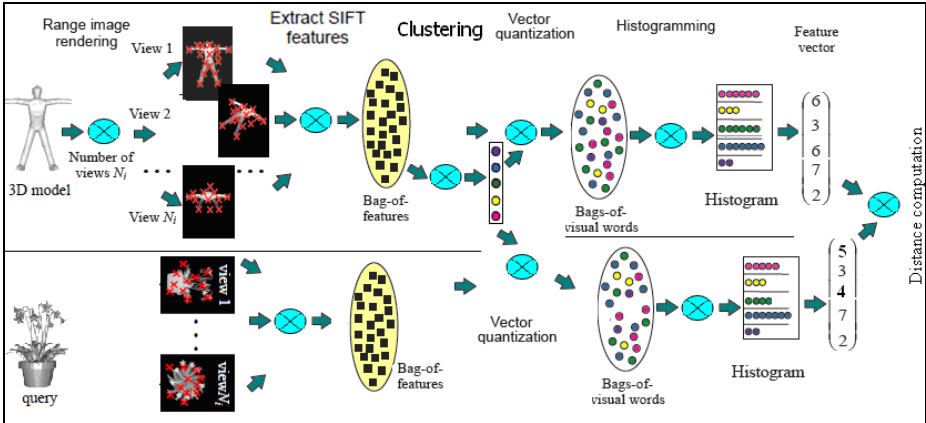


Fig. 1. Processing of comparison between a 3D object-query and a 3D model in database basing on local codebook of the 3D model

4 Experiments and Results

Our tests are made on the Princeton 3D Shape Benchmark database [3] with a set of different and rigid shapes. To fairly compare the retrieval effectiveness and performance of our improvement of the BF-SIFT method, we show the top 10 objects matched between the query models and the retrieved models, both for the original BF-SIFT original and our variant.

For both implementation (original BF-SIFT and our variant), we proceed as follows:

- To extract local feature from a depth-buffer view, the SIFT is implemented with the VLFeat MATLAB source of Veldaldi [19].
- To learn the codebook, we use the k-means MATLAB implementation, also in the VLFeat MATLAB source of Veldaldi [19], in order to cluster the set of local features by setting N_V to the size of vocabulary.
- For vector quantification, we use the MATLAB implementation of the linear k-nearest neighbor (KNN) search.

On figure 2, we report the top 6 objects. Experimental results show that as our method performs better than the BF-SIFT basing on global codebook but at the expense of more computational cost.

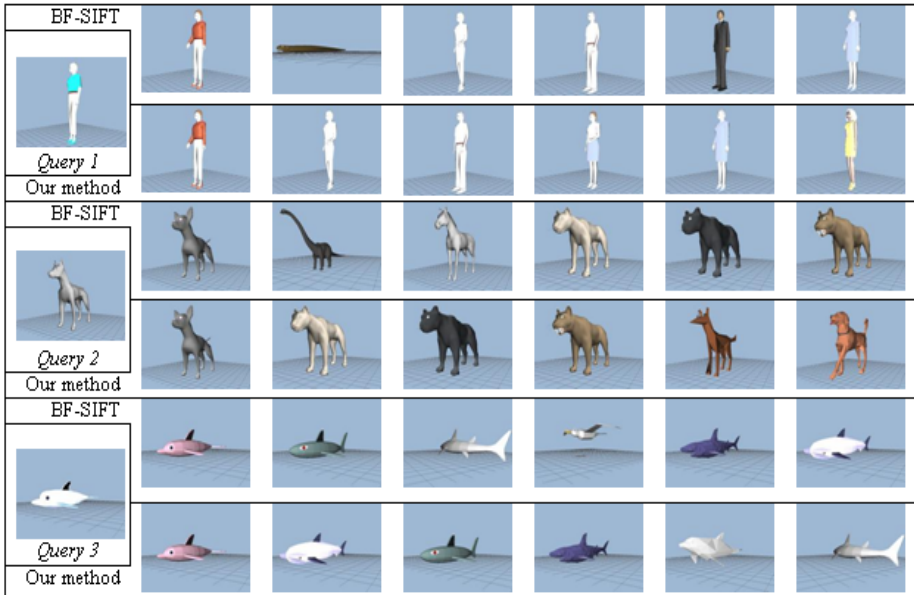


Fig. 2. The 6 top 3D objects retrieved from different “Class” query, using the BF-SIFT and our method

5 Conclusion and Perspectives

In this paper the principal object was to improving the retrieval performance of BF-SIFT. We presented an improvement of the 3D shape retrieval method using Bag-of-Features and SIFT. The key idea is to use local codebooks to vector quantization of salient local features, extracted from a given 3D object, basing on its associated codebook. We are compared the performance of the original BF-SIFT method with our improvement basing on the top-k result of the retrieval in Shape Benchmark database. Experimental results show that our method performs better than BF-SIFT.

References

1. Axenopoulos, A., Daras, P., Dutagaci, H., Furuya, T., Godil, A., Ohbuchi, R.: SHREC 2009 - Shape Retrieval Contest of Partial 3D Models. In: Eurographics Workshop on 3D Object Retrieval (2009)
2. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int'l Journal of Computer Vision* 60(2) (November 2004)
3. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The Princeton Shape Benchmark. In: *Proc. SMI 2004*, pp. 167–178 (2004), <http://shape.cs.princeton.edu/search.html>
4. Tangelder, J.W.H., Velkamp, R.C.: A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications* 39(3), 441–471 (2008)

5. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The Princeton shape benchmark. In: Shape Modeling and Applications Conference, SMI 2004, Genova, Italy, pp. 167–178. IEEE (June 2004)
6. Zaharia, T., Prêteux, F.: 3D versus 2D/3D shape descriptors: A comparative study. In: SPIE Conf. on Image Processing: Algorithms and Systems III - IS &T/SPIE Symposium on Electronic Imaging, Science and Technology 2003, San Jose, CA, vol. 5298 (January 2004)
7. Bustos, B., Keim, D.A., Schreck, T., Vranic, D.: An experimental comparison of feature-based 3D retrieval methods. In: 2nd Int. Symp. on 3D Data Processing, Visualization, and Transmission (3DPVT 2004), Thessaloniki, Greece (September 2004)
8. Del Bimbo, A., Pala, P.: Content-based retrieval of 3D models. *ACM Trans. Multimedia Com.*
9. Daras, P., Axenopoulos, A.: A 3D shape retrieval framework supporting multimodal queries. *International Journal of Computer Vision* (2010)
10. Chaouch, M., Verroust-Blondet, A.: A new descriptor for 2D depth image indexing and 3D model retrieval. In: Proc. ICIP 2007, vol. 6, pp. 373–376 (2007)
11. Lian, Z., Rosin, P.L., Sun, X.: Rectilinearity of 3D meshes. *International Journal of Computer Vision* 89, 130–151 (2010)
12. Lian, Z., Godil, A., Sun, X.: Visual Similarity based 3D Shape Retrieval Using Bag-of-Features. In: IEEE International Conference on Shape Modeling and Applications, SMI (2010)
13. Ohbuchi, R., Osada, K., Furuya, T., Banno, T.: Salient local visual features for shape-based 3D model retrieval, 93-102. In: Proc. IEEE Shape Modeling International, SMI (2008)
14. Tangelder, J., Veltkamp, R.C.: A Survey of Content Based 3D Shape Retrieval Methods. In: Proc. SMI 2004, pp. 145–156 (2004)
15. Iyer, M., Jayanti, S., Lou, K., Kalyanaraman, Y., Ramani, K.: Three Dimensional Shape Searching: State-of-the-art Review and Future Trends. *Computer Aided Design* 5(15), 509–530 (2005)
16. Veltkamp, R.C., et al.: SHREC 2006 3D Shape Retrieval Contest, Utrecht University Dept. Information and Computing Sciences Technical Report UU-CS-2006-030 (ISSN: 0924-3275)
17. Veltkamp, R.C., ter Harr, F.B.: SHREC 2007 3D Shape Retrieval Contest, Dept. of Info. and Comp. Sci., Utrecht University, Technical Report UU-CS-2007-015
18. Chaouch, M., Verroust-Blondet, A.: Alignment of 3D models. *Graphical Models* 71, 63–76 (2009)
19. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms, <http://www.vlfeat.org/>

Segmentation of Prostate Using Interactive Finsler Active Contours and Shape Prior

Foued Derraz^{1,3}, Abdelmalik Taleb-Ahmed³, Azzeddine Chikh⁴, Christina Boydev³,
Laurent Peyrodie², and Gerard Forzy¹

¹ Faculté Libre de Médecine, Institut Catholique de Lille, France

² HEI, LAGIS UMR CNRS 3304, Lille, France

Laurent.peyrodie@hei.fr

³ LAMIH UMR CNRS 8201, Le Mont Houy, 59313 Valenciennes, France

⁴ Biomedical Engineering Laboratory, Technology College, Abou Bekr Belkaid University

foued.derraz@icl-lille.fr, taleb@univ-valenciennes.fr,

az_chikh@hotmail.com, laurent.peyrodie@hei.fr,

gerard.forzy@ghcl.net

Abstract. We present a new interactive segmentation framework to segment the prostate from MR prostate imagery. We first explicitly address the segmentation problem based on fast globally Finsler Active Contours (FAC) by incorporating both statistical and geometric shape prior knowledge. In doing so, we are able to exploit the more global aspects of segmentation by incorporating user feedback in segmentation process. In addition, once the prostate shape has been segmented, a cost functional is designed to incorporate both the local image statistics as user feedback and the learned shape prior. We provide experimental results, which include several challenging clinical data sets, to highlight the algorithm's capability of robustly handling supine/prone prostate segmentation task.

Keywords: Finsler Active contours, characteristic function, shape prior, user interaction.

1 Introduction

Segmentation of the prostate boundary on clinical images is useful in a wide spread range of applications including calculation of prostate volume pre- and post-treatment, radiotherapy planning [1, 2, 16], dosimetry [14], and for creating patient-specific anatomical models [18]. Prostate volume is routinely asked as part of imaging evaluation as it helps in clinical decision making [2, 15]. However, manual segmentation of the prostate boundary is highly time-consuming and subject to inter- and intra-reader variability. Automatic segmentation based on deformable models such as Active Contour (AC) models have been widely used for prostate segmentation [7,12] and can be split into two classes, those which fully rely on image data [9], and those which incorporate prostate prior shape information [6, 10, 11]. To deal with the complex prostate anatomy and partially missing boundaries, the shape of prostate is

approximately assumed to be elliptical. In many cases, this segmentation is a two-part problem. First, one must properly align a set of training shapes such that any variation in shape is not due to alignment. Then, the segmentation based on deformable models can be performed under the constraint of the learned prostate shapes. However, the alignment of prostate shapes becomes increasingly difficult for a large variation in training shapes and when the training sets increase, and this is not readily allowed by existing methods. To overcome this problem, we investigated an interactive FAC model to boost the performance of segmentation results. The FAC model has been proposed as a natural way for adding directionality to the AC model [5]. This allows the AC to favor appropriate locations and suitable directions [5, 12].

In this paper, we proposed a bi-stage interactive prostate segmentation method based on fast globally active contours incorporating prior shape to segment the prostate from MR images. Then, our segmentation method based fast FAC incorporating shape prior is applied to delineate prostate boundary. User intervention is then needed to guide our method to fine-tune the final segmented shape of prostate. Finally, we apply some post-processing operations to further refine prostate boundaries. This paper organized as follows. Section 2. describes the basics of our method. In Section III, the experimental results obtained using our method are illustrated. Conclusions and future work are presented in Section 4.

2 Segmentation Method

The The new framework which we proposed for prostate delineation consists of two main steps: the first step is applying a FAC model incorporating prostate shape prior. The second step is refining the prostate shape by user feedback. Each of these steps is described in detail in next section.

2.1 Finsler Active Contours in the Total Variation Framework

We are interested in a globally interactive segmentation of an object Ω in Total Variation (TV) framework through the characteristic function [3, 9]. We proposed to segment prostate shape using a fast version of the interactive Finsler Active contours in the TV framework. We proposed a two stage fast globally Finsler Active Contours (FAC). In the first step, our fast globally FAC incorporated both statistical region, geometric shape and in the second step, the final segmentation is completed by adding an interactive user feedback. Our fast interactive segmentation can be modelled as the following energy criterion [5, 9,12]:

$$\begin{aligned}
 E_{FAC}(\chi, \Omega, \Omega_{ref}, \Omega_{ROI}) = & \underbrace{\int_{\partial\Omega} k_f(\mathbf{x}, \partial\Omega) da(\mathbf{x})}_{E_{finsler}(\partial\Omega)} + \lambda_1 \underbrace{\int_{\Omega_I} k_r(\mathbf{x}, \Omega) \chi(\mathbf{x}) d\mathbf{x}}_{E_{data}(I, \Omega)} \\
 & + \lambda_2 \underbrace{\int_{\Omega_I} k_s(\mathbf{x}, \Omega_{ref}) \chi(\mathbf{x}) d\mathbf{x}}_{E_{shape}(\Omega, \Omega_{ref})} + \lambda_3 \underbrace{\int_{\Omega_I} k_{user}(\mathbf{x}, \Omega_{ROI}) \chi(\mathbf{x}) d\mathbf{x}}_{E_{user}(I, \Omega, \Omega_{ROI})}
 \end{aligned} \tag{1}$$

Where k_f is the anisotropic boundary descriptor, $da(\mathbf{x})$ is surface element, λ_1, λ_2 and λ_3 are the calibration factors and the characteristic function χ framework is defined as:

$$\chi(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \Omega \\ 0 & \text{if } \mathbf{x} \notin \Omega \end{cases} \quad (2)$$

The region descriptor k_r is defined in the same manner as in [4], k_s is the shape prior descriptor defined in the same manner as in [6] and Ω_{ref} the reference prostate shape (section 2.3) and k_{user} the interactive user feedback descriptor which will be defined in section 2.4.

To allow our segmentation model to detect both edge and its direction, we assume that anisotropic boundary descriptor is defined as follows:

$$k_f(\mathbf{x}, \partial\Omega) = \max_{|\mathbf{p}| \leq 1, \chi} \langle \mathbf{p}, \nabla \chi(\mathbf{x}) \rangle \quad (3)$$

Where the potential field, \mathbf{p} , is defined as $\mathbf{p} = [\vec{N}, \vec{T}]$ and \vec{N} is the unit inward normal vector and \vec{T} the unit tangential vector (Fig. 1). This potential field allows the FAC to deform and move toward object of interest. This propriety makes the proposed segmentation method much faster since the topology of the deformed curve is more like the object to be segmented. The energy of FAC is given by:

$$\begin{aligned} E_{FAC}(\chi, \mathbf{p}) = & \int_{\Omega} \underbrace{\max_{|\mathbf{p}| \leq 1} \langle \mathbf{p}, \nabla \chi(\mathbf{x}) \rangle}_{\psi_{\mathbf{x}, \mathbf{p}}(\mathbf{x})} da(\mathbf{x}) + \lambda_1 \int_{\Omega_0} k_r(\mathbf{x}, \Omega) \chi(\mathbf{x}) d\mathbf{x} \\ & + \lambda_2 \int_{\Omega_0} k_s(\mathbf{x}, \Omega_{ref}) \chi(\mathbf{x}) d\mathbf{x} + \lambda_3 \int_{\Omega_0} k_{user}(\mathbf{x}, \Omega_{ROI}) \chi(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (4)$$

and the respective gradient descent (for χ) and ascent (for the potential field \mathbf{p}) equations are:

$$\begin{cases} \frac{\partial \chi(\mathbf{x}, \tau)}{\partial \tau} = \text{div}(\mathbf{p}) - \lambda_1 k_r(\mathbf{x}, \Omega) - \lambda_2 k_s(\mathbf{x}, \Omega_{ref}) - \lambda_3 k_{user}(\mathbf{x}, \Omega_{ROI}), \chi(\mathbf{x}, \tau = 0) = \mathbf{1} \\ \frac{\partial \mathbf{p}}{\partial \tau} = -\nabla \chi(\mathbf{x}, \tau) \end{cases} \quad (5)$$

where τ is an artificial time parameter, χ_0 is the initialized characteristic function corresponding to the initial contour curve $\partial\Omega_0$. In the next section, we introduced region based term used in our fast segmentation based model.

2.2 Region Energy Term

The second energy term incorporated in our Finsler AC is defined as region energy based term. We proposed in to incorporate this term as statistical Bhattachryya distance [4]. The region energy based term is usually defined as a domain integral of the region descriptor k_r :

$$E_{data}(I, \Omega) = \int_{\Omega_0} k_r(\mathbf{x}, \Omega) d\mathbf{x} = \sqrt{\int_{\Omega_0} p_f(I, \Omega) p_b(I, \Omega) d\mathbf{x}} \tag{6}$$

In this study we maximize the statistical Bhattachryya distance between the foreground density probability $p_f(I, \Omega)$ of the object to be segmented and the background density probability $p_b(I, \Omega)$ such as in [12]. The region velocity $k_r(\mathbf{x}, \Omega)$ is expressed [10] as:

$$k_r(\mathbf{x}, \Omega) = \frac{1}{2} \left\{ \begin{aligned} & \left(\frac{1}{|\Omega_f|} - \frac{1}{|\Omega_b|} \right) \sqrt{p_f(I, \Omega_f) p_b(I, \Omega_b)} + \frac{1}{|\Omega_b|} \int_{R^+} \sqrt{\frac{p_f(I, \Omega_f)}{p_b(I, \Omega_b)}} \\ & \left(G_{\sigma_{ker}}(I - I(\Omega_f)) - \sqrt{p_b(I, \Omega_b)} \right) dI \\ & - \frac{1}{|\Omega|} \int_{R^+} \sqrt{\frac{p_b(I, \Omega_b)}{p_f(I, \Omega_f)}} \left(G_{\sigma_{ker}}(I - I(\Omega_f)) - \sqrt{p_f(I, \Omega_f)} \right) dI \end{aligned} \right\} \tag{7}$$

We therefore estimate probability density by Parzen kernel, which can better describe the regions (see Fig. 1.b):

$$p_{f/b}(I, \Omega) = \frac{1}{|\Omega_{f/b}|} \int_{\Omega_{f/b}} G_{\sigma}(I - I(\Omega_{f/b})) d\mathbf{x} \tag{8}$$

where G_{σ} denote the Gaussian kernel and σ^2 the variance.

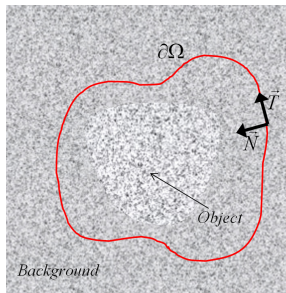


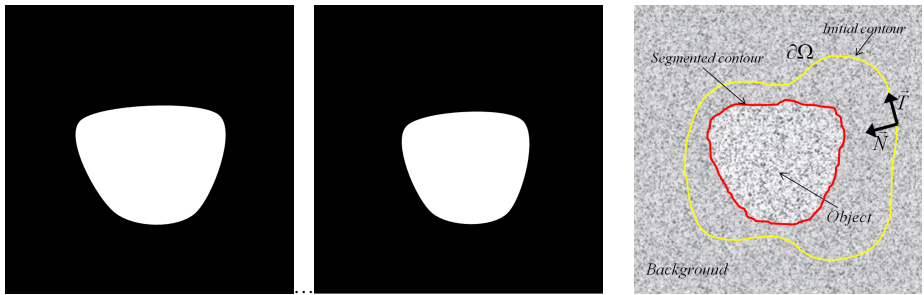
Fig. 1. Evolving FAC curve around a synthetic object guided by both normal and tangential components

2.3 Shape Prior Energy Term

The shape prior descriptor is defined as the Euclidean distance between the evolving Legendre moment [6] region $\eta(\Omega)$ and the reference shapes $\{\Omega_{ref}^i, i = 1, \dots, N\}$:

$$E_{shape}(\Omega, \Omega_{ref}) = \int_{\Omega} k_{shape}(\mathbf{x}, \Omega_{ref}) d\mathbf{x} = \sum_{\substack{p+q \leq N \\ p, q}} \left| \eta_{pq}(\Omega) - \eta_{pq}(\Omega_{ref}^i) \right|^2 \quad (9)$$

where the η_{pq} are defined as follows, using the geometric moments and coefficients of the Legendre polynomials [6]. In the figure 3, the object shape is segmented using 20 learned shape prior and the segmentation is done using statistical and geometric shape prior. In the next section we introduced our interactive user term used to suitability segmentation method (see Fig. 2).



a) shapes learned as binaries surfaces b) segmentation of the object using both statistical and geometric shape, yellow color initial contour curve and in red final segmentation done by our FAC

Fig. 2. Automatic segmentation using FAC based statistical and geometric shape prior

2.4 User Energy Term

Let $\mathbf{x}_i, i = 1, \dots, n$ denote the set of n user feedback points. We define the user feedback function $L: \Omega \rightarrow \mathbb{R}$ as follows:

$$L(\mathbf{y}) = \chi(\mathbf{y}) + \{1 - \chi(\mathbf{y})\} \sum_{i=1}^n \int_{\mathbf{z} \in \Omega_{ROI}} \delta(\mathbf{z} - \mathbf{x}_i) d\mathbf{z} \quad (10)$$

where $\delta(\mathbf{z})$ is the Dirac function and Ω_{ROI} is local region. Hence, for each $\mathbf{y} \in \{\mathbf{y}_i\}_{i=1}^n$:

$$L(\mathbf{y}) = \begin{cases} 0 & \mathbf{y} \in \Omega_{\chi} \\ 1 & \mathbf{y} \in \Omega \setminus \Omega_{\chi} \\ \chi(\mathbf{y}) & \text{not marked} \end{cases} \quad (11)$$

$L(\mathbf{x})=0$ if the feedback point is within the segmented region of the first phase and $L(\mathbf{x})=1$ if the feedback point is situated in the background. Finally, if the pixel \mathbf{x} is not marked, then the indicator function is identical to indicator function ($L(\mathbf{x}) = \chi(\mathbf{x})$). (see Fig. 3.) The indicator function $L(\mathbf{x})$ is used in the formulation of the energy term which incorporates the user feedback:

$$E_{user}(I, \Omega, \Omega_{ROI}) = \int_{\mathbf{x} \in \Omega} \int_{\mathbf{y} \in \Omega_{ROI}} |L(\mathbf{y}) - \chi(\mathbf{x})|^2 e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma^2}} d\mathbf{x}d\mathbf{y} \tag{12}$$

The algorithm supports two modes of user feedback. The user may either draw a cross such that its eccentricity and orientation determines the entries of the variance coefficient σ or can provide a point-wise mouse click. The interactive energy functional E_{user} is minimized, w.r.t the evolving domain $\Omega(\tau)$, is done with the shape derivative tool [10,12]. Thus, the user Eulerian derivative of E_{user} in the normal direction is as follows:

$$\left\langle \frac{\partial E_{user}(I, \Omega(\tau), \Omega_{ROI})}{\partial \tau}, \xi \right\rangle = \int_{\partial \Omega} k_{user}(\mathbf{x}, \Omega_{ROI}) \langle \xi, N_{\partial \Omega} \rangle da(\mathbf{x}) \tag{13}$$

where the interactive velocity is expressed as:

$$k_{user}(\mathbf{x}, \Omega) = \int_{\mathbf{y} \in \Omega_{ROI}} |L(\mathbf{y}) - \chi(\mathbf{x})|^2 e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma^2}} d\mathbf{y} \tag{14}$$

In the figure below (Fig. 3) we present final segmentation based interactive Finsler AC in blue color superposed on the automatic segmentation based Finsler AC (red color) using statistical and geometric shape

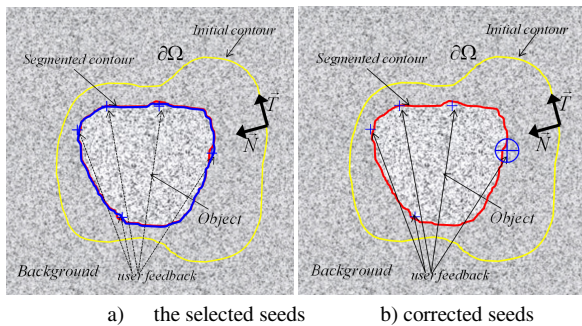


Fig. 3. Segmentation refinement using 5 user feedback, each point selected uses 4 pixel neighbors in Ω_{ROI} .

3 Results

3.1 Data and Protocol

In this section, we provide prostate segmentation results for two data sets obtained from Saint Philibert Hospital Lille France. The MR images are pre-processed through the following pipeline: 1) spatial registration, 2) noise removal and 3) intensity standardization. We use the T1 weighted and T2 weighted MR sequences. The image sizes are 256x256 pixels, each slice thickness is 3.5mm with spacing between slices of 3.9 mm.

3.2 Segmentation Results

The first row shows the segmentation result by the proposed method, the second row illustrates the ground truth outlined by an expert radiologist, and the third row provides a comparison between our result and ground truth. We observe that our segmentation result is sufficiently close to the result provided by a radiologist. In addition to visual evaluation, we use Dice measure (DSC) to quantitatively evaluate the segmentation result. The Dice measure is defined as:

$$DSC(A, B) = 2 \frac{|A \cap B|}{|A| + |B|} \quad (15)$$

where A is the segmentation result, B is the ground truth provided by an expert radiologist, and $|\cdot|$ denotes the number of segmented pixels.

To quantify the accuracy of the segmentation, we measured the Dice Similarity Coefficient (DSC) between the manually segmented prostate and our segmentation method. We provide not only qualitative results, but also give quantitative results in the form of the DSC to illustrate the viability of the proposed method in the context of prostate segmentation (see Fig 4.).

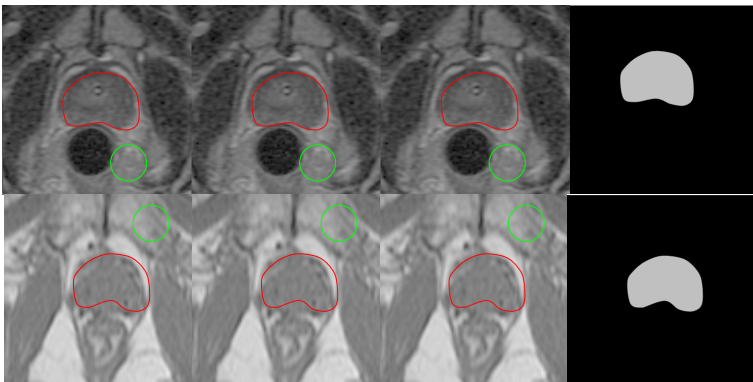


Fig. 4. Segmentation by TV FAC. In yellow color traditional FAC, red color segmentation results of our method. The DSC for the first image 85,63% and the second image of 82,69%.

4 Conclusion

This paper has presented an interactive prostate segmentation MR images method based active contours. The segmentation is achieved in two stages. In the first stage, the patient prostate is segmented using a fast globally FAC incorporating statistical and shape prior knowledge. The position and orientation are dependent on prior for the boundary segmentation in Finsler metrics. Finsler active contours provide an alternative approach to integrating image-based priors on the location and orientation of the traditional boundary descriptor. Future work will address extending other classes of energies that can be optimized in TV framework.

References

- [1] Zwiggelaar, R., Zhu, Y., Williams, S.: Semi-automatic Segmentation of the Prostate. In: Perales, F.J., Campilho, A.C., Pérez, N., Sanfeliu, A. (eds.) *IbPRIA 2003*. LNCS, vol. 2652, pp. 1108–1116. Springer, Heidelberg (2003)
- [2] Villeirs, G., De Meerleer, G.: Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning. *Eur. J. Radiol.* 63(3), 361–368 (2007)
- [3] Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J., Osher, S.: Fast Global Minimization of the Active Contour/Snake Model. *JMIV* 28(2) (2007)
- [4] Michailovich, O., Rathi, Y., Tannenbaum, A.: Image Segmentation Using Active Contours Driven by the Bhattacharyya Gradient Flow. *IEEE Trans. IP* 16(11), 2787–2801 (2007)
- [5] Melonakos, J., Pichon, E., Angenent, S., Tannenbaum, A.: Finsler active contours. *IEEE Trans. PAMI* 30(3), 412–423 (2008)
- [6] Foulonneau, A., Charbonnier, P., Heitz, F.: Affine-Invariant Geometric Shape Priors for Region-Based Active Contours. *IEEE Trans. PAMI* 28(8), 1352–1357 (2006)
- [7] Pasquier, D., Lacornerie, T., Vermandel, M., Rousseau, J., Lartigau, E., Betrouni, N.: Automatic Segmentation of Pelvic Structures From Magnetic Resonance Images for Prostate Cancer Radiotherapy. *Int. Jnl. of Radiation Oncology, Biology, Physics* 68(2), 592–600 (2007)
- [8] Mahdavi, S., Chng, N., Spadinger, I., Morris, W.J., Salcudean, S.E.: Semi-automatic segmentation for prostate interventions. *Medical Image Analysis* 15(2), 226–237 (2011)
- [9] Derraz, F.: Optimal segmentation by fast binary geometric active contours, PhD Thesis (2010)
- [10] Duay, V., Houhou, N., Thiran, J.P.: Atlas-based segmentation of medical images locally constrained by level sets. In: *IEEE ICIP 2005*, vol. 2, pp. 1286–1289 (2005)
- [11] Martin, S., Daanen, V., Troccaz, J.: Atlas-based prostate segmentation using an hybrid registration. *Int. J. CARS* 3, 485–492 (2008)
- [12] Aubert, G., Barlaud, M., Faugeras, O., Jehan-Besson, S.: Image segmentation using active contours: Calculus of variations or shape gradients? *SIAM Applied Mathematics* 63 (2002)
- [13] Klein, S., Staring, M., Pluim, J.P.W.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans. Image Process.* 16(12), 2879–2890 (2007)

- [14] Pasquier, D., Peyrodie, L., Denis, F., Pointreau, Y., Bera, G., Lartigau, E.: Segmentation automatique des images pour la planification dosimetrique en radiotherapie. *Cancer/Radiotherapie* 14(S.1), 6–13 (2010)
- [15] Vikal, S., Haker, S., Tempany, C., Fichtinger, G.: Prostate contouring in MRI guided biopsy. In: *SPIE Conf.*, vol. 7259, p. 144 (2009)
- [16] Pasquier, D., Lacornerie, T., Vermandel, M., Rousseau, J., Lartigau, E., Betrouni, N.: Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *Int. J. Radiat. Oncol., Biol., Phys.* 68(2), 592–600 (2007)
- [17] Liu, X., Langer, D.L., Haider, M.A., Van der Kwast, T.H., Evans, A.J., Wernick, M.N., Yetik, I.S.: Unsupervised Segmentation of the Prostate Using MR Images Based on Level Set with a Shape Prior. In: *IEEE EMBC 2009* (2009)

Tracking Moving Objects in Road Traffic Sequences

Salma Kammoun Jarraya¹, Najla Bouarada Ghrab¹,
Mohamed Hammami², and Hanene Ben-Abdallah¹

¹MIRACL-FSEG, Sfax University, Rte Aeroport Km 4, 3018 Sfax, Tunisia

²MIRACL-FS, Sfax University, Rte Soukra Km 3, 3018 Sfax, Tunisia
{Salma.Kammoun, Hanene.Benabdallah}@fsegs.rnu.tn,
Najla.Bouarada@yahoo.fr, Mohamed.Hammami@fss.rnu.tn

Abstract. In this paper, we present an algorithm for tracking objects in road traffic sequences which is based on coherent strategy. This strategy relies on two times processing. Firstly, a Short-Term Processing (STP) based on spatial analysis and multilevel region descriptors matching allows identification of objects interactions and particular objects states. Secondly, a Long-Term Processing (LTP) is applied to cope with track management issues. In fact LTP feedbacks objects and their corresponding regions in each frame to update tracked object attributes. In case of merging objects, attributes are obtained using Template matching. An experimental study by quantitative and qualitative evaluations shows that the proposed approach can deal with multiple rigid objects whose sizes vary over time. The obtained results prove that our method can provide an effective and stable road objects tracks.

Keywords: Tracking moving object, foreground segmentation, point descriptors, template matching.

1 Introduction

Road traffic monitoring has become a very important research area. Such system is based essentially on tracking road objects. The aim of tracking object is to estimate the trajectory of moving objects over time. The information gathered by road objects tracking can help to identify their behavior in the observed scene and allows building statistical information about road traffic.

The purpose of our contributions is to track multiple rigid moving objects (road objects) with different sizes and speeds. Note that moving objects are detected automatically. The proposed method for object tracking takes in consideration the possibly states changes of moving objects and interactions between them. In addition, appearance of a new object and disappearance of existing object are managed automatically.

The reminder of this paper is divided into 4 sections. In Section 2, we describe a brief state of art in object tracking. Section 3 presents our proposed method. Section 4 outlines the results of a quantitative and a qualitative evaluation. Finally, Section 5 recapitulates the presented method and outlines future work.

2 State of the Art in Object Tracking

Several methods [1]-[4] deal with object tracking; however their accuracy depends on, both, constraints and context of the application. Constraints are related to the tracked object(s) (single or multiple, rigid or non-rigid), to the camera (single or multiple, mobile or fixed) and to the observed scene (indoor and/ or outdoor). Dealing with context, we distinguish different applications as person tracking, road object tracking, ball tracking, etc. In this paper, we focus particularly on road object tracking. In addition to constraints of application context, methods reported in literature differ by their manner to represent object. We can classify these methods in two categories of approaches which are (1) Points based approach [*cf.* 1,2] and, (2) Model based approach [*cf.* 3,4] (silhouettes or kernel). In these approaches, tracking strategy relies on matching information provided by points/models over times. Points based methods are fast and can deal with partial occlusions. However, cannot usually cope with complex deformation of nonrigid objects. In model based approach, silhouettes model allows tracking of both nonrigid and rigid objects but their computations is very expensive and lack of generality. Unlike silhouettes model, kernel based model can be obtained without knowledge about object nature or shape but cannot resolve occlusions. Within the context of road traffic, methods aim to track unlimited number of rigid road objects in video stream. Thus, we adopt points based approach.

In literature, several methods [1,2][5]-[8] are based on descriptors points. Among of techniques to compute descriptor points are Harris detector [9], KLT (Kanade-Lucas-Tomasi) detector [10] and SIFT descriptor (Scale Invariant Feature Transform)[11]. Our comparative study between these techniques, shows that descriptors from SIFT are invariant to different invariance criteria (*Translation, Scale Changes, Image Rotation, Illumination changes, Image Locale Deformations, and Affine Transformation*). In addition, from theatrical point of view, SIFT technique can (1) produce a great number of descriptor points, (2) give a local image measurement that is robust to noises and to partial occlusions and (3) give distinctive points.

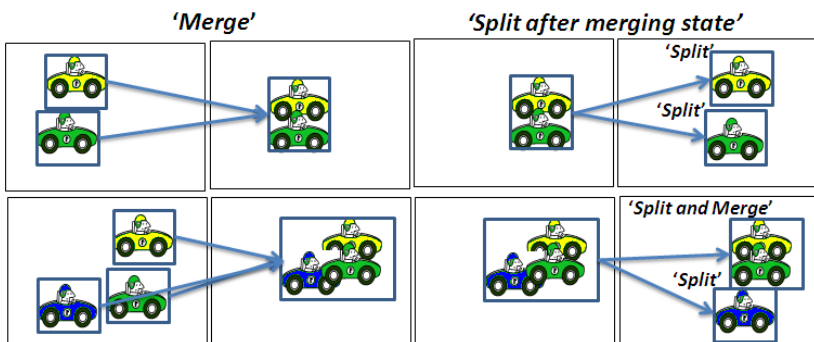


Fig. 1. Examples of interactions between road objects

Note that the success of such tracking field relies on the management of both frequently object state changes (life cycle) and interactions. Life cycle of objects road start by their appearance in the scene (state 'Entry') and ended by their disappearance

(state 'Exit'). During its presence in the scene, a road object can be in normal state ('Normal'), normal state with a high speed ('Normal HS'), stopped ('Stopped'), restart motion after stopped ('Re-moving').

During a life cycle, two types of interactions between objects road can occur (Fig.1). The first interaction happens when two or many objects appear close to each other ('Merge') causing partial or total occlusion. The second interaction results of two or many objects fragmentation ('Split') after merging state.

The most recent tracking methods based on SIFT technique (cf. [5]-[8]) track pre-selected (single or limited number) specific object(s). In addition, states changes of moving objects are not considered. Furthermore, appearance and disappearance of objects are managed according to a region of interest drawing manually.

3 Proposed Method

Our proposed method for tracking road objects is based on three main steps: (1) Foreground segmentation, (2) Short-Term Processing (STP) and (3) Long-Term Processing (LTP). In fact, let $R_t^{cc=1...m}$ and $R_{t-1}^{c=1...n}$ denote respectively the segmented regions from frames F_t and F_{t-1} with $cc \in \{1, \dots, m\}$ and $c \in \{1, \dots, n\}$, n and m , are respectively the number of region in F_t and F_{t-1} . Foreground segmentation is done by the method presented in our previous work [12]. This method is based on background modeling approach; it demonstrates robust and accurate results under most of the common problem in foreground segmentation. In STP, $R_t^{cc=1...m}$ and $R_{t-1}^{c=1...n}$ are used to manage objects states and interactions for each input frame, thus produces region correspondence ($R_{j=(t-1,t)}^{i=\{c,cc\}}.Cor$) and state ($R_{j=(t-1,t)}^{i=\{c,cc\}}.State$). LTP establish all objects tracks ($TrackingObject\{O_{i=1...ObjectCount}\}$) between $t=0$ and t based on $R_{j=(t-1,t)}^{i=\{c=1...n,cc=1...m\}}.Cor$ and $R_{j=(t-1,t)}^{i=\{c=1...n,cc=1...m\}}.State$ to generate objects trajectories. In the following subsections, we detail the STP and the LTP steps.

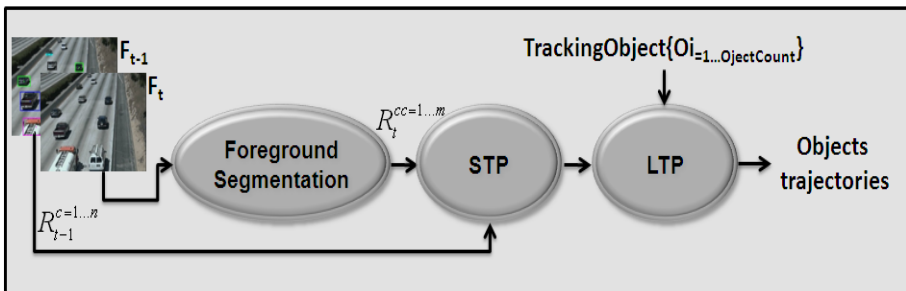


Fig. 2. Flowchart of the proposed tracking method

3.1 Short-Term Processing (STP)

Short-Term Processing takes into account, (1) the spatial analysis and, (2) Multilevel region descriptors matching of $R_t^{cc=1...m}$ and $R_{t-1}^{c=1...n}$. Each region R is represented by a

set of attributes ($Z(R) = (\beta^{1\dots 5}(R), \phi_k^{128}(R))$). Where $\beta^{1\dots 5}(R)$ are 2D spatial attributes (cf. Fig. 3) and $\phi_k^{128}(R)$ is a K -by-128 matrix, each row gives an invariant descriptor for one of the K keypoints. The descriptor is a vector of 128 values normalized to unit length. Regions correspondences ($R_{j=\{t-1,t\}}^{i=\{c=1\dots n, cc=1\dots m\}}.Cor$) are initialized by -1.

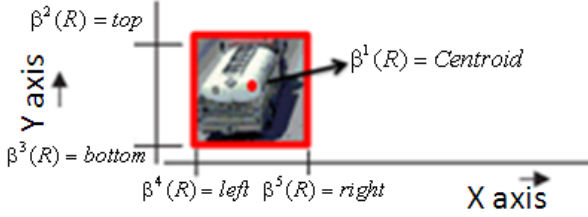


Fig. 3. 2D Spatial attributes ($\beta^{1\dots 5}(R)$)

Spatial Analysis. We project $\beta_t^l(R_t^{cc=1\dots m})$ onto area from $\beta_{t-1}^{2\dots 5}(R_{t-1}^{c=1\dots n})$, thus provides correspondence for regions in 'Normal' states and/or in 'Split' interactions according. Region in state 'Normal' corresponds to the case where $\beta_t^l(R_t^{cc})$ belongs to only one R_{t-1}^c area. The Split interaction corresponds to the case where β_t^l of two or more R_t^{cc} (i.e. $R_t^{cc=1, cc=2\dots}$) belong to one R_{t-1}^c area. We associate regions R_t^{cc} and R_{t-1}^c according to Equation 1.

$$\begin{cases} \text{If 'Normal' state then} \\ R_t^{cc}.Cor = c \\ \text{Else If 'Split' interaction then} \\ R_t^{cc=1, cc=2\dots}.Cor = c \end{cases} \quad (1)$$

Multilevel Region Descriptors Matching. A multilevel region descriptors matching is proposed for $R_{j=\{t-1,t\}}^{i=\{c=1\dots n, cc=1\dots m\}}$ with ($R_{j=\{t-1,t\}}^{i=\{c=1\dots n, cc=1\dots m\}}.Cor = -1$). This step allows us to cope with region interaction ('Merge') and states ('Entry', 'Exit', 'Normal VE', 'Stopped' and 'Re-Moving'). We aim to select, for each region descriptors $\phi_{i=1\dots k_1}^{128}(R_t^{cc})$, its match to $\phi_{k_2}^{128}(R_{t-1}^c)$ (Equation (2)). There is matching (R_Match) between two regions in case of at least one descriptor match (Des_Match). Decision to select matched descriptors from $\phi_{i=1\dots k_1}^{128}(R_t^{cc})$ is given by Equation (3). In our work, SIFT descriptors matching is based on dot products ($DP^{i=1\dots k_1}$) between unit vectors of descriptors (Equation (4)). Generic rules of the multilevel region descriptors matching is presented by Algorithm 1.

$$\begin{cases} R_Match(\phi_{k_1}^{128}(R_t^{cc}), \phi_{k_2}^{128}(R_{t-1}^c)) = 1 \\ \text{If any}(Des_Match > 0) \end{cases} \quad (2)$$

$$\begin{cases} Des_Match(i) = 1 \\ \text{if}(DP^{i=1\dots k_1}(1) < 0.6 * DP^{i=1\dots k_1}(2)) \end{cases} \quad (3)$$

$$DP^{i=1\dots k_1} = \text{sort}(\arccosine(\phi_{i=1\dots k_1}^{128}(R_t^{cc}) * \phi_{k_2}^{128}(R_{t-1}^c)^T)) \quad (4)$$

Algorithm 1: Multilevel region descriptors matching

Input: $f_{k_1}^{128}(R_t^{cc=1\dots m})$, $f_{k_2}^{128}(R_{t-1}^{c=1\dots n})$, Stopped($O_{j=1\dots h}$).f

Output: $R_{j=(t-1,t)}^{i=(c=1\dots n, cc=1\dots m)}$.State, $R_{j=(t-1,t)}^{i=(c=1\dots n, cc=1\dots m)}$.Cor,

Stopped($O_{j=1\dots h}$).f

If $R_Match(f_{k_1}^{128}(R_t^{cc=1\dots m}), f_{k_2}^{128}(R_{t-1}^{c=1\dots n})^T)$ then

 If $(f_{k_1}^{128}(R_t^{cc}), f_{k_2}^{128}(R_{t-1}^c)^T)$ then

R_t^{cc} .Cor = c

 Else If $(f_{k_1}^{128}(R_t^{cc}), f_{k_2}^{128}(R_{t-1}^{c_1, c_2 \dots c_q})^T)$ then

R_t^{cc} .Cor = $c_1, c_2 \dots c_q$

 End

Else

 If $R_Match(f_{k_1}^{128}(R_t^{cc=1\dots m}), \text{Stopped}(h).f)$ then

R_t^{cc} .Cor = Stopped(h).Cor

 Else

R_t^{cc} .Cor = *

 End

 If $R_Match(f_{k_2}^{128}(R_{t-1}^{c=1\dots n}), f_{k_3}^{128}(R_t^{c=1\dots m}))$ then

$\left\{ \begin{array}{l} \text{Stopped}(j+1).Cor = c \\ \text{Stopped}(j+1).f = f_{k_3}^{128}(R_t^c) \end{array} \right.$

 Else

R_{t-1}^c .Cor = *

 End

End

Three level matching levels are proposed: the first one is between $\phi_{k_1}^{128}(R_t^{cc=1\dots m})$ and $\phi_{k_2}^{128}(R_{t-1}^{c=1\dots n})$ to identify regions with state ‘Normal HS’ in case of $\phi_{k_1}^{128}(R_t^{cc})$ matches to only one $\phi_{k_2}^{128}(R_{t-1}^c)$ or prevent merging interaction (‘Merge’) in case of $\phi_{k_1}^{128}(R_t^{cc})$ matches to two or more $\phi_{k_2}^{128}(R_{t-1}^{c_1, c_2 \dots c_q})$. The second one is between $\phi_{k_1}^{128}(R_t^{cc=1\dots m})$ and

$Stopped(h).\phi$. $Stopped(h).\phi$ corresponds to region of stopped objects in previous frames, thus, if they match, $R_t^{cc=1\dots m}$ are in state 'Re-Moving', otherwise they are in state 'Entry'. The third matching is between $\phi_{k_2}^{128}(R_{t-1}^{c=1\dots n})$ and $\phi_{k_3}^{128}(R_t^{c=1\dots m})$ to identify stopped objects, otherwise means disappearance of $R_{t-1}^{c=1\dots n}$ ('Exit'). $\phi_{k_3}^{128}(R_t^{c=1\dots m})$ correspond to SIFT descriptors of $\beta_t^{2\dots 5}(R_{t-1}^{c=1\dots n})$ projection onto current frame.

3.2 Long- Term Processing (LTP)

We recall that the LTP rule feedbacks objects (O_i) in $TrackingObject\{O_{i=1\dots ObjectCount}$ and their corresponding regions in each frame to update tracked object attributes. Spatiotemporal attributes and descriptors of tracked object ($Z_t^{O_i} = (\beta^{1\dots 5}(O_i), Cor, \phi(O_i))$) are updated according to region/object association. The association between objects and their corresponding regions is based essentially on $R_{j=\{t-1,t\}}^{i=\{c=1\dots n, cc=1\dots m\}}.Cor$. In fact, attributes of objects in states 'Entry', 'Split' and 'Normal VE' are updated according to Equation (5). Objects in state 'Stopped' are controlled by $Stopped(O_{j=1\dots h}).\phi$ and objects in state 'Exit' ($TrackingObject\{O_i.Z_t^{O_i}(Cor)\} = c$ AND $R_{t-1}^c.Cor = *$) are killed.

$$\left\{ \begin{array}{l} \text{if } (TrackingObject\{O_i.Z_t^{O_i}(Cor)\} = R_t^{cc}.Cor) \text{ OR } (R_t^{cc}.Cor = *) \text{ Then} \\ TrackingObject\{O_i.Z_t^{O_i}(Cor)\} = cc \\ TrackingObject\{O_i.Z_t^{O_i}(\beta^{1\dots 5})\} = \beta^{1\dots 5}(R_t^{cc}) \\ TrackingObject\{O_i.Z_t^{O_i}(\phi)\} = \phi(R_t^{cc}) \end{array} \right. \quad (5)$$

Attributes of objects in merging region are hardly obtained since several objects shared the same region (cf. Fig. 4 (A)). To deal with this problem, we use template matching (cf. Fig. 4 (B)) based sum of squared difference to find 2D spatial attributes of each object (cf. Fig. 4 (C)), then, we compute their SIFT descriptors. Sum of squared difference is implemented using FFT (Fast Fourier Transform) based correlation.

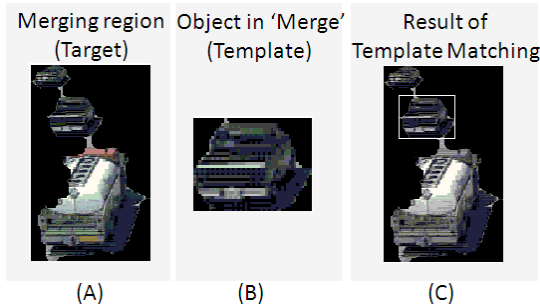


Fig. 4. Example of object localization in merging region. (A) Merging region (Target), (B) Object in 'Merge' and, (C) Result of Template Matching

4 Experimental Results

To evaluate our method, we used a corpus of 2 road traffic sequences¹ recorded in typical conditions (*HighwayII* and *HighwayIII*). *HighwayII* includes several interactions between road objects. *HighwayIII* include a dense traffic of road objects (different speed and size). The evaluation is made through the calculation of the rates of ‘Centroid Error’ [13] regard to Ground-Truth (GT) of the two sequences parts (4 parts for each one). ‘Centroid Error’ rates are computed according to two-pass matching scheme: first pass matching from system track to GT (*distanceSy*) to find false positive track (*FPT*) and second pass matching from GT to system track (*distanceTrack*) to find false negative track (*FNT*). In typical results, ‘Centroid Error’ rates from the two pass are the same. In addition to the above quantitative metric, we also consider in our evaluation a second metric ‘Two-pass many-to-many system to ground truth track matching’ [14] to measure how the system can deal with ‘Merge’ and ‘Split’ interactions. A GT/system track is matched to the system/GT track if there is both temporal overlap and spatial overlap. Temporal overlap is with respect to the duration of the system track. Spatial overlap is based on the centroid of the system lying inside the bounding box of the ground truth track. If multiple GT-matches, then this system track has ‘Merge Error’ equal to matched GT tracks. If multiple system-matches, then this GT track has ‘Split Error’ equal to matched system tracks.

As we can see in Fig.5, the four *HighwayII* parts (first line) echoed a low average *distanceSys/distanceTrack* rates peer part respectively between 0 and 6.163 pixels while *FPT* and *FNT* are between 0 and 6.19 percent. The four *HighwayIII* parts (second line) echoed a low average *distanceSys/distanceTrack* rates peer part respectively between 1.928 and 8.795 pixels while *FPT* and *FNT* are between 0 and 20,44%.

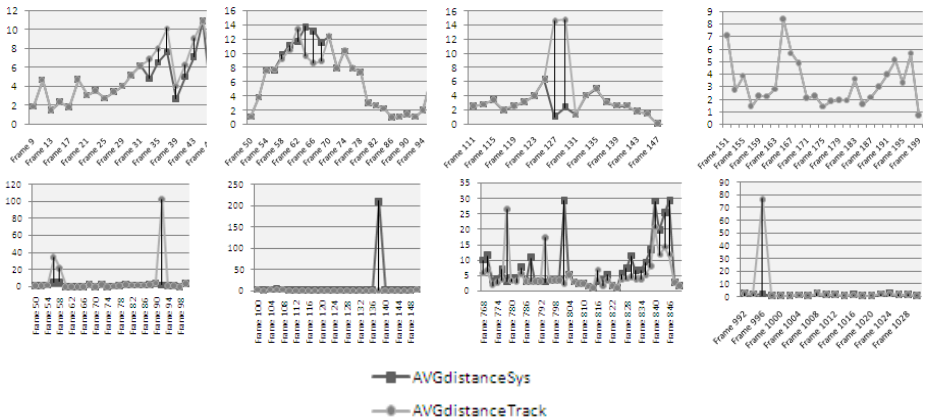


Fig. 5. Average distanceSys/distanceTrack curves of each frame of 4 part from *HighwayII* (first line) and *HighwayIII* (second line).

¹ Video sequences are courtesy of the Computer Vision and Robotics Research Laboratory at UCSD

We have performed the experimental study to know how our system can deal with 'Merge' and 'Split' interactions on 11 tracks from *HighwayII*. Temporal overlap and Spatial overlap curves for 4 of 11 tracks are depicted in Fig.6. For each track, both measures are computed firstly (A) from *GT-Track-Matching* and secondly (B) from *System-Track-Matching*. There is a 'Merge Error'/'Split Error' in case of multiple GT-matches/system-matches, more explicitly, if a curve from *GT-Track-Matching*/*System-Track-Matching* show more than peak with temporal overlap greater than 0.5 (cf. Track 4 for 'Merge Error'). Our system achieves a 'Merge Error rate' of 9.09 percent and a 'Split Error rate' of 0 percent.

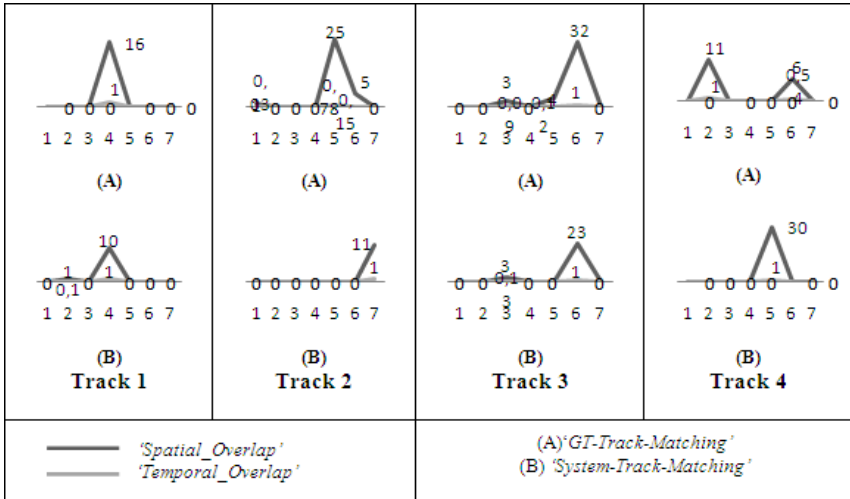


Fig. 6. Temporal overlap and Spatial overlap of 5 tracks from *HighwayII*

5 Conclusions

In this paper, we presented a novel method to track road objects. Our method is based on two prior processing: (1) Short-Term Processing (STP) that is based on spatial analysis and multilevel region descriptors matching. (2) Long-Term Processing (LTP) that is based on data association from STP. In these processing, both region and object information are used to establish objects correspondence over times.

The proposed algorithm was evaluated by a qualitative and quantitative experimental study on a corpus of road traffic sequences. The obtained results are rather satisfactory. In the near future, we plan to evaluate our method with computer vision applications like highway control and management system.

References

1. Peleshko, D., Ivanov, Y., Kustra, N., Kovalchuk, A.: An application of combined detector algorithm to extract the interest points of foreground objects in videostreams. In: 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, p. 262 (2011)

2. Dan, L., Jian-sheng, Q.: Sift-based object matching and tracking of coal mine. In: IET 3rd International Conference on Wireless, Mobile and Multimedia Networks, pp. 327–330 (2010)
3. Lin, X., Zhang, J., Liu, Z., Shen, J.: Semi-automatic road tracking by template matching and distance transform. In: Joint Urban Remote Sensing Event, pp. 1–7 (2009)
4. Cremers, D., Schnörr, C.: Statistical shape knowledge in variational motion segmentation. *Image and Vision Computing* 21(1), 77–86 (2003)
5. Rahman, M., Saha, A., Khanum, S.: Multi-object Tracking in Video Sequences Based on Background Subtraction and SIFT Feature Matching. In: Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 457–462 (2009)
6. Yan, Y., Wang, J., Li, C.: Object tracking using SIFT features in a particle filter. In: IEEE 3rd International Conference on Communication Software and Networks (ICCSN), pp. 384–388 (2011)
7. Liu, Y., Wang, X., Yang, J., Yao, L.: Multi-objects tracking and online identification based on SIFT. In: International Conference on Multimedia Technology (ICMT), pp. 429–432 (2011)
8. Cheng-bo, Y., Jing, Z., Yu-xuan, L., Ting, Y.: Object tracking in the complex environment based on SIFT. In: 3rd International Conference on Communication Software and Networks, pp. 150–153 (2011)
9. Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In: *Alvey Vision Conference*, vol. 15(Manchester), pp. 147–151 (1988)
10. Tomasi, C., Kanade, T.: Detection and Tracking of Point Features Technical Report CMU-CS-91-132, pp. 1–22 (1991)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
12. Hammami, M., Jarraya, S., Ben-Abdallah, H.: On line Background Modeling For Moving Object Segmentation in Dynamic Scenes. *Multimedia Tools and Applications Journal* (available first on-line) (2011)
13. Senior, A., Hampapur, A., Tian, Y.-L., Brown, L., Pankanti, S., Bolle, R.: Appearance Models for Occlusion Handling. In: *IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance* (2001)
14. Lisa, M.B., Andrew, W.S., Tian, Y.-L., Connell, J., Hampapur, A.: Performance Evaluation of Surveillance Systems Under Varying Conditions. In: *IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance* (2005)

Eigen Combination of Colour and Texture Informations for Image Segmentation

D. Attia, C. Meurie, and Y. Ruichek

Université de Technologie de Belfort-Montbéliard,
Institut Régional Supérieur du Travail Educatif et Social de Bourgogne,
Laboratoire Systèmes et Transports,
13 rue Ernest-Thierry Mieg, 90010 Belfort Cedex, France
{dhouha.attia,cyril.meurie,yassine.ruichek}@utbm.fr

Abstract. In this paper, we present a new combination of colour and texture informations for image segmentation. This technique is based on principal components analysis of a 3D points cloud, followed by an eigenvalues analysis. A set of colour gradients (morphological, Di-Zenzo) and texture gradients (Gabor, three Haralick attributes, Alternative Sequential Filter (ASF)) are used to test the proposed combination. The segmentation is performed using a hybrid gradient based watershed algorithm. The major contribution of this work consists in combining locally colour and texture informations using an adaptive and non parametric approach. The proposed method is tested on 100 images from the Berkley dataset [1] and evaluated with the Mean Square Error (MSE), the Variation of Information (VI) and the Probabilistic Rand Index (PRI).

1 Introduction

One of the most challenging problems in computer vision concerns image segmentation. If we consider colour and texture informations, the segmentation methods can be divided into three classes. The first one regroups those using only texture information [2,3,4], while the second class contains segmentation methods which use only colour information [5,6,7,8]. In spite of satisfactory results of these approaches, the combination of colour and texture informations leads generally to obtain best results and increases the robustness of segmentation [9,10,11]. In this paper, we propose a non parametric method to combine colour and texture attributes. This combination allows defining a structural gradient that will be used in watershed algorithm. This approach is based on principal components analysis (PCA) of a 3D points cloud formed by colour and texture descriptors, followed by an eigenvalues analysis. This paper is organized as follows : in section 2, the segmentation step by watershed algorithm based on a structural gradient is presented. Then, colour and texture combination approach is detailed. In section 3, we describe extensive experiments carried out on 100 images from the BSD8 dataset to test and validate the proposed method.

2 Image Segmentation Using Colour and Texture Informations Collectively

In this section, a short introduction of watershed algorithm based on colour and texture gradients is presented. Then, an existing fixed combination based approach is briefly described. Finally the proposed approach based on PCA and eigenvalues analysis is detailed.

2.1 Watershed Algorithm Based on Colour-Texture Gradients

Watershed algorithm segments image into watershed regions [12,13]. Considering the input image (gradient image) as a topographic surface, each seed of the input region (calculated with an optimal density) can be viewed as the point to which water falling on the surrounding region drains. The boundaries of watersheds lie on tops of ridges. In this paper, a new combination of colour and texture informations is proposed, in order to provide a hybrid gradient which will be the input of watershed algorithm. Before detailing the combination, different colour and texture gradients that will be used for tests are introduced.

Texture is an important perceptual information that is generally extracted using mathematical tools. In literature, four main classes of texture descriptors are discriminated: geometric attributes, descriptors based on spatial texture models, spatio-frequency and statistic attributes. In the present paper, we choose to use the following descriptors: Gabor transformation (spatio-frequency attribute) [14,15], three Haralick parameters (second order statistic attributes) [16,17] and an Alternate Sequential Filter (a transformation from mathematical morphology to calculate a texture gradient) [18]. Considering colour information, we use Di-Zenzo gradient based on the first derivative of the initial image (see [19] for more details) and morphological gradient which corresponds to the subtraction between dilatation and erosion of the initial image (using a lexicographic order).

