

Roumen Kountchev
Rumen Mironov
Shengqing Li *Editors*



New Approaches for Multidimensional Signal Processing

Proceedings of International Workshop,
NAMSP 2020

Smart Innovation, Systems and Technologies

Volume 216

Series Editors

Robert J. Howlett, Bournemouth University and KES International,
Shoreham-by-sea, UK

Lakhmi C. Jain, KES International, Shoreham-by-Sea, UK

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago, DBLP.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <http://www.springer.com/series/8767>

Roumen Kountchev · Rumen Mironov ·
Shengqing Li
Editors

New Approaches for Multidimensional Signal Processing

Proceedings of International Workshop,
NAMSP 2020

 Springer

Editors

Roumen Kountchev
Technical University of Sofia
Sofia, Bulgaria

Rumen Mironov
Technical University of Sofia
Sofia, Bulgaria

Shengqing Li
Hunan University of Technology
Changsha, China

ISSN 2190-3018

ISSN 2190-3026 (electronic)

Smart Innovation, Systems and Technologies

ISBN 978-981-33-4675-8

ISBN 978-981-33-4676-5 (eBook)

<https://doi.org/10.1007/978-981-33-4676-5>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

International Program Committee

Honorary Chair

Prof. Lakhmi C. Jain, University of Technology Sydney, Australia; Liverpool Hope University, UK and KES International, UK

General Chair

Prof. Roumen Kountchev, Technical University of Sofia, Sofia, Bulgaria

General Co-chair

Prof. Srikanta Patnaik, SOA University, Bhubaneswar, India

Co-chairs

Prof. Rumen Mironov, Technical University of Sofia, Sofia, Bulgaria
Prof. Parvinder Singh, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India

International Advisory Chairs

Prof. Shengqing Li, Hunan University of Technology, China
Prof. Yikui Zhai, Wuyi University, China

Chair Members

Silai Zhou, Founder of IRnet International Academic Communication Center, China

Bin Hu, Co-founder of IRnet International Academic Communication Center, China

Publicity Chair

Dr. Roumiana Kountcheva, T&K Engineering, Bulgaria

International Program Committee

Prof. K. Rao, University of Texas at Arlington, USA

Prof. K. Nakamatsu, University of Hyogo, Japan

Prof. M. Milanova, University of Arkansas at Little Rock, USA

Prof. A. Salem, Ain Shams University, Egypt

Prof. B. Iantovics, University of Medicine, Pharmacy, Sciences and Technology of Targu Mures, Romania

Prof. K. Kpalma, INSA de Rennes, France

Prof. J. Ronsin, INSA de Rennes, France

Prof. I. Kralov, Technical University of Sofia, Bulgaria

Prof. P. Kervalishvili, Georgian Technical University, Georgia

Prof. Yo-Sung Ho, Gwangju Institute of Science and Technology, South Korea

Prof. M. Favorskaya, Siberian State Aerospace University, Russian Federation

Prof. Badrul Khan, Virginia International University, USA

Prof. P. Koprinkova-Hristova, Bulgarian Academy of Sciences, Bulgaria

Dr. R. Kountcheva, T&K Engineering, Bulgaria

Prof. V. Georgieva, Technical University of Sofia, Bulgaria

Prof. Jair Minoru Abe, University of Sao Paulo, Brazil

Prof. A. Elmaghraby, University of Louisville, USA

Prof. I. Draganov, Technical University of Sofia, Bulgaria

Prof. Omer Jasim, University of Fallujah, Iraq

Prof. A. Bekiarsky, Technical University of Sofia, Bulgaria

Prof. H. Chouiyakh, Université Internationale de Rabat, Morocco

Prof. Pl. Pavlov, Technical University of Sofia, Bulgaria

Prof. S. Pleshkova-Bekiarska, Technical University of Sofia, Bulgaria

Prof. A. Manolova, Technical University of Sofia, Bulgaria

Prof. St. Rubin, Space and Naval Warfare Systems Center, San Diego, USA

Prof. Nasreddine Taleb, Djillali Liabes University of Sidi Bel Abbes, Algeria

Prof. N. Hikal, Mansoura University, Egypt

Prof. S. Nagy, Széchenyi István University, Győr, Hungary

Prof. Bratislav Milovanović, University of Niš, Serbia
Prof. Z. Bojković, University of Belgrade, Serbia
Prof. Mihai Talmaciu, University of Bacau
Prof. Elena Nechita, University of Bacau
Prof. Anna Saro Vijenran, Bharathiar University, Coimbatore, India
Prof. Turki Obaidat, Al-Zaytoonah University of Jordan
Prof. L. Yaroslavsky, Tel Aviv University, Israel
Prof. I. Iliev, Technical University of Sofia, Bulgaria
Prof. V. Balyan, Cape Peninsula University of Technology, Capetown, South Africa
Prof. Atanas Gotchev, Tampere University of Technology, Finland
Prof. Boris Mirkin, Higher School of Economics University, Moscow, Russian Federation

Preface

This book presents the result of the International Workshop “New Approaches for Multidimensional Signal Processing—NAMSP 2020” which was carried out online, during July 9–11, 2020, at the Technical University of Sofia, Bulgaria. The prevailing part of the authors is the team members of the bilateral Bulgarian-Indian project KII-06-India-04 “Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications” between Technical University of Sofia and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India. The workshop was supported by the Bulgarian National Science Fund (BNSF) and the Ministry of Education and Science of Bulgaria. Co-organizers of NAMSP 2020 are Interscience Research Network (IRNet) International Academy Communication Center, China; Hunan University, China; and Interscience Institute of Management and Technology-Bhubaneswar, India. In the workshop, the participated authors were from Egypt, Morocco, USA, Hungary, India, China and Bulgaria.

The main objective in the presented publications is the creation and implementation of ideas, aimed at new approaches in the development of the intelligent processing and analysis of multidimensional signals in various application areas. The advance of the contemporary computer systems for processing, analysis and recognition of patterns and situations opens new abilities, beneficial to practice. As a result is got a synergic combination of various theoretical investigations and approaches.

The aim of this book is to present the latest achievements of the authors in the processing and analysis of multidimensional signals and the related applications, to a wide range of readers: IT specialists, engineers, physicians, Ph.D. students and other specialists.

The book comprises 21 chapters, related to the following areas:

- Computational Intelligence for Brain Tumors Detection;
- Video-Based Monitoring and Analytics of Human Gait for Companion Robot;
- Comparative Analysis of the Hierarchical 3D-SVD and Reduced Inverse Tensor Pyramid in Regard to Famous 3D Orthogonal Transforms;
- Tracking of Domestic Animals in Thermal Videos by Tensor Decompositions;

- Vision-Based Line Tracking Control and Stability Analysis of Unicycle Mobile Robots;
- Markerless 3D Virtual Glasses Try-on System;
- Web-based Virtual Reality for Planning and Simulation of Lifting Operations Performed by a Hydraulic Excavator;
- Partial Contour Matching based on Affine Curvature Scale Space Descriptors;
- On Metrics Used in Colonoscopy Image Processing for Detection of Colorectal Polyps;
- Copy–Move Forgery Detection by Using Key-Point-based Harris Features and CLA Clustering;
- MedSecureChain: Applying Blockchain for Delegated Access in Health Care;
- Cosine and Soft Cosine Similarity-based Anti-phishing Model;
- Enhanced Image Steganography Technique Using Cryptography for Data Hiding;
- Comparative Analysis of Various Recommendation Systems;
- Finger Knuckle Print Feature Extraction Using Artificial Intelligence Algorithm;
- The Using of Deep Neural Networks and Acoustic Waves Modulated by Triangular Waveform for Extinguishing Fires;
- Electronic Information Image Processing Technology Based on Convolutional Neural Network;
- Research on Related Problems of Intelligent Medical Logistics Distribution Based on Particle Swarm Optimization;
- Study of Wool Image Recognition Based on Texture Features;
- The Observation and Simulation of Dynamic Diffraction Patterns Caused by a Cylindrical Liquid Diffusion Pool for Diffusivity Measurement;
- Multidimensional Graphic Objects Filtration Using HoSVD Tensor Decomposition.

Sofia, Bulgaria
 Sofia, Bulgaria
 Changsha, China
 October 2020

Roumen Kountchev
 Rumen Mironov
 Shengqing Li

Acknowledgments The book editors express their special thanks to book chapter reviewers: Ivo Draganov, Roumiana Kountcheva, Mariofanna Milanova, Bin Hu, Bailing Wang, Zhongding Zhou, Ankur Singh and Omer Jasim for their efforts and goodwill to help for the successful preparation of the book. Special thanks to Prof. Lakhmi Jain (Honorary Chair) and Prof. Srikanta Patnaik (General Chair) and Prof. Parvinder Singh (Co-chair) of workshop NAMSP 2020.

The editors express their warmest thanks to the excellent Springer team which made this book possible.

Contents

1	Computational Intelligence for Brain Tumors Detection	1
	Abdel-Badeeh M. Salem	
2	Video-Based Monitoring and Analytics of Human Gait for Companion Robot	15
	Xinyi Liu, Md Imran Sarker, Mariofanna Milanova, and Lawrence O’Gorman	
3	Comparative Analysis of the Hierarchical 3D-SVD and Reduced Inverse Tensor Pyramid in Regard to Famous 3D Orthogonal Transforms	35
	Roumen Kountchev and Roumiana Kountcheva	
4	Tracking of Domestic Animals in Thermal Videos by Tensor Decompositions	57
	Ivo Draganov and Rumen Mironov	
5	Partial Contour Matching Based on Affine Curvature Scale Space Descriptors	73
	Sinda Elghoul and Faouzi Ghorbel	
6	Vision-Based Line Tracking Control and Stability Analysis of Unicycle Mobile Robots	83
	Plamen Petrov and Veska Georgieva	
7	Markerless 3D Virtual Glasses Try-On System	99
	Mariofanna Milanova and Fatima Aldaeif	
8	Copy–Move Forgery Detection by Using Key-Point-Based Harris Features and CLA Clustering	113
	Kavita Rathi and Parvinder Singh	
9	Web-Based Virtual Reality for Planning and Simulation of Lifting Operations Performed by a Hydraulic Excavator	125
	Boris Tudjarov, Rosen Mitrev, and Daniela Gotseva	

10 On Metrics Used in Colonoscopy Image Processing for Detection of Colorectal Polyps 137
Raneem Ismail and Szilvia Nagy

11 MedSecureChain: Applying Blockchain for Delegated Access in Health Care 153
Tripti Rathee and Parvinder Singh

12 Cosine and Soft Cosine Similarity-Based Anti-Phishing Model 165
Bhawna Sharma, Parvinder Singh, and Jasvinder Kaur

13 Enhanced Image Steganography Technique Using Cryptography for Data Hiding 175
Jasvinder Kaur and Shivani Sharma

14 Comparative Analysis of Various Recommendation Systems 187
Ekta Dalal and Parvinder Singh

15 Finger Knuckle Print Feature Extraction Using Artificial Intelligence Algorithm 195
Chander Kant, Sheetal Chaudhary, Sukhdev Singh, and Parvinder Singh

16 The Using of Deep Neural Networks and Acoustic Waves Modulated by Triangular Waveform for Extinguishing Fires 207
Stefan Ivanov, Stanko Stankov, Jacek Wilk-Jakubowski, and Paweł Stawczyk

17 Electronic Information Image Processing Technology Based on Convolutional Neural Network 219
Xiao Min and Guo Mei

18 Research on Related Problems of Intelligent Medical Logistics Distribution Based on Particle Swarm Optimization 225
Guo Mei and Xiao Min

19 Study of Wool Image Recognition Based on Texture Features 233
Yao Juan, Xu Wang, Zhang Cheng, and Tian Fang

20 The Observation and Simulation of Dynamic Diffraction Patterns Caused by a Cylindrical Liquid Diffusion Pool for Diffusivity Measurement 243
Licun Sun, Yuanfangzhou Wang, Linhai Li, Jie Feng, Ya Liu, and Shuwu Sheng

21 Multidimensional Graphic Objects Filtration Using HoSVD Tensor Decomposition 255
Rumen Mironov and Ivo Draganov

Author Index 267

About the Editors

Prof. Roumen Kountchev, Ph.D., D.Sc. is Professor at the Faculty of Telecommunications, Department of Radio Communications and Video Technologies, Technical University of Sofia, Bulgaria. His areas of interest are digital signal and image processing, image compression, multimedia watermarking, video communications, pattern recognition and neural networks. Prof. Kountchev has 350 papers published in magazines and proceedings of conferences; 15 books; 46 book chapters; 20 patents (3 intern.). He had been Principle Investigator of 38 research projects. At present, he is a member of Euro Mediterranean Academy of Arts and Sciences (EMAAS) and President of Bulgarian Association for Pattern Recognition (member of Intern. Association for Pattern Recognition). He is an Editorial board member of International Journal of Reasoning-based Intelligent Systems; International Journal Broad Research in Artificial Intelligence and Neuroscience; KES Focus Group on Intelligent Decision Technologies; Egyptian Computer Science Journal; International Journal of Bio-Medical Informatics and e-Health; and International Journal of Intelligent Decision Technologies. He has been a plenary speaker at WSEAS International Conference on Signal Processing, 2009, Istanbul, Turkey; WSEAS International Conference on Signal Processing, Robotics and Automation, University of Cambridge 2010, UK; WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision 2012, Istanbul, Turkey; International Workshop on Bioinformatics, Medical Informatics and e-Health 2013, Ain Shams University, Cairo, Egypt; Workshop SCCIBOV 2015, Djillali Liabes University, Sidi Bel Abbes, Algeria; International Conference on Information Technology 2015 and 2017, Al-Zaytoonah University, Amman, Jordan; WSEAS European Conference of Computer Science 2016, Rome, Italy; The 9th International Conference on Circuits, Systems and Signals (CSS'17), London, UK, 2017; IEEE International Conference on High Technology for Sustainable Development (HiTech'18) and (HiTech'19), Sofia, Bulgaria; The 8th International Congress of Information and Communication Technology (ICICT'18), Xiamen, China, 2018.

Prof. Rumen Mironov is working at Technical University of Sofia, Sofia, Bulgaria. Dr. Rumen Mironov received his M.Sc. and Ph.D. in Telecommunications from Technical University of Sofia and M.Sc. in Applied Mathematics and Informatics

from Faculty of Applied Mathematics and Informatics. He is currently Head of the Department of Radio Communications and Video Technologies, Technical University of Sofia, Bulgaria. His current research focuses on digital signal and image processing, pattern recognition, audio and video communications, information systems, computer graphics and programming languages. He is a member of Bulgarian Association of Pattern Recognition (IAPR) and Bulgarian Union of Automation and Automation Systems. Rumen Mironov is the author of more than 60 scientific publications.

Prof. Shengqing Li is working at Hunan University of Technology, China. He is Doctor of Engineering, Professor, Doctoral Supervisor and Expert entitled to Government Special Allowance (GSA). Professor Li is Dean in the School of Electrical and Information Engineering at Hunan University of Technology. Meanwhile, he is serving as Chairman of Hunan Engineering Research Center, the academic leader of provincial key discipline as well as the permanent member of Hunan Electrotechnical Society and Hunan Electrical Engineering Society. Professor Li's extensive experiences with electricity and engineering lead to his broader interest in Power Quality Control Strategy for Distribution Network, New Energy Grid-Connected System Power and Electrical Energy-saving Technology. He currently hosted and participated in 20 important research projects of the Provincial Natural Science Foundation, the National Natural Science Foundation and National Key R&D Program of China, etc. He received 9 ministerial second prizes and honoured prizes of Progress in Science and Technology, Science and Technology Award of China Electrotechnical Society, etc., with his papers, researches, 14 authorized patents and software copyright. Professor Li is the author of 6 books (2 of them were published by China Science Press). He has published numerous papers, 45 of which are included by SCI/EI consisting of proceedings of the CSEE and International Journal of Robotics and Automation, etc.

Chapter 1

Computational Intelligence for Brain Tumors Detection



Abdel-Badeeh M. Salem

Abstract Recently, computational intelligence (CI) techniques have become efficient intelligent tools for brain tumor detection. It has become one of the major research subjects in medical imaging and diagnostic radiology. In the area of processing the brain images, computer-aided diagnosis (CAD) systems are basically relied on different CI techniques in all its stages to implement a smart consultation system that can help the radiologists by providing a second opinion that can assist in detection and diagnosis of brain tumors. This paper presents a comprehensive and up-to-date research in the area of digital medical imaging covering a wide spectrum of CI methodological and intelligent algorithm. The paper discusses the current research of the CI techniques for developing smart CAD systems. We present two applications for a hybrid intelligent technique for automatic detection of brain tumor through MRI. The technique is based on the following CI methods: the feedback pulse-coupled neural network for image segmentation, the discrete wavelet transform for features extraction, the principal component analysis for reducing the dimensionality of the wavelet coefficients, and the feed-forward back-propagation neural network to classify inputs into normal or abnormal.

1.1 Introduction

Magnetic resonance imaging (MRI) is an imaging technique that plays a vital role in detection and diagnosis of brain tumors in both research and clinical care for providing detailed information about the brain structure and its soft tissues. The image-processing techniques can provide great help in analyzing the tumor area. Computer-aided detection (CAD) has been developing fast in the last two decades.

A.-B. M. Salem (✉)

Artificial Intelligence and Knowledge Engineering Research Labs, Faculty of Computer and Information Sciences, Ain Shams University, Abbassia, Cairo, Egypt

e-mail: abSalem@cis.asu.edu.eg

URL: <http://staff.asu.edu.eg/Badeeh-Salem>; <http://aiasulab.000webhostapp.com/>

The main idea of CAD is to assist radiologists in interpreting medical images by using dedicated computer systems to provide ‘second opinions.’ The current research reveals that the CAD systems of human brain MRI images are still an open problem [1, 2].

On the other side, AI and CI offer robust, intelligent algorithms and smart methods that can help to solve problems in a variety of healthcare and life sciences areas [3–6]. This paper presents some of the CI techniques for managing and engineering knowledge in digital computer-aided diagnosis (CAD) systems. Some applications of the author and his colleagues that have been carried out in last years are discussed. The paper is organized as follows: Sect. 1.2 presents the computational intelligence paradigms. Section 1.3 presents the subareas of the intelligent of healthcare informatics. Section 1.4 discusses the applications of deep learning technique in health informatics. In Sect. 1.5, we discuss the medical imaging techniques for human brain. Section 1.6 presents our applications of brain tumors diagnosis using CI techniques, and then, we conclude in Sect. 1.7.

1.2 Computational Intelligence Paradigms

Computational intelligence (CI) is the study of intelligent computer algorithms that improve automatically through experience. CI aims to enable computers to learn from data and make improvements without any dependence on commands in a program. This learning could eventually help computers in building smart models for specific task for prediction purpose [6–8]. Figure 1.1 shows the inherently interdisciplinary field of research of CI. From the figure, it can be seen that CI includes the following disciplines: neurobiology, information theory, probability, statistics, AI, control theory, Bayesian methods, physiology, and philosophy.

Figure 1.2 shows the computational intelligence paradigms and models from intelligent computing and AI perspectives [9–19].

- (a) Cognitive computing: This model is based on cognitive sciences (human memory, languages, vision, stress, learning, and sleep phenomena)

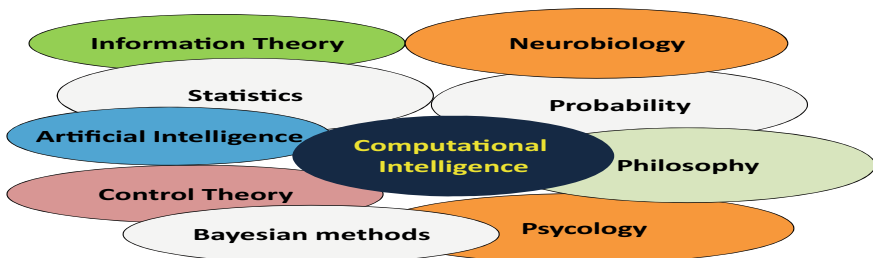


Fig. 1.1 Interdisciplinary field of research of CI

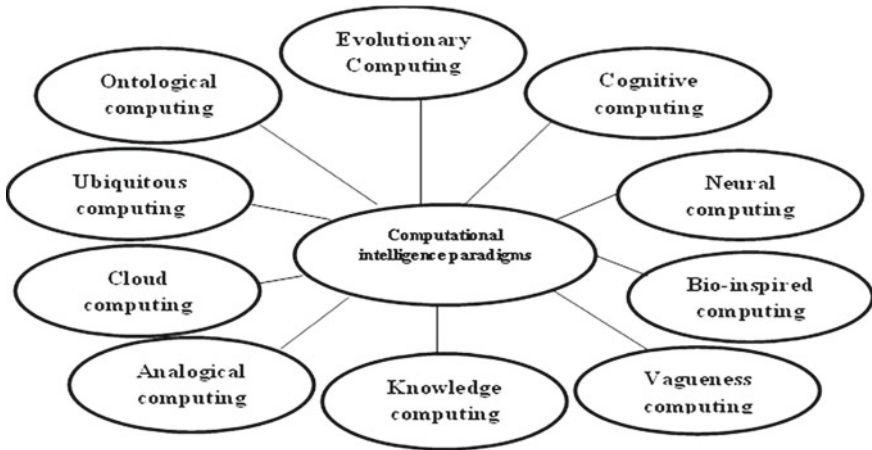


Fig. 1.2 Computational intelligence paradigms

- (b) Neural computing: This is based on neuroscience, biological sciences, and mathematics
- (c) Bio-inspired computing: This is based on life sciences (biology and medicine)
- (d) Vagueness computing: This is based on fuzzy logic theory and rough sets
- (e) Knowledge computing: This is based on knowledge engineering issues (management, discovery, acquisition, and modeling)
- (f) Analogical computing: This is based on case-based reasoning methodology
- (g) Cloud computing: This is based on distributed processing and Internet of things (IoT)
- (h) Ubiquitous computing: In this, model computing is made to appear anytime and everywhere
- (i) Ontological computing: This is based on the ontological engineering concepts
- (j) Evolutionary computing: This is based on simulated annealing, swarm optimization, and ant colony.

1.3 Computational Intelligence Applications in Intelligent Health Informatics

Figure 1.3 shows the main and subareas of the intelligent health informatics. From the figure, it can be seen that it is a multidisciplinary field of research and covered many digital healthcare areas, namely dental, neuro, biological, medical, nursing, and clinical. From the informatics perspective, each are composed of many fields of research, e.g. the medical informatics contains, knowledge engineering, medical imaging, expert systems, robotic surgery, education, learning, and training [20–22].