In literature, one can find an interesting approach combining colour and texture informations [20]. The combination process is based on a set of operations derived from mathematical morphology. This technique uses collectively colour and texture informations to generate a structural gradient used in image segmentation. As expressed in equation 1, this gradient is obtained by a fixed combination of colour and texture gradients. In equation 1, Q_{col} and Q_{tex} represent respectively colour and textural gradients. The textural gradient is obtained using following steps: filtering, definition of texture layer, and granulometric analysis. α represents the combination parameter, which is chosen between 0 and 1. Even if this approach gives satisfactory results, its major drawback concerns the choice of the optimal value of the parameter α for each image. Furthermore, α is global for the entire image, and thus the combination does not take into account pertinent local information.

$$Q_{struc}(I) = (1 - \alpha)Q_{tex}(I) + \alpha Q_{col}(I) \quad (1)$$

Recent work in combining colour and texture descriptors were presented in [21]. In this work, the authors were mostly concerned with detecting contours then image segmentation. Their contour detection is based on combining multiple local cues into a globalization framework based on spectral clustering. The contour detection results are injected into the segmentation step. The segmentation algorithm consists of a generic approach for transforming any contour detector into a hierarchical region tree. The authors use four features in their approach : after transforming the input image into CIE Lab colorspace, they extract brightness, colour a and colour b channels. The fourth feature channel is a texture channel, which assigns to each pixel a texon id . Then, for each feature channel, an oriented gradient $G(x, y, \theta)$ is computed by placing at location (x, y) a circular disc split into two half-discs by a diameter at angle θ . The combination step of these local cues is given by equation [2]:

$$mPb(x, y, \theta) = \sum_s \sum_i \alpha_{i,s} G_{i,s,\sigma(i,s)}(x, y, \theta) \quad (2)$$

where mPb is the multiscale predicted boundary detector, s refers to scales, i indexes feature channels (brightness, colour a , colour b , texture), and $G_{i,s,\sigma(i,s)}(x, y, \theta)$ measures the histogram difference in channel i between two halves of a disc of radius $\sigma(i, s)$ centred at (x, y) and divided by a diameter at angle θ . The parameters $\alpha_{i,s}$ weight the relative contribution of each gradient signal. Contrary to this linear combination of local cues, the optimal mixing function could be non linear. Thus, in [22], each cue was treated as an expert for a certain class of boundary and a set of classifiers (such as Classification Trees and SVM) were used to combine the various cues. Even if this approach gives satisfactory results, its major difficulty consists of the dependency of the segmentation step to the boundaries detection. A second limit concerns the choice of the optimal mixture function of the various local cues.

2.2 PCA Based Colour-Texture Combination : Eigen Combination

We propose a novel combination based on Principal Components Analysis (PCA) of a 3D points cloud, followed by an eigenvalues analysis. Principal Components Analysis is a technique which uses geometric and graphic representations to describe the dispersion of a dataset (observations). This dataset is assimilated to a points cloud P composed by m quantitative variables having n unities (called also subjects):

$$P^s = \{(G_{i1}^s, \dots, G_{ij}^s, \dots, G_{im}^s), \forall i \in 1 \dots n\} \quad (3)$$

where index i corresponds to subject i and index j corresponds to variable j such as :

$$p_{.j} = (p_{1j} \cdots p_{nj}) \quad (4)$$

By representing the subjects, we can determine which ones are similar. On the other hand if we represent the variables, we can study structures of linear links

within the 3D points cloud, and then determine correlated variables [23,24]. Based on linear algebra, PCA technique aims also to extract axes which conserve the maximum of information [24]. The first step consists in calculating the covariance matrix of the points cloud. Then, eigenvalues and associated eigenvectors of the covariance matrix are extracted. Eigenvalues give the variation of the points cloud along principal components which are obtained by eigenvectors. Principal Components Analysis is generally used to eliminate the correlation of initial data and to reduce their size. In computer vision, PCA is used for image classification [25], image compression [26,27] and objects recognition [28]. In the present paper, PCA technique permits to determine axis which conserve the maximum of information (pertinent colour and texture informations). Thus, the proposed approach generates a novel structural hybrid gradient which will be used as an input of watershed algorithm.

Let I be the initial image, s a pixel in the image I and P^s is the associated 3D points cloud generated for the pixel s . P^s contains the colour and texture values of n gradients (in our case, n represents the number of the used colour and texture descriptors). Each gradient $(G^s_i)_{i=1,\dots,n}$ is calculated in a colour space $E_1E_2E_3$. $m = 3$, is the number of the variables, and $\forall i \in \{1, \dots, n\} G^s_i \in \mathbb{R}^3$. Therefore, the local 3D points cloud can be defined as below :

$$P^s = \{(G^s_{i1}, G^s_{i2}, G^s_{i3}), \forall i \in 1 \dots n\} \tag{5}$$

$$P^s = (G^s_1, G^s_2, \dots, G^s_n.) \tag{6}$$

The covariance matrix of the 3D points cloud P^s is generated using the following equation :

$$C^s = (cov(G^s_j, G^s_{j'}))_{\substack{j=1,\dots,3 \\ j'=1,\dots,3}} \tag{7}$$

where :

$$cov(G^s_j, G^s_{j'}) = \frac{1}{n} \sum_{i=1}^n (G^s_{ij} - \bar{G}^s_j)(G^s_{ij'} - \bar{G}^s_{j'}) \tag{8}$$

$$\bar{G}^s_j = \frac{1}{n} \sum_{i=1}^n G^s_{ij} \tag{9}$$

According to linear algebra rules, there are two local matrices L^s and D^s such as :

$$(L^s)^{-1}C^sL^s = D^s \tag{10}$$

Let λ_1^s, λ_2^s and λ_3^s be the eigenvalues of the local covariance matrix C^s such as $\lambda_3^s \leq \lambda_2^s \leq \lambda_1^s$. Let V_1^s, V_2^s, V_3^s be the associated eigenvectors. Therefore, the matrices D^s and L^s are expressed by the following equations :

$$D^s = \begin{pmatrix} \lambda_1^s & 0 & 0 \\ 0 & \lambda_2^s & 0 \\ 0 & 0 & \lambda_3^s \end{pmatrix} \tag{11}$$

$$L^s = (V_1^s V_2^s V_3^s) \quad (12)$$

In order to determine the principal components that maximize local information, the eigenvalues are compared according to three cases listed as follows :

Case 1 $\lambda_1^s \gg \lambda_2^s$: One can conclude that it exists only one axis (direction of V_1^s) around which the 3D points cloud is concentrated. In this case, only the third principal component maximizing the local information is chosen. Thus, the new value $\hat{I}(s)$ of the pixel s is given by the following mathematical formulation:

$$\hat{I}(s) = \operatorname{argmax}_{i \in \{1, \dots, n\}} \{G_{i1}\}$$

Case 2 $\lambda_3^s \ll \lambda_2^s$: In this case, λ_1^s and λ_2^s have the same order of the magnitude and the 3D points cloud is assimilated to a plane formed by the eigenvectors V_1^s and V_2^s (respectively associated to the eigenvalues λ_1^s and λ_2^s). The eigenvector V_3^s constitutes the normal of the plane ($\widehat{V_1^s, V_2^s}$). Therefore, there are two principal components conserving the maximum of local information. In this case, the new value $\hat{I}(s)$ of the pixel s is given by the following equation:

$$\hat{I}(s) = \operatorname{argmax}_{i \in \{1, \dots, n\}} \left\{ \frac{1}{2} (G_{i1}^s + G_{i2}^s) \right\}$$

Case 3 $\lambda_1^s \simeq \lambda_2^s \simeq \lambda_3^s$: In this case, the local 3D points cloud is dispersed according to all directions. Thus, no information is privileged. All principal components are fairly considered, and the new value $\hat{I}(s)$ of the pixel s is given by the following mathematical formulation:































$$\hat{I}(s) = \operatorname{argmax}_{i \in \{1, \dots, n\}} \left\{ \frac{1}{3} (G_{i1}^s + G_{i2}^s + G_{i3}^s) \right\}$$

3 Experiments and Discussion

In this section, an evaluation of the proposed colour and texture combination is presented. The results are compared to those obtained by the fixed combination [20], described before. Extensive experiments are carried out on 100 images from the BSD3 dataset [1] using Mean Square Error (MSE) and two other evaluation metrics used in [21]: the variation of information (VI) [29] and the Probabilistic Rand Index (PRI) [30]. In order to conclude on the effectiveness and the robustness of the proposed approach, tests including two colour gradients (Di-Zenzo and morphological) and five texture gradients (ASF, Gabor filter, Second Angular Moment (SAM) attribute, Coherence attribute and Variance attribute) are realized. Table 1 presents the evaluation by MSE criteria of segmentation results obtained with both the fixed and the proposed combinations on ten images of the database.

For a better visualization, we present only the obtained results with a texture gradient calculated with the Variance attribute. This example shows the difficulty to choose the best value of the parameter α when a fixed combination is

Table 1. Segmentation results of the fixed and proposed approaches on 10 images of the Berkley database (α^* corresponds to optimal value of α parameter)

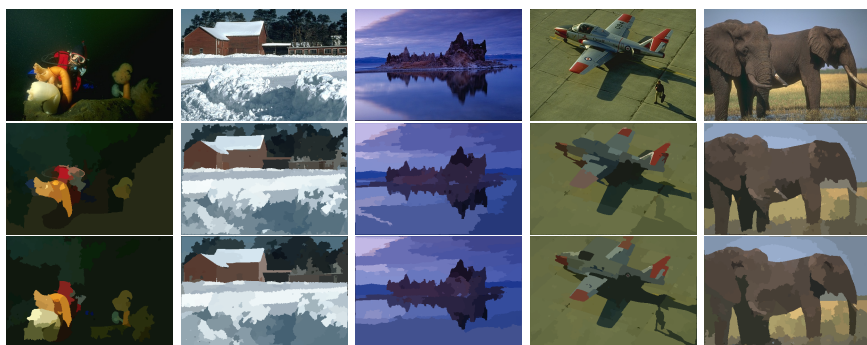
Initial Image	Fixed combination						Proposed approach	
	$\alpha = 0$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	α^*		
	189.7	198.6	199.4	204.4	190.5		179.9	
	242.9	220.8	223.2	220.9	264.5		203.5	
	85.4	86.3	76.7	90.1	89.9		82.5	
	103.8	95.9	82.9	81.6	92.5		90.3	
	106	195.6	141.2	145.3	721.3		126.5	
	94.7	93.4	110.5	101.4	106.6		91.9	
	169.3	159.8	114.4	127.1	116.4		140.2	
	244.1	138.3	131.6	117.7	135.1		112.8	
	123.2	115.4	112.6	79.1	78.7		84.7	
	75.2	68.1	65.8	66.6	73.4		41.9	

applied. Indeed, the value of this parameter is not the same for all images and must be redefined for each image to obtain an acceptable segmentation. Considering different values of the parameter α , one can notice that the proposed approach obtains the second position or the third one when it is not in the first position, when MSE of the two methods are compared. Moreover, when the fixed combination obtains the first position, the gap with the proposed method is very low. For more details, table 2 presents the evaluation with the mean of MSE, the mean of PRI and the mean of VI on all images of the database for the proposed approach according to the used colour and texture gradients. One can conclude that the segmentation results are similar even if the best one is obtained with the texture variance gradient and Di-Zenzo colour gradient if we consider MSE criteria. If we consider measures of PRI metric, best results are obtained with the texture coherence gradient and the morphological colour gradient. Finally, considering VI metric, best results are obtained by texture variance gradient and morphological colour gradient.

In figure 1, segmentation results are illustrated for the two combinations (the fixed and proposed approaches) using the best colour and texture gradients (Di-Zenzo colour gradient and texture variance gradient considering the MSE metric of evaluation). Even if the results of the two combinations are similar for the second and third images, one can notice that the proposed method provides better results for the other images. For example, the segmentation of different

Table 2. Segmentation results of the proposed approach on 100 images of the Berkley database

Texture gradient	Colour gradient					
	Morphological gradient			Di-Zenzo gradient		
	MSE	PRI	VI	MSE	PRI	VI
Gabor	388,4	0.69	1.33	340,2	0.69	1.64
ASF	394,8	0.69	1.31	334,5	0.70	1.66
SAM	394,8	0.69	1.65	344,8	0.69	1.84
Coherence	390,1	0.71	1.29	334,5	0.70	1.59
Variance	381	0.70	1.25	324,7	0.70	1.57

**Fig. 1.** Segmented images with the combination using Di-Zenzo colour gradient and the texture variance gradient (from top to bottom): initial images, segmented images with the fixed combination, segmented images with the proposed approach)

parts of algae (figure 1-1) obtained with the proposed approach is better than the segmentation obtained with the fixed combination. One can note the same remark for the man and the aircraft (figure 1-4), and the proboscis of the elephant (figure 1-5).

4 Conclusion

In this paper, we presented a novel segmentation method combining colour and texture informations. The proposed method is based on an eigenvalues analysis and principal components analysis of a 3D points cloud formed by colour and texture attributes. The contribution of this technique is the definition of an adaptive combination of colour and texture gradients. The proposed combination method provides good segmentation results by watershed algorithm. This technique is local since we assign to each pixel the maximum of information provided by combination of the colour and the texture. Furthermore, the proposed combination is non parametric since it does not require any parameter. In future work, we will expand the number of colour and texture gradients and will show the influence of colour spaces.

Acknowledgements. This research work is funded in the framework of the ViLoc project and supported by the Regional Council of Franche-Comté (France).

References

1. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision (2001)
2. Lo, E.H.S., Pickering, M.R., Frater, M.R., Arnold, J.F.: Image segmentation using invariant texture features from the double dyadic dual-tree complex wavelet transform. In: IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. I-609–I-612 (2007)
3. Paulhac, L., Makris, P., Gregoire, J.-M., Ramel, J.-Y.: Descripteurs de textures pour la segmentation d'images échographiques 3d. Congrès des Jeunes Chercheurs en Vision par Ordinateur, ORASIS 2009 (2009)
4. Liu, B., Xian: A texture feature fusion-based segmentation method of sar images. In: Second International Conference on Intelligent Human-Machine Systems and Cybernetics (2010)
5. Meyer, F.: Color image segmentation. In: International Conference on Image Processing and its Applications, pp. 303–306 (1992)
6. Kurugollu, F., Sankur, B., Harmanci, A.: Color image segmentation using histogram multithresholding and fusion. *Image Vision Computing* 19, 915–928 (2001)
7. Delon, J., Desolneux, A., Lisani, J.L., Petro, A.B.: Color image segmentation using acceptable histogram segmentation. In: Pattern Recognition and Image Analysis, pp. 239–246 (2006)
8. Kiranyaz, S., Birinci, M., Gabbouj, M.: Perceptual color descriptor based on spatial distribution: a top-down approach. *Image and Vision Computing* 28, 1309–1326 (2010)
9. Dubuisson-Jolly, M.-P., Gupta, A.: Color and texture fusion: application to aerial image segmentation and gis updating. *Image and Vision Computer* 18, 823–832 (2000)
10. Chen, J., Pappas, N., Mojsilovic, A., Rogowitz, B.E.: Adaptive perceptual color-texture image segmentation. *IEEE Transactions on Image Processing* 14, 1–13 (2005)
11. Ilea, D.E., Whelan, P.F.: Image segmentation based on the integration of colour-texture descriptors. *Pattern Recognition* 44, 2479–2501 (2011)
12. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersions simulations. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)* 13(16), 583–598 (1991)
13. Shafarenko, L., Petrou, M., Kittler, J.: Automatic watershed segmentation of randomly textured color images. *IEEE Transactions On Image Processing* 6(11), 1530–1543 (1997)
14. Palm, C., Keyser, D., Lehmann, T., Spitzer, K.: Gabor filtering of complex hue saturation images for color texture classification. In: Proceedings of the 5th Joint Conference on Information Science, pp. 45–49 (2000)
15. Palm, C., Lehmann, T.: Classification of color textures by gabor filtering. *Machine Graphics & Vision International Journal* 11, 195–219 (2002)
16. Haralick, R., Shanmugan, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3(6), 610–621 (1973)

17. Majdoulayne, H.: Extraction de caractéristiques de texture pour la classification d'images satellites. Ph.D. dissertation, Université de Toulouse (2009)
18. Soille, P.: Morphological image analysis, 2nd edn. Springer, Heidelberg (1999)
19. Zeno, S.D.: A note on the gradient of multi-image. *Computer Vision, Graphics, and Image Processing* 33, 116–125 (1986)
20. Angulo, J.: Morphological texture gradients. Application to colour+texture watershed segmentation. In: Proc. of the 8th International Symposium on Mathematical Morphology, pp. 399–410 (October 2007)
21. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. University of California at Berkeley, Tech. Rep., February 16 (2010), <http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-17.html>
22. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions On Pattern Analysis and Machine Intelligence* 26, 530–549 (2004)
23. Duby, C., Robin, S.: Analyse en composantes principales. Institut National Agronomique Paris - Grignon, Tech. Rep., département O.M.I.P, Institut National Agronomique Paris - Grignon (2006)
24. Gergaud, J.: Unité fondamentale: Algèbre linéaire: une application l'analyse en composantes principales. INP ENSAT, Tech. Rep. (2006)
25. Zeng, W., Zhang, Y.: A novel improvement to pca for image classification. In: International Conference on Computer Science and Service System (CSSS), pp. 1964–1967 (2011)
26. Du, Q., Zhu, W., Yang, H., Fowler, J.E.: Segmented principal component analysis for parallel compression of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters* 6, 713–717 (2009)
27. Ho, P.-M., Wong, T.-T., Leung, C.-S.: Compressing the illumination-adjustable images with principal component analysis. *IEEE Transaction on Circuits and Systems for Video Technology* 15, 355–364 (2005)
28. Ilin, A., Raiko, T.: Practical approaches to principal component analysis in the presence of missing values. *Journal of Machine Learning Research* 11 (2010)
29. Meila, M.: Comparing clusterings: An axiomatic view. In: ICML (2005)
30. Unnikrishnan, R., Pantofru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. PAMI (2007)

A Graph Based Approach for Heterogeneous Document Segmentation

Fattah Zirari^{1,2}, Driss Mammass¹, Abdellatif Ennaji², and Stephane Nicolas²

¹Laboratory IRF-SIC Agadir Maroc

²Laboratory LITIS Rouen France

{zirari_fattah, driss_mammass}@yahoo.fr,
{Abdel.Ennaji, stephane.nicolas}@univ-rouen.fr

Abstract. In the field of document image processing, the text/graphic separation is a major step that conditions the performance of the recognition and indexing systems. That involves identifying and separating the graphical and textual components of a document image. In this context, it is important to implement approaches that effectively address these problems. This paper presents a method for separating textual and non textual components in document images using a graph-based modeling and structural analysis. This is a fast and efficient method to separate adequately the graphical and the textual areas of a document. Some examples obtained on technical documents and magazines issued from the databases approved by the community make it possible to validate the approach.

Keywords: Segmentation, text/no text Separation, Document Image, Graph, modelization, structural analysis.

1 Introduction

Segmentation is a crucial basic step in a document image processing and analysis workflow, because in fact it precedes any other operation of identification or classification. This step depends on the type of image that differs from both the acquisition system and the image formation process. In the case of document images, this consists in locating and possibly identifying the elements constituting the document at different granularity levels. Thus, a first segmentation task may consist in locating and identifying the text areas and the areas of different natures. If we consider for example the document images presented in Figure 1, which are pages from technical documents and magazines, the first segmentation task may consist in differentiating the areas containing text from the areas representing tables, curves or graphs.

In the literature, three families of approaches are possible for document image segmentation: the bottom-up, the top-down and hybrid approaches.

In top-down techniques, document images are recursively divided to smaller regions. These techniques are often fast, but the efficiency depends on a prior knowledge about the class of documents to be processed. Among the developments

produced in early times, the most well known methods are projection methods [1] and space transforms [2] (Fourier transform, Hough transform, ect).

Though these top-down methods generally perform well, they have a major drawback which is the need to have a prior knowledge about the document class and layout (number of columns, width of margins, etc) for them to be effective.

Bottom-up methods start with the thinnest elements (pixels), merging them recursively in connected components or regions, and then in larger structures. They are more flexible but may suffer from accumulation of errors. They make use of methods like connected components analysis [3], [4], region growing methods, run-length smoothing (RLSA) [5], neural networks [6] and active contours [7].

The advantage of bottom-up methods is that they are very flexible. Another thing is that these methods make use of a lot of parameters that need to be adjusted precisely for good results.

Many other methods that do not fit into either of these categories are therefore called hybrid methods that combine and make use of both bottom-up and top-down approaches. For example, connected components analysis for shape information and block separation for background block map have been used in [8] in a hybrid segmentation approach. Classification of these blocks is achieved according to the scenarios defined by the user.

As part of this paper, we propose a segmentation method based on a modeling of the image document by graphs and applying structural layout rules. The main advantage of using graphs to represent images is the integration of spatial information in the model. Indeed classical representations provide no information on how the regions of interest of the image are organized. On the contrary, the representation by the graphs makes to describe the structure of image as the way in which the areas are laid out the one compared to the others. Our method is insensitive to low skew and adapted to the text / non-text segmentation.

This paper is organized as follows. In Section 2, we first describe the steps used by our approach. Some experimental results are given and discussed in Section 3 and the last section concludes the paper.



Fig. 1. Examples of structured documents containing textual and graphical information

2 Proposed Method

The approach we propose consists in modeling the document image by a graph that will enable us to establish the neighborly relations and connexity according to a homogeneity criterion based on the pixels intensity. This will be a first step to extract the connected components of the document image using a graph modelling approach. In a second step, the connected components thus formed will be categorized into graphical regions and text areas by applying structural layout rules. First introduce the graph formalism that we have adopted in our system.

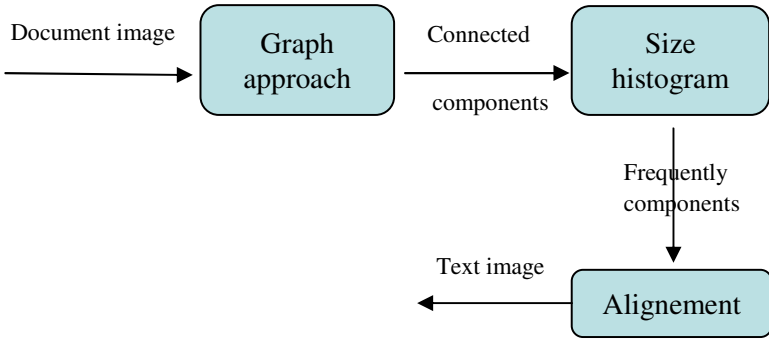


Fig. 2. Proposed Method

2.1 Used Formalism

The graphs constitute a mode of representation very frequently used in image processing and pattern recognition. They indeed make it possible to describe naturally in a unified formalism some objects and the relations between these objects.

A graph G consists of a set of nodes, denoted by V, linked by a set of edges, denoted by E:

$$\begin{aligned}
 G &= (V, E) \\
 V &= \{v_1, v_2, \dots, v_n\} \\
 E &= \{(v_i, v_j) | v_i \in E, v_j \in E\}
 \end{aligned}$$

Finally let us give for memory some definitions that we will need later.

A connected graph is a graph where for any two nodes i and j we can find a walk which begins at i and ends at j.

An undirected graph is one in which the edges have no orientation. The edge (i, j) is identical to the edge (j, i), i.e., they are not ordered pairs.

A tree is a connected graph without cycles which connects a subset of all nodes. A spanning tree is a tree which connects all nodes.

A minimum spanning tree (MST) of an edge-weighted graph is a spanning tree whose weight (the sum of the weights of its edges) is not larger than the weight of any other spanning tree.

All these concepts are illustrated on the example of a simple graph modeling the image given in Figure 3.

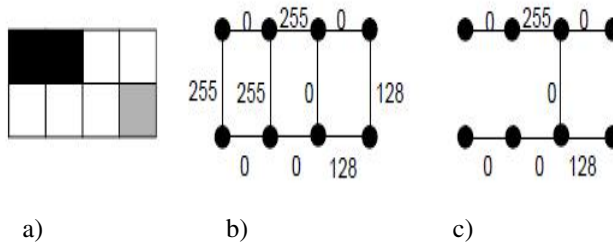


Fig. 3. a) initial image; b) associated graph ; c) Minimum spanning tree

In our approach we model the image by undirected related graph. The nodes of the graph represent the pixels of the image and will be balanced by their intensity, whereas the edges represent the relationships of connexity and are balanced by the sum of the intensities of the pixels at the ends.

We seek to combine the pixels to form homogeneous regions and then label them. To measure the homogeneity of intensity we adopt the concept of internal difference of a component defined in [9] as follows:

The internal difference of a component $C \subseteq V$ is the largest weight in the Minimum Spanning Tree of the component, (MST). That is,

$$Int(C) = \max_{e \in MST(C, E)} p(e) \tag{1}$$

$$\text{with } P(e) = (I(P_i) + I(P_j)) / 2, \quad e = (v_i, v_j) \in E \tag{2}$$

and $I(P_i)$ pixel intensity.

The motivating argument is that since the MST spans a region C through a set of edges of minimal cost, any other connected set of same cardinality will have at least one edge with weight superior to $Int(C)$.

Initially, a graph is constructed over the entire image, with each pixel p being its own unique region {p}. Subsequently, regions are merged by traversing the edges in a sorted order by increasing weight and evaluating whether the edge weight is smaller than the internal variation of both regions incident to the edge. If true, the regions are merged and the internal variation of the compound region is updated.

Now that we have defined the formalism we will explain the algorithm to find a partitioning of the image in homogeneous regions related. This algorithm is:

Algorithm:

The input is a graph $G = (V, E)$, with n vertices and m edges. This graph is formed according to the rules of 8-connected neighborhood classically used to model images. The output is a segmentation of V into components $S = (C_1; \dots; C_r)$. The proposed algorithm is iterative:

- 1- Sort E into $\Pi = (o_1, \dots, o_m)$, with $o_q = (v_i; v_j)$, by non-decreasing edge weight.
- 2- Start with a segmentation S^0 , where each vertex v_i corresponds to exactly one unique component.
- 3- Construct S^q given S^{q-1} as follows:
 Let v_i and v_j denote the vertices connected by the q^{th} edge in the order list Π , i.e., $o_q = (v_i, v_j)$. If v_i and v_j are in disjoint components of S^{q-1} and $p(o_q)$ is small compared to the internal mean of both components, then merge the two components otherwise do nothing. More formally, let C_i^{q-1} be the component of S^{q-1} containing v_i and C_j^{q-1} the component containing v_j . If $C_i^{q-1} \neq C_j^{q-1}$ and $p(o_q) \leq \text{MINT}(C_i^{q-1}, C_j^{q-1})$ with $\text{MINT}(C_i^{q-1}, C_j^{q-1}) = \min(\text{Int}(C_i^{q-1}), \text{Int}(C_j^{q-1}))$, then S^q is obtained from S^{q-1} by merging C_i^{q-1} and C_j^{q-1} . Otherwise $S^q = S^{q-1}$.
- 4- Repeat the step 3 for $q = 1, \dots, m$.
- 5- Return $S = S^m$.

At the end of the algorithm we obtain a set of homogeneous regions (Figure 4) we have to label as graphical or textual elements. We describe this labeling process in the next section.

2.2 Labelling of the Segmented Components

The aim is to label the components resulting from the segmentation obtained at the previous step, in two classes: "text" (or textual components) or "non-text" class (graphics, tables, lines ...). For that, we sought to exploit the fact that the text zones are often characterized by an alignment of characters of very similar size. Thus, we developed a simple approach to identify text areas based on the filtering of the components provided by our first stage, based on a size criterion and then the overlapping between components is analyzed

Thus, to detect the textual components we apply the following two steps:

- A first step consists in calculating the histogram of the frequencies of the components size. Only the components belonging to the most significant peaks of this histogram are retained (figure 5). For that we use a detection threshold set empirically up to now. This threshold can also be determined by a machine learning procedure. The idea of this first step is to filter the majority of the non textual components.

- The second phase consists in eliminating the frequent noise components and graphics that were not filtered at the preceding step. For that, we use the notion of alignment components between them. This alignment is determined by the vertical overlap between the components according to a given threshold, allowing a certain inclination.

This first preliminary approach for the identification of text zones showed good performance despite its simplicity. It should nevertheless be completed in order to answer some weaknesses as shown in the results section.

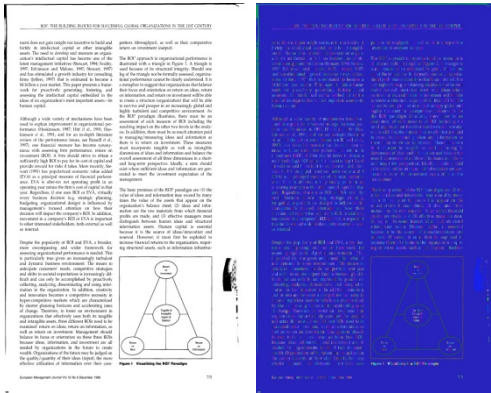


Fig. 4. original document ; segmented document

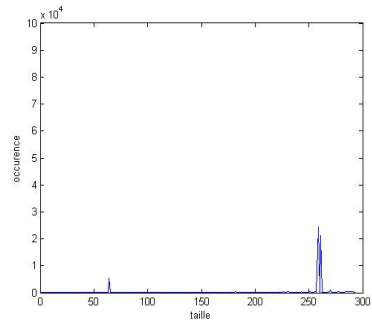


Fig. 5. Frequency histogram of the components size of the document result in Figure 3

3 Experiment and Results

To validate the effectiveness of our approach, we conducted tests on two databases of technical documents and magazines. These are documents that contain text, tables, charts, graphics, or inserts. These documents issued from the databases Prima [10] and de Washington University III (UWIII) [11]. These two databases are available with ground truth information. Thus, to evaluate our approach, we conducted a comparison between the pixel to pixel documents results and documents ground truth.

In the detection of textual components, we obtained a detection percentage of 96,7%. This rate of correct detection increases to 98,5% if one takes into account the textual components that are integrated with graphic blocks (Figure 6). Indeed, the ground truth provided for database for Prima does not consider this textual information as such but combines with graphic blocks in which they are understood. Similarly, we obtained a percentage of detection of 97% for non-textual components.

The error rate of 3% is due to the presence of textures in some graphic components that have similarities with textual components (Figure 7).

The figures 8 and 9 illustrate the results obtained by our method on a sample of 2 documents chosen in the base of the documents treated so as to illustrate the

capacities and the limits of our approach. These figures show in the order the original document image, the image truth ground, and the 2 images corresponding to the textual zones and the graphic zones identified by our approach. In the truth ground the graphic zones are represented in red and the textual zones in blue.

These results illustrate the good performance of our approach.

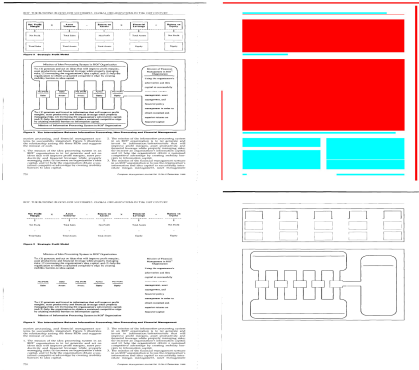


Fig. 6. Example of graphic elements with text included in the diagrams well detected by our method

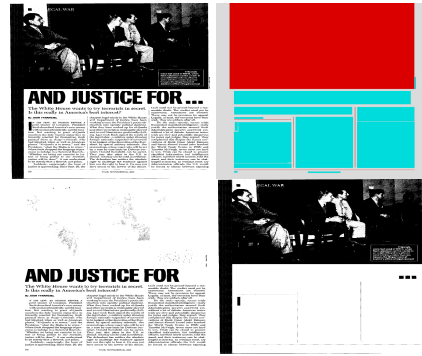


Fig. 7. Example of detection error of the graphic parts produced by our method

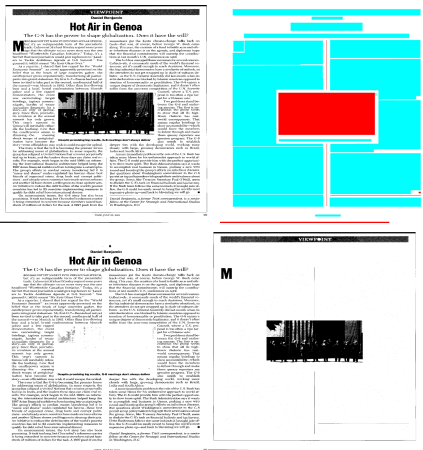


Fig. 8. Example of good results obtained by our method

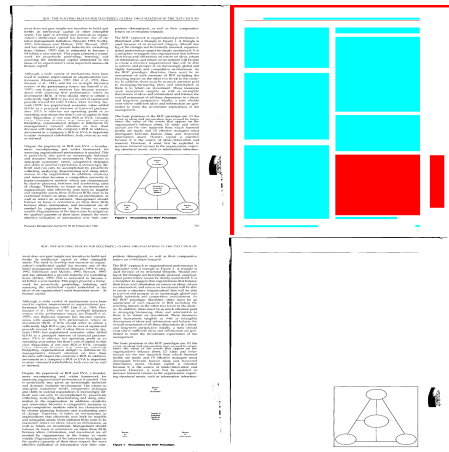


Fig. 9. Example of good results obtained by our method

4 Conclusion

We have presented a method for document image segmentation to identify the textual and the non textual zones being able to be either graphics or any other type of illustrations. This method is based on modeling of the various blocks of the document image by a graph approach. The blocks resulting from this step of modeling are then classified by a simple method which exploits the concept of alignment of the forms. Extensions of this approach for segmenting text blocks into words are underway for the development of a system for document indexing by content. Additional validations on more complex documents and/or degraded are also envisaged. The exploitation of the whole of this information is considered thereafter for the treatment of the non textual zones.

Acknowledgment. We would like to acknowledge the financial support of our project by the “action intégrée Maroc-française” n° MA/10/233 and the AIDA project, program Euro Mediterranean 3+3 : n° M/09/05.

References

1. Antonacopoulos, A., Karatzas, D.: Semantics based content extraction in typewritten historical documents. In: 8th International Conference on Document Analysis and Recognition, pp. 48–53 (2005)
2. Jain, A.K.: Fundamentals of digital image processing. Prentice Hall (1989)
3. Mitchell, P.E., Yan, H.: Newspaper document analysis featuring connected line segmentation. In: Sixth International Conference on Document Analysis and Recognition, pp. 1181–1185 (2001)
4. Faure, C., Vincent, N.: Simultaneous detection of vertical and horizontal text lines based on perceptual organization. In: 16th Document Recognition and Retrieval Conference, DRR 2009, USA (2009)
5. Wong, K.Y., Casey, R.G., Wahi, F.M.: Document analysis system. IBM Journal of Research Development 26, 647–656 (1982)
6. Caponetti, L., Castiello, C., Gorecki, P.: Document page segmentation using neurofuzzy approach. Applied Soft Computing (2007) (in press, corrected proof)
7. Bukhari, S.S., Shafait, F., Breuel, T.M.: Segmentation of curled textlines using active contours. In: The Eighth IAPR Workshop on Document Analysis Systems (2008)
8. Ramel, J., Leriche, S.: Segmentation et analyse interactive de documents anciens imprimés. In: Traitement du Signal (TS), pp. 209–222 (2005)
9. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient Graph-Based Image Segmentation. International Journal of Computer Vision 59(2), 167–181 (2004)
10. Antonacopoulos, A., Bridson, D., Papadopoulos, C., Pletschacher, S.: Performance Analysis Framework for Layout Analysis Methods. In: Proceedings of The 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Catalonia, Spain, pp. 296–300 (September 2009)
11. Guyon, I., Haralick, R.M., Hull, J.J., Phillips, I.T.: Data sets for OCR and document image understanding research. In: Bunke, H., Wang, P. (eds.) Handbook of Character Recognition and Document Image Analysis, pp. 779–799. World Scientific, Singapore (1997)

Rotation Invariant Fuzzy Shape Contexts Based on Eigenshapes and Fourier Transforms for Efficient Radiological Image Retrieval

Alaidine Ben Ayed¹, Mustapha Kardouchi¹, and Sid-Ahmed Selouani²

¹ Université de Moncton, Campus de Moncton, 18 avenue Antonine-Maillet, Moncton, NB, Canada E1A 3E9

alaidine.ben.ayed@umoncton.ca

² Université de Moncton, Campus de Shippagan, 218 boulevard J-D Gauthier, Shippagan, NB, Canada E8S 1P6

Abstract. This paper proposes a new descriptor for radiological image retrieval. The proposed approach is based on fuzzy shape contexts, Fourier transforms and Eigenshapes. First, fuzzy shape context histograms are computed. Then, a 2D FFT is performed on each 2D histogram to achieve rotation invariance. Finally, histograms are projected onto a lower dimensionality feature space whose basis is formed by a set of vectors called Eigenshapes. They highlight the most important variations between shapes. The proposed approach is translation, scale and rotation invariant. Classes of the medical IRMA database are used for experiments. Comparison with the known approach rotation invariant shape contexts based on feature-space Fourier transformation proves that the proposed method is faster, more efficient, and robust to local deformations.

Keywords: Image retrieval, Fuzzy Shape Contexts, Fourier transform, Eigenshapes, Radiological images.

1 Introduction

One of the most vivid fields of computer vision research is medical image processing [1] [2]. Medical image tools are used by physicians for diagnosis. So many works proposed new techniques of medical image processing [3] [4] [5] [6].

Medical image retrieval is a branch of medical image processing. The concept of content based image retrieval is used in many applications such as breast cancer diagnosis systems [7] [8] [9]. Each image in the database needs to be described by features providing its signature. Features extraction is based on visual characteristics. The best features when dealing with simple radiological image retrieval is shape information. In fact, using gray level based approaches is not sufficient. They are in most of cases coupled with edge detection techniques [10]. This work deals with shape descriptors. It improves the rotation invariant shape contexts based on feature-space Fourier transformation [11]. First, fuzzy shape context histograms are computed. Then, a 2D FFT is performed on each

2D histogram. Next, data is projected onto a more representative feature space highlighting the most important variations between shapes. Eigenshapes form the basis of the new space. This will be more detailed in the next section.

This paper is organized as follows: Section 2 presents the proposed approach. Section 3 presents experimental results. The conclusion comes in section 4.

2 Rotation Invariant Fuzzy Shape Contexts Based on Eigenshapes and Fourier Transforms

2.1 Previous Work

S. Belongie et al. initially proposed the Shape context feature descriptor used for shape matching and object recognition [12] [13] [14]. The proposed approach inspired many authors to propose variants of this descriptor [15] [16]. S. Yang and Y. Wang proposed the rotation invariant shape contexts based on feature-space Fourier transformation [11]. The main idea is to extract the shape of the object and pick up n points. They do not need to be key points such as corners. Shape context at a given point p_i is an histogram providing the distribution of vectors originating from p_i over relative positions by considering p_i as the center of a log polar coordinate system (Figure 1). This distribution provides a reach description of the shape localized at that point.

Shape context at a given point p_i is defined as follows:

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in \text{bin}(k)\} \tag{1}$$

Indeed, coordinates and tangents at each point are used to compute a set $\{(r_{ij}, \alpha_{ij}) | i, j = 1, 2, \dots, n\}$ of magnitudes and angles. For a point p_i , magnitudes are obtained by first computing distances l_{ij} between p_i and the remaining points:

$$l_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{2}$$

Then, a log scale is performed on all distances. Note that a length normalization is needed. Thus, every magnitude is divided by the mean distance r_0 . Finally r_{ij} is determined as:

$$r_{ij} = \frac{\log(l_{ij})}{r_0} \tag{3}$$

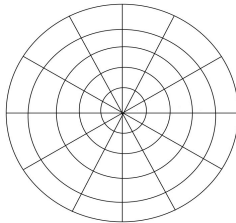


Fig. 1. Log polar grid with 60 bins used to compute shape contexts

Angles α_{ij} are defined as follows:

$$\alpha_{ij} = \arctan \frac{y_j - y_i}{y_j + y_i} \tag{4}$$

The obtained set $\{(r_{ij}, \alpha_{ij}) | i, j = 1, 2, \dots, n\}$ is used to compute the 2D histogram defining the shape context. Application of a 2D FFT on this 2D histogram provides rotation-invariance.

2.2 Fuzzy Shape Contexts

The main idea behind fuzzy shape contexts concept consists in considering that the belonging of a contour pixel to a given bin is not absolute. It also belongs to the surrounding bins with smaller weights. This makes the descriptor more robust to local deformations. Figure 2 shows the case of a local shape deformation supposing that a log-polar grid of four bins is used to compute shape context histograms. A pixel belongs to a given bin with weight $w_1 = 0.7$. It also belongs to the previous and next bins with weights $w_2 = w_3 = 0.15$.

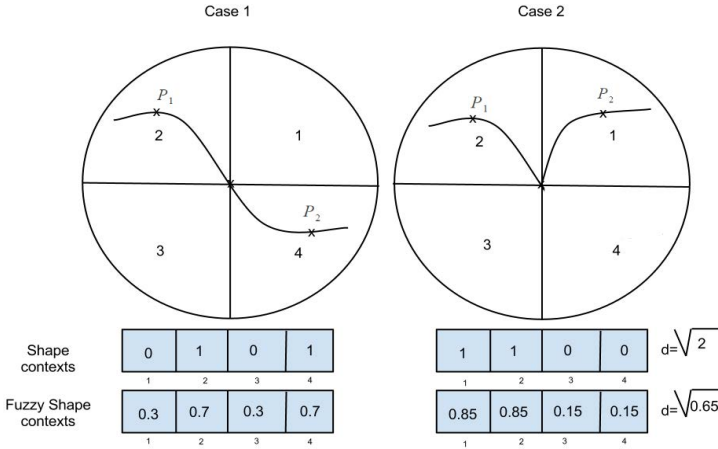


Fig. 2. A comparison between Shape contexts and Fuzzy Shape Contexts (illustration with one level-four bins)

The Euclidean distance is used to measure similarity between histograms. It is equal to $\sqrt{2}$ when shape contexts are used. However, it is equal to $\sqrt{0.65}$ when dealing with fuzzy shape contexts which are proven more robust to local deformations. Note that the difference $\delta = \sqrt{2} - \sqrt{0.65}$ is note huge. This is due to the fact that we are dealing with a local deformation.

In the rest of this work, a set of 12 equally log bins and 5 equally log radius bins is used to compute fuzzy shape contexts (Figure 3). Weights of belonging to a given bin are set empirically. There are three cases:

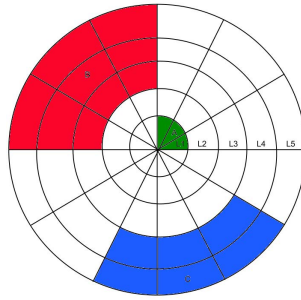


Fig. 3. Weight assignation

- a given pixel belongs to a bin of level L1 with weight $w_1 = 0.8$ and belongs to the next and the precedent bins with weight $w_1 = 0.1$ for each (Eg. Bin A).
- a given pixel belongs to a given bin of level L2, L3 or L4 with weight $w_2 = 0.6$ and belongs to all the surrounding bins with weight $w'_2 = 0.05$ for each (Eg. Bin B).
- a given pixel belongs to a given bin of level L5 with weight $w_3 = 0.75$ and belongs to all the surrounding bins with weight $w'_3 = 0.05$ for each (Eg. Bin C).

2.3 Eigenshapes

For a given image, a reference point corresponding to the closest pixel to the top left image corner is fixed. The next step is to pick up other $n - 1$ equidistant points. Every point is described via its fuzzy shape context which is a 2D histogram. Each histogram is then reshaped onto 1D vector which is added as a new line to the matrix representing the signature of that image. The signature is so an $n \times l$ matrix where n denotes the number of picked points and l denotes the number of bins. The next two sub-sections describe the training and recognition procedures.

Training. A set $S = \{S_1, S_2, \dots, S_m\}$ of m images is used for training. Each image is represented by an $n \times l$ matrix. Each matrix is converted onto a column vector ζ_i . ζ_i is a $z \times 1$ vector where $z = n \times l$. Then, the average shape vector τ is computed as follows:

$$\tau = \frac{1}{m} \sum_{i=1}^m \zeta_i \tag{5}$$

Next, each ζ_i is normalized by subtracting the mean shape:

$$\Theta_i = \zeta_i - \tau \tag{6}$$

Then, the covariance matrix C is computed as follows :

$$C = \frac{1}{m} \sum_{n=1}^m \Theta_n \Theta_n^t = AA^t \tag{7}$$

Where $A = [\Theta_1, \Theta_2, \dots, \Theta_m]$. Note that C in (7) is a $z \times z$ matrix and A is a $z \times m$ matrix. Eigenshapes are the eigenvectors U_i of AA^t .

Note that the matrix AA^t is very large so it is not practical for computations because of its dimension. Also, note that AA^t and A^tA have the same eigenvalues and their eigenvectors are related as follows: $U_i = AV_i$. Next, eigenvectors of A^tA are computed. Finally, m eigenvectors of AA^t are obtained following the relation: $U_i = AV_i$. Only k eigenvectors corresponding to the largest eigenvalues are kept. They form the basis of the new eigenshape space:

$$\Xi_k = [U_1, U_2, \dots, U_k] \tag{8}$$

Each normalized shape in the training database is so projected in the new space. It is represented as a linear combination of k eigenshapes:

$$\Theta_i^{proj} = \sum_{j=1}^k W_j U_j \tag{9}$$

where $W_j = U_j^t \Theta_i$. Next, every normalized training shape Θ_i is represented by a vector ω^i providing its coordinates in the new eigenshape space where:

$$\omega^i = \begin{pmatrix} W_1^i \\ W_2^i \\ \vdots \\ W_k^i \end{pmatrix} \tag{10}$$

Retrieval. Now, given a query image, the goal is to retrieve the most similar image to it in the database. First, it is reshaped onto a column vector ψ . Then, it is normalized: $\theta = \psi - \tau$. The next step is to project it on the eigenshape space.

$$\theta^{proj} = \sum_{i=1}^k W_i U_i \tag{11}$$

where $W_i = U_i^t \theta$. Finally, θ is represented as:

$$\Omega = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_k \end{pmatrix} \tag{12}$$

The last step is to compute $d = \min_l \|\Omega - \omega^l\|$. The corresponding image to vector ω^l is considered as the most similar one to the query image.

3 Experiments

3.1 Image Collection

The radiological IRMA database is used for experiments. It includes images of several body parts. Figure 4 shows some IRMA database samples.

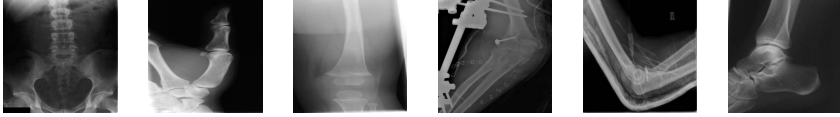


Fig. 4. IRMA samples

A set of 1000 images belonging to four classes (Hands, Breasts, Chests and Heads) is used. The number of images per class is the same. Figure 5 shows sample images of these classes.

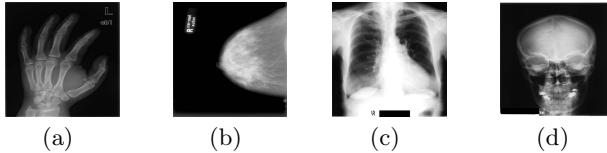


Fig. 5. Four classes used for performance measurement

Images in Figure 5 are randomly picked. They are used in the next sub-section as targets to evaluate the performance of the proposed approach. The Euclidean distance is used to measure the similarity between images.

3.2 Experimental Results

To evaluate the proposed approach, recall and precision measurements are used. Recall is defined as the ratio between the number of correctly retrieved images and the total number of images retrieved while precision is defined as the ratio between the number of correctly retrieved images by search and the total number of images used for test. For each measure of recall precision, the 10, 20, 40, 60, 80, 100, 150, 200 and 250 most similar images are taken in consideration. Figures 6, 7 and 8 shows recall versus precision for three tested approaches:

- FFT-RISC: Rotation-invariant shape contexts based on FFT [11].
- RISC-FFT-EIG: Rotation invariant shape contexts based on Fourier transforms and eigenshapes: histograms obtained by FFT-RISC are projected onto a new eigenshape space.

- Fuzzy RISC-FFT-EIG: Fuzzy Rotation invariant shape contexts based on Fourier transforms and Eigenshapes: Histograms obtained by FFT-RISC when using fuzzy bins are projected onto a new eigenshape space.

Figure 6 shows the recall precision curve for the Hand sample image (a) showing that Fuzzy RISC-FFT-EIG and RISC-FFT-EIG outperform significantly the FFT-RISC approach. Even when considering the best 250 retrieved images, precision rate remains superior to 90 %.

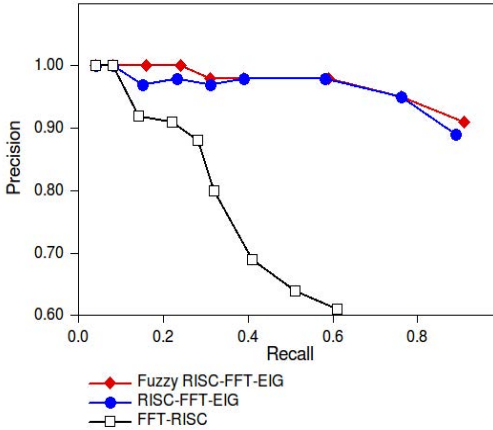


Fig. 6. Recall Vs. Precision for the Hand sample (a)

Recall and precision curve for the Breast sample image (b) is illustrated by Figure 7. The precision rate is equal to 100 % for the five first measurements for all of the three approaches. Fuzzy RISC-FFT-EIG and RISC-FFT-EIG provide better recognition rates than FFT-RISC when recall is higher than 0.4

Figure 8 shows recall precision curve for the Chest image sample (d). For the first measure, the precision rate is equal to 100 % for all of the tested approaches. Then, it is higher when using FFT-RISC. However the Fuzzy RISC-FFT-EIG and RISC-FFT-EIG outperform when recall is higher than 0.5.

To further prove the performance of the proposed approach, the average of the precision rate per class is computed considering the best 200 images retrieved. Results are illustrated by Table 1 showing that Fuzzy RISC-FFT-EIG and RISC-FFT-EIG outperform the FFT-RISC approach. This is due to elimination of noisy data.

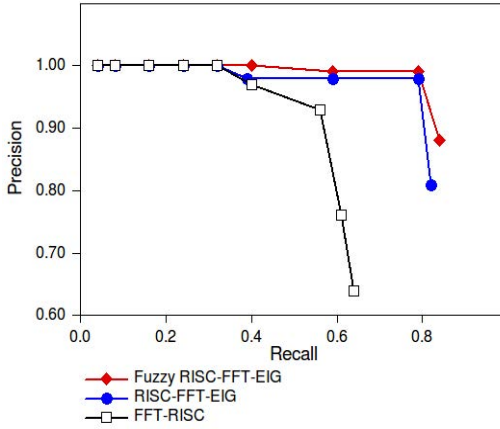


Fig. 7. Recall Vs. Precision for the Breast sample (b)

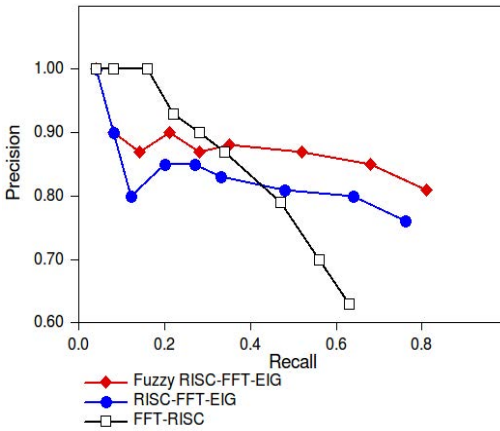


Fig. 8. Recall Vs. Precision for the Chest sample (c)

Table 1. Average of the precision rate per class considering the best 200 images retrieved

Image	FFT-RISC	RISC-FFT-EIG	Fuzzy-RISC-FFT-EIG
Hands	83.09	98.08	98.64
Breasts	85.65	93.47	94.1
Chests	98.49	97.01	97.92
Heads	95.67	94.06	94.02
Average	90.72	95.65	96.16

3.3 Discussion

Experimental results show that better results are obtained when histograms are projected in a new eigenshape space. The average of precision rate per class considering the best 200 images retrieved with RISC-FFT-EIG reaches 95.65 % while it is equal to 90.72 % with FFT-RISC. Other point to note is that using fuzzy shape contexts ameliorates results. In this case, the recognition rate is 96.16 %. Indeed, fuzzy shape contexts are more robust to local deformations. Note that there is no significant gap between results obtained by RISC-FFT-EIG and Fuzzy RISC-FFT-EIG approaches. In fact, local deformations do not affect significantly the performance of retrieval.

4 Conclusion

Shape context has been proven a very powerful shape descriptor. It is translation and scale invariant. Rotation invariance is achieved by application of 2D FFTs on the 2D histograms.

This work proves that using fuzzy bins makes the descriptor more robust to local deformations. Also, projecting data onto a lower dimensionality space highlighting the most important variations between shapes reduces time execution. In addition to that, better recognition rates are obtained. This is due to elimination of noisy data. Note that the major limitation of the proposed descriptor is the fact that it can not be used when dealing with images having many textures. The proposed approach can be improved if weights are set in respect to the linear distance between each pixel and the surrounding bins.

Acknowledgment. This work was supported by the New Brunswick Innovation Foundation (NBIF) and the Natural Sciences and Engineering Research council of Canada (NSERC). Authors would like to thank University of Ashen for providing the IRMA database.

References

1. Akl, C.B., Rubin, D.L., Napel, S., Beaulieu, C.F., Greenspan, H., Acarl, B.: Content-based image retrieval in radiology: current status aExperimentsnd future directions. *Digit. Imaging* 24, 208–222 (2011)
2. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of Content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics* 73(1) (2004)
3. Šajin, L., Kukar, M.: Image processing and machine learning for fully automated probabilistic evaluation of medical images. *Computer Methods and Programs in Biomedicine* 104(3), 75–86 (2011)
4. Krefting, D., Vossberg, M., Hoheisel, A., Tolxdorff, T.: Simplified implementation of medical image processing algorithms into a grid using a workflow management system. *Future Generation Computer Systems* 26(4), 681–684 (2010)

5. Mahmoudi, S.E., Akhondi-Asl, A., Rahmani, R., Faghih-Roohi, S., Taimouri, V., Sabouri, A., Soltanian-Zadeh, H.: Web-based interactive 2D/3D medical image processing and visualization software. *Computer Methods and Programs in Biomedicine* 98(2), 172–182 (2010)
6. Martínez, A., Jiménez, J.J.: Tracking by means of geodesic region models applied to multidimensional and complex medical images. *Computer Vision and Image Understanding* 115(8), 1083–1098 (2011)
7. Wei, L., Yang, Y., Nishikawa, R.M.: Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognition* 42(6), 1126–1132 (2009)
8. Chen, D.R., Huang, Y.L., Lin, S.H.: Computer-aided diagnosis with textural features for breast lesions in sonograms. *Computerized Medical Imaging and Graphics* 35(3), 220–226 (2011)
9. Kuo, W.J., Chang, R.F., Lee, C.C., Moon, W.K., Chen, D.R.: Retrieval technique for the diagnosis of solid breast tumors on sonogram. *Ultrasound in Medicine and Biology* 28(7), 903–909 (2002)
10. Bottigli, U., Golosio, B.: Feature extraction from mammographic images using fast marching methods. *Nuclear Instruments and Methods in Physics* 487(1-2), 209–215 (2002)
11. Yang, S., Wang, Y.: Rotation invariant shape contexts based on feature-space Fourier transformation. In: *Fourth International Conference on Image and Graphics* (2007)
12. Belongie, S., Malik, J.: Matching with Shape Contexts. In: *IEEE on Content based Access of Image and Video Libraries, CBAIVL 2000* (2000)
13. Belongie, S., Malik, J., Puzicha, J.: Shape Context: A new descriptor for shape matching and object recognition (2001)
14. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 24(4) (2002)
15. Diplaros, A., Gevers, T., Patras, I.: Color-Shape Context for Object Recognition. In: *IEEE Workshop on Color and Photometric Methods in Computer Vision, in Conjunction with the 9th Int. Conf. Computer Vision* (2003)
16. Kortgen, M., Park, G.-J., Novotni, M., Klein, R.: 3D Shape Matching with 3D Shape Contexts

A 2D Rigid Point Registration for Satellite Imaging Using Genetic Algorithms

Fatiha Meskine, Nasreddine Taleb, and Ahmad Almhdie-Imjabber

Laboratoire RCAM, University of Sidi-bel-Abbes 22000, Algeria
{me_fatiha, ne_taleb} @univ-sba.dz

Laboratoire PRISME Université d'Orléans 45067 Orléans, France
ahmad.almhdie@univ-orleans.fr

ISTO Institute, CNRS, 45071 Orleans, France
ahmad.almhdie@univ-orleans.fr

Abstract. Image registration is an important step for a great variety of applications such as remote sensing, medical imaging, and multi-sensor fusion-based target recognition. The objective is to find, in a huge search space of geometric transformations, an acceptable accurate solution in a reasonable time to provide better registered images for high quality products. In the broad area of global optimization methods, Genetic Algorithms form a widely accepted trade-off between global and local search strategies. They are well-investigated and have proven their applicability in many fields. In this paper, we present an efficient 2D point based rigid image registration method integrating the advantage of the robustness of GAs in finding the best transformation between two images. The algorithm is applied for registering SPOT images and the results show the effectiveness of this approach.

Keywords: Image registration, point registration, feature points, satellite images, Genetic Algorithms.

1 Introduction

The process of image registration can be formulated as a problem of optimizing a function that quantifies the match between the original and the transformed image. Several image features have been used for the matching process, depending on the modalities used, the specific application and the implementation of the transformation. The registration process can be divided into three main categories: point-based, surface-based and volume-based methods. *Point-based* registration involves the determination of the co-ordinates of corresponding points in different images and the estimation of geometrical transformation using these corresponding points. Then, the task of registration is to place the data into a common reference frame by estimating the transformations between the datasets. What makes the problem difficult is that correspondences between the point sets are unknown a-priori. A popular approach to solving the problem is the class of algorithms based on the Iterated Closest Point (ICP).

The ICP algorithm described by Besl and McKay [1] is well known for aligning 3D object models. Originally ICP starts with two data sets (mostly points) and an

initial guess for their rigid body motion. Then the transformation is refined by repeatedly generating pairs of corresponding points of the sets and minimizing an error metric. ICP algorithms are mostly applied to 2D or 3D point sets [2]. ICP is attractive because of its simplicity and its performance. Although the initial estimate does need to be reasonably good, the algorithm converges relatively quickly. This algorithm is composed of two basic procedures. The first one is to find matching points, and the second one is to estimate the transformations iteratively for these points until some stop distance criteria is satisfied. Another approach to the registration of images consists in determining a set of matches through a search process instead of the classical approach based on distances. This approach consists in finding a solution close to the global minimum in a reasonable time. This can be done by means of a Genetic Algorithm (GA).

In recent years, GAs have been intensively investigated and applied to many optimization problems [3]. GAs are especially appropriate for the optimization in large search spaces, which are unsuitable for exhaustive search procedures. GAs do a trade-off between the exploration of the search space and the exploitation of the best solutions found so far. A number of authors have used GAs for full-view image matching in various forms. Jacq and Roux [4] use GAs for registration of 3D medical images. Brunnström and Stoddard [5] used a GA to find an initial guess for the free-form matching problem that is finding the translation and the rotation between an object and a model surface. In contrast to the 2D–3D registration, numerous methods exist to precisely register 3D data by iterative algorithms like the Iterative Closest Point and its variants [6].

In this paper, a novel approach is developed based on the application of GAs for registration of two data sets from satellite images. The remainder of the paper is organized as follows: the second topic gives an overview of genetic algorithms and their basics. The third topic describes the registration strategy used in this work. The feature point extraction algorithm based on the NSCT method is given in the fourth topic. The simulation results are presented in the fifth topic, and in the last we finalize with a conclusion.

2 Genetic Algorithms Overview

The GA is a well-known efficient global optimization algorithm, introduced by Holland [7] in 1975, that utilizes the concept of biological structure to natural selection and survival of the fittest. Due to the fact that the method requires no previous experience on the problem, it is applied on various problems whereof some characteristics are mentioned in [8].

The general principle of a genetic algorithm is to subject a population of individuals to an evolutionary process, encoded as chromosomes, which represent some possible solutions to a searching problem. During evolution, an aptitude value is assigned to each individual obtained from a specifically defined function for the problem to be solved. This function, called aptitude or fitness function, should be designed in such a way that it favors the most apt or adequate individuals as the

solution to the problem. The aptitude assigned to each individual is taken into account in the selection of the parents who will take part in the reproduction process. Here there is an exchange of genetic material or content of a pair of selected individuals to generate two new individuals or two new possible solutions to the problem that, according to a replacement mechanism, are incorporated into the population. The new descended individuals are also subjected to a mutation process which is a random perturbation of its genetic material in order to offer variability and also to enrich the exploration of the possible solutions to the problem. These are represented as chromosomes. Finally, after having completed a certain number of cycles of aptitude assignation, reproduction, mutation, and replacement (called generations), the individual with better aptitude is chosen as the best solution to the problem. The GA cycle is shown in the following figure.

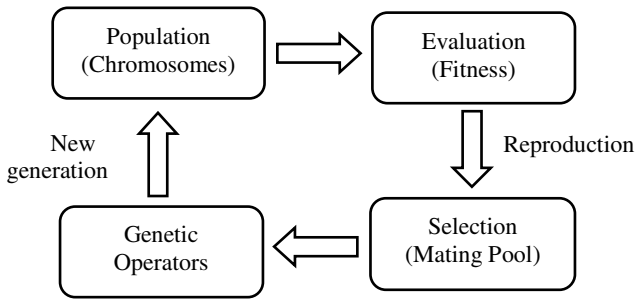


Fig. 1. A GA scheme

3 Registration Strategy

For the GAs to be successful, how to formulate the chromosome and fitness function is very important. The GAs will have better convergence behavior if the fitness function is generally continuous and the chromosome with the optimal fitness value corresponds to the target solution. In the following, formulations of the chromosomes and the fitness function for point registration are described.

3.1 Similarity Metric

One of the most important components of any image registration method is the similarity metric. This is considered as a function F that measures the goodness of a given registration solution, that is, of a registration transformation f .

A genetic algorithm uses a fitness function to determine the performance of each artificially created chromosome; therefore the fitness function should measure the registration quality of each chromosome. A GA should try to find a chromosome with the minimum Euclidean distance between each correspondence pair [9].

Assume that the given two data sets to be matched are $P = \{p_1, p_2, \dots, p_m\}$ and $Q = \{q_1, q_2, \dots, q_n\}$ where m is not necessarily equal to n .

If the registration parameters are given, then for any point p_i in P , we can use the following criterion to determine its possible correspondent q_i in Q :

$$Q_i = \operatorname{argmin} \|q - (Rp_i + T)\|. \quad (1)$$

Thus the objective is to minimize the Euclidean distance between the transformed point $Rp_i + T$ and q in the Q . A suitable transform means the distance error between P and Q is minimized. Therefore, the *fitness function* of GA to be *minimized* is described in the following equation

$$F = \operatorname{median} \|p_i - q_i\|. \quad (2)$$

We assume that the type of transformation is rigid. Then, for the data point in the model image with coordinates x , y and intensity value I , its image is x' , y' , and I' , and then they are related by the mapping:

$$Pt = R * p_i + T. \quad (3)$$

Where, R is the rotation matrix and T is a translation vector in both x and y directions.

3.2 Chromosome Encoding

The geometric transformation between two models can be defined by three parameters, defined as a chromosome. Each parameter corresponds to one of the genes in the chromosome.

Using a bit encoding scheme for the chromosome string, the rotational transform (R), x -axis translational transform (X), and y -axis translational transform (Y) are encoded. An 8-bit field is used to represent the possible relative rotation of the input image to the reference image. Likewise, 6 bits are used to express the translation in the x -axis and 6 more for the y -axis. Thus, the total length of the chromosome is 20 bits.

All representations are signed magnitude, using one bit for the sign and the rest of the bits to represent the magnitude of the rotation or translation. Thus, the relative rotation has a range of ± 128 degrees, while relative translation in the x (or y) direction has a range of ± 32 pixels. Every individual represents a combination of all transformation parameters which describe an image transformation.

4 Feature Points Extraction

The purpose of the feature extraction is to derive features that describe image characteristics that are relevant in a co-registration process and which can be used to select a subset of regions and choose an appropriate method for each. The feature extraction approach used in this paper exploits a non-sampled directional multi-resolution image representation to capture significant image features across spatial and directional resolutions.

Recently, Cunha et al. [10] proposed the non-sampled contourlet transform (NSCT) which is a shift-invariant version of the contourlet transform and

multidirectional expansion that has a fast implementation. The NSCT eliminates the downsamplers and the upsamplers during the decomposition and the reconstruction of the image. Instead, it is built ahead the nonsubsampling filter banks which provide multiscale decomposition and the nonsubsampling directional filter banks which provide directional decomposition.

The proposed feature extraction method is described in the following algorithm [11]:

Step 1: Compute the NSCT coefficients of both images for N levels and L directional subbands.

Step 2: Compute the difference between each directional subband at one level and the corresponding one at another level. L difference subbands will be obtained at the end.

Step 3: At each pixel location, compute the maximum magnitude of all obtained difference subbands. These points are called “maxima of the NSCT coefficients”.

Step 4: A hard thresholding procedure is then applied on the NSCT maxima image in order to eliminate non significant feature points. A point is recorded if NSCT maxima $> Th$,

Where $Th = c(\sigma + \mu)$, c is a parameter whose value is defined by the user, and σ and μ are the standard deviation and mean of the NSCT maxima image, respectively.

Step 5: Take a block neighborhood of size $w \times w$ and find one local maximum in each neighbourhood, this will eliminate maxima that are very close to each other. The locations of the obtained thresholded NSCT maxima are taken as the extracted feature points.

After the feature points are detected from the images to be registered, a correspondence mechanism between these two feature points sets must be established in order to refine the control points. The objective is that each feature point in the reference image is paired with its correspondent in the sensed image. In this work, correlation based similarity measure is used to establish the correspondence between the two feature point sets.

5 Simulation Results

The parameters of GAs used in this test are: The population size in each generation is restricted to 100 individuals with a crossover probability of 0.75 and a mutation probability of 0.05. GA meets the criterion within 200 generations. To improve the performance of GAs, we have used two techniques named elitism and fitness sharing. Elitism consists of preserving the best individuals at each generation and fitness sharing to keeps the population diversity.

We have applied our algorithm on SPOT satellite images. The transformed image to be corrected is rotated by 7 degrees and displaced by 13 and 9 pixels in X and Y directions from the center of the reference image. The sensed (reference) image is warped using bilinear transformation.

Evaluation of the fitness function described above requires a search on the closest point from a data set given an input data point. The corresponding searching time will be very long and becomes a major obstacle in utilizing the GA approach for practical applications. Therefore, in the first experiment and in order to limit the point set representing the image, we choose a window of size 40×40 pixels from the center of both images in order to have about 1600 points at each model.

Figures 2 and 3 illustrate the performance of GAs process during the run. In figure 2, we see the evolution of the best fitness value at each generation. This value which is median (dist) is minimized from generation to another until found the optimal fitness value which corresponds to the optimal solution. Figure 3 depicts the evolution of the parameters (R,X,Y) during the generations. The red dashed lines show the initial parameters and blue lines show the optimal parameters found during the run the GAs. We see that the optimal parameters values are closer to the initial parameters.

The results of the parameters transformation found with this technique of GAs based point registration noted by 'GAs proposed' is compared with other registration methods as the ICP algorithm (noted ICP) and the intensity registration based method (noted GAs intensity) for which the objective is it to maximize the correlation coefficient of the two images.

The analytical results are depicted in table 1. The results found with 'GAs proposed' are similar to those of ICP. However, the results of the GAs intensity method are slightly different particularly for the X translation. So, we can say that the point registration is more robust and accurate than those of intensity methods.

Table 1. Analytical results of the parameters found with different methods

Methods	Rotation R	Translation X	Translation Y
ICP	-7	-13	-8
GAs intensity	-7	-10	-8
GAs proposed	-7	-13	-8

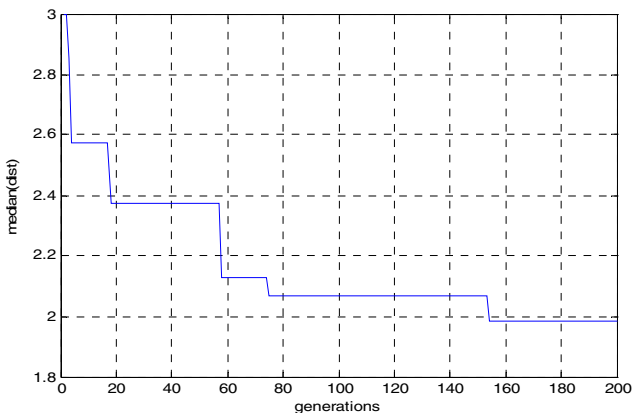


Fig. 2. Evolution of the best fitness during the run of GAs

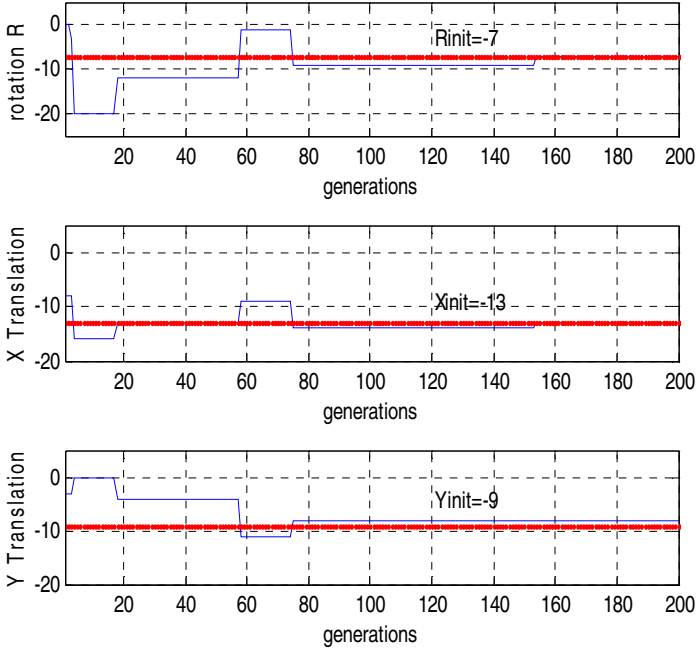


Fig. 3. Evolution of the parameters transform during the generations

In the case of important sizes where the data sets are very large and time consuming is very important, we have suggested to employ the NSCT method for extraction of the feature points as cited in section IV (the second experiment). The NSCT decomposition of images was performed with the following parameters: $N=4$ levels and $L=4$ sub-bands at each level; $c=1$ and the block neighborhood is of size $w = 32$. An example of interest points extracted using the NSCT method is shown in figure 4 for both images of size 512×512 pixels.

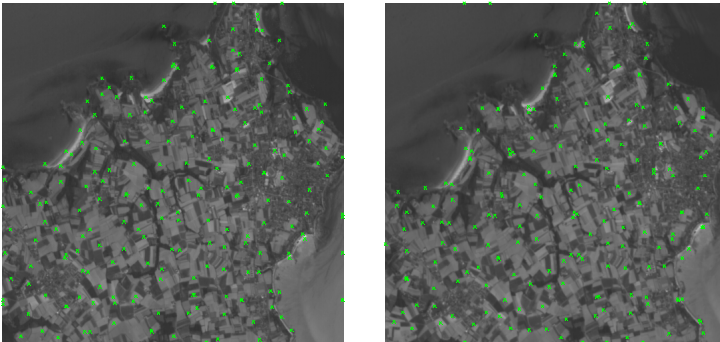


Fig. 4. Feature points extraction using the NSCT method of the reference image at the left and the transformed image at the right

After selecting the corresponding feature points with the NSCT method, we apply the GAs process for registration of the corresponding point sets pair. The registered image obtained is shown in the following figure.

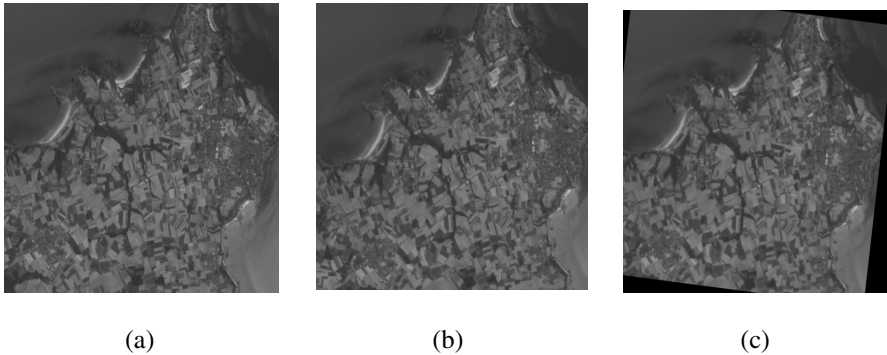


Fig. 5. Image registration results: (a) reference image which corresponds to model M, (b) transformed image which corresponds to model P to be registered, and (c) the resulting image registered with the GAs process

6 Conclusion

Point set registration is among the most fundamental problems in vision research. It is widely used in areas such as range data fusion, medical image alignment, object localization, tracking, object recognition, just to name a few. The goal of the registration task is to find the transformation that best represents the relative transformation between two sets data. In this paper, we present an efficient point based rigid 2D image registration method. The registration optimization problem is solved by the Genetic algorithms method.

GAs represent an intelligent exploitation of a random search used to solve optimization problems. In this work, GAs have been used to estimate the rotation angle and displacement values at x-axis and y-axis. We have considered an image registration algorithm based on the alignment of a set of feature points. Our interest in this problem stems from its application in remote sensing, and in particular in the alignment of satellite images. We have presented a novel approach of a 2D point registration based on the GAs for which the results have proven its accuracy compared to the intensity based methods.

References

1. Besl, P.J., McKay, N.D.: A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), 239–256 (1992)
2. Stoddart, A.J., Hilton, A.: Registration of multiple point sets. In: *Proc. ICPR*, pp. 40–44 (1996)

3. Goldberg, D.E.: Genetic Algorithm in search, optimization and machine learning. Addison Wesley (1989)
4. Jacq, J., Roux, C.: Registration of 3D images by genetic optimization. *Pattern Recognition Letters* 16, 823–841 (1995)
5. Brunnström, K., Stoddart, A.J.: Genetic algorithms for free-form surface matching. In: *Proc. 13th International Conference on Pattern Recognition*, vol. 4, pp. 689–693 (1996)
6. Chetverikov, D., Stepanov, D., Krsek, P.: Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing* 23(3), 299–309 (2005)
7. Holland, J.H.: *Adaptation in Natural and Artificial System*. University of Michigan Press (1975)
8. Coley, D.: *An Introduction to Genetic Algorithms for Scientists and Engineers*. World Scientific Press (1999)
9. Chow, C.K., Tsui, H.T., Lee, T.: Fast Free-form Surface Registration by A New Genetic Algorithm. In: *The Fifth Asian Conference on Computer Vision*, Melbourne (2002)
10. da Cunha, A.L., Zhou, J., Do, M.N.: The Nansubsampled Contourlet Transform: Theory, design, and applications. *IEEE Trans. on Image Processing* 15(10), 3089–3101 (2006)
11. Serief, C., Barkat, M., Bentoutou, Y., Benslam, M.: Robust feature points extraction for image registration based on the nonsubsampling contourlet transform. *International Journal of Electronics and Communications* (63), 148–152 (2009)

Image Quality Assessment Measure Based on Natural Image Statistics in the Tetrolet Domain

Abdelkaher Ait Abdelouahad¹, Mohammed El Hassouni²,
Hocine Cherifi³, and Driss Aboutajdine¹

¹ LRIT URAC- University of Mohammed V-Agdal-Morocco
{a.abdelkher,mohamed.elhassouni}@gmail.com

² DESTEC, FLSHR- University of Mohammed V-Agdal-Morocco
hocine.cherifi@u-bourgogne.fr

³ Le2i-UMR CNRS 5158 -University of Burgundy, Dijon-France
aboutaj@fsr.ac.ma

Abstract. This paper deals with a reduced reference (RR) image quality measure based on natural image statistics modeling. For this purpose, Tetrolet transform is used since it provides a convenient way to capture local geometric structures. This transform is applied to both reference and distorted images. Then, Gaussian Scale Mixture (GSM) is proposed to model subbands in order to take account statistical dependencies between tetrolet coefficients. In order to quantify the visual degradation, a measure based on Kullback Leibler Divergence (KLD) is provided. The proposed measure was tested on the Cornell VCL A-57 dataset and compared with other measures according to FR-TV1 VQEG framework.

Keywords: RRIQA, Tetrolet transform, natural image statistics, Gaussian Scale Mixture.

1 Introduction

Recently, several RR methods have been introduced but few of them are general-purpose. The first general-purpose RR methods was introduced by Wang [1] in the steerable pyramids domain named WNISM. The KLD was used to quantify the difference between two subband coefficient histograms. The first histogram is computed from the distorted image while the second is summarized using the Generalized Gaussian Density (GGD) model parameters instead of sending all histogram bins. Promising results were obtained for five distortions in the LIVE dataset. Tao et al [2] have proposed the contourlet transform which is effective in dealing with directional information like edges. After CSF masking, the JND is applied to remove visually insensitive coefficients. A histogram is formed from the remaining coefficient. Finally, the histogram is normalized and considered as RR feature. Results were presented for two distortions from the LIVE dataset : JPEG and JPEG2000 compressions. Li *et al* [3] investigated the Divisive Normalization Transform (DNT) to take into account the dependencies between wavelet coefficients which were ignored in the WNISM. The measure based on the DNT improved the WNISM, specially when it was tested on a set formed by different distortions. Nevertheless, its performances can change significantly since it depends on some parameters which need to be trained. In [4]

the construction of the Strongest Component Map (SCM) is proposed. The Weibull distribution parameters are estimated from the SCM coefficients histograms. Finally, only the scale parameter β is involved in a measure called β W-SCM. Experiments with the LIVE dataset show significant correlation between the model predictions and the subjective scores, nearly the same as WNISM. In [5] grouplets have been used to capture image geometric structures and orientations. To incorporate HVS characteristics, a Contrast Sensitivity Function (CSF) is applied before measuring the changes between the reference and the distorted images. Their results show some significant improvements for JPEG distorted images as compared to WNISM.

Inspired by the work of Wang, we have proposed the use of the BEMD (Bi-dimensional Empirical Mode Decomposition) in the general scheme as an adaptive decomposition. Although the BEMD-based method outperforms the WNISM over several distortions in the TID 2008 dataset, low correlations with human judgment were obtained.

In this work, we propose a joint probability distribution of tetrolet coefficients using GSM model. This allows us to exploit the dependencies between tetrolet coefficients. A GSM model is defined as the product of zero mean Gaussian vector and positive random variable called multiplier. Here, we propose Weibull distribution to model the multiplier distribution. Then, assuming the independency between GSM components (the multiplier and the Gaussian vector) we derived an expression for the KLD in order to evaluate the visual quality of a processed image.

The rest of this paper is organized as follows. In section 2 we give a brief review of the tetrolet transform, in section 3 we explain how we model the dependencies between tetrolet coefficients using the GSM model, we present the distortion measure in section 4. Section 5 is reserved for experimental results and finally a conclusion ends the paper.

2 Tetrolet Transform

Nowadays, a sparse representation is required in image processing techniques. In such representation the energy of the signal is concentrated in few number of coefficients not null. This facilitates the feature extraction step used in image retrieval, image classification and RRIQA algorithms. Although wavelets were introduced for this aim, they can take advantage only of singularity points. Thus directional information like edge is disregarded. The idea of tetrolet transform [6] is to allow more general partitions which capture the image local geometry by bringing the "tiling by tetrominoes" problem into play.

Tetrominoes are derived from the well know game "tetris". They were introduced by Golomb [7]. We can obtain a tetromino by connecting four equal sized square. Disregarding rotation and isometric we have five free tetrominoes as shown in Figure 1. The Haar transform is a special case, since it considers only the first tetromino (square). To use other tetrominoes we should have at least a 4×4 blocks (Figure 2) which will give 117 possibility, whereas a 8×8 blocks gives $117^4 > 10^8$ possibilities. From computational complexity standpoint, it's clear that the first choice is the reasonable one. Therefore, tetrominoes ensure more directions when rotations and reflections are considered. To illustrate this let's take from Figure 2 (Line 4) the third covering (from left to right), eight other coverings are possible with different directions are shown in Figure 3.



Fig. 1. The five free tetrominoes

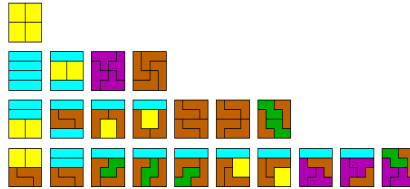


Fig. 2. The 22 fundamental forms tiling a 4 × 4 board

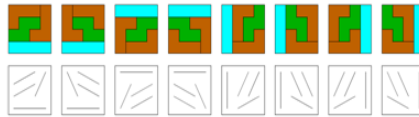


Fig. 3. Different directions covered by the same tetrominoes

2.1 Computing Tetrolet Transform

The computation of the tetrolet transform consists in two stages. First, the tiling by tetrominoes is achieved ensuring an optimal covering for each 4×4 block $Q_{i,j}$ in the image. Second, the Haar transform is applied to the tetrominoes of each covering. More precisely, let us take an image $\mathbf{a}^0 = [a(i, j)]_{i,j=0}^{N-1}$, N is a power of 2, i.e $N = 2^p, p \in \mathbb{N}$ and we suppose that we are in the r^{th} level. The image is decomposed into 4×4 blocks, for each block we consider the 117 possible covering $c = 1, \dots, 117$. The Haar transform is then applied to the tetrominoes forming the investigated covering. This leads to four low-pass coefficients and 12 tetrolet coefficients as follows :

$$\mathbf{a}^{r,(c)} = (a^{r,(c)}[s])_{s=0}^3 \quad \text{and} \quad \mathbf{w}_l^{r,(c)} = (w_l^{r,(c)}[s])_{s=0}^3$$

c and r refer to the actual covering and the actual level of decomposition respectively, while s refers to the tetrominoes of the covering and l refers to the three high-pass parts. The optimal covering C_{op} is then qualified as the one whose tetrolet coefficients provide the minimal l^1 :

$$\begin{aligned} C_{op} &= \underset{c}{\operatorname{argmin}} \sum_{l=1}^3 \|\mathbf{w}_l^{r,(c)}\|_1 \\ &= \underset{c}{\operatorname{argmin}} \sum_{l=1}^3 \sum_{s=0}^3 |w_l^{r,(c)}[s]| \end{aligned} \tag{1}$$

In other words, the smaller is the magnitude of the 12 tetralet coefficients, the minimal is the l^1 norm. Thus we obtain the optimal covering and a sparse image representation. Once we get the optimal covering C_{op} , we store the corresponding four low-pass coefficients and 12 tetralet coefficients : $[a^{r,(c_{op})}, w_1^{r,(c_{op})}, w_2^{r,(c_{op})}, w_3^{r,(c_{op})}]$. Doing this for all blocks $Q_{i,j}$ in the image we achieve the tetralet transform. Before applying further levels of the tetralet transform, we should rearrange the components of the vector $a^{r,(c_{op})}$ into 2×2 matrix using a reshape function :

$$a^r_{Q_{i,j}} = R(a^{r,(c_{op})}) = \begin{pmatrix} a^{r,(c_{op})}[0] & a^{r,(c_{op})}[2] \\ a^{r,(c_{op})}[1] & a^{r,(c_{op})}[3] \end{pmatrix} \tag{2}$$

3 Joint Statistics of Tetralet Coefficients

The tetralet transform provides a multi-resolution representation with three orientations since it is derived from the Haar wavelet transform. Here, we propose to exploited the dependencies between tetralet coefficients as it was done for wavelet coefficients [8] as the same as for the curvelet coefficients [9]. The Gaussian Scale mixture (GSM) model has been used to model both marginal and joint statistics of natural image wavelet coefficients [10]. Let us consider a N -length random vector Y . we assume that Y in our study is formed from coefficients clustered around a given coefficient $y^{s,o}$ at scale s and orientation o . Y is a GSM if it can be written as the product of a zero mean Gaussian random vector U with covariance matrix M and a positive scalar random variable x called multiplier:

$$Y \doteq x.U \tag{3}$$

\doteq denotes equality in probability. U and x are independent. If we denote $p_x(x)$ as the density of the variable x the density of Y can be expressed as [10]:

$$p_Y(Y) = \int \frac{1}{[2\pi]^{\frac{N}{2}} |x^2 M|^{\frac{1}{2}}} \exp\left(-\frac{Y^T M^{-1} Y}{2x^2}\right) p_x(x) dx \tag{4}$$

To obtain an explicit expression of the PDF of Y we should specify the density of the multiplier x . Since the multiplier variable is positive, several distributions can be considered. Here, we choose Weibull density. To this end, we should estimate first the multiplier. As this later is unknown, we can estimate it by maximum-likelihood method [10] of the observed coefficients given by :

$$\begin{aligned} \hat{x} &= \underset{x}{\operatorname{argmax}} \{ \log p(Y|x) \} \\ &= \underset{x}{\operatorname{argmin}} \left\{ N \log x + \frac{Y^T M^{-1} Y}{2x^2} \right\} \\ &= \sqrt{\frac{Y^T M^{-1} Y}{N}} \end{aligned} \tag{5}$$

where M is the covariance matrix of the Gaussian vector estimated from the tetralet coefficients and N is the length of the vector Y . Figure 4 illustrates Weibull fitting to the estimated multiplier.

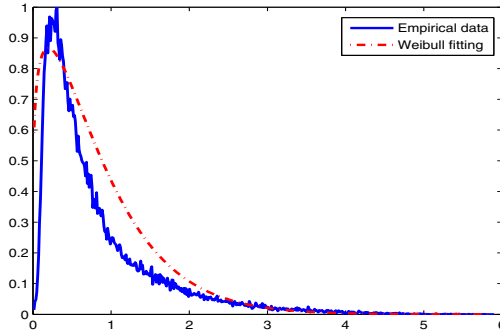


Fig. 4. Weibull distribution fitted to empirical histogram of the estimated multiplier

The PDF of Weibull distribution is given by:

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-(x/\lambda)^k} \quad (6)$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. Inserting the equation (6) in equation (4) the PDF of Y becomes :

$$p_Y(Y) = \int \frac{kx^{k-1}}{[2\pi]^{\frac{N}{2}} |x^2 M|^{\frac{1}{2}} \lambda^k} \exp \left(- \left(\frac{Y^T M^{-1} Y}{2x^2} + \left(\frac{x}{\lambda} \right)^k \right) \right) dx \quad (7)$$

4 Distortion Measure

In the previous section we have represented the joint statistics of tetrolet coefficients using a univariate Weibull distribution and a multivariate Gaussian distribution. Considering a neighborhood of dimension equals to 9 (3×3). At the sender side, we apply two levels tetrolet transform to the reference image. This leads to six tetrolet coefficients subbands (2 scales $\times 3$ orientations). From each subband three features are extracted : the covariance matrix M and the Weibull parameters (λ, k) . The extracted features are considered as RR side information. Similarly, the same features are extracted from the distorted image at the receiver side and we consider them as reduced description (RD). A dissimilarity measure is required to compare the RR to the RD and thus quantify the visual degradation. According to our knowledge a closed analytical form of the KLD for the proposed joint distribution in equation (7) does not exist. To resolve this problem, let us consider two joint distributions $P_1(Y; M_1, k_1, \lambda_1)$ and $P_2(Y; M_2, k_2, \lambda_2)$, where Y is a GSM vector. Since the components of the GSM (the multiplier and the Gaussian vector) are independent, we can derive an expression for the KLD between two joint distributions as the sum of the KLD between two multivariate Gaussian densities and the KLD between two Weibull distributions. In other words :

$$KLD(P_1(Y; M_1, k_1, \lambda_1) || P_2(Y; M_2, k_2, \lambda_2)) = KLD(P_1(x; k_1, \lambda_1) || P_2(x; k_2, \lambda_2)) \\ + KLD(P_1(U; M_1) || P_2(U; M_2)) \quad (8)$$

Now, that we have a closed analytical form for the KLD for both, Weibull distribution and the multivariate Gaussian density we can easily derive the KLD for the proposed joint distribution as:

$$\begin{aligned}
 KLD(P_1(Y; M_1, k_1, \lambda_1) || P_2(Y; M_2, k_2, \lambda_2)) &= \Gamma\left(\frac{\lambda_2}{\lambda_1} + 1\right) \left(\frac{k_1}{k_2}\right)^{\lambda_2} + \ln(k_1^{-\lambda_1} \lambda_1) - \ln(k_2^{-\lambda_2} \lambda_2) \\
 &+ \ln(k_1)(\lambda_1 - \lambda_2) + \gamma \frac{\lambda_2}{\lambda_1} - \gamma - 1 \\
 &+ 0.5 \left[\text{tr}(M_2^{-1} M_1) + \ln\left(\frac{|M_2|}{|M_1|}\right) - N \right] \quad (9)
 \end{aligned}$$

where γ denotes the Euler-Mascheroni constant ($\gamma \approx 0.57721$) and $\Gamma(\cdot)$ is the Gamma function.

First, the distance in equation (9) is computed to quantify the dissimilarity between two tetrolet coefficient subbands, the first from the reference image and the second is its correspondent from the distorted image. Finally, the dissimilarities between the subbands are combined to produce a global dissimilarity as follows :

$$Q = \log_2\left(1 + \frac{1}{D_0} \sum_{i=1}^L D_i\right) \quad (10)$$

where L is the number of the subbands, D_0 is a constant to control the scale of the distortion measure and it is equal to 0.1. The log function is involved here to reduce the difference between a high values and a low values of D , so that we can have values in the same order.

5 Experimental Results

Our experimental test was carried out using the Cornell VCL-A 57 [11] dataset. It provides 60 distorted images. Three reference images are altered with six distortions labeled : FLT, NOZ, JPG, JP2, DCQ and BLR. The labels refer to quantization of the LH subbands of a five-level DWT of the image using the 9/7 filters, additive Gaussian white noise, baseline JPEG compression, JPEG2000 compression using the 9/7 filters, JPEG2000 compression using the 9/7 filters with the dynamic contrast-based quantization algorithm, blurring by using a Gaussian filter, respectively. Each image in the Cornell VCL-A57 has its Mean Opinion Score (MOS). The subjective scores must be compared in term of correlation with the objective scores. These objective scores are computed from the values generated by the objective measure, using a non linear function according to the Video Quality Expert Group (VQEG) Phase I FR-TV [12]. Here, we use a four parameters logistic function.

$$\text{logistic}(\gamma, Q) = \frac{\gamma_1 - \gamma_2}{1 + e^{-\left(\frac{D - \gamma_3}{\gamma_4}\right)}} + \gamma_2 \quad (11)$$

where $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$.

Thus, the predicted MOS is given by :

$$MOS_p = \text{logistic}(\gamma, Q) \quad (12)$$

Once the nonlinear mapping is achieved, we obtain the predicted objective quality scores. To compare the subjective and objective quality scores, several metrics were introduced by the VQEG. In our study, we compute the correlation coefficient to evaluate the accuracy prediction and the Rank order coefficient to evaluate the monotonicity prediction. Table 1 shows the results for the Cornell VCL A-57 dataset.

Table 1. Performance evaluation for the proposed measure using Cornell VCL A-57 dataset

Dataset	FLT	JPG	JP2	DCQ	BLR	NOZ	All
Correlation Coefficient							
<i>Proposed</i>	0.71	0.96	0.83	0.95	0.91	0.86	0.70
<i>DNT</i>	0.76	0.91	0.81	0.90	0.93	0.99	0.66
Method in [13]	0.49	0.85	0.78	0.93	0.76	0.62	0.31
<i>PSNR</i>	0.91	0.70	0.79	0.56	0.59	0.93	0.63
<i>MSSIM</i>	0.92	0.91	0.87	0.94	0.79	0.88	0.72
Rank-Order Correlation Coefficient							
<i>Proposed</i>	0.46	0.96	0.81	0.90	0.90	0.80	0.74
<i>DNT</i>	0.50	0.76	0.80	0.66	0.80	0.98	0.70
Method in [13]	0.10	0.76	0.53	0.80	0.66	0.73	0.29
<i>PSNR</i>	0.90	0.63	0.80	0.50	0.46	0.95	0.62
<i>MSSIM</i>	0.96	0.93	0.86	0.96	0.90	0.91	0.78

As we can see, results reported in table. 1 concern the proposed measure as well as some FR and RR methods. In comparison with RR methods, the proposed measure outperforms the DNT-based methods for JPG, JP2 and DCQ distortions, and the method in [13] for JPG, JP2, DCQ, BLR and NOZ distortions. The proposed measure outperforms also the PSNR for JPG, JP2, DCQ and BLR distortions, and MSSIM [14] for JPG, DCQ and BLR distortions. However, the proposed measure fails for the FLT distortion.

6 Conclusion

In this paper we have introduced a RR measure in the tetrolet domain. The GSM model was used to characterize the dependencies between tetrolet coefficients. We have proposed the Weibull distribution to model the multiplier of the GSM model, this leads to a new joint distribution. Assuming the independence between GSM components we have derived a closed expression of the KLD for the propose joint distribution. Significant improvements were remarked for the proposed measure when it was tested on the Cornell VCL-A57 dataset.

References

1. Wang, Z., Simoncelli, E.P.: Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In: Proc. of SPIE Human Vision and Electronic Imaging, vol. 5666, pp. 149–159 (2005)

2. Tao, D., Li, X., Lu, W., Gao, X.: Reduced-reference IQA in contourlet domain. *IEEE Transactions on Systems, Man, and Cybernetics, PartB: Cybernetics* 39(6), 1623–1627 (2009)
3. Li, Q., Wang, Z.: Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing* 3(2), 202–211 (2009)
4. Xue, W., Mou, X.: Reduced reference image quality assessment based on weibull statistics. In: *The International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (2010)
5. Maalouf, A., Larabi, M.C., Fernandez-Maloigne, C.: A grouplet-based reduced reference image quality assessment. In: *The International Workshop on Quality of Multimedia Experience(QoMEX)*, pp. 59–63 (2009)
6. Krommweh, J.: Tetrolet transform: A new adaptive haar wavelet algorithm for sparse image representation. *Journal of Visual Communication and Image Representation* 21(4), 364–374 (2010)
7. Golomb, S.W.: *Polyominoes: puzzles, patterns, problems, and packings*. Princeton Univ. Pr. (1996)
8. Simoncelli, E.P.: Modeling the joint statistics of images in the wavelet domain. In: *Proc. SPIE*, vol. 3813, pp. 188–195 (1999)
9. Boubchir, L., Fadili, J.M.: Multivariate statistical modeling of images with the curvelet transform. In: *Proc. IEEE Conf. on Signal Processing and Its Applications*, pp. 747–750 (2005)
10. Wainwright, M.J., Simoncelli, E.P.: Scale mixtures of gaussians and the statistics of natural images. *Advances in Neural Information Processing Systems* 12(1), 855–861 (2000)
11. Chandler, D.M., Hemami, S.S.: Cornell-vcl a57 database (2007), <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>
12. Rohaly, A.M., Libert, J., Corriveau, P., Webster, A., et al.: Final report from the video quality experts group on the validation of objective models of video quality assessment, ITU-T Standards Contribution COM, pp. 9–80
13. Wang, Z., Wu, G., Sheikh, H.R., Simoncelli, E.P., Yang, E.H., Bovik, A.C.: Quality-aware images. *IEEE Transactions on Image Processing* 15(6), 1680–1689 (2006)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612 (2004)

Real Time Door Access Event Detection and Notification in a Reactive Smart Surveillance System

Gaetano Di Caterina, Nurulfajar Abd Manap,
Masrullizam Mat Ibrahim, and John J. Soraghan

Department of Electronic and Electrical Engineering, University of Strathclyde,
Royal College Building, 204 George Street, Glasgow, G1 1XW, UK
gaetano.di-caterina@strath.ac.uk

Abstract. The effectiveness of modern video surveillance systems critically depends on camera image quality and human operators' reactivity. In this paper we present a door access event detection application in the context of a reactive smart surveillance system, which automatically notifies in real time the occurrence of events to registered users, through SMS alerts. The system utilizes two fixed IP cameras and a high resolution PTZ camera to acquire high quality images of the face of people entering the room. System users can access a web-based interface to review the event details, along with a short video clip and the high quality face images acquired. Experimental results demonstrate that the final system allows the PTZ camera to automatically acquire high-resolution images of faces and deliver them to system operators in real time.

1 Introduction

CCTV is not always as effective as expected, due to two important issues, namely (i) image quality and (ii) reactivity of the surveillance personnel in spotting events of interest. To address these issues, digital video surveillance systems are beginning to incorporate megapixel IP cameras, which can deliver high quality images over IP networks, at high frame rate. Secondly, smart technologies can be used to analyze the video feeds and detect events of interest in real time for effective use. In smart surveillance systems [1], video analytics, which is the semantic analysis of video data through signal and image processing techniques, is used to extract and process only the relevant information, to reduce both processing time and storage space.

The main contribution of this paper is the incorporation of a door access event detection application in the context of the reactive smart surveillance system described in [2]. In particular the proposed system can automatically detect and record faces of people entering a room, and notify the door access events in real time to registered users, through SMS alerts. Two low resolution IP cameras are used to obtain the 3D location of the object of interest, which is the face of people entering the room, with stereo matching techniques [3]. Such positional information is passed to a high resolution pan-tilt-zoom (PTZ) camera to locate the detected face and acquire high quality images of it. The presented system builds on the work in [4]. However, the system in [4] only focused on static objects, while the surveillance system proposed in this paper is integrated within a door access monitoring framework, wherein it can detect and record

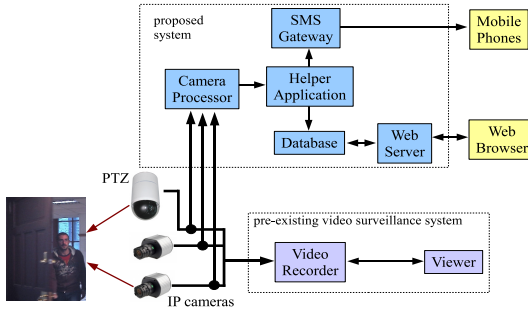


Fig. 1. System design with two IP cameras and a PTZ camera

moving objects, such as the faces of people entering the room. The remainder of the paper is organized as follows. Section 2 gives a brief overview of the system architecture. Section 3 provides a detailed description of the system camera processor. Section 4 contains experimental results and discussion, while section 5 concludes the paper.

2 System Architecture

A block diagram of the overall reactive smart surveillance system is depicted in Fig. 1. The system hardware includes two 1.3 megapixel Arecont AV1300 fixed IP cameras, and a 5 megapixel ACTi IP Speed Dome (CAM-6510) PTZ camera. The system software components are: one camera processor, which analyzes the input video feeds; a web server and associated database to store details of the detected events; a helper application which saves event data received from the camera processor into the database and sends SMS alerts to registered users. The camera processor is implemented in Matlab, Java and C, and it includes the video analytics algorithms, the PTZ controller and the event notification block. The two IP cameras are set up in a stereo configuration and have the door in their field of view. When the door opens, the IP cameras acquire real time images from two different angles. Such images are combined to produce stereo vision and compute the 3D location of the object of interest, i.e. the face of the person entering. This information is fed to the PTZ controller, which pans and tilts the PTZ camera to point at the targeted face and acquire a high resolution image of it. The door access event is also notified in real time to registered users, through SMS alerts.

3 Camera Processor Description

3.1 Door Open Detection

The main objective of this block is to detect in each new frame whether the door is open or closed. Door open detection is performed only on the left image, for simplicity. Since the two IP cameras are fixed, it is reasonable to select a region of interest (ROI) for the door in the $W \times H$ image, either manually or automatically [5], as shown in Fig. 2(a), with x_0 , x_1 , y_0 and y_1 being the horizontal and vertical coordinates of the ROI.

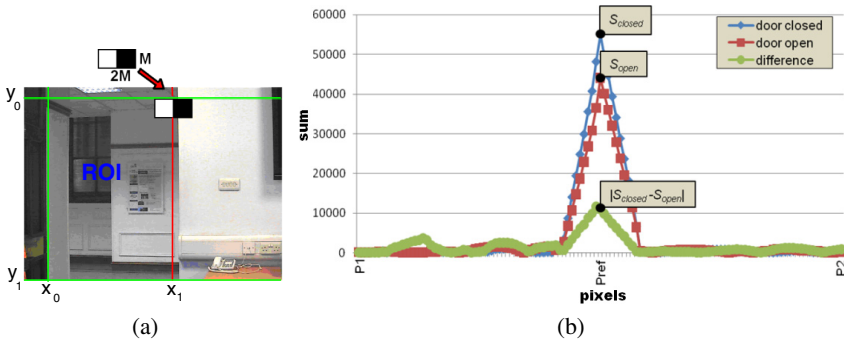


Fig. 2. Door open detection. (a) $2M \times M$ binary mask applied to the door image; (b) Behaviour of the sum S_i in both ‘door open’ and ‘door closed’ images.

The vertical side of the ROI where the hinges of the door are, is identified as ‘hinge side’, while the other vertical side is identified as ‘free side’. In order to detect whether the door is open in the i^{th} frame, a $2M \times M$ binary mask as in Fig. 2(a) is overlapped across the ‘free side’ at the top, in position $\mathbf{P}_{ref} = (x_1 - M, y_0)$, so that no object can ever occlude this part of the ROI. In usual video surveillance setups, cameras are mounted from the ceiling or at the very top of side walls, therefore the line of sight between camera and top edge of the door is never occluded. The pixel values in the binary mask are multiplied with the corresponding pixel values in the i^{th} frame and summed together to obtain the sum S_i at position \mathbf{P}_{ref} . As an experiment, if the binary mask scans the ‘door closed’ and ‘door open’ images horizontally, with its position going from $\mathbf{P}_1 = (x_1 - 3M, y_0)$ to $\mathbf{P}_2 = (x_1 + 3M, y_0)$, the graph in Fig. 2(b) is obtained. It is possible to see that in position \mathbf{P}_{ref} the sum S_i can assume two very different values S_{open} and S_{closed} , when the door is respectively open and closed. The only assumption here is that the door, the wall beside it and the background behind it do not all have the same colour. This suggests that a threshold χ can be set as $\chi = |S_{open} - S_{closed}|/2$. For the i^{th} frame, S_i is computed and if $|S_i - S_{closed}| > \chi$, then the door is considered to be open and the face detection algorithm is run. The presented door open detector is simple and fast and it can be seen as an improved motion detector that works on the underlying image structure: in a conventional motion detector, the pixel-wise difference between frames is thresholded to detect motion; in the proposed detector, the strength of the vertical edge on the door ‘free side’ is analyzed instead. Therefore, while a conventional motion detector could also be triggered by shadows and light changes, the presented door open detector is triggered only when the door is actually open, i.e. the strength of the vertical edge on its ‘free side’ varies.

3.2 Face Detection

There are four stages in the face detection step: skin colour segmentation, morphological processing, bounding rectangle forming and SVM classification. The obvious advantages of skin colour segmentation are fast processing and high robustness to geometric variation of head pose and orientation. For this purpose three colour spaces have

been employed: RGB, YCbCr and HSV. These three colour spaces are widely used in skin detection research [6, 7, 8, 9]. RGB is the most used one, although it is not very robust to light changes. Therefore Kovac *et al.* [7] used gray world method as an adaptation technique, to correct the images before applying skin detection. To adapt to light changes, Pai *et al.* [8] modulated the range of YCbCr skin colour distribution. The last colour space, i.e. HSV, represents colours in terms of depth, purity and brightness [6, 9]. From these three colour spaces, a combination rule for segmentation is formulated as in (1), to overcome sensitivity to illumination changes, ethnicity skin colour and different characteristics of cameras.

$$\begin{aligned}
 & \text{if } (r > 95 \wedge g > 40 \wedge b > 20) \\
 & \quad \wedge ((\max(r, g, b) - \min(r, g, b)) > 15) \\
 & \quad \wedge (|r - g| > 15) \wedge (r > g) \wedge (r > b) \\
 & \quad \wedge (140 < c_b < 195) \wedge (140 < c_r < 165) \\
 & \quad \wedge (0.01 < hue < 0.1) \\
 & \text{then } (\textit{selected pixel is skin})
 \end{aligned} \tag{1}$$

To obtain well segmented skin regions, mathematical morphology is used to remove noise and fill small holes. Bounding rectangles are formed by using the connected components labeling operator. Each bounding rectangle is then examined in terms of size and pattern. The pattern shape describes whether the rectangle bounds a face or a non-face object, and it is measured by the width-to-height ratio of the rectangle defined as:

$$0.83 < \frac{\textit{width}}{\textit{height}} < 1.27 \tag{2}$$

The range values in (2) are obtained from experiments carried out on 98 images containing 561 faces. Fig. 3(a) shows example of experimental results after skin colour segmentation and rectangle bounding formation. In these images, the rectangles bound all the regions segmented as skin. Rectangles that are too small or do not comply with (2) are discarded, as in Fig. 3(b). The remaining bounding rectangles are then classified as whether containing face or non face by using SVM on horizontal projection features. The horizontal projection of a face has a distinctive pattern, which is used as features for SVM training and classification. Fig. 4 shows three different poses of face, with horizontal projection profiles of eyes, nose and mouth. The values of peak and valley projected by the horizontal profile are used as features to differentiate between face and non-face objects. For this purpose, the image regions included in the remaining bounding rectangles are converted to gray scale. However, due to noise, such regions project an indistinctive horizontal graph projection, from which it is difficult to extract features. Therefore a Gaussian filter is employed to smoothen such face candidate regions. These smoothened regions are finally classified using SVM. Face regions in output from the SVM classifier in the left image are then processed in the stereo matching step, to find their corresponding regions in the right image.

3.3 Stereo Matching and 3D Location Estimation

Stereo matching determines which parts of the left and right images correspond to the same scene element. The central block from the detected human face in the left image



Fig. 3. Experimental results after skin colour segmentation and rectangle bounding formation

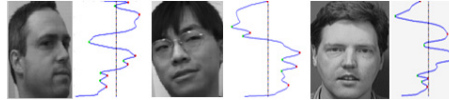


Fig. 4. The horizontal projection profile of three faces

is taken as a reference and compared with blocks in a search area in the right image. The block size is constrained to $\psi \times \psi$ pixels, while the size of the search area is of $\xi \times \xi$ pixels. The actual values of ψ and ξ depend on the application and also on the stereo camera setup. In the proposed system these values are $\psi = 32$ and $\xi = 128$. The matching between blocks in the left and right cameras is determined by the value of a cost function. Here, any matching measure could be used; however for low computation, we use the Sum of Absolute Differences (SAD). Minimizing the SAD measure gives the position in the right image of the best match for the reference block selected in the left image. To calculate the accurate 3D location of the detected human face, basic geometry rules are used. The simplest geometry of stereo video system consists of two parallel cameras with horizontal displacement, i.e. along the X axes, as shown in Fig. 5. Such geometry is derived from the pinhole camera model [10] and the same horizontal line is referred to as epipolar line. The symbol f is the focal length of the camera lens and B is the baseline distance, i.e. the distance between the two camera optical centres. If $\mathbf{O}_L = (U_L, V_L)$ and $\mathbf{O}_R = (U_R, V_R)$ are the projections in the left and right images, relative to the respective camera centre points, of the 3D point \mathbf{P}_L , as illustrated in Fig. 5 it holds $V_L = V_R$ and the disparity of the stereo images is obtained as difference between U_L and U_R :

$$d = U_L - U_R = \left(f \frac{x_L}{z_P} - f \frac{x_R}{z_P} \right) = \left(f \frac{x_L}{z_P} - f \frac{x_L - B}{z_P} \right) \quad (3)$$

The location of correct projections of the same point \mathbf{P} on the two image planes can determine the exact depth of \mathbf{P} in the real world. From (3), the depth z_P of the point \mathbf{P} is computed as $z_P = (fB)/d$. Therefore, the equations used to calculate the exact location $\mathbf{P} = (x_P, y_P, z_P)$ of the target object are:

$$x_P = \frac{Bx_L}{d}, \quad y_P = \frac{By_L}{d}, \quad z_P = \frac{Bf}{d} \quad (4)$$

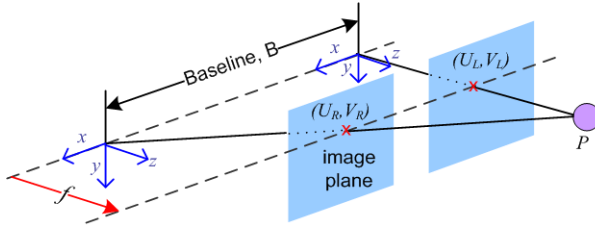


Fig. 5. Stereo camera configuration

3.4 PTZ Controller

The PTZ controller module deals with the PTZ hardware, firmware and communication protocols. First, it applies a homogeneous transformation to compute the 3D location $\mathbf{P}_{PTZ} = (x_{PTZ}, y_{PTZ}, z_{PTZ})$ of the target with respect to the PTZ. If \mathbf{T} is a transformation matrix that transforms from the stereo cameras coordinate frame to the PTZ coordinate frame, the location \mathbf{P}_{PTZ} is computed as:

$$[x_{PTZ}, y_{PTZ}, z_{PTZ}, 1]^T = \mathbf{T} [x_P, y_P, z_P, 1]^T \tag{5}$$

The PTZ controller converts the target location \mathbf{P}_{PTZ} into pan and tilt angles, and zoom factor for the PTZ. These values are incorporated into commands for the PTZ, in the form of standard HTTP requests, over the network. The panning angle θ and the tilting angle β are calculated as:

$$\theta = \tan^{-1} \left(\frac{z_{PTZ}}{D - x_{PTZ}} \right) \tag{6}$$

$$\beta = \tan^{-1} \left(\frac{y_{PTZ}}{\sqrt{(D - x_{PTZ})^2 + z_{PTZ}^2}} \right) \tag{7}$$

where D is the distance between IP cameras and PTZ along the X axes. The zoom ratio instead is proportional to the Euclidean distance between PTZ camera and target object.

3.5 Event Notification

When the door is detected as open as described in section 3.1, a timer is started and after 10s a door access event is triggered. At this point a low frame rate (2 – 5fps) video clip of the past 10s is created and asynchronously sent to the helper application, along with event details, such as time, date and camera ID. The helper application saves the event data in the web server database and issues an SMS alert to a list of pre-registered users, who can access the remote interface, to review event details and short video clip in real time, along with the high resolution face images recorded by the PTZ camera. The time delay before triggering a door access event is needed to make sure that the short video clip includes also images of the actual person entering the room. Within such interval, no other events are triggered. This is to prevent events from being triggered at every frame. However, if the door stays open for more than 10s, the timer is started again and a new event is triggered when the timer expires again.

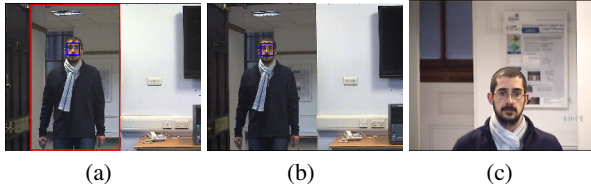


Fig. 6. Visual results. (a) left image; (b) right image; (c) high resolution image from the PTZ.

Table 1. Mean and standard deviation of absolute differences

<i>Axes</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
μ	0.047m	0.099m	0.357m
σ	0.027m	0.011m	0.077m

Table 2. Average execution times

<i>Operation</i>	μ	σ
Acquisition	0.090s	0.004s
Face detection	0.032s	0.002s
Stereo matching	0.028s	0.001s
Location estimation	0.001s	0.000s

4 Results and Discussion

Fig. 6 shows results of the face detection and high resolution face image acquisition. Fig. 6(a) and (b) are the left and right camera views and they are in the same epipolar line. The searching area is minimized to the door mask region only, instead of all the pixel images. With this approach, the execution of stereo matching and face detection is faster. The distance between IP cameras and the target in Fig. 6 is 4m. The face detection algorithm was tested with the CMU face colour images database [11], which contains a variety of faces in normal room lighting conditions. 346 face images with a variety of skin colour tones and different facial poses were used. The face detection described in this paper correctly detected human faces in 327 images (94.5%), with 19 images (5.5%) erroneously detected. The main cause of the errors was due to pieces of clothing classified as skin.

The face detection result is processed in the block matching and 3D location estimation steps, to obtain the depth and location of the targeted object. With this information, the coordinates of the object are calculated and transmitted to PTZ camera controller. The coordinates are converted into pan and tilt angles for the PTZ. The PTZ camera captures the targeted object as shown in Fig. 6(c), where the distance between object and PTZ camera is calculated as 8.13m. The object detected with the PTZ can be tracked and images of it are recorded automatically. The system has been developed and tested using different test vectors, by placing the cameras at different locations with respect to the PTZ, and with different people as target. The PTZ response upon changes of the coordinates has been found to be quick.

For the location estimation test, the system is fed with the 22 sets of stereo images, to evaluate the accuracy of the target location estimated by the proposed system, with respect to the exact target location in the 3D space. The error between each set of estimated and exact locations is computed as Euclidean distance. Table 1 shows means

μ and standard deviations σ of the absolute differences between exact and estimated values, for each coordinate axis. The error in X and Y coordinates are very small, while the error in Z coordinate is slightly higher.

The mean and standard deviation profile of the recorded execution times are presented in Table 2. The results show that face detection, stereo matching and location estimation steps accounts for less than 50% of the total execution time of 133ms. The high image acquisition time is due to the transmission of both left and right images over the network, from the IP cameras. The average frame rate is about 8fps. It is expected that an implementation on a dedicated DSP board would significantly speed up the total execution time.

5 Conclusion

A fully automated reactive smart surveillance system using stereo images has been designed and developed. It automatically detects door access events and uses multiple cameras to localize and zoom in on the faces of people entering the room, to acquire high resolution images of them. System features include door detection, face detection, high quality face image acquisition and real time notification to registered users. The overall system makes extensive use of IP technologies, to ensure communication among components and remote availability of the system resources, such as IP cameras, event database and user front-end. Despite its simplicity, the proposed system performs well and it is suitable for real time execution. As future work, the presented video analytics algorithms will be ported to a DSP board for fast ‘in-camera’ processing.

References

1. Valera, A., Velastin, S.A.: Intelligent distributed surveillance systems: a review. *IEE Proceedings - Vision, Image and Signal Processing* 152, 192–204 (2005)
2. Di Caterina, G., Soraghan, J.J.: An abandoned and removed object detection algorithm in a reactive smart surveillance system. In: *DSP 2011* (2011)
3. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal of Computer Vision* 47, 7–42 (2002)
4. Manap, N.A., Di Caterina, G., Soraghan, J.J., Sidharth, V., Yao, H.: Smart surveillance system based on stereo matching algorithms with IP and PTZ cameras. In: *3DTV-Con 2010*, pp. 1–4 (2010)
5. Yang, X., Tian, Y.: Robust door detection in unfamiliar environments by combining edge and corner features. In: *IEEE CVPR Workshops*, pp. 57–64 (2010)
6. Chaves-Gonzalez, J.M., Vega-Rodriguez, M.A., Gomez-Pulido, J.A., Sanchez-Perez, J.M.: Detecting skin in face recognition systems: a colour spaces study. *Digital Signal Processing* 20, 806–823 (2010)
7. Shovic, J., Peer, P., Solina, F.: Human skin color clustering for face detection. In: *IEUROCON 2003 - Computer as a Tool*, pp. 144–148 (2003)
8. Pai, Y.T., Ruan, S.J., Shie, M.C., Liu, Y.C.: A simple and accurate color face detection algorithm in complex background. In: *IEEE ICME*, pp. 1545–1548 (2006)

9. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. *Pattern Recognition* 40, 1106–1122 (2007)
10. Bovik, A.: *Handbook of image and video processing*, 2nd edn. Elsevier, Academic Press (2005)
11. CMU: Image database: face,
http://vasc.ri.cmu.edu/idb/html/face/frontal_images/

Optical Flow Estimation on Omnidirectional Images: An Adapted Phase Based Method

Brahim Alibouch¹, Amina Radgui^{1,2},
Mohammed Rziza¹, and Driss Aboutajdine¹

¹ LRIT associated unit with CNRST (URAC29), Mohammed V-Agdal University,
B.P 1014, Rabat, Morocco

alibouch_brahim@yahoo.fr, rziza@fsr.ac.ma, aboutaj@fsr.ac.ma

² INPT, Madinat AL Irfane, Rabat, Morocco
radgui@inpt.ac.ma

Abstract. Omnidirectional vision is one of emerging areas of research. Omnidirectional images offer a large field of view compared to conventional perspective images. However, these images contain important distortions, and classical optical flow estimation are thus not appropriate. In this paper, we propose to estimate optical flow on omnidirectional images using a phase based method which proved its robustness and its accuracy on the perspective images. We will adapt different treatments that this method involve in order to take into account the nature of omnidirectional images.

Keywords: optical flow, omnidirectional vision, phase based methods, component velocity, Gabor filters.

1 Introduction

A fundamental problem in images processing is the computation of optical flow [1]. Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image [2]. The information given by the optical flow can be used in many applications [3] such as object detection and tracking [4][5], robot navigation [6], video surveillance [7], ego-motion estimation [8] or visual odometry [9]. To estimate the optical flow there are several methods. A selection of those methods was tested and compared in [10] and grouped in four different classes: differential methods [11][12][2], phase-based methods [13][14], region-based methods [15] and energy based methods [16][17]. In optical flow estimation, Phase based methods are among techniques which proved their robustness and their accuracy [10]. Those techniques were introduced the first time by Fleet and Jepson [13]. Their method is based on the assumption that the level contours of constant phase provide a good approximation to the motion field [13]. They propose to use spatiotemporal filters to decompose the image sequence according to scale and orientation [10], and then normal components of 2D velocity are calculated at each location in the different filters outputs. Finally, the full velocity is estimated by integrating all reliable normal components.

Based on this approach, Gautama et al [18] introduced a new technique based on spatial filters instead of spatiotemporal ones. They consider phase nonlinearity as a criterion of reliability instead instability [18].

In this paper, we will adapt this last approach to omnidirectional images. The remainder of the paper is as follows: in the next section we present the phase-based approach proposed by Gautama et al to estimate the optical flow when using perspective images. Then, in section 3, we describe how to adapt this approach to estimate optical flow on omnidirectional images. Section 4 shows experiment results. We present our conclusions in section 5.