In the last years, various computational intelligence techniques and methodologies have been proposed by the researchers in order to develop digital healthcare systems

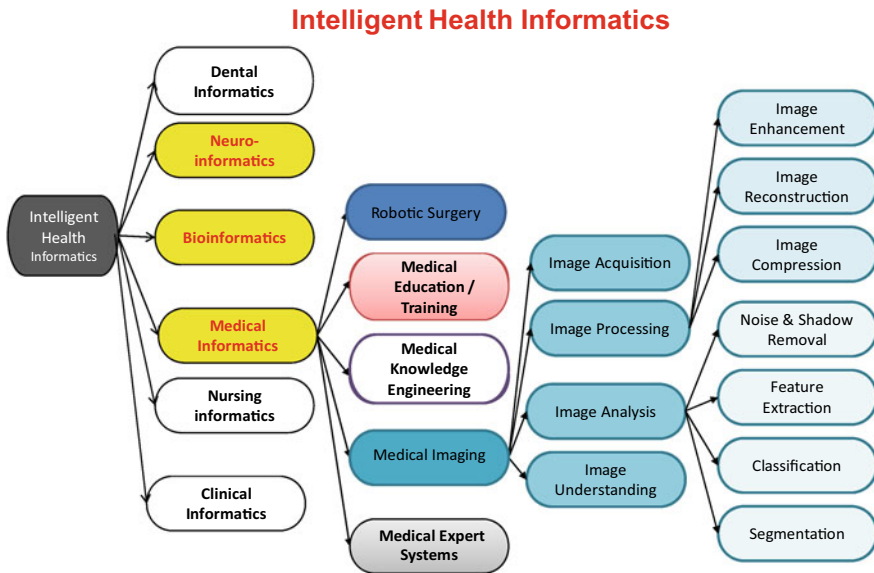


Fig. 1.3 Intelligent healthcare informatics

for different medical and healthcare tasks. These systems are based on the knowledge engineering paradigms and artificial intelligence (AI) concepts and theories. Many types of such systems are in existence today and are applied to different healthcare domains and tasks.

CI is successfully used on a wide variety of medical problems and data. Medicine is largely an evidence-driven discipline where large quantities of relatively high-quality data are collected and stored in databases. The medical data are highly heterogeneous and are stored in numerical, text, image, sound, and video formats. In addition, CI techniques are also used to modify medical procedures in order to reduce cost and improve perceived patient's experience and outcomes.

From the medical informatics point of view, medical data includes [9, 10, 12]:

(a) Clinical data (symptoms, demographics, biochemical tests, diagnoses and various imaging, video, vital signals, etc.), (b) logistics data (e.g., charges and cost policies, guidelines, clinical trials), (c) bibliographical data, and (d) molecular data.

On the other side, bioinformatics concerns with biological data, conceptualizes biology in terms of molecules, and applies CI techniques to understand and organize the information associated with these molecules on a large scale [14, 15]. Bioinformatics encompasses analysis of molecular data expressed in the form of nucleotides, amino acids, DNA, RNA, peptides, and proteins. The huge amount and breadth of biological data requires development of efficient methods for knowledge/information extraction that can cope with the size and complexity of the accumulated data.

There are numerous examples of successful applications of CI in areas of diagnosis and prevention, prognosis, and therapeutic decision making. CI algorithms are

used for the following tasks: (a) discovering new diseases, (b) finding predictive and therapeutic biomarkers, and (c) detecting relationships and structure among the clinical. CI contributes to the enhancement of management and information retrieval processes leading to development of intelligent (involving ontologies and natural language processing) and integrated (across repositories) literature searches. Moreover, applications of CI in bioinformatics include the following areas of research: (a) microarray analysis, chromosome and proteome databases, modeling of inhibition of metabolic networks, (b) signal analysis (echocardiograph images and electroencephalograph time series), and (c) drug delivery, information retrieval, software for pattern recognition in biomedical data.

1.4 Deep Learning Paradigm in Health Informatics

Deep learning is a subarea of CI covering a spectrum of current exciting research and industrial innovation that provides more efficient algorithms to deal with large-scale data in healthcare, recommender systems, learning theory, robotics, games, neurosciences, computer vision, speech recognition, language processing, human–computer interaction, drug discovery, biomedical informatics, act [23–25]. DL is a branch of AI covering a spectrum of current exciting research and industrial innovation that provides more efficient algorithms to deal with large-scale data in many areas (see Fig. 1.4).

In the last decade, with the development of artificial neural networks, many researchers have tried to develop further studies using DL methods. DL studies are investigated in popular areas to illuminate the paths of researchers working in DL Tables 1.1, 1.2, 1.3, and 1.4 and to summarize the different DL methods and applications of health informatics.

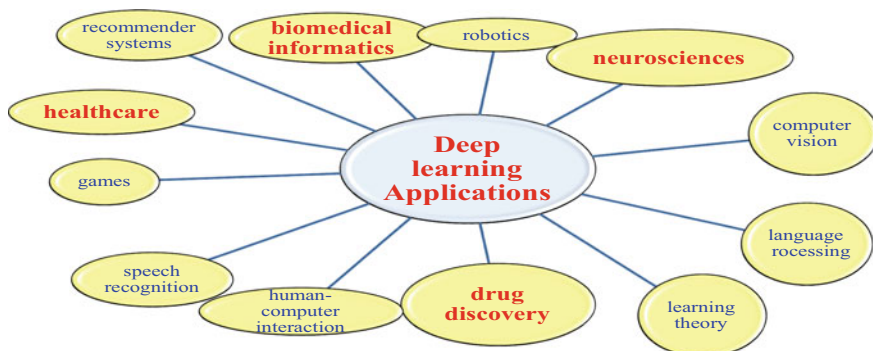


Fig. 1.4 Deep learning (DL) applications

Table 1.1 Deep learning methods and applications for bioinformatics

Application	Input data	DL method
Cancer diagnosis	Gene expression	Deep autoencoders
Gene selection/classification	Micro RNA	Deep belief network
Gene variants	Microarray data	Deep neural network
Drug design	Molecule compounds	Deep neural network
Compound–protein interaction	Protein structure	Deep belief network
RNA-binding protein	Molecule compounds	Deep neural network
DNA methylation	Genes/RNA/DNA sequences	

Table 1.2 Deep learning methods and applications for medical informatics

Application	Input data	DL method
Prediction of disease	Electronic health records	Deep autoencoders
Human behavior monitoring	Big medical dataset	Deep belief network
Data mining	Blood/laboratory tests	Convolutional neural network
		Recurrent neural network
		Convolutional deep belief network
		Deep neural network

Table 1.3 Deep learning methods and applications for public health

Application	Input data	DL method
Predicting demographic info	Social media data	Deep autoencoders
Lifestyle diseases	Mobile phone metadata	Deep belief network
Infectious disease epidemics	Geo-tagged images	Convolutional neural network
Air pollutant prediction	Text messages	Deep neural network

1.5 Medical Imaging Techniques for Human Brain

1.5.1 Medical Aspects

A brain tumor is a mass or growth of abnormal cells in the brain. Many types of brain tumors exist (see Fig. 1.5). Some brain tumors are noncancerous (benign), and some brain tumors are cancerous (malignant). Brain tumors can begin in the brain (primary brain tumors), or cancer can begin in other parts of the body and spread to the brain (secondary, or metastatic, brain tumors). The growth rate as well as location of a brain tumor determines how it will affect the function of nervous system. Brain tumor treatment options depend on the type of brain tumor as well as its size and

Table 1.4 Deep learning methods and applications for pervasive sensing

Application	Input data	DL method
Anomaly detection Biological parameters monitoring	EEG ECG Implantable device	Deep belief network
Human activity recognition	Video Wearable device	Convolutional neural network Deep belief network Deep neural network
Hand gesture recognition Obstacle detection Sign language recognition	Depth camera RGB-D camera Real-sense camera	Convolutional neural network Deep belief network
Food intake Energy expenditure	Wearable device RGB image Mobile device	Convolutional neural network Deep neural network


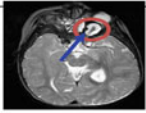
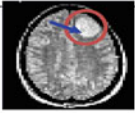
Tumor	Acoustic	Optic glioma	Astrocytoma
Location	Hearing Nerve	Optic nerve	Temporal Lobe
Symptom	1-hearing loss 2-ringing & headache 3-weakness of face 4-balance problems	1-visual loss 2-double vision 3-rapid eye movement	1- seizure. 2-paralysis 3-problems with language
MRI			

Fig. 1.5 Some types of brain tumors

location. Tables 1.5 and 1.6 show applications of deep learning methods of medical imaging as well as the well-known brain imaging techniques (BIT), respectively.

1.5.2 Magnetic Resonance Imaging (MRI)

Brain MRIs have three types: T1, T2, and PD which differ in the contrast characteristics of the brain tissues. These three types have three orientations which are axial, corona, and sagittal. All these types are interpreted by radiologists, physicians,

Table 1.5 Deep learning methods and applications for medical imaging

Application	Input data	DL method
3D brain reconstruction	MRI/fMRI	Deep autoencoders
Neural cells classification	Fundus images	Convolutional neural network
Brain tissues classification	PET scans	Deep belief network
Alzheimer/MCI diagnosis		Deep neural network
Tissue classification	MRI/CT images	Convolutional deep belief network
Organ segmentation	Endoscopy images	Convolutional neural network
Cell clustering	Microscopy	Deep autoencoders
Hemorrhage detection	Fundus images	Group method of data handling
Tumor detection	X-ray images	Deep neural network
	Hyperspectral images	

Table 1.6 Well-known brain imaging techniques

No.	Brain imaging technique
1	Computed tomography scan
2	MRI brain sagittal and coronal MRI scans
3	Positron emission tomography (PET) scan
4	Single positron emission computed tomography (SPECT) scan
5	Functional magnetic resonance imaging (fMRI) scan
6	Electroencephalography scan
7	Magneto encephalography scan

and researchers for diagnosing brain tumors and putting a treatment. Also, for the research purpose, a number of datasets are available online for research use [26].

1.5.3 General Methodology of Brain Images Processing

Based on our comprehensive study of the published literatures [26], one can conclude that the general methodology of brain images processing consists of seven processes, namely (a) image acquisition and preprocessing, (b) segmentation of ROI, (c) feature extraction and selection, (d) dimensionality reduction, (e) classification of the selected ROI, (f) performance evaluation, and (g) interpretation by the expert radiologists. Figure 1.6 shows the general methodology of brain image processing.

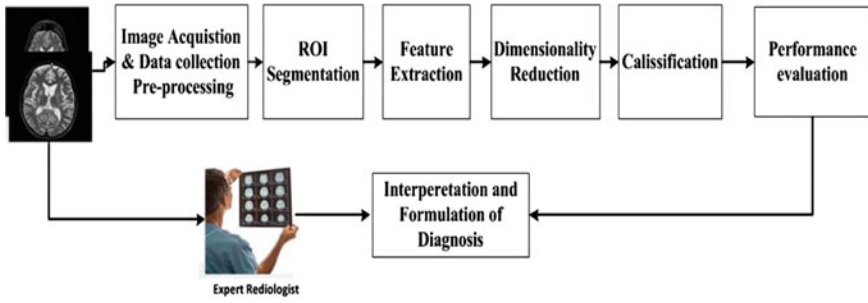


Fig. 1.6 General methodology of brain image processing

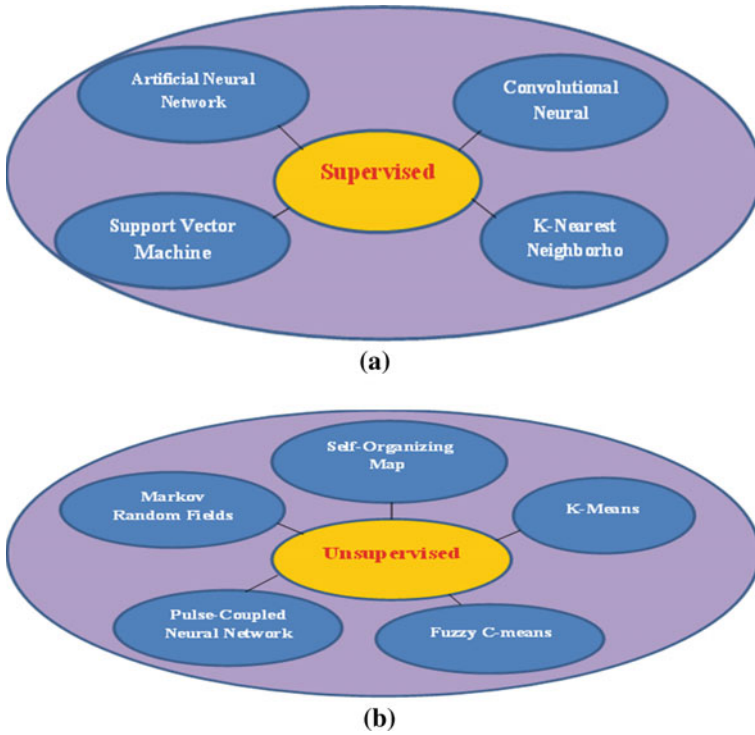


Fig. 1.7 a Different supervised CI techniques for the image segmentation. b Different unsupervised CI techniques for the image segmentation

1.5.4 CI Techniques for Segmentation and Classification Processes

Figure 1.7 shows the different supervised and unsupervised CI techniques for the image segmentation. For more technical information about these techniques, see Refs. [27–30]. For classification process, see Refs. [31–36].

1.6 Brain Tumor Diagnosis Using CI Techniques

1.6.1 1st Application: Developing CAD Systems for Brain Tumor Diagnosis

This study concerned with developing a CAD system that can process the brain MR images for detection and diagnosis of different brain tumors using several computational intelligence techniques. Commonly, the implementation of any CAD system for classification of brain tumors based on brain MRIs involved the following three stages: (1) segmentation, (2) feature extraction and selection, and (3) training/testing of the classification model

In this study, two types of CAD systems are implemented.

- (A) For the first type of CAD system, three CAD systems with several models are presented. The three CAD systems included three stages: segmentation, feature extraction and selection, and classification. For segmentation process, K-means and fuzzy C-means techniques that have been used separately. While for feature extraction and selection processes, gray-level co-occurrence matrix (GLCM) and discrete wavelet transform (DWT) integrated with principal component analysis (PCA) have been used separately.
- (B) The second type is to differentially diagnose cognitive normal (CN) brain from Alzheimer's disease (AD) brain subjects using brain MRIs from two real online datasets of brain MRIs. This study is based on linear discriminate analysis (LDA) classifier.

Extracting the features from the input brain MRIs of the two datasets used is done for each dataset separately using DWT integrated with PCA for reducing the number of features to avoid classification complications and reduce the computation time and costs.

The developed system is tested using two different datasets obtained from online datasets of real human brain MRIs. The performance of the system proved its efficiency and reliability in the problem which it is used for according to different performance measures.

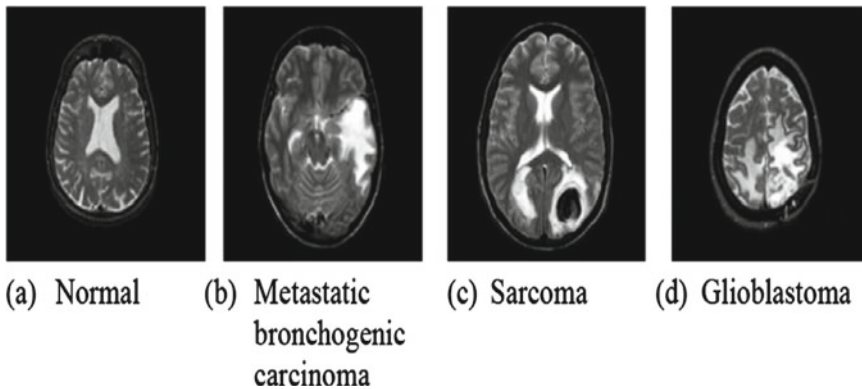


Fig. 1.8 Sample of brain MRIs' dataset

1.6.2 2nd Application: Classification Using Deep Learning for Brain Tumors

In this application, we used DL neural network for classifying a dataset of 66 brain MRIs into four classes, e.g., normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors (see Fig. 1.8). The classifier was combined with the discrete wavelet transform (DWT) and principal components analysis (PCA). The evaluation of the performance was quite good over all the performance measures.

We can summarize the main features of this study as follows:

1. The methodology combines the discrete wavelet transform (DWT) with the deep neural network (DNN) to classify the brain MRIs into normal and three types of malignant brain tumors: glioblastoma, sarcoma, and metastatic bronchogenic carcinoma.
2. The architecture of this developed system resembles the convolution neural network (CNN) architecture but requires less hardware specifications and takes a convenient time of processing for large-size images (256×2563 —The developed DNN classifier shows high accuracy compared to traditional classifiers).

1.7 Recommendations and Conclusions

1.7.1 Recommendation

From our comprehensive analysis, one can recommend the following recommendations:

- (a) The cooperation between physicians and AI communities is essential to produce efficient computing systems for medical purposes. The physicians will have

more information to deliver a better service and dynamic guidelines to improve quality and reduce risks.

- (b) The industry of intelligent medical decision support systems is a promising area of research for developing successful telemedicine projects.
- (c) Mobile devices can feed real-time medical data directly to patients and doctors via secure computing networks and IoT. The web-based and IoT medical systems can enhance the online education/learning/training processes.
- (d) The use of ICT technologies also improves the quality of patient care and reduces clinical risk. At the same time, the patient will be part of the healthcare process, having more information about diseases and access to his/her electronic health record.
- (e) The pharmaceutical industries can get more accurate information about drug's effects and supply chain delivery systems.
- (f) Public health authorities can get more accurate information and develop dashboards to make better and fast decisions.
- (g) Hospital management benefits from a more updated meaningful data. This data is used by management systems to deliver key performance indicators (KPI).

1.8 Conclusions

AI technologies and techniques play a key role in developing intelligent tools for medical tasks and domains. This paper analyzes the main paradigms and applications of the computational intelligence (CI) in health care from the artificial intelligence perspective. CI offers potentially powerful tools for the development a novel digital healthcare system. The variety of such technique enabling the design of robust and efficient intelligent healthcare systems. CI techniques (e.g., CBR, data mining, rough set, and ontology) can cope with medical noisy data, subsymbolic data, and complex structure data. In addition, CI offers intelligent computational methods for accumulating, changing, and updating medical knowledge in IHS, and in particular learning mechanisms that will help us to induce knowledge from medical information or data. The key to the success of such systems is the selection of the CI technique that best fits the domain knowledge and the problem to be solved. That choice depends on the experience of the knowledge engineers. On the other side, the development of such systems is a very difficult and complex process that raises a lot of technological and research challenges that have to be addressed in an interdisciplinary way.

References

1. Arimura, H., Magome, T., Yamashita, Y., Yamamoto, D.: Computer-aided diagnosis systems for brain diseases in magnetic resonance images. *Algorithms* 2(3), 925–952 (2009)

2. Arimura, H., Tokunaga, C., Yamashita, Y., Kuwazuru, J.: Magnetic resonance image analysis for brain CAD systems with machine learning. In: Suzuki, K. (ed.) *Machine Learning. Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis*, pp. 258–296. The University of Chicago, IGI Global, USA (2012)
3. Nordlinger, B., Villani, C., Rus, D.: *Healthcare and Artificial Intelligence*. Springer, Switzerland (2020). ISBN 978-3-030-32160-4. ISBN 978-3-030-32161-1 (eBook). <https://doi.org/10.1007/978-3-030-32161-1>
4. Marquis, P., Papini, O., Prade, H.: *A Guided Tour of Artificial Intelligence Research, Volume III: Interfaces and Applications of Artificial Intelligence*. Springer, Switzerland (2020). ISBN 978-3-030-06169-2. ISBN 978-3-030-06170-8 (eBook). <https://doi.org/10.1007/978-3-030-06170-8>
5. Russell, S.J., Norvig, P.: *Artificial Intelligence: A Modern Approach*, 2nd edn. Prentice Hall (2003)
6. Rayan, Z., Alfonse M., Salem, A-B.M.: Machine learning approaches in smart health. *Procedia Comput. Sci.* **154**, 361–368 (2019). www.sciencedirect.com. www.elsevier.com/locate/procedia
7. Said, H.M., Salem, A.B.M.: Exploiting computational intelligence paradigms in e-technologies and activities. *Procedia Comput. Sci.* **65**, iii–vii, 396–405 (2015)
8. Cook, D.J., Das, S. (eds.) *Smart Environments: Technologies, Protocols and Applications*. Wiley, New York, NY (2004) (eBook)
9. Dix, A., Finlay, J., Abowd, G.D., Beale, R.: *Human–Computer Interaction*, 3rd edn. Prentice Hall (2003)
10. Simon, S.: *Neural Networks and Learning Machines*, 3rd edn. (2009) (originally published: 1993). <http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed>
11. Pan, J-S., Lin, J.C.-W., Liang, Y., Chu, S.-C.: Genetic and evolutionary computing. In: *Proceedings of the Thirteenth International Conference on Genetic and Evolutionary Computing*, 1–3 Nov 2019, Qingdao, China. Springer, Singapore (2020). ISSN 2194-5357/ISSN 2194-5365 (electronic). *Advances in Intelligent Systems and Computing*. ISBN 978-981-15-3307-5/ISBN 978-981-15-3308-2 (eBook). <https://doi.org/10.1007/978-981-15-3308-2>
12. Salem, A.B.M., Aref, M.M., Moawad, I.F., Alfonse, M., Marzoqi, G.: Ontological engineering in medical informatics. *Int. J. Genomics Proteomics Metabol. Bioinform. (IJGPMB)* **1**(2), 9–13 (2016)
13. Cios, K.J., Kurgan, L.A., Reformat, M.: How it is used on a wide variety of medical problems and data, machine learning in the life sciences. *IEEE Eng. Med. Biol. Mag.* 14–16 (2007)
14. Mitchell, T.: *Machine Learning*. McGraw-Hill, New York (1997)
15. Paun, G.H., Rozenberg, G., Salomaa, A.: *DNA Computing: New Computing Paradigms* (trans: Yokomori, T.). Springer, Berlin, Heidelberg (1999)
16. Huang, Y., He, L.: DNA, computing research progress and application. In: *The 6th International Conference on Computer Science & Education*, pp. 232–235 (2011)
17. Kolonder, J.: *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, CA (1993)
18. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. *A. I. Commun.* **7**(1), 39–52 (1994)
19. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht (1991)
20. Voskoglou, M.: Special issue fuzzy sets, fuzzy logic and their applications. In: *Mathematics* (2020). ISBN 978-3-03928-520-4 (Pbk); ISBN 978-3-03928-521-1 (PDF). <https://doi.org/10.3390/books978-3-03928-521-1>
21. Salem, A.B.M.: Artificial intelligence technology in intelligent health informatics. In: Borzemski, L., Świątek, J., Wilimowska, Z. (eds.) *Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology—ISAT 2018*. ISAT 2018. *Advances in Intelligent Systems and Computing*, vol. 85. Springer (2019). https://doi.org/10.1007/978-3-319-99981-4_1
22. Nariman, N.-S.: Artificial intelligence transforms the future of healthcare. *Am. J. Med.* **132**(7), 795–801 (2019)

23. Salem, A.B.M.: Knowledge engineering paradigms in the intelligent health informatics and medical knowledge-based systems. *J. Global Res. Comput. Sci.* ISSN: 2229-371X
24. Miotto, R., et al.: Deep learning for healthcare: review, opportunities and challenges. *Brief. Bioinform.* **19**(6), 1236–1246 (2017)
25. Kwak, G.H.-J., Pan, H.: DeepHealth: Deep Learning for Health Informatics reviews, challenges, and opportunities on medical imaging, electronic health records, genomics, sensing, and online communication health. arXiv preprint [arXiv:1909.00384](https://arxiv.org/abs/1909.00384) (2019)
26. Mohsen, H., El-Dahshan, E.A., El-Horbaty, E.M., Salem, A.B.M.: Intelligent methodology for brain tumors classification in magnetic resonance images. *Int. J. Comput.* **11**, 1–5 (2017)
27. Ayachi, R., Ben Amor, N.: Brain tumor segmentation using support vector machines; symbolic and quantitative approaches to reasoning with uncertainty. *Lect. Notes Comput. Sci.* **5590**, 736–747 (2009)
28. Balafar, M.A., Ramli, A.R., Sariipan, M.I., Mahmud, R., Mashohor, S.: Medical image segmentation using fuzzy c-mean (FCM), learning vector quantization (LVQ) and user interaction. In: Huang, D.-S., et al. (eds.) *Proceedings of International Conference on Intelligent Computing, ICIC 2008, CCIS15*, pp. 177–184. Springer, Berlin (2008)
29. Chao, W.H., Chen, Y.Y., Lin, S.H., Shih, Y.Y.I., Tsang, S.: Automatic segmentation of magnetic resonance images using a decision tree with spatial information. *Comput. Med. Imaging Graph.* **33**(2), 111–121 (2009)
30. Mohsen, H., El-Dahshan, E.A., El-Horbaty, E., Salem, A.M.: A comparative study of segmentation techniques for brain magnetic resonance images. *ATINER'S Conference Paper Series, Athens*, pp. 2–12. No: COM2016-1994. ISSN: 2241-2891 (2016)
31. Chaplot, S., Patnaik, L.M., Jagannathan, N.R.: Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Signal Process. Control* **1**(1), 86–92 (2006)
32. Alfonse, M., Salem, A.-B.M.: An automatic classification of brain tumors through MRI using support vector machine. *Egypt. Comput. Sci. J.* **40**(3), 11–21 (2016). ISSN-1110-2586
33. Mohsen, H., El-Dahshan, E.A., El-Horbaty, E.M., Salem A.B.M.: Classification using deep learning neural networks for brain tumors. *Future Comput. Inform. J.* 68–71 (2017). <https://doi.org/10.1016/j.fcij.2017.12.001>
34. Salem, A.B.M., Mahmoud, A.M.A.: Hybrid genetic algorithm-decision tree classifier. In: *Proceedings of the 3rd International Conference on New Trends in Intelligent Information Processing and Web Mining*, Zakopane, Poland, 2–5 June, pp. 221–232 (2003)
35. Salem, A.-B.M., Mahmoud, A.M.: Applying the genetic algorithms approach for data mining classification task. In: *IFIP WG12.6, First IFIP Conference on Artificial Intelligence Applications and Innovations*, Toulouse, France (2004)
36. Abdrabou, E.A.M., Salem, A.B.: A breast cancer classifier based on a combination of case-based reasoning and ontology approach. In: *Proceedings of 2nd International Multi-conference on Computer Science and Information Technology (IMCSIT 2010)*, Wisła, Poland (2010)

Chapter 2

Video-Based Monitoring and Analytics of Human Gait for Companion Robot



Xinyi Liu, Md Imran Sarker, Mariofanna Milanova,
and Lawrence O’Gorman

Abstract Human gait is essential for long-term health monitoring as it reflects physical and neurological aspects of a person’s health status. In this paper, we propose a non-invasive video-based gait analysis system to detect abnormal gait, and record gait and postural parameters framework on a day-to-day basis. It takes videos captured from a single camera mounted on a robot as input. Open Pose, a deep learning-based 2D pose estimator is used to localize skeleton and joints in each frame. Angles of body parts form multivariate time series. Then, we employ time series analysis for normal and abnormal gait classification. Dynamic time warping (DTW)-based support vector machine (SVM)-based classification module is proposed and developed. We classify normal and abnormal gait by characterizing subjects’ gait pattern and measuring deviation from their normal gait. In the experiment, we capture videos of our volunteers showing normal gait as well as simulated abnormal gait to validate the proposed methods. From the gait and postural parameters, we observe a distinction between normal and abnormal gait groups. It shows that by recording and tracking these parameters, we can quantitatively analyze body posture. People can see on the display results of the evaluation after walking through a camera mounted on a companion robot.

X. Liu · M. I. Sarker · M. Milanova (✉)
University of Arkansas at Little Rock, Little Rock, AR, USA
e-mail: mgmilanova@ualr.edu

X. Liu
e-mail: xxliu8@ualr.edu

M. I. Sarker
e-mail: misarker@ualr.edu

L. O’Gorman
Nokia Bell Labs, Murray Hill, NJ, USA
e-mail: larry.o_gorman@nokia-bell-labs.com

2.1 Introduction

With the rise of aging populations, desire for independent living, high healthcare costs, non-invasive health monitoring, and smart personal and home communications, the time for “healthy-living” companion robots are approaching. Every 11 s, an older adult is treated in the emergency room for a fall; every 19 min, an older adult die from a fall. Healthcare cost of falls in the USA, including hospitalization, surgery, therapy, etc., is expected to increase as the population ages and may reach \$67.7 billion by 2020 [1, 2].

According to Selke [3], *lifelogging* is understood as different types of digital self-tracking and recording of everyday life. Another feature of *lifelogging* is that it is a continuous process that requires no user interaction. In the context of active-assisted living (AAL), sensors used for lifelogging can also be *ambient-installed* as opposed to wearable sensors, for instance, video surveillance or other cameras installed in nursing or smart homes to monitor and support older and fragile people [4–6]. Recently, there is a growing research on robots and their amplification in AAL Healthcare humanoid robots are designed and used by individuals at home or healthcare centers to analyze, treat and improve their medical conditions.

The quantitative gait analysis requires specific devices such as a 3D motion capture system, accelerometer, or force plate, which are time-, labor-, space-, and cost-consuming to use daily. Furthermore, monitoring gait in home using monocular camera without any annotations is interesting and not explored (practically unavailable for gait analyses).

The goal of this study is to develop a video-based marker-less system for non-invasive healthcare monitoring, using skeleton and joint location from pose estimation to extract gait features and generate alerts for abnormal gait, indicating needs for further medical attention. We propose use of Open Pose [7], a deep learning-based 2D keypoint framework that estimates the joint coordinates of persons in the image or videos obtained using a monocular camera, as it does not require external scales or markers. Using this estimator, we can automatically obtain the joint coordinates of persons in each image/movie recorded, thus enabling the calculation of joint angles or other spatial parameters useful for further gait analysis.

We first designed experiments to measure consistency of the system on a healthy population, then monitoring gait of subjects walking carrying the weight to replicate gait decay. The second set of experiments is “nudge” human posture monitoring. The system guides individuals in regular walking, freely assesses gait states and provides real-time personalized feedback to evaluate correct body posture. The robot will also connect individuals with family and friends through a virtual connection, and if needed, it will set up alarms.

Hypothesis Can we evaluate gait and motion dynamic integrating deep learning-based 2D skeleton estimator and time series analysis?

Research involves two complementary tasks:

1. Designing set of experiments and creating own dataset.

2. Analyzing gait changes using lower body skeleton representation and calculate falling risk.

Our contribution

1. We designed system to evaluate gait deviation and classification of normal and abnormal gait non-invasively over time.
2. Perform experiments on subject walking normally and the same subject walking carrying the weight.
3. Feature extraction and machine learning evaluation of human pose changes.

In Sect. 2.2, we describe evidence primarily from the medical and health-related literature on the correlation between gait and health. We also give an overview of other video gait recognition systems. In Sect. 2.3, we describe the methodology, which combines use of a deep learning gait and pose recognition engine, and classification module to determine significant health-related changes. Section 2.4 contains description of the experiment which simulated an attempt to understand how the system would work for people with changing gait or emotional state. Finally, in Sect. 2.5, we discuss results and current and future utility of this approach.

2.2 Related Work

In terms of data modalities, there are mainly two categories of gait analysis approaches: sensor based and vision based. Although sensor-based approach has shown its ability to reflect human kinematics, the requirement of certain sensors or devices and needs to be worn on human body for some approach has made it less convenient to be applied. Vision-based approaches are more unobtrusive and only requires cameras for data collection [8].

Recently, skeleton has been widely used [9–14]. Some researchers employed Microsoft Kinect camera to generate 3D skeleton using its camera SDK. Gait abnormality index is estimated using 3D skeleton in [11], joint coordinates are used as input of auto-encoders. Then, reconstruction errors from auto-coders are used to differentiate abnormal gaits. Jun et al. [12] proposed to extract features from 3D skeleton data using a two recurrent neural network-based auto-encoder for abnormal gait recognition and then evaluated the performance by feeding the features extracted to discriminative models. While Kinect RGB-D camera provided additional depth information, in our preliminary experiment of comparing skeleton output from Azure Kinect body tracking SDK and Open Pose, it is not as robust as Open Pose.

Similar to our work, Xue et al. [14] presented a system for senior care using gait analysis. They accurately calculated gait parameters, including gait speed, stride length, step length, step width, and swing time from 2D skeleton. Compare to their work, besides gait parameters, we also employed time series analysis techniques to characterize motion dynamics. In addition, our system gives classification result of normal or abnormal gait.

There are several gait-related public datasets. CASIA gait database [15] includes video files of the subject walking with variations in view angle, clothing, whether or not carrying a bag. However, it does not have samples of the subject walking with abnormal gait patterns. And the resolution is quite low, 320×240 , and captured at 25 fps. INIT Gait database [16] is designed for gait impairment research; it consists of one normal gait pattern and seven simulated abnormal gait styles. But only binaries silhouettes sequences were released. The walking gait dataset by Nguyen et al. is designed for abnormal gait detection [11]. It includes point cloud, skeleton, and frontal silhouette captured by Microsoft Kinect 2 camera. Nine subjects performed nine different gaits on a treadmill. Aside from one normal gait, they simulated abnormal gaits by padding a sole with three different levels of thickness and attaching a weight to ankle. However, it is captured in front view, while our method is designed on sagittal plane from the side view. Thus, we captured our own dataset for this research.

2.3 Methodology

2.3.1 Pipeline Overview

Our system is set up to capture people walking left-right or right-left through a camera view, termed an event. A camera is placed to capture the movement in sagittal plane of human body while walking (see Fig. 2.7).

The processing system contains the three modules as shown in Fig. 2.1. The system continually monitors a camera view, and when a person walks through the camera view, that event is detected and captured to a video clip. Pose estimation is performed on the clip to obtain a sequence of skeleton which contains location of body joints. In feature extraction and gait classification stage, we extracted gait and postural parameters and employed time series analysis to classify if the gait is normal or abnormal.

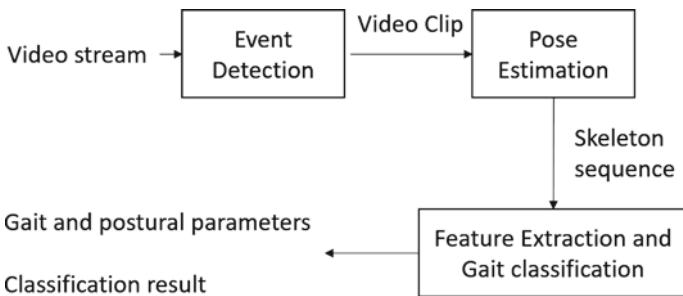


Fig. 2.1 Three processing modules of the system

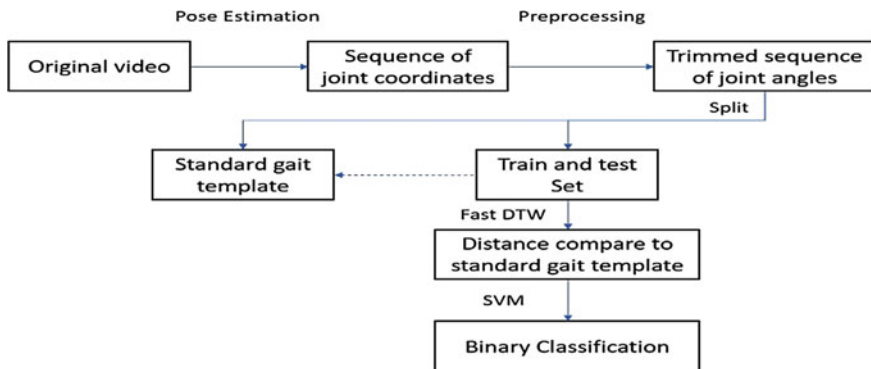


Fig. 2.2 Data processing flow

Event detection is performed in tripwire areas which located in both edges of the frame. When motion is detected in one side of the tripwire area, it starts recording an event; when motion is detected in another side, it ends the recording. We use a motion edge method (akin to optical flow) described in [17, 18].

After we captured the video clip of the subject walking perpendicular to the camera, we applied pose estimation on each frame of the video. The output of the pose estimation is a sequence of 25 anatomical joint coordinates in each frame. As shown in Fig. 2.2, in order to prepare the data for the binary classification, the following steps are performed: Trim the sequence to ensure all videos start from the same walking position. Keep a few samples of walking videos of normal gait as the standard gait template. Split the rest of the dataset into training and test set. We trained and evaluated two different algorithms of time series analysis to classify whether the gait in a video is normal or abnormal.

If the gait is classified as abnormal, it means the gait has deviated from their normal gait. To better analyze the gait, we also extract gait and posture features to help understand how the gait has changed.

2.3.2 Pose Estimation and Skeleton Extraction

Open Pose is an open-source real-time human 2D pose estimation deep learning model. It introduces a novel bottom-up approach to pose estimation using Part Affinity Fields (PAFs) to learn to associate body parts of a person in images or videos. Figure 2.3 shows Open Pose architecture (TensorFlow-based framework) [7].

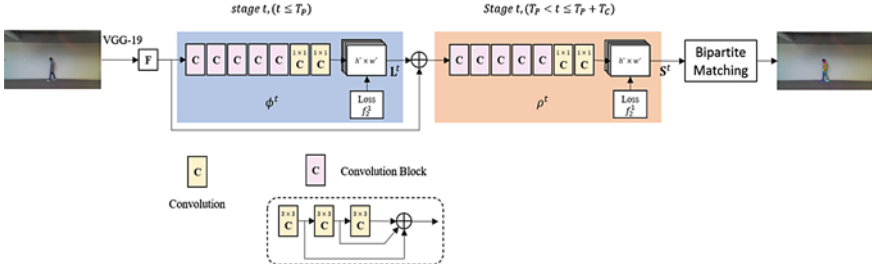


Fig. 2.3 Architecture of the multi-stage CNN

2.3.2.1 The Architecture of Open Pose

1. The system takes a color image as input, then analyzes it with a convolutional neural network (CNN) initialized and fine-tuned based on the first 10 layers of Visual Geometry Group-19 model (VGG-19).
2. A set of feature maps generated from CNN are feed into another multi-stage CNN.
3. The first set of stages predicts and refines PAFs, which is a set of 2D vector fields that encode the degree of association between body parts.
4. The last set of stages generates confidence maps of body part locations.
5. Finally, the confidence maps and the PAFs are parsed by bipartite matching to obtain 2D key points for each person in the image.

2.3.2.2 Part Affinity Fields for Part Association

PAFs contain location and orientation information across the region of support of the limb. It is a set of flow fields that encodes the unstructured pairwise relationship between body parts. Each pair of body parts have one PAF. PAFs are represented as set $\mathbf{L} = (\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_C)$, where $\mathbf{L}_c \in \mathbb{R}^{w \times h \times 2}$, $c \in \{1, \dots, C\}$. C denotes the number of pairs of body parts, $w \times h$ is the size of the input image. Each image location in \mathbf{L}_c encodes a 2D vector, if it lies on the limb c between body parts j_1 and j_2 , the value of PAF at that point is a unit vector that points from j_1 to j_2 ; otherwise, the vector is zero-valued. The ground truth PAF $\mathbf{L}_{c,k}^*$ at a point \mathbf{p} for person k as

$$\mathbf{L}_{c,k}^*(\mathbf{p}) = \begin{cases} \frac{\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}}{\|\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}\|_2} & \text{if } p \text{ on limb } c \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

where $\mathbf{x}_{j,k}$ is the ground truth position of the body part j of person k .

2.3.2.3 Confidence Map for Part Detection

Each confidence map is a 2D representation of the belief that a particular body part can be located in any given pixel. Each body part has one corresponding confidence map. Confidence maps are represented as set $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_J)$, where $\mathbf{S}_j \in \mathbb{R}^{w \times h}$, $j \in \{1, \dots, J\}$ and J denotes the number of body parts. Individual confidence maps $\mathbf{S}_{j,k}^*$ for each person k is defined as

$$\mathbf{S}_{j,k}^*(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{x}_{j,k}\|_2^2}{\sigma^2}\right) \quad (2.2)$$

where $\mathbf{x}_{j,k}$ is the ground truth position of the body part j of person k . The ground truth confidence map is an aggregation of individual confidence maps:

$$\mathbf{S}_j^*(\mathbf{p}) = \max_k \mathbf{S}_{j,k}^*(\mathbf{p}) \quad (2.3)$$

2.3.2.4 Multi-stage CNN

Stage $t = 1$: Given the feature maps \mathbf{F} generated from VGG-19, the network computes a set of part affinity fields, $\mathbf{L}^1 = \phi^1(\mathbf{F})$, where ϕ^1 refers to the CNN at stage 1.

Stage $2 \leq t \leq T_P$: Original feature maps \mathbf{F} and the PAF prediction from previous stage are concatenated to refine the prediction,

$$\mathbf{L}^t = \phi^t(\mathbf{F}, \mathbf{L}^{t-1}), \forall 2 \leq t \leq T_P \quad (2.4)$$

T_P refer to the number of PAF stages, ϕ^t is the CNN at stage t .

Stage $T_P < t \leq T_P + T_C$: After T_P iterations, starting from the most updated PAF prediction \mathbf{L}^{T_P} , the process is going to be repeated for T_C iterations to refine confidence map detection.

$$\mathbf{S}^{T_P} = \rho^t(\mathbf{F}, \mathbf{L}^{T_P}), \quad \forall t = T_P \quad (2.5)$$

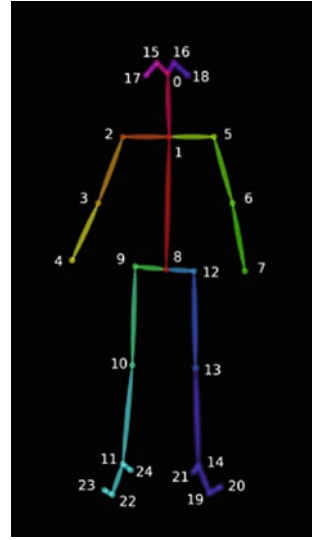
$$\mathbf{S}^t = \rho^t(\mathbf{F}, \mathbf{L}^{T_P}, \mathbf{S}^{t-1}), \quad \forall T_P < t \leq T_P + T_C \quad (2.6)$$

T_C refer to the number of confidence map stages, ρ^t is the CNN at stage t .

An L_2 loss function is applied at the end of each stage; it is specially weighted to tackle the case when people in some images are not completely labeled. Loss function at PAF stages t_i is

$$f_{\mathbf{L}}^{t_i} = \sum_{c=1}^C \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{L}_c^{t_i}(\mathbf{p}) - \mathbf{L}_c^*(\mathbf{p})\|_2^2 \quad (2.7)$$

Fig. 2.4 BODY_25 skeleton output. *Source* [19]



where \mathbf{L}_c^* is the ground truth PAF, \mathbf{W} is a binary mask. If pixel \mathbf{p} is not labeled, $\mathbf{W}(\mathbf{p}) = 0$. Loss function at confidence map stages t_k is

$$f_S^{t_k} = \sum_{j=1}^J \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{S}_j^{t_k}(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p})\|_2^2 \quad (2.8)$$

where \mathbf{S}_j^* is the ground truth part confidence map. The overall objective is to minimize the total loss.

$$f = \sum_{t=1}^{T_P} f_L^t + \sum_{t=T_P+1}^{T_P+T_C} f_S^t \quad (2.9)$$

The output of the Open Pose is BODY-25 output format as Fig. 2.4 shows; it consists of an (x, y) coordinate pair and confidence score for each of 25 joints [19].

2.3.3 Gait and Postural Feature Extraction

After getting pose estimation of each frame in the videos, we calculated angles of back and each lower leg respect to the vertical axis in each frame. The vertical axis in the coordinate system oriented downwards. Back angle is the angle of the vector start from “Neck” (key point 1) to “Mid Hip” (key point 8) respect to vertical. The angle of lower legs uses the vector point from knee to ankle, respect to vertical. Left lower leg: “LKnee” (key point 13) and “LAnkle” (key point 14); Right lower leg: “RKnee” (keypoint 10) and “RAnkle” (key point 11) (Table 2.1).

Table 2.1 List of body parts angles

Angle name	Joints
Left lower leg angle	Left knee (key point 13), left ankle (key point 14)
Right lower leg angle	Right knee (key point 10), right ankle (key point 11)
Back angle	Neck (key point 1), mid hip (key point 8)

Each video has three corresponding time series representing the angles mentioned above. And preprocessing is performed to clean up missing data, crop, and align time series.

In Fig. 2.5, left and right lower leg (knee-ankle) angles are colored in blue and orange, respectively. The first row (a–d) is a sample of normal gait, while the second row (e–h) is for abnormal gait. First column shows the original frame, the skeleton was extracted and displayed as shown in (b) and (f). (c) And (g) shows the vector, point from knee to ankle. The lower leg angle is the angle between the vector and vertical axis. (d) and (h) are the sequence of lower leg angle in the video, the horizontal axis of the plot is the frame index, and the vertical axis is angle in degree.

From the pose estimation result of the video, we also extracted gait and postural features listed below (see Figs. 2.6 and 2.7):

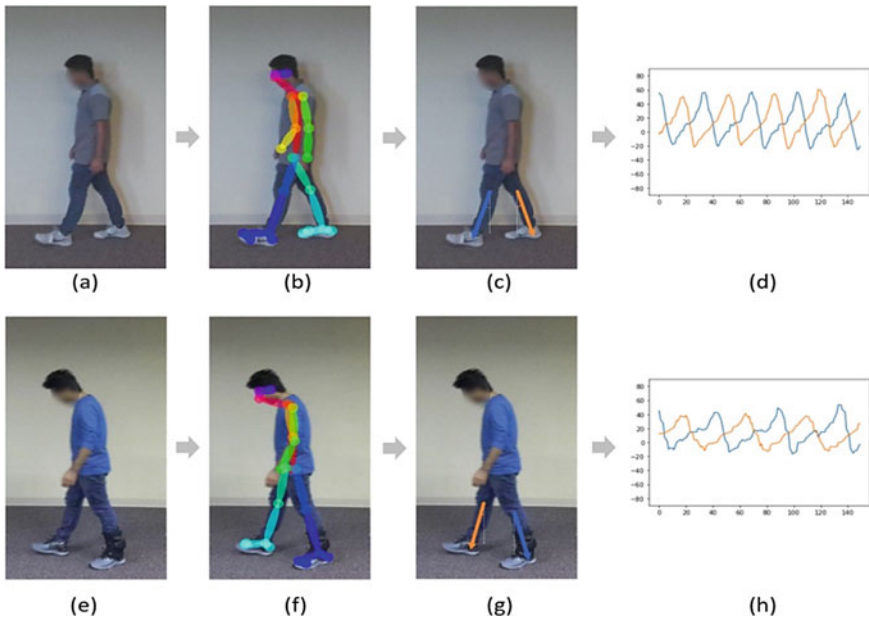


Fig. 2.5 Time series of lower leg angles

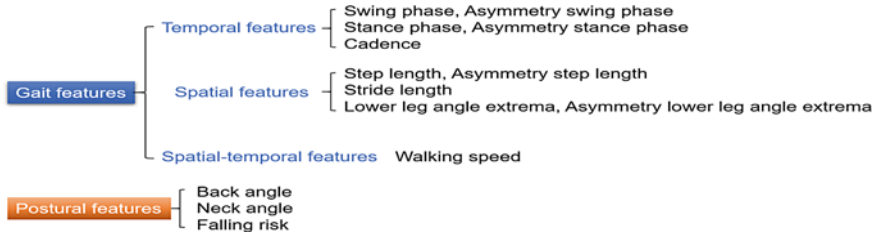


Fig. 2.6 Gait and postural feature

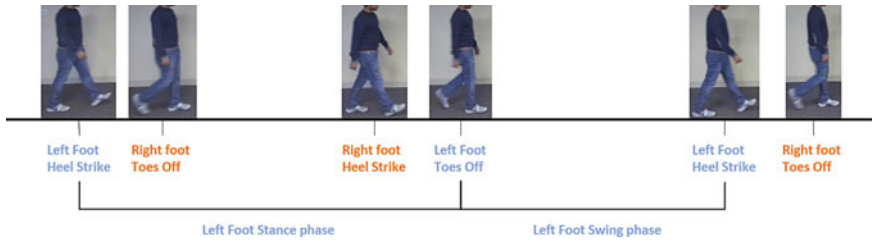


Fig. 2.7 Phases of the gait cycle

Stance Phase: The period starts from the heel strike, the heel of the same foot strike floor to toe-off, the foot is lifted from the floor. **Swing Phase:** The period that foot left the floor and swung forward in the air until the heel strikes the floor again [20].