2 Phase Based Method for Optical Flow Estimation

The phase-based technique proposed by Gautama et al [18] uses a set of 2D complex filters to extract spatial phase. Then, temporal phase gradient is estimated at every position in image sequence and a reliability measure is applied to determine valid component velocities. These component velocities are thereafter combined to generate the optical flow field.

2.1 Filters Setting

To extract the phase in [18] Gabor filters are used to proceed to the multichannel decomposition. Gabor filter's impulse response is given by :

$$G(\mathbf{x}) = \frac{1}{2\pi\sigma} e^{-\frac{|\mathbf{x}|^2}{\sigma^2}} e^{i2\pi f} \quad (1)$$

With $\mathbf{x} = (x, y)$ is the pixel position, $f = (f_x, f_y)$ are center frequencies which define filter orientation θ , and σ is the standard deviation of the elliptical Gaussian which defines scale parameter. Once an image $I(\mathbf{x})$ is filtered by such filter, the response is given by:

$$\begin{aligned} R(\mathbf{x}) &= I(\mathbf{x}) * G(\mathbf{x}) \\ &= \rho(\mathbf{x}) e^{i\phi(\mathbf{x})} \end{aligned} \quad (2)$$

$\rho(\mathbf{x})$ and $\phi(\mathbf{x})$ are respectively the amplitude and the phase component of the image convolved with the Gabor filter.

2.2 Optical Flow Estimation

Starting from the hypothesis that surfaces of constant phase provides a good approximation to the motion field [13], we can deduce the phase gradient constraint equation. Indeed, such surfaces satisfy:

$$\phi(\mathbf{x}, t) = c \quad (3)$$

Differentiating this equation with respect to t yields:

$$\nabla\phi \cdot \mathbf{V} + \frac{\partial\phi}{\partial t} = 0 \quad (4)$$

Where $\nabla\phi$ is the spatial phase gradient, $\frac{\partial\phi}{\partial t}$ the temporal phase gradient and $V = (v_x, v_y)$ is the velocity vector. As in the brightness constancy equation, the aperture problem appears also in the phase gradient constraint equation. In fact, we can estimate only the velocity component in the direction of the spatial phase gradient V_c . Equation (4) yields :

$$(\nabla\phi \cdot V) \frac{\nabla\phi}{|\nabla\phi|} = -\frac{\partial\phi}{\partial t} \frac{\nabla\phi}{|\nabla\phi|} \quad (5)$$

Given that:

$$V_c = (V \cdot \frac{\nabla\phi}{|\nabla\phi|}) \frac{\nabla\phi}{|\nabla\phi|} \quad (6)$$

This gives:

$$V \cdot \nabla\phi = \frac{V_c}{|\nabla\phi|} |\nabla\phi|^2 \quad (7)$$

Upon substituting equation (7), equation (5) become :

$$V_c = -\frac{\partial\phi}{\partial t} \frac{\nabla\phi}{|\nabla\phi|^2} \quad (8)$$

The spatial phase gradient $\nabla\phi = (\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y})$ can be substituted with the local instantaneous frequency $(2\pi f_x, 2\pi f_y)$ [19] :

$$V_c(x, y) = -\frac{\partial\phi}{\partial t} \frac{1}{2\pi(f_x^2 + f_y^2)} (f_x, f_y) \quad (9)$$

The temporal phase gradient $\frac{\partial\phi}{\partial t}$ is obtained from the temporal evolution of the phase by a accomplishing a linear regression in the least-squares sense [19] [20] on the next equation:

$$\phi(x, t) = c + \frac{\partial\phi}{\partial t} t \quad (10)$$

Note that the phase is unwrapped along the image sequence to deal with the phase periodicity. To determine the reliability of each component velocity, we calculate the mean squared error:

$$MSE = \frac{\sum_t (\Delta\phi(x, t))^2}{N} \quad (11)$$

Where N is the number of images and $\Delta\phi(x, t) = (c + \frac{\partial\phi}{\partial t}(x, t) \cdot t) - \phi(x, t)$

Thereafter, valid component velocities are combined to estimate the full velocity:

$$V^*(x) = \arg \min \sum \left(\|V_{c,i}(x)\| - V(x, t)^T \frac{V_{c,i}(x)}{\|V_{c,i}(x)\|} \right)^2 \quad (12)$$

Where $V_{c,i}$ is the component velocity at pixel x corresponding to the i^{th} filter.

3 Optical Flow in Omnidirectional Images

Omnidirectional images offer a large field of view compared to conventional perspectives images, although they are distorted due to the non-linear projection of the scene points in the image [3]. Consequently calculating optical flow on such images in the same way as on perspectives images will lead to mistaken results. One of the most used techniques to avoid this problem is to project omnidirectional images on the sphere and using image processing in that new domain.

3.1 Projection on the Sphere

The equivalence between the catadioptric projection and the projection on the sphere has been proved by Geyer and Daniilidis [21]. In their work, they have presented a unifying theory for central panoramic systems. This equivalence is shown in Fig. 1.

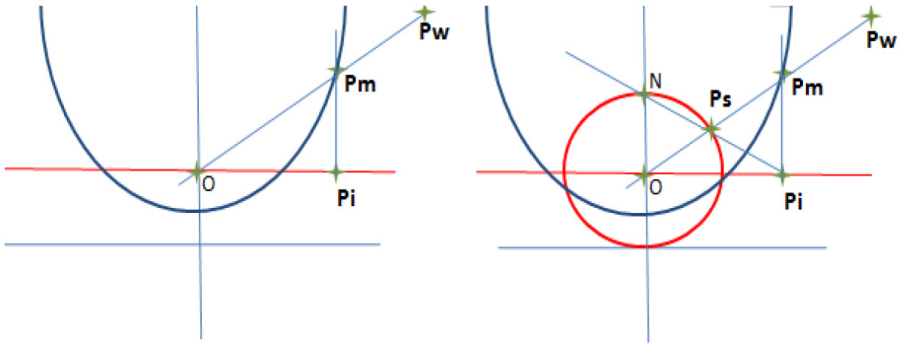


Fig. 1. Equivalence between the catadioptric projection and the two-step mapping via the sphere

The 3D point $P_w(X_w, Y_w, Z_w)$ is first projected in the mirror on a point $P_m(X_m, Y_m, Z_m)$, then reflected to the image plane on a point $P_i(x, y)$, such that it is parallel to the optical axis.

Let $P_s(X_s, Y_s, Z_s) = P_s(\theta, \varphi)$ be the equivalent point on the unit sphere. The Cartesian coordinates of this point are given by :

$$\begin{cases} X_s = \sin \theta \cos \varphi \\ Y_s = \sin \theta \sin \varphi \\ Z_s = \cos \theta \end{cases} \tag{13}$$

The stereographic projection of P_s on the image plane yields point $P_i(x, y)$ given by:

$$\begin{cases} x = \frac{X_s}{1-Z_s} \\ y = \frac{Y_s}{1-Z_s} \end{cases} \tag{14}$$

By combining Equations (12) and (13) we obtain the spherical coordinates of point P_i :

$$\begin{cases} x = \cot \frac{\theta}{2} \cos \varphi \\ y = \cot \frac{\theta}{2} \sin \varphi \end{cases} \quad (15)$$

3.2 Optical Flow on the Sphere

To adapt the phase based method to omnidirectional images, we need to reformulate the phase gradient constraint equation in the sphere. Let $\phi_s(\theta, \varphi)$ be the spherical phase in the unit sphere, and $\nabla\phi_s = \left(\frac{\partial\phi_s}{\partial\theta}, \frac{1}{\sin\theta} \frac{\partial\phi_s}{\partial\varphi} \right)$ the spatial phase gradient on the sphere, the phase gradient constraint given in equation (4) becomes :

$$\frac{1}{\sin\theta} \frac{\partial\phi_s}{\partial\varphi} V_\varphi + \frac{\partial\phi_s}{\partial\theta} V_\theta + \frac{\partial\phi_s}{\partial t} = 0 \quad (16)$$

Where (V_θ, V_φ) are the components of the flow vector in the tangential coordinates system. As for perspective images this equation provides only normal velocity component:

$$V_c(\theta, \varphi) = -\frac{\partial\phi_s}{\partial t} \frac{\nabla\phi_s}{|\nabla\phi_s|^2} \quad (17)$$

4 Experiment and Results

To test our approach we use real sequences of omnidirectional images, and we compared it to the classical phase based method proposed by Gautama [18]. To extract phase we used a filterbank consisting of spherical Morlet wavelets [22], tuned at the same orientations as in Gautama method. The sequences are captured using a catadioptric camera embedded on a mobile robot as shown in Fig. 2.

The resolution of our images is 1280*960 pixel and the intrinsic parameters are: $\alpha_u = 243$, $\alpha_v = 236$ and $h = 0.86$.

We estimate the optical flow for two different motions kinds : a rotation of the camera around the Z-axis as shown in Fig. 3 , and object movement in the scene with a fixed camera as shown in Fig. 4. Since in the case of real images we do not have the ground truth, we will just present the 2D motion fields that illustrated the amelioration given by our adapted method.

In Fig. 3, the image on the bottom left represents the optical flow obtained by applying Gautama approach without adaptation on the omnidirectional sequence corresponding to rotation. Overall, the optical flow field is correct with some minor irregularities. The image on the bottom right represents the optical flow obtained by applying our adapted method. This optical flow field is much better and more regular than the first one.



Fig. 2. Top: omnidirectional sensor embedded on a mobile robot. Bottom: omnidirectional image.

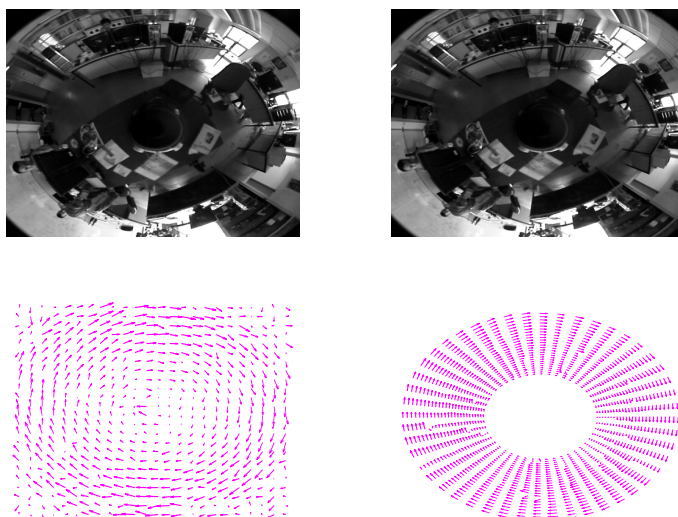


Fig. 3. Top: a sequence depicting a rotation. Bottom: optical flow obtained using classical Gautama method (Left) and using our approach (right).

Fig. 4 shows, on the bottom left, the optical flow obtained by applying Gautama approach without adaptation on the omnidirectional sequence corresponding to the object movement. This image shows an optical flow field who doesn't reflect the real motion on the sequence, and therefore a wrong one. On the other side the optical flow obtained by applying our adapted method is much closer to the real movement in the left of scene.

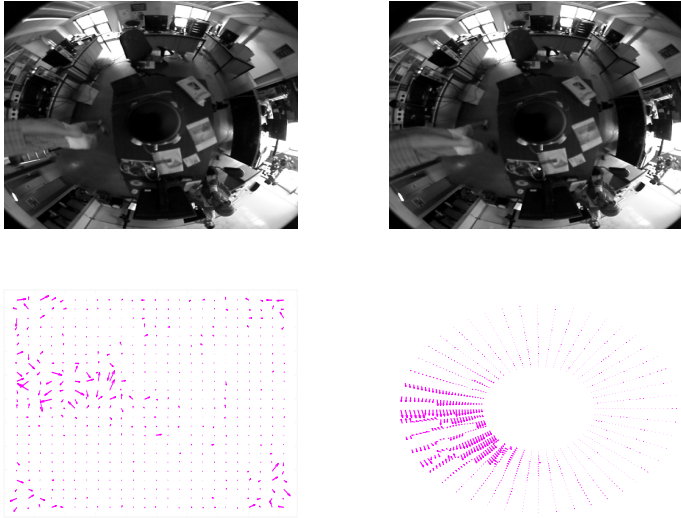


Fig. 4. Top: a sequence depicting an object movement. Bottom: optical flow obtained using classical Gautama method (Left) and using our approach (right).

5 Conclusion

Omnidirectional images are rich in information since they depict almost the whole scene. Unfortunately, they include severe distortions. That is why classical methods used to estimate the optical flow that work for perspectives images need to be adapted for omnidirectional ones. In this paper we have proposed an adaptation to a phase based method proposed by Gautama [18] which is one of the most robust optical flow methods. We applied our approach in real images and we compared it to the classical Gautama method. The comparison shows that our adapted Gautama method provides a correct local motion fields.

References

1. Beauchemin, S.S., Barron, J.L.: The Computation of Optical Flow. *ACM Comput. Surv.* 27, 433–467 (2003)
2. Horn, B., Schunck, B.: Determining optical flow. *Artificial Intelligence* 17, 185–203 (1981)

3. Radgui, A., Demonceaux, C., Mouaddib, E., Rziza, M., Aboutajdine, D.: Optical flow estimation from multichannel spherical image decomposition. *Computer Vision and Image Understanding* 115, 1263–1272 (2011)
4. Kim, J., Suga, Y.: An omnidirectional vision-based moving obstacle detection in mobile robot. *International Journal of Control Automation and Systems* 5, 663–673 (2007)
5. Yoshizaki, W., Mochizuki, Y., Ohnishi, N., Imiya, A.: Catadioptric omnidirectional images for visual navigation using optical flow. In: *OMNIVIS 2008* (2008)
6. Winters, N., Gaspar, J., Lacey, G., Santos-Victor, J.: Omni-directional vision for robot navigation. In: *IEEE Workshop on Omnidirectional Vision*, pp. 21–28 (2000)
7. Wang, M.L., Huang, C.C., Lin, H.Y.: An intelligent surveillance system based on an omnidirectional vision sensor. In: *IEEE Conference on Cybernetics and Intelligent Systems*, pp. 1–6 (2006)
8. Gluckman, J., Nayar, S.: Ego-motion and omnidirectional cameras. In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 999–1005 (1998)
9. Bunschoten, R., Krose, B.: Visual odometry from an omnidirectional vision system. In: *IEEE International Conference on Robotics and Automation (ICRA 2003)*, vol. 1, pp. 577–583 (2003)
10. Barron, J.L., Fleet, D.J., Beauchemin, S.: Performance of optical flow techniques. *Int. J. Comput. Vis.* 12, 43–77 (1994)
11. Kanade, T., Lucas, B.: An iterative image registration technique with an application to stereo vision. In: *IJCAI 1981*, pp. 674–679 (1981)
12. Nagel, H.H.: On a constraint equation for the estimation of displacement rates in image sequences. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 11, 13–30 (1989)
13. Fleet, D.J., Jepson, A.D.: Computation of component image velocity from local phase information. *Int. J. Comput. Vis.* 5, 77–104 (1990)
14. Tsao, T., Chen, V.: A neural scheme for optical flow computation based on Gabor filters and generalized gradient method. *Neurocomputing* 6, 305–325 (1994)
15. Anandan, P.: A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision* 2, 283–310 (1989)
16. Adelson, E., Bergen, J.: Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America* 2, 284–299 (1985)
17. Heeger, D.: Optical flow using spatiotemporal filters. *International Journal of Computer Vision* 1, 279–302 (1988)
18. Gautama, T., Van Hulle, M.M.: A phase-based approach to the estimation of the optical flow field using spatial filtering. *IEEE Trans. Neural Networks* 13, 1127–1136 (2002)
19. Pauwels, K., Van Hulle, M.M.: Optic Flow from Unstable Sequences containing Unconstrained Scenes through Local Velocity Constancy Maximization. In: *BMVC*, pp. 397–406 (2006)
20. Pauwels, K., Van Hulle, M.M.: Realtime phase-based optical flow on the GPU. In: *Computer Vision and Pattern Recognition Workshops*, pp. 1–8 (2008)
21. Geyer, C., Daniilidis, K.: Catadioptric projective geometry. *Int. J. Comput. Vis.* 43, 223–243 (2001)
22. Demanet, L., Vandergheynst, P.: Gabor wavelets on the sphere. In: *SPIE Conference on Wavelets: Applications in Signal and Image Processing* (2003)

DWT Based-Approach for Color Image Compression Using Genetic Algorithm

Aldjia Boucetta¹ and Kamal Eddine Melkemi²

¹ Department of Computer Science, Faculty of Science, University of Batna, 05000 Batna, Algeria

boucetta_batna@yahoo.fr

² Department of Computer Science, Faculty of Science, University of Biskra, 07000 Biskra, Algeria

melkemi2002@yahoo.com

Abstract. This paper describes a color image compression technique based on Discrete Wavelet Transform (DWT) and Genetic Algorithm (GA). High degree of correlation between the RGB planes of a color image is reduced by transforming them to more suitable space by using the GA. This GA would enable us to find $T_1T_2T_3$ representation, in which T_1 energy is more maximized than that of T_2 and T_3 .

The result of the proposed method is compared with previous similar published methods and the former is found superior in terms of quality of the reconstructed image.

Further, proposed method is efficient in compression ability and fast in implementation.

Keywords: Color image compression, Color space, Discrete wavelet transform, Arithmetic encoder, Two-role encoder, Genetic algorithm.

1 Introduction

Compression/coding of digital image is done by detecting and removing redundant information from the image. Image compression algorithm consists of two basic categories:

Methods of the first category are called direct image compression [1], [2] methods which are applied directly on the samples of an image in the spatial domain. Block Truncation Coding (BTC) and vector quantization are two widely used spatial domain compression techniques [3].

The second category contains methods called transform methods [1], [7], which transform the image to frequency representations suitable for detecting and removing redundancies, such as Discrete Fourier Transform (DFT) [1], Discrete Cosine Transform (DCT) [2], [8] and Discrete Wavelet transform (DWT) [12].

Of all the transform methods, the wavelet transform achieves better energy compaction than the DCT and hence can help in providing better compression for the same Peak Signal to Noise Ratio (PSNR).

A comparative study of DCT and wavelet based image coding can be found in [2].

In this paper, we propose a new color image compression method based on DWT and an appropriate GA [13], [14]. The RGB system color representation is the most commonly used in computer graphics. In fact, there are an infinite number of possible color spaces instead of this common RGB channels. Many of these other color spaces are derived by applying linear functions of R, G, B.

Recently, Douak et al. [4] proposed a color image compression algorithm based on the DCT transform and the RGB to YCbCr transformation. However, in our proposed approach we move from the RGB space to more suitable space for each image, by using an appropriate GA. This suitable space is referred to as $T_1T_2T_3$.

Indeed, our GA would enable us to find these $T_1T_2T_3$ color space, in which T_1 energy is more maximized than that of T_2 and T_3 . This allows a more effective compression because the information is condensed in the plan T_1 . Thus, compress T_2 and T_3 more effectively .

In the remaining of this paper, the proposed method is referred to as GA-DWT based compression approach.

The rest of this paper is organized as follows.

Section 2 presents fundamental and methodological concepts needed in this work, and describes the performance criteria used to elaborate the GA-DWT based compression approach. Section 3 gives more details to explain the GA-DWT based compression approach. Section 4 presents and discusses some experimental results. Section 5 gives a general conclusion and some ideas for future research.

2 Basic Concepts

2.1 Genetic Algorithm

A GA (see [13] and [14]) is a probabilistic research algorithm that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to search problems such as image compression [15], [16]. GAs, which generate solutions to optimization problems using methods inspired by inheritance, mutation, selection, and crossover.

In this paper, using the GA to find the $T_1T_2T_3$ color space, bearing in mind that T_1 represents the luminance; T_2 and T_3 represent the chrominance as:

$$\begin{aligned} T_1 &= a_{11} \times R + a_{12} \times G + a_{13} \times B . \\ T_2 &= a_{21} \times R + a_{22} \times G + a_{23} \times B . \\ T_3 &= a_{31} \times R + a_{32} \times G + a_{33} \times B . \end{aligned} \tag{1}$$

To solve the problem, we must find $a = a_{ij}$ that maximizes the energy in T_1 than that of the two other channels T_2 and T_3 .

GA processes:

Figure 1 shows the GA scheme used in this approach.

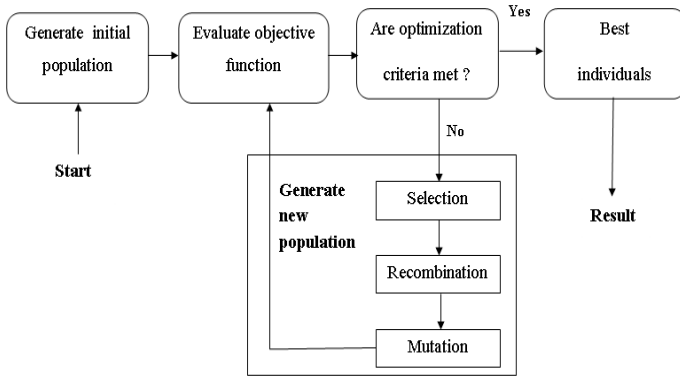


Fig. 1. The GA scheme

1. Generate Initial Population: create real-valued initial population of chromosomes. A chromosome in our algorithm is constituted by nine genes and each gene is encoded as a real number. Figure 2 shows our chromosome representation.



Fig. 2. The chromosome codification

2. Objective function: is used to calculate the effectiveness of each chromosome. For a more complete review, see [15].

We define T_1SE , T_2SE , and T_3SE , respectively, as T_1 , T_2 , T_3 space energy and $TE_{T_1T_2T_3}$ express the total energy. Their definitions are:

$$T_1SE = 100 \times \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T_{1ij}^2}{TE_{T_1T_2T_3}} \quad (2)$$

$$T_2SE = 100 \times \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T_{2ij}^2}{TE_{T_1T_2T_3}} \quad (3)$$

$$T_3SE = 100 \times \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T_{3ij}^2}{TE_{T_1T_2T_3}} \quad (4)$$

$$TE_{T_1T_2T_3} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T_1^2 + \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T_2^2 + \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} T_3^2 \quad (5)$$

$$f = T_1SE - (T_2SE + T_3SE) . \quad (6)$$

The problem is to maximize f .

3. Genetic Operators:

- (a) **Recombination (crossover):** The crossover is the operator that exchanges genetic material between two individuals by selecting a point at which pieces of the parents are swapped in order to generate two new individuals [15]. In our implementation we use high-level recombination operator (recombin) [19].
- (b) **Mutation:** Mutation operator modifies the chromosome genes randomly according to the mutation probability. We use real-value mutation (mutbga) [19].
- (c) **The parameters of the algorithm:** The behavior of the GA can be controlled using many initial conditions and parameters. The various parameters of GA are shown in Table 1.

Table 1. Genetic algorithm parameters

Parameter	Value
Population Size	50
Maximum generations	20
Crossover probability	0.8
Mutation probability	0.1

2.2 Discrete Wavelet Transform

The DWT (see [10] and [8]) is applied independently to the image components and decorrelates the image into different scale sizes, preserving much of its spatial correlation. A one-dimensional (1-D) DWT consists of a low (L) and high (H) pass filter splitting a line of pixels into two lines of half the size. Application of the filters to two-dimensional (2-D) images in horizontal and vertical directions produces four subbands (LL, LH, HL, and HH). The LL subband is a lower resolution representation of the original image, and the missing details are filtered into the remaining subbands. The subbands contain the horizontal (LH), vertical (HL), and diagonal (HH) edges on the scale size defined by the wavelet.

2.3 Performances Criterion

The performances of compression technique are based on two widely used essential criteria, the compression ratio, and the quality of the reconstructed image. Here, compression ratio is measured in terms of Bits Per Pixel (bpp) and the image quality in terms of PSNR [3]. The bpp is given by:

$$\text{bpp} = \frac{\text{size of compressed image in bits}}{\text{number of pixels}} . \quad (7)$$

The PSNR is given by [3]:

$$PSNR = 10 \times \log_{10} \left(\frac{255^2 \times 3}{MSE(R) + MSE(G) + MSE(B)} \right) . \quad (8)$$

3 GA-DWT Based Compression Approach

The different steps of transformation, compression and decompression are summarized in Fig 3.

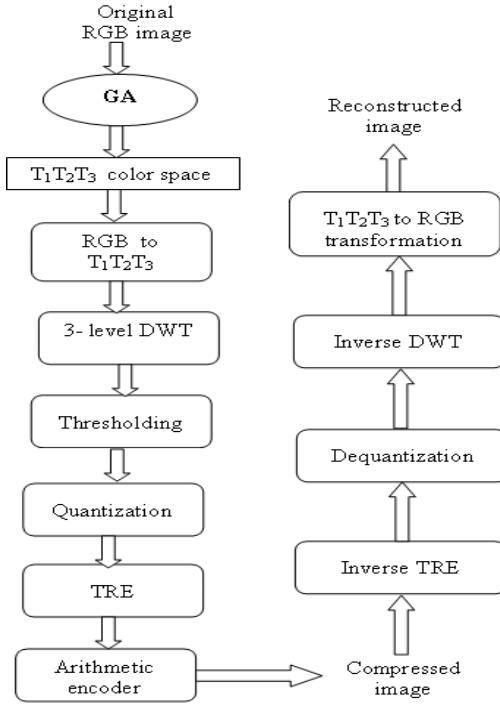


Fig. 3. The process of compression

3.1 GA-DWT Based Compression Phase

In this phase, the proposed GA-DWT based compression technique is built around several steps. Each step will be explained in more details as follows:

1. Genetic Algorithm: we use the GA to find the $T_1T_2T_3$ color space. T_1 represents luminance information and T_2, T_3 are chrominance information.
2. RGB to $T_1T_2T_3$ transformation: The reason of this process is that most of the image energy lies within T_1 component and also human eye is more sensitive towards luminance change than color changes. So is good to work in $T_1T_2T_3$ domain and to treat these three components separately.

3. DWT procedure: In this step, the process applies the DWT on the original image up to three levels in order to obtain the vector of wavelet coefficients. We note that we used mother wave bior4.4 detailed in [10].
4. Thresholding: simply, if the absolute values of NonZero Wavelet Coefficients (NZWC) are less than a given THreshold (TH), these coefficients are eliminated.

In this step we employed the bisection algorithm similar to that in [5] to control the PNSR in advance with a precision of convergence ϵ (it is chosen equal to 0.01).

5. Quantification: In this step, the NZWC are quantified by a linear quantification of size Q bits. The objective of this quantification is to reduce the number of bits necessary to the representation of these coefficients. The quantification of the NZWC is done according to the formula [5]:

$$\text{QNZWC} = \left\lfloor 1 + \frac{\text{NZWC} - \text{NZWC}_{\text{Min}}}{\text{NZWC}_{\text{Max}} - \text{NZWC}_{\text{Min}}} \times (2^Q - 2) \right\rfloor. \quad (9)$$

where: $\lfloor \cdot \rfloor$ represents the nearer value, NZWC_{Min} is the minimal value of NZWC, NZWC_{Max} maximal value of NZWC and Q is the lowest quantizer resolution.

In this step, every NZWC is quantized to become a Quantified NZWC (QNZWC) with the lowest possible resolution.

6. The Two Role Encoder (TRE): In this step, the process encodes the quantified coefficients in the zig-zag sequence by lossless encoding TRE [5]. The QNZWC is coded by a non negative integer of width equal $(Q + 1)$ bits .

The thresholding step yields to many long run of zeros, each one is replaced by only one TRE code of $(Q + 1)$ bits. The minimum run of zeros that is allowed to be coded by a TRE code is 1, the maximum value is $2^Q - 1$.

7. Arithmetic encoder: The concatenation of all vectors produces a global vector that is compressed by means of the arithmetic encoder.

3.2 GA-DWT Based Decompression Phase

Decompression is just the inverse process of compression as indicated in Fig. 3.

4 Experimental Results and Performance Comparison

In order to assess and test the robustness and the efficiency of the proposed GA-DWT based approach, we have used the well-known color images : Airplane, Peppers, Lena of size 512×512 for each one and Girl, Couple and House of size 256×256 for each one.

The results reported in Table 2; show the efficiency in performance of our GA-DWT based approach in the $T_1T_2T_3$ color space.

The curves given on the Fig.4 illustrate that the bpp and the PSNR obtained in the $T_1T_2T_3$ space are better than of the direct application on the RGB space. Therefore these results confirm that the $T_1T_2T_3$ color space is more suitable for compression.

Table 2. Performances in the RGB and $T_1T_2T_3$ space for the different quantizer width

Q	RGB color space						$T_1T_2T_3$ color space					
	7 bits		8 bits		9 bits		7 bits		8 bits		9 bits	
Images	PSNR	bpp	PSNR	bpp	PSNR	bpp	PSNR	bpp	PSNR	bpp	PSNR	bpp
Airplane	30.90	0.83	31.21	0.90	31.51	1.02	31.84	0.57	31.85	0.56	31.74	0.63
Peppers	30.95	0.84	30.87	0.97	30.97	0.95	31.94	0.95	31.94	0.87	31.74	0.97
Lena	32.85	1.08	32.85	1.18	33.00	1.21	33.51	0.81	33.84	0.75	33.15	0.82
Girl	35.75	0.90	35.48	0.96	35.88	1.10	36.25	0.57	35.74	0.56	35.87	0.58
Couple	33.57	1.50	32.87	1.53	33.87	1.72	32.36	0.89	32.90	1.01	33.99	1.12
House	32.27	1.14	32.17	1.25	32.17	1.36	32.87	0.90	32.09	0.93	32.87	1.06
Zelda	31.67	1.21	31.57	1.29	31.57	1.46	32.77	0.89	31.86	0.87	31.96	1.06
Average	32.57	1.07	32.43	1.15	32.71	1.26	33.08	0.80	33.03	0.79	33.05	0.89

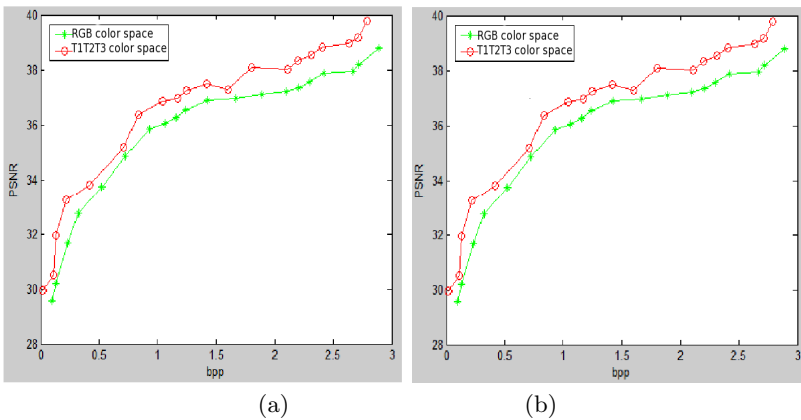


Fig. 4. Performances in the RGB and $T_1T_2T_3$ space applied for Lena and Girl color images: (a) Lena, (b) Girl

Figure 5 gives visual and quantitative results of the proposed method.

Comparative results of the recent published algorithms in [3][4] and our proposed algorithm are presented in Table 3.



Fig. 5. Reconstructed images: (a) Airplane (PSNR= 31.16, bpp= 0.49), (b) Peppers (PSNR=31.20, bpp =0.83), (c) Lena (PSNR=32.76, bpp=0.66)

Table 3, shown clearly that the results of our approach are particularly powerful compared to the CBTC-PF and CDABS.

Table 3. Performances comparison between the proposed method, CDABS and the CBTC-PF algorithm

Images	Proposed method		CDABS [4]		CBTC-PF	
	PNSR	bpp	PNSR	bpp	PNSR	bpp
Airplane	31.16	0.49	30.38	0.59	30.36	1.04
Peppers	31.20	0.83	30.05	0.80	30.15	1.50
Lena	32.76	0.66	31.97	0.81	31.93	1.17
Cirl	35.90	0.41	35.00	0.45	35.13	0.60
Couple	32.87	0.89	32.28	0.92	32.44	1.00
House	32.10	0.83	31.72	0.82	31.79	1.20
Zelda	31.98	0.76	31.33	0.87	31.31	1.12
Average	32.57	0.69	31.82	0.75	31.87	1.09

5 Conclusion

In this paper, we have proposed a new color image compression method based on DWT and an appropriate GA. This approach is based on the fact that there are an infinite number of possible color spaces instead of the RGB channels. The best of these other color spaces are deriving by using a GA based on the DWT transform. Indeed, we apply our proposed GA-DWT approach in order to find a more suitable space referred to as $T_1T_2T_3$ for each image from the given RGB image. Thus, this new GA-DWT approach has the ability to build this base $T_1T_2T_3$, which T_1 energy is more maximized than that of T_2 and T_3 , which permits a more effective compression because the information is concentrated in the plan T_1 . Thus, our GA-DWT have the ability to compresses more effectively T_2 and T_3 in order to eliminate the intercolor planes correlation.

The evaluation tests and experimental results obtained on different color images, showed clearly that the $T_1T_2T_3$ color space is better than RGB in general. In addition, the obtained results are rather satisfactory compared to the CBTC-PF and CDABS.

Acknowledgments. The authors thank Mr. Redha Benzid for his original ideas and encouragement which helped in obtaining the results.

References

1. Khalid, S.: Introduction to data compression. Elsevier, San Francisco (2006)
2. Zixiang, X., Kannan, R., Orchard, M.T., Ya-Qin, Z.: A Comparative study of DCT- and Wavelet-based image coding. IEEE Transactions on Circuits and Systems for Video Thechnology 9, 692–695 (1999)
3. Chandra Dhara, B., Chanda, B.: Color image compression based on block truncation coding using pattern fitting principle. Pattern Recognition 40, 2408–2417 (2007)

4. Douak, F., Benzid, R., Benoudjit, N.: Color image compression algorithm based on the DCT transform combined to an adaptive block scanning. *AEU - International Journal of Electronics and Communications* 65, 16–26 (2011)
5. Benzid, R., Marir, F., Bouguechal, N.-E.: Electrocardiogram compression method based on the adaptive wavelet coefficients quantization combined to a modified Two-Role Encoder. *IEEE Signal Processing Letters* 14, 373–376 (2007)
6. Varun, S., Vinod, K.: Coding of DWT coefficients using Run-length coding and Huffman coding for the purpose of color image compression. *World Academy of Science, Engineering and Technology* 62, 696–699 (2012)
7. Bhardwaj, A., Ali, R.: Image compression using modified fast haar wavelet transform. *World Applied Sciences Journal* 7, 647–653 (2009)
8. Elharar, E., Stren, A., Hadar, O., Jvidi, B.: A Hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform. *Journal of Display Technology* 5, 1–5 (2007)
9. Piotr, P., Agnieszka, L.: The haar wavelet transform in digital image processing: Its status and achievements. *Machine Graphics and Vision* 13, 79–98 (2004)
10. RajKumar, T.M.P., Mrityunjaya Latte, V.: Performance evolution of various wavelet families in spihl image compression technique. *European Journal of Scientific Research* 59, 14–21 (2011)
11. Ghamisi, P., Santha devi, P., Phil, M.: Efficient wavelet based image compression technique for wireless communication. *International Journal of Innovative Technology and Creative Engineering* 1, 53–59 (2011)
12. Alkholidi, A., Cottour, A., Alfalou, A., Hamam, H., Keryer, G.: Real-time optical 2D wavelet transform based on the JPEG 2000 standards. *European Physical Journal Applied Physics (EPJ AP)* 44, 261–272 (2008)
13. Holland, J.H.: *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor (1975)
14. Goldberg David, E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston (1989)
15. Chakrapani, Y., Soundara Rajan, K.: Genetic algorithm applied to fractal image compression. *ARNP Journal of Engineering and Applied Sciences* 4, 53–58 (2009)
16. Lifeng, X., Liangbin, Z.: A study of fractal image compression based on an improved genetic algorithm. *International Journal of Nonlinear Science* 3, 116–124 (2007)
17. Alkholidi, A., Alfalou, A., Hamam, H.: A new approach for optical colored image compression using the JPEG standards. *Signal Processing* 87, 569–583 (2007)
18. Chan, K.Y., Stich, D., Voth, A.G.: Real-time image compression for high-speed particle tracking. *Review of Scientific Instruments* 78 (2007)
19. Andrew, C., Peter, F., Hartmut, P., Carlos, F.: *Genetic Algorithm TOOLBOX For Use with MATLAB, Version 1.2 Users Guide*, <http://www.shef.ac.uk/acse/research/ecrg/gat.html>

Accelerator-Based implementation of the Harris Algorithm^{*}

Claude Tadonki¹, Lionel Lacassagne²,
Elwardani Dadi³, and Mostafa El Daoudi³

¹ Mines ParisTech - Centre de Recherche en informatique,
Mathématiques et systèmes 35, rue Saint Honoré,
77305 Fontainebleau Cedex, France
claude.tadonki@mines-paristech.fr

² Institute of Fundamendal Electronics,
University of Orsay, Faculty of Sciences, Bat. 220,
91405 Orsay Cedex, France

³ Université Mohamed Premier Oujda,
Boulevard Mohamed VI, Oujda, Morocco

Abstract. Real-time implementations of *corner detection* is crucial as it is a key ingredient for other image processing kernels like *pattern recognition* and *motion detection*. Indeed, *motion detection* requires the analysis of a continuous flow of images, thus a real-time processing implies the use of highly optimized subroutines. We consider a tiled implementation of the Harris corner detection algorithm on the CELL processor. The algorithm is a chain of local operators applied to each pixel and its periphery. Such a special memory access pattern clearly exacerbates on the hierarchy transition penalty. In order to reduce the consequent time overhead, tiling is a commonly considered way. When it comes to *image processing filters*, incoming tiles are overdimensioned to include their neighborhood, necessary to update border pixels. As the volume of "extra data" depends on the tile shape, we need to find a good tiling strategy. On the CELL, such investigation is not directly possible with native DMA routines. We overcome the problem by enhancing the DMA mechanism to operate with non conventional requests. Based on this extension, we proceed with experiments on the CELL with a wide range of tile sizes and shapes, thus trying to confirm our intuition on the optimal configuration.

Keywords: Accelerator, CELL BE, Harris, image processing, tiling, DMA.

1 Introduction

The common characteristic of image processing algorithms is the heavy use of basic operators. Indeed, the typical scheme is a repetitive application of

^{*} Work jointly supported by ANR projects Ocelle and PetaQCD, also by the Excellence Grant of Moroccan Ministry of Higher Education. Grant No. G 08/004.

linear and local kernels at the pixel level. The fact that each output pixel is obtained from the corresponding input pixel and its periphery breaks any hope of regular memory access, thus making it hard to achieve real-time performance implementations.

The *Harris* algorithm [4] for corner detection is an interesting case study application because it allows various implementation and optimization strategies [6]. Among these possibilities, tiling [10] is potentially attractive as it can be considered inside any valid scheduling as an additional (memory) optimization. However, tiling on the CELL cannot be directly implemented because of data alignment constraints when using native DMA routines. Because of this constraint, tiles corresponding to contiguous memory region (full row tiles for instance) are used most of the time. Thus, there is no choice for the tile shape.

Tile shape restriction is particularly frustrating with image processing operators because either it does not allow the use of a predicted optimal tile shape, or it acts as a runtime bottleneck. The latter could occur, for instance, with an image so large that the SPE local store cannot hold three of its entire rows (one active row plus its top and bottom neighborhoods). *Data alignment* is another critical requirement. On this paper, we focus on the problem and provide a seamless effective solution. We study the effect of tiling and report experimental results driven by theoretical predictions. Our approach is more general for an accelerated-based computation, we chose CELL BE to illustrate our strategy.

The rest of the paper is organized as follows. The next section presents an overview of the CELL Broadband Engine. We describe the *Harris-Stephens algorithm* and some implementation considerations in Section 3. In Section 4, we discuss about tiling and predict the optimal tile shape. This is followed in Section 5 by an outline of the DMA issue when considering general tile shapes and what we provide to overcome the problem. Section 6 presents and analyses our experimental result according to our predictions. Section 7 concludes the paper.

2 The CELL Broadband Engine

Designed to provide a real-time processing response and a high-bandwidth network, the CELL [15] is a multi-core chip composed of nine processing elements. One core, the *POWER Processing Element* (PPE), is a 64-bit Power Architecture acting as a kind of master. The remaining eight cores, the *Synergistic Processing Elements* (SPEs), RISC architecture with SIMD organization with 128-bit vector registers and 256 KB of local memory, referred to as local store (LS). Each SPE has a clock speed of 4 Ghz (3.2 Ghz in average), with a peak performance of 256 GFlops (single precision) and 26 GFlops (double precision). The chip can handle 128 concurrent transactions to memory per processor. Figure 1 provides a synthetic view of the CELL architecture [5].

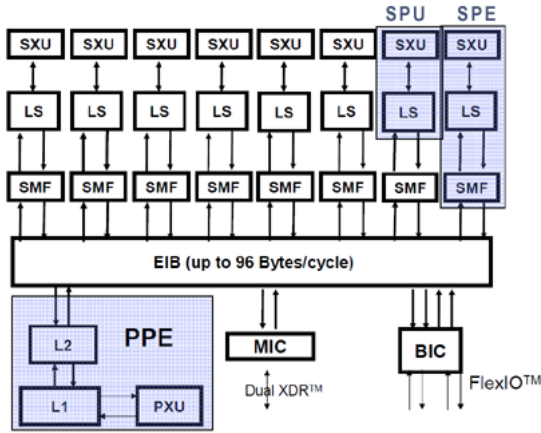


Fig. 1. Cell Chip Block Diagram

Programming the CELL is mainly a mixture of single instruction multiple data parallelism, instruction level parallelism and thread-level parallelism. The chip was primarily intended for digital image/video processing, but was immediately considered for general purpose scientific programming (see [9] for an exhaustive report on the potential of the CELL BE for several key scientific computing kernels). A specific consideration for QR factorization is presented in [2]. Nevertheless, exploiting the capabilities of the CELL in a standard programming context is really challenging. The programmer has to deal with hardware and software constraints like *data alignment*, *local store size*, *double precision penalty*, *different level of parallelism*. Efficient implementation on the CELL is commonly a conjunction of a good computation/DMA overlap and a heavy use of the SPU intrinsics.

3 The Harris-Stephen Algorithm

Harris and Stephen [4] *interest point detection algorithm* is an improved variant of the *Moravec corner detector* [7], used in computer vision for feature extraction like *motion detection*, *image matching*, *tracking*, *3D reconstruction* and *object recognition*. Figure 2 illustrates the use of the algorithm.



Fig. 2. Illustration of the Harris-Stephens procedure

The algorithm is mainly a succession of local operators implementing a discrete form of an autocorrelation S , given by

$$S(x, y) = \sum_{u, v} w(u, v)[I(x, y) - I(x - u, y - v)]^2, \tag{1}$$

where (x, y) is the location of a pixel with color value $I(x, y)$, and $u, v \in 1, 2, 3$ model the move on each dimension. At a given point (x, y) of the image, the value of $S(x, y)$ is compared to a suitable *threshold*, and the decision follows on the nature of the pixel at (x, y) . Roughly speaking, the process is achieved by applying four discrete operators, namely *Sobel* (S), *Multiplication* (M), *Gauss* (G), and *Coarsity* (C). Figure 3 displays an overview of the global workflow.

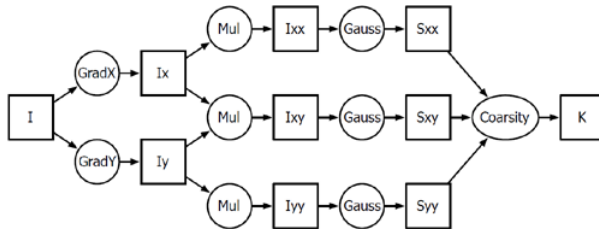


Fig. 3. Harris algorithm diagram

Multiplication and *Coarsity* are point to point operators, while *Sobel* and *Gauss*, which approximate the first and second derivatives, are $9 \rightarrow 1$ or 3×3 operators defined by

$$S_x = \frac{1}{8} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad S_y = \frac{1}{8} \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{2}$$

$$G = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \tag{3}$$

Applying a 3×3 operator to a given pixel (x, y) consists in a point-to-point multiplication of the corresponding 3×3 matrix by the following pixels matrix

$$\begin{pmatrix} I(x - 1, y - 1) & I(x - 1, y) & I(x - 1, y + 1) \\ I(x, y - 1) & I(x, y) & I(x, y + 1) \\ I(x + 1, y - 1) & I(x + 1, y) & I(x + 1, y + 1) \end{pmatrix} \tag{4}$$

Here comes the notion of *border*. In order to compute an output pixel $O(x, y)$, we need the pixel $I(x, y)$ and its immediate periphery. We say the operator is of *depth* 1. Operator depth is additive, means that if two operators f and g are of depth p and q respectively, then the depth of $f \circ g$ is $p + q$. Three problems are raised by the way operators are applied:

- accessing the points at the periphery yields an irregular memory access pattern, which is a serious performance issue
- computing two consecutive points involves some reused pixels (those on their common border). This yields a memory access redundancy, thus another performance issue
- applying each operator separately implies several read and write operations on the main memory (same location or not), yet another source of performance penalty

There are several ways to deal with the above problems. One way is to fuse or chain consecutive operators whenever possible. This overcome the repetitive read/write of the entire image, at the price of data and computation redundancy (more border pixels), thus should be done under a certain compromise. The first two issues are well tackled by tiling, which could be considered with fused operators. Although tiling is a more general technique, we really need a specific analysis in order to understand how the extra data that covers each incoming tile affect the global performance when dealing with operator-based algorithms.

4 Tiling Consideration

When applying an operator to a given tile, we need some *extra pixels* for the calculation of *border pixels*. This means that, applying the *Sobel* operator to a $a \times b$ tile yields a $(a - 1) \times (b - 1)$ tile. This aspect is usually referred in the reverse side, means that we require a $(a + k) \times (b + k)$ tile in order to produce a $a \times b$ tile, where k is the *depth* of the operator. Redundant reads/writes and computations occur within the *border*, whose the volume depends on the tile shape. Indeed, since $(a + k) \times (b + k) = ab + k(a + b) + k^2 \approx ab + k(a + b)$, we see that the volume of the *border* is $k(a + b)$ for a $a \times b$ tile. Here comes the question about the optimal tile shape for a fixed tile volume (typically derived from memory constraints). We give the answer in proposition [□](#)

Proposition 1. *The optimal tile shape over the set of tiles with equal volume is a square tile.*

Proof. We need to minimize

$$M(a, b) = (a + b) \frac{W \times H}{ab}, \quad (5)$$

where $ab = \lambda$ (constant). Indeed, $\frac{W \times H}{ab}$ is the number of $a \times b$ tiles on the $W \times H$ region, and the border of a $a \times b$ tile is proportional to $(a + b)$. Reporting $b = \frac{\lambda}{a}$ in [\(5\)](#) yields

$$M(a, \lambda) = \left(a + \frac{\lambda}{a}\right) \frac{W \times H}{\lambda}, \quad (6)$$

and we get

$$\frac{\partial M}{\partial a} = \left(1 - \frac{\lambda}{a^2}\right) \frac{W \times H}{\lambda}. \quad (7)$$

Thus, $\frac{\partial M}{\partial a} = 0$ gives $a = \sqrt{\lambda}$ and then $b = \frac{\lambda}{\sqrt{\lambda}} = \sqrt{\lambda}$, i.e. $a = b$.

This result is important, provided the possibility to use any expected tile shape. We made this possible by encapsulating the necessary DMA into a single and generic routine. The result is general, but we need to check it for the case of the CELL because of the special access to the main memory. We use a scalar implementation of the operators and consider a full fused form of the Harris-Stephens algorithm.

5 DMA Issue with Standard Tiles

The problem we want to solve can be stated as follows. Given M_p , a $n_p \times m_p$ matrix on the main memory, and M_s , a $n_s \times m_s$ matrix on the local store. We need to copy the $a \times b$ submatrix of A_p located at (i_p, j_p) into A_s at the location (i_s, j_s) . Figure 4 depicts the task.

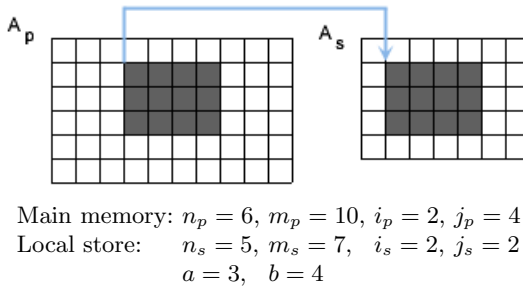


Fig. 4. Generic DMA pattern

Performing the transfer expressed in figure 4 raises number of problems:

- the region to be transferred is not contiguous on memory, thus list DMAs are used most of the time
- the address of one given row is not aligned, thus the global list DMA is not possible
- the (address, volume) pair of a row does not match the basic DMA rules (the above two ones), thus the entire list DMA cannot be carried out
- misalignment could come from both sides (main memory and/or local store)
- the target region on the local store might be out of the container limits

It is important to overcome the above problems at the minimum cost, since the consequent (pre/post)processing is an overhead for the programmer. To do so, we encapsulate all necessary (pre/post)processing into a single generic DMA subroutine. Roughly speaking, we perform either a direct DMA or a list DMA (one DMA per line of the tile), taking care of the above issues.

6 Experimental Results

We now proceed to some experimentations. The goal is to validate our implementation over various tile shapes, and see how close we are regarding our prediction on the optimal tile shape. Our program runs from the PPE and cooperate with one SPE. For each image, we chose a fixed tile volume and iterate on various shapes.

Table 1. Timings on a 512×512 image

$tile_h$	$tile_w$	total time(s)
8	512	0.0494
16	256	0.0598
32	128	0.0485
64	64	0.0345
128	32	0.0517
256	16	0.0699
512	8	0.0734

Table 2. Timings on a 2048×512 image

$tile_h$	$tile_w$	total time(s)
8	512	0.198
16	256	0.238
32	128	0.187
64	64	0.110
128	32	0.180
256	16	0.218
512	8	0.352

Table 3. Timings on a 1200×1200 image

$tile_h$	$tile_w$	total time(s)
5	1200	0.494
10	600	0.360
20	300	0.264
40	150	0.235
80	75	0.183
160	37	0.247
320	18	0.275

Table 4. Timings on a 2048×2048 image

$tile_h$	$tile_w$	total time(s)
8	512	0.985
16	256	0.726
32	128	0.643
64	64	0.438
128	32	0.692
256	16	0.866
512	8	1.422

We see that the most squared tile always gives the best global performance. The difference is marginal with closest shapes, but we should keep in mind that the typical use of the algorithm is with a flow of images. Our implementation does not overlap DMA with computations because of memory postprocessing due to misalignment. For wider images (Figures 3 and 4), we see that the improvement is more than 50% compared to full row tiles. We emphasize on the extra cost for managing irregular DMAs, although our implementation seems to perform well. The main difference between full row tiles and the others is that, for the later, DMA list is always necessary. Thus, the compromise here is between irregular DMAs and redundancies. Our experimental results clearly show that it still advisable to consider tiles with balanced dimensions.

7 Conclusion

The Harris-Stephens algorithm is a classical procedure in computer vision. From a programming point of view, it offers a wide range of optimization possibilities, each of them being appropriate for specific architecture. Since the CELL processor suits for image/video processing, investigating on the implementation of the Harris-Stephens algorithm is quite relevant, having in mind the impact on a stream processing context. In our work, we consider a tiled implementation based on a fused version of the algorithm. Using on our implementation of "irregular" DMAs, we provide a blocked implementation of the algorithm and we validate the optimal tile shape prediction. For absolute performances, we need to optimize our implementation of the basic operator (mainly with SPU intrinsics) and study how to overlap DMAs and computations. Due to the current status of the CELL BE, we plan to test our method on GPUs, and then consider the aforementioned issues in a more global context.

References

1. Cell SDK 3.0, www.ibm.com/developerworks/power/cell
2. Kurzak, J., Dongarra, J.: QR factorization for the Cell Broadband Engine. *Scientific Programming* 17(1-2), 31–42 (2009)
3. <http://www.gnu.org/software/octave/>
4. Harris, C., Stephens, M.: A combined corner and edge detector. In: 4th ALVEY Vision Conference (1988)
5. Peter Hofstee, H.: Power Efficient Processor Design and the Cell Processor, http://www.hpcacconf.org/hpca11/slides/Cell_Public_Hofstee.pdf
6. Saidani, T., Lacassagne, L., Falcou, J., Tadonki, C., Bouaziz, S.: Parallelization Schemes for Memory Optimization on the Cell Processor: A Case Study on the Harris Corner Detector. *HIPEAC Journal* (2009)
7. Moravec, H.: Obstacle avoidance and navigation in the real world by a seeing robot rover. In: Tech. report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University & doctoral dissertation, Stanford University (September 1980)
8. Sen, S., Chatterjee, S.: Towards a theory of cache-efficient algorithms. In: *SODA* (2000)
9. Williams, S., Shalf, J., Oliker, L., Kamil, S., Husbands, P., Yelick, K.: Scientific Computing Kernels on the Cell Processor. *International Journal of Parallel Programming* (2007)
10. Xue, J.: Loop tiling for parallelism. Kluwer (2000)

Writer Recognition on Arabic Handwritten Documents

Chawki Djeddi¹, Labiba Souici-Meslati², and Abdellatif Ennaji³

¹ Laboratoire LAMIS, Département de Mathématiques et d'Informatique,
Université de Tébessa, Route de Constantine, 12000, Tébessa, Algérie

² Laboratoire LRI, Département d'Informatique,
Université Badji Mokhtar d'Annaba, B.P 12, 23000. Annaba, Algérie

³ Laboratoire LITIS, UFR des Sciences, Université de Rouen, 76800,
Saint-Etienne du Rouvray, France

c.djeddi@mail.univ-tebessa.dz

Abstract. Recognizing the writer of a handwritten document has been an active research area over the last few years and is at the heart of many applications in biometrics, forensics and historical document analysis. In this paper, we present a novel approach for text-independent writer recognition from Arabic handwritten documents. To characterize the handwriting styles of different writers involved in the evaluation of our approach, we have used two texture methods based on edge hinge features and run-lengths features. The efficiency of the proposed approach is demonstrated experimentally by the classification of 1375 handwritten documents collected from 275 different Arabic writers.

Keywords: writer identification, writer verification, run-lengths, edge hinge, Arabic handwriting.

1 Introduction

Writer recognition based on handwritten documents is a hot and promising research topic in the field of pattern recognition due to its various applications; it is a classical pattern recognition problem [1]. The classification task in pattern recognition is to assign a pattern to one class out of a set of classes. In this paper, a pattern is a sample of handwritten text and a class represents a writer.

Writer recognition is the process of automatically recognizing who is writing on the basis of individual information included in handwritten documents. Writer recognition refers to two different tasks: Writer identification and writer verification. Writer identification determines which writer provides a given handwriting form amongst a set of known writers. Writer verification consists to decide on two handwritten documents and determine if they are written by the same writer or by two different writers.

Writer recognition approaches can be categorized into two distinct families: text-dependent approaches and text-independent approaches: In text-dependent approaches, the writer must write exactly a predefined or a given text. The text-independent writer recognition is a process of identifying or verifying the identity of the writer without constraint on the text content.

Writer recognition systems are involved in many applications such as biometric recognition [2, 3, 4, 5], personalized handwriting recognition systems [6], automatic forensic document examination [7], classification of ancient manuscripts [8] and smart meeting rooms [9]. That is the reason why many efforts have been made in order to improve writer recognition methods.

Up to now, researchers in the field of Text-Independent Writer Recognition have mainly focused on the statistical approach. This has led to the specification and extraction of statistical features such as slant distribution, entropy, and edge-hinge distribution. We found that the edge-hinge distribution feature outperforms all other statistical features [2]. Therefore, the aim of this paper is to compare our improved run-lengths features with edge hinge features.

The remaining of the paper is organized as follows: in the first section, we give a brief overview on some significant recent contributions to Arabic writer recognition. In the next part, we introduce the database used in our study, followed by the description of our proposed approach. The following section presents the experimental results and their analysis. Finally, we give a conclusion with some future research directions.

2 Arabic Writer Recognition: A Survey

Writer recognition from Arabic handwritten documents has not been addressed as extensively as writer recognition from Latin or Chinese handwritten documents until the last few years. The first study dates back to 2005 when Al-Zoubeidy et al [10] proposed the use of multichannel Gabor filtering and gray-scale co-occurrence matrices to characterize the writing style of writers. Gazzah et al [11] combined local and global features. Global features were extracted with 2D DWT using lifting scheme but the local features describe the morphological variations of writing (lines height, ascenders slant and diacritical dots features).

Al-Dmour et al [12] presented a feature extraction technique based on hybrid spectral-statistical measures (SSMs) of texture. Bulacu et al [2] proposed an approach based on the combination of textural with allographic features. Joint directional probability distributions and grapheme-emission distributions are extracted independently of the textual content of the written samples. The authors conducted an analysis of the combination of textural and allographic features and showed that the combination of these features improves the performances.

Abdi et al [13] proposed a method based on the combination and the cooperation of six feature vectors computed from the minimum perimeter polygon (MPP) contours of Arabic words. These feature vectors are in the form of probability distribution functions (PDFs), and are based on the length, direction, angle and curvature measurements. In [14] the authors calculate the fractal dimensions for images by using the "Box-counting" method, and calculate the multi-fractal dimensions of images by using the method of DLA (Diffusion Limited Aggregates).

Awaida et al [15] addresses writer identification of Arabic handwritten digits. A combination of Gradient, curvature, density, horizontal and vertical run lengths, stroke, and concavity features is used for characterization of the writing samples.

Al-Ma'adeed et al [16] evaluated the performance of edge-based directional probability distributions as features and moment invariants and words' measurements such as area, length, height, length from baseline to upper edge and length from baseline to lower edge in Writer identification.

Chen et al [17] proposed a method for detecting and removing ruling lines from the handwritten documents and tested its utility for Arabic Writer identification through series of experiments. Their preliminary results show that, under realistic assumptions where ruling lines are expected to have different properties across the collection, e.g., thickness, spacing, etc., removing them significantly improves identification performances.

3 Feature Extraction

In our work, two texture analysis methods are implemented and used for characterizing Arabic handwritings, these methods are : Run-Length distribution [18] and edge-hinge distribution [19].

3.1 Run-Length Features

To characterize the writing style of different writers involved in the evaluation of our writer recognition methods, we compute the probability distribution of run-lengths features, which are determined on a binary image taking into consideration the black pixels corresponding to the ink trace and the white pixels corresponding to the background.

There are four scanning methods: horizontal, vertical, left-diagonal and right-diagonal. We calculate the runs-lengths features using the grey level run-length matrices and the histogram of run-lengths is normalized and interpreted as a probability distribution function (PDF). The method considers horizontal, vertical, left-diagonal and right-diagonal white run-lengths as well as horizontal, vertical, left-diagonal and right-diagonal black run-lengths extracted from the image of the handwritten document.

The run-lengths features we propose to use here give information on the average width of the letters, the density of writing, the structure of the letters, the average size of the letters, the ink width, the characters position, the regions enclosed inside the letters and also the empty spaces between letters and words, the regularity and irregularity of handwriting and finally the slope in handwriting.

We have used the set of features proposed here in the ICDAR'2011 Writer Identification Contest [20], we have also used a part of these features in the ICDAR'2011 Arabic Writer Identification Contest [21] and in the ICDAR'2011 Music Scores Competition: Staff Removal and Writer Identification [22]. We have obtained interesting results in these competitions.

3.2 Edge-Hinge Features

Edge-hinge distribution is a feature that characterizes the changes in direction of a writing stroke in handwritten text [19]. The edge-hinge distribution is extracted by

means of a window that is slid over an edge-detected binary handwriting image. Whenever the central pixel of the window is on, the two edge fragments (i.e. connected sequences of pixels) emerging from this central pixel are considered. Their directions are measured and stored as pairs. A joint probability distribution $P(\varphi_1, \varphi_2)$ is obtained from a large sample of such pairs. An example of an angle pair is shown in Figure 1.

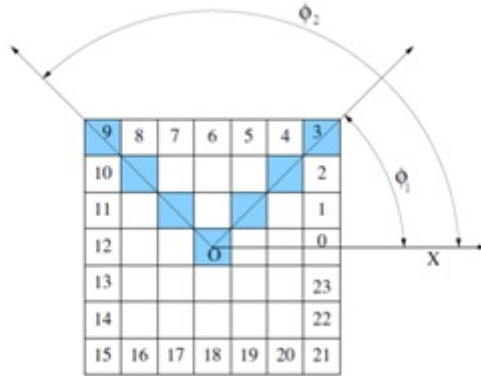


Fig. 1. Example of an edge-hinge distribution (image reproduced from [19])

4 Writer Recognition

Once the handwriting samples have been represented by their respective features, we need to compute the distances between respective features to define a (dis)similarity between two handwriting samples. We tested three distance measures including: Euclidean distance, Chi-square distance and Manhattan distance. In our experimentations, Manhattan distance performed the best.

In writer identification task, the efficiency of the considered features has been evaluated using nearest-neighbor classification [2] in a leave-one-out strategy. Explicitly, one document (a query document) is chosen and extracted from the total of 1375 documents (note that the experimental dataset contains 5 documents written by each of 275 writers), then the distances between the features vector of the chosen document and the features vectors of all of the remaining 1374 documents are computed. For a query document, we don't only find the Top-1 but a longer list up to a given rank (Top-10) thus increasing the chance of finding the correct writer in the retrieved list.

For writer verification, we compute the distance between two given documents and consider them as being written by the same writer if the distance falls within a predefined decision threshold. Beyond the threshold value, we consider that the documents are written by different writers. By varying the acceptance threshold, the ROC curves are computed and the verification performance is quantified by the Equal Error Rate (EER): the point on the curve where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). The lower is EER value, the higher is the accuracy of the system.

5 Experimental Study

The experimental study was carried out on the writing samples from the IFN/ENIT database [23] which is the unique Arabic handwriting publicly available database. It consists of forms with handwritten Arabic town/village names collected from 411 subjects (binary images at 300 dpi resolution). Most writers filled in 5 forms. This database was designed for training and testing recognition systems for handwritten words and was used for the ICDAR 2005 Arabic OCR competition [24].

The IFN/ENIT database was used also in [2, 3, 4, 5, 13] for writer identification and verification because the writer information was recorded. We have extracted the handwriting from the scanned forms. The text content is variable and the samples contain a limited amount of handwriting: only 12 names and 12 zip codes of Tunisian towns/villages. In our writer identification and verification experiments, we used the data concerning 275 writers with 5 samples per writer.

Table 1. Overview of proposed features and their dimensions

Feature	Description	Dimension
<i>f1</i>	Horizontal run-lengths on white pixels	120
<i>f2</i>	Left-diagonal run-lengths on white pixels	120
<i>f3</i>	Vertical run-lengths on white pixels	120
<i>f4</i>	Right-diagonal run-lengths on white pixels	120
<i>f5</i>	Horizontal run-lengths on black pixels	264
<i>f6</i>	Left-diagonal run-lengths on black pixels	264
<i>f7</i>	Vertical run-lengths on black pixels	264
<i>f8</i>	Right-diagonal run-lengths on black pixels	264
<i>f9</i>	Edge-hinge with fragment of 5 pixels	1024
<i>f10</i>	Edge-hinge with fragment of 6 pixels	1600
<i>f11</i>	Edge-hinge with fragment of 7 pixels	2304
<i>f12</i>	Edge-hinge with fragment of 8 pixels	3136

To evaluate the proposed approach, we have conducted two types of experiments: the first one is designed to evaluate the result we can reach by using individually each studied feature vector. Whereas the second type aims at testing the result we can reach by combining the studied feature vectors. For writer identification task, we report the Top 1, Top 5 and Top 10 identification rates while for writer verification task, we present the Equal-Error-Rate (EER).

For each feature, Table 1 summarizes the corresponding number, the description and the dimension, whereas Table 2 presents the performance of the individual features detailed in the above sections. Although the feature performances vary significantly, it can be noticed that the edge-hinge features (*f9-f12*) outperform the run-lengths features (*f1-f8*), with *f12* (Edge-hinge with fragment of 8 pixels) achieving the best results both on identification and verification tasks.

Table 2. Writer recognition performance on individual features

Feature	Top 1	Top 5	Top 10	EER
<i>f1</i>	14,27%	36,51%	50,91%	17,58%
<i>f2</i>	16,07%	41,31%	56,44%	15,02%
<i>f3</i>	8,94%	26,25%	38,47%	21,29%
<i>f4</i>	17,89%	42,40%	56,00%	16,07%
<i>f5</i>	28,65%	56,51%	69,16%	14,20%
<i>f6</i>	28,65%	54,62%	67,27%	13,83%
<i>f7</i>	29,96%	53,16%	65,02%	15,62%
<i>f8</i>	30,69%	54,98%	65,96%	15,14%
<i>f9</i>	83,56%	95,49%	97,45%	6,58%
<i>f10</i>	84,36%	95,34%	97,45%	7,08%
<i>f11</i>	87,49%	97,02%	97,82%	6,30%
<i>f12</i>	89,16%	97,45%	98,84%	5,49%

Table 3. Writer recognition performance on features combination

Features combinations	Top 1	Top 5	Top 10	EER
<i>f1, f2, f3, f4</i>	47,20%	76,14%	86,54%	10,56%
<i>f5, f6, f7, f8</i>	75,42%	90,25%	93,82%	9,56%
<i>f1, f2, f3, f4, f5, f6, f7, f8</i>	88,07%	96,87%	98,54%	5,80%
<i>f1, f2, f3, f4, f5, f6, f7, f8, f12</i>	93,53%	98,47%	99,13%	4,78%

Table 3 summarizes some of the combinations we have tested. For writer identification, the highest rate we have reached stands at 93.53% in Top 1, 98.47% in Top 5 and 99.13% in Top 10 when combining run-lengths on white and black pixels with edge hinge with fragment of 8 pixels (*f1-f8, f12*). For the verification task, we achieve an EER of 4.78% when combining run-lengths on white and black pixels with edge hinge with fragment of 8 pixels (*f1-f8, f12*). The ROC curves for some of the feature combinations have been illustrated in figure 2.

When comparing the recognition performance across the two types of features, it can be seen that the identification and verification results are much poor when using run-lengths with individual features but that is comparable with the edge hinge features when we combine all the run-lengths features. Since the IFN/ENIT database has been widely used in evaluating writer identification and verification tasks, it would be interesting to present a comparative overview of the proposed methods.

Table 4 summarizes the performance of the most recent studies on writer identification and verification on this dataset. Bulacu & al [2] currently hold the best performance results with 88% in Top 1 and 99% in Top 10 on 350 writers in identification task and EER of 5.8% in verification task. We have achieved an identification rate of 88.07% in Top 1, 96.87% in Top 5 and 98.54% in Top 10 by using the run-lengths features and we have improved the results by combining the run-lengths features with edge hinge features to achieve an identification rate of 93.53% in Top 1, 98.47% in Top 5 and 99.13% in Top 10 and an EER of 4.78%.

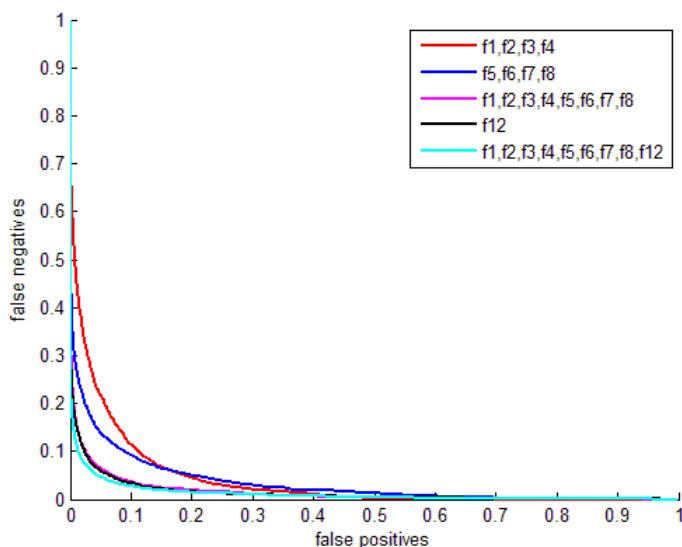


Fig. 2. ROC curves for some of the feature combinations

Table 4. Comparison of writer recognition methods

Reference	Writers	Top 1	Top 5	Top 10	ERR
Abdi & al [13]	82	90.20%	96.30%	97.50-%	-
Bulacu & al [2]	350	88.00%	-	99.00%	5.80%
Our method	275	93,53%	98,47%	99,13%	4,78%

6 Conclusion and Future Work

We have proposed here a new writer recognition method based on Arabic handwriting. The strength of this method is demonstrated experimentally by the classification of 1375 Arabic handwriting images from 275 different writers. Comparisons of improved run-lengths features with the edge hinge features demonstrate that the run-lengths features possess good discriminatory information and that a good method of extracting such information is the key to success of the classification.

The method that has been proposed here is mainly text-independent. In our future work, text-dependent writer identification will be considered and can include signature verification methods. A comparison between the two approaches will then be conducted. Currently, our work is based on the extraction of global features, but further work will focus on the use of local features. An integrated system will be considered by combining both local and global features to produce more reliable classification accuracy.

References

1. Schlapbach, A.: Writer Identification and Verification. PhD Thesis, Bern University (2007)
2. Bulacu, M., Schomaker, L., Brink, A.: Text-Independent Writer Identification and Verification on Off-Line Arabic Handwriting. In: 9th International Conference on Document Analysis and Recognition, Brazil, vol. 2, pp. 769–773 (2007)
3. Djeddi, C., Souici-Meslati, L.: Une approche locale en mode indépendant du texte pour l'identification de scripteurs: Application à l'écriture arabe. In: Colloque International Francophone sur le Document et l'Écrit, Rouen, France, pp. 151–156 (2008)
4. Djeddi, C., Souici-Meslati, L.: A texture based approach for Arabic Writer Identification and Verification. In: IEEE International Conference on Machine and Web Intelligence, Algiers, Algeria, pp. 88–93 (2010)
5. Djeddi, C., Souici-Meslati, L.: Artificial Immune Recognition System for Arabic Writer Identification. In: 4th IEEE International Symposium on Innovation in Information and Communication Technology, Amman, Jordan, pp. 159–165 (2011)
6. Nosary, A., Heutte, L., Paquet, T.: Unsupervised writer adaption applied to handwritten text recognition. *Pattern Recognition* 37, 385–388 (2004)
7. Van Erp, M., Vuurpijl, L., Franke, K., Schomaker, L.: The WANDA measurement tool for forensic document examination. *Journal of Forensic Document Examination* 16, 103–118 (2005)
8. Siddiqi, I., Cloppet, F., Vincent, N.: Contour Based Features for the Classification of Ancient Manuscripts. In: 14th Conference of the International Graphonomics Society, Dijon, France (2009)
9. Liwicki, M., Schlapbach, A., Bunke, H., Bengio, S., Mariéthoz, J., Richiardi, J.: Writer Identification for Smart Meeting Room Systems. IDIAP research report IDIAP-RR 05-70 (2005)
10. Al-Zoubeidy, L.M., Al-Najar, H.F.: Arabic writer identification for handwriting images. In: International Arab Conference on Information Technology, Amman, Jordan, pp. 111–117 (2005)
11. Gazzah, S., Ben Amara, N.E.: Arabic Handwriting Texture Analysis for Writer Identification using the DWT-lifting Scheme. In: 9th International Conference on Document Analysis and Recognition, vol. 2, pp. 1133–1137 (2007)
12. Al-Dmour, A., Abu Zitar, R.: Arabic Writer Identification based on Hybrid Spectral-Statistical Measures. *Journal of Experimental and Theoretical Artificial Intelligence* 19, 307–332 (2007)
13. Abdi, M.N., Khemakhem, M., Ben-Abdallah, H.: Off-Line Text-Independent Arabic Writer Identification using Contour-Based Features. *International Journal of Signal and Image Processing* 1, 4–11 (2010)
14. Chaabouni, A., Boubaker, H., Kherallah, M., Alimi, A.M., El Abed, H.: Fractal and Multi-fractal for Arabic Offline Writer Identification. In: International Conference on Pattern Recognition, Istanbul, Turkey, pp. 1051–4651 (2010)
15. Awaida, S.M., Mahmoud, S.A.: Writer Identification of Arabic Handwritten Digits. In: 1st International Workshop on Frontiers in Arabic Handwriting Recognition, Istanbul, Turkey (2010)
16. Al-Ma'adeed, S., Mohammed, E., Al Kassis, D., Al-Muslih, F.: Writer Identification using Edge-Based Directional Probability Distribution Features for Arabic Words. In: IEEE/ACS International Conference on Computer Systems and Applications, pp. 582–590 (2008)

17. Chen, J., Lopresti, D., Kavallieratou, E.: The Impact of Ruling Lines on Writer Identification. In: 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, pp. 439–444 (2010)
18. Tang, X.: Texture Information in Run-Length Matrices. *IEEE Transactions on Image Processing* 7(11), 1602–1609 (1998)
19. Bulacu, M.: Statistical Pattern Recognition for Automatic Writer Identification and Verification. PhD thesis, University of Groningen (2007)
20. Louloudis, G., Stamatopoulos, N., Gatos, B.: ICDAR 2011 - Writer Identification Contest. In: 11th International Conference on Document Analysis and Recognition, Beijing, China, pp. 1475–1479 (2011)
21. Hassaine, A., Al-Maadeed, S., Alja'am, J.M., Jaoua, A., Bouridane, A.: The ICDAR 2011 Arabic Writer Identification Contest. In: 11th International Conference on Document Analysis and Recognition, China, pp. 1470–1474 (2011)
22. Fornés, A., Dutta, A., Gordo, A., Lladós, J.: The ICDAR 2011 Music Scores Competition: Staff Removal and Writer Identification. In: 11th International Conference on Document Analysis and Recognition, Beijing, China, pp. 1511–1415 (2011)
23. Pechwitz, M., Maddouri, S., Margner, V., Ellouze, N., Amiri, H.: IFN/ENIT-database of handwritten arabic words. In: Colloque International Francophone sur le Document et l'Écrit, pp. 129–136 (2002)
24. Margner, V., Pechwitz, M., El Abed, H.: ICDAR 2005 arabic handwriting recognition competition. In: 8th International Conference on Document Analysis and Recognition, pp. 70–74 (2005)

Outline Matching of the 2D Shapes Using Extracting XML Data

Noredidine Gherabi and Mohamed Bahaj

Hassan 1st University, FSTS,
Department of Mathematics and Computer Science
{gherabi,mohamedbahaj}@gmail.com

Abstract. This paper presents an efficient shape matching method based on XML data, we extract the contour of the shape and this one is represented by set of points. Using corner detection method for representing the contour by a sequence of convex and concave segments. After, each segment is described by local and global features, this features are coded in string of symbols and stored in a XML file. Finally, using the dynamic programming, we find the optimal alignment between sequences of symbols. Results are presented and compared with existing methods using MATLAB for KIMIA-25 database and MPEG7 databases.

Keywords: XML, DOM, Shape descriptor, Shape matching, Dynamic Programming.

1 Introduction

Matching 2D shapes and measuring the similarity between shapes are important problems in Computer vision.

A large body of research has been devoted to shape matching, comparison and recognition. The most commonly used shape representation primitives are curves, point sets, and medial axes.

Many traditional curve matching approaches [1], [2], [3] use local invariant features (e.g., curvature) as descriptors.

Shapes have several properties that can be used for recognition and categorization, like shape, color, texture and brightness. Biederman [4] suggested that edge-based representations mediate object recognition. In his approach, color and texture of surfaces are used to define edges which are then used for recognition.

The goal of our work is to develop a descriptor based on the local and global information of the shape. These information are coded in string of symbols, these are compared using the Dynamic Programming approach. In [5, 6] dynamic programming is used to minimize a cost function that accounts for displacement of a contour in a pair of images from an image sequence. In [7] a DP approach has been used for shape matching and retrieval. The basic idea behind this approach is to represent each shape by a sequence of convex and concave segments using the inflection points extracted

from the curvature and to allow the matching of merged sequences of small segments in a shape with larger segments in the other shape.

This paper looks into developing a shape descriptor for a contour of any shape and transforms it into string of symbols; they will be stored in an XML file. For each shape there is an XML file that corresponds to the features of the shape. Man Hing [8] uses this technique to extract the features information of the shape and represent this information in an XML format, this proposed system will use the XML (standard language) for querying different image databases.

Our approach aims to develop a simple and fast method of shape matching based on the transformation semantic data of the shape into XML format and compute the similarity between XML Files using Dynamic programming.

2 Proposed Method

Our approach to shape recognition is based on several steps summarized in Fig 1.

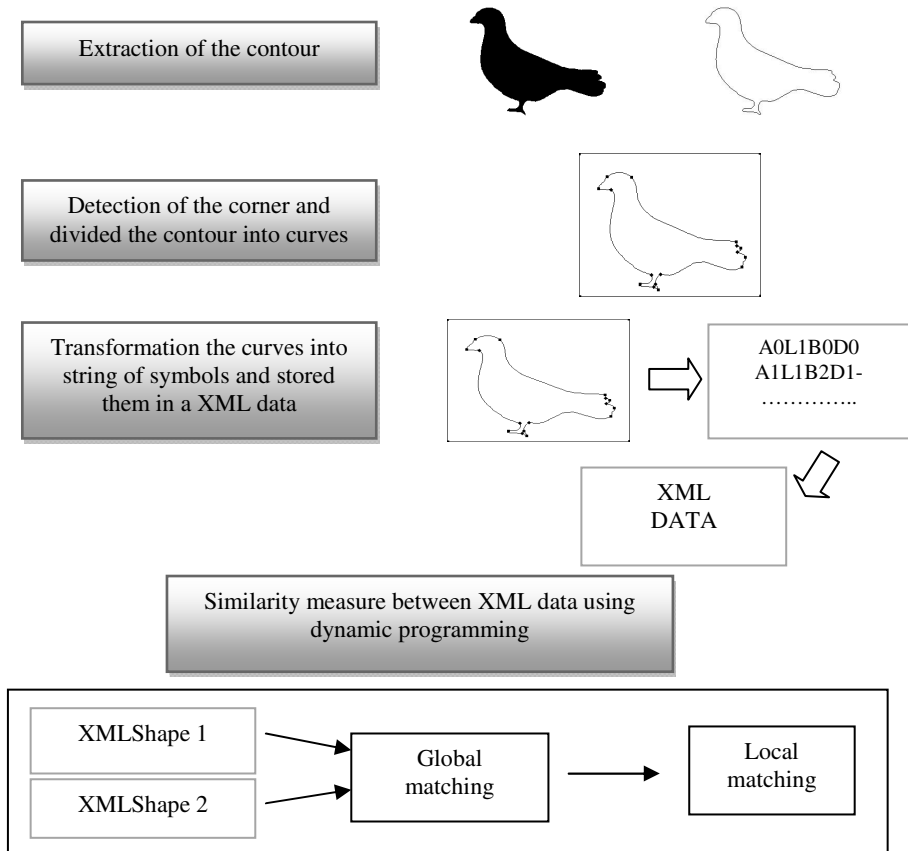


Fig. 1. A block representation of the proposed algorithm for shape matching

The first step is to analyze the contour of the shape to be studied. The contour is retrieved and represented by sets of points. After, the contour is divided into curves by using the technique of detection corner [11], this technique is detailed in section 2.1. Then each curve is transformed into a string of symbols. The string of symbols of each curve is stored in a XML file (Section 2.2). Finally, we use the technique of dynamic programming for computing the similarity between the set of symbols stored in XML file (Section 2.3).

2.1 Corner Detector

Corners in images represent critical information in describing object features that are essential for pattern recognition and identification. There are many applications that rely on the successful detection of corners, including motion tracking, object recognition, and stereo matching [9,10,11]. As a result, a number of corner detection methods have been proposed in the past. In this paper, we use an algorithm developed by Xiao Chen and Nelson H. C. Yung [12], it works in two passes and defines a corner in a simple and intuitively appealing way, as a location where a triangle of specified size and opening angle can be inscribed in a curve. The curve has to be generated previously using an edge detector. It is not required to be a closed curve. In the first pass the sequence of points is scanned and candidate corner points are selected. In each curve point p the detector tries to inscribe in the curve a variable triangle (p^- , p , p^+). The triangle varies between a minimum and a maximum square distance on the curve from p^- to p , from p to p^+ and the angle $\alpha \leq \alpha_{max}$ (the value of α_{max} is defined) between the two lines a and b in Figure 2. Triangles are selected starting from point p outward and the number of admissible triangles is defined. At a neighborhood of points only one of these admissible triangles is selected (See Fig.2)

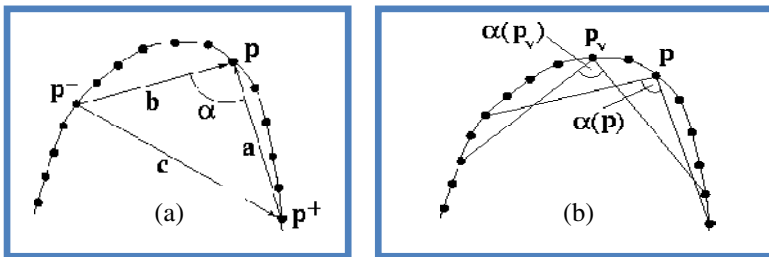


Fig. 2. Detecting high curvature points. (a) Determining if p is a candidate point. (b) Testing p for sharpness

2.2 Symbolic Representation and Storing XML

Our Approach for Symbolic Representation

The contour of the shape is retrieved and normalized to a set of points. This normalized contour used for feature extraction.

Our method uses local and global features to transform shape data into a new structure that supports measuring the similarity between shapes in an efficient manner,

using the corner detection, the contour is segmented into a set of primitives (line, convex and concave curves) and described by the features: A_i, l_i, Dg_i, β_i , where :

A_i is the area of the triangle enclosed the chord and the arc between the inflection points P_i and P_{i+1} , this area is calculated using Heron's formula (See Fig 2):

$$Area = \sqrt{s(s - a)(s - b)(s - c)} \tag{1}$$

Where: $s = \frac{a+b+c}{2}$ is the Semiperimeter, or half of the triangle's perimeter.

l_i is the length of Curve (C_i). Dg_i is the Degree of concavity or convexity ($Dg_i=d_i/l_i$) is computed as the ratio of the maximum of distances from points on the curve to associated chord and the distance of the chord of (C_i) .

β_i is the angle traversed by the tangent to the segment from inflection point P_i to inflection point P_{i+1} and shows how strongly a section is curved(See Fig. 3).

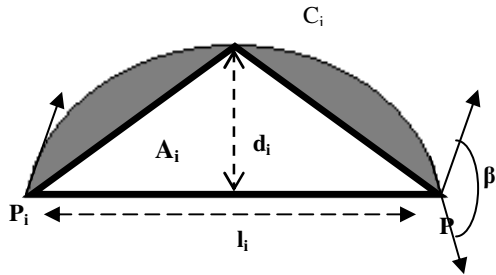


Fig. 3. shape descriptor of the curve C_i

Now we will see how to transform the shape into a set of symbols.

The area A_i is computed and quantized in three bins (A_0, A_1, A_2) corresponding to a zero, small and large area. The same, for each curve of the contour, the length of the segment $[P_i, P_{i+1}]$ is computed, this normalized distance l_i is quantized in three bins (L_1, L_2, L_3) corresponding to a small, medium and large distance of l_i . Next, the angle β_i is computed and quantized in different five bins between $[0, \pi]$ (B_0, B_1, B_2, B_3, B_4), B_0 for $\beta_i=0$. Finally, the values Dg_i of each curve is computed and quantized in three bins (D_0, D_1, D_2) corresponding to zero, a small and large values of Dg_i .

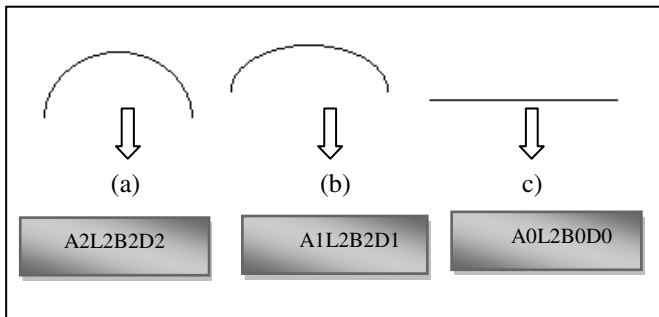


Fig. 4. The mapping obtained by the algorithm of a given contour into a string of symbols

Our algorithm converts the contour of the shape into sequences of symbols, for example the Mapping obtained by the algorithm of a curve (a) illustrated in figure 4 is : A2L2B2D2, A2 corresponding to a large area, L2 corresponding to a medium distance of L_i , β_i is quantized in the bin B2 and D2 corresponding to a large value of D_{g_i} .

Writing in XML Data

In this paper we propose XML language for describing the features of the shape structured in order where each contour is associated one descriptor written in a specific file XML. This technique for converting into XML is already used in our previous work [12].

To write the shape descriptor, each curve is represented by a set of parameters, these parameters are encoded using XML tags. to construct the XML structure we used the technique of DOM in Matlab. We use the syntax of XML to write our outline of the shape as follows:

The curve is defined by its type and described by the parameters (A_i , L_i , D_{g_i} , β_i)

- Concave/ Convex/Right curve

`< C number =' ' > // curve and its number in the outline`

`<Type > CC or CV or R <Type/> // Type of curve (concave/convex) or right line`

`<Ai> </Ai> area A_i`

`< li > length of curve`

`<beta_i> </ beta_i > angle β_i`

`< Dgi > </Dgi > degree of gravity`

`</C>`

An iterative process is presented to describe the shape using XML, this process is in the following algorithm:

Begin

Open an XML file F

NP: Compute the number of points in the contour;

NC: Compute number of curves

Storing NC and NP in XML file F

For i=1 to NC

$C(i) \leftarrow$ current curve

If ($C(i)$ is concave) then

 Compute the values $\{A_i, \text{Length } l_i, \text{Angle } i, D_{g_i}\}$

}

Else If($C(i)$ is convex) then

 Compute the values $\{A_i, \text{Length } l_i, \text{Angle } i, D_{g_i}\}$

}

Else // Right line

 Compute the value $\{\text{Length } l_i\}$

End If

Quantifying the values of A_i , l_i , i and D_{g_i} in its symbols.

Storing the symbols in an XML file (F).

End for(i)

Close the file F // Description symbolic of the shape is stored in F

End.

Our algorithm creates an XML file for each shape, for example the XML descriptor computed for curve (a) illustrated in figure 4 is:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<SHAPE>
<Name>Exemple d'un descripteur XML</Name>
<C Number="1">
<TYPE>CV</TYPE>
<A>A2</A>
<l>L2</l>
<beta>B2</beta>
<Dg>D2</Dg>
</C>
</SHAPE>
```

2.3 Matching Shape

Global Matching

At this level, we are interested only in global information's which characterize the general aspect of an object. At first, matching is done with comparing the number of different components of the outline shape descriptor.

From the XML descriptor it is easy to extract the following indices:

- Number of inflection points in the contour and Number of curves: Computed as the number of tag <C>.
- The order of each Curve defined by its number.
- Number of convex/ concave curves: computed as the number of type CV/ CC.
- Number of right lines: computed as the number of type R.

In some cases both of the contours have the same global descriptor; in this case the global matching is not possible. At the issue of global matching using an XML descriptor is the obtaining of a list of candidate couples of contours that constitute the input for the next step.

Using an XML descriptor reduces the execution time and research to a large database

Local Matching

After the outlines stored in XML files, their similarity can be evaluated by an appropriate comparison of each string of symbols stored in the tags <C>.

All string of symbols stored in the tags <C> be extracted from the XML file of the first shape and then compare them with other strings of the other XML file of the second shape.

We use the technique of dynamic programming for a good matching between strings of symbols. The dynamic programming can find the best alignment between two strings with different lengths. When sequences of strings are aligned, sequence

alignment scores are computed. The system can find similar sequences by sorting the alignment score. In this paper, we use the algorithm of Levenshtein Edit distance [14], this technique was modified by adding cost of similarity between the symbols. The edit distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is either an insertion, deletion, or a substitution of a single character.

We construct a matrix $D[0,--,m;0,--,n]$. The matrix D is computed using the recurrent equation:

$$D(i; j) = \min \left\{ \begin{array}{l} D(i-1; j-1) + F; //a \text{ substitution} \\ D(i-1; j) + w; //a \text{ deletion} \\ D(i; j-1) + w; //an \text{ insertion} \end{array} \right\} \quad (2)$$

$D(i,j)$ represents the score for the matrix position, W represents a gap of penalty score its value equal to "2" and F represents the match/mismatch score.

This is a dynamic programming algorithm in Matlab language:

```
for i = 1:n1
    D(i+1,1) = D(i,1) + DelCost;
end;

for j = 1:n2
    D(1,j+1) = D(1,j) + InsCost;
end;