Cadence: Cadence is the number of steps taken in a given period of time, expressed in steps per minute.

Step length: Distance between the contact points of two heels.

Stride length: Distance between two consecutive heel contact points of the same leg [21].

Lower leg angle extrema: The maximum and minimum of the lower leg angle time series, which reflects knee flexion while walking.

Asymmetry measure: To represent feature's asymmetry between left and right leg, we calculate the asymmetry measure of each feature. Let f denote feature, f_L and f_R denote feature extracted from left and right leg, respectively. Asymmetry measure A_f is defined as [22]:

$$A_f = \frac{|f_L - f_R|}{\max(f_L, f_R)} \quad (2.10)$$

Walking speed: Walking speed is calculated by dividing walking distance with the time period taken.

Back Angle: Back angle is between the mid-hip (key point 8) and bottom of the neck (key point 1)

Neck Angle: Neck angle is between the bottom of the neck and the nose (key point 0).

Falling risk: Body posture at left heel strike phase is used to calculate falling risk. Similar to [22], falling risk Fr is defined as:

$$Fr = \frac{\left| Nose_x - \frac{(LHeel_x + RHeel_x)}{2} \right|}{\frac{|LHeel_x - RToe_x|}{2}} \tag{2.11}$$

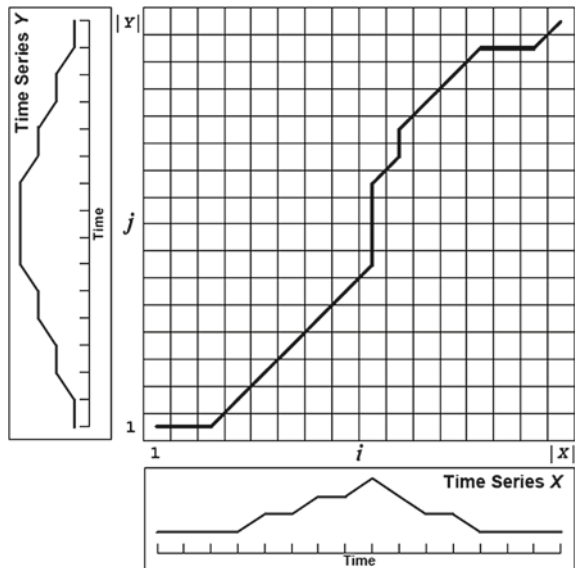
where $Nose_x$ is the x -coordinate of Nose (key point 0), $LHeel_x$, $RHeel_x$ and $RToe_x$ are x -coordinate of Left heel (keypoint 21), Right heel (keypoint 24), and Right big toe (keypoint 22).

2.3.4 Time Series Analysis

In order to measure the similarity between time series derived from pose estimation, we use fast dynamic time warping (DTW) [23] method to help calculate Euclidean distance with optimal alignment (see Fig. 2.8). And, support vector machine (SVM) is used for binary classification of normal and abnormal gait.

Due to the nature of the human walk, there might have shifts and distortions in gait data between each walk in the time axis, caused by subtle difference in walking speed or cadence. It is hard to have sequences aligned perfectly. Even between multiple samples of the normal walk, the slight time shift causes the distance to be considerably

Fig. 2.8 Cost matrix with the minimum-distance wrap path traced through it. Source [23]



large, making it unable to differentiate with abnormal gait, as their distances are both large. So directly calculating Euclidean distance is going to give poor results.

Dynamic time warping (DTW) [24] algorithms are commonly used to overcome shifts in the time dimension. Assume we have two time series X and Y . The value of a cell $D(i, j)$ in cost matrix D is the minimum-distance warp path of sequence $X' = x_1, \dots, x_i$ and $Y' = y_1, \dots, y_j$, $\text{Dist}(i, j)$ denote distance of two data points x_i and y_j .

$$D(i, j) = \text{Dist}(i, j) + \min[D(i-1, j), D(i, j-1), D(i-1, j-1)] \quad (2.12)$$

Fast DTW approximates DTW, using a multi-level approach to achieve linear time and space complexity, in contrast to quadratic time and space complexity in standard DTW algorithm. It first produces different lower resolutions of the time series, by taking an average of adjacent pairs of points. Then, project the minimum distance calculated from lower resolution to higher resolution as an initial guess. Finally, refine the wrap path by local adjustments.

After calculation of DTW distance for each time series, we used support vector machine (SVM) on a multidimensional DTW distance vector for classification.

2.4 Experiments and Results

2.4.1 Data Collection

In our experiment, we aim to simulate gait in different health states within the laboratory. Because it is difficult to change health status of our volunteers, we propose to use different levels of physical ankle weights in the experiment to help them demonstrate abnormal gait. Intuitively, when additional weights added on human body, gait will change accordingly as mobility and stability of walking are affected.

The adjustable ankle weights strap with removable sand packets were used because it can easily adjust the weight by adding or removing sand packets on the strap. Each strap has five slots to hold sand packets, and each packet weighs around 0.6 lbs.

We designed the experiment capturing the same person walking across the camera normally for 10 times, walking with three different levels of weights 10 times, respectively, and walking with the 3rd level of weights plus carrying a heavy box for 10 times (see Fig. 2.9).

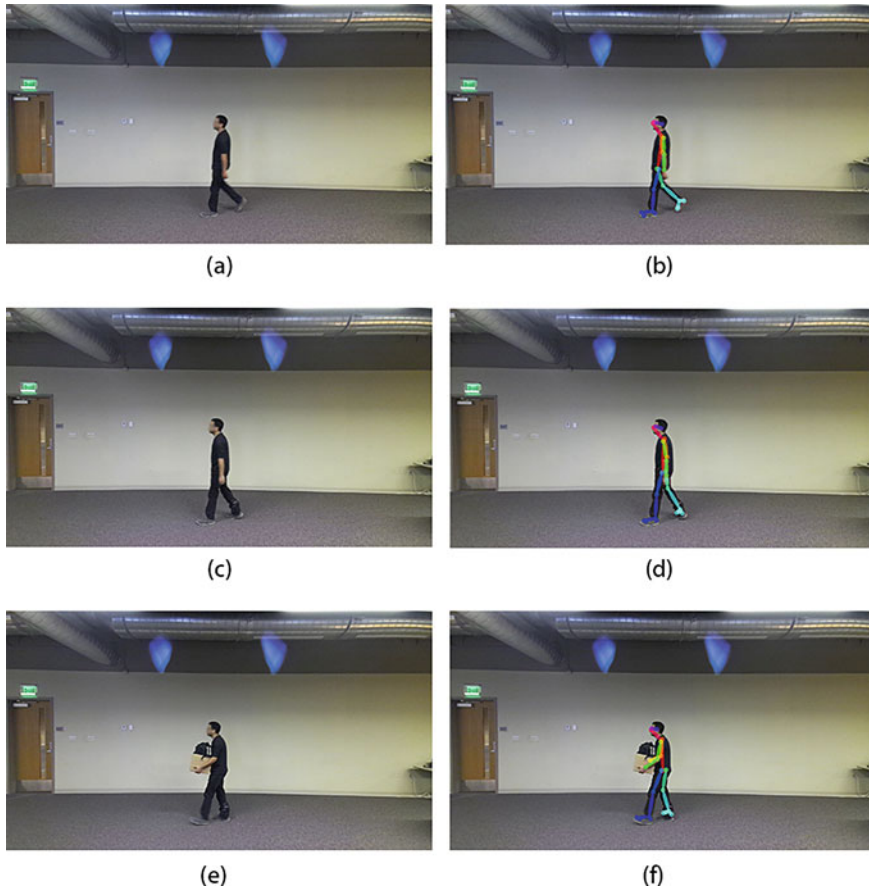


Fig. 2.9 Sample frames in the dataset: **a** original frame in video of normal walk; **b** normal walk frame with skeleton overlay; **c** original frame in video of walk with 3rd level of weights; **d** frame of 3rd level of weights experiment with skeleton overlay; **e** original frame in video of walk with 3rd level of weights and carrying a heavy box; **f** frame of 3rd level of weights and a heavy box with skeleton overlay

2.4.2 Experiment Setting

Videos we captured are with resolution of 1920×1080 and frame rate of 30 frames per second. In pose estimation stage, Open Pose was also processed at 30 FPS. We dropped frames with incomplete or low confidence joints to filter out the low-quality frames. Failure of pose estimation in these frames usually caused by motion blur in the frame, especially at lower leg and foot area where the amplitude of motion is largest.

In the preprocessing stage, we first performed imputation using interpolation to fill the missing data. Then, we cropped all the sequences to start from the first peak

(maximum) point of the right lower leg angle's time series, so that all of the sequences should start from nearly the same gait phase which minimized the noise from data misalignment.

In our experiment, for each person, we set aside three videos of normal gait as standard template gait data to be compared with. The rest of the videos were split into training and test set, and for each one of them, we calculate the distance with three standard template gait data using fast DTW then take an average to measure how close it is comparing to normal gait. Linear kernel is used in SVM for classification.

2.4.3 Results for Gait Classification

2.4.3.1 Evaluation Metrics

Gait classification is a binary classification task to predict if an unlabeled video shows normal or abnormal gait. Detection of abnormal gait is defined as "positive." If the ground truth label of the data matches detection result, it is defined as "true," otherwise it is "false."

We employed accuracy rate, precision rate, recall rate and F_1 score in our experiment.

2.4.3.2 Intra-subject Cross Validation

To evaluate the performance of our proposed methods, leave-one-out cross validation is performed within each subject. As Fig. 2.10 shows, for each subject's data, we reserve one as test set and use the rest of this subject's data to train the algorithms. The same process is performed on each of four subjects.

Results of leave-one-out cross validation are listed as below (Table 2.2).

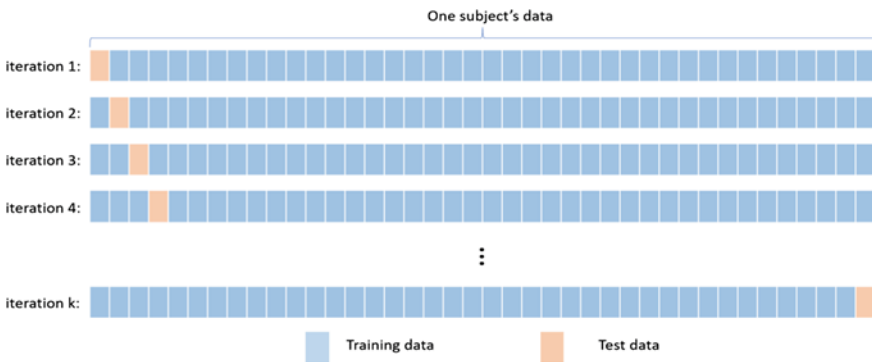


Fig. 2.10 Leave-one-out cross validation

Table 2.2 Results of leave-one-out cross validation

Subject	Precision	DTW-SVM Recall	Accuracy	F_1 score
1	1.000	1.000	1.000	1.000
2	0.974	0.925	0.915	0.949
3	1.000	0.975	0.979	0.987
4	1.000	1.000	1.000	1.000
Average	0.993	0.975	0.973	0.984

2.4.3.3 Inter-subject Cross Validation

We also performed leave-one subject-out cross validation to validate the performance when applied on a different subject that is not included in training set. As Fig. 2.11 shows, in each iteration, one subject’s data is used as test set. All other subjects’ data is used for training.

From the result of leave-one-subject-out cross validation, we can see that both methods successfully classified normal and abnormal gaits. The DTW-SVM-based method achieved 0.982 in F_1 score (Table 2.3).



Fig. 2.11 Leave-one-subject-out cross validation

Table 2.3 Results of leave-one subject-out cross validation

Subject	Precision	DTW-SVM Recall	Accuracy	F_1 score
1	1.000	1.000	1.000	1.000
2	0.907	0.975	0.894	0.940
3	1.000	0.975	0.979	0.987
4	1.000	1.000	1.000	1.000
Average	0.977	0.988	0.968	0.982

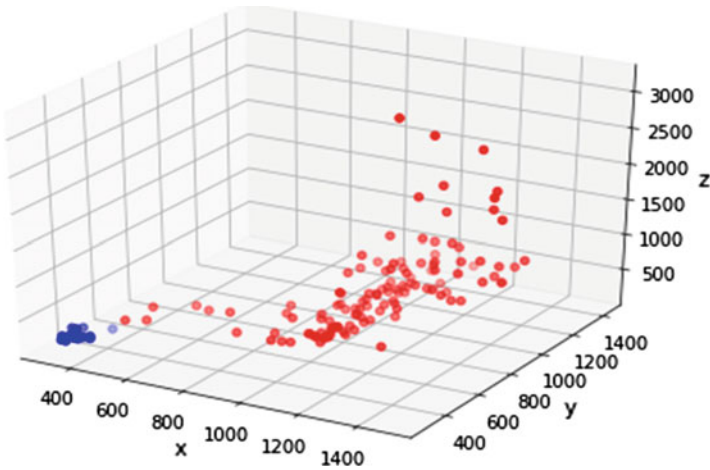


Fig. 2.12 DTW distance to standard gait template; data points of normal gait were marked with blue dots; data points of abnormal gait were marked with red dots

2.4.4 Discussion

The proposed DTW-SVM-based method performed classification by measuring the deviation from the standard gait template. As Fig. 2.12 shows, the x -, y -, z -axis denotes the DTW distance of left and right lower leg angle, back angle compares to standard gait template, respectively. Normal gait's data is marked with blue while abnormal gait's data is marked with red. Intuitively, normal gait is closer to the standard gait template, yet the distance should be small, so the data points are clustered near origin of the coordinate system.

As DTW-SVM-based method measures deviation, so it has better performance while applied on inter-subject prediction. It only requires a few samples as standard gait template to be compared with.

2.4.5 Graphical User Interface and Companion Robot

The gait monitoring system based on body posture and walking speed is implemented as a module in companion robot. On a robot's monitor directly in line with the walking trajectory (see Fig. 2.7), we displayed a graphical "reward" of the person's gait with respect to average values among the population (see Fig. 2.13). So, people can see the results of the evaluation after they walked through the room. We have heard from users that this system motivates them to improve body posture.

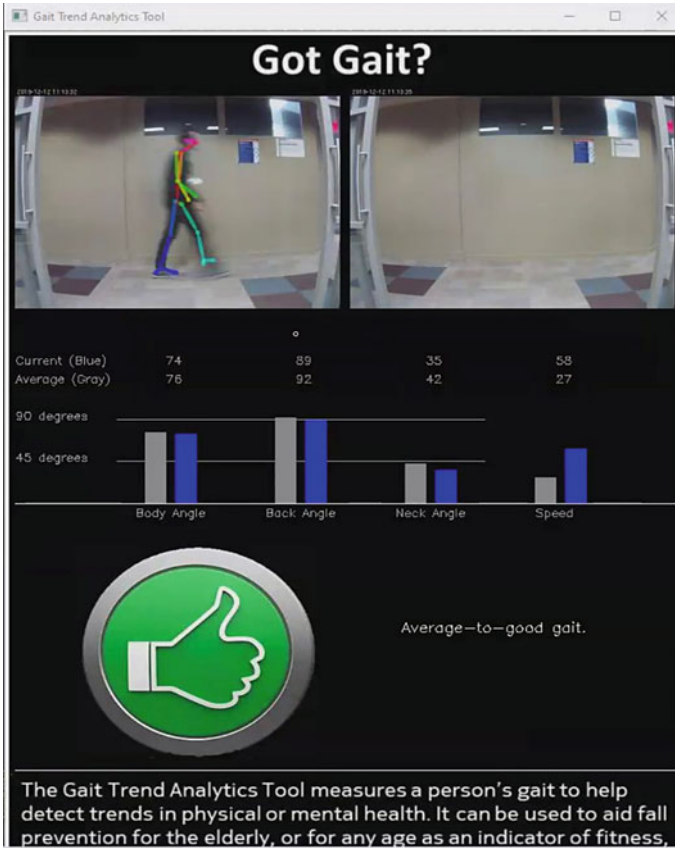


Fig. 2.13 Graphical user interface (GUI) of the gait analysis system

2.5 Conclusions and Future Work

This paper proposed a video-based gait analysis system to detect abnormal gait and capture gait features non-invasively over time. Gait analysis and gait abnormality detection allows early intervention and treatment to prevent underlying conditions develops and cause a fall. It can also evaluate the recovery progress of the physical therapy.

The system consists of three stages; first, the event detection module records the gait event video clip while a person walks through the camera. Second, pose estimation is applied to extract sequence of skeleton and joints in the video. Then, feature extraction and gait classification is performed to calculate gait and postural parameters and to classify if the gait is normal or abnormal.

In gait classification task, we proposed DTW-SVM-based method, using time series of lower leg angle and back angle. Experiment results shows that DTW-SVM-based method achieved higher accuracy in inter-subject classification. Gait and postural parameters extracted shows distinction between normal and abnormal gait. With tracking the parameters on a day-to-day basis, we can quantitatively monitor the gait changes in long term.

In the future, this work can be extended by adding another camera, located in front view, to analyze the motion and gait asymmetry in the coronal plane or frontal plane of the human body. The system can be integrated with a face recognition module to fit the need of a multi-person household or facility so that it can automatically recognize the identity of the person in the video and then add the extracted gait features to their record. When the system detected gait abnormality of a person, an alert will be generated for medical attention.

Acknowledgements This work was supported by the National Science Fund of Bulgaria: KP-06-H27/16 *Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas* and NOKIA Corporate University Donation number NSN FI (85) 1198342 MCA.

References

1. NCOA: Falls Prevention Facts. Retrieved from: <https://www.ncoa.org/news/resources-for-reporters/get-the-facts/falls-prevention-facts/>. 12 May 2020
2. MedlinePlus: Walking Problems. <https://medlineplus.gov/walkingproblems.html>. Accessed 12 May 2020
3. Selke, Stefan: Book Lifelogging. Springer, Heidelberg (2016)
4. Jalal, A., Kamal, S., Kim, D.: Depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* **14**(7), 11735–11759 (2014)
5. Climent-Pérez, P., Spinsante, S., Mihailidis, A., Florez-Revuelta, S.: A review on video-based active and assisted living technologies for automated lifelogging. *Expert Syst. Appl. J.* **139** (2020)
6. Liu, X., Milanova, M.: An image captioning method for infant sleeping environment diagnosis. In: IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction, pp. 18–26. Springer, Cham (2018)
7. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: Open Pose: real-time multi-person 2D pose estimation using Part Affinity Fields. arXiv preprint [arXiv:1812.08008](https://arxiv.org/abs/1812.08008) (2018)
8. Muro-De-La-Herran, A., Garcia-Zapirain, B., Mendez-Zorrilla, A.: Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors* **14**(2), 3362–3394 (2014)
9. Meng, M., Drira, H., Daoudi, M., Boonaert, J.: Detection of abnormal gait from skeleton data. In: 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016), vol. 3: VISAPP, pp. 133–139 (2016)
10. Khokhlova, M., Migniot, C., Morozov, A., Sushkova, O., Dipanda, A.: Normal and pathological gait classification LSTM model. *Artif. Intell. Med.* **94**, 54–66 (2019)
11. Nguyen, T.N., Huynh, H.H., Meunier, J.: Estimating skeleton-based gait abnormality index by sparse deep auto-encoder. In: 2018 IEEE Seventh International Conference on Communications and Electronics (ICCE), pp. 311–315 (2018)

12. Jun, K., Lee, D.W., Lee, K., Lee, S., Kim, M.S.: Feature extraction using an RNN autoencoder for skeleton-based abnormal gait recognition. *IEEE Access* **8**, 19196–19207 (2020)
13. Dolatabadi, E., Zhi, Y.X., Flint, A.J., Mansfield, A., Iaboni, A., Taati, B.: The feasibility of a vision-based sensor for longitudinal monitoring of mobility in older adults with dementia. *Arch. Gerontol. Geriatr.* **82**, 200–206 (2019)
14. Xue, D., Sayana, A., Darke, E., Shen, K., Hsieh, J.T., Luo, Z., Li, L.-J., Lance Downing, N., Milstein, A., Fei-Fei, L.: Vision-based gait analysis for senior care. arXiv preprint [arXiv:1812.00169](https://arxiv.org/abs/1812.00169) (2018)
15. CASIA Gait Database. <https://medlineplus.gov/walkingproblems.html>. Accessed 12 May 2020
16. INIT Gait Database homepage. <https://www.vision.uji.es/gaitDB/>. Accessed 12 May 2020
17. O’Gorman, L., Yin, Y., Ho, T.K.: Motion feature filtering for event detection in crowded scenes. *Pattern Recogn. Lett.* **44**, 80–87 (2014)
18. O’Gorman, L., Yang, G.: Orthographic perspective mappings for consistent wide-area motion feature maps from multiple cameras. *IEEE Trans. Image Process.* **25**(6), 2817–2832 (2016)
19. OpenPose Output Format. Retrieved from: <https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md>. 12 May 2020
20. Dale, R.B.: Clinical gait assessment. *Physical Rehabilitation of the Injured Athlete*, pp. 464–479 (2012). <https://doi.org/10.1016/b978-1-4377-2411-0.00021-6>
21. Richards, J., Chohan, A., Erande, R.: Biomechanics. *Tidy’s Physiotherapy*, pp. 331–368 (2013). <https://doi.org/10.1016/b978-0-7020-4344-4.00015-8>
22. Ortells, J., Herrero-Ezquerro, M.T., Mollineda, R.A.: Vision-based gait impairment analysis for aided diagnosis. *Med. Biol. Eng. Comput.* **56**(9), 1553–1564 (2018)
23. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **11**(5), 561–580 (2007)
24. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: *KDD Workshop*, vol. 10, No. 16, pp. 359–370 (1994)

Chapter 3

Comparative Analysis of the Hierarchical 3D-SVD and Reduced Inverse Tensor Pyramid in Regard to Famous 3D Orthogonal Transforms



Roumen Kountchev and Roumiana Kountcheva

Abstract In this work are presented two new approaches for hierarchical decomposition represented as tensors of size $N \times N \times N$ for $N = 2^n$, based on algorithms which (unlike the famous similar approaches) do not require iterative calculations. Instead, they use repetitive simple calculations in each hierarchical decomposition level. As a result, the computational complexity (CC) of the new hierarchical algorithms is lower than that of the iteration-based. In general, hierarchical decompositions are divided into two basic groups: statistical and deterministic. To the first group is assigned the algorithm hierarchical tensor SVD (HTSVD) based on the multiple calculation of the two-level SVD for the elementary tensor of size $2 \times 2 \times 2$. The decomposition is executed by using the HTSVD in three orthogonal spatial directions simultaneously. The deterministic decompositions have lower CC than the statistical, but they do not ensure full decorrelation between the components of the 3D decompositions. In this group are the famous orthogonal transforms 3D fast Fourier transform (3D-FFT), 3D discrete cosine transform (3D-DCT), 3D discrete wavelet transform (3D-DWT), 3D contourlet discrete transform (3D-CDT), 3D shearlet discrete transform (3D-SDT), etc., and also, the algorithm 3D reduced inverse spectrum pyramid (3D-RISP). The last is distinguished by its lower CC and the high energy concentration in the first decomposition components. To achieve this, for the basic tensor of size $2^n \times 2^n \times 2^n$ is executed the 3D fast truncated Walsh–Hadamard transform (3D-FTWHT). Significant advantage of 3D-RISP compared to the famous pyramidal decompositions of the kind 3D-DWT, 3D-CDT, 3D-SDT, etc., is the absence of 3D decimation and 3D interpolation which produce distortions in the restored tensor.