for i = 1:n1
    for j = 1:n2
        if s1(i) == s2(j)
            Subst = 0;
        else
            Subst = SubstCost;
        end;
        D(i+1,j+1) = min([D(i,j)+Subst, D(i+1,j)+DelCost,
            D(i,j+1)+InsCost]);
    end;
end;
```

The first step for DP algorithm is to create a matrix with $M+1$ columns and $N+1$ rows where M and N correspond to the size of the sequences to be compared. This DP algorithm has been modified to take into account the differences resulting from the quantification of areas, distances and angles. A smaller weight or penalty (with a value lower than one) for the substitution of two adjacent symbols was introduced; for example the distance between $A1$ and $A2$ was taken to be equal to 0.5 and $A1$ and $A3$ equal to 1, and similarly the distance between $B1$ and $B2$ or $L1$ and $L2$ or $D1$ and $D2$ was taken to be equal to 0.5. The compute starting in the upper left hand corner in the matrix and finding the minimal score for each position in the matrix. The minimal score is calculated using the formula (2). Therefore, the algorithm helps the system avoid computing an exponentially large number of comparisons. When sequences of

strings are aligned, sequence alignment scores are computed. String sequences are matched well for lower alignment scores. The system can find sequences that are similar to a query key sequence and the minimum score is selected.

Consider the sequence of the shape (a) presented in figure 4 is a query key sequence and comparing him with sequences of the two shapes (b) and (c) respectively:

The score matrix for two cases is as follows:

Contour (a) with (b)					Contour (a) with (c)				
	A2	L2	B2	D2		A0	L2	B0	D0
A1	0.5	2	2	2	A1	0.5	2	2	2
L2	2	0.5	2.5	4	L2	2	0.5	2.5	4
B2	2	2.5	0.5	2.5	B2	2	2.5	1.5	3.5
D1	2	4	2.5	1	D1	2	4	3.5	2

Fig. 5. An example showing how to compute the edit distance between two strings. The last cell shows the distance computed for these two strings.

After filling the score matrix, the minimum alignment score for the sequence (a) with sequence (b) is 1 and the minimum alignment score for the sequence (a) with sequence (c) is 2, so the shape (a) is more matched with shape (b) .

3 Experiments

The method has been tested on a set of MPEG-7 and KIMIA-25 shapes illustrated by Fig.6. For each shape, contours of objects have been extracted. After that, we determine for each contour, the inflection points using corner detection. In Fig.7 we illustrate an example of dividing the contour of the shape into set of primitives (convex or concave curves or lines)

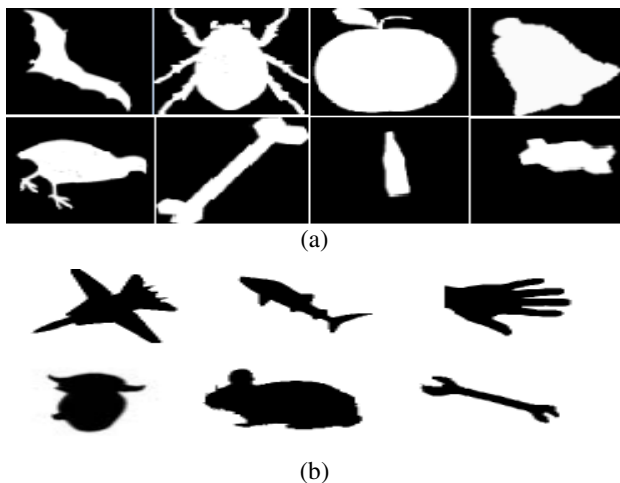


Fig. 6. Some of the objects in the (a) MPEG-7 database and (b) KIMIA-25 database

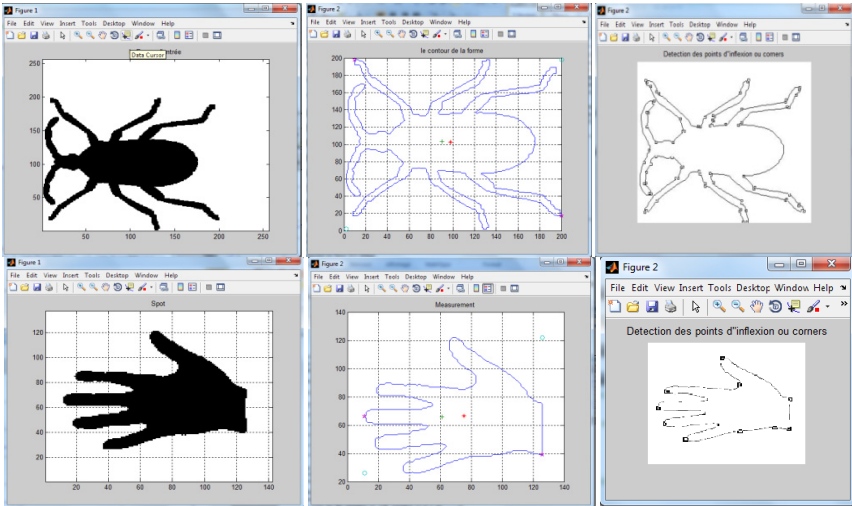


Fig. 7. Extraction the contour from two shapes in databases MPEG-7 and KIMIA-25 and detection the inflection points

The system creates an XML file for each shape and each shape is indexed by its xml file.

We took the XML file for some shapes and were used as reference XML files for experimentation; the number of reference XML files is defined as K. The percentage of matches between the reference XML and other files is obtained, and we computed the percentage of matching for different databases and different values of K.

The number of iteration using in this paper is 10 and the value of the parameter K is taking equal (3, 7, 15, 20, 50).

Results are presented as a percentage. The graph in figure 8 shows the score of matching for different values of K in two databases MPEG-7 and KIMIA-25.

The best match is achieved for $k = 20$ (98.7% for KIMIA-25) and (98.9% for MPEG-7). These results are compared with some old methods and techniques in MPEG-7.

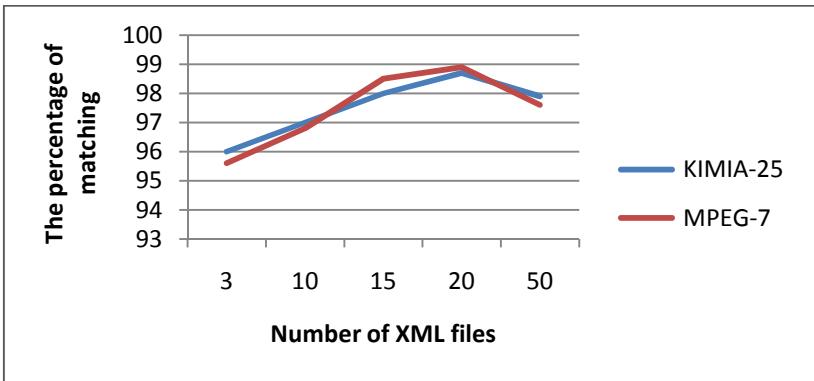


Fig. 8. Percentage of matching for five values of K in MPEG-7 and KIMIA-25

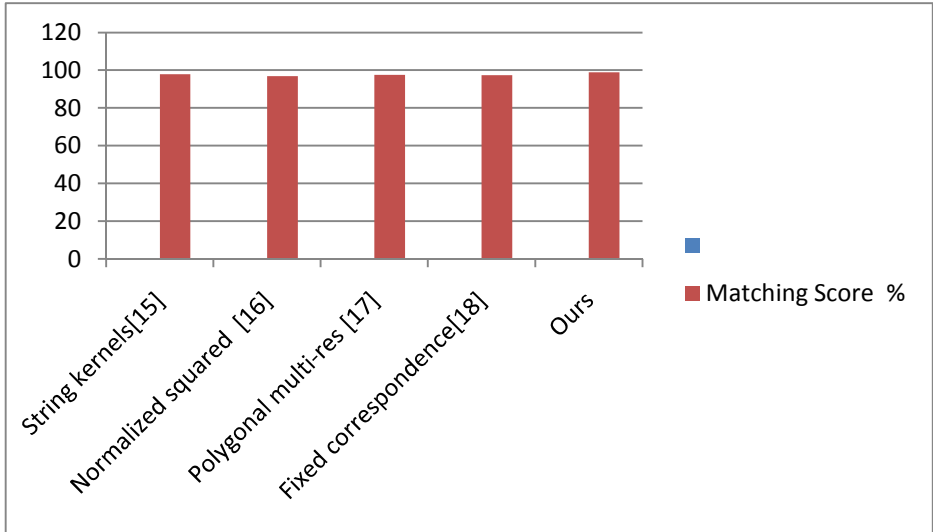


Fig. 9. Comparison of results in MPEG-7database

Our algorithm was compared with some old methods and our result is advanced with a little percentage compared to other solutions (Fig.9).

4 Conclusion

We have presented a new technique for shape matching. A key characteristic of our approach is the transformation of the shape features in XML file and then compare these XML files using dynamic programming. After different kinds of experimentation on MPEG -7 and KIMIA-25 Shape database, the proposed method has given interesting results over the existing methods.

References

1. Wolfson, H.J.: On Curve Matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 12, 483–489 (1990)
2. Kishon, E., Hastie, T., Wolfson, H.J.: 3-D Curve Matching using Splines. *J. of Robotic Systems* 8, 723–743 (1991)
3. Barequet, G., Sharir, M.: Partial Surface Matching by Using Directed Footprints. In: *Symposium on Computational Geometry*, pp. C-9–C-10 (1996)
4. Biederman, I., Ju, G.: Surface versus edge-based determinants of visual recognitions. *Cognit. Psychol.* 20, 38–64 (1988)
5. Geiger, D., Gupta, A., Costa, L.A., Vlontzos, J.: Dynamic Programming for Detecting, Tracking and Matching Deformable contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 17(3), 294–302 (1995)

6. Floreby, L.: A Multiscale Algorithm for Closed Contour Matching in Image Sequence. In: IEEE Intern. Conf. on Pattern Recognition, pp. 884–888 (1996)
7. Petrakis, E.G.M., Diplaros, A., Milios, E.: Matching and retrieval of distorted and occluded shapes using dynamic programming. IEEE Trans. Pattern Anal. Mach. Intell. 24(11), 1501–1516 (2002)
8. Yu, M.H., Lim, C.C., Jin, J.S.: Shape similarity using XML and portal technology. In: Visual Information Processing, VIP, Sydney, Australia (2006)
9. Fung, G.S.K., Yung, N.H.C., Pang, G.K.H.: Vehicle shape approximation from motion for visual traffic surveillance. In: Proc. IEEE 4th Int. Conf. on Intelligent Transportation Systems, pp. 201–206 (2001)
10. Manku, G.S., Jain, P., Aggarwal, A., Kumar, A., Banerjee, L.: Object tracking using affine structure for point correspondence. In: Proc. 1997 IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition, pp. 704–709 (1997)
11. Serra, B., Berthod, M.: 3-D model localization using highresolution reconstruction of monocular image sequences. IEEE Trans. Image Process. 6(1), 175–188 (1997)
12. He, X.C., Yung, N.H.C.: Corner detector based on global and local curvature properties. Optical Engineering 47(5), 057008-1–057008-12 (2008)
13. Gherabi, N., Bahaj, M.: A new shape descriptor using XML. IJCSSE 3, 1369–1376 (2011)
14. Ristad, E.S., Yianilos, P.N.: Learning string edit distance. IEEE Trans. Pattern Anal. Mach. Intell. 20(5), 522–532 (1998)
15. Daliri, M.R., Delponte, E., Verri, A., Torre, V.: Shape Categorization Using String Kernels. In: Yeung, D.-Y., Kwok, J.T., Fred, A., Roli, F., de Ridder, D. (eds.) SSPR 2006 and SPR 2006. LNCS, vol. 4109, pp. 297–305. Springer, Heidelberg (2006)
16. Super, B.J.: Learning chance probability functions for shape retrieval or classification. In: Proceedings of the IEEE Workshop on Learning in Computer Vision and Pattern Recognition (June 2004)
17. Attalla, E., Siy, P.: Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. Pattern Recognition 38(12) (2005)
18. Super, B.J.: Retrieval from shape databases using chance probability functions and fixed correspondence. Int. J. Pattern Recognition Artif. Intell. 20(8), 1117–1137 (2006)

Texture Classification Based on Lacunarity Descriptors

João Batista Florindo and Odemir Martinez Bruno

Instituto de Física de São Carlos (IFSC)
Universidade de São Paulo (USP)
Avenida do Trabalhador São-carlense, 400
13560-970 São Carlos SP Brazil
jbflorindo@ursa.ifsc.usp.br, bruno@ifsc.usp.br

Abstract. The present work presents a novel solution to provide descriptors of a texture image with application in the classification of such images. The proposed method is based on the lacunarity measure of an image. We apply a multiscale transform over the power-law relation of lacunarity and extract the descriptors from a window of the multiscale transform selected whose limits are determined empirically. We compare the classification accuracy of the proposed method with other state-of-the-art and classical texture descriptors found in the literature. We also do a brief theoretical summary of lacunarity definition, explaining its excellent performance comproved in the results.

Keywords: Pattern Recognition, Fractal Theory, Texture Descriptors, Lacunarity.

1 Introduction

Nowadays, we see a growing use of fractal theory in many applied areas, such as Biology [12,15], Medicine [16,13], Engineering [18], among many others. Indeed, fractal theory provides a solid and rich framework for the analysis of structures presenting some kind of self-similarity patterns. This is plentifully found in natural objects and scenarios, studied in natural and physical sciences.

An interesting aspect of fractal literature is that most applications is still based only on the fractal dimension concept. Despite the fact that this metric may model many problems with a good efficiency, it still suffers from serious drawbacks, for instance, the fact that quite distinct structures may present the same fractal dimension once it follows the same self-similarity law. Another point is the questionable efficiency of a single parameter dictating the whole model of a complex system. This situation is evident in digital image analysis, the problem focused here. In this application we usually find complex patterns, often turned still more complex due to noises and other artifacts we must deal with.

Lacunarity was defined in [8] as an alternative metric for fractal objects and posteriorly generalized to other “fractal-like” structures. Roughly speaking, while fractal dimension measures the spatial filling of a fractal, lacunarity

measures the spatial gapping of the same object. This measure is capable of distinguishing in an elegant fashion between two objects with the same fractal dimension but with different aspect. Thus, although literature still explores maidenly lacunarity concept [6,5,10], it is a worth complement of fractal dimension as a descriptor of structures, like those present in the images analyzed here.

For the same reason mentioned as a possible cause of failure of the fractal dimension use, the use of the simple lacunarity measure may demonstrate to be inefficient in most image analysis problems. With the goal of filling this gap, we propose the development and study of a novel approach capable of providing a set of descriptors based on the lacunarity measure. This is achieved by applying a space-scale transform to the lacunarity power-law associated to the self-similar aspect of the object. In this way, we obtain a group of measures representing the lacunarity computed over different scales, emphasizing at each scale, different patterns and sub-patterns, details and irregularities. This constitutes a rich source of information about the composition and pixel distribution inside the image. Moreover, the multiscale transform binds the mathematical fractal model to the biological visual system, once such system employs widely a multi-scale paradigm to extract details which will allow the discrimination of different objects.

The proposed methodology was tested in a discrimination task over the well-known Brodatz texture image dataset [3]. As waited from the theoretical context, the novel method achieved the best results when compared with classical texture descriptors methods. These results confirmed the efficiency of the proposed model as a powerful discriminator of objects even presenting a high level of complexity and noises inherent to the image generation process. Finally, the results point to the need for a deeper study of lacunarity concept in its possible applications in fractal modelling.

2 Lacunarity

Lacunarity is a concept defined in [8] to characterize fractal objects which despite having the same fractal dimension, present a quite dissimilar aspect, relative to their spatial distribution.

The literature shows a lot of algorithms employed in the estimation of lacunarity from the digital image representation of a shape [1]. Here we apply an adaptation of gliding-box method, originally proposed in [1] to the computation of lacunarity of texture images. The idea is initially to map the gray-level representation $I : [M, N] \rightarrow \mathfrak{R}$ onto a surface S , in the following way:

$$S = \{i, j, f(i, j) | (i, j) \in [1 : M] \times [1 : N]\}, \quad (1)$$

where:

$$f(i, j) = \{1, 2, \dots, I_M\} | f = I(i, j), \quad (2)$$

where I_M is the maximum gray-level intensity present in the image.

In the following step, we apply a three-dimensional version of gliding-box algorithm to the surface. In this method, we construct a rectangular prism $R : [1 : M] \times [1 : N] \times [1 : I_M]$ supporting the surface. Thus we divide such space into cubes with sidelength r , varying this value of r . For each r value, we can calculate the distribution $Q(s, r)$ corresponding to the mass probability distribution:

$$Q(s, r) = \frac{n(s, r)}{N(r)},$$

where $n(s, r)$ is the number of boxes, with side r , containing s points of the surface representing the object whose lacunarity we must estimate and $N(r)$ is the total number of box with side r . The number s is also known as the mass of the box. The Figure 1 illustrates the gliding-box process.

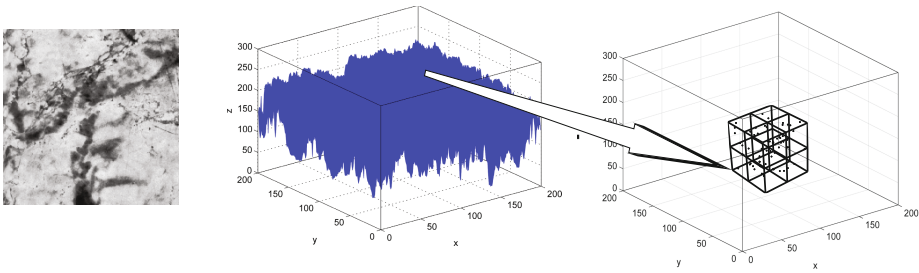


Fig. 1. Steps in the gliding-box method. From left to right, the original texture, the mapped 3D surface and the division of 3D space into cubes with sidelength r .

From the probability distribution, we may calculate the first and second moment Z_1 and Z_2 in a straightforward way:

$$Z_1(r) = \sum_{s=1}^{s_{max}} sQ(s, r)$$

and

$$Z_2(r) = \sum_{s=1}^{s_{max}} s^2Q(s, r).$$

From the moments we obtain the quotient $\Lambda(r)$:

$$\Lambda(r) = Z_1(r)/Z_2(r).$$

Finally, the lacunarity is calculated through the derivative:

$$\lambda = \frac{d \log(\Lambda(r))}{d \log(r)}.$$

In practice, the lacunarity is usually estimated by plotting the curve $\Lambda(r) \times r$ in a log-log scale and taking the slope of the straight line which may be fit to this curve.

3 Lacunarity Texture Descriptors

The proposed idea is based on the concept of fractal descriptors, presented in [47]. Thus, the method consists in computing the lacunarity of an object, here a texture image, at different scales and taking these values to compose the object descriptors.

More formally, the values at multiple scales may be represented as a function $u(t)$:

$$u(t) : \log(r) \rightarrow \log(\Lambda(r)),$$

where t now is the independent variable analog to $\log(r)$. In fractal descriptors approach, this function might be used directly or after some kind of specific post-processing depending on the particular application.

Here, in order to emphasize nuances in the function $u(t)$ which provides rich information of the texture image, we apply a multiscale transform to this function. Such kind of transform is represented by $U(b, a)$ where b is related to t and a is the scale at which the measure is taken. In this work, we use a particular multiscale method named space-scale. This is based on the derivative of $u(t)$ followed by the convolution with a Gaussian filter aiming at attenuating possible noises emphasized by the derivative. Then, the descriptors function \mathfrak{D} is provided through:

$$\mathfrak{D}(\sigma) = \frac{du}{dt} * g_\sigma(t).$$

In this expression, g_σ states for the well-known Gaussian function:

$$g_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-t^2/2\sigma),$$

where σ is the smoothing parameter.

Finally, the lacunarity descriptors are extracted from the set of values of \mathfrak{D} at a specific value of σ and thresholded at a specific point, once the last descriptors are more susceptible to noise influence. Both values of σ and the threshold delimiter τ of the descriptors are determined empirically in each particular application. The Figure 2 shows the potential of the proposed descriptors in discriminating among texture samples from different classes.

4 Experiments

The experiments to verify the efficiency of the proposed technique is carried out over a classical gray-level texture dataset, namely, the Brodatz basis [3]. This

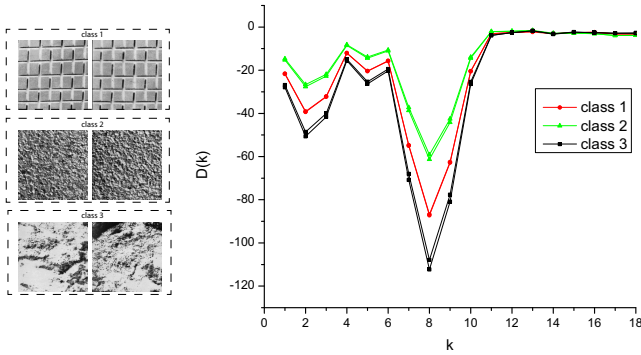


Fig. 2. Discrimination power of lacunarity descriptors. Three classes with 2 textures in each one and respective descriptors. Observe the visual distinction provided by the proposed descriptors.

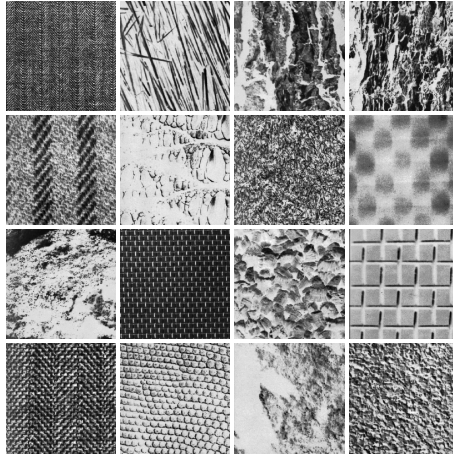


Fig. 3. Some texture samples from Brodatz dataset. Each image is from one different class.

is composed by photographs of natural scenes split into 200×200 windowed images. Each original photograph corresponds to a class and each window image to a sample. The databasis contains 111 classes with 10 samples in each one. The Figure 3 shows some image samples from the dataset.

We extract the proposed descriptors as well as other texture descriptors well-known in the literature from each image. Beyond the lacunarity descriptors, we also applied Gabor wavelets [14], Fourier [9], Bouligand-Minkowski [2], GLDM [17] and multifractal [11] descriptors. We classify such descriptors by K-Nearest

Neighbor classifier method, in a hold-out scheme, with $K = 1$, determined empirically. Finally, we compare the classification accuracy of each tested approach.

5 Results

The Table 1 shows the results in terms of success rate for each compared descriptor. For the proposed method, we used $\sigma = 0.1$ and a threshold $\tau = 18$. We notice that lacunarity descriptors presented a significant advantage over the state-of-the-art Gabor wavelets descriptors. Another interesting point is that the proposed method uses only 18 descriptors. This reduced amount constitutes an important statistical and computational advantage once turns possible a faster computational performance and attenuates significantly any effect related to the dimensionality curse, when a large number of descriptors dissipates the discrimination ability.

Table 1. Correctness rate for Brodatz dataset

Method	Correctness Rate (%)	Number of descriptors
Gabor	81.2613	20
Fourier	63.7838	74
GLDM	52.2523	20
Multifractal	35.1351	101
Bouligand-Minkowski	47.5676	85
Proposed method	85.5856	18

Actually, the good performance of lacunarity descriptors was predictable, given that lacunarity is a fundamental measure associated to fractal characteristics broadly present in real-world images. As we are dealing with objects which are not real fractals strictly speaking, the lacunarity presents, in some sense, an irregular behavior along different scales. Thus, the multiscale transform highlights this aspect of imperfect power-law, providing, in this way, a valuable information of levels of lacunarity along the scales of the image. Ultimately, such lacunarity scale pattern is directly related to the psycho-visual and physical characteristics of the object represented in the texture image. So, the discrimination power is an immediate consequence of such inherent characteristics.

Finally, in the Figure 4 we show the confusion matrices for the two best methods, that is, Lacunarity and Gabor descriptors. In this figure, each point color corresponds to the number of samples pertaining to the class in vertical axis and classified as being from the class in the horizontal axis. So, the success predictions are represented in the principal diagonal while the errors are outside the diagonal. Observe that, in this case, both matrices are not so different, but the Gabor matrix shows a greater number of brighter points outside the diagonal, indicating a higher number of missclassifications, confirming the results in the Table 1.

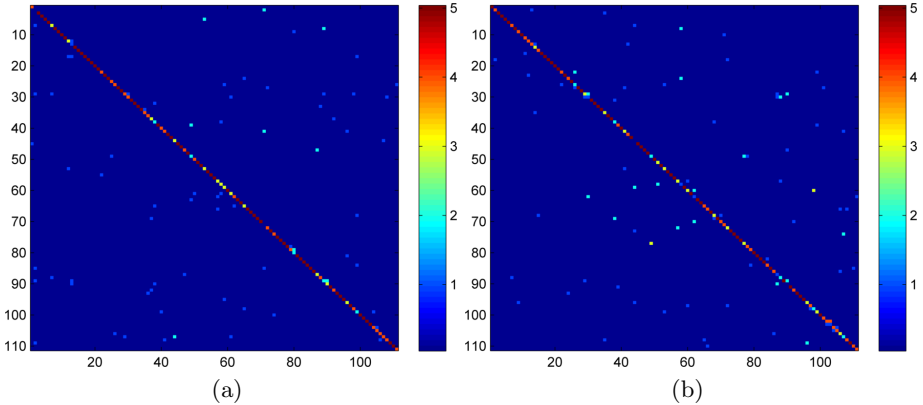


Fig. 4. Confusion matrices of the methods presenting the best performances. At left, the proposed lacunarity descriptors. At right, Gabor descriptors. Each color represent the number of samples pertaining and assigned to a class, following the colorbar notation.

6 Conclusion

This work proposed a novel texture descriptor technique based on the concept of lacunarity. We obtained such descriptors by applying a multiscale transform to the lacunarity computation, followed by selecting empirically a region from the multiscale response.

We compared the efficiency of the novel method with other classical and state-of-the-art texture descriptors in the classification of a benchmark dataset. The proposed descriptors presented the best performance in terms of classification accuracy. The results confirm the expectation from fractal theory. Indeed, the proposed technique demonstrates in practice that has a large potential for modelling, discriminating and describing the most complex patterns present in a real-world image.

The efficiency of the novel descriptors encourages to a deeper research for the properties of lacunarity measure in the analysis of digital images. Besides, it is comproved that lacunarity descriptors have a large potential to be tested in applications involving image pattern recognition and computer vision in many areas of the science.

Acknowledgements. J.B.F. acknowledges support from CNPq (National Council for Scientific and Technological Development, Brazil) (Grant 140624/2009-0). O.M.B. acknowledges support from CNPq (Grant 308449/2010-0 and 473893/2010-0) and FAPESP (Grant 2011/01523-1).

References

1. Allain, C., Cloitre, M.: Characterizing the lacunarity of random and deterministic fractal sets. *Phys. Rev. A* 44, 3552–3558 (1991)
2. Backes, A.R., Casanova, D., Bruno, O.M.: Plant leaf identification based on volumetric fractal dimension. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)* 23(6), 1145–1160 (2009)
3. Brodatz, P.: *Textures: A photographic album for artists and designers*. Dover Publications, New York (1966)
4. Bruno, O.M., de Oliveira Plotze, R., Falvo, M., de Castro, M.: Fractal dimension applied to plant identification. *Information Sciences* 178(12), 2722–2733 (2008)
5. Dong, P.: Test of a new lacunarity estimation method for image texture analysis. *International Journal of Remote Sensing* 21(17), 3369–3373 (2000)
6. Feagin, R.A.: Relationship of second-order lacunarity, Hurst exponent, Brownian motion, and pattern organization. *Physica A: Statistical Mechanics and its Applications* 328(3-4), 315–321 (2003)
7. Florindo, J.B., De Castro, M., Bruno, O.M.: Enhancing Multiscale Fractal Descriptors Using Functional Data Analysis. *International Journal of Bifurcation and Chaos* 20(11), 3443–3460 (2010)
8. Gefen, Y., Meir, Y., Mandelbrot, B.B., Aharony, A.: Geometric Implementation of Hypercubic Lattices with Noninteger Dimensionality by Use of Low Lacunarity Fractal Lattices. *Physical Review Letters* 50(3), 145+ (1983)
9. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Prentice Hall, Upper Saddle River (2002)
10. Greenhill, D.R., Ripke, L.T., Hitchman, A.P., Jones, G.A., Wilkinson, G.G.: Characterization of suburban areas for land use planning using landscape ecological indicators derived from IKONOS-2 multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 41(9), 2015–2021 (2003)
11. Harte, D.: *Multifractals: theory and applications*. Chapman and Hall/CRC (2001)
12. Lebedev, D., Filatov, M., Kuklin, A., Islamov, A., Kentzinger, E., Pantina, R., Toperverg, B., Isaev-Ivanov, V.: Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties. *FEBS Letters* 579(6), 1465–1468 (2005)
13. Lorthois, S., Cassot, F.: Fractal analysis of vascular networks: Insights from morphogenesis. *Journal of Theoretical Biology* 262(4), 614–633 (2010)
14. Manjunath, B., Ma, W.: Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 837–842 (1996)
15. Quevedo, R., Jaramillo, M., Diaz, O., Pedreschi, F., Miguel Aguilera, J.: Quantification of enzymatic browning in apple slices applying the fractal texture Fourier image. *Journal of Food Engineering* 95(2), 285–290 (2009)
16. Tian-Gang, L., Wang, S., Zhao, N.: Fractal Research of Pathological Tissue Images. *Computerized Medical Imaging and Graphics* 31(8), 665–671 (2007)
17. Weszka, J., Dyer, C., Rosenfeld, A.: A comparative study of texture measures for terrain classification. *SMC* 6(4), 269–286 (1976)
18. Wool, R.P.: Twinkling Fractal Theory of the Glass Transition. *Journal of Polymer Science Part B-Polymer Physics* 46(24), 2765–2778 (2008); Annual Meeting of the American-Physical-Society, New Orleans, LA, March 10 (2008)

Real-Time Fall Detection Method Based on Hidden Markov Modelling

Alban Meffre, Christophe Collet, Nicolas Lachiche, and
Pierre Gançarski

University of Strasbourg-CNRS, LSIIT UMR 7005
Pôle API - Bd Sébastien Brant - BP 10413
67412 Illkirch CEDEX - France
{ameffre,c.collet,nicolas.lachiche,gancarski}@unistra.fr

Abstract. In the next few decades the increase of the number of elderly people will be of major concern, so that solutions must be found in order to maintain them at home. However such a population is exposed to the risk of falls, that can lead to dependency. This paper recalls some approaches used for fall detection and focuses on a method based on an uncalibrated camera. Motion detection uses a combination of simple Gaussian background modelling and interframe difference for person shape detection and features extraction. These features feed a Hidden Markov Model dedicated to fall detection. The algorithm has been tested on real data and we show that simple techniques can be used in order to obtain a fast and reliable fall detection system.

1 Introduction

1.1 Context

In the future the western European and American countries will have to meet the important challenge of elderly care. In 2050 the number of elderly people will rise from 9% to 15% in France, thanks to the stability of birthrate and the increase of life expectancy. Thus it will be necessary to increase the number of old people homes, and/or allow them to stay at home as long as possible.

Elderly people need to keep their partial autonomy in order to stay at home and falling is among the risks which can make them become dependent. After a fall elderly people often end up in hospital, and if not seriously injured, they lose their self-confidence and become more exposed to falling again. The aggravating factor with the fall is the time during which a person stays on the floor. The longer a person lies, the worse his situation becomes with a high risk of mortality.

Today a lot of devices exist for warning when elderly people fall. These devices can be automated and based on various technologies, such as accelerometers and inclinometers¹, or infrared motion sensors². Some devices are just remote alarm

¹ <http://www.fallsaver.net/>

² <http://www.pervaya.com/produit.html>

control and the person needs to manually activate the device after a fall, provided that he remains conscious, is able to move, and the device is properly worn.

Along with the expansion of information technologies and wireless communication systems, elderly people's care at home becomes an important subject for researchers and every improvement in this domain is a step to overcoming the problem of dependency. Over the past few years many studies have been carried out concerning fall detection by means of sensors integrated within the home.

In this way, the fall detection by video camera can be achieved by different means. Some methods are based on 3D reconstruction but need the use of one or several calibrated cameras. Nevertheless we will focus in this paper on 2D video image analysis with a single uncalibrated camera, for simplicity and cost reasons.

We present in the following different approaches related to single camera fall detection, whereas Sect. 2 presents video processing and detection with a Markovian model. Next we give the overview of the method we use and then we validate it on real data (Sect. 3). Finally we conclude with a discussion about the possible improvements of our work.

1.2 Related Work

The first step for fall detection consists in detecting the motion using a camera. The interframe difference is a straightforward method that consists in building a binary mask with the thresholded absolute difference between two subsequent images. It is used by authors such as Hsieh et al. [9] or Lee et al. [11] in combination with another background subtraction method. The simple Gaussian background subtraction method, used by Töreyin et al. [16] and Dahmane et al. [6], is another fast and straightforward method and usually sufficient for indoor scenes. Easy to implement, this method does not take into account a moving background. Some authors use the mixture of Gaussian by Stauffer [14] or the codebook background subtraction by Kim et al. [10]. These methods are more suitable for outdoor scenes. Hsieh et al. [9] combine image difference with a simple Gaussian background, and Lee et al. [11] combine it with a Gaussian mixture. Shadow suppression can be done in the HSV color space [4] but is time consuming. Shadow suppression methods in RGB color space are presented by Dahmane et al. [6] and Kim et al. [10]. Tracking can be implemented in order to follow several people in the scene. Lee et al. [11] use a variant of the Kalman filter. Tracking can be useful for occlusion handling but it requires complex models [5].

The second step is the feature extraction needed for fall detection. This is performed on the binary mask obtained with motion detection or background segmentation. The straightforward method consists in surrounding the moving shape (or blob) with a bounding box as in [17][11][7], and/or a fitted ellipse [12][7]. Thus the height to width ratio and inclination angle are retrieved. In

Rougier et al. [13] the 3D spatio-temporal trajectory of the head is used for detecting the fall. Sound can be used as well for detecting events [16] and Foroughi et al. [8] use projection histograms, along with Discrete Fourier Transforms (DFT). These methods require dimensionality reduction by Principal Component Analysis (PCA) for easier handling and computing. In certain cases some information about motion quantity is needed. The Motion History Image (MHI) and Motion Energy Image (MEI) were introduced by Bobick et al. [3].

The final decision taking of fall detection systems is made by a classification algorithm. Neural networks can be used [7], as well as Support Vector Machines [8]. These techniques need a supervised learning. Hidden Markov Models (HMM) are well suited for fall detection based on video. Anderson et al. [1] use one HMM per posture detection, and Töreyn et al. [16] detect falls from both audio and video.

2 Fast Fall Detection Method

The general principle of our algorithm (Fig. 1) is separated into two distinct parts. The first (Sect. 2.1) is based on a simple Gaussian background subtraction and interframe difference based motion detection, followed by a shape analysis and feature extraction with a fitted ellipse and a bounding box. In the second one (Sect. 2.2) the features are undersampled and classified with a Bayesian detection process, and the label sequence combined with ground truth for performance evaluation.

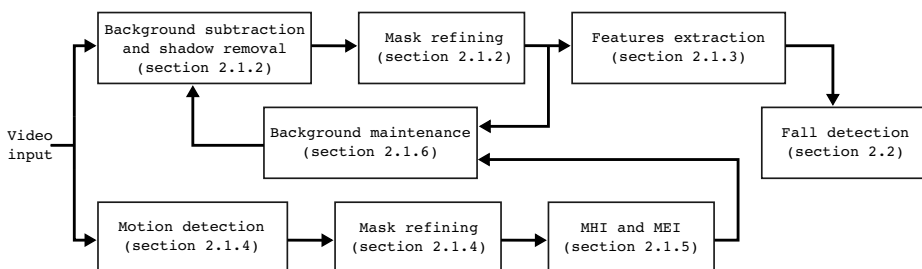


Fig. 1. Diagram of the video analysis

2.1 Video Processing and Feature Extraction

2.1.1 Initialization

Both the background model and interframe difference need to be initialized at program start. This is done without any movement in the scene, because only pixel noise has to be measured.

The background model initialization consists in accumulating the average and variance of input video images over time [6], in our work we choose the learning rate $\alpha = 0.95$ and thus initialization requires a few tens of images.

We assume that global illumination remains constant over time or changes very slowly. Cases with strong illumination changes are special cases that need to be processed separately and induce resetting the background model.

2.1.2 Background Subtraction and Mask Refining

The Gaussian background subtraction method results in a binary mask. The background maintenance is achieved by computing the temporal average and variance of the pixel intensities over the three RGB channels (1), (2). Let $\alpha \in]0 ; 1[$ represent the learning rate at which the input image I_n is added into the background model B_n . V_n can be considered as the variance of pixel s values in each RGB channel. The background model is updated recursively depending on the mask \mathcal{UM} .

$$B_n(s) = \begin{cases} \alpha \cdot B_{n-1}(s) + (1 - \alpha) \cdot I_n(s) & \text{if } s \notin \mathcal{UM} \\ B_{n-1}(s) & \text{if } s \in \mathcal{UM} \end{cases} \quad (1)$$

$$V_n(s) = \begin{cases} \alpha \cdot V_{n-1}(s) + (1 - \alpha) \cdot (B_n(s) - I_n(s))^2 & \text{if } s \notin \mathcal{UM} \\ V_{n-1}(s) & \text{if } s \in \mathcal{UM} \end{cases} \quad (2)$$

The segmentation is done by comparing each pixel s in the input image I_n with the background model B_n in the three RGB channels. If the difference is greater than 3 times the standard deviation $\sqrt{V_n(s)}$ for at least one of the RGB channels, the pixel is marked as foreground in the segmentation mask, and background otherwise. A pixel marked as foreground can correspond to a shadow. The shadow detection follows the method described in [6], wherein thresholds are respectively 0.44 and 1.09 for the lower and higher brightness distortion limits, and 7 for the color distortion.

Noise removal is performed by an erode-dilate morphological operation using a 3 by 3 square structural element and all the blobs that have an area smaller than 300 pixels are removed.

2.1.3 Feature Extraction

The features needed for fall detection are extracted in the mask (Fig 2). The fitted ellipse gives the angle α_n of the main axis of the blob, and the bounding box around the blob gives the height to width ratio ρ_n . The topmost point of the blob gives the displacement speed Vh_n of the head of the person. After the feature extraction the blob is replaced by its convex hull by redrawing the mask.

2.1.4 Motion Detection and Mask Refining

To detect the motion we use the binarized absolute image difference D_n . Let G be the gain and Ig_n the gray levels input image, so $D_n = (G \cdot |Ig_n - Ig_{n-1}|) \begin{matrix} \geq 255 \\ 0 \end{matrix}$

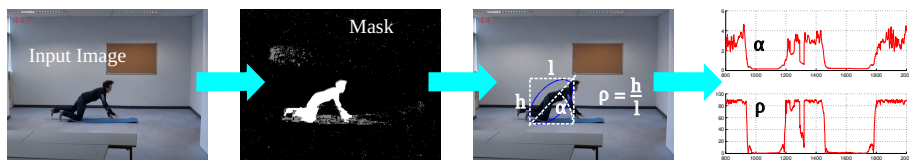


Fig. 2. Extraction of the features α_n and ρ_n

128. The interframe difference shows zones that have changed, i.e. only the outline of a moving object appears in the mask. The empty zones left are partially filled by expanding the outline with a dilate operation performed at quarter resolution by double down-sampling with Gaussian filtering. Then the mask is up-sampled twice toward the original resolution. The gain G is initialized in order to keep a ratio of 0.1% of white pixels when there is no movement in the scene.

2.1.5 MHI and MEI

The refined motion mask is used for updating the MHI (Bobick et al. [3]). The decay rate between two subsequent images is of 10 gray levels. The MEI is obtained by thresholding the MHI at gray level 1.

2.1.6 Motion Quantification and Background Maintenance

In Sect. 2.1.2, the background is updated with respect to the update mask \mathcal{UM} . When motion is present in the scene, the MEI is copied to the update mask \mathcal{UM} and thus the background model “absorbs” slight changes in the scene like a previously displaced object. When no motion is present in the scene the background subtraction mask is copied to \mathcal{UM} , preventing the stationary person becoming integrated in the background model.

The motion is quantified by measuring the ratio m of the number of white pixels in MEI to the number of pixels in the background subtraction mask. The threshold between the two cases is set to 2.

2.2 Fall Detection

The fall detection is achieved in four steps and separately from the video processing. The whole feature data is processed at once after it has been extracted from the video sequences (Fig 3).

2.2.1 Pre-filtering of the Features

The video processing and feature extraction provide features at a rate of 25 datasets per second. Actually we do not need to detect falls at such a rate so the features are filtered and under-sampled before the detection. The angle α_n and height to width ratio ρ_n are averaged over 50 samples every 25 samples and hence we obtain α_t and ρ_t at a rate of 1 dataset per second. In order to

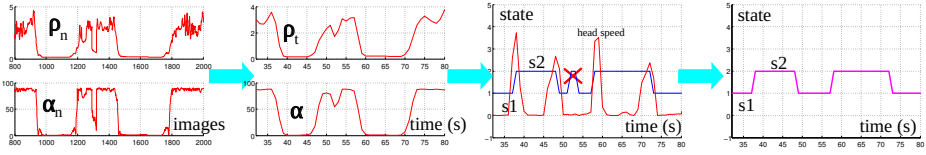


Fig. 3. Data pre-processing, classification and post-processing of states S_1 = "non-fall" and S_2 = "fall"

remove spurs the speed of the head Vh_n is first filtered by a Butterworth filter using the general formula $y_n = b_0x_n + b_1x_{n-1} - a_1y_{n-1}$ where $a_1 = -0.7265$ and $b_0 = b_1 = 0.13675$, then $\max(Vh_n)$ is kept over 50 samples every 25 samples so that we obtain Vh_t at a rate of one dataset per second.

2.2.2 Detection

In our work we assume that the postures of a person with time follows a Markov process. The observation is formed by the ρ_t and α_t information and the two hidden states (or classes) are labelled "fall" and "non-fall" with two parameters per class corresponding to the average and standard deviation of the normal distribution of the observation given the class. The labeling process is performed using the Baum-Welch procedure [2] and an Expectation Maximization (EM) algorithm [15] is used for the hyper-parameters estimation. These hyper-parameters are initialized with $\{\mu_\rho = 3; \sigma_\rho^2 = 1; \mu_\alpha = 90; \sigma_\alpha^2 = 10\}$ for the "non-fall" label and $\{\mu_\rho = 0; \sigma_\rho^2 = 1; \mu_\alpha = 0; \sigma_\alpha^2 = 100\}$ for the "fall" label, where μ and σ^2 are the average and variance of the features. Each iteration of the loop is performed as following. The forward probabilities are calculated using scaling. Then the backward probabilities are calculated and the product of forward and backward variables is maximized in order to retrieve the hidden states chain. This is the "Maximization" step. The Log-Likelihood of the data driven term is calculated by summing the scaling variables over the whole chain. The "Expectation" step consists in re-evaluating the parameters of the Markov model. First the gaussian parameters of the data driven term are evaluated given the new sequence of hidden states (i.e. the average values and covariances matrices), and second the transition probabilities and state initial probabilities are calculated.

The loop of the EM algorithm ends when the Log-Likelihood has converged to a steady value or when 100 iterations have been performed. Then the sequence of labels that maximize the likelihood of each observation is returned.

This procedure returns the maximum a posteriori hidden states sequence given the observation $\{\rho_t; \alpha_t\}$.

2.2.3 Data Post-Processing and Performance Evaluation

The classification takes into account only static labels such as “non-fall” and “fall”. Hence the speed of the head Vh_t and the duration of a posteriori occurrence of “fall” states are used after classification in order to remove false positives. A true positive is obtained when the duration of the occurrence of the “fall” state is greater than 5 seconds, and when the speed Vh_t is greater than 5 at the very beginning of the state occurrence.

As the video sequences are manually annotated, the result of the classification process is compared with the ground truth. This is done by combining them with a finite state machine, in order to prevent multiple detection of a fall during an actual “fall” state. So the real number of actual falls in video sequences is estimated.

3 Experimental Results

The video processing and feature extraction part of the program is written in C++ using the *OpenCV2.2* library and the classification and fall detection is performed using *Matlab*. The method has been tested on a set of 24 video clips available at <http://lsiit-miv.u-strasbg.fr/lsiit/perso/collet/Images/videosENSPS/>. The videos have a standard format of 640×480 pixels at a framerate of 25 images per second. We obtained a 34 images per second process rate on an *intel core i5 750* processor. Different kinds of situations have been simulated with the help of actors, such as walking, falling with different angles, bending down, sitting on a chair, carrying an object, etc. These actions are combined with lighting variations and partial occlusions.

We performed the classification on the whole video set and we obtained 90% sensitivity and 100% specificity. Some falls have not been detected because our system is not able to deal with complete occlusion. If we do not take into account the speed of the head Vh_t , the specificity drops to 96%. Therefore the speed of the head helps to eliminate false positives.

The Fig 4 gives examples of falls properly detected. Even if the fall occurs in the camera axis, the fall is detected as long as the head speed is high enough. The fall will be detected too if the fall occurs in another direction and the body is masked by small object only, so the angle of the fitted ellipse will correspond to the angle of the entire body. Small object displaced in the scene will do not affect the behavior of the background segmentation algorithm because they do not move once put down, so they are quickly absorbed in the background model. However strong illumination changes or hard shadows can lead to false positives or false negatives because in these cases the fitted ellipse surrounds both the body of the person and the undesirable artifacts.

An example of fall positive is given with Fig 5 where the person fall behind a big object. As the body is partially or totally hidden, the ellipse fitting can not perform properly.

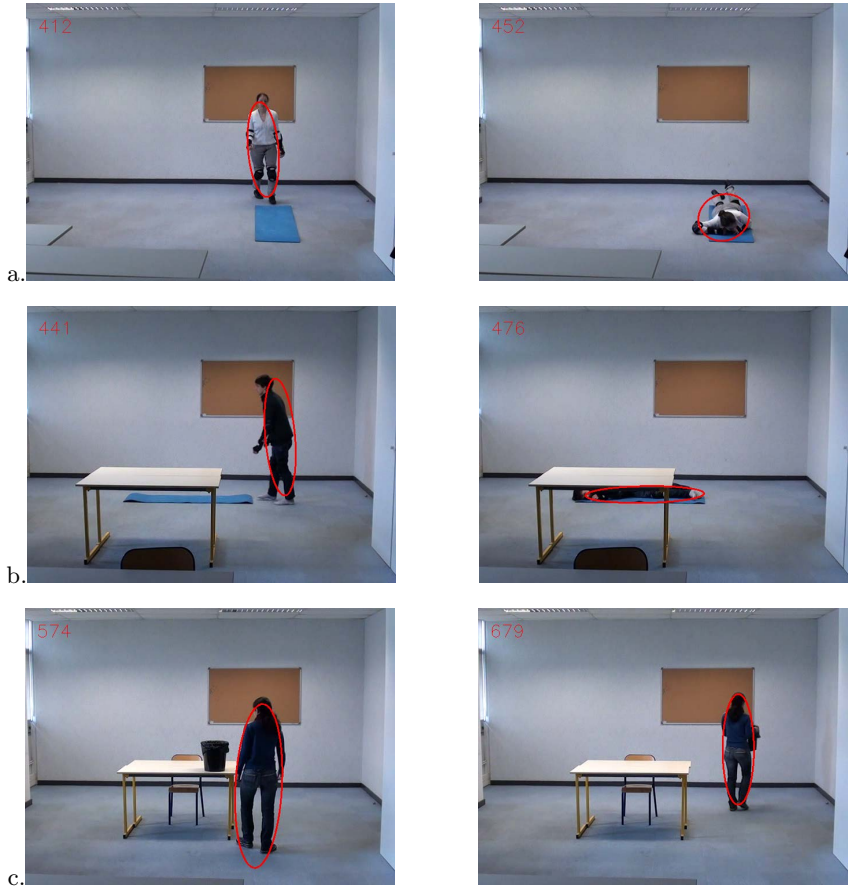


Fig. 4. Examples of true positives and true negatives: falling in the camera axis(a), falling with slight occlusion(b) carrying an object(c)



Fig. 5. Example of false negative: falling with strong occlusion

4 Discussion and Conclusion

After we read the different papers related to our work, we retain that the complexity of the methods varies a lot depending on which goal the authors expected to reach. If we want to follow several people in the scene or deal with strong occlusions, hard shadows and strong illumination changes, we have to implement tracking and/or pattern recognition algorithms. These techniques tend to use a lot of computing and memory resources.

In our case the aim was to verify that we can implement a reliable fall detection system with robust and fast techniques and thus we use the simple background modelling and motion detection, in combination with a Bayesian classification. Therefore we show that combining them wisely gives good classification results and a high video processing rate. In the meantime a video system cannot be used alone for reliable care of elderly people. Good performance and security can be achieved by a use of our algorithm in combination with other external sensors integrated into the home environment such as Passive-InfraRed (PIR) detectors.

We would like to thank Region Alsace for funding and partial support of this work.

References

1. Anderson, D., Keller, J.M., Skubic, M., Chen, X., He, Z.: Recognizing falls from silhouettes. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2006, pp. 6388–6391 (2006)
2. Baum, L.E.: An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities* 3, 1–8 (1972)
3. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(3), 257–267 (2001)
4. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with HSV color information. In: Proc. IEEE Intl. Conf. Intelligent Transportation Systems, pp. 334–339 (2001)
5. Cucchiara, R., Vezzani, R.: Assessing temporal coherence for posture classification with large occlusions. In: IEEE Workshop on Motion and Video Computing, WACV/MOTIONS 2005, vol. 2, pp. 269–274 (2005)
6. Dahmane, M., Meunier, J.: Real-time moving object detection and shadow removing in video surveillance. In: 3rd International Conference of SETIT, Tunisia (2005)
7. Froughi, H., Aski, B.S., Pourreza, H.: Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: 11th International Conference on Computer and Information Technology, ICCIT 2008, pp. 219–224 (2008)
8. Froughi, H., Rezvani, A., Pazirae, A.: Robust fall detection using human shape and multi-class support vector machine. In: Sixth Indian Conference on Computer Vision, Graphics & Image Processing, ICVGIP 2008, pp. 413–420 (2008)
9. Hsieh, C.-C., Hsu, S.-S.: A simple and fast surveillance system for human tracking and behavior analysis. In: Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, SITIS 2007, pp. 812–818 (2007)

10. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. *Real-Time Imaging* 11(3), 172–185 (2005)
11. Lee, Y.-S., Lee, H.J.: Multiple object tracking for fall detection in real-time surveillance system. In: 11th International Conference on Advanced Communication Technology, ICACT 2009, vol. 03, pp. 2308–2312 (2009)
12. Nait-Charif, H., McKenna, S.J.: Activity summarisation and fall detection in a supportive home environment. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 4, pp. 323–326 (2004)
13. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Monocular 3d head tracking to detect falls of elderly people. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2006, pp. 6384–6387 (2006)
14. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, p. 252 (1999)
15. Tanner, M.A.: Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions. Springer (1996)
16. Ugur Toreyin, B., Dedeoglu, Y., Enis Cetin, A.: HMM based falling person detection using both audio and video. *Computer Vision in Human-Computer Interaction*, 211–220 (2005)
17. Vishwakarma, V., Mandal, C., Sural, S.: Automatic Detection of Human Fall in Video. In: Ghosh, A., De, R.K., Pal, S.K. (eds.) *PRMI 2007*. LNCS, vol. 4815, pp. 616–623. Springer, Heidelberg (2007)

Extracting Buildings by Using the Generalized Multi Directional Discrete Radon Transform

I. ELouedi¹, A. Hamouda¹, H. Rojbani¹, R. Fournier², and A. Nait-Ali²

¹The Sciences institute, Computer technology department, Tunis, Tunisia
ineselouedi@yahoo.com, atef_hamouda@yahoo.fr,
hmida.rjb@gmail.com

²The University Paris Est-Creteil, Laboratory of Images,
Signal and Intelligent systems(LISSI), Paris, France
{rfournier,naitali}@u-pec.fr

Abstract. This paper presents a new method to detect and accurately locate a rectangular form object in any given image. In order to find the right coordinates of those objects in the image, we develop the Generalized Multi Directional Discrete Radon Transform (GMDRT). The GMDRT can detect any given shape whatever its form and orientation are. Experimental results on high resolution QuickBird image to extract rectangular buildings form show the efficiency of our method.

Keywords: Generalized multi Directional Discrete Radon Transform, High-Resolution QuickBird images, Rectangular Buildings.

1 Introduction

The recognition of objects whatever its sizes, scales, positions or orientations in images like humans do, is still a challenge for computer vision systems. In recent years, the Radon Transform has received much attention. This transform projects a two dimensional image along straight lines within different directions and then transforms the image into a parameters space where each line in the initial image gives a peak positioned at the corresponding line parameters [1]. This have lead to many line detection applications on image processing [3, 4,8], medical imaging [5] and seismic applications [2]. The Radon Transform was also widely used in satellite image area such as ship wakes detection [5,6], or buildings detection[14].

Especially, the authors in [9] extract buildings from high resolution images by applying the classical Radon Transform. Then, they use the Forstner operator in the Radon transform parameters space to detect peaks. The inconvenience of this approach is its dependence on the buildings size to detect peaks. In addition, this method needs a post-treatment to extract building contours due to the use of the classical Radon Transform which detects only lines.

Here in, we use our Generalized Multi Directional Discrete Radon Transform (GMDRT) to extract directly the rectangular form buildings from an

image. The extraction problem is reduced only to peaks selection without any post-treatment. The paper is outlined as follows: Section 2 presents the definition of the GMDRT approach. Then, section3 details the process of the buildings extraction from high resolution images. Section 4 presents and evaluates experimental results. Finally, we summarize our research and conclude the paper in Section 5.

2 Generalized Discrete Radon Transform

The Generalized Multi Directional Radon transform (GMDRT), as detailed in [13], is defined to detect from the initial image geometric objects by precisely locating its positions, its parameters and its rotation angles. The GMDRT method consists of projecting the image over parameterized curves ϕ that are rotated according to an angle θ and translated according to both the horizontal and the vertical axis as it is shown in figure1:

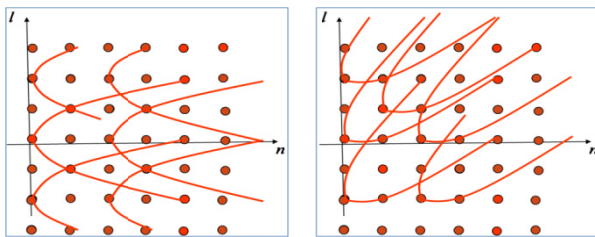


Fig. 1. (a): the parabolic curves $\phi(l)=l^2$ follow the horizontal direction (b):the rotated parabolic curves follow the $\pi/4$ direction .

The GMDRT is an algebraic exactly invertible method inspired from [2] based on the multiplication of a selection matrix $R_{m,\theta}$ and an image column $I(n)$. The transform matrix $R_{m,\theta}$ is defined to select from a column $I(n)$, pixels belonging to a θ -rotated curve. Let the following expression be the GMDRT formalism:

$$y_\theta(n) = \sum_{m=-M}^{m=M} R_{m,\theta} \times I(n+m) = \begin{pmatrix} y_\theta(-L,n) \\ \dots \\ y_\theta(0,n) \\ \dots \\ y_\theta(L,n) \end{pmatrix} \text{ and } y_\theta = \begin{pmatrix} y_\theta(0) \\ y_\theta(1) \\ \dots \\ \dots \\ y_\theta(N-1) \end{pmatrix} \tag{1}$$

Where $I(n)$ is a vector presenting a fixed column n of the two dimensional image $I(l,n)$ with $l, -L \leq l \leq L$ and $n, 0 \leq n \leq N-1$:

$$I(n) = \begin{pmatrix} I_0(n) \\ \dots \\ I_l(n) \\ \dots \\ I_{L-1}(n) \end{pmatrix} \text{ and the intire image as } I = \begin{pmatrix} I(0) \\ I(1) \\ \cdot \\ \cdot \\ I(N-1) \end{pmatrix}$$

One condition on $I(n)$ is that it must be periodic, with period N , such that:

$$I(n) = I(n + N) \tag{2}$$

M represents the number of neighboring vectors (i.e., columns of the image) either side of the input vector $I(n)$ involved in calculating $y_\theta(n)$ where $N = 2M + 1$.

$\theta, \theta \in [0, 2\pi[$ denotes the projection angle of GMDRT. $y_\theta(j, n)$ presents the sum of the pixels centered on the curve starting at the position (j, n) and rotated according to the angle θ . $R_{m,\theta}$ are $(2L+1) \times (2L+1)$ matrices whose non-zero entries select which elements of $I(n+m)$ contribute to the projection $y_\theta(n)$. Each row $j, -L \leq j \leq L$ in $R_{m,\theta}$ samples the pixels from $I(n+m)$ belonging to the curve starting at the position (j, n) in the image I . The principle of the transform is illustrated in the figure 2. The transform matrix $R_{m,\theta}$ is constructed as follows:

$$R_{m,\theta}(j, l) = \delta(m' - \phi(l')) \tag{3}$$

where

$$(l', m') = \begin{cases} Rt_\theta^{-1}(\langle l - j \rangle_{-(2L+1)}, m) & \text{if } L < l - j \leq 2L \\ Rt_\theta^{-1}(\langle l - j \rangle_{(2L+1)}, m) & \text{if } -2L \leq l - j < -L \\ Rt_\theta^{-1}(l - j, m), & \text{if } -L \leq l - j \leq L \end{cases}$$

Where $l, -L \leq l \leq L$ and $j, -L \leq j \leq L$. $\delta(\mu)$ is the kronecker delta function and $\langle \mu \rangle_x$ denotes $\mu \pmod{x}$ since $l - j$ is restricted to be included in the $[-L, L]$ interval to avoid to exceed the image borders. j presents the vertical translation step of the curve ϕ along y-axis. Rt_θ^{-1} means the inverse plane rotation according to a certain angle θ . More details about the GMDRT method is provided in [13] and it is to note that for $j=0$ and for $\theta=0$, the GMDRT formalism is equivalent to the formalism proposed by Beylkin [2] where the curves are uniquely projected according to the horizontal direction and is not vertically translated. Figure2 presents $R_{0,0}$ and $R_{l,0}$ selecting pixels belonging to a parabolic curves from respectively the columns $I(0)$ and $I(l)$ and starting both at columns 0.

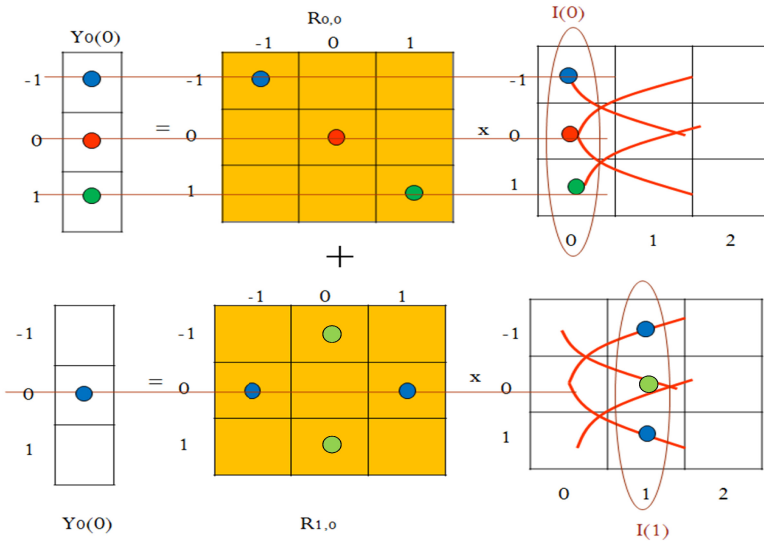


Fig. 2. A projection of horizontal parabolic curves $\phi(l)=l^2$ translated vertically, $\theta=0, M=1, L=1, N=3, n=0$.

The GMDRT then can be seen as a mapping between the image space and the Radon space. The coordinates of a point in the latter space correspond to the coordinates of an object oriented according to an angle θ in the image space. The intensity at that point corresponds to the amount of evidence for that object. The following figure represents the peaks resulting from the GMDRT on a simple image where rectangles are distributed and rotated in various directions. The peaks in the Radon space correspond to the coordinates of the upper left corner of the rectangle.

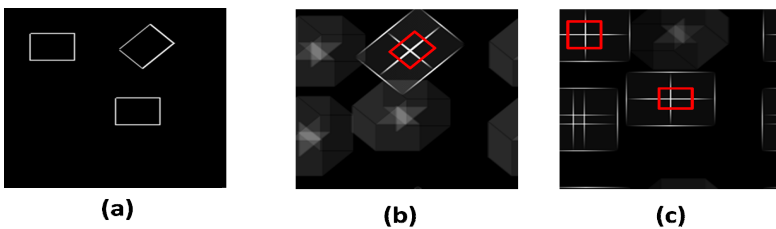


Fig. 3. (a): initial image, (b): The peak in red square resulting from projecting the initial image by $\pi/4$ -rotated rectangles. (c): The peaks projecting (a) with horizontal rectangles.

3 Buildings Extraction

The method developed in this study was applied to extract buildings from a Quickbird image of Strasbourg city with a spatial resolution 0.6 metre/pixel. The initial image is pretreated to remove noise. Then, we have extracted from the image edges by

means of the Perwitt operator. After that, we have applied on the edge image the GMDRT approach and in final, we have extracted the local maximas to detect the peaks.

3.1 Pretreatment Phase

The initial image is treated by the multi-scale mathematical morphological filter. This filter [10] operates with four structure elements of various scales and was used to preserve building boundaries while removing noise such as thinner lines and spots. In addition, This operation decreases the spectral variability of the building regions.

3.2 Extracting Contours

We detect the edges of the original image by means of the Perwitt operator [14] which finds edges using the Perwitt approximation to the derivative. It returns edges at points where the gradient of the image is the maximum.

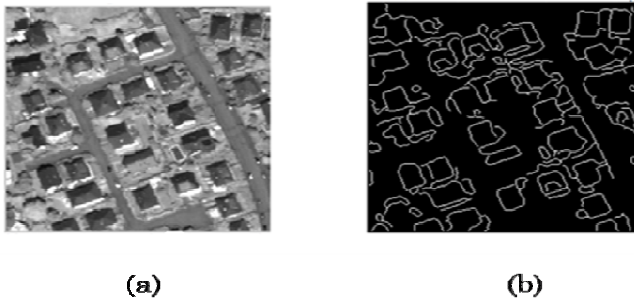


Fig. 4. (a) the initial image. (b) the contours image.

3.3 Applying GMDRT

We have applied iteratively the GMDRT transform on the contour image by projecting it with rectangular objects. For that, we vary at each iteration, the parameters i.e. the length l and the width w of the projected rectangles and its orientation angle θ . Thus, the result of a GMDRT transform is the Radon images $y_{\theta,l,w}$. Taking in account the image resolution and the general information about buildings, we have varied l and w in the [19-32] interval and θ in $[0^\circ, 180^\circ]$ applied with 5° step.

3.4 Local Maximas

$y_{\theta,l,w}$ is a two dimensional Radon space where each pixel denotes the projection result of a rectangle having as parameters l , w and oriented according to θ angle. Therefore, a peak in the Radon image testifies of the coordinates and parameters of a building in that position. To detect the peaks in the Radon images, we have used a

local maximas filter. In our tests, the size of the local maxima filter was fixed to be 21x21 taking in consideration the minimum distance between buildings. After selecting the local maximas in all $y_{\theta,i,w}$ images, we check the peaks coordinates. If more than one peak are at the same position or at a very approximate positions, we select the peak with the highest value.

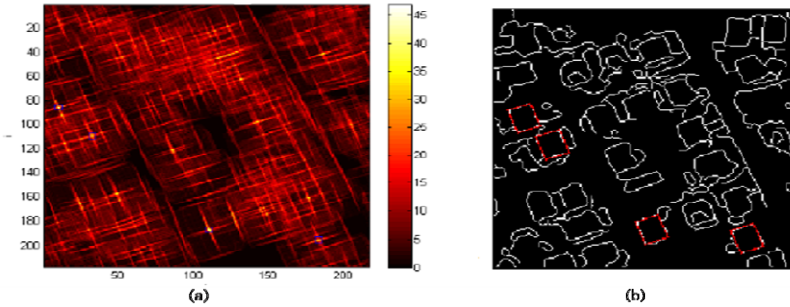


Fig. 5. A GMDRT projection via rectangles sized of 23x20 and rotated according to 20° angle. (a)some of peaks in $y_{\pi/4,23,20}$.(b): The related rectangles in the contour image.

4 Experiments

To evaluate the efficiency of the proposed approach in the buildings extraction, we used a metric defined in [12]:

For a quantitative assessment:

$$BER = \frac{BCE}{TB}$$

Where BER is the Building Extraction Rate, BCE is the Correctly Extracted Buildings number and TB is the total buildings number existing in the area of test.

And for a qualitative metric, we use these formalisms:

$$exactness = \frac{BCE}{BCE + FA} \tag{4}$$

$$quality = \frac{BCE}{BCE + BPE + BNE + FA} \tag{5}$$

Where BPE is the number of Partially Extracted Buildings, BNE is the number of Not Extracted Buildings and FA is the False Alarms denoting the wrongly identified buildings. We applied our approach on five test images. We present here some of our results(Figure6 and Figure7).The table1 presents the measures of GMDRT on each of the five images. We have compared our results with those of three other approaches. The first was developed using the *R-θ signature*[15]which presents a new shape descriptor based on Classical Radon transform, the second ,“*DCB*”, extracts buildings with reference to geodatabase and prior knowledge[12] and the last method “*DRV*” which detects buildings with the help of photometric, geometric and morphological informations[11]. The choice of these methods was guided by the similarity between their images characteristics and ours.

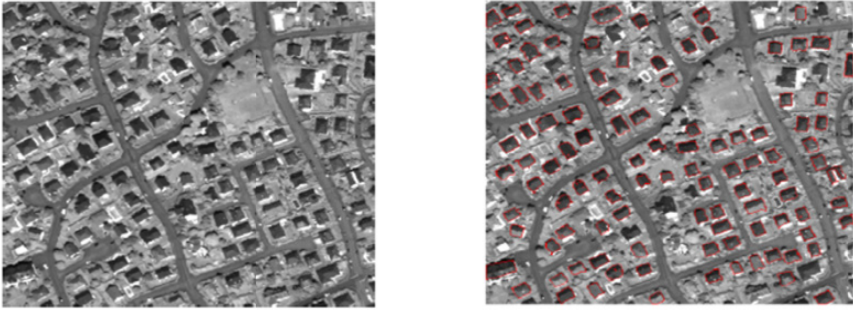


Fig. 6. On the left: initial image I1, on the right: the extracted buildings in red contours

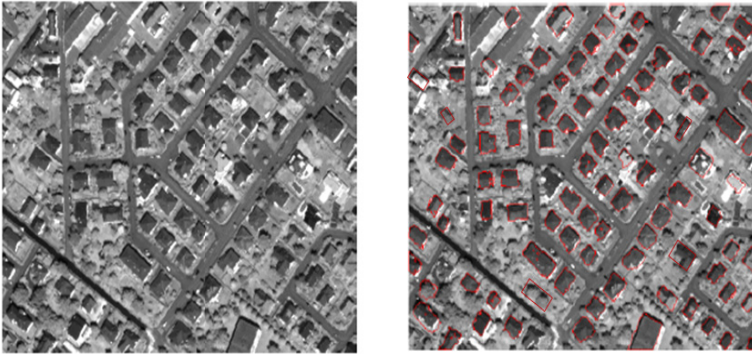


Fig. 7. On the left: initial image I2, on the right: the extracted buildings.

Table 1. (a) The measures of the GMDRT results. (b) The results rates of DCB, DRV, $R-\theta$ Signature and GMDRT methods.

	I1	I2	I3	I4	I5	Total
BCE	100	83	65	43	184	475
BPE	2	4	0	0	2	8
BNE	8	5	13	3	17	46
FA	4	3	3	0	5	15
TB	110	92	78	46	203	529

(a)

methods	BER(%)	Exactness	Quality
DCB	94	0.903	0.855
DRV	91	0.940	0.824
$R-\theta$ signature	83	0.926	0.808
GMDRT	89	0.969	0.873

(b)

We show that the Buildings extraction rate of our method is less than the other DCB and DRV approaches but this result can be improved if the quality of the contour image is perfectly performed since the GMDRT process is closely related to the extraction contours results. In another hand, our method seems to be more perfect regarded to the quality and the exactness rates. In fact, the GMDRT method tends to limit the number of the buildings partially extracted and the false alarms.

5 Conclusion

This paper describes the GMDRT approach and its application in building extraction. The results showed that our approach has good performance in detecting rectangular building shapes in the image but the user interfere to setting the dimensions interval of rectangles is a problem itself. In the future work we will try to make the Rectangular GMDRT fully automatic method in the way that the sizes of the rectangles in the image will be extracted with the help of the GMDRT itself.

References

1. Tofts, P.: The Radon Transform: Theory and implementation. Ph.D.Thesis (1996)
2. Beylkin, G.: Discrete Radon transform. *IEEE Transactions of Acoustics, Speech and Signal Processing* 35, 162–172 (1987)
3. Magli, E., Olmo, G., LoPresti, L.: Pattern recognition by means of the Radon transform and the continuous wavelet transform. *Signal Processing* 73 (1999)
4. Milanfar, P.: A model of the effect of image motion in the Radon transform domain. *IEEE Transactions on Image Processing* 8, 1276–1281 (1999)
5. Shepp, L.A., Krystal, J.B.: Computerized tomography: The new medical X-ray technology. *Am. Math. Monthly* 85, 420–439 (1978)
6. Courmontagne, P.: An improvement of ship wake detection based on the radon transform. *Signal Processing* 85 (2005)
7. Krishnaveni, M., Kumar Thakur, S., Subashini, P.: An optimal method for wake detection in SAR images using Radon transformation combined with wavelet filters. *International Journal of Computer Science and Information Security* 6, 066–069 (2009)
8. Zhang, Q., Couloigner, I.: Accurate Centerline Detection and Line Width Estimation of Thick Lines using the Radon Transform. *IEEE Transactions On Image Processing* 16, 310–316 (2007)
9. Wang, L., Hao, Y.: Radon Transform and Forstner Operator Applying in Buildings Contour Extraction. In: *Sixth International Conference on FSKD*, pp. 415–419 (2009)
10. Mukhopadhyay, S., Chanda, B.: An edge preserving noise smoothing technique using multi-scale morphology. *Signal Processing* 82, 527–544 (2002)
11. Lhomme, S., He, D.C., Weber, C., Morin, D.: A new approach to building identification from very-high-spatial-resolution images. *International Journal of Remote Sensing* 30, 1341–1354 (2009)
12. Bouziani, M., Goita, K., He, D.-C.: Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge. *ISPRS J. of Photogrammetry and Remote Sensing* 65, 143–153 (2010)
13. Elouedi, I., Fournier, R., Nait-Ali, A., Hammouda, A.: Generalized Multi Directional Discrete Radon Transform. *Signal Processing* (paper in revision)
14. Zhang, W., Bergholm, F.: Multi-scale blur estimation and edge type classification for scene analysis. *International Journal of Computer Vision* 24, 219–250 (1997)
15. Hamouda, A., Rojbane, H., Elouedi, I.: A new shape descriptor based on the Radon transform: the $R\theta$ -signature. Accepted paper in *International Conference on Signal, Image Processing and Applications, ICSIA* (2011)

Speaker Tracking Using Multi-modal Fusion Framework

Anwar Saeed, Ayoub Al-Hamadi, and Michael Heuer

Institute for Electronics, Signal Processing and Communications (IESK),
Otto-von-Guericke-University Magdeburg,
D-39016 Magdeburg, P.O. Box 4210, Germany
{Anwar.Saeed,Ayoub.Al-Hamadi}@ovgu.de

Abstract. This paper introduces a framework by which multi-modal sensory data can be efficiently and meaningfully combined in the application of speaker tracking. This framework fuses together four different observation types taken from multi-modal sensors. The advantages of this fusion are that weak sensory data from either modality can be reinforced, and the presence of noise can be reduced. We propose a method of combining these modalities by employing a particle filter. This method offers satisfied real-time performance.

Keywords: Speaker tracking, Skin detection, Face detection, Particle filter, Time difference of arrival.

1 Introduction

This work represents an example of using multiple modalities within a particle filter. We then describe a particle filter by which multi-modal sensory data is fused together to track a speaker in a scene. For designing a particle filter for an application, it is necessary to introduce feedback to the system. This feedback should describe the scene. In the case of speaker tracking, we have two types of sensory input: video and audio input devices. Image processing methods operate on image sequence captured from the video input to detect the speaker and other approaches use the audio input for the speaker localization.

Tracking the speaker using audio-visual information is an active research topic in the computer vision due to its importance to various applications such as smart video-conferencing and surveillance security systems. Shivappa et al. [1] explored different strategies for audio-visual fusion. Vermaak et al. [2] described a method in which a standard contour tracking algorithm, consisting of an edge detector and a particle filter, is used in conjunction with a Time Difference of Arrival (TDOA) calculation to deduce the speaker location from auditory data received by a pair of microphones. The audio data are used for initialization and video data for localization in an attempt to utilize the strengths of each modality. This method enhances the existing visual tracking successfully and can detect speaker 'ping-pong'. However, this implementation is not a real-time solution. Zhou et al. [3] employed a histogram matching based technique for

image based speaker detection and a TDOA algorithm once again for audio localization. The audio undergoes pre-processing to remove noise and a Kalman filter is used to further reduce spurious detections before both audio and video observations are passed to a Weighted Probabilistic Data Association (WPDA) filter for fusion and tracking. The aforementioned approach fused deferent types of sensory data; however, it did not provide real-time level of performance. In this paper, we address the speaker tracking with the help of audio-visual sensory data in real-time level.

We extract two observation data types from the video modality. The first one represents skin blobs that fit human face shape. The second observation type is the output of a human face detector. Additionally, two observation types are extracted from audio modality: the first one utilizes the time-difference of arrival of audio signal at two microphones, and the second one uses Received Signal Strength (RSS) to estimate the speaker location. These four observation types from the two modalities will then be combined in a useful and meaningful fashion using a particle filter and then used to detect and track the speaker in the scene.

The remainder of this paper is structured as follows. In section 2, we describe the video modality. Audio modality is explained in section 3. Overview of the particle filter implementation is detailed in section 4. Experimental results are discussed in section 5. Finally, the conclusion and future perspectives are given in section 6.

2 Video Modality

The sensor configuration used in this work is a mono-camera centered between two microphones. In the video modality, we segment the human face using two methods. The first one utilizes human skin color, while the second approach uses texture features.

2.1 Face Detection Using Skin Color

Several approaches were proposed to classify each pixel as skin or non-skin. Multivariate Gaussian mixture model (GMM) is an example of those approaches. Two GMMs could be built for skin and non-skin pixels. Then, each pixel with likelihood ratio exceeding an experimentally set threshold value is classified as skin [4]. However, this method is time-consuming in contrast to other methods that compare the pixel value with pre-learnt threshold values. This comparison could be carried out in different color spaces such as RGB, HSL, HSV, YCrCb, etc. [5]. In this paper, a combination of threshold filters in HSV and YCrCb is used to segment the skin pixels. The threshold values define boundaries for the pre-learnt skin color [6]. As we mentioned, we chose this skin detection technique due to its efficiency and to meet our requirement for building real-time speaker tracking approach. We chose chrominance channels (cr, cb) from YCrCb color space and the Hue channel (h) from HSV color space [6]. Each pixel $I(h, cr, cb)$ is classified as follows.

$$I(h, cr, cb) = \begin{cases} 1 & HSV(h) \wedge CbCr(cr, cb) \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $HSV(h)$ represents the skin segmentation in HSV color space. The Hue component is proven to be a good discriminator for the human skin tone. $HSV(h)$ is calculated by

$$HSV(h) = (h < 25) \vee (h > 230). \quad (2)$$

$CbCr(cr, cb)$ represents the skin segmentation in YCrCb color space. The luminance component (Y) is ignored to have skin detection immune to the luminance variation. $CbCr(cr, cb)$ is calculated by

$$CbCr(cr, cb) = \left\{ \begin{array}{l} (cr \leq 1.5862 \times cb + 20) \wedge \\ (cr \geq 0.3448 \times cb + 76.2069) \wedge \\ (cr \geq -4.5652 \times cb + 234.5652) \wedge \\ (cr \leq -1.15 \times cb + 301.75) \wedge \\ (cr \leq -2.2857 \times cb + 432.85) \end{array} \right\}. \quad (3)$$

At the end of skin pixel classification, we operate morphology operations to remove the outlier skin detection and to close misdected pixels. The resulting skin regions are then examined for contours which may describe faces in the scene. The contours which do not fit the facial characteristics will be discarded.

2.2 Face Detection Using Texture Features

The face detected by skin color is error-prone due to cluttered environments and due to the existence of objects that are human skin colored with same face shape. Hence, we enhance the face detection by the use of texture features. This detection is more accurate; however, it detects the face only in a frontal upright pose with ± 20 head rotation angles (yaw, pitch, and roll). Thus, the two methods of face detection will complement each other using the particle filter framework. We assign the detection using texture features more weight than that using skin color. We pass the same video frame, which is processed by skin segmentation, into a well-trained Haarcascade classifier [7,8]. This classifier utilizes the Haar-like features, which are defined as the ratio of intensities of adjacent rectangles of different locations [9].

3 Audio Modality

The audio data are processed by fast but rather imprecise Received Signal Strength RSS based location scheme and by much slower but far more accurate TDOA algorithm. Different precision and confidence factors are assigned to the locations estimated by TDOA and RSS according to the algorithms accuracy and reliability.

3.1 Audio TDOA

It is possible to locate the origin of a sound source by observing the TDOA of signal at audio sensors placed in differing locations. The number of the used audio sensors determines the accuracy and dimensionality as well as the performance of the system. The processing using TDOA method is relatively costly, increasing exponentially with sensor count. To maintain real-time levels of performance, only two audio sensors are used in this implementation. This limits the output of the modality to a single dimension (the x-coordinate). However, as this dimension has the most variance, the chosen configuration gave an excellent balance of performance versus contribution. The algorithms used to estimate (TDOA) exploit the fact that the sound arriving at each microphone will be delayed according to the speaker position, given that the position of the microphones is known. To find this time delay, a cross-correlation function is used. Due to the nature of audio sensory information, multiple detections are possible for a single source. These ghost detections are the result of reverberations being interpreted falsely as signals originating directly from the source. To minimize the effect of these reverberations, we used a Generalized Cross Correlation function using a Phase Transform (GCC-PHAT), which was introduced by Knapp et al. [10]. This is a computationally expensive operation so the Discrete Fourier Transform (DFT) is used for efficiency. Given two signals S_1 and S_2 , the weighted correlation function at time delay τ is

$$GCC_{PHAT}(\tau) = \mathbf{FFT}^{-1} \left(\frac{F_1(f)[F_2(f)]^*}{|F_1(f)[F_2(f)]^*|} \right), \quad (4)$$

where F_1 and F_2 are the Fourier transforms of S_1 and S_2 , respectively. \mathbf{FFT}^{-1} is the inverse Fourier Transform and $[\]^*$ denotes the complex conjugate. Only the frames containing speech will be processed. To determine the speech frames, we used Signal to Noise Ratio (SNR) along with Zero Crossing Rate [11].

3.2 Audio RSS

This Audio signal strength is far simpler than the previously described TDOA routine. However, in most cases it gives a decent approximation of the speaker position. It is not precise as TDOA. Consequently, this will be reflected in the particle filter parameters. Similar to TDOA, audio RSS is used to estimate the speaker position just in one dimension (x-coordinate). To measure audio RSS, the total energy is calculated for the signals contained in both left and right audio buffers. A signal contained in a buffer of size N samples is defined as

$$\mathbf{s} = (s(0), \dots, s(N)).$$

Hence, the energy of a signal of size N samples is

$$e(\mathbf{s}) = \sum_{n=0}^N [s(n)]^2. \quad (5)$$

By comparing the energy of each buffer, we can arrive at x-coordinate representing the relative position of a speaker in the scene. While being less precise, a greater confidence is placed in this Audio RSS compared to TDOA. This is due to the fact that RSS is less prone to noise such as reverberations.

The audio signals are continuously measured and buffered for (10ms-41ms) before using them for the location estimation by TDOA and RSS. Obviously, the buffers are cleared after each estimation, and the process iterates.

4 Particle Filter Implementation

Before discussing the fusion of multiple modalities, it is necessary to understand the means by which sensory data is incorporated into a particle filter. The points within a filter where sensory data is used are the measurement and selection stages. In the measurement stage, the particle set generated by the filter is examined and each particle is given a score, or weight, based on how accurately said the particle describes the scene. During the selection stage, the particle set is refined to accentuate those particles which best describe the scene, and new particles are introduced based on the current sensory data. The parameters of these steps vary on what the purpose of the filter is. In the case of the filter implementation given in this paper, it is intended that the particles should describe a bounding box relative to a camera snapshot showing where the current speaker is. In this case, each particle needs to be given some form of score based on how accurate the bounding box describes the real speaker position. The location of this real speaker from the raw sensory data and the choice of what weight to assign to each particle is our concern.

The general particle filtering scheme in [12][13] is used in this work. The approach is particularly effective at tracking objects in substantial cluttered environments, which is very desirable when dealing with multiple modalities each potentially generates many observations and hence noise. The problem of speaker tracking can be formulated as a continues estimation of speaker state \mathbf{x}_t at each time t . The state forms a vector containing a union of relevant parameters obtained from all modalities of feature extractors. The state \mathbf{x}_t is defined here as

$$\mathbf{x}_t = (x_t(1), x_t(2), x_t(3), x_t(4), x_t(5)). \quad (6)$$

$(x_t(1), \dots, x_t(5))$ denote x-position, y-position, head width, head height, and head orientation angle, respectively. Particle Filters, a Quasi-Monte Carlo solution, solve the tracking issue. Let state \mathbf{x}_t in the Markov state-space model represent a possible speaker configuration at time t . And \mathbf{y}_t is the observation obtained using the two modalities, as discussed in sections (3) and (2). Then, the distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ can be calculated by

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (7)$$

Eq. (7) consists of likelihood function $p(\mathbf{y}_t|\mathbf{x}_t)$ multiplied by an integral representing the prediction step. The particle filtering method used here is an

approximation to Eq. 7. At each time t , there are N samples $\{\mathbf{x}_t^{(i)}, i = 1, \dots, N\}$ each associated with weight value $w_t^{(i)}$. To avoid the degeneracy problem caused in re-sampling, the weight value is calculated as follows [14].

$$w_t^{(i)} \propto p(\mathbf{y}_t | \mathbf{x}_t^{(i)}). \quad (8)$$

In the particle prediction step, the state of each particle is altered according to an underlying temporal model, as given in Eq. 9.

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{n}_t. \quad (9)$$

Where \mathbf{A} and \mathbf{B} are experimentally set parameters for the model. \mathbf{n} is a normalized white Gaussian noise vector. Within the update step, we assign each sample $\mathbf{x}_t^{(i)}$ new weight value according to the observation likelihood as in Eq. 8. Where the observation likelihood function is defined as an averaged sum of likelihood functions of the particle components,

$$p(\mathbf{y}_t | \mathbf{x}_t) = \frac{1}{5} \sum_{i=1}^5 p(\mathbf{y}_t | x_t(i)). \quad (10)$$

As we mentioned in sections 3 and 2, we have four different observation types. In addition, we could have multiple observations from each type. For example, more than one face could be detected and may be many locations could be estimated for the speaker in each observation type. The component $x(1)$ is fused from the four observation types, while the other four components are fused only from the video modality. Each component $x(i)$ has precision factors $\sigma_{i,m}$ and confidence factors $\gamma_{i,m}$ reflecting the observation type (m) accuracy and reliability, respectively. Let us have Z observations for the component $x_t(i)$. These observations are denoted by $(y_{t,1}^m(i), \dots, y_{t,Z}^m(i))$, where m denotes the observation type. Then, we formulate the observation likelihood function for each component as GMM given by

$$p(\mathbf{y}_t | x_t(i)) = K \sum_{o=1}^Z \frac{\gamma_{i,m}}{\sqrt{2\pi\sigma_{i,m}^2}} \exp\left(-\frac{(y_{t,o}^m(i) - x_t(i))^2}{2\sigma_{i,m}^2}\right), \quad (11)$$

where K is a normalization factor, which depends on the observation number Z . Obviously, K is the reciprocal of the sum of the confidence factors $\gamma_{i,m}$ for all observations. Finally, the density distribution $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ in Eq. 7 is approximated by a sum of N Dirac functions as follows.

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) \approx \sum_{i=1}^N w^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}). \quad (12)$$

5 Experimental Results

The performance of the particle filter is more than sufficient for real-time operation. This filter is employing the four input data, as described in sections 2 and 3.

An internal timer of the application iterates the filter every 10ms, which is enough for most iterations; however, when a slowdown occurs due to the presence of more observations, it has never been significant enough to go beyond the 24 frames per second (41ms per iteration) limit. The main observable distinction between the multi-modal approach as opposed to single modal one is that the system is much more robust when dealing with noise or non detection. For example, the Haar-like face detection observations are not always present due to the orientation of speaker face in the scene; however, both the audio and skin color modalities allow tracking to continue. Conversely, over abundance of observations generated by the skin color segmentation is complemented by the accuracy Haar-like face detection meaning that the tracker does not get distracted by the multiple false positives. Fig. 1 (a,b) shows two samples of our experiment. The two images are overlaid with data extracted from the audio and video modalities. The probability density distribution of x-component is given below each image. The speaker was detected and tracked successfully.

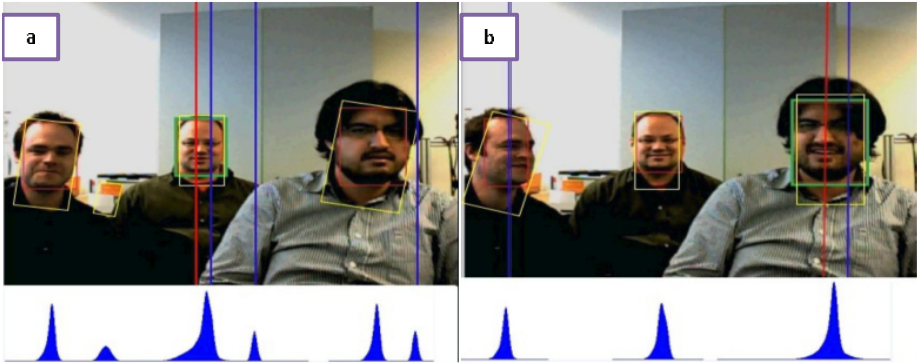


Fig. 1. Images showing the particle filter performance. The red and yellow rectangles show measurements from the video modality and the vertical blue and red lines from the audio modality. Below each image is the probability density distributions for the x-component given the measurements shown. The green rectangle is the filter output corresponding to the greatest peak in the probability distribution. In (a), the center person is speaking whereas in (b) the right person is doing the majority of the speaking with some sound coming from the left.

6 Conclusions and Future Work

We have shown an approach for speaker tracking. We combined four observation types from two modalities (audio-visual). Each observation could not be used alone; however, when we combined them within a particle filter, we achieved a robust performance. The observations are modeled by GMMs with experimentally set parameters. Our proposed approach is able to perform in real time on standard workstation. In future, we will consider adding more microphones. Hence, we will be able to estimate the speaker location in y-coordinate as well.

Acknowledgments. This work is supported by Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by DFG and OvG-University Magdeburg.

References

1. Shivappa, S., Trivedi, M., Rao, B.: Audiovisual information fusion in human computer interfaces and intelligent environments: A survey. *Proceedings of the IEEE* 98, 1692–1715 (2010)
2. Vermaak, J., Gangnet, M., Blake, A., Perez, P.: Sequential monte carlo fusion of sound and vision for speaker tracking. In: *ICCV*, pp. 741–746 (2001)
3. Zhou, H., Taj, M., Cavallaro, A.: Target detection and tracking with heterogeneous sensors. *IEEE Journal of Selected Topics in Signal Processing* 2, 503–513 (2008)
4. Saeed, A., Niese, R., Al-Hamadi, A., Michaelis, B.: Coping with hand-hand overlapping in bimanual movements. In: *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 238–243 (2011)
5. Schettini, R., Gasparini, F.: Skin segmentation using multiple thresholding. In: *Internet Imaging VII, IS and T/SPIE*, pp. 60610F-1–60610F-8. SPIE (2006)
6. Rahman, N.A., Wei, K.C., See, J.: RGB-H-CbCr Skin Colour Model for Human Face Detection. In: *Proceedings of The MMU International Symposium on Information & Communications Technologies, M2USIC 2006* (2006)
7. Saeed, A., Niese, R., Al-Hamadi, A., Panning, A.: Hand-face-touch measure: a cue for human behavior analysis. In: *2011 IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS)*, vol. 3, pp. 605–609 (2011)
8. Bradski, G.: *The OpenCV Library*. Dr. Dobb's Journal of Software Tools (2000)
9. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features, pp. 511–518 (2001)
10. Knapp, C.H., Carter, G.C.: The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust., Speech, Signal Processing* 24, 320–327 (1976)
11. Bachu, R.G., Kopparthi, S., Adapa, B., Barkana, B.D.: Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. In: *American Society for Engineering Education ASEE Zone Conference Proceedings*, pp. 1–7 (2008)
12. Blake, A., Isard, M.: The CONDENSATION algorithm - conditional density propagation and applications to visual tracking. In: *NIPS*, pp. 361–367 (1996)
13. Steer, M., Al-Hamadi, A., Michaelis, B.: Audio-visual data fusion using a particle filter in the application of face recognition. In: *2010 20th International Conference on Pattern Recognition (ICPR)*, pp. 4392–4395 (2010)
14. Doucet, A., De Freitas, N., Gordon, N. (eds.): *Sequential Monte Carlo methods in practice* (2001)

New Encoding Algorithm for Distributed Speech Recognition Based on DTFS Transform

Azzedine Touazi* and Mohamed Debyeche

University of Sciences and Technology Houari Boumediene, El Alia 16111, Bab Ezzouar,
Algiers, Algeria
{touazi.azzedine,mdebyeche}@gmail.com

Abstract. The paper presents a new algorithm for efficient compression of front-end feature extracted parameters used in distributed speech recognition systems (DSR). In the proposed method the source encoder is mainly based on discrete time Fourier series (DTFS) by interpolation using Fourier coefficients with conventional vector quantization. The system provides a compression bit rate as low as 4 kbps; the experiments were carried out on the TIDigits Aurora2 database [1]. The simulation results show good recognition performance without dramatic change comparing with ETSI STQ-AURORA standard front-end feature compression algorithm with quantized features at 4.4 kbps [2].

Keywords: Distributed speech recognition, Vector quantization, Discrete time Fourier series, Aurora2 database.

1 Introduction

The growth in wireless communication and mobile devices has supported the development of distributed speech recognition systems. Being developed and standardized by ETSI [2]. The basic idea of DSR consists of using a local Front end (FE) from which speech features are extracted and transmitted through a data channel to a remote Back end (BE) or remote server recognizer. The speech features used for recognition are the first 12 mel frequency cepstral coefficients (MFCCs) (c1-c12), the zeroth cepstral coefficient (c0) and the log energy (log E) in the frame. The 14-dimensional feature vector is split into seven sub vectors; each of the sub vectors is encoded with a different 2-dim VQ.

The standard computes a feature vector every 10ms and allocates 44 bits to each vector to achieve a total bit rate of 4400 bps [2]. The number of bits allocated to each sub vector is shown in Table 1; with 8 bits are allocated to the (c0-log E) sub vector and 6 bits are allocated to the rest of each sub vector.

The Aurora 2 database consists of connected digit sequences for American English Talkers. It provides speech samples and scripts to perform speaker independent speech recognition experiments in clean and noisy conditions. This database has been prepared by down sampling to 8 kHz, filtering with the G.712 characteristic; noise is

* Azzedine Touazi, Faculty of Electronic and Computer Sciences, Signal and Communication Laboratory (LCPTS).

Table 1. Bits allocation used by Aurora

Sub-vector	c1, c2	c3, c4	c5, c6	c7, c8	c9, c10	c11, c12	c0, log E
Bits allocated	6	6	6	6	6	6	8

artificially added to the filtered TIDigits at a desired SNR (clean, 20, 15, 10, 5, 0, -5dB) and eight different noise conditions - Subway - Babble - Car - Exhibition hall - Restaurant - Street - Airport - Train station. Furthermore, a full description of the Aurora2 database is given in [1].

Various schemes for compressing the MFCC vectors have been proposed in the literature. Among these methods are the coding based on discrete cosine transforms (DCT) [3] [4], another method that uses the predictive vector quantization [5]. Also, by analysis the statistical properties of the MFCC vectors a series of quantization schemes have been described in [6].

In this paper we have derived an interpolation method which operates in discrete time Fourier series domain (DTFS). This transform is widely used in signal processing such as spectral analysis and filter design.

In the proposed algorithm we exploit the temporal correlation characteristic between consecutive MFCC vectors extracted at regular period and transformed into DTFS domain, suggests that we do not have to transmit every spectral component (magnitudes and phases) of MFCC vector to the decoder; instead, we could transmit only one part of the spectral component at regular interval. However, at the decoder, the non transmitted spectral component could then be derived by means of linear interpolation from the adjacent components.

2 Overall Description of the Algorithm

At the client side, speech is first segmented into frames; features are computed for each frame of 10 ms. As shown in figure 1 a normalized DTFS is performed and converted to the polar form with $N=14$ ($c1, \dots, c12, c0, \log E$) and $k=0, \dots, N/2$, where the phase spectrum is discarded according to the following equations [7] [8] and [9]:

$$S_k = \sqrt{A_k^2 + B_k^2}, \quad \Phi_k = \arctan \left[\frac{B_k}{A_k} \right] \tag{1}$$

With:

$$\left. \begin{aligned} A_k &= \frac{2}{N} \sum_{n=0}^{N-1} s(n) \cos\left(\frac{2\pi kn}{N}\right) \\ B_k &= \frac{2}{N} \sum_{n=0}^{N-1} s(n) \sin\left(\frac{2\pi kn}{N}\right) \end{aligned} \right\} k=1,2,\dots,N/2-1 \quad \text{and} \quad \left. \begin{aligned} A_k &= \frac{1}{N} \sum_{n=0}^{N-1} s(n) \cos\left(\frac{2\pi kn}{N}\right) \\ B_k &= \frac{1}{N} \sum_{n=0}^{N-1} s(n) \sin\left(\frac{2\pi kn}{N}\right) \end{aligned} \right\} k=0, N/2 \tag{2}$$

The polar components are $S=(s_0, s_1, \dots, s_7)$ and $\Phi=(\varphi_0, \varphi_1, \dots, \varphi_7)$. Since the equation (2), for $k=0$, we can demonstrate that $\varphi_0 = 0$. For the set of two successive feature vectors $MFCC_n$ and $MFCC_{n+1}$ we transmit both phase and amplitude spectra for $MFCC_n$ and just the phase spectrum for $MFCC_{n+1}$.

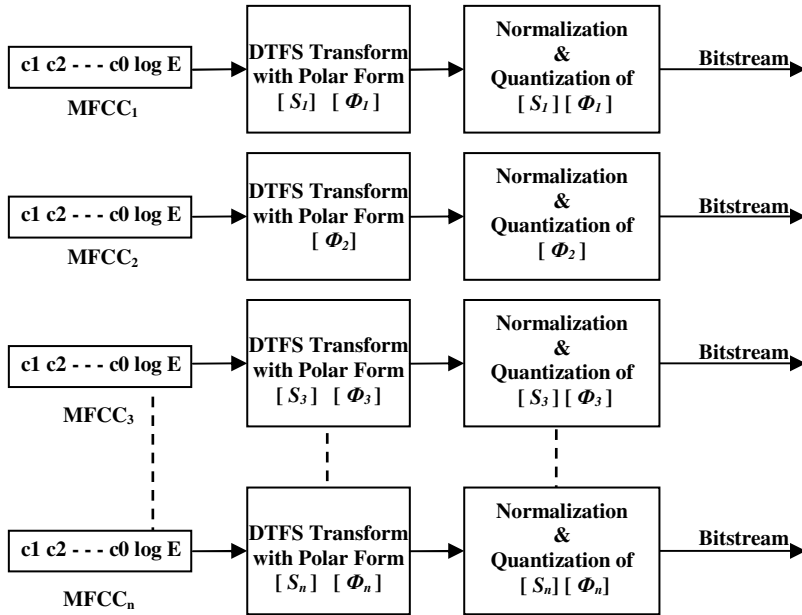


Fig. 1. DTFS transforming and quantization block

The choice to transmit the phase instead of the amplitude in the case of $MFCC_{n+1}$ is approved by an experiment with comparing the SNRs average (sets A, B and C) for interpolated DTFS components without quantization; as shown in figure below for the following cases: 1) Both amplitude and phase interpolation. 2) Amplitude interpolation. 3) Phase interpolation.

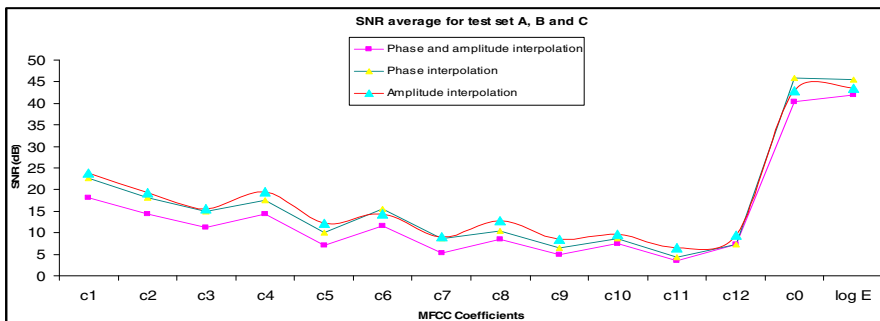


Fig. 2. SNR measurements average without quantization

The SNRs average for the case of amplitude interpolation shows a greatest value comparing with the two other cases. We have also compared the SNRs average between interpolated amplitude spectra and Aurora encoder [2]. It can be seen from figure 3 that the SNR degrees in the case of amplitude interpolation are higher than the Aurora encoder for the first four coefficients (c1-c4) and are decreasing from the coefficient c5; for (c0, log E) the values are smoothly increased comparing with Aurora encoder. As it is well known that the lower feature coefficients provide the greatest contribution to the recognition performance [10]; therefore a DTFS transform with amplitude interpolation can lead to a minor influence in the recognition performance.

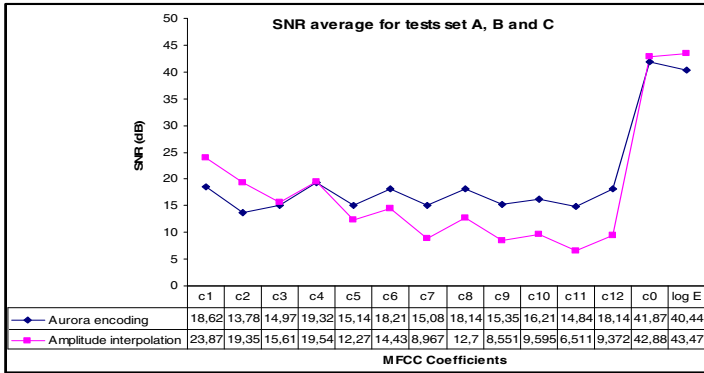


Fig. 3. SNR measurements average comparing with aurora encoder

In the quantization step, for two successive MFCC vectors the amplitude spectrum vectors (s_0, s_1, \dots, s_7) are encoded using split vector quantizer SVQ with the same codebooks, in which each vector is split equally into four sub-vectors and each one is quantized using its own VQ codebook trained with LBG algorithm [11] with codebooks of size 128 each.

The phase spectrum $(\varphi_1, \dots, \varphi_7)$ is encoded using SVQ quantizer in which the vector is split into three sub vectors of ranks 2, 2 and 3 respectively, with codebooks of size 256 each in the case of 3.8 kbps and 512 each in the case of 4.1 kbps.

The reason to choose more bits for encoding the phase of Fourier transform coefficients than encoding of spectral amplitude it's proved in reference [12]. The Table 2 shows the bits allocation for both cases of 3.8 and 4.1 kbps.

The decoding process at the back end consists of the inverse operations of the encoding in reverse order. However, the non-transmitted spectral component could then be derived by means of linear interpolation from adjacent components by:

$$S_{MFCC_n} = \frac{S_{MFCC_{n-1}} + S_{MFCC_{n+1}}}{2} \tag{3}$$

Table 2. Bits allocation at 3.8 & 4.1 kbps

		Bits allocation at 3.8 kbps		Bits allocation at 4.1 kbps	
Polar form	Sub-vector	MFCC _n	MFCC _{n+1}	MFCC _n	MFCC _{n+1}
Phase	(φ_1, φ_2)	8	8	9	9
	(φ_3, φ_4)	8	8	9	9
	($\varphi_5, \varphi_6, \varphi_7$)	8	8	9	9
Amplitude	(s_0, s_1)	7	-	7	-
	(s_2, s_3)	7	-	7	-
	(s_4, s_5)	7	-	7	-
	(s_6, s_7)	7	-	7	-

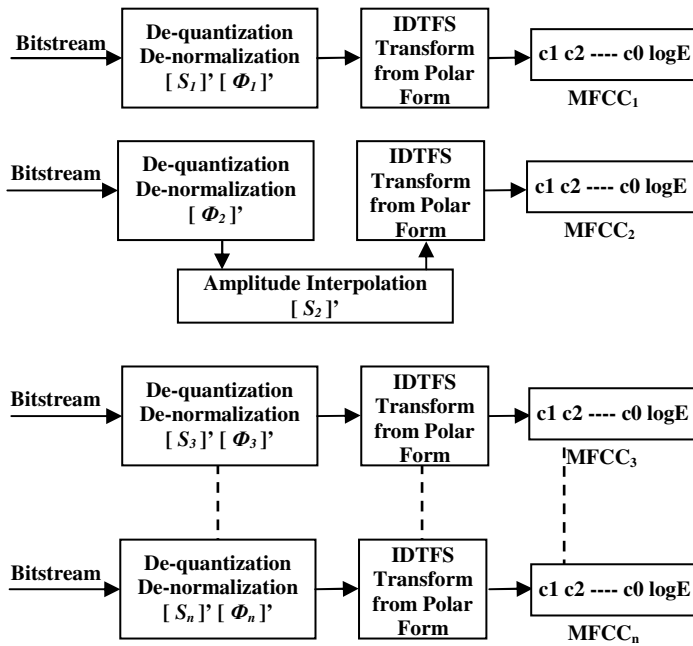


Fig. 4. DTFS de-quantization and interpolation block

3 Recognition Results

The experiments were carried out on the TIDigits Aurora corpus (sets A, B and C) with MFCCs extracted using the Aurora2 front-end [2] for both multi-condition and clean trainings. In the figure 5 we compared the SNR results for the following cases:

- Aurora encoder working at 4.4 kbps [2].
- Proposed DTFS working at 3.8 kbps.
- Proposed DTFS working at 4.1 kbps.

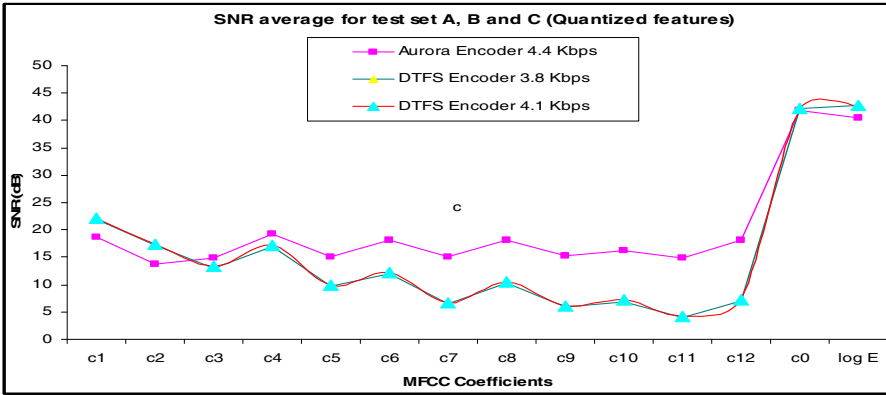


Fig. 5. SNR measurements average with quantization

It can be noticed a minor degradation from SNR levels after quantization; but for the first five MFCC coefficients we got an acceptable SNR values comparing with Aurora encoder. We note likewise when comparing (c0, log E).

The recognition were done using HTK 3.4 speech recognizer [13] to the coded MFCCs, while the c0 and log E coefficients are both used in the recognition task; however, the results are compared for both compressed and uncompressed Aurora recognition.

In order to confirm our alternative the recognition accuracies average were firstly preformed on the uncompressed DTFS features, the Tables 3 and 4 show that transmitting DTFS coefficients with amplitude interpolation can lead to the good recognition performance.

Table 3. Recognition Accuracies in multi condition (uncompressed features)

	Test set A	Test set B	Test set C
Aurora	89.60	88.31	86.24
Aurora encoding	89.58	87.91	85.30
Amplitude interpolation	89.02	88.00	85.56
Phase interpolation	80.09	78.13	80.60
Amplitude & phase interpolation	79.56	77.84	79.81

Table 4. Recognition Accuracy in clean condition (uncompressed features)

	Test set A	Test set B	Test set C
Aurora	67.62	62.96	71.62
Aurora encoding	66.65	62.29	69.80
Amplitude interpolation	66.92	62.31	70.46
Phase interpolation	57.63	52.36	65.41
Amplitude & phase interpolation	57.49	52.40	64.62

We can see from Table below, in comparison with the compressed Aurora features, the proposed DTFS encoder with amplitude interpolation working at 3.8 and 4.1 kbps maintains the same word level accuracies in the case of multi-condition, and the recognition accuracies are slightly inferior in the case of clean-condition.

Table 5. Word accuracy average (SNR: 0-20 dB), for test Sets A,B and C

Set	Training mode	Aurora standard	Aurora at 4.4 kbps	DTFS at 3.8 kbps	DTFS at 4.1 kbps
A	Clean	67.62	66.65	65.70	65.70
	Multi	89.60	89.58	88.70	88.69
B	Clean	62.96	62.29	61.07	61.15
	Multi	88.31	87.91	87.49	87.54
C	Clean	71.62	69.80	68.90	69.17
	Multi	86.24	85.30	85.22	85.34

The experiment results show that the proposed algorithm at low bit rates slightly affect the final speech recognition accuracy only by less than 1% , so there is no significant difference in term of recognition accuracy.

4 Conclusion

The proposed algorithm working on DTFS domain focuses on reducing bit rate lower than 4 kbps; generally this encoder maintains the recognition performance without dramatic change comparing with Aurora encoder, with relatively more computational cost.

Also, the proposed technique can be extended to compress another kind of parameters that highly correlated with each other like LPC coefficients. Further work will involve improving both computational cost by proposing a new quantization techniques for the DTFS coefficients and recognition performance.

Acknowledgments. The authors would like to thank the LCPTS team, for providing contribution to carry out this research work.

References

1. Hirsch, H. G., D. Pearce.: The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions, [6th International Conference on Spoken Language Processing], ICSLP,China, October (2000)
2. ETSI Standard Document.: Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Front-end Feature Extraction algorithm Compression Algorithm, ETSI ES 201 108 V 1.1.3, Sep. (2003)
3. Kiss, I., Kapanen, P.: Robust feature vector compression algorithm for distributed speech recognition. In Eurospeech (1999)

4. Zhu, Q., Alwan, A.: An efficient and scalable 2D-DCT based feature coding scheme for remote speech recognition, in Proc. IEEE Int. Conf. Acoustic, Speech. Signal processing. (2001)
5. Jose Enrique, Garcia., Alfonso, Ortega., Antonio, Miguel., Eduardo, Lleida.: Predictive vector quantization using the M-algorithm for distributed speech recognition, VI Jornadas en Tecnología del Habla and II Iberian SLTech Workshop, FALA (2010)
6. So, Stephen., Kuldip, Paliwal, K.: Quantization of speech features, source coding, Automatic Speech Recognition on Mobile Devices and over Communication Networks Advances; <http://www.springerlink.com/content/u1t465157615k202/>, (2008)
7. Julius, O, Smith.: Mathematics of the discrete Fourier transform (DFT) with audio applications, Center for Computer Research in Music and Acoustics (CCRMA), Department of Music, Stanford University, <https://ccrma.stanford.edu/~jos/mdft/>; Viewed May (2011)
8. Eddie L, T, Choy.: Waveform Interpolation Speech Coder at 4 kb/s, Department of Electrical and Computer Engineering, McGill University Montreal, Canada August (1998)
9. Mohamed, Elfataoui., Gagan, Mirchandani.: A Frequency- Domain Method for Generation of Discrete-Time Analytic Signals., IEEE trans on signal processing, vol. 54, no. 9, september (2006)
10. Yongxin, Zhang.: Acoustic model and pronunciation adaptation in automatic speech recognition, University of Miami, (2006)
11. Linde, Y., Buzo, A., R.M, Gray.: An algorithm for vector quantizer design, IEEE Trans on Communications, Vol. 28, pp. 84-95, Jan. (1980)
12. Pearlman, W.A., Gray, R.M.: Source coding of the discrete Fourier transform. IEEE trans. on information theory. IT-24 (6): 683-692. (1978)
13. Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D, Povey, D., Valtchev, V., Woodland, P.: The HTK Book, HTK Version 3.4, Cambridge University Engineering Department, (2006)

Satellite Image Classification Using a Divergence-Based Fuzzy c-Means Algorithm

Dong-Chul Park

Dept. of Electronics Engineering, Myong Ji University, Korea
parkdc@mju.ac.kr

Abstract. A satellite image classifier scheme by using a Fuzzy c-Means (FcM) algorithm is proposed in this paper. The FcM algorithm adopted in this paper is a Gradient-based FcM with Divergence measure (GFcM(D)) and it utilizes the Divergence measure to exploit the statistical nature of the image data and thereby improves the classification accuracy. Experiments and results on a set of satellite images demonstrate that the proposed GFcM(D)-based classifier outperforms conventional algorithms such as the traditional Self-Organizing Map (SOM) and Fuzzy c-Means (FcM) in terms of classification accuracy.

1 Introduction

Traditionally, conventional clustering algorithms such as the Self Organizing Map (SOM) [1] and the k-means algorithm [2] have seen the widest use in practice. However, they assign an object to a single class and ignore the possibility that the object may also belong to other classes. Fuzzy clustering techniques have also been proposed for clustering problems. One of the most widely used algorithms employing fuzzy clustering techniques is the Fuzzy c-Means (FcM) algorithm. The FcM algorithm was originally introduced by Bezdek in 1981 as an improvement on earlier clustering algorithms such as the SOM and the k-Means [3]-[5]. In the FcM, an object can belong to several classes at the same time but with different degrees of certainty, which are measured by the membership function. The FcM algorithm has more robust capabilities in comparison with the SOM and k-means and has been successfully applied to many clustering applications.

The Gradient-based Fuzzy c-Means (GFcM) algorithm introduced by Park [6,7] overcomes the drawback that each iteration requires the use of all the data at once. GFcM combines the characteristics of the SOM (presenting one datum at a time and applying the gradient descent method) and the FcM algorithm (continuous values of the membership grades in the range $[0, 1]$). In the FcM algorithm, all the data are present in the objective function, and the gradients are set to zero in order to obtain the equations necessary for minimization. In contrast, only one datum at a time is required in the GFcM.

In this paper, a classification method for satellite image data by employing the GFcM algorithm with Divergence measure (GFcM(D)) [8,9] is proposed. While

the GFcM algorithm has been shown to give high clustering accuracy [6], it is also been demonstrated that the divergence measure can provide better modeling of statistic data such as image data. Therefore, this combination is expected to yield an improvement for image classification in terms of classification.

The remainder of this paper is organized as follows. Section 2 summarizes the Fuzzy c-Means and Gradient-based Fuzzy c-Means algorithms. Section 3 describes the Gradient-based Fuzzy c-Means algorithm with Divergence measure. Section 4 presents experiments and results on satellite image data sets including comparisons with other conventional algorithms. Conclusions are presented in Section 5.

2 Gradient-Based Fuzzy c-Means(GFcM) Algorithm

2.1 Fuzzy c-Means(FcM) Algorithm

Bezdek first generalized the *fuzzy ISODATA* by defining a family of objective functions $J_m, 1 < m < \infty$, and established a convergence theorem for that family of objective functions [3,4]. For FcM, the objective function is defined as :

$$J_m(U, \mathbf{v}) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ki})^m (d_i(\mathbf{x}_k))^2 \tag{1}$$

where $d_i(\mathbf{x}_k)$ denotes the distance from the input data \mathbf{x}_k to \mathbf{v}_i , the center of the cluster i , μ_{ki} is the membership value of the data \mathbf{x}_k to the cluster i , and m is the weighting exponent, $m \in 1, \dots, \infty$, while n and c are the number of input data and clusters, respectively. Note that the distance measure used in FcM is the Euclidean distance.

Bezdek defined a condition for minimizing the objective function with the following two equations [3,4]:

$$\mu_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{d_j(\mathbf{x}_k)}{d_j(\mathbf{x}_k)}\right)^{\frac{2}{m-1}}} \tag{2}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (\mu_{ki})^m \mathbf{x}_k}{\sum_{k=1}^n (\mu_{ki})^m} \tag{3}$$

The FcM finds the optimal values of group centers iteratively by applying Eq. (2) and Eq. (3) in an alternating fashion.

2.2 Gradient-Based Fuzzy c-Means(GFcM) Algorithm

The FcM in Eq. (2) and Eq. (3) uses all data to update the center value of the cluster, but the GFcM that is used in this paper was developed to update the center value of the cluster with a given individual data sequentially [6,7]. Given one datum \mathbf{x}_k and c clusters with centers at $\mathbf{v}_j, (j = 1, 2, \dots, c)$, the objective function to be minimized is:

$$J_k = \mu_{k1}^2 (\mathbf{v}_1 - \mathbf{x}_k)^2 + \mu_{k2}^2 (\mathbf{v}_2 - \mathbf{x}_k)^2 + \dots + \mu_{kc}^2 (\mathbf{v}_c - \mathbf{x}_k)^2 \tag{4}$$

with the following constraint:

$$\mu_{k1} + \mu_{k2} + \dots + \mu_{kc} = 1 \tag{5}$$

The basic procedure of the gradient descent method is that starting from an initial center vector, $\mathbf{v}_i(0)$, the gradient ΔJ_k of the current objective function can be computed. The next value of \mathbf{v}_i is obtained by moving to the direction of the negative gradient along the error surface such that:

$$\mathbf{v}_i(n + 1) = \mathbf{v}_i(n) - \eta \frac{\partial J_k}{\partial \mathbf{v}_i(n)}$$

where n is the iteration index and

$$\frac{\partial J_k}{\partial \mathbf{v}_i(n)} = 2\mu_{ki}^2(\mathbf{v}_i(n) - \mathbf{x}_k)$$

Equivalently,

$$\mathbf{v}_i(n + 1) = \mathbf{v}_i(n) - 2\eta\mu_{ki}^2(\mathbf{v}_i(n) - \mathbf{x}_k) \tag{6}$$

where η is a learning constant.

A necessary condition for optimal positions of the centers for the groups can be found by the following:

$$\frac{\partial J_k}{\partial \mu} = 0 \tag{7}$$

After applying the condition of Eq. (7) , the membership grades can be found as:

$$\mu_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{d_i(\mathbf{x}_k)}{d_j(\mathbf{x}_k)}\right)^2} \tag{8}$$

More detailed explanation about GFcM can be found in [6][7].

3 GFcM with Divergence Measure

In addition to the advantages of GFcM, GFcM was extended to another version that can deal with the probabilistic data. In the extended version, the selection of a proper distance measure between two data vectors should be extremely important since the performance of the algorithm largely depends on the distance measure to be adopted[4]. After evaluating various distance measures, the Divergence distance (*Kullback-Leibler Divergence*) between two Gaussian Probability Density Functions(GPDFs), $\mathbf{x} = (x_i^\mu, x_i^{\sigma^2})$ and $\mathbf{v} = (v_i^\mu, v_i^{\sigma^2})$, $i = 1, \dots, d$, has been chosen as the distance measure in GFcM(D) [4][11]:

$$\begin{aligned} D(\mathbf{x}, \mathbf{v}) &= \sum_{i=1}^d \left(\frac{x_i^{\sigma^2} + (x_i^\mu - v_i^\mu)^2}{v_i^{\sigma^2}} + \frac{v_i^{\sigma^2} + (x_i^\mu - v_i^\mu)^2}{x_i^{\sigma^2}} - 2 \right) \end{aligned}$$

$$= \sum_{i=1}^d \left(\frac{(x_i^{\sigma^2} - v_i^{\sigma^2})^2}{x_i^{\sigma^2} v_i^{\sigma^2}} + \frac{(x_i^\mu - v_i^\mu)^2}{x_i^{\sigma^2}} + \frac{(x_i^\mu - v_i^\mu)^2}{v_i^{\sigma^2}} \right) \tag{9}$$

where x_i^μ and $x_i^{\sigma^2}$ denote μ and σ^2 values of the i^{th} component of \mathbf{x} , respectively, while v_i^μ and $v_i^{\sigma^2}$ denote μ and σ^2 values of the i^{th} component of \mathbf{v} , respectively.

The GFcM used in this paper is based on the FcM algorithm. However, instead of calculating the center parameters of the clusters after applying all the data vectors in the FcM, the GFcM updates their center parameters at every presentation of data vectors. By doing so, the GFcM can converge faster than the FcM [6,7]. To deal with probabilistic data such as the GPfD, the GFcM(D) updates the center parameters, mean and variance, according to the distance measure shown in Eq. (9). That is, the membership grade for each data vector \mathbf{x} to the cluster i is calculated by the following equation:

$$\mu_i(\mathbf{x}) = \frac{1}{\sum_{j=1}^c \left(\frac{D(\mathbf{x}, \mathbf{v}_i)}{D(\mathbf{x}, \mathbf{v}_j)} \right)^2} \tag{10}$$

After finding the proper membership grade from an input data vector \mathbf{x} to each cluster i , the GFcM(D) updates the mean and variance of each center as follows:

$$\mathbf{v}_i^\mu(n+1) = \mathbf{v}_i^\mu(n) - \eta \mu_i^2(\mathbf{x})(\mathbf{v}_i^\mu(n) - \mathbf{x}^\mu) \tag{11}$$

$$\mathbf{v}_i^{\sigma^2}(n+1) = \frac{\sum_{k=1}^{N_i} (\mathbf{x}_{k,i}^{\sigma^2}(n) + (\mathbf{x}_{k,i}^\mu(n) - \mathbf{v}_i^\mu(n))^2)}{N_i} \tag{12}$$

where

- $\mathbf{v}_i^\mu(n)$ and $\mathbf{v}_i^{\sigma^2}(n)$: the mean and variance of the cluster i at the time of iteration n
- $\mathbf{x}_{k,i}^\mu(n)$ and $\mathbf{x}_{k,i}^{\sigma^2}(n)$: the mean and variance of the k^{th} data in the cluster i at the time of iteration n
- \bullet η and N_i : the learning gain and the number of data in the cluster i

GFcM(D) has been successfully applied to various data classification problems [8]- [10].

4 Experiments and Results

For the evaluation of the proposed image data classifier based on GFcM(D), satellite image data sets collected. The satellite image data set consists of different image classes (categories) in which each class contains different areas [15,15].

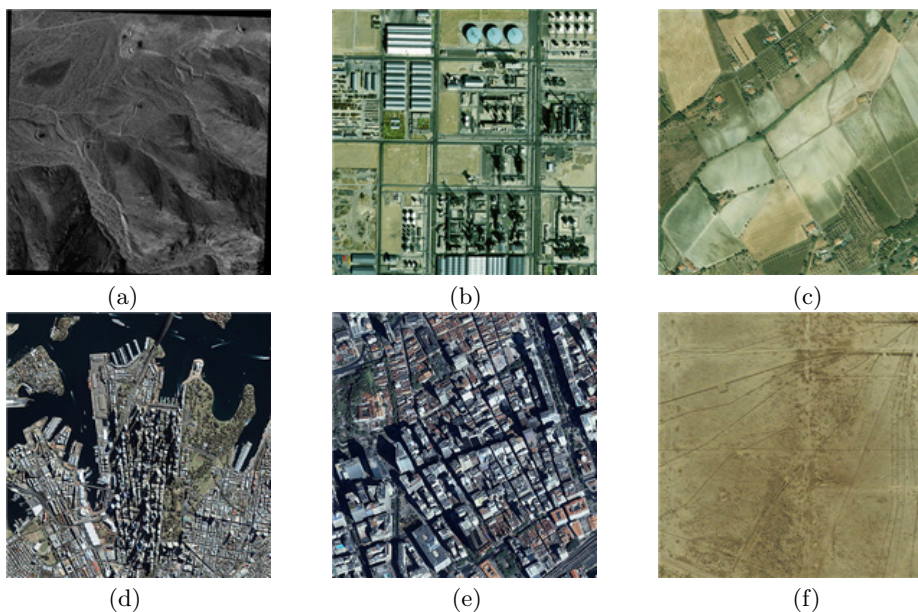


Fig. 1. Examples of satellite image data set:(a) Mountain area, (b) Factory area (c)Farm area, (d) Harbor area, (e) Urban area, and (f) Desert Area

Fig. 1 shows examples of mountain area, factory area, farming area, harbor area, urban area, and desert area. Each class consists of 100 images with different views resulting in a total of 600 images in the data set.

In order to obtain the texture information from the image, conventional texture descriptors based on a frequency domain analysis such as Gabor filters [12] and wavelet filters [13] are often used. However, these algorithms often induce a high computational load for feature extraction and are not suitable for real-time applications. In this paper, the Discrete Cosine Transform (DCT) is adopted for extracting the texture information from each block of the image [14]. The DCT transforms the image from the spatial domain into the frequency domain.

For the localized representation, images are transformed into a collection of 8×8 blocks. Each block is then shifted by an increment of 2 pixels horizontally and vertically. The DCT coefficients of each block are then computed and returned in 64 dimensional coefficients. Only the 32 lowest frequency DCT coefficients that are visible to the human eye are kept. Finally, a GPDF with a 32-dimensional mean vector and a 32×32 covariance matrix is used to represent the content of the image.

For each image class, the distribution of its feature vectors by a number of code vectors is calculated. Each code vector represents a group with its own mean and covariance matrix. During testing, the class of each image is decided by using a Bayesian classifier. To evaluate the proposed algorithm, its performance is compared with the performances of conventional algorithms such as SOM and

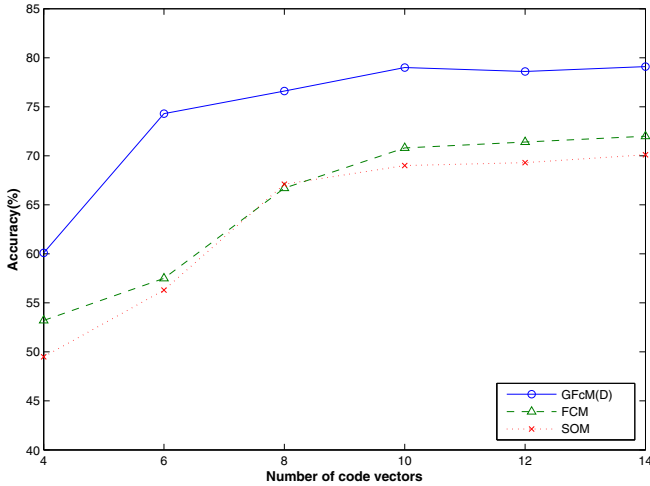


Fig. 2. Overall classification accuracies using different algorithms

Table 1. Classification accuracy (%) of different algorithms using 12 code vectors

	Mountain	Factory	Farm	Harbor	Urban	Desert	Overall
SOM	58.3	67.1	66.3	73.2	81.5	69.2	69.3%
FcM	63.6	70.2	65.7	75.6	83.2	70.0	71.4%
GFcM(D)	70.3	72.5	70.4	85.6	90.4	82.1	78.6%

FcM. Fig. 2 shows the performance in terms of the correct classification for three algorithms with several numbers of code vectors in a range from 4 to 14. Table 1 shows the performance for each image class, using different algorithms with 12 code vectors. The classification performances for different algorithms are fairly saturated with 12 code vectors. Note that the SOM and the FcM algorithm do not use the covariance information. From the result shown in Fig. 2 and Table 1, we can infer that the algorithm that uses the covariance information, GFcM(D), usually outperforms the SOM and the FcM which use the Euclidean distance as their distance measure. The results also show that GFcM(D) provides far better accuracy over the other two algorithms.

5 Conclusion

In this paper, a new approach for classification of satellite images using a clustering algorithm is proposed. This paper shows how the mean and variance information of satellite image data are utilised using the Gradient-Based Fuzzy c-Means algorithm with Divergence measure. Experiments are successfully performed on a database with each image class such as mountain area, factory area,, farm area, harbor area, urban area, and desert area. Based on the results of this

experiment and previous experiments, it is expected that the GFcM(D) algorithm will have broad applicability. However, the classification results between mountain area and farm area or between urban area and factory area are far from satisfactory. More research on the selection of feature values in addition to the DCT values will help to discriminate some areas. Future work will include some standard evaluation protocols and baseline algorithms for the object recognition task in addition to FcM and SOM.

Acknowledgments. This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (2010-0009655) and by the IT R&D program of The MKE/KEIT (10040191, The development of Automotive Synchronous Ethernet combined IVN/OVN and Safety control system for 1Gbps class).

References

1. Kohonen, T.: The Self-Organizing Map. Proc. of IEEE 78, 1464–1480 (1990)
2. Hartigan, J.: Clustering Algorithms. Wiley, New York (1975)
3. Bezdek, J.C.: A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. Pattern Anal. Mach. Intel., 1–8 (1980)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York (1981)
5. Windham, M.P.: Cluster Validity for the Fuzzy cmeans clustering algorithm. IEEE Trans. Pattern Anal. Mach. Intel., 357–363 (1982)
6. Park, D.-C., Dagher, I.: Gradient Based Fuzzy c-Means Algorithm. In: Proc. of IEEE Int. Conf. on Neural Networks, pp. 1626–1631 (1994)
7. Looney, C.: Pattern Recognition Using Neural Networks, pp. 252–254. Oxford University Press, New York (1997)
8. Park, D.-C.: Classification of MPEG VBR Video Data Using Gradient-Based FCM with Divergence Measure. In: Wang, L., Jin, Y. (eds.) FSKD 2005. LNCS (LNAI), vol. 3613, pp. 475–483. Springer, Heidelberg (2005)
9. Park, D.-C., Nguyen, D.-H., Beack, S.-H., Park, S.: Classification of Audio Signals Using Gradient-Based Fuzzy c-Means Algorithm with Divergence Measure. In: Ho, Y.-S., Kim, H.-J. (eds.) PCM 2005. LNCS, vol. 3767, pp. 698–708. Springer, Heidelberg (2005)
10. Park, D.-C., Woo, D.-M.: Image Classification Using Gradient-Based Fuzzy c-Means with Divergence Measure. In: Proc. of IJCNN, pp. 2520–2524 (2008)
11. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press Inc. (1990)
12. Daugman, J.G.: Complete Discrete 2D Gabor Transform by Neural Networks for Image Analysis and Compression. IEEE Trans. Acoust., Speech, and Signal Processing 36, 1169–1179 (1988)
13. Pun, C.M., Lee, M.C.: Extraction of Shift Invariant Wavelet Features for Classification of Images with Different Sizes. IEEE Trans. Pattern Anal. Mach. Intel. 26(9), 1228–1233 (2004)
14. Huang, Y.L., Chang, R.F.: Texture Features for DCT-Coded Image Retrieval and Classification. In: Proc. of IEEE ICASSP, vol. 6, pp. 3013–3016 (1999)
15. Park, D.-C.: Classification of Satellite Images Using Partitioned Feature-based Classifier Model. In: Proc. of ICISA (2011)
16. Park, D.-C., Jeong, T., Lee, Y., Min, S.-Y.: Satellite Image Classification Using a Classifier Integration Model. In: Proc. of AICCSA, pp. 90–94 (2011)

Classifiers Combination for Arabic Words Recognition: Application to Handwritten Algerian City Names

Soulef Nemouchi¹, Labiba Souici Meslati², and Nadir Farah³

¹ EPSECG, Ecole Préparatoire des Sciences Economiques,
Commerciales et Sciences de Gestion, Annaba, Algeria

² LRI Laboratory, Badji Mokhtar - Annaba University, Algeria

³ LabGED Laboratory, Badji Mokhtar - Annaba University, Algeria
soulef_inf@yahoo.fr

Abstract. In this paper, we present a global recognition system for Arabic handwritten words; we focus on the two phases of feature extraction and classification. In our system, we have retained three feature sets. The Zernike moments and the structural features of the word are extracted from the binary image, the Freeman code is established from the contour image of the word and the zoning is given from the skeleton image. These features, representing the words, are extracted to be used as input, in an individual or combined way, of the four classifiers used in our system: the Fuzzy C-Means algorithm (FCM), the K-Means algorithm, the K Nearest Neighbor algorithm (KNN) and a Probabilistic Neural Network (PNN). The system architecture is a parallel one where each expert (classifier) gives his point of view and we combine the results to make a final decision. The classifier results are combined using two methods: the simple vote and the weighted sum.

Keywords: Arabic handwriting recognition, Fuzzy C-Means (FCM), K Nearest Neighbor algorithm (KNN), K-Means algorithm, Probabilistic Neural Network (PNN), zernike moments, zoning, Freeman chain code.

1 Introduction

Communicate by writing has always been a first concern for humans who want to create an easy and direct interaction with a computer. This gives labor to the researchers in the “writing recognition” field, especially for handwriting recognition, which is the dream of all those who need to enter data in a computer.

The first research in this field was done more than thirty years ago. Nowadays, there are several applications in which the recognition of handwritten writing is required like the automatic mail sorting, the automatic processing of administrative documents, the investigation forms or the automatic reading of postal addresses and bank checks [1]. Contrary to Latin, the recognition of the handwritten or printed Arabic writing is, till now, in the research and experimentation level.

Unconstrained off-line handwriting recognition remains a challenging problem. Word recognition algorithms suffer from two major problems. One is the

segmentation error given by the word segmentation process, especially for cursive handwriting documents. The other is that the accuracy of recognition drops when the size of the lexicon increases [2].

On the other hand, given the number and variety of methods used in pattern recognition, there is no single method that can be called the *best*. Each approach has strengths and weaknesses, good ideas and bad. One way to take advantage of this variety is to build multiple sources of information based systems. This direction is given more attention in pattern recognition and more work is being done, especially for handwriting recognition applications. The reported results show the efficiency of such techniques including hybrid approaches and multiple classifier schemes, especially for Arabic recognition [3, 4].

The hybrid approaches are represented by those recognition systems that use different sources of information either at the feature extraction level, by using several types of primitives to better describe the input word/character, or at the classification stage by integrating two (or more) complementary classification paradigms, or at both feature extraction and classification levels. The multiple classifier approach is defined as a system consisting of a set of classifiers and a decision combination function. It applies a number of generally independent classifiers and combines their results to generate a single decision.

In our previous works, we dealt with the recognition of handwritten Arabic words in literal amounts [5] using single classifiers (structural, neural, statistical...) having as input several kinds of features. We focused later on multiple classifiers and hybrid systems; we were particularly attracted by the integration of neural and symbolic approaches. We built neuro-symbolic hybrid and multiple classifiers for Arabic literal amounts [6, 7, 8].

In this article, we focus on classifier combination, we propose a system (see figure 1) to recognize handwritten words in Algerian city names lexicon. This system combines several types of classifiers and features to enhance the recognition accuracy.

This article is structured into seven sections. We first describe the overall architecture of the proposed system in the Arabic handwritten words recognition. In the third section, we present the preprocessing operations performed on word images. The fourth section is devoted to the description of our choices for the feature extraction phase while section six is dedicated to the recognition phase. In the last sections, we present the results followed by some conclusions and perspectives.

2 Overview of the Proposed System

An operational pattern recognition system contains a set of processing modules where feature extraction and classification stages are the most important for its overall performance. The feature extraction methods are generally specific to each particular pattern recognition problem, whereas the same classification algorithms can be used in various applications.

There are three ways to approach word recognition problems [9]. The system recognizes the word as an entire and indivisible entity, it is the *global* or *holistic* approach, or it recognizes the word starting from its previously segmented characters, it is the *analytical* approach. The third way consists in the one used by *human reading based systems* which use only some properties and refine gradually, in loops, the word

description [10]. In our work, we have focused on the holistic approach because the considered vocabulary is limited to the Algerian city-names.

Many works show that the combination of classifiers (sequential, parallel or hybrid) improves significantly the performances of the recognition system compared with each classifier separately [7, 11, 12]. In our work, we focus on the parallel combination because it is the most used one in the pattern recognition problems and it proved its effectiveness in many classification tasks [8, 12, 13]. This success is due to its implementation simplicity and its capacity to explore the answers of different classifiers to be combined, by taking into account (or not) the behavior of each classifiers. In this parallel combination, each classifier has to recognize the entire word using its global features. The difference between these classifiers is their word processing manner according to their functional principles. The figure 1 gives the general scheme of the proposed system components which will be described in the following sections. This classifier fusion is made either by a democratic way where there is no superiority of any classifier compared to another, or in a directed manner where the answer of each classifier is weighted according to its performances.

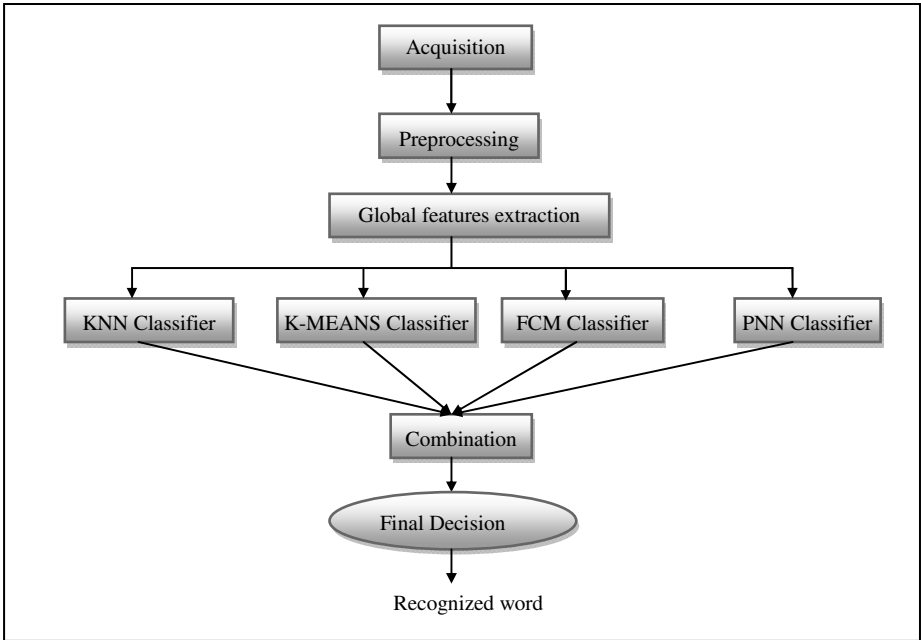


Fig. 1. Overview of the proposed handwritten word recognition system

3 Acquisition and Preprocessing

In our work, we use the Algerian city-names images database, built in the LRI laboratory at Annaba University [14]. The images acquisition was done a scanner and, before being analyzed, the images are submitted to some preprocessing operations.

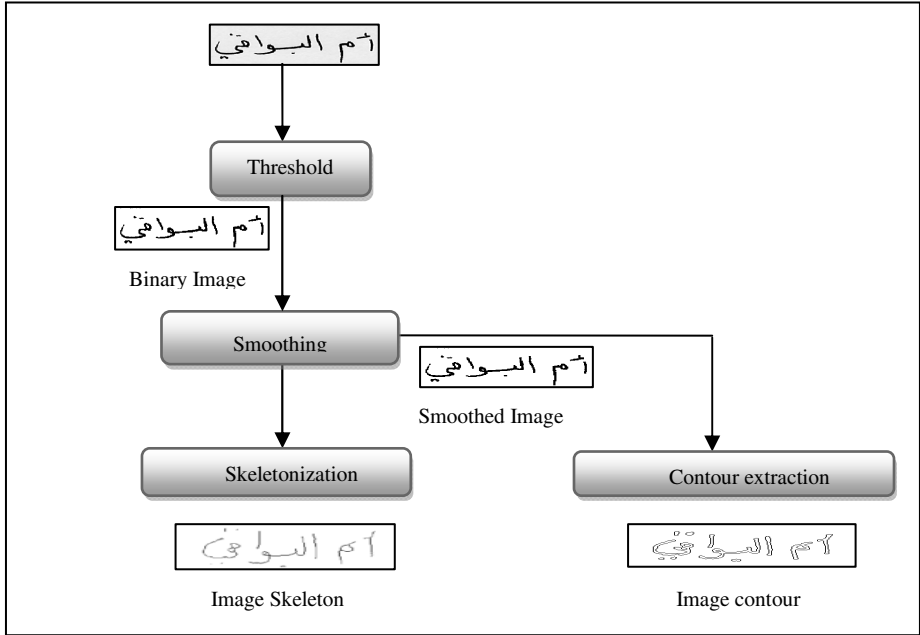


Fig. 2. Image preprocessing

Preprocessing poses major problems: some loops can be opened or not detected and some diacritic dots can be eliminated or confused with noise.

4 Extraction of Global Features and Word Description

The feature extraction problem consists in extracting, from data, the information which is most relevant for classification purposes, in the sense of minimizing the intra-class pattern variability while enhancing the inter class pattern variability [2, 4].

Handwriting recognition systems typically involve two steps: feature extraction in which the patterns are represented by a set of features and classification in which decision rules for separating pattern classes are defined.

The feature extraction phase must ensure a maximum of reliability, because the later phases will not handle the original image but use the results provided by this module.

In the literature, several works concern the elaboration of new features which have an increasing discriminative capacity while minimizing intra-class variability. These features are generally classified in two families: structural features (like strokes, concavities, end points, intersections of line segments, loops, stroke relations . . .) and the statistical features which derive from spatial measurements of the pixels (zoning, invariants moments, Fourier descriptors, Freeman chain code... etc).

We can represent image in different forms: gray-scale, binary, contour, skeleton... and different features can be extracted from each form. Our goal, in this work is to find

the best image representation for the considered Arabic words. Three sets of features are retained (see figure 3), our choice was based on the work done by Arrivault [10].

In our system, we retained the global features from the three image representation forms, for each one we choose the appropriate feature extraction methods. The Zernike moments are extracted from binary image, the Freeman chain code is extracted from the image contour, and zoning is done on the image skeleton. We have also retrained 9 global structural features:

- The number of ascenders in each connected component.
- The number of descenders in each connected component.
- The number of loops in each connected component.

And the diacritic dots which are:

- The number of a high single dot in each connected component.
- The number of two high dependent dots in each connected component.
- The number of three high dependent dots in each connected component.
- The number of a low single dot in each connected component.
- The number of two low dependent dots in each related component.
- And a statistical characteristic which represents the percentage of each component connected in the word (the number of the connected components).

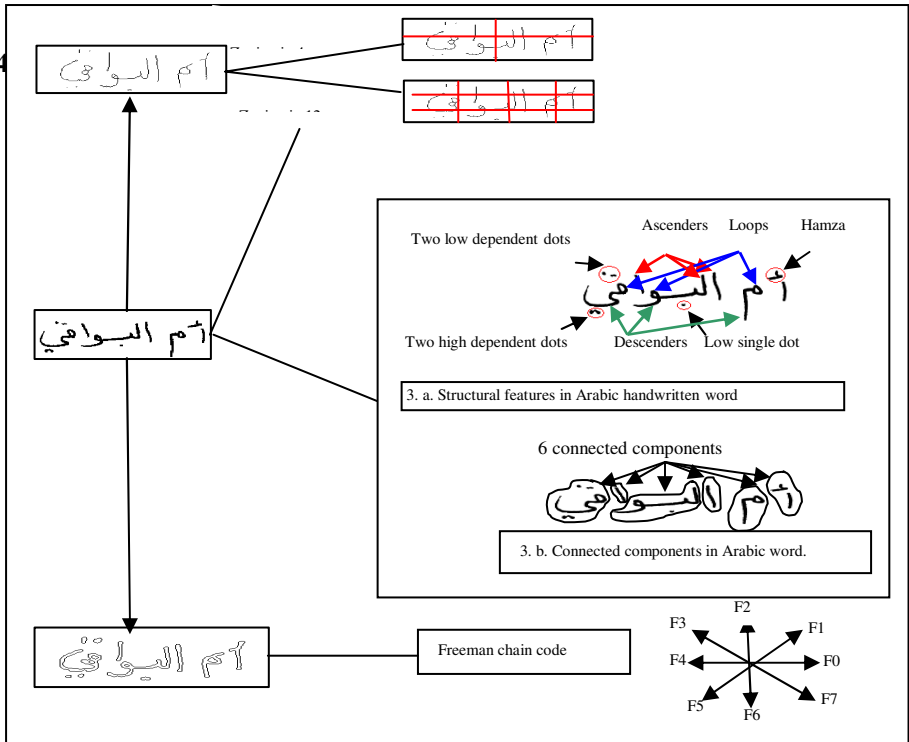


Fig. 3. Chosen features for Arabic handwritten word description

In our work, the extraction of these structural features consists in processing on connected component detected in the preprocessing phase, on the labeled image, from right to left according to the orientation of Arabic writing. The figure 3 illustrates these features. At the end of the feature extraction phase, a word is described by a features vector containing 133 elements:

- 9 structural features.
- 100 Zernike moments.
- 8 Freeman chain code.
- 16 zoning features.

5 Recognition

There are broadly two main approaches for classification: the statistical approach consisting in representing a pattern as an ordered, fixed-length list of numerical values and the structural approach describing the pattern as an unordered, variable-length list of simple shapes. The statistical approach relies on firmly established elements of statistical decision theory, even though viewing a pattern in an n -dimensional space is rather difficult. On the contrary, the structural approach is intuitively appealing because it appears closer to human recognition strategy. Unfortunately, this approach is usually difficult to implement in a fast, trainable and robust way over a large variety of shapes.

As a structural classifier is naturally well-suited to the use of structural features but cannot easily handle statistical features, we have chosen the frame of statistical classification to investigate the combination of both structural and statistical features in a one-shot classifier. In the other way we have chosen the neural networks because several neural network models have been proposed for various difficult problems, especially classification ones. Traditional classifiers test the competing hypothesis sequentially, whereas neural network classifiers test the competing hypothesis in parallel, thus providing high computational rates [15].

As each classification method has its advantages and shortcomings, we can deduce that the performance of system can be increased significantly by combining multiple classifiers. Thus, as shown in figure 1, we propose a system which combines, with two parallel combination methods, four classifiers:

- A K Nearest Neighbor (KNN) classifier.
- A K-Means classifier.
- A Fuzzy C-Means (FCM) classifier.
- A Probabilistic Neural Network (PNN).

We have chosen these four types of classifiers for several reasons. We used the FCM method because it doesn't require prior information about the classes; this algorithm has been already used successfully in image segmentation. In addition, FCM is a

classification method which allows a data sample belonging to two classes or more (with a membership degree for each class) contrary to the K-MEANS algorithm where the data sample must belong or not to one class. We thought to implement these two methods for comparing performances between the obtained results to deduce the influence of fuzzy logic in the area of images classification (in our case, the images represent handwriting Arabic words). The third implemented algorithm is KNN, opposing to FCM, it requires a reference base (word images already classified). The latter chosen approach is the neural one because neural networks have a large capacity of classification and showed their skills in handwriting recognition.

6 Results and Discussion

In this paragraph, we give some classification results with different pairs of (features types/ classifiers). The aim of our work is to find the most interesting combinations at feature and classification levels. The results are summarized in tables 1 and 2.

Note that we have divided our base into two parts, one for training (270 images for each city name) and the second for the test (30 images for each city name).

Table 1. Recognition rates for individual classifiers

Features \ Classifiers	FCM(%)	KMeans(%)	KNN (%)	PNN (%)
Zoning features	65.24	67.32	76.76	78.00
Zernike moments	59.95	58.91	61.80	74.50
Freeman chain code	69.50	66.63	74.87	77.21

The table 1 gives the recognition rates for the four individual classifiers using different feature sets. From these results, we can conclude that the most interesting features are the structural and zoning features. In addition, we note that Gaussian parametric evaluation (PNN) gets good results with zoning. Overall, the two classifiers KNN and PNN reach comparable results which are more interesting than the two others.

At the end of our work, we combined the four classifiers. This combination is made either in a democratic way, in the sense that it does not prefer any classifier compared to another, or by attributing, to the answer of each classifier, a weight according to its performances generally based on the rate obtained in training phase. According to the obtained results (Table 1) we have given priority to the PNN and KNN classifiers while the two others have the same priority. The results are summarized in the following table:

Table 2. Classifiers combination results

Features \ Combination	Without priority between classifiers (%)	With priority between classifiers (%)
Zoning features	70.80	72.09
Zernike moments	61.00	62.03
Freeman chain code	69.65	70.18
All features	77.50	79.80

We have evaluated the performance of our system; we have tested it on database containing 1440 words images. Approximately 80% of these words were properly assigned to the correct class when using all features. This result is very encouraging in handwriting word recognition.

7 Conclusion and Perspectives

In this paper, we have presented a system for holistic (global) handwritten Arabic words recognition, which combines the strengths of both statistical and structural feature extractors thanks to a combination of four complementary families of features (ranging from pure structural to pure statistical and including both local and global features). In the classification phase, our system combines four different types of classifiers in order to get a better recognition rate. The obtained results are interesting and encouraging. They experimentally confirm the assumption that the combination of multiple classifiers decisions and the use of different feature types enhance the overall accuracy of a recognition system. However, for each new pattern recognition systems, problems still remain: How many classifiers and what kind of classifiers should be used? For each classifier, what types of features should be chosen?. Moreover, this involves multiple tedious learning steps for both the chosen features, classifiers and combination rules. We have several possibilities for the evolution of our work: we think that the performance of the system can be increased by combining an analytical approach with the holistic proposed one. We can also increase the performance of our system by automatically selecting the most relevant features with the use of feature selection methods. We can also investigate the field of classifiers ensembles and dynamic classifier selection.

References

1. Bunke, H.: Recognition of Cursive Roman Handwriting - Past, Present and Future. In: International Conference on Document Analysis and Recognition, ICDAR, Edinburgh, Scotland (2003)
2. Steinherz, T., Rivlin, E., Intrator, N.: Offline cursive script recognition: a survey. IJDAR, International Journal on Document Analysis and Recognition 2, 90–110 (1999)
3. Essoukhri Ben Amara, N., Bouslama, F.: Classification of Arabic script using multiple sources of information: state of the art and perspectives. IJDAR, International Journal on Document Analysis and Recognition 5, 195–212 (2003)

4. Lorigo, L.M., Govindaraju, V.: Offline Arabic Handwriting Recognition: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(5), 712–724 (2006)
5. Souici, L., Aoun, A., Sellami, M.: Global recognition system for Arabic literal amounts. In: *ICCTA 1999*, Alexandria, Egypt (1999)
6. Souici-Meslati, L., Sellami, M.: A hybrid approach for Arabic literal amounts recognition. *AJSE* 29(2B), 177–194 (2004)
7. Farah, N., Ennaji, A., Khadir, T., Sellami, M.: Benefit of multiclassifier systems for Arabic handwritten words recognition. In: *International Conference on Document Analysis and Recognition, ICDAR*, Seoul, Korea, vol. 1, pp. 222–226 (2005)
8. Farah, N., Souici, L., Sellami, M.: Classifiers combination and syntax analysis for Arabic literal amount recognition. *Engineering Applications of Artificial Intelligence* 19(1), 29–39 (2006)
9. Vinciarelli, A.: A survey one off-line cursive Word recognition. *Pattern Recognition* 35(7), 1433–1446 (2002)
10. Arrivault, D.: *Apport des graphes dans la reconnaissance non-contrainte de caractères manuscrits anciens*. PhD Thesis, University of Poitiers, France (2002) (in French)
11. Al-Ohali, Y.: *Handwritten Word Recognition – Application to Arabic Cheque Processing*. PhD Thesis, Concordia University, Montreal, Canada (2002)
12. Azizi, N., Farah, N., Sellami, M.: Off-line handwriting word recognition using ensemble of classifier selection and features fusion. *JATIT, Journal of Theoretical and Applied Information Technology* 14(2), 141–150 (2010)
13. Zouari, H., Heutte, L., Lecourtier, Y., Alimi, A.: Building Diverse Classifier Outputs to Evaluate the Behavior of Combination Methods: The Case of Two Classifiers. In: Roli, F., Kittler, J., Windeatt, T. (eds.) *MCS 2004*. LNCS, vol. 3077, pp. 273–282. Springer, Heidelberg (2004)
14. Souici-Meslati, L., Sellami, M.: Toward a generalization of neuro-symbolic recognition: an application to arabic words. *KES, International Journal of Knowledge-Based and Intelligent Engineering Systems* 10(5), 347–361 (2006)
15. Prema, K.V., Subba Reddy, N.V.: Two-tier architecture for unconstrained handwritten character recognition. *Sadhana* 27, Part 5, 585–594 (2002)

Robust Arabic Multi-stream Speech Recognition System in Noisy Environment

Anissa Imen Amrous and Mohamed Debyeche

Speech Communication and Signal Processing Laboratory (LPCTS),
Faculty of Electronics and Computer Sciences, USTHB
P.O. Box 32, Bab Ezzouar, Algiers, Algeria
amrous_im@hotmail.fr, mdebyeche@gmail.com

Abstract. In this paper, the framework of multi-stream combination has been explored to improve the noise robustness of automatic speech recognition systems. The main important issues of multi-stream systems are which features representation to combine and what importance (weights) be given to each one. Two stream features have been investigated, namely the MFCC features and a set of complementary features which consists of pitch frequency, energy and the first three formants. Empiric optimum weights are fixed for each stream. The multi-stream vectors are modeled by Hidden Markov Models (HMMs) with Gaussian Mixture Models (GMMs) state distributions. Our ASR is implemented using HTK toolkit and ARADIGIT corpus which is data base of Arabic spoken words. The obtained results show that for highly noisy speech, the proposed multi-stream vectors leads to a significant improvement in recognition accuracy.

Keywords: Multi-stream speech recognition, HMM, noisy environments.

1 Introduction

Improve the robustness of automatic speech recognition in presence of additive noise has become an active topic and a number of techniques has been proposed to improve word accuracies in noisy environments. The use of multi-stream models is one such technique [1]. A multi-stream speech recognizer is based on the combination of multiple feature streams each containing complementary information. The performance of such system depends on the fact that the selected features for every stream must not go through the same distortion in presence of noise. The weight given to each stream is another important aspect in multi-stream combination system. The rule should be such that the streams that are reliable should get more weight compared to the stream corrupted by noise [2], [3], [4].

We can refer to many works that tried to improve the robustness of ASR system by using several streams of features that rely on different underlying assumptions and exhibit different properties. Shimmer and jitter are used in [5], and formant and auditory-based acoustic cues are used together with MFCC in [6], [7]. In [8], [9], a multi-stream approach is used to combine MFCC features with formant estimates and

a selection of acoustic cues such as acute/grave, open/close, tense/lax, etc. Pitch has been also taken into account in many works for the recognition of tonal languages [10], [11]. For the same purpose, many works in audio-visual domain have investigated the contribution of the visual information on the acoustic recognition system in noisy environments [12], [13].

This work aims to improve ASR system in noisy environments by using a new multi stream vector based on MFCC, pitch, energy and the three first formants. The remainder of the paper is organized as follows: the multi-stream HMM based ASR systems are presented in section 2. In section 3, the experiments setup and results are given. Finally, we draw conclusions in Section 4.

2 Multi-stream HMM Based ASR System

The schematic overview of the multi-stream system is shown in Fig.1. Where γ_i ($i=1,2,..N$) is the stream weight of stream i , and it can be fixed statically [14] or estimated dynamically [3], [4]. Each stream is composed of a set of features and the N streams are combined to form a multi-stream vector at the input of the multi-stream modeling unit. In the test step the multi-stream features are decoded by a usual decoding ASR system unit.

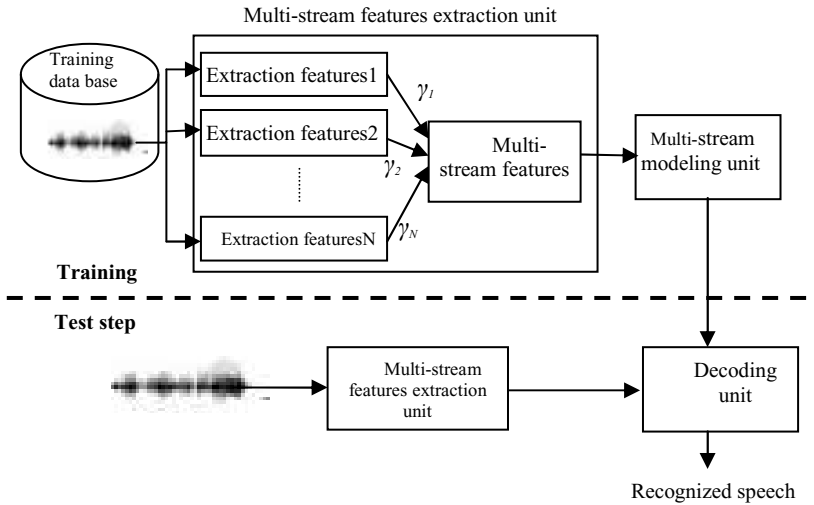


Fig. 1. Multi-stream HMM based ASR system

2.1 Multi stream Features

We describe in this section some of the theoretical background of the two stream features used in this work.

2.1.1 Stream1: MFCC Features

Our first stream is made up from the Mel-Frequency Cepstral Coefficients (MFCCs) [15] and their first (Δ) and second ($\Delta \Delta$) derivatives. For each analysis window, the MFCCs coefficients are calculated by equation (1), as follows:

$$MFCC(n) = \sum_{m=0}^{M-1} E[m] \cos\left(\frac{\pi m(m+\frac{1}{2})}{M}\right) \quad 0 \leq n \leq M \quad (1)$$

where M is the number of filter bank channels and $E[m]$ is the energy of a given filter.

2.1.2 Stream 2: Complementary Features

The second stream consists of three kinds of features, namely pitch, energy and the first three formants. To complete the stream, the first and the second order derivatives of the five features are added.

According to the literature [16], [17], [18], those features are less affected by noise comparing to the usual features such as MFCC [15], PLP[19] and LPC [20] which represent the vocal tract characteristics and are very susceptible to noise.

2.1.2.1 Pitch. Its estimation is based on autocorrelation function [21]. Given a speech window $\{s(n), n = 0, 1, \dots, N_s - 1\}$ the autocorrelation function is defined as

$$R(k) = \frac{1}{N} \sum_{n=0}^{N_s-1-k} s(n)s(n+k), \quad k = 0, \dots, N_s - 1 \quad (2)$$

where N_s is the number of autocorrelation points to be computed.

2.1.2.2 Formant frequencies. In this paper we choose to use the frequencies of the first three formants which are estimated from the maxima of the LPC spectrum model [22]. These maxima are defined as the complex roots of the following polynomial:

$$1 + \sum_{i=1}^P a_i z^{-i} = 0 \quad (3)$$

where p is the LPC order.

2.1.2.3 Energy. Is defined as the variation of the signal amplitude caused by the force coming from the pharynx. The energy was computed by taking the logarithm of the windowed signal $(s_t)_{t=1,T}$ [23]:

$$E = \frac{1}{T} \sum_{t=1}^T s_t^2 \quad (4)$$

where T is the window signal $(s_t)_{t=1,T}$ size.

2.2 Multi-stream Modeling

A multi-stream model is a product model of the different feature streams. For S independent streams, the output distribution for state j using a Gaussian mixture is defined as

$$b_j(o_t) = \prod_{s=1}^S \left[\sum_{m=1}^{M_s} c_{j sm} N(o_{st}; \mu_{j sm}, \Sigma_{j sm}) \right]^{\gamma_s} \quad (5)$$

where M_s is the number of mixture components for stream s , $c_{j sm}$ is the weight of the m -th component and $N(o; \mu, \Sigma)$ is a multivariate Gaussian with mean μ and covariance Σ [23]. The exponent γ_s is the weight for stream s .

2.3 Decoding

The decoding unit calculates the likelihood between the word to recognize and all the acoustic models which are already trained in the training step. The recognized word is the one which corresponds to the acoustic model according to the maximum likelihood. This likelihood was performed using the Viterbi algorithm [23].

3 Experimental Setup

This section presents the database and the experimental setup used for the evaluation of the proposal multi-stream HMM based ASR system.

3.1 Database Description

The speech database used in this work is the isolated ARADIGIT corpus [24]. It is composed of Arabic isolated digits from 0 until 9. This database is divided into the following corpora:

- Train corpus: consisting of 1800 utterances pronounced by 60 speakers including the two genders, where, each speaker repeats the same digit 3 times.
- Test corpus: consisting of 1000 utterances pronounced by 50 speakers including the two genders, where, each speaker repeats the same digit 2 times.

3.2 Muti-stream Feature Extraction

- Stream1: For the first stream, MFCC features are extracted by HTK [23]. The speech signal is divided into a number of overlapping time windows of 25 ms with a frame period of 10 ms. For each analysis window, 12 MFCC features with their delta and acceleration coefficients, resulting in a feature vector of 36 acoustic features (MFCC_D_A) has been extracted.
- Stream2: The complementary features of the second stream which are: pitch, energy and the first three formants are extracted by the Praat package [25] based on algorithms described in section 2.1.2. Delta and accelerations coefficients are added to this stream by HTK, making a total vector stream size of 15 coefficients (Comp_D_A).

3.3 Experimental Methodology

Our experiments were developed using HTK package (Hidden Markov Toolkit) [23], from Cambridge University. With the aim to show the advantage of using multi-stream features in speech recognition under real-life test conditions, we carried out a set of experiments. Four ASR systems are built:

- 1. Single stream1 ASR system:** uses as observation vectors, features of stream1.
- 2. Single stream2 ASR system:** uses as observation vectors, features of stream1.
- 3. Equally-weighted multi-stream ASR system:** uses as observation vectors, features of stream1 concatenated to stream2 features. The two stream are equally-weighted ($\gamma_1 = \gamma_2 = 1$).
- 4. Optimally-weighted multi-stream ASR system:** uses as observation vectors, features of stream1 concatenated to stream2 features. The two streams are optimally weighted. The optimum weights are chosen empirically from experiences. Stream1 weights for each of the SNR's are as shown in Table 1. The weights of the second stream may be computed from this table using $\gamma_2 = 2 - \gamma_1$.

Table 1. Stream1 weights

SNR(dB)	20dB	15dB	10dB	5dB	0dB	-5dB
γ_1	1.1	1.1	1.1	1.1	0.9	0.9

The HMM models used for the all systems are a left-to-right HMM with continuous observation densities. Each model consists of 3 states, in which, each state is modeled by 1 Gaussian mixture with a diagonal covariance matrices defined as in equation (5).

To simulate the adverse conditions of test, we have corrupted the database by an airport noise extracted from the NOISEX92 database [26] and added to the speech signal with SNR ranging from -5 dB to 20 dB.

The acoustic models' training uses the clean speech database; the noise is only added for testing the recognition performance.

Table 2. Comparative speech recognition results

SNR (dB)	20dB	15dB	10dB	5dB	0dB	-5dB
Single stream1						
ASR system	80.81%	67.99%	49.91%	30.35%	17.99%	9.96%
Single stream2						
ASR system	59.32%	56.09%	52.82%	41.97%	26.66%	18.36%
Equally-weighted multi-stream ASR system	84.96%	75.92%	65.41%	46.68%	32.75%	20.76%
Optimally-weighted multi-stream ASR system	85.89%	76.38%	66.05%	47.51%	32.93%	21.49%

3.4 Results

Table 2 gives the results for the implemented ASR systems in different test conditions. Best results in terms of word recognition accuracy are edited in bold. For single stream systems, the ASR system based on stream1 (MFCC) outperform the one based on stream2 in quite noisy environments (20dB, 15 dB). In highly noisy environments, it is the single stream2 system which performs better than the single stream1 one. for instance, at 5 dB 41.97vs. 30.35. This is due to the fact that the proposed complementary features were more robust to noise comparing to the MFCC features.

As it can be observed, overall (SNR = -5 to 20 dB), the multi-stream systems, either equally-weighted or optimally-weighted, shows an improvements in word accuracy over the single stream systems. Another interesting aspect of these results is that the improvement in word accuracies is more pronounced in cases of low SNRs. For example, with 5dB : 47.51%% vs. 30.35%, i.e., an improvement of 17.16%is noticed.

It can be seen that the optimally-weighted system gives a better word accuracy when compared to the equally weighted system by about 1%. This shows the important role of weights in multi-stream framework.

4 Conclusions

In this paper, we have studied the contribution of a new multi-stream vector for Arabic speech recognition system based on Hidden Markov Model. The new multi-stream vector is consisted of the standard cepstral features MFCC, and a set of

complementary features namely, pitch, energy and the first three formants. Results show that with these complementary features we can get significant word accuracy improvement over both single and multi stream ASR systems.

References

1. Janin, A., Ellis, D., Morgan, N.: Multi-stream speech recognition: ready for prime time. In: Proc. of Eurospeech, Budapest (1999)
2. Guo, H., Chen, Q., Huang, D., Zhao, X.: A Multi-stream Speech Recognition System Based on The Estimation of Stream Weights. In: Proc. ICISP, pp. 3479 – 3482 (2010)
3. Sanchez-soto, E., Potamianos, A., Daoudi, K.: Unsupervised stream weights computation in classification and recognition Tasks. IEEE Trans. Audio, Speech and Language Processing 17(3), 436–445 (2009)
4. Potamianos, A., Sánchez-Soto, E., Daoudi, K.: Stream weight computation for multi-stream classifiers. In: Proc. ICASSP, pp. 353–356 (2006)
5. Li, X., Tao, J., Johanson, M.T., Soltis, Savage, J.: Stress and emotion classification using jitter and shimmer features. In: Proc. ICASSP, vol. 4, pp. IV-1081–IV-1084(2007)
6. Holmes, J.N., Holmes, W.J.: Using formant frequencies in speech recognition. In: Proc. Eurospeech, Rhodes, pp. 2083–2086 (1997)
7. Selouani, S.A., Tolba, H.: Distinctive features, formants and cepstral coefficients to improve automatic speech recognition. In: Proc. IASTED, pp. 530–535 (2002)
8. Selouani, S.A., Tolba, H., O’Shaughnessy, D.: Auditory-based acoustic distinctive features and spectral cues for automatic speech recognition using a multi-stream paradigm. In: Proc. of ICASSP, pp. 837–840 (2002)
9. Tolba, H., Selouani, S.A., O’Shaughnessy, D.: Comparative experiments to evaluate the use of auditory-based acoustic distinctive features and formant cues for robust automatic speech recognition in low snr car environments. In: Proc. of Eurospeech, pp. 3085–3088 (2003)
10. Chongjia, N.I., Wenju, L., Xu, B.: Improved Large Vocabulary Mandarin Speech Recognition Using Prosodic and Lexical Information in Maximum Entropy Framework. In: Proc. CCPR 2009, pp. 1–4 (2009)
11. Ma, B., Zhu, D., Tong, R.: Chinese Dialect Identification Using Tone Features Based on Pitch Flux. In: Proc. ICASSP, p. I (2006)
12. Gurbuz, S., Tufekci, Z., Patterson, E., Gowdy, John, N.: Multi-stream product modal audio-visual integration strategy for robust adaptive speech recognition. In: Proc. ICASSP, pp. II-2021–II-2024 (2002)
13. Guoyun, L.V., Dongmei, J., Rongchun, Z., Yunshu, H.: Multi-stream Asynchrony Modeling for Audio-Visual Speech Recognition. In: Proc. ISM, pp. 37–44 (2007)
14. Addou, D., Selouani, S.A., Boudraa, M., Boudraa, B.: Transform-based multi-feature optimization for robust distributed speech recognition. In: Proc. GCC, pp. 505– 508 (2011)
15. Davis, S.B., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. Proc. IEEE Trans. ASSP 28, 357–366 (1980)
16. Mary, L., Yegnanarayana, B.: Extraction and representation of prosodic features for language and speaker recognition. Proc. Speech Communication 50, 782–796 (2008)
17. Doss, M.: Using auxiliary sources of knowledge for automatic speech recognition. Ph.D Theses; École Polytechnique Fédérale de Lausanne (2005)

18. Ververidis, D., Kotropoulos, C.: Emotional speech recognition: resources, features, and methods. *Proc. Speech Communication* 48, 1162–1181 (2006)
19. Hermansky, H.: Perceptual linear predictive (PLP) analysis of speech. *The Journal of the Acoustical Society of America* 87, 1738–1752 (1990)
20. Slifka, J., Anderson, T.R.: Speaker modification with lpc pole analysis. In: *Proc. of ICASSP*, pp. 644–647 (1995)
21. Rabiner, L.R.: On the Use of Autocorrelation Analysis for Pitch Detection. *IEEE Transaction on Acoustics, Speech, and Signal Processing* 25, 1 (1977)
22. Davis, S.B., Mermelstein, P.: Comparison of parametric representations for monosyllable word recognition in continuously spoken sentences. *IEEE Trans. on Speech and Audio Processing* 28(4), 357–366 (1980)
23. Young, S., Odell, J., et al.: *The HTK Book Version 3.3*. Speech group, Engineering Department. Cambridge University Press (2005)
24. Amrouche, A.: *Reconnaissance automatique de la parole par les modèles connexionnistes*. Ph.D Theses, Faculty of Electronics and Computer Sciences, USTHB (2007)
25. Boersma, P., Weenink, D.: *Praat: doing phonetics by computer* (2008), <http://www.praat.org/>
26. Varga, A.P., Steeneken, H.J.M., et al.: The NOISEX-92 study on the effect of additive noise on automatic speech recognition. In: *NOISEX 1992 CDROM* (1992)

SVM Based GMM Supervector Speaker Recognition Using LP Residual Signal

Dalila Yessad and Abderrahmane Amrouche

Speech Communication and Signal Processing Laboratory,
Faculty of Electronics and Computer Sciences, USTHB,
P.O. Box 32, El Alia, Bab Ezzouar, 16111, Algiers, Algeria
yessad.dalila@gmail.com, namrouche@usthb.dz

Abstract. Feature extraction is an important step for speaker recognition systems. In this paper, we generated MFCC (Mel Frequency Cepstral Coefficients) and LPCC (Linear Predictive Cepstral Coefficients) from LP residual of speech signal, instead their calculation directly from speech samples. These features represent complementary vocal cord information's. In this work, Universal Background Gaussian Mixture Models (GMM-UBM) and Gaussian Supervector (GMM-SVM) based speaker modeling have been used. Experimental results, using, ARADIG-ITS data-base, show the efficiency of the GMM-SVM based approach associated with feature vectors issued from LP residual signal.

Keywords: LPC, LPCC, LP residual, MFCC, GMM-UBM, GMM-SVM, Speaker Recognition.

1 Introduction

Voiced speech is usually used for speaker recognition. But in text-independent speaker recognition it would be better to use special voiced phonemes which are present in all words. In the source-filter model of human speech production, the speech signal is modeled as the convolutional output of a vocal source excitation signal and the impulse response of a vocal tract filter system [1]. Cepstral features [2] such as Mel-frequency cepstral coefficients (MFCCs) and linear predictive cepstral coefficients (LPCCs) have been the dominant features for a long time in speaker recognition. These features are believed to provide pertinent cues for phonetic classification and have been successfully implemented in most existing speaker recognition systems [3]. This indicates that MFCC and LPCC features capture properties of vocal tract and contain important speaker-specific information. Since MFCCs capture a mixture of phonemic and speaker-related information, their use has resulted in good performance in speaker recognition. In [4], the standard procedures for extracting MFCC and LPCC features were applied to LP residual signals, resulting in a set of residual features for speaker recognition. In linear predictive (LP) modeling of speech signals, the vocal tract system is represented by an all-pole filter. The prediction error, which is named the LP residual signal, contains useful information about the source excitation.

In [5], the speaker information present in LP residual signals was captured using an auto-associative neural network model and in [6] features extracted from linear predictive (LP) analysis were used. Despite these investigations, state-of-art systems are mostly based on the Mel cepstral frequency coding (MFCC) or the linear predictive cepstral coding (LPCC). Indeed, these short-term features have proven their efficiency in terms of performances and are adapted for the Gaussian mixture models (GMMs).

Current state of the art systems for text-independent speaker recognition use cepstral coefficients as base features, and speaker modeling techniques, such as Universal Backgrounds Gaussian Mixture Models (GMM-UBM) and Gaussian Supervector (GMM-SVM). These later are two successful approaches recently used. The first approach uses a speaker model which is formed by MAP adaptation of the means of the UBM. In the second approach, the GMM supervector is formed by stacking all mean vectors of the adapted model and is classified using a Support Vector Machines (SVM)[7], [8], [9].

This paper deals with the MFCC and LPCC feature extraction techniques based on LP residual signal. Section 2 provides feature extraction technique. Then, Sections 3 elaborates speaker modeling principles. In Section 4, we discuss the evaluation of speaker recognition performance, followed by conclusion in Section 5.

2 Feature Extraction

2.1 Linear Prediction (LP) Residual

Linear prediction (LP) is the process of predicting future sample values of a digital signal from a linear system. It is therefore about predicting the signal $x(n)$ at instant n from p previous samples as in Eq. (1)

$$x(n) = \sum_{i=1}^p a_i x_{n-i} + G\epsilon(n) \quad (1)$$

Where a_1, a_2, \dots, a_p are the Linear Prediction Coefficients (LPCs), p is the model order, G and $\epsilon(n)$ are the excitation gain and source, respectively. The LPCs are derived adaptively for each 20-30 ms speech frame by minimization of excitation mean square energy. For simplicity, we will assume that the order of LP model is uneven, $p = 2m - 1$. The LPC spectrum or the transfer function of the LP filtering is defined by:

$$H(z) = \frac{G}{A(z)} \quad (2)$$

Where

$$A(z) = 1 - \sum_{i=1}^{2m-1} a_i z^{-i} \quad (3)$$

2.2 Cepstral Linear Prediction Coding (LPCC)

The cepstrum coefficients $\{ceps_q\}_{q=0}^Q$ can be estimated from the LPC coefficients $\{a_q\}_{q=1}^p$ using a recursion procedure:

$$ceps_q = \begin{cases} \ln(G), & q = 0 \\ a_q + \sum_{k=1}^{q-1} \frac{k-q}{q} a_k ceps_{q-k}, & 1 \leq q \leq p \\ \sum_{k=1}^p \frac{k-q}{q} a_k ceps_{q-k}, & p < q \leq Q \end{cases} \quad (4)$$

Where G is the gain term in the LPC model, p the LPC model order, and $Q + 1$ the number of cepstrum coefficients.

2.3 Mel Frequency Cepstral Coefficients (MFCC)

The most commonly used feature vector in speech recognition is composed of Mel-Frequency Cepstral Coefficients (MFCC). The MFCC extraction is done in three steps:

1. Step 1-a: Cut up the signal in several overlapping windows;
2. Step 1-b: To decrease the spectral distortion, a Hamming windowing is applied to signal frames;

$$W(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi n}{N-1}\right) \quad (5)$$

Where N is the window size.

3. Step 2-a: Apply the FFT ;
4. Step 2-b: The Mel-frequency scale is applied using the following transformation formula;

$$Mel(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \quad (6)$$

5. Step 2-c: Apply the logarithm after the Mel scale;
6. Step 3: Finally, obtain the discrete cosine transform (DCT) of the output signal.

3 Classifiers

3.1 Gaussian Mixture Model Universal Background (GMM-UBM)

The speaker recognition system is a Gaussian mixture model-universal background. The GMM-UBM approach is the state of the art system in text-independent speaker recognition [10]. This approach is based on a statistical modeling paradigm, where a hypothesis is modeled by a GMM model:

$$p(x|\lambda) = \sum_{i=1}^{i < m} \alpha_i N(x|\mu_i, \Sigma_i) \tag{7}$$

Where α_i , μ_i and Σ_i respectively, the weights, the mean vectors, and the covariance matrices (generally diagonal) of the mixture components. During a test, the system has to determine whether the recording Y was pronounced by a given speaker S . This question is modeled by the likelihood ratio;

$$\frac{p(x|\lambda_{hyp})}{p(x|\lambda_{\overline{hyp}})} \geq \tau \tag{8}$$

Where Y is the test speech recording, λ_{hyp} is the model of the hypothesis where S pronounced Y , $\lambda_{\overline{hyp}}$ corresponds to the model of the negated hypothesis (S did not pronounce Y), $p(y|m)$ is the GMM likelihood function, and τ is the decision threshold. The model $\lambda_{\overline{hyp}}$ is a generic background model, the so-called UBM, and is usually trained during the development phase using a large set of recordings coming from a large set of speakers. The model λ_{hyp} is trained using a speech record obtained from the speaker S . It is generally derived from the UBM by moving only the mean parameters of the UBM, using a Bayesian adaptation function.

In this study The GMM-UBM system is the LIA SpkDet system [11] based on the ALIZE platform3 and distributed under an open source license. This system produces speaker models using MAP adaptation by adapting only the means from a UBM. The UBM component was trained on a selection of 60 corpus. For all the experiments, the model size is 128 and the performances are assessed using DET plots and measured in terms of equal error rate (EER) and minimum of detection cost (minDCF).

3.2 Support Vector Machines (SVM)

The support vector machine (SVM) [8] is a two-class classifier constructed from sums of a kernel function $k(\cdot, \cdot)$,

$$f(x) = \sum_{i=1}^L \alpha_i t_i k(x, x_i) + d \tag{9}$$

Where t_i are the ideal outputs, $\sum_{i=1}^L \alpha_i t_i = 0$, $i = 0$ and $\alpha > 0$. The vectors x_i are support vectors and obtained from the training set by an optimization process [11]. The ideal output are either 1 or -1 , depending upon whether the corresponding support vector is in class 0 or class 1, respectively. For classification, a class decision is based upon whether the value, $f(x)$, is above or below a threshold. The kernel $k(\cdot, \cdot)$ is constrained to have certain properties, so that $k(\cdot, \cdot)$ can be expressed as :

$$k(x, y) = b(x)^t b(y) \tag{10}$$

Where $b(x)$ is a mapping from the input space (where x lives). For a separable data set, SVM optimization chooses a hyperplane in the expansion space with maximum margin [7], [8]. The data points from the training set lying on the boundaries are the support vectors in equation (1). The focus of the SVM training process is to model the boundary between classes in [7], [8].

3.3 GMM Supervector (GMM-SVM)

Gaussian mixture models with universal backgrounds is constructed by MAP adaptation of the means of the UBM. A GMM supervector is constructed by stacking the means of the adapted mixture components. We assume we are given a Gaussian mixture model universal background model (GMM-UBM):

$$p(x|\lambda) = \sum_{i=1}^{i < m} \alpha_i N(x|\mu_i, \sum_i) \quad (11)$$

Where α_i are the mixture weights, m indicates a Gaussian density and μ_i and \sum_i are the corresponding mean and covariance. From a speaker utterance, the GMM-UBM model is adapted by Maximum A Posteriori (MAP) adaptation to provide the speaker GMM model. Generally, only the means μ_i of Gaussian components are adapted. In this case, all GMMs have the same covariance matrices \sum_i and differ only in means. As a consequence, for SVM classification, each model is represented only by the concatenation of all GMM Gaussians mean vectors, that is, a GMM supervector [12].

4 Results and Discussions

4.1 Speech Database and Features Extraction

Arabic digits, which are polysyllabic, can be considered as representative elements of language, because more than half of the phonemes of the Arabic language are included in the ten digits. The speech database used in this paper is a part of the database ARADIGITS [13]. It consists of a set of 10 digits of the Arabic language (zero to nine) spoken by 60 speakers of both genders with three repetitions for each digit. This database was recorded by Algerian speakers from different regions aged between 18 and 50 years in a quiet environment with an ambient noise level below 35 dB, in WAV format, with a sampling frequency equal to 16 kHz. In this work we used the "long training / short test" for speaker recognition on ARADIGITS. The features corresponding to the six digits (from zero to five, with concatenation of three repetitions) are used for training each speaker model. Only 60 speakers of the database are used in the speaker identification system for testing. Four digits (from six to nine) of every speaker is tested separately (60x4=240 test patterns of seconds each, in average). The experiments are totally text independent. Speaker utterances were represented by 19 coefficients LPCC or MFCC, with their first derivatives and the delta energy. Altogether, a 40 coefficients vector is extracted from each LP residual, based speech signal frame. Mean subtraction

and variance normalization were applied to all features. Figure 1 shows the speech waveforms and the corresponding LP residual signals, of the vowel /a/ from the sound /wahid/ uttered by two different female speakers. We can see the differences between the two segments of residual signals. In addition to the difference between their pitch periods, the residual signal of speaker A shows much stronger periodicity than that of speaker B.

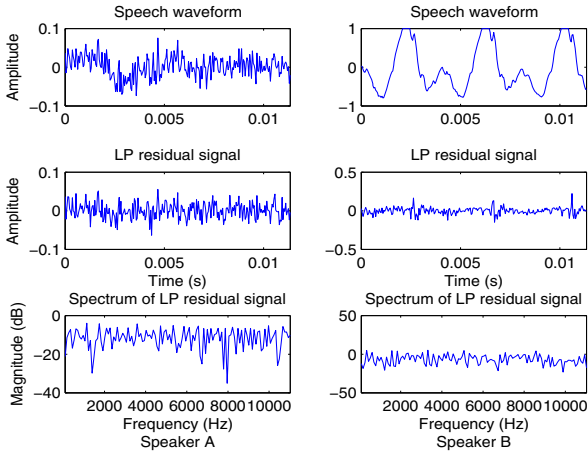


Fig. 1. Speech waveform, LP residual signals and Fourier spectra of LP residual signal of two female speakers; Speaker A in the left and speaker B in the right

4.2 Experimental Results

We evaluate the speaker recognition performances of MFCC and LPCC individually like baseline system, using both GMM-UBM and GMM-SVM classifiers. In addition, we evaluate these classifiers with MFCC or LPCC extracted from LP residual signal, using the same evaluation database. The EER performance of the baseline system are shown in Figure 2. The best performance are obtained with MFCC based GMM-SVM, 91% in average. Figure 3 shows the recognition performance of MFCC and LPCC features extracted from LP residual signal. The MFCC extracted from LP residual and based GMM-SVM achieved the best performance (it is found at 88% in average). Experimental results show that the GMM-SVM using MFCC features gives the best performances, and MFCC features outperform the MFCC extracted from LP residual, because in frequency domain, the useful temporal information, the amplitudes and the time locations of pitch pulses, are not represented in the Fourier spectra of LP residual. To characterize the time-frequency characteristics of the pitch pulses, others transformations like wavelet are more appropriate than the short-time Fourier transform.

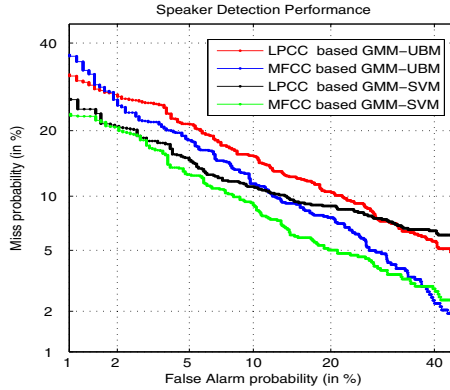


Fig. 2. The performance of GMM-UBM and GMM-SVM systems with MFCC and LPCC features extracted from ARADIGITS database

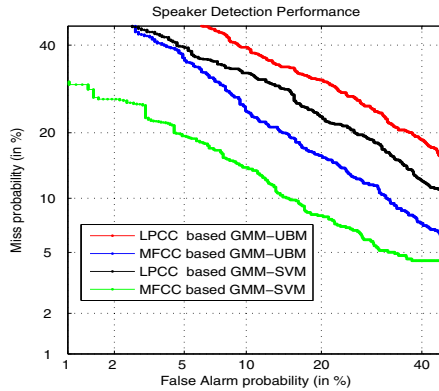


Fig. 3. The performance of GMM-UBM and GMM-SVM systems with MFCC and LPCC features extracted from LP residual signal

5 Conclusion

This paper investigates MFCC and LPCC features extraction from LP residual signal based on both GMM-UBM and GMM-SVM classifiers. We have shown that the MFCC and LPCC features based LP residual contain speaker-specific information for speaker recognition applications, and the MFCC features provide additional information in speaker recognition. This work shows the possibility of performing speaker recognition by extracting features directly from LP residual signal.

References

1. Rabiner, L.R., Schafer, R.W.: Digital Processing of Speech Signals. Prentice-Hall, Englewood Cliffs (1978)
2. Quatieri, T.F.: Discrete-Time Speech Signal Processing - Principles and Practice. Prentice-Hall (2002)
3. Reynolds, D.A.: An overview of automatic speaker recognition technology. In: Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP), pp. 4072–4075 (2002)
4. Chen, S.H., Wang, H.C.: Improvement of speaker recognition by combining residual and prosodic features with acoustic features. In: Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 93–96 (2004)
5. Yegnanarayana, B., Reddy, K.S., Kishore, S.P.: Source and system features for speaker recognition using AANN models. In: Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 409–413 (2001)
6. Mary, L., Sri Rama Murty, K., Mahadeva Prasanna, S.R., Yegnanarayana, B.: Features for speaker and language identification. In: Proc. of the ISCA Tutorial and Research Workshop on Speaker and Language Recognition (Odyssey 2004), pp. 323–328 (2004)
7. Dong, X., Zhaohui, W.: Speaker Recognition using Continuous Density Support Vector Machines. Electronics Letters 37(17), 1099–1101 (2001)
8. Wan, V., Renals, S.: SVM SVM: Support Vector Machine Speaker Verification Methodology. In: Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Hong Kong, vol. 2, pp. 221–224 (2003)
9. Karam, Z.N., Campbell, W.M.: A Multi-Class MLLR Kernel for SVM Speaker Recognition. In: Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 4117–4120 (April 2008)
10. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing 10(1-3), 19–41 (2000)
11. <http://www.lia.univ-avignon.fr/heberges/ALIZE/LIA/~RAL>
12. Campbell, W.M., Sturim, D.E., Reynolds, D.A.: Support Vector Machines using GMM supervectors for Speaker Verification. IEEE Signal Process. Lett. 13(5), 308–311 (2006)
13. Amrouche, A., Debyeche, M., Taleb Ahmed, A., Rouvaen, J.M., Ygoub, M.C.E.: Efficient System for Speech Recognition in Adverse Conditions Using Nonparametric Regression. Engineering Applications on Artificial Intelligence 23(1), 85–94 (2010)

Plugin of Recommendation Based on a Hybrid Method for the Ranking of Documents in the E-Learning Platforms

Hicham Moutachaouik¹, Hassan Douzi¹,
Abdelaziz Marzak², Hicham Behja³, and Brahim Ouhbi³

¹ Laboratory IRF-SIC, Faculty of Sciences Agadir,
University Ibn Zohr,
BP.8106 Hay Dakhla, Agadir, Morocco
gotohicham@gmail.com,
douzi_h@yahoo.fr

² Laboratory of Information Technologies and Modeling,
Faculty of Science Ben M'sik Casablanca,
BP.7955 Casablanca, Morocco
marzak@hotmail.com

³ Laboratory Command and Control of Production System,
ENSAM-Meknes,
BP.4024 Beni M'hamad, Meknes, Morocco
h_behja@yahoo.com,
ouhbib@yahoo.co.uk

Abstract. The objective of this work is the conception and the realization of a recommendatory system, using concepts of the web usage mining and being inspired by approaches to information filtering. This system includes a new hybrid method to rank documents web, in order to propose to the Webmaster (or admin) of platform e- learning the best available documents based of the historical to research done by learners.

It is, actually a meta-search engine on the web, integrated into the e-learning platform to keep surfing traces of the learner during his searching. This will permit to have a usage basis that will be used by the system to help webmaster (admin) to make decisions about the documents to be added to the platform. The elaborated system will make it possible to propose help and assistance to learners of the platform.

Keywords: e-learning, web usage mining, information filtering, ranking, recommendation system, Moodle, classification.

1 Introduction

Due to the exponential increase in the amount of resources available and accessible on the web, Recommendation systems have seen their popularity grow in recent years. Combining techniques of information filtering, personalization, artificial intelligence,

social networks and human-computer interaction, recommendation systems provide users with suggestions to meet their informational needs and preferences. Indeed, recommendation systems are particularly in demand in e-commerce applications. For example, the Amazon site recommends all kinds of products (movies, music, books, etc...) [1, 2].

To Produce recommendations, a number of approaches is possible: (i) the approach by content [3] which makes recommendations by comparing the semantic content of resources with the user's tastes, (ii) the approach based on knowledge [4] that makes recommendations by exploiting knowledge about the user and pre-established heuristics, and (iii) the approach by collaborative filtering [5], which makes recommendations by analyzing, at the same time, the user's opinions and those of other users about the resources they have consulted.

All these approaches require a ranking of documents before presented to the user system. For example, in the latest approach that we interests the most (collaborative filtering), documents is presented in order of decreasing evaluation, the latter is usually given as a vote.

This paper, proposes a Plugin what can integrate on any platform e-Learning (in our case, we used the platform open source moodle) to keep tracks of the web searches made by learners (student), in order to use it for the recommendation. This Plugin also includes a new hybrid method to rank documents web before presenting to the learner.

This ranking is given according to order of decreasing relevance score of documents deemed relevant. This score is calculated according three phases. First, we inspired from the collaborative filtering to calculate the score of vote for the document. Second, we use the formula chan to measure the degree of appreciation of the document by the learner. Finally, we develop a new function called λ method to combine these two measures (score of vote, formula chan) to calculate the relevance score.

The results show the quality of recommendations by contribution to our former work [6, 7] mainly through the addition of other criteria that have improved the results in order to help learners in their learning while trying to overcome the major problems of recommender systems(Critical Mass, Cold Start, Principle of induction) [7].

The rest of the article is organized as follows:

In section 2, we describe the working process of the proposed solution as well as the methodologies of ranking. In Section 3, we evaluated our Plugin by analyzing the learners' behaviors of the platform e-learning moodle. We terminate by conclusions and perspectives of our work in the section 4.

2 Process of Operation and Methodologie for Ranking of Document

2.1 Process of Operation

The system (Plugin) that was integrated in the platform e-learning to study the behavior of learners and produce recommendations is named MX-Search.

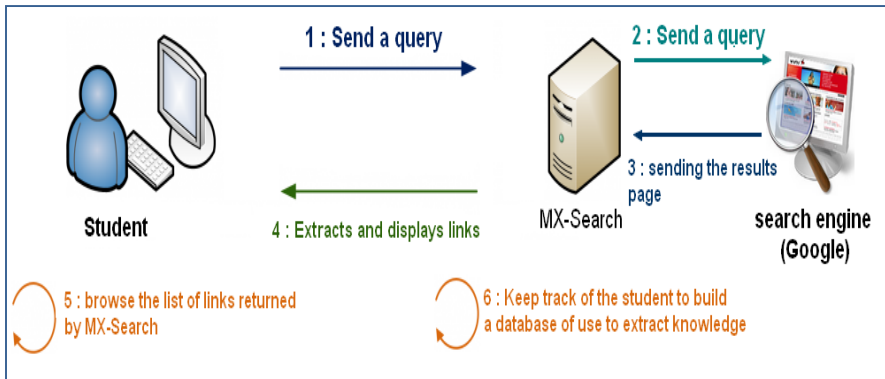


Fig. 1. Process of operation of MX-Search

MX-Search retrieves and displays the search results returned by the search engines (Google, Yahoo, Bing) while keeping track of the learner to make recommendations (figure1).

2.2 Allocation of the Link Visited in a Category of Courses

Before raking documents, we must classify the links returned by the search engine web. This will allow categorizing all the documents requested by all learners of the platform during their navigation.

To achieve this categorization, we utilize the method of classification using K-means [8]. We consider the categories of courses as classes that we attribute objects (web documents visited by the learner).

To calculate the similarity between a document and a category of course, we utilize the lexical similarity [9].

Algorithm of attribution of a link to a category of courses:

Algorithm classification