R. Kountchev (✉)
Technical University of Sofia, Sofia 1756, Bulgaria
e-mail: rkountch@tu-sofia.bg

R. Kountcheva
TK Engineering, Drujba 2 404/2, Sofia 1582, Bulgaria

3.1 Introduction

Tensor decompositions became recently an important object of various investigations presented in many publications [1, 2], and the ideas were implemented as new software algorithms and programs [3, 4]. Significant interest attracts the tensor decomposition of 3D images which comprise sequences of 2D images and are more or less correlated in the spatial–temporal areas. Example 3D images of the kind are: the hyperspectral image (HSI), computer tomography (CTI), magnetic resonance (MRI), multi-view (MVI), video sequences, etc. As a result of the decomposition, each third-order tensor (respectively, 3D image) is represented as a sum of components—tensors of the same order. The optimum tensor decomposition has several important qualities: (1) the correlation between tensor components for the three mutually orthogonal directions, is zero—i.e., full decorrelation is achieved; (2) the tensor energy is concentrated mainly in the first several decomposition components; (3) in result of the low-energy components “truncation,” the restored tensor calculated by using the retained components only, has minimum mean square error; (4) the global number of parameters defined by the full (non-reduced) sum of the parameters for each tensor component, is not larger than that of the original tensor, i.e., the decomposition is “non-overcomplete.” All known methods for tensor decomposition could be divided into two basic groups: statistical and deterministic. In the group of the *statistical methods* for image decomposition are various different multilinear extensions of the matrix-SVD, called multilinear SVD (MSVD), or generalizations of the SVD matrix for higher-order tensors, called higher-order SVD (HOSVD) [5–7]. Such are also the famous methods: CANDECOMP/PARAFAC or canonical polyadic decomposition (CPD) where the tensor is represented as a sum of rank-one tensors; the tucker decomposition (TD) [1, 3, 6]; the tensor train decomposition [8]; the Kruskal decomposition, etc. The statistical methods are implemented through applying various algorithms for calculation of the tensors eigenvectors, which have relatively high computational complexity. The tensor decomposition components are usually calculated by using iterative methods whose iterations stop, when the pre-defined accuracy is achieved. Such are: the tensor power iteration; the QR-factorization followed by the householder transforms (or the Gram–Schmidt process), the Givens rotations; the Jacobi method; the higher-order eigenvalue decomposition (HOEVD); the SVD calculation based on its relation to PCA, etc. The tensor decomposition based on the use of iterative SVD methods needs significant number of computational operations. To overcome the problem, various hierarchical methods are already developed, based on the hierarchical tucker decomposition (HTD) [9], the sequentially truncated HOSVD (ST-HOSVD) [10], and the sequential unfolding SVD (SUSVD) [11]. In the same group is also the non-iterative Hierarchical SVD algorithm for tensor decomposition offered in [12]. It has lower computational complexity and is based on SVD for elementary tensor of size $2 \times 2 \times 2$.

In the group of the *deterministic methods* for image decomposition are the pyramidal 3D transforms: the 3D discrete wavelet transform (3D-DWT) [13], the 3D curvelet and the 3D contourlet discrete transform (3D-CDT) [14, 15] and the shearlet

discrete transform (SDT) [16]. The methods from the first group overcome these from the second in respect of the decomposition components' decorrelation degree, but these in the second group have much lower computational complexity. The deterministic methods for tensor decomposition are usually executed by using various kinds of 3D orthogonal transforms. Publications [4, 13, 15] are proposed algorithms for cubical decomposition based on the 3D separable discrete transforms: the 3D discrete Fourier transform (3D-DFT), the 3D discrete Hartley transform (3D-DHT), the 3D discrete cosine transform (3D-DCT), etc.; the algorithm for hierarchical third-order tensor decomposition with low CC, based on the 3D inverse spectrum pyramid (3D-ISP), is offered in [17].

The choice of a method from the two basic groups for a certain application is done on the basis of the requirements which the algorithm for 3D image decomposition must satisfy in restoration accuracy, execution time, etc. In this work are analyzed and compared the approaches for hierarchical third-order tensors decomposition: the 3D truncated hierarchical SVD (3D-THSVD) [18], and the 3D-ISP combined with the 3D truncated Walsh–Hadamard Transform (3D-TWHT) [17]. In the next sections are given: the decorrelation of image sequences through 3D-HAPCA; tensor decomposition through 3D-THSVD; the pyramidal decomposition based on the 3D-ISP and on the truncated divisible 3D-TWHT, comparison of the CC of the algorithms for hierarchical tensor decomposition and conclusions.

3.2 Tensor Decomposition Through 3D Truncated Hierarchical SVD

3.2.1 Decomposition of Elementary Tensor of Size $2 \times 2 \times 2$ Through 3D Truncated Hierarchical SVD

The algorithm 3D-THSVD is based on the calculation of $SVD_{2 \times 2}$ for the matrix image $[X]$ of size 2×2 , in correspondence with the relation below [18]:

$$\begin{aligned}
 [X] &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \sigma_1 \vec{U}_1 \vec{V}_1^T + \sigma_2 \vec{U}_2 \vec{V}_2^T \\
 &= [C_1] + [C_2] = \frac{1}{2A} \left\{ \sigma_1 \begin{bmatrix} B & C \\ D & E \end{bmatrix} + \sigma_2 \begin{bmatrix} E & -D \\ -C & B \end{bmatrix} \right\}, \quad (3.1)
 \end{aligned}$$

where a, b, c, d are the elements of the matrix $[X]$; $\sigma_{1,2} = \sqrt{(\omega \pm A)/2}$ —the singular values of the matrix $[X]$; $\vec{U}_1 = (1/\sqrt{2A})[\sqrt{r}, \sqrt{q}]^T$; $\vec{U}_2 = (1/\sqrt{2A})[-\sqrt{q}, \sqrt{p}]^T$ —the left eigenvectors of the matrix $[X]$; $\vec{V}_1 = (1/\sqrt{2A})[\sqrt{p}, \sqrt{s}]^T$; $\vec{V}_2 = (1/\sqrt{2A})[-\sqrt{s}, \sqrt{p}]^T$ —the right eigenvectors of the matrix $[X]$;

$$A = \sqrt{v^2 + 4\eta^2}; B = \sqrt{rp}; C = \sqrt{sp}; D = \sqrt{rq}; E = \sqrt{sq};$$

$$[C_1] = \frac{\sigma_1}{2A} \begin{bmatrix} B & C \\ D & E \end{bmatrix}; [C_2] = \frac{\sigma_2}{2A} \begin{bmatrix} E & -D \\ -C & B \end{bmatrix};$$

$$\omega = a^2 + b^2 + c^2 + d^2; \nu = a^2 + c^2 - b^2 - d^2; \mu = a^2 + b^2 - c^2 - d^2;$$

$$\eta = ab + cd; r = A + \nu; q = A - \mu; p = A + \mu; s = A - \nu.$$

The elements of matrices $[C_1]$ and $[C_2]$ are defined by four parameters only (ω , ν , μ , η), and from this, it follows that the decomposition represented by Eq. (3.1) is not “overcomplete.”

In Fig. 3.1 is shown the algorithm 3D-THSVD for a two-level decomposition of the elementary tensor \mathbf{X} of size $2 \times 2 \times 2$ (ET), which is calculated on the basis of $\text{SVD}_{2 \times 2}$ for a matrix image of size 2×2 , in correspondence with Eq. (3.1).

In the *first level* of the algorithm, after mode-1 unfolding of the tensor $\mathbf{X}_{2 \times 2 \times 2}$, on the matrices $[X_1]$ and $[X_2]$ is applied the $\text{SVD}_{2 \times 2}$. In result, are calculated the matrices $[C_{11}]$ and $[C_{21}]$ which have high weights σ_{11} , σ_{21} (their elements are colored in red), and the matrices $[C_{12}]$ and $[C_{22}]$ whose weights σ_{12} , σ_{22} are small (their elements are colored in blue).

After the rearrangement of the matrices in accordance with the lessening of their weights, the tensors $\mathbf{X}_{1(2 \times 2 \times 2)}$ and $\mathbf{X}_{2(2 \times 2 \times 2)}$ are reconstructed and as a result is got the decomposition:

$$\mathbf{X}_{2 \times 2 \times 2} = \mathbf{X}_{1(2 \times 2 \times 2)} + \mathbf{X}_{2(2 \times 2 \times 2)}. \quad (3.2)$$

In the *second level* of the algorithm, on the tensors $\mathbf{X}_{1(2 \times 2 \times 2)}$ and $\mathbf{X}_{2(2 \times 2 \times 2)}$ are applied mode-2 unfolding and are got the couples of matrices $[X_{11}]$, $[X_{12}]$ and $[X_{21}]$, $[X_{22}]$, for which are calculated the corresponding $\text{SVD}_{2 \times 2}$.

$[X_{11}]$ and $[X_{12}]$ are got four matrices: $[C_{111}]$, $[C_{121}]$ —with high weights σ_{111} , σ_{121} , and $[C_{112}]$, $[C_{122}]$ —with low weights, σ_{112} , σ_{122} . $[X_{21}]$, $[X_{22}]$ are got another four matrices: $[C_{211}]$, $[C_{221}]$ —with high weights σ_{211} , σ_{221} , and $[C_{212}]$, $[C_{222}]$ —with low weights, σ_{212} , σ_{222} . After rearrangement in accordance with their weights lessening, from these eight matrices $[C_{111}]$, $[C_{121}]$, $[C_{112}]$, $[C_{122}]$, $[C_{211}]$, $[C_{221}]$, $[C_{212}]$, $[C_{222}]$ are reconstructed four tensors $\mathbf{X}_{r,1(2 \times 2 \times 2)}$, $\mathbf{X}_{r,2(2 \times 2 \times 2)}$, $\mathbf{X}_{r,3(2 \times 2 \times 2)}$, $\mathbf{X}_{r,4(2 \times 2 \times 2)}$, each of size $2 \times 2 \times 2$. Then the decomposition is represented as follows:

$$\mathbf{X}_{2 \times 2 \times 2} = \mathbf{X}_{1(2 \times 2 \times 2)} + \mathbf{X}_{2(2 \times 2 \times 2)} + \mathbf{X}_{3(2 \times 2 \times 2)} + \mathbf{X}_{4(2 \times 2 \times 2)}, \quad (3.3)$$

where

$$\begin{aligned} \mathbf{X}_{1(2 \times 2 \times 2)} &\Rightarrow \text{fold}_{\text{mod } e-2} ([C_{111}], [C_{121}]); \\ \mathbf{X}_{2(2 \times 2 \times 2)} &\Rightarrow \text{fold}_{\text{mod } e-2} ([C_{211}] \text{ or } [C_{112}], [C_{221}] \text{ or } [C_{122}]); \\ \mathbf{X}_{3(2 \times 2 \times 2)} &\Rightarrow \text{fold}_{\text{mod } e-2} ([C_{112}] \text{ or } [C_{211}], [C_{122}] \text{ or } [C_{221}]); \\ \mathbf{X}_{4(2 \times 2 \times 2)} &\Rightarrow \text{fold}_{\text{mod } e-2} ([C_{212}], [C_{222}]). \end{aligned}$$

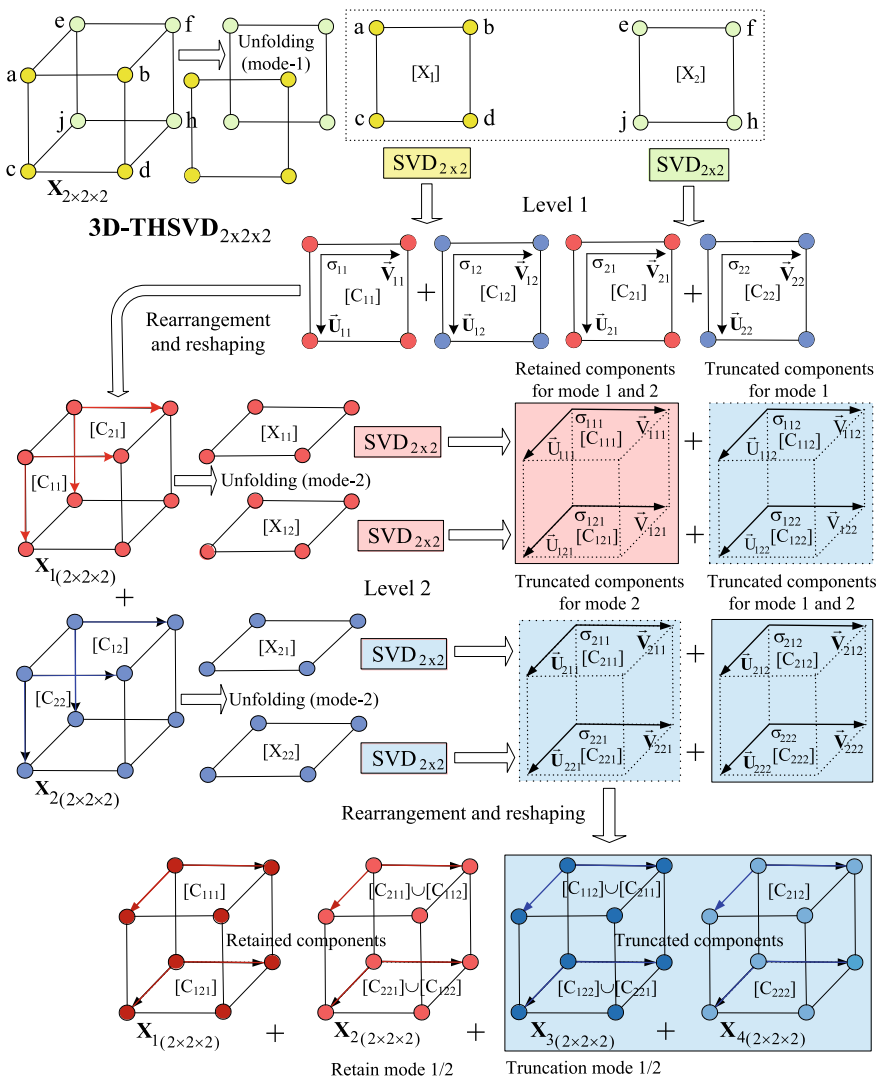


Fig. 3.1 Structure of the two-level adaptive 3D-THSVD for the tensor $X_{2 \times 2 \times 2}$ based on $SVD_{2 \times 2}$

The tensors $X_{1(2 \times 2 \times 2)}$ and $X_{4(2 \times 2 \times 2)}$ are always reconstructed on the basis of the couples of matrices $[C_{111}], [C_{121}]$ and $[C_{212}], [C_{222}]$, respectively. The tensor $X_{2(2 \times 2 \times 2)}$ is reconstructed by using the couples of matrices $[C_{211}]$ and $[C_{221}]$, or respectively $[C_{112}]$ and $[C_{122}]$, and the tensor $X_{3(2 \times 2 \times 2)}$ —by using the couples of matrices $[C_{112}], [C_{122}]$ or, respectively— $[C_{211}], [C_{221}]$. The choice of the couple of matrices is defined by the selected truncation for mode-1 or mode-2 unfolding. the decomposition from Eq. (3.3) is retained two tensors only: $X_{1(2 \times 2 \times 2)}$ and $X_{2(2 \times 2 \times 2)}$.

For their reconstruction, from the calculated eight matrices in the second level are used the four matrices of highest weights only.

For a given approximation accuracy ε of the tensor $\mathbf{X}_{2 \times 2 \times 2}$, the remaining two tensors $\mathbf{X}_{3(2 \times 2 \times 2)}$ and $\mathbf{X}_{4(2 \times 2 \times 2)}$ in Eq. (3.3) could be neglected, and in result is got:

$$\begin{aligned} \mathbf{X}_{2 \times 2 \times 2} &\approx \hat{\mathbf{X}}_{2 \times 2 \times 2} = \sum_{r=1}^2 \mathbf{X}_{r(2 \times 2 \times 2)} \text{ for } \|\mathbf{X}_{2 \times 2 \times 2} - \hat{\mathbf{X}}_{2 \times 2 \times 2}\|_F \\ &= \left\| \sum_{r=3}^4 \mathbf{X}_{r(2 \times 2 \times 2)} \right\|_F \leq \varepsilon. \end{aligned} \quad (3.4)$$

where $\|\cdot\|_F$ is the Frobenius norm.

The decomposition of tensor $\mathbf{X}_{2 \times 2 \times 2}$ in correspondence with Eq. (3.4) represents the algorithm THSVD $_{2 \times 2 \times 2}$. In the second decomposition level, the SVD $_{2 \times 2}$ is executed four times and each SVD $_{2 \times 2}$ is defined by four parameters (three—for the first matrix component $[\mathbf{C}_1]$, and one—for the second $[\mathbf{C}_2]$). Each of the tensors from Eq. (3.3) contains the components of two selected SVD $_{2 \times 2}$ (shown in Fig. 3.1 colored in red and blue). Then, the total number of parameters needed for the description of all four tensors from Eq. (3.3) is $4 \times 4 = 16$. For the “truncated” decomposition represented by Eq. (3.4), which contains the tensors $\mathbf{X}_{1(2 \times 2 \times 2)}$ and $\mathbf{X}_{2(2 \times 2 \times 2)}$ only, the total number of needed parameters is 20 or 14, depending on the kind of truncation chosen to neglect the components of low weights for the selected four SVD $_{2 \times 2}$. In the first way of truncation (mode-1), the second retained tensor $\mathbf{X}_{2(2 \times 2 \times 2)}$ is defined by matrices $[\mathbf{C}_{211}]$ and $[\mathbf{C}_{221}]$, and the first “truncated” tensor $\mathbf{X}_{3(2 \times 2 \times 2)}$ —by the matrices $[\mathbf{C}_{112}]$ and $[\mathbf{C}_{122}]$. In this case, the “truncated” decomposition is represented by 20 parameters. In the second way of truncation (mode-2), the second retained tensor $\mathbf{X}_{2(2 \times 2 \times 2)}$ is defined by the matrices $[\mathbf{C}_{112}]$ and $[\mathbf{C}_{122}]$, and the first “truncated” tensor $\mathbf{X}_{3(2 \times 2 \times 2)}$ —by the matrices $[\mathbf{C}_{211}]$ and $[\mathbf{C}_{221}]$. The tensor $\mathbf{X}_{4(2 \times 2 \times 2)}$ is cutoff in both cases (mode-1 and mode-2). Hence, for THSVD mode-2 are needed $2 \times 3 + 2 \times 4 = 14$ parameters in total, because in the first decomposition level is calculated the first matrix component of each SVD $_{2 \times 2}$ only. Then, the total number of parameters, needed for the ET decomposition based on the THSVD $_{222}$ truncation mode-2 (equal to 14), is the smaller, when compared to algorithm HSVD $_{222}$ (with 24 parameters needed). For the execution of THSVD $_{2 \times 2 \times 2}$ truncation mode-2, in the second level must be calculated only the two SVD $_{2 \times 2}$, colored in red in Fig. 3.1.

The reconstruction of tensors $\mathbf{X}_{1(2 \times 2 \times 2)}$ and $\mathbf{X}_{2(2 \times 2 \times 2)}$ depends on the truncation mode (1 or 2), applied on the eight matrices in the second decomposition level. The “truncation” could be fixed, or adaptive. The adaptive truncation criterion is defined on the basis of the total weight of the matrices used for the reconstruction of tensor \mathbf{X}_2 , when compared to that of the matrices used for the tensor \mathbf{X}_3 . The *adaptive truncation rule* is defined by the relation:

$$\text{Truncation mode} \begin{cases} 1, & \text{if } (\sigma_{211} + \sigma_{221}) > (\sigma_{112} + \sigma_{122}), \\ 2, & \text{if } (\sigma_{211} + \sigma_{221}) \leq (\sigma_{112} + \sigma_{122}). \end{cases} \quad (3.5)$$

Here $(\sigma_{211} + \sigma_{221})$ is the sum of the weights of the couple of matrix components in the second level $[C_{211}]$ and $[C_{221}]$, and $(\sigma_{112} + \sigma_{122})$ —the sum of the weights for the couple of matrices $[C_{112}]$ and $[C_{122}]$. In the case when the ET represents a sequence of two images of size 2×2 , its elements (pixels) have non-negative values, and hence, its rank is $R = 2$. The proposed adaptive 3D-THSVD algorithm is based on the decomposition of elementary tensors (ET) of size 222.

3.2.2 Decomposition of a Tensor of Size $N \times N \times N$ Through 3D Truncated Hierarchical SVD

The procedure explanation starts here with the decomposition of a tensor of size $4 \times 4 \times 4$ based on the 3D-THSVD, and after that it is generalized for a tensor of size $N \times N \times N$, where $(N = 2^n)$.

In the first level of the 3D-THSVD $_{4 \times 4 \times 4}$, the tensor $T_{4 \times 4 \times 4}$ (for $N = 4$) as shown in Fig. 3.2 a is divided into eight ETs $T_{i(2 \times 2 \times 2)}$ for $i = 1, 2, \dots, 8$ (cubes of size $2 \times 2 \times 2$). Figure 3.2b shows the front cut of the tensor $T_{4 \times 4 \times 4}$ in the first level of 3D-THSVD $_{4 \times 4 \times 4}$. Each ET is then decomposed into 2 new tensors $T_{ij(2 \times 2 \times 2)}$ for $j = 1, 2$, and of same size. In the first level of 3D-THSVD $_{4 \times 4 \times 4}$ for each group, composed of 8 elements of same color (yellow, red, green, blue, white, purple, light blue, or orange), is calculated the two-level adaptive 3D-THSVD $_{2 \times 2 \times 2}$. After the rearrangement of ETs and their reshaping are obtained two new tensors $T_{j(4 \times 4 \times 4)}$, shown in Fig. 3.3a in pink and light blue.