```

Var A,B,C,D,E : table of strings;
      H : table of table of strings ;
Begin
  H ← ∪ hi ; /; i ∈ [1.. number of modules] where hi is the
      whole of keywords of each module
  A ← extract keywords (query learning);
  B ← extract keywords (title of the link visited) ;
  C ← extract keywords ( link visited) ;
  D ← extract keywords (description of the link visited) ;
  E ← A ∪ B ∪ C ∪ D ;
  Calculation_of_similarity_between (E, hi) (we used
  Jaccard's index [9]) ;
  assign the link visited at the class (category of courses
  or module) which corresponds to hi ;
End.
  
```

2.3 Methodology Followed for Ranking of Documents

How to Calculate the Score of Vote?

To get the preferences of users, the collaborative filtering approach uses either the preference relation or the utility function (eg voting). Indeed in the second case which interests, us most is proposed to the user to give his opinion on a scale of integer values and relatively reduced task (usually a value between 1 and 5 or 1 and 7) [1]. But this voting task with note is very hard for the learner. This prompted us to propose a new scale of assessment instead of the whole scale of values.

To do this, we ask the learner for example:

How do you find the document?

Useless, Poor, Average, Good or Excellent

The response will be closer to the real context and facilitates the assessment of the learner. Then, we assign for each assessment a note (Useless=1, Poor=2, Passable =3, Good=4, Excellent=5).

Finally, we calculate the score of vote defined by:

$$\text{score of vote}(doc) = \text{sum}(ai \text{ vote}(i)) / \text{sum}(ai) \quad (1)$$

With ai : weight of vote according to the level of the learner

How to Calculate Formula Chan?

Explicit evaluations require more users' effort. As a result, users often tend to avoid this burden by leaving the system permanently or providing erroneous assessments.

In contrast, the deduction of such assessments by the single observation of user behavior is much less intrusive.

A real example of inference implicit evaluations is the formula proposed by Chan (1999) [2], to predict whether a web page has been appreciated or not.

This formula is based largely on information that can be harvested from the data protocol communication. Indeed, it is calculated based on the history, the bookmark, the contents of pages and the access log.

Finally, Chan (1999) defines the degree of interest in a page:

$$\text{Interest}(Page) = \text{Frequency}(Page) * (1 + \text{IsBookmark}(Page) + \text{Duration}(Page) + \text{Recency}(Page) + \text{LinkVisitPercent}(Page)) \quad (2)$$

In our system this formula was modified by adding weights to the variables in order to promote and give prominence to a variable contribution to the other. Formula Chan becomes:

$$\text{Interest}(Page) = \text{Frequency}(Page) * (1 + \alpha * \text{IsBookmark}(Page) + \beta * \text{Duration}(Page) + \delta * \text{Recency}(Page) + \gamma * \text{LinkVisitPercent}(Page)) \quad (3)$$

With $\alpha, \beta, \delta, \gamma$: the weight of the variable and $0 < \alpha, \beta, \delta, \gamma < 1$ and $\alpha + \beta + \delta + \gamma = 1$

By experimentation, to have a better recommendation, we offer $\alpha=\delta= \gamma=0.3$ and $\beta=0.1$.

Raking Using the λ Method?

Our contribution in this regard is made using our method that we have developed called the λ method, which consisting of combining the explicit criterion (score of vote) with the implicit criterion (Formula Chan) to produce recommendations.

$$score\ of\ pertinence\ (doc) = \lambda\ score\ of\ vote(doc) + (1 - \lambda)\ Interest(doc) \tag{4}$$

With λ : the weight of the variable and $0 < \lambda < 1$

To keep the weight distribution of the variables and for given the same importance, we give 0.5 to λ value.

3 Exploration and Analysis of the Behavior of Students of the Platform E-Learning Moodle: LP Java/C++

To test the functioning of our Plugin MX-Search, we integrated it in the platform e-learning moodle [10, 11] of the license professional Java/C++ of the faculty of sciences of Casablanca, Morocco.

3.1 Assist the Learner to Browse the Result of His Research on the Web

The learner from the platform e-learning LP Java/C++ can carry out research on the web while remaining on the platform (figure2, 3, 4).

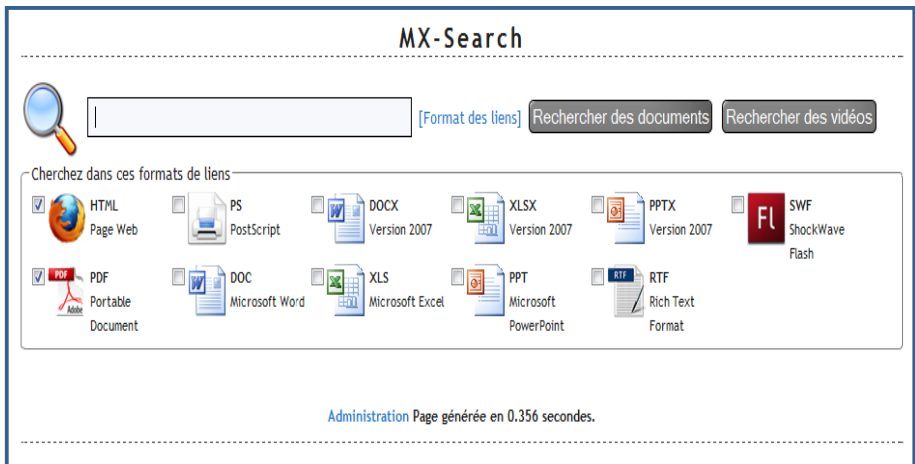


Fig. 2. Interface to conduct research on the web

Numéro	Titre	Nombre de clic	Appréciations
#1	Apprendre Java	0	
#2	Du C/C++ à Java : Table des matières	76	Bien : 6 Moyen : 12 Nul : 3 Pas Mal : 6
#3	Java - page de ressources et cours	81	Bien : 9 Excellent : 2 Moyen : 15 Nul : 2 Pas Mal : 3
#4	Cours Java de Patrick Iley - INRIA Sophia Antipolis	51	
#5	Cours et applets Java	77	Bien : 8 Excellent : 3 Moyen : 8 Pas Mal : 4
#6	Programmation en Java	118	Bien : 10 Excellent : 14 Moyen : 12
#7	Java [Cours/Formation à télécharger. pdf. zip ou .rar	3	
#8	Support de cours JAVA	78	Bien : 1 Moyen : 11 Pas Mal : 3
#9	Cours document langage de programmation Java C C++	76	
#10	Apprentissage de Java (tm) - F. Rossi	75	
#11	Cours POO-JAVA	74	
#12	Cours sur C C++ JAVA en français	75	Bien : 3 Excellent : 20 Moyen : 6 Nul : 3 Pas Mal : 6
#13	improve-technologies.com - Formation objet - Support de cours, pdf ...	0	
#14	Cours de JavaScript et DHTML [L'éditeur JavaScript]	74	
#15	Livre cours bases de l'informatique, Java et C#	74	

Fig. 3. Interface for the result of the search for a learner

Site Web : www.dailymotion.com
 URL : http://www.dailymotion.com/video/x6y9eu_ic2t-html-cours-2-creer-un-lien-ver_lifestyle

Votre appréciation ? Nul Pas Mal Moyen Bien Excellent

[Retour à la recherche](#)

http://ic2t.unblog.fr/

Flash Marionnettes : voyage
 Par Gejal
 ★★★★★ 396 vues
 Découvrez la nouvelle création théâtrale de la compagnie "Flash"

IC2T HTML cours #2 : créer u
 Par Gejal
 ★★★★★ 598 vues
 Deuxième volet de videocours IC consacré à l'apprentissage du

IC2T HTML cours #1 : créer u
 Par Gejal
 ★★★★★ 1278 vues

Générique gg02 echo TV
 Par Gejal
 ★★★★★ 41 vues
 Bienvenue sur gge2echo/TV

Administration

Fig. 4. Interface to view the contents of the link web and vote

3.2 Guide the Administrator of the Platform to Extract Knowledge

The administrator (web master or teacher) the platform e-learning moodle of the license professional Java/C++ can choose the chain, the desired module as well as the criterion of ranking to make decisions concerning web documents to add in the platform e-learning (figure5).

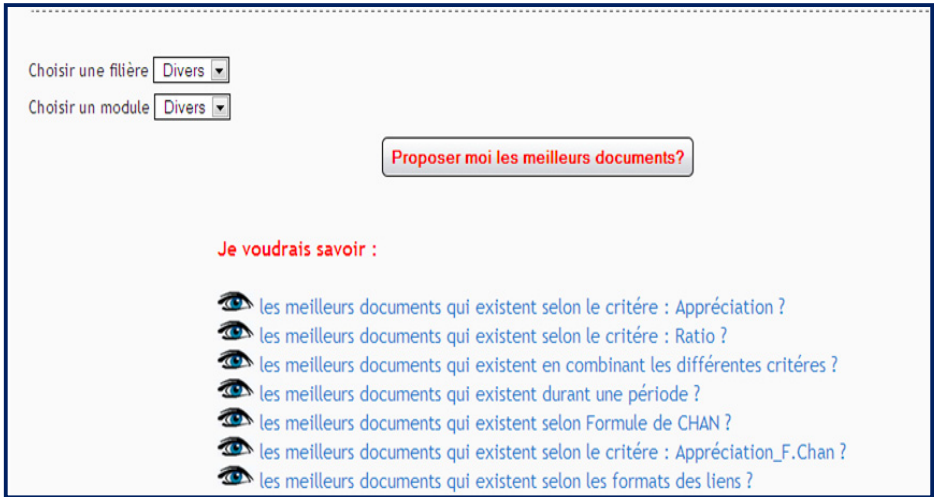


Fig. 5. Interface to choose the criteria of knowledge extraction

After having chosen the java module for example, the system ranks the documents according to their relevance according to the point of view of the learner and l or formula chan (figure 6, 7, 8.). This ranking will allow learners to have additional resources to understand the module.



Fig. 6. Result obtained for the module java according to the criterion: Score of vote



Fig. 7. Result obtained for the module java according to the combination of criteria and document type (html)

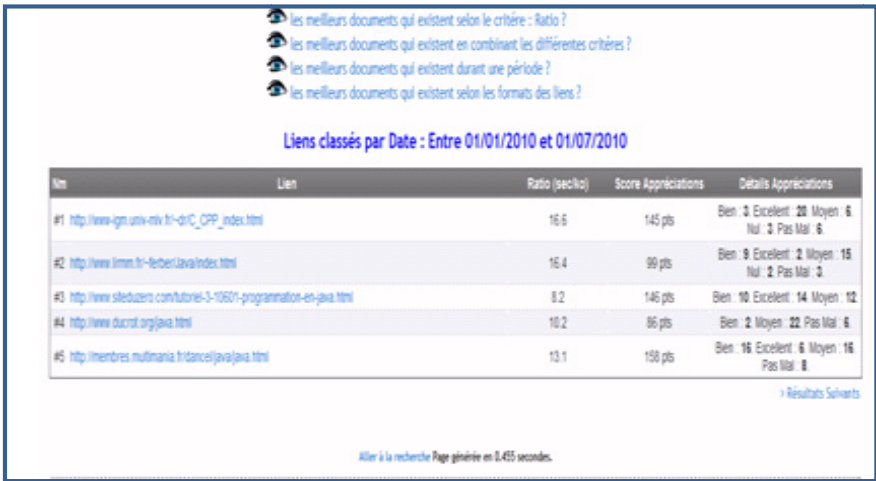


Fig. 8. Result obtained for the module java according to the combination of criteria: Score of vote, Interest (page) and period of search

4 Conclusion and Perspective

In this work, we have shown how the techniques of web usage mining and the ranking method can contribute to produce recommendations for improving the services offered by the platforms e-learning to guide and facilitate learning of learners.

The solution suggested MX-Search is based on a system of search of information web, on the concepts of web usage mining, on novel method hybrid of ranking of documents, K-means, and the filtering of information to keep track about navigation learners during their research in the web while involving them: (vote, click ...).

This will have a basis of use that will be used by the system to assist the administrator (Webmaster or teacher) to take decisions on the best documents existing on the web so to add them as additional resources in the platform, so that these learners or other promotions can benefit of them.

The hybrid method of ranking that we propose also increases the quality of recommendations in relation to use only techniques of the collaborative filtering.

This work is still being improved, as regards the topics and the features.

References

1. Brun, A., Hamad, A., Buffet, O., Boyer, A.: Vers l'utilisation de relations de préférence pour le filtrage collaboratif, Actes du dix-septième congrès francophone AFRIF-AFIA sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA 2010), Caen, France (2010)
2. Zaier, Z.: These Modèle multi-agents pour le filtrage collaboratif de l'information (Janvier 2010)
3. Pazzani, M.J., Billsus, D.: Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)
4. Burke, R., Hammond, K., Cooper, E.: Knowledge based navigation of complex information spaces. In: Proc. of the 13th National Conference on Artificial Intelligence (AAAI 1996), Menlo Park, Canada, pp. 462–468 (1996)
5. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Communications of the ACM 35(12), 61–70 (1992)
6. Moutachaouik, H., Marzak, A., Behja, H., Douzi, H., Ouhbi, B.: Système de Recommandation pour améliorer le service de recherche d'information dans les plates formes E-learning: Application sur la plate forme E-learning Moodle. In: Proc. of the the Second Edition of the International Conference on Next Generation Networks and Services (NGNS 2010), Marrakesh, Morocco, July 8-10 (2010)
7. Moutachaouik, H., Marzak, A., Behja, H., Douzi, H., Ouhbi, B.: Recommendation system to improve service to search for information in e-learning platforms: Application on E-learning platform Moodle. In: Proc. of the 2nd International Conference on Multimedia Computing and Systems (ICMCS 2011), Ouarzazate, Morocco, April 7-9 (2011)
8. Hartigan, J.A.: Clustering Algorithms. John Wiley & Sons (1975)
9. Sahami, M., Heilman, T.: A web-based kernel function for measuring the similarity of short text snippets. In: Proceedings of WWW 2006, pp. 377–386 (2006)
10. Al-Ajlan, A., Zedan, H.: Why Moodle. In: 2008 12th IEEE International Workshop on Future Trends of Distributed Computing Systems, FTDCS, pp. 58–64 (2008)
11. Dougiamas, M.: Moodle (2008), <http://www.Moodle.org>

Author Index

- Abadi, Mohamed 333
Abdelouahad, Abdelkaher Ait 451
Aboutajdine, Driss 36, 77, 175, 261, 451, 468
Adamo, Alessandro 245
Adib, Abdellah 282
Agliz, Dris 166
Aherrahrou, Noura 307
Akhouayri, Es-Said 166
Al-Hamadi, Ayoub 298, 539
Alibouch, Brahim 468
Amrouche, Abderrahmane 579
Amrous, Anissa Imen 571
Aouat, Saliha 370
Atmani, Abderrahman 166
Atri, Mohamed 85
Attia, D. 415
Ayed, Alaidine Ben 432
- Badri, Hicham 77
Bahaj, Mohamed 502
Bebis, George 235
Behja, Hicham 587
Bekkari, Aissam 17
Bekkarri, Aissam 56
Belkasmi, M. 149
Ben-Abdallah, Hanene 226, 406
Ben Jemaa, Salma 226
Bentabet, Layachi 1
Berthoumieu, Yannick 36
Boucetta, Aldjia 476
Bouziani, Mrahi 142
Boydev, Christina 397
Bruno, Odemir Martinez 343, 513
- Casanova, Dalcimar 343
Chaibi, H. 149
Cherifi, Hocine 451
Chikh, Azzeddine 397
Collet, Christophe 521
- Dadi, El Wardani 391, 485
Daoudi, El Mostafa 391
Debyeche, Mohamed 547, 571
Deger, Ferdinand 9
- Derraz, Foued 397
Deville, Yannick 157, 191
Di Caterina, Gaetano 459
Djeddi, Chawki 493
Djeradi, Amar 131
Douzi, Hassan 326, 587
Driss, Aboutajdine 200
Ducrot, Danielle 17, 56
- El Amine, Sonia 157
El Aroussi, Mohammed 261
El Daoudi, Mostafa 485
Elfazziki, Aziz 103
El Hajji, Mohamed 326
El Hassani, Ahmed Salmi EL Boumnini 290
El Hassouni, Mohammed 36, 77, 290, 451
Elhassouny, Azeddine 17, 56
EL Haziti, Mohamed 269
El Maliani, Ahmed Drissi 36
ELouedi, I. 531
El Rhabi, Mohammed 65
El yassa, Mostafa 17
Ennaji, Abdellatif 424, 493
- Faez, Karim 253
Fakhar, Khalid 261
Falek, Leila 131
Faqihi, My A. 149
Farah, Nadir 562
Florindo, João Batista 513
Forzy, Gerard 397
Fournier, R. 531
Frikha, Mayssa 226
- Gadermayr, Michael 362
Gançarski, Pierre 521
Gherabi, Noredine 502
Ghouzali, Sanaa 235, 269
Ghrab, Najla Bouarada 406
Golea, Nour El-Houda 316
Grossi, Giuliano 245

- Haddad, Boualem 183
 Hakim, Abdelilah 65
 Hamdouni, Nawal El 282
 Hammami, Mohamed 226, 406
 Hamouda, Atef 531
 Hamoudi, Hocine 183
 Handrich, Sebastian 298
 Harba, Rachid 326
 Hardeberg, Jon Yngve 9, 45
 Heuer, Michael 539
 Hosseini, Shahram 157, 191

 Ibrahim, Masrullizam Mat 459
 Idbraim, Soufiane 17, 56
 Imjabbeer, Ahmad 442

 Jarraya, Salma Kammoun 406
 Jennane, Rachid 290

 Kanan, Hamidreza Rashidy 113
 Kardouchi, Mustapha 432
 Khalil, Mohammed 282
 Khare, Ashish 93
 Khoudeir, Majdi 333
 Krzeslowski, Jakub 27

 Laasri, El Hassan Ait 166
 Lacassagne, Lionel 485
 Lachiche, Nicolas 521
 Lafkih, Maryam 269
 Lanzarotti, Raffaella 245
 Larabi, Slimane 370
 Le Moan, Steven 9
 Lespessailles, Eric 290
 Lognonné, Philippe 183

 Mączkowski, Grzegorz 27
 Mahani, Zouhir 65
 Mahjoub, Mohamed Ali 380
 Mammass, Driss 17, 56, 424
 Manap, Nurulfajar Abd 459
 Mansouri, Alamin 9
 Marchand, Sylvie 333
 Marzak, Abdelaziz 587
 Meffre, Alban 521
 Melkemi, Kamal Eddine 476
 Meskine, Fatiha 442
 Meslati, Labiba Souici 562
 Meurie, C. 415
 Mezgar, Rabeb 380

 Mikram, Mounia 235, 269
 Moalla, Imen 226
 Mohammadi, Zakaria 175
 Moujahdi, Chouaib 235
 Moutachaouik, Hicham 587
 Mtibaa, Abdellatif 380

 Nafchi, Hossein Ziaei 113
 Nait-Ali, A. 531
 Nemouchi, Soulef 562
 Nibouche, Mokhtar 122
 Nicolas, Stephane 424

 Ouhbi, Brahim 587

 Park, Dong-Chul 555
 Peyrodie, Laurent 397

 Qaffou, Issam 103

 Radgui, Amina 468
 Rashid, Omer 298
 Rathgeb, C. 217
 Rojbani, H. 531
 Ruichek, Y. 415
 Rziza, Mohammed 235, 290, 468

 Saadane, Rachid 149, 175
 Sabre, Rachid 209
 Sadgal, Mohamed 103
 Saeed, Anwar 539
 Saïd, El Abdellaoui 200
 Said, Yahia 85
 Saidani, Taoufik 85
 Saïdi, Mohamed Nabil 261
 Salahshoor, Sadegh 253
 Saoud, Sahar 65
 Saylani, Hicham 157, 191
 Selouani, Sid-Ahmed 432
 Serir, Amina 122
 Shrestha, Raju 45
 Singh, Rajiv 93
 Sitnik, Robert 27
 Skoudarli, Abdellah 122
 Smach, Fethi 85
 Snoussi, Hichem 85
 Soraghan, John J. 459
 Souici-Meslati, Labiba 493
 Srivastava, Richa 93

- Tadonki, Claude 391, 485
Tahri, Layla 352
Tairi, Hamid 307
Taleb, Nasreddine 442
Taleb-Ahmed, Abdelmalik 397
Tayebi, Mohamed 142
Teffahi, Hocine 131
Touazi, Azzedine 547
Uhl, Andreas 217, 362
Voisin, Yvon 9
Wakrim, Mohamed 352
Yessad, Dalila 579
Youssef, Fakhri 200
Zahid, Jalal 65
Zhang, Hui 1
Zirari, Fattah 424