In the *second level* of the 3D-THSVD $_{4 \times 4 \times 4}$, each of the two tensors $T_{j(4 \times 4 \times 4)}$ is divided into 8 sub-tensors $T_{j,k(2 \times 2 \times 2)}$ for $k = 1, 2$ in the way, defined by the spatial net

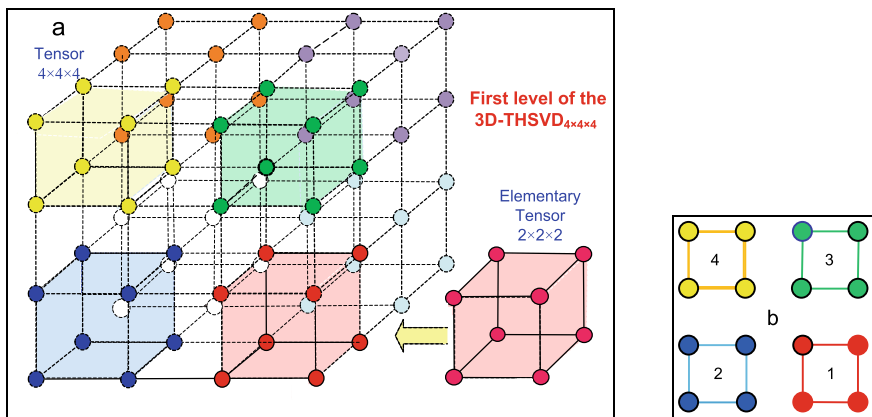


Fig. 3.2 Tensor $T_{4 \times 4 \times 4}$, divided into eight ETs $T_{i(2 \times 2 \times 2)}$ for $i = 1, 2, \dots, 8$ in the first 3D-THSVD $_{4 \times 4 \times 4}$ level where 3D-THSVD $_{2 \times 2 \times 2}$ is applied on each group of elements of same color

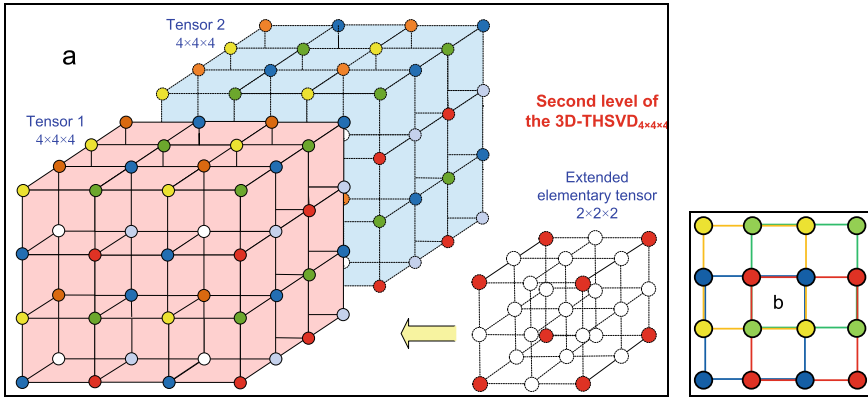


Fig. 3.3 Division of tensors $T_{j(4 \times 4 \times 4)}$ into extended ETs $T_{i,j(2 \times 2 \times 2)}$ in the second 3D-THSVD $_{4 \times 4 \times 4}$ level where the 3D-THSVD $_{2 \times 2 \times 2}$ is applied on each group of elements of same color (8 in total)

for elements interlacing, as shown in Fig. 3.3b for one front cut of the tensor $T_{1(4 \times 4 \times 4)}$. The color of the elements in each cube corresponds to that from the first level of the 3D-THSVD $_{4 \times 4 \times 4}$ algorithm. For each expanded elementary tensor (double-size tensor), shown in Fig. 3.3, is calculated once again the 3D-THSVD $_{2 \times 2 \times 2}$. After execution of the 1st decomposition level, the tensor $T_{4 \times 4 \times 4}$ is approximated by the two-component sum below:

$$T_{4 \times 4 \times 4} \approx \sum_{j=1}^2 T_{j(4 \times 4 \times 4)} \quad (3.6)$$

The so calculated two tensors $T_{j(4 \times 4 \times 4)}$ are arranged in correspondence with the mean singular values lessening for the elementary tensors $T_{j,k(2 \times 2 \times 2)}$ when $j = 1, 2$ and $k = 1, 2, \dots, 8$. The tensors $T_{j(4 \times 4 \times 4)}$ are rearranged in accordance with the energy decrease of the ETs $T_{j,k(2 \times 2 \times 2)}$, which build them. In accordance with Fig. 3.3, on each ET is applied 3D-THSVD $_{2 \times 2 \times 2}$ again. After the execution of the second decomposition level, the tensor $T_{4 \times 4 \times 4}$ is approximated by the sum of four third-order tensors:

$$T_{4 \times 4 \times 4} \approx \sum_{j_1=1}^2 \sum_{j_2=1}^2 T_{j_1, j_2(4 \times 4 \times 4)} \quad (3.7)$$

The computational graph of the two-level 3D-THSVD $_{4 \times 4 \times 4}$ decomposition is shown in Fig. 3.4. In the initial decomposition level for each ET is executed unfolding mode-1 and in the next level—unfolding mode-2. The so calculated two tensors $T_{j(4 \times 4 \times 4)}$ are arranged in accordance with the decreasing of the singular values of kernels $T_{j,k(2 \times 2 \times 2)}$, which compose them, for $j = 1, 2$ and $k = 1, 2, \dots, 8$. If the

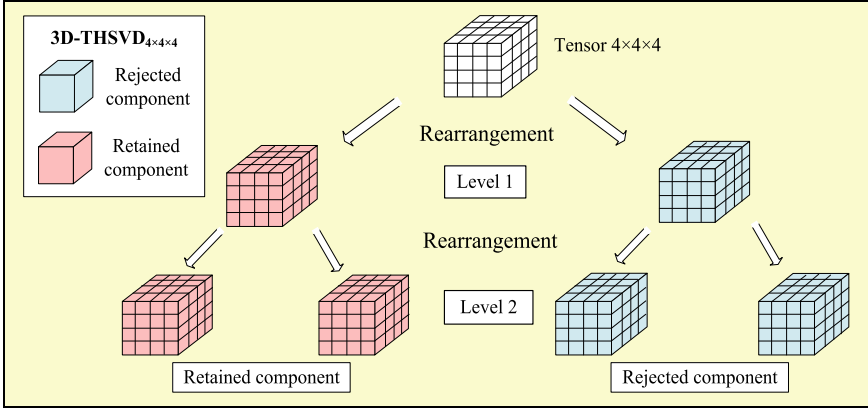


Fig. 3.4 Structures of the full and truncated 2-level binary tree computational graphs for the 2-level 3D-THSVD₄₄₄ based on the 3D-HSVD_{2x2x2}

possible truncations are not taken into account for the low-energy branches of the computational graph (colored in blue in Fig. 3.4) in the first level of the full 3D-THSVD₄₄₄, the SVD₂₂ is executed 32 times, and in the second level—64 times.

3.2.3 Calculation of 3D-THSVD for Third-Order Tensor of Size $N \times N \times N$

The decomposition of tensors $T_{4 \times 4 \times 4}$ could be generalized for the case when the tensor $T_{N \times N \times N}$ is of size $N \times N \times N$, for $N = 2^n$. As a result of the use of the 3D-THSVD _{$N \times N \times N$} full algorithm and for the pre-defined approximation accuracy, the tensor $T_{2^n \times 2^n \times 2^n}$ is represented as a sum of 2^n eigen-tensors [19]:

$$T_{2^n \times 2^n \times 2^n} \approx \sum_{j_1=1}^2 \sum_{j_2=1}^2 \cdots \sum_{j_n=1}^2 T_{j_1, j_2, \dots, j_n(2^n \times 2^n \times 2^n)}. \quad (3.8)$$

The eigen-tensors $T_{j_1, j_2, \dots, j_n(2^n \times 2^n \times 2^n)}$ of size $2^n \times 2^n \times 2^n$ are arranged in accordance with tensors energies decreasing. The maximum number of hierarchical levels needed for the execution of the 3D-THSVD _{$N \times N \times N$} for $N = 2^n$, is n . The number of retained eigen-tensors in the decomposition is reduced at least two times, compared to 3D-HSVD _{$N \times N \times N$} , and in the last level this number is 2^n .

In case that the rank of the tensor $T_{2^n \times 2^n \times 2^n}$ is $R = 2^r \leq N = 2^n$, Eq. (3.8) becomes:

$$T_{2^n \times 2^n \times 2^n} \approx \sum_{j_1=1}^2 \sum_{j_2=1}^2 \cdots \sum_{j_r=1}^2 T_{j_1, j_2, \dots, j_r(2^n \times 2^n \times 2^n)} \text{ for } r = \log_2 R. \quad (3.9)$$

In this case, the decomposition comprises R eigen-tensors. In the first level of 3D-THSVD $_{2 \times 2 \times 2}$ for each ET is applied “unfolding mode-1,” as shown in Fig. 3.4. The next levels of 3D-THSVD $_{N \times N \times N}$ are alternatively applied “unfolding mode-1, 2, 3” for level 1 of ET. As a result, the elements of the decomposition tensors are mutually decorrelated in three orthogonal directions.

The 3D-THSVD algorithm does not require iterative calculations. The number of operations “addition” and “multiplication” $O(*)$ needed to calculate the SVD $_{2 \times 2}$ in correspondence with Eq. (3.1) is $O_{SVD}(2 \times 2) = 40$.

From Fig. 3.4, it follows that $O_{3D-TSVD}(2 \times 2 \times 2) = 40.4 = 160$. The computational complexity of the full 3D-THSVD algorithm for a tensor of size $2^n \times 2^n \times 2^n$, represented by the number of needed arithmetic operations, could be evaluated in accordance with the relation:

$$\begin{aligned} O_{3D-THSVD}(2^n \times 2^n \times 2^n) &= 2^{n-1} 8 \left(1 + \sum_{i=1}^{n-1} 2^i \right) \times O_{3D-THSVD}(2 \times 2 \times 2) \\ &= 2^{n+2} (2^{n-1} - 1) \times 160 \approx O(2^{2n+1}). \end{aligned} \quad (3.10)$$

According to Eqs. (3.4) and (3.9), ETs decomposition is repeatedly executed in each hierarchical level. This permits to implement the decomposition by using similar sets of calculations, executed in parallel.

The advantage of 3D-THSVD is the need of multiple calculation of SVD $_{2 \times 2}$ in each hierarchical level, instead of iterative calculations. This permits parallel implementation of 3D-THSVD based on processors of relatively simple structure designed to calculate the SVD $_{2 \times 2}$. However, the high computation speed of this decomposition does not ensure lossless (full) restoration of the original tensor \mathbf{X} , due to the truncation of all components whose sequential number is larger than the tensor rank, R .

The pseudocode of the adaptive 3D-THSVD algorithm for the n -level decomposition of third-order tensor of size $N \times N \times N$ is given below.

Algorithm 1: Adaptive 3D-THSVD for Tensor of Size $N \times N \times N$

Input: Third-order tensor \mathbf{X} of size $N \times N \times N$ ($N=2^n$) with elements $x(i,j,k)$, accuracy ε (a small constant), and $d=0$.

Output: Adaptive 3D-THSVD for n-level decomposition of the tensor \mathbf{X}

- 1 **begin**
- 2 Divide the tensor \mathbf{X} into $R=N^3/2^3$ sub-tensors \mathbf{X}_r of size $2 \times 2 \times 2$ (ET) and distances of size 2^d between the neighbor elements;
- 3 **for** $r=1,2,\dots,R$ in the level $p=1$ of the local 3D-THSVD $_{r,2 \times 2 \times 2}$ for $\mathbf{X}_{r,2 \times 2 \times 2}$ **do**
- 4 Execute unfolding mode-1 for the tensor $\mathbf{X}_{r,2 \times 2 \times 2}$:
 $\text{unfold}_{\text{mode-1}}(\mathbf{X}_{r,2 \times 2 \times 2}) = [[\mathbf{X}_{r,1}] \quad [\mathbf{X}_{r,2}]]$
 On each matrix $[\mathbf{X}_{r,1}]$ and $[\mathbf{X}_{r,2}]$ apply SVD $_{2 \times 2}$ in accordance with Eq. (1):
 $[\mathbf{X}_{r,1}] = [\mathbf{C}_{r,11}] + [\mathbf{C}_{r,21}]$, $[\mathbf{X}_{r,2}] = [\mathbf{C}_{r,12}] + [\mathbf{C}_{r,22}]$
- 5 Execute rearrangement 1 of matrices $[\mathbf{C}_{r,i,j}]$ for $i,j=1,2$ in accordance with their weights: $\sigma_{11}^2 \geq \sigma_{12}^2 \geq \sigma_{21}^2 \geq \sigma_{22}^2$.
- 6 Execute folding mode-1 for tensors $\mathbf{X}_{r,1(2 \times 2 \times 2)}$ and $\mathbf{X}_{r,2(2 \times 2 \times 2)}$:
 $\mathbf{X}_{r,1(2 \times 2 \times 2)} \Rightarrow \text{fold}_{\text{mode-1}}([\mathbf{C}_{r,11}], [\mathbf{C}_{r,21}])$; $\mathbf{X}_{r,2(2 \times 2 \times 2)} \Rightarrow \text{fold}_{\text{mode-1}}([\mathbf{C}_{r,12}], [\mathbf{C}_{r,22}])$;
- 7 Reconstruct the tensor $\mathbf{X}_{r,2 \times 2 \times 2}$ on the basis of $\mathbf{X}_{r,1(2 \times 2 \times 2)}$ and $\mathbf{X}_{r,2(2 \times 2 \times 2)}$:
 $\mathbf{X}_{r,2 \times 2 \times 2} = \mathbf{X}_{r,1(2 \times 2 \times 2)} + \mathbf{X}_{r,2(2 \times 2 \times 2)}$
- 7 Execute unfolding mode-2 for tensors $\mathbf{X}_{r,1(2 \times 2 \times 2)}$ and $\mathbf{X}_{r,2(2 \times 2 \times 2)}$:
 $\text{unfold}_{\text{mode-2}}(\mathbf{X}_{r,1(2 \times 2 \times 2)}) = [[\mathbf{C}_{r,11}] \quad [\mathbf{C}_{r,21}]]$, $\text{unfold}_{\text{mode-2}}(\mathbf{X}_{r,2(2 \times 2 \times 2)}) = [[\mathbf{C}_{r,12}] \quad [\mathbf{C}_{r,22}]]$
- 8 In the level $p=2$ of the 3D-THSVD $_{r,2 \times 2 \times 2}$ on each matrix $[\mathbf{C}_{r,i,j}]$ for $i,j=1,2$ apply SVD $_{2 \times 2}$:
 $[\mathbf{C}_{r,11}] = [\mathbf{C}_{r,111}] + [\mathbf{C}_{r,112}]$, $[\mathbf{C}_{r,21}] = [\mathbf{C}_{r,211}] + [\mathbf{C}_{r,212}]$,
 $[\mathbf{C}_{r,12}] = [\mathbf{C}_{r,121}] + [\mathbf{C}_{r,122}]$, $[\mathbf{C}_{r,22}] = [\mathbf{C}_{r,221}] + [\mathbf{C}_{r,222}]$.
- 9 Execute folding mode-2 or 1 of the tensors $\mathbf{X}_{r,1(2 \times 2 \times 2)}$, $\mathbf{X}_{r,2(2 \times 2 \times 2)}$, $\mathbf{X}_{r,3(2 \times 2 \times 2)}$, $\mathbf{X}_{r,4(2 \times 2 \times 2)}$:
 $\mathbf{X}_{r,1(2 \times 2 \times 2)} \Rightarrow \text{fold}_{\text{mode-2}}([\mathbf{C}_{r,111}], [\mathbf{C}_{r,112}])$;
 $\mathbf{X}_{r,2(2 \times 2 \times 2)} \Rightarrow \begin{cases} \text{fold}_{\text{mode-1}}([\mathbf{C}_{r,211}], [\mathbf{C}_{r,221}]), & \text{if } (\sigma_{r,211} + \sigma_{r,221}) > (\sigma_{r,112} + \sigma_{r,122}) \\ \text{fold}_{\text{mode-2}}([\mathbf{C}_{r,112}], [\mathbf{C}_{r,122}]), & \text{if } (\sigma_{r,211} + \sigma_{r,221}) \leq (\sigma_{r,112} + \sigma_{r,122}) \end{cases}$;
 $\mathbf{X}_{r,3(2 \times 2 \times 2)} \Rightarrow \begin{cases} \text{fold}_{\text{mode-1}}([\mathbf{C}_{r,112}], [\mathbf{C}_{r,122}]), & \text{if } (\sigma_{r,211} + \sigma_{r,221}) > (\sigma_{r,112} + \sigma_{r,122}) \\ \text{fold}_{\text{mode-2}}([\mathbf{C}_{r,211}], [\mathbf{C}_{r,221}]), & \text{if } (\sigma_{r,211} + \sigma_{r,221}) \leq (\sigma_{r,112} + \sigma_{r,122}) \end{cases}$;
 $\mathbf{X}_{r,4(2 \times 2 \times 2)} \Rightarrow \text{fold}_{\text{mode-2}}([\mathbf{C}_{r,212}], [\mathbf{C}_{r,222}])$.
- 10 Reconstruct the tensor $\mathbf{X}_{r,2 \times 2 \times 2} = \sum_{s=1}^4 \mathbf{X}_{r,s(2 \times 2 \times 2)}$

- 11 Calculate the truncated tensor $\hat{X}_{r,2 \times 2 \times 2}$:
- $$\hat{X}_{r,2 \times 2 \times 2} = \sum_{s=1}^2 X_{r,s(2 \times 2 \times 2)}, \text{ if } \left\| \sum_{s=3}^4 X_{r,s(2 \times 2 \times 2)} \right\|_F \leq \varepsilon$$
- 12 Reconstruct the tensors $X_s = \bigcup_{r=1}^R X_{r,s(2 \times 2 \times 2)}$ for $s=1,2,3,4$ of size $N \times N \times N$.
- 13 **end**
- 14 Enlarge distances between the neighbor ET elements, $X_{r,s,2 \times 2 \times 2}$, which comprise 8 elements, up to 2^d for $d \rightarrow d+1$;
- 15 If $d=n$, **go to** step 18, **else go to** Step 16;
- 16 Divide each tensor X_s for $s=1,2,\dots,4^d$ into $R=N^3/2^3$ Elementary tensors.
- 17 Replace $X_{r,2 \times 2 \times 2} \rightarrow X_{r,s,2 \times 2 \times 2}$ and **go back** to Step 3.
- 18 **end**

Result of the algorithm execution without truncation is obtained:

$$X_{2^n \times 2^n \times 2^n} = \sum_{s=1}^{2^n} \bigcup_{r=1}^R X_{r,s(2 \times 2 \times 2)} \quad (3.11)$$

3.3 Tensor Decomposition Based on the 3D Reduced Inverse Spectrum Pyramid

The second group of hierarchical decompositions for 3D images is the deterministic, which have low CC but do not ensure full decorrelation. The tensor decomposition, called reduced inverse spectrum pyramid (3D-RISP) [17], was presented for the case when the tensor X is of size 888 ($N = 2^n$ for $n = 3$). In this case, the inverse spectrum pyramid is of 3 hierarchical levels. For the decomposition implementation, in each level could be applied each of the well-known 3D orthogonal transforms: 3D-FFT, 3D-DCT, 3D-WHT, etc. Here the 3D-WHT was chosen because of its low CC. In this case, the three-level 3D-RISP pyramid based on the truncated 3D-WHT (3D-TWHT) for levels $p = 0, 1, 2$ is defined by the relation below:

$$X = \tilde{X} + \tilde{E}_0 + E_1 \quad (3.12)$$

where

$$\tilde{X} = (1/8^3) \sum_{u=0}^1 \sum_{v=0}^1 \sum_{l=0}^1 s(u, v, l) W_{u,v,l}; E_0 = X - \tilde{X}; \tilde{E}_0 = \bigcup_{t=1}^8 \tilde{E}_0^t;$$

$$\begin{aligned} \tilde{E}_0^t &= (1/4^3) \sum_{u=0}^1 \sum_{v=0}^1 \sum_{l=0}^1 s_0^t(u, v, l) W_{u,v,l} \text{ for } t = 1, 2, \dots, 8, \\ E_1^t &= E_0^t - \tilde{E}_0^t \text{ for } t = 1, 2, \dots, 8; E_1 = \bigcup_{t=1}^8 E_1^t. \end{aligned} \quad (3.13)$$

The tensors tagged with a small wave, approximate these without a wave, with E are tagged the difference tensors, and with $W_{u,v,l}$ —the 3D basic Walsh–Hadamard functions. The coefficients of the 3D-WHT in levels $p = 0, 1, 2$ are defined by the relations:

$$s(u, v, l) = \sum_{i=0}^7 \sum_{j=0}^7 \sum_{k=0}^7 x(i, j, k) \text{wal}(i, u, 8) \text{wal}(j, v, 8) \text{wal}(k, l, 8) \text{ for } p = 0; \quad (3.14)$$

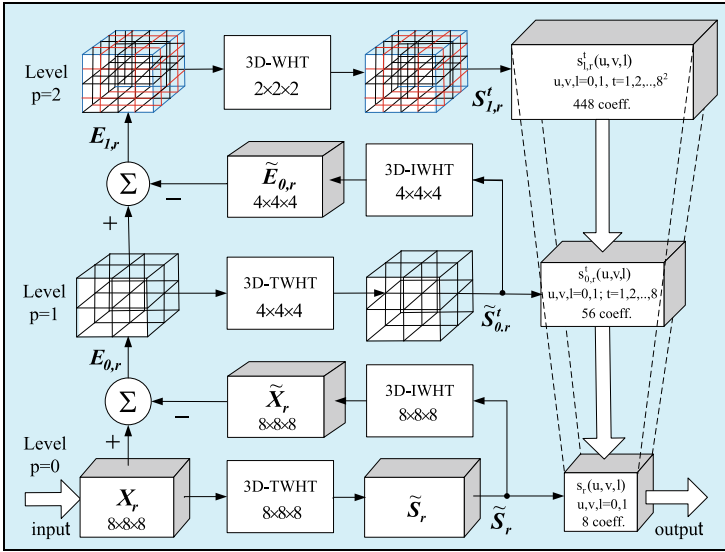
$$\begin{aligned} s_p^t(u, v, l) \\ = \sum_{i=0}^{2^{3-p}-1} \sum_{j=0}^{2^{3-p}-1} \sum_{k=0}^{2^{3-p}-1} \tilde{e}_p^t(i, j, k) \text{wal}(i, u, 2^{3-p}) \text{wal}(j, v, 2^{3-p}) \text{wal}(k, l, 2^{3-p}) \end{aligned} \quad (3.15)$$

for $t = 1, 2, \dots, 8^{p+1}$ and $p = 1, 2$.

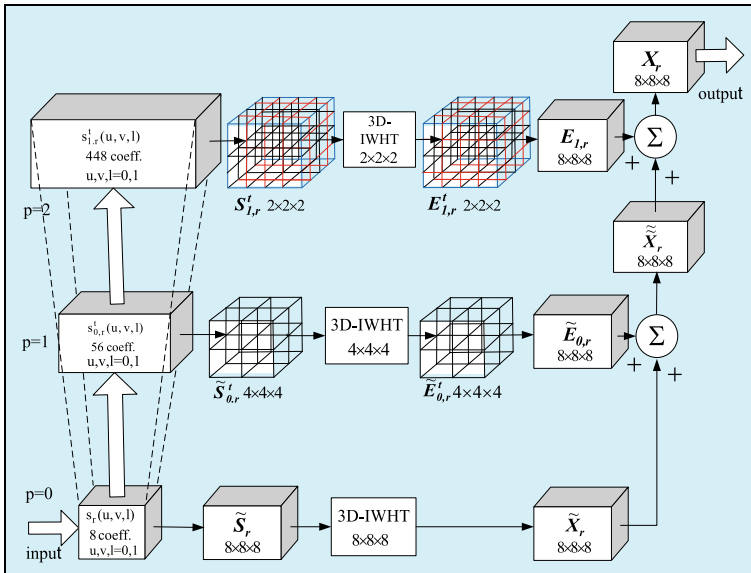
Here $x(i, j, k)$ and $\tilde{e}_p^t(i, j, k)$ are correspondingly the elements of the tensor X and of the difference tensors \tilde{E}_p^t for $t = 1, 2, \dots, 8^{p+1}$ and $p = 1, 2$; $i(\cdot)$ is one-dimensional basic Walsh–Hadamard function which first argument is its discrete value in the horizontal or vertical direction; the second argument—the spatial frequency in horizontal or vertical direction respectively, and the third argument—the size of the function.

After the truncation of coefficients $s_p^t(0, 0, 0) = 0$ in levels $p = 1, 2$, is got the 3D Reduced ISP (3D-RISP) [18]. The block diagrams for tensor decomposition/reconstruction through three-level 3D-RISP/TWHT are shown in Fig. 3.5.

In result of the truncated 3D-WHT execution, in the level $p = 0$ are retained 8 coefficients only $s(u, v, l)$ for $u, v, l = 0, 1$, while in levels $p = 1, 2$ are retained coefficients $s_p^t(u, v, l)$, for $u, v, l = 0, 1$ and $t = 1, 2, \dots, 8^{p+1}$ except coefficients $s_p^t(0, 0, 0)$ which are always equal to zero. For this reason, the 3D-RISP decomposition is not “overcomplete.” As it is seen from Fig. 3.5, the total number of spectrum coefficients $s(u, v, l)$ in all pyramid levels ($8 + 56 + 448$) is equal to the number of elements of the tensor X , of size $8 \times 8 \times 8 = 512$. To accelerate the calculation of the retained coefficients, in each RISP level is executed the fast direct and inverse 3D-TWHT [18]. Due to transform separability, it is executed in three consecutive steps—first for the vectors (fibers) of the tensor X in the horizontal direction, after that—for the



a 3D-RISP for the decomposition of tensor $X_{8 \times 8 \times 8}$



b 3D-RISP for the reconstruction of tensor $X_{8 \times 8 \times 8}$

Fig. 3.5 Block diagrams for tensor decomposition/reconstruction through 3D-RISP/TWHT

vectors of the transformed tensor in the vertical direction, and at the end—in the direction z —for the vectors of the tensor which was transformed in the second step. To execute the direct transform, for each vector on the selected direction is applied one-dimensional fast Walsh–Hadamard transform (1D-FTWHT). Before the inverse transform, the “truncated” coefficients of each spectrum vector are replaced by zeros and then is executed the one-dimensional inverse fast WHT (1D-IFWHT).

In case that the tensor \mathbf{X} is of size larger than $N \times N \times N$, it must be divided into R sub-tensors \mathbf{X}_r of size $8 \times 8 \times 8$, for $r = 1, 2, \dots, R$ and then for each to be executed the already presented pyramidal decomposition. After that, the related sub-tensors of size $8 \times 8 \times 8$ are to be combined into the corresponding tensors of size $N \times N \times N$.

The decomposition of the third-order tensor \mathbf{X} calculated through 3D-RISP/WHT of 3 hierarchical levels ($m = 3$) is given in the **Pseudocode of Algorithm 2**. After its execution is got the full octotree of 3 levels which contains the coefficients of the 3D-WHT. The coefficients which have zero values in the pyramid levels $p = 1, 2$ are cutoff and in result is got the reduced octotree which corresponds to the 3D-RIDP/TWHT decomposition of the third-order tensor.

Algorithm 2: 3D Reduced Inverse Spectrum Pyramid based on the 3D-TWHT

Input: Third-order tensor \mathbf{X} of size $N \times N \times N$ ($N=2^n$) and with elements $x(i, j, k)$

Output: Three level 3D-RISP/TWHT with $8R$ spectrum coefficients in the level $p=0$; $56R$ coefficients in the level $p=1$, and $446R$ coefficients in the level $p=2$ (R – the number of sub-tensors of size $8 \times 8 \times 8$, which build the tensor \mathbf{X}).

1 **begin**

2 Divide the tensor \mathbf{X} into $R=N^3/8^3$ sub-tensors \mathbf{X}_r of size $8 \times 8 \times 8$ for $r=1, 2, \dots, R$.

3 **for** $r=1$ to R in the level $p=0$ of the local 3D-RISP(r) **do**

4 Calculate the direct 3D-TWHT for $N=8$ through the elements $x_r(i, j, k)$ of the sub-tensor \mathbf{X}_r :

$$s_r(u, v, l) = \sum_{i=0}^7 \sum_{j=0}^7 \sum_{k=0}^7 x_r(i, j, k) (-1)^{\sum_{r=0}^2 [q_r(i)u_r + q_r(j)v_r + q_r(k)l_r]} \quad \text{for } u, v, k=0, 1 \text{ and}$$

$$i = \sum_{r=0}^2 i_r 2^r, u = \sum_{r=0}^2 u_r 2^r; \quad q_0(i) = i_2, \quad q_1(i) = i_2 \oplus i_1, \quad q_2(i) = i_j \oplus i_0;$$

$$j = \sum_{r=0}^2 j_r 2^r, v = \sum_{r=0}^2 v_r 2^r; \quad q_0(j) = j_2, \quad q_1(j) = j_2 \oplus j_1, \quad q_2(j) = j_j \oplus j_0;$$

$$k = \sum_{r=0}^2 k_r 2^r, l = \sum_{r=0}^2 l_r 2^r; \quad q_0(k) = k_2, \quad q_1(k) = k_2 \oplus k_1, \quad q_2(k) = k_l \oplus k_0.$$

5 Calculate the inverse 3D-WHT for $N=8$ through the coefficients $s_r(u, v, l)$:

$$\tilde{x}_r(i, j, k) = (1/8^3) \sum_{u=0}^2 \sum_{v=0}^2 \sum_{l=0}^2 s_r(u, v, l) (-1)^{\sum_{r=0}^2 [q_r(i)u_r + q_r(j)v_r + q_r(k)l_r]}$$

6 Calculate the elements $e_{0,r}(i, j, k)$ of zero-difference sub-tensor $\mathbf{E}_{0,r} = \mathbf{X}_r - \tilde{\mathbf{X}}_r$

7 **end**

8 Divide each zero-difference sub-tensor $\mathbf{E}_{0,r}$ into 8 cubes of size $4 \times 4 \times 4$.

9 **for** $r=1$ to R in the level $p=1$ of the local 3D-RISP(r) **do**

10 Calculate the direct 3D-TWHT for $N=4$ through the elements $e_{0,r}^t(i, j, k)$ of each sub-tensor $\mathbf{E}_{0,r}^t$ of size $4 \times 4 \times 4$ for $t=1, 2, \dots, 8^2$:

$$s_{0,r}^t(u, v, l) = \sum_{i=0}^3 \sum_{j=0}^3 \sum_{k=0}^3 e_{0,r}^t(i, j, k) (-1)^{\sum_{r=0}^l [q_r(i)u_r + q_r(j)v_r + q_r(k)l_r]} \quad \text{for } u, v, k=0, 1 \text{ and}$$

$$i = \sum_{r=0}^l i_r 2^r, u = \sum_{r=0}^l u_r 2^r; \quad j = \sum_{r=0}^l j_r 2^r, v = \sum_{r=0}^l v_r 2^r; \quad k = \sum_{r=0}^l k_r 2^r, l = \sum_{r=0}^l l_r 2^r;$$

$$q_0(i) = i_1, \quad q_1(i) = i_j \oplus i_0; \quad q_0(j) = j_1, \quad q_1(j) = j_l \oplus j_0; \quad q_0(k) = k_1, \quad q_1(k) = k_l \oplus k_0.$$

11 Calculate the inverse 3D-WHT for $N=4$ through the coefficients $s_{0,r}^t(u, v, l)$:

$$\tilde{e}_{0,r}^t(i, j, k) = (1/4^3) \sum_{u=0}^1 \sum_{v=0}^1 \sum_{l=0}^1 s_{0,r}^t(u, v, l) (-1)^{\sum_{r=0}^l [q_r(i)u_r + q_r(j)v_r + q_r(k)l_r]}$$

12 Calculate the elements $e_{1,r}^t(i, j, l)$ of second-difference sub-tensor $\mathbf{E}_{1,r}^t = \mathbf{E}_{0,r}^t - \tilde{\mathbf{E}}_{0,r}^t$

13 **end**

14 Divide each second-difference sub-tensor $\mathbf{E}_{1,r}^t$ into 8 cubes of size $2 \times 2 \times 2$.

15 **for** $r=1$ to R in the level $p=2$ of the 3D-RISP(r) **do**

16 Calculate the direct 3D-TWHT for $N=2$ based on the elements $e_{1,r}^t(i, j, k)$ for $t=1, 2, \dots, 8^3$ of each sub-tensor $\mathbf{E}_{1,r}^t$:

$$s_{1,r}^t(u, v, l) = \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^1 e_{1,r}^t(i, j, k) (-1)^{\sum_{r=0}^l [q_r(i)u_r + q_r(j)v_r + q_r(k)l_r]}$$

17 **end**

18 Remove the coefficient $s_{0,r}^l(0, 0, 0)$ in the level $p=1$ from the group of coefficients

$s'_{0,r}(0,0,0)$ for $t=1,2,\dots,8^2$ and $r=1,2,\dots,R$ (reduce the number of coefficients from 8 to 7) in accordance with the relation:

$$s'_{0,r}(0,0,0) = -\sum_{t=2}^8 s'_{0,r}(0,0,0).$$

The total number of the spectrum coefficients is reduced from $64R$ to $56R$.

- 19 Remove the coefficient $s'_{l,r}(0,0,0)$ in the level $p=2$ from the group of coefficients $s'_{l,r}(0,0,0)$ for $t=1,2,\dots,8^3$ and $r=1,2,\dots,R$ in accordance with the relation:

$$s'_{l,r}(0,0,0) = -\sum_{t=2}^8 s'_{l,r}(0,0,0).$$

In result, the total number of the spectrum coefficients is reduced from $512R$ to $446R$.

- 20 Gather the spectrum coefficients from each level $p=0,1,2$ of the local 3D-IDP(r) for $r=1,2,\dots,R$ and build of the global 3D-RISP/WHT of 3 levels.
21 **end**
-

3.4 Comparison of the CC of the Famous Deterministic and Statistical Hierarchical 3D Decompositions

3.4.1 CC of the Deterministic Hierarchical 3D Decompositions: RISP, SDT, FFT and DWT

The CC of the n -level 3D-RISP/3D-FTWHT is defined by the number of needed additions/multiplications [19]:

$$A_{3D-RISP}(n) = \sum_{p=0}^{n-1} A_p = 2^{3n} \{(n-1)[(5/2)n + 7] + 4\}; \quad (3.16)$$

$$M_{3D-RISP}(n) = \sum_{p=0}^{n-1} M_p = (1/7) [2^{3(n-1)} - 1]. \quad (3.17)$$

Then, the number of operations O (additions and multiplications) needed for the calculation of the 3D-RISP/TWHT is:

$$\begin{aligned} O_{3D-RISP/TWHT}(n) &= 2^{3n} \{(n-1)[(5/2)n + 7] + 4\} + (1/7) [2^{3(n-1)} - 1] \\ &\approx 2^{3n} \times 2.5n. \end{aligned} \quad (3.18)$$

The number of additions/multiplications which define the operations O needed for the calculation of 3D shearlet discrete transform (3D-SDT), 3D-FFT and 3D discrete wavelet transform (3D-DWT), in accordance with [19], is:

- for 3D-SDT:

$$A_{3D-SDT} = 2^{3n}(5n + 43); M_{3D-SDT} = 2^{3n} \times 48.6n; O_{3D-SDT}(n) \approx 2^{3n} \times 53.6n; \quad (3.19)$$

- for 3D-FFT:

$$A_{3D-FFT} = 2^{3n} \times 6.75n; M_{3D-FFT} = 2^{3n} \times 0.75n; O_{3D-FFT}(n) = 2^{3n} \times 7.5n; \quad (3.20)$$

- for 3D-DWT:

$$A_{3D-DWT} = 2^{3n} \times 20.5; M_{3D-DWT} = 2^{3n} \times 27.4; O_{3D-DWT}(n) = 2^{3n} \times 47.9. \quad (3.21)$$

3.4.2 *CC of the Statistical Hierarchical 3D Decompositions HSVD, THSVD, and H-Tucker*

From Eq. (3.10) used for the definition of $O(n)$ for THSVD, and from Eq. (3.49) in [12] for the HSVD, it follows that:

- for 3D-HSVD:

$$O_{3D-HSVD}(n) = 54 \times (2^{2n} - 1)2^{2n} \approx 2^{4n} \times 54; \quad (3.22)$$

- for 3D-THSVD:

$$O_{3D-THSVD}(n) = 2^{n+2}(2^{n-1} - 1) \times 160 \approx 2^{2n} \times 320. \quad (3.23)$$

The H-Tucker transform [9] was selected for the CC comparison with the analyzed 3D transforms. The needed number of operations for the H-Tucker cubical tensor decomposition of rank $R = 2^n$ (as defined in [3]), size $N = 2^n$ and order $d = 3$, is:

$$O_{HT}(n) = (3 \times 2^{3n} + 2 \times 2^{4n}) \approx 2^{2n+1} \quad (3.24)$$

3.4.3 Comparison Evaluation of the CC for the Analyzed 3D Decompositions

The CC evaluation for each decomposition is done on the basis of the corresponding normalized value of the number of operations $O^0(\cdot)$ for one voxel. In accordance with Eqs. (3.18)–(3.24), in Table 3.1 is given the evaluation of the number $O_i^0(n)$ normalized for 2^{3n} operations for the analyzed 3D decompositions.

The data in Table 3.2 were calculated on the basis of Table 3.1 for the values of $O_i^0(n)$ when $i = 1, 2, \dots, 8$ and $n = 2, 3, \dots, 8$.

Figure 3.6 shows in a graphic form the values of the function $O_i^0(n)$ in correspondence with Table 3.2. For tensors of size $N \times N \times N$ when $N = 16$, the number of levels is 4. In this case $n = 4$ and from Table 3.2, it follows that in correspondence with the larger number of normalized operations $O_i^0(n)$, the analyzed tensor decompositions are arranged as follows:

$$\begin{aligned} O_1^0(4) &= 10 \text{ for RISP}; O_2^0(4) = 214.4 \text{ for SDT}; O_3^0(4) = 30 \text{ for FFT}; \\ O_4^0(4) &= 47.9 \text{ for DWT}; O_5^0(4) = 43.2 \text{ for HSVD}; O_6^0(4) = 20 \text{ for THSVD}; \\ O_7^0(4) &= 32 \text{ for H - Tucker}; \end{aligned}$$

In accordance with the mutual correlation lessening between tensor decomposition components, their arrangement is changed: the statistical are first (H-Tucker, HSVD, THSVD) and then follow the deterministic (RISP, SDT, DWT, and FFT). The choice of the 3D decomposition must be done on the basis of a compromise between controversial requirements for the minimum number of $O_i^0(n)$, and the maximum decorrelation between the decomposition components.

3.5 Conclusions

In this work are presented two new approaches for hierarchical decomposition of a tensor of size $N \times N \times N$ for $N = 2^n$: 3D-THSVD and 3D-RISP/3D-FTWHT. Their CC, evaluated on the basis of the needed mathematical operations $O^0(n)$, is compared with that of the decompositions 3D-HSVD, H-Tucker, 3D-DWT, 3D-SDT, and 3D-FFT. The obtained results show the CC advantage of the pyramidal decomposition 3D-RISP, based on the 3D-FTWHT. There is also certain advantage of this decomposition compared to 3D-DWT and 3D-SDT in respect of the quality of the restored image sequence represented as a third-order tensor because there are no operations of the kind 3D decimation and interpolation. These qualities of the 3D-RISP decomposition open new abilities for its future application and development in such areas as the multidimensional processing and analysis of 3D images.

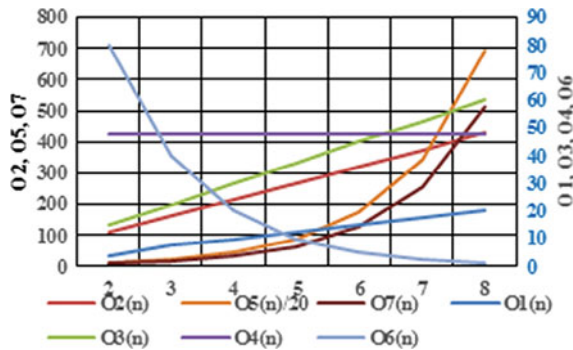
Table 3.1 Normalized number of operations $O_i^0(n)$ for the 3D decompositions

$O_{3D-RISP}^0(n)$ or $O_1^0(n)$	$O_{3D-SDT}^0(n)$ or $O_2^0(n)$	$O_{3D-FFT}^0(n)$ or $O_3^0(n)$	$O_{3D-DWT}^0(n)$ or $O_4^0(n)$	$O_{3D-HSVD}^0(n)$ or $O_5^0(n)$	$O_{3D-THSVD}^0(n)$ or $O_6^0(n)$	$O_{HT}^0(n)$ or $O_7^0(n)$
$2.5n$	$53.6n$	$7.5n$	47.9	$2^n 54$	$320/2^n$	2^{n+1}

Table. 3.2 Normalized number of operations $O_i^0(n)$ as a function of n

n	2	3	4	5	6	7	8
$O_1^0(n)$	4	7.5	10	12.5	15	17.5	20
$O_2^0(n)$	107.2	160.8	214.4	268	321.6	375.2	428.8
$O_3^0(n)$	15	22.5	30	37.5	45	52.5	60
$O_4^0(n)$	47.9	47.9	47.9	47.9	47.9	47.9	47.9
$O_5^0(n)/20$	10.8	21.6	43.2	86.4	172.8	345.6	691.2
$O_6^0(n)$	80	40	20	10	5	2.5	1.25
$O_7^0(n)$	8	16	32	64	128	256	512

Fig. 3.6 Normalized number of operations $O_i^0(n)$ for $i = 1, 2, \dots, 8$, depending on n



Acknowledgements This work was supported by the National Science Fund of Bulgaria: Bilateral Project KII-06-India-04: *Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications*.

References

1. Ji, Y., Wang, Q., Li, X., Liu, J.: A survey on tensor techniques and applications in machine learning. IEEE Access, pp. 1–39. <https://doi.org/10.1109/ACCESS.2019.2949814> (2019).
2. Cichocki, A., Lee, N., Oseledets, I., Phan, A., Zhao, Q., Mandic, D.: Tensor networks for dimensionality reduction and large-scale optimization: Part 1. Low-rank tensor decompositions. Found. Trends Mach. Learn. **9**(4/5), 249–429 (2016). <https://doi.org/https://doi.org/10.1561/22000000059>.
3. Sidiropoulos, N., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E., Faloutsos, C.: Tensor decomposition for signal processing and machine learning. IEEE Trans. Signal Process. **65**(13), 3551–3582 (2017)
4. Sakai, T., Sedukhin, S.: 3D Discrete Transforms with Cubical Data Decomposition on the IBM Blue Gene/Q. Technical Report 2013–001. Graduate School of Computer Science and Engineering, University of Aizu Tsuruga, Aizu-Wakamatsu City, Japan, pp. 1–31 <https://doi.org/https://doi.org/10.1103/PhysRevLett.111.10700> (2013).

5. Bergqvist, G., Larsson, E.: The higher-order singular value decomposition: theory and an application. *IEEE Signal Process. Mag.* **27**(3), 151–154 (2010). <https://doi.org/10.1109/MSP.2010.936030>
6. Kolda, T., Bader, B.: Tensor decompositions and applications. *SIAM Rev.* **51**(3), 455–500 (2009). <https://doi.org/10.1137/07070111X>
7. Zare, A., Ozdemir, A., Iwen, M., Aviyente, S.: Extension of PCA to Higher Order Data Structures: An Introduction to Tensors, Tensor Decompositions, and Tensor PCA. *arXiv:1803.00704v2 [eess.SP]* (2018).
8. Oseledets, I.: Tensor-train decomposition. *SIAM J. Sci. Comput.* **33**(5), 2295–2317 (2011). <https://doi.org/10.1137/090752286>
9. Grasedyck, L.: Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.* **31**(4), 2029–2054 (2010). <https://doi.org/10.1137/090764189>
10. Vannieuwenhoven, N., Vandebril, R., Meerbergen, K.: A new truncation strategy for the higher-order singular value decomposition. *SIAM J. Sci. Comput.* **34**(2), A1027–A1052 (2012). <https://doi.org/10.1137/110836067>
11. Salmi, R., Richter, A., Koivunen, V.: Sequential unfolding SVD for tensors with applications in array signal processing. *IEEE Trans. Signal Process.* **57**(12), 4719–4733 (2009). ISSN: 1053–587X.
12. Kountchev, R., Kountcheva, R.: 3D image representation through hierarchical tensor decomposition based on SVD for elementary tensor of size 222. *Int. J. WSEAS Trans. Signal Process.* **12**, 199–207 (2016)
13. Vetterli, M., Kovacevic, J.: *Wavelets and Subband Coding*. Prentice Hall, NJ (2007)
14. Candes, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. *Multiscale Model. Simul.* **5**(3), 861–899 (2006)
15. Rao, K., Kim, D., Hwang, J.: *Fast Fourier transform: Algorithms and Applications*. Springer, Heidelberg, London, New York (2010)
16. Goossens, B., Luong, H., Aelterman, J., Pizurica, A., Philips, W.: Efficient multiscale and multi-directional representation of 3D data using the 3D discrete shearlet transform. In: Papadakis, M., Ville, D., Goyal, V. (eds.) *Proceedings of SPIE*, Vol. 8138, pp. 81381Z–1–81381Z–13. Bellingham (2011). <https://hdl.handle.net/1854/LU-2015272>
17. Kountchev, R., Kountcheva, R.: Third-order tensor representation through Reduced inverse difference pyramid. In: *International Conference on Creative Business for Smart and Sustainable Growth (CREBUS'19)*, Sandanski, Bulgaria, IEEE Xplore Digital Library (2019). <https://doi.org/10.1109/CREBUS.2019.8840045>
18. Kountchev, R., Kountcheva, R.: Image sequence decomposition based on the truncated hierarchical SVD. *Int. J. Multimedia Image Process. (IJMIP)* **7**(1), 352–361 (2017). <https://doi.org/10.20533/ijmip.2042.4647.2017.0043>
19. Kountchev, R., Mironov, R., Kountcheva, R.: Complexity evaluation of tensor decomposition using 3D inverse spectrum pyramid in respect of deterministic orthogonal transforms. *Int. J. WSEAS Trans. Signal Process.* **15**, 142–148 (2019)

Chapter 4

Tracking of Domestic Animals in Thermal Videos by Tensor Decompositions



Ivo Draganov and Rumen Mironov

Abstract In this paper, we present a comparative analysis of the performance of the Tucker-ALS, CP-ALS, Tucker-ADAL, and the HoRPCA-S tensor decomposition algorithms, applied for tracking of domestic animals in video. Decomposition and full processing time, detection rate, precision, and F-measure are the evaluating parameters revealing the efficiency of each algorithm. Promising results suggest the applicability of the investigated decompositions but also demonstrate particular differences among them in terms of decomposition time and detection rate. In order to increase the detection rate of systems of parallel type employing multiple decomposition algorithms we propose a score fusion with fair voting which performs better than some of the tested algorithms alone.

4.1 Introduction

There are numerous studies on spotting wildlife species in thermographic videos with various applications such as population estimation, migration analysis, behavior investigation, and others [1–6]. Typically, these implementations use input videos from cameras with wide field of view, covering vast areas in which significant number of animals reside. Often, the movement animals do are fast and some of them vanish out from the scene while others come into it.

Ward et al. [1] propose to use predictive navigation for unmanned aerial vehicles (UAV) to detect feral animals. Heat signatures of target brutes are stored in a database easing the spotting procedure. Blob analysis is the foundation of the incorporated algorithm for feature generation and then detection. It performed fast and reliable enough, taking data from a thermal sensor with resolution of 60×80 pixels at a

I. Draganov (✉) · R. Mironov
Technical University of Sofia, 8 Kliment Ohridski Blvd, Sofia, Bulgaria
e-mail: idadranov@tu-sofia.bg

R. Mironov
e-mail: rmironov@tu-sofia.bg

distance of around 10 m from moving animals, but no results are reported for greater distances or when more animals are present.

Thermal motion imaging has been proposed to monitor the flight of bats and birds with the explicit aim of risk analysis in wind turbine fields [2]. Extraction of full flight tracks becomes possible with the discovering of video peaks and then applying background masking, followed by perceptual grouping. Bird taxonomy is one of the possible applications of this approach in a wide scale. Detection rate of 81% is reported with 17% of false positives over video database containing 184 flight tracks. No results are gathered from videos with terrestrial animals, moving slowly or occupying larger portions of the frames.

Aerial surveillance, accomplished from small elevations, is also used for koalas detection as a completely automated process, proposed by Corcoran et al. [3]. Heat signatures in this case are extracted from the input frames by the Faster-RCNN (RCNN—recurrent convolutional neural network) and YOLO (“You Only Look Once”) neural networks. The probability of correct detection is found to be no less than 68% with a precision rising up to 71%. Still, there are grounds for increase of achieved accuracy although the conditions of the experiment involve lots of obscures due to the dense vegetation in the surrounding environment.

Higher accuracy, in comparison with [3], of 77.3% achieved Oishi et al. [4] detecting wild animals, primarily swimming ducks. The relief displacement effect was employed as a simple but powerful technique when there is a stationary part from the body of an animal while rest of the parts are moving to detect it. Two frames over predetermined period of time could be used to make that kind of analysis. It is done after binarization with the Otsu algorithm and edge detection in parallel to it on the same frame in order to get overlapped objects. Although fast, the technique needs detection rate enhancement.

RCNN finds its dual application for both spotting poachers and animals in the SPOT system proposed by Bondi et al. [5]. During the test phase, after training, this particular system performed close to real time. Training was done offline on a dedicated machine using Faster-RCNN. It was connected to module for transferring labels via the AzureBasic through the Internet a connection to local machine for detection assures update of training data. There, the pre-trained now Faster-RCNN finds the bounding boxes of the objects of interest. Achieved precision is around 60% with recall of around 0.7% for animals alone.

Automatic counting of hippos from a distance 38–155 m combining UAV and thermal camera was done by Lhoest et al. [6]. They used isolines and polygonization, analyzing local maximums in the input image of size 640×480 pixels. Then polygon selection takes place based on few geometrical criteria, and from emerged parts, it could be deduced if an animal occupies the overlooked area. Deviation in accuracy from manual counting is enclosed between -9.8% and $+13.7\%$. Although the algorithm has been adapted to hippo counting, it comes closer to realistic scenario of large area screening with animals, large in size and relatively slowly moving, from images with close to the standard resolution (SD).

Another group of applications, also incorporating thermal imaging, aims assessment of the health condition of domestic animals [7–9]. Most of the systems of this

type work in highly controlled environment, some of them indoor, with a single animal occupying dominant part of the video frame during examination, making none or slow movements with small displacements.

Multiple vital signs in cattle has been tried for registering accurately enough by Jorquera-Chavez et al. [7] using thermal infrared images. Heart and respiration rate along with eye and ear-base temperatures were automatically measured with average accuracy of 92% when compared to conventional methods. Tracking of the face of a dairy cow by the Kanade–Lucas–Tomasi detector with modified features allows simultaneous measurement of selected parameters based on secondary features. The experimental setup and the videos with the close-ups of the animals do not allow for freely surveilling of multitude of species. Somewhat similar is the case with the implementation of Scoley et al. [8] where they monitor calves examining various anatomical areas from their body using thermography. Significant simplification of the experimental setup comes with the selection of only measuring temperature remotely and not trying to derive any other vital sign. Accuracy of 80.48% was achieved for the most repeatable parameter, that is the right eye temperature, but still the method does not allow for complex scene analysis and counting of multiple exemplars. Respiratory disorders associated with certain diseases in pigs, found early enough, could considerably improve the effect of treatment [9]. Measuring eye and ear-base temperature together with heart and respiratory rate could indicate in certain combination of levels the presence of such a respiratory disorder. Butterworth filtering with subsequent Fourier transform of captured images allows further the extraction of features applicable for analysis of the parameters of interest by statistical evaluation. The method could also be described as limited in covered area for general purpose monitoring.

Tensor decomposition algorithms are generalization of the singular value decomposition (SVD). SVD gets as input a matrix while tensor processing is associated with three- and higher dimensional array representations of input data. Grayscale videos, including thermographic ones, could be represented as 3D arrays, or third-order tensors, with rows and columns indexes as 2 independent variables and the sequential number of the current frame—as third independent variable. The intensity of each pixel is the dependent variable. Decomposing such tensors from video data leads to low-rank and sparse representations from which it becomes straightforward to make background subtraction [10]. The result of this operation is another video with all moving objects being segmented. Various techniques for approximation of the complete decompositions depending on application type to get reasonable execution time are developed [11–15]. Kasai [11] proposes the recursive least squares as approximation tool during the canonical decomposition over parallel factors (CP) which is used to perform subspace tracking at low ranks. Experimentation shows that this approach converges faster than other algorithms, such as GROUSE, GRASTA and others, but the residual error when processing video with stationary background varies between 10 and 40%. Support tensor machine [12] of linear type has been proposed by Biswas and Milanfar for detecting pedestrians from thermographic images. Local steering kernels prove to be more noise resistant than oriented histograms when selecting low-level features in this case and were preferred in this

implementation. At 10% false positives rate, the algorithm has miss rate of around 13%.

Tracking of multiple targets could be hard to perform when using pairwise comparison as the study of Pang et al. suggests [13]. The authors undertook high-order association with its affinity as a measure to select discriminative information for tracking. They selected the multiway histogram intersection for their implementation to track vehicles in videos from cameras with wide field of view. The correct matching rate over one of the tested databases, including some heavy traffic at moments, is around 93%. False alarms rate is considerable for another database—exceeding 3 times the real targets.

In order to overcome some of the difficulties related to occlusions, noise and varying lighting conditions Javed et al. [14] use SVD on tensor-tensor product and modified norm of a nuclear type. According to their results, this leads to better exploitation of the correlation between the outer structures of targeted objects from one position to new one. The overlap score for various tested sequences ranges between 80% and 90%.

Tucker, CP, higher-order principal component analysis (PCA) and others decompositions in their multiple versions are deeply investigated in a survey [15] with regard to various applications. One of them is video surveillance with anomaly detection based on object appearance, primarily humans, within the scene. All three are found to be promising for such kind of applications considering wider covered areas and multiple instances of movement.

In the current study, a comparative analysis is presented among 4 of the most time efficient tensor decomposition algorithms with background subtraction [10]. These are the Tucker-ALS (ALS—Alternating Least Squares), CP-ALS, Tucker-ADAL (ADAL—Alternating Direction Augmented Lagrangian), and HoRPCA-S (High-order Robust Principal Component Analysis solved by Singleton model). They process thermographic videos in order to track domestic animals recorded in different numbers and at various distances. Based on these results, the aim of the research is to lead to a scheme which will increase the accuracy. A new, score fusion algorithm, based on 3 of the tested implementations, proved to improve significantly the precision of tracking. In Sect. 4.2, the evaluated tensor decomposition algorithms are described, followed by experimental results in Sect. 4.3. In Sect. 4.4, a discussion over the results, both quantitatively and qualitatively over the test video set is presented. Then, in Sect. 4.5, a conclusion is made.

4.2 Evaluated Algorithms

4.2.1 Tucker-ALS

Tucker tensor decomposition [16] takes as input array with dimensions $I \times J \times K$ which may be a sequence of video frames and gives as output core array \mathbf{T}_c with

dimensions $R_1 \times R_2 \times R_3$. The core is all-orthogonal and decomposition relies on coordinates transformation which is orthogonal as well. In general, R_p represents mode- p rank while in the current case $p = 1, 2, 3$. The following expression holds [16]:

$$\mathbf{T} = (\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \mathbf{Q}^{(3)}) \cdot \mathbf{T}_c \Leftrightarrow \mathbf{T}_c = (\mathbf{Q}^{(1)\text{T}}, \mathbf{Q}^{(2)\text{T}}, \mathbf{Q}^{(3)\text{T}}) \cdot \mathbf{T}, \quad (4.1)$$

where $\mathbf{Q}^{(l)}$ are matrices of orthogonal type. They could be found by singular value decomposition over the unfolding matrices of \mathbf{T} , which also count in 3. Finally, the right relation from (4.1) leads to \mathbf{T}_c . Particular singular values could be put to zero when falling below certain threshold, which will truncate the components of the decomposition. In such a way proper, close to the optimal, low-rank estimate of a matrix could be calculated. Truncation is possible in every mode, and resulting ranks (r_1, r_2, r_3) are typically lower than (R_1, R_2, R_3) . As a sequence of that, the tensor rank is also reduced; but in the general case, it may not be predicted. The reason for that is the lack of proof for a generalized Eckart-Young theorem applicable to tensor computations [16].

Alternating Least Squares (ALS) is a technique to find various tensor decompositions [17], among them the Tucker one. In a preliminary stage of this algorithm, the initial values of the factor matrices are set [18]. They could be randomly selected. ALS is implemented then in iterative fashion using the tensor matrix-chain multiplication along every possible node n . The process leads to new tensor for which matricization takes place over the mode n . A possible approach into getting the product in mode n , e.g., $n = 1$ for tensor of third order is to use [18]:

$$\mathcal{Y} = ((\mathcal{X} \times_2 \mathbf{U}_2^T) \times_3 \mathbf{U}_3^T), \quad (4.2)$$

where \mathcal{X} is the input tensor, \mathcal{Y} —the product over the n -mode, \mathbf{U}_2 and \mathbf{U}_3 —are the factor matrices. Pseudocode of the ALS algorithm incorporating (4.2) is given in [18]. Another way of product calculation is to perform it in an element-wise fashion where \mathcal{Y} fibers are found according to [18]:

$$y_{:,j,k} = \mathcal{X} \times_2 \mathbf{U}_2(j, :) \times_3 \mathbf{U}_3(k, :). \quad (4.3)$$

In (4.3) $j = 1, \dots, J$ and $k = 1, \dots, K$ and this approach leads to change in the sequence of steps of the algorithm as shown, again as a pseudocode, in [18].

4.2.2 Tucker-ADAL

Tucker tensor decomposition could be achieved using the alternating direction augmented Lagrangian (ADAL) method. It comes as natural solution to the semidefinite programming (SDP) problem [19]:

$$\min_{X \in S^n} \langle C, X \rangle, \text{ given } A(X) = b, X \geq 0, \quad (4.4)$$

where X is unknown positive semidefinite matrix, S^n —set of symmetric matrices with dimensions $n \times n$; C —known matrix belonging to S^n , b —known real vector with m components, $A(\cdot)$ —linear map that gives $S^n \rightarrow \mathbb{R}^m$ and it relates to [19]:

$$A(X) := \langle A^{(1)}, X \rangle, \dots, \langle A^{(m)}, X \rangle. \quad (4.5)$$

The matrix X could be rearranged in a vector with its columns put one over the next in their order from left to right. Then (4.5) becomes [19]:

$$A := (\text{vec}(A^{(1)}), \dots, \text{vec}(A^{(m)}))^T \in \mathbb{R}^{m \times n^2}, \quad (4.6)$$

where $\text{vec}(\cdot)$ represents the rearrangement of a matrix into a vector. The augmented Lagrangian [19] is defined as:

$$\mathcal{L}_\mu(X, y, S) := -b^T y + \langle X, A^*(y) + S - C \rangle + \frac{1}{2\mu} \|A^*(y) + S - C\|_F^2, \quad (4.7)$$

where $\mu > 0$ and (4.7) satisfies linear constraints. The following condition needs to be met [19]:

$$\min_{y \in \mathbb{R}^m, S \in S^n} \mathcal{L}_\mu(X^k, y, S), \text{ given } S \geq 0, \quad (4.8)$$

where k is the current iteration of repeating process starting from $k = 0$; y — m -dimensional vector with real components. At $(k + 1)$ -th iteration the update of X is [19]:

$$X^{k+1} := X^k + \frac{A^*(y^{k+1}) + S^{k+1} - C}{\mu}. \quad (4.9)$$

The optimization procedure (4.8) leads to minimal augmented Lagrangian taking into account both y and S , but in practice they are considered consecutively which is not the precise solution but considered accurate enough [19]. Pseudocode of the complete set of steps for ADAL is presented in [20].

4.2.3 CP-ALS

Canonical decomposition (CANDECOMP) with analysis procedure over parallel factors (PARAFAC) or simply CP decomposition represents a tensor in the form of summation comprised by tensors of rank 1 [21]. In the case of third-order tensor which is in interest within this study due to the analysis of video and assuming R

components of decomposing the target function takes the following form [21]:

$$\min_{\hat{\mathcal{X}}} \|\mathcal{X} - \hat{\mathcal{X}}\|, \text{ when } \hat{\mathcal{X}} = \sum_{r=1}^R \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [[\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]], \quad (4.10)$$

where \mathbf{A} , \mathbf{B} , and \mathbf{C} are factor matrices and $\hat{\mathcal{X}}$ —the approximated tensor. They consist of vectors combination gathering rank 1 components, that is $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R]$ and similar expressions for \mathbf{B} and \mathbf{C} taking \mathbf{b}_r and \mathbf{c}_r . The length of each vector is normalized to 1 using the absorption vector $\boldsymbol{\lambda} \in \mathbb{R}^R$. During the ALS procedure, \mathbf{A} and \mathbf{B} are fixed in order to find \mathbf{C} , then from setting \mathbf{B} and \mathbf{C} unchanged, \mathbf{A} could be found and at last \mathbf{B} is calculated from putting the components of \mathbf{A} and \mathbf{C} to be constant. This sequence lasts as many times as some initially set criterion is satisfied. When only one matrix changes, the solution could be found using the linear least-squares algorithm. For instance, if \mathbf{A} and \mathbf{B} are fixed, the following minimization problem needs to be solved [21]:

$$\min_{\hat{\mathbf{C}}} \|\mathbf{X}_{(1)} - \hat{\mathbf{C}}(\mathbf{B} \odot \mathbf{A})^T\|_F, \quad (4.11)$$

where $\mathbf{X}_{(1)}$ is the mode 1 matricization of \mathcal{X} . The matrix norm in (4.5) is Frobenius type denoted by F . Also, $\hat{\mathbf{A}} = \mathbf{A} \cdot \text{diag}(\boldsymbol{\lambda})$. The result in optimal form comes from [21]:

$$\hat{\mathbf{C}} = \mathbf{X}_{(1)} [(\mathbf{B} \odot \mathbf{A})^T]^\dagger. \quad (4.12)$$

In its turn (4.6) could be represented also as [21]:

$$\hat{\mathbf{C}} = \mathbf{X}_{(1)} (\mathbf{B} \odot \mathbf{A}) (\mathbf{B}^T \mathbf{B} * \mathbf{A}^T \mathbf{A})^\dagger. \quad (4.13)$$

Thus, the pseudoinverse of $R \times R$ matrix instead of that of $JK \times R$ has to be calculated [21]. Normalization for each of the approximated tensors should be implemented by columns which mean that the components of the absorption vectors being equal to the norm of the components of the respective column from the approximated matrix will divide them. The complete set of steps as pseudocode of the CP-ALS algorithm could be found in [22].

4.2.4 HoRPCA-S

Principal component analysis (PCA) has certain disadvantages when processing two-way data such as the presence of outliers and errors with non-normal distribution [23]. It was extended to Robust PCA (RPCA). A sparse model, most often containing noisy elements, comes as addition to the low-rank model from the input data matrix.

The optimization is targeting [23]:

$$\min_{L,M} \{\text{rank}(L) + \lambda \|M\|_0 | L + M = X \}, \quad (4.14)$$

where X is input matrix of data, L —low-rank matrix and M —a sparse matrix taking place with its l_0 norm as a second component in (14). The same relation may be expressed in convex form [23]:

$$\min_{L,M} \{ \|L\|_* + \lambda \|M\|_1 | L + M = X \}, \quad (4.15)$$

where $*$ denotes the nuclear norm and the index 1—the l_1 norm. Retaining M as sparse during the optimization with regards to the rank of L leads to optimal result. RPCA has been extended to tensors as well, according to [23]:

$$\min_{\mathcal{L}, \mathcal{M}} \{ \text{Trank}(\mathcal{L}) + \lambda \|\mathcal{M}\|_1 | \mathcal{L} + \mathcal{M} = \mathcal{X} \}, \quad (4.16)$$

where Trank is the Tucker rank and all components in the decomposition are now tensors.

The Singleton model [24] is one of few models, which leads to the solution of this optimization problem. The algorithm is known as HoRPCA-S. It substitutes the tensor nuclear norm with the sum of all the nuclear norms of the same tensor after it has been mode- n matricized [23]:

$$\min_{\mathcal{L}, \mathcal{M}} \left\{ \sum_{i=1}^N \|L_{(n)}\|_* + \lambda \|\mathcal{M}\|_1 | \mathcal{L} + \mathcal{M} = \mathcal{X} \right\}. \quad (4.17)$$

4.2.5 Score Fusion

In order to improve the performance of a tracking system of the type investigated in the current study, we propose the use of a score fusion. It takes the scores from numerous tensor decomposition algorithms with background subtraction, running simultaneously over the same input video. Fair voting helps for taking the final decision whether a pixel from the current video frame is part of the foreground or it belongs to the background. At least half plus one of the scores must be positive in order to classify a pixel as foreground (Algorithm 1). The various *Video* variables are considered 3D arrays comprising of grayscale frames ($k = 1, \dots, K$) with pixel intensities over the rows ($i = 1, \dots, M$) and columns ($j = 1, \dots, N$).

Algorithm 1

```
program TScoreFusion (OutVideo)
  {Assuming T tensor decomposition algorithms};
```

```

var
  i, j, k, M, N, K, T, InVideo, OutVideo: Integer;
begin
  i := 1; j := 1; k := 1;
  ReadLn(T); ReadLn(InVideo);
  OutVideo = 0;
  OutVideo1 = Alg1(InVideo);
  OutVideo2 = Alg2(InVideo);
  ...
  OutVideoT = AlgT(InVideo);
  OutVideoAdd = OutVideo1 + OutVideo2 + ... + OutVideoT;
  repeat
    if OutVideoAdd(i, j, k) >= (N/2 + 1) then
      begin
        OutVideo(i, j, k) = 1;
      end;
  until (i <= M) && (j <= N) && (k <= K)
end.

```

4.3 Experimental Results

Test database consists of 4 uncompressed videos with dimensions 480×360 pixels and 2 with 640×360 pixels. All of them are in RGB color space and have 30 frames per second. The number of frames for each video is 200 resulting in duration of 6.67 s. Four of the videos contain birds and two—rabbits.

All tests are completed over IBM PC compatible desktop computer with Intel Core i5-3450 CPU with 4 cores operating at 3.1 GHz and 12 GB of RAM. The operating system is Linux Ubuntu 14.04 LTS 64-bit and the simulation environment is MATLAB R2016A. Video decompositions use tensor representations using the LRS library [10].

The evaluation methodology includes the following set of parameters as proposed in [25]. First, the detection rate DR is estimated according to:

$$DR = TP/(TP + FN), \quad (4.18)$$

where TP is the True Positives defined as the entity of pixels belonging to the foreground and correctly found as such, FN—False Negatives consisting from the pixels classified as background but are part of the foreground. Then, the Precision is calculated:

$$\text{Precision} = TP/(TP + FP), \quad (4.19)$$

where FP is the False Positives comprising from the pixels incorrectly detected as part of the foreground but belong to the background. Finally, the F-measure could be found from:

Table 4.1 Tucker-ALS results

Video	1	2	3	4	5	6
DR	0.9999	0.9998	0.9999	0.9997	0.9998	0.9998
Precision	0.9873	0.9694	0.9865	0.9765	0.9865	0.9918
F-measure	0.9936	0.9844	0.9931	0.9880	0.9931	0.9958
DT · 10 ⁻⁷ , s/px	3.40	3.54	2.09	3.19	1.87	3.11
FT · 10 ⁻⁷ , s/px	8.54	8.08	4.27	7.69	4.24	7.27

$$F - \text{measure} = 2 \times DR \times \text{Precision} / (DR + \text{Precision}). \quad (4.20)$$

The pixels which are considered correctly as part of the background form a group measured by the True Negatives (TN).

The tracking results for each of the 6 input videos using the Tucker-ALS decomposition are given in Table 4.1.

During experimentation, two time intervals have been measured—Detection Time (DT) and Full Processing Time (FT) in seconds/pixels (s/px). The first represents only the time during the actual tensor decomposition with background subtraction while the second include all input–output operations.

Tucker-ADAL and CP-ALS were the next tensor decompositions tried with results shown in Tables 4.2 and 4.3, respectively.

HoRPCA-S is the last decomposition tested over the video database (Table 4.4).

Table 4.2 Tucker-ADAL results

Video	1	2	3	4	5	6
DR	0.9999	0.9998	0.9999	0.9997	0.9998	0.9998
Precision	0.9873	0.9694	0.9865	0.9765	0.9865	0.9918
F-measure	0.9936	0.9844	0.9931	0.9880	0.9931	0.9958
DT · 10 ⁻⁵ , s/px	1.32	1.23	1.71	1.51	1.74	1.33
FT · 10 ⁻⁵ , s/px	1.37	1.27	1.73	1.55	1.77	1.37

Table 4.3 CP-ALS results

Video	1	2	3	4	5	6
DR	0.8061	0.8838	0.4522	0.8201	0.7503	0.6858
Precision	0.6795	0.8234	0.7474	0.7036	0.5376	0.7199
F-measure	0.7374	0.8526	0.5635	0.7574	0.6263	0.7024
DT · 10 ⁻⁷ s/px	7.76	5.91	2.91	3.11	1.14	5.36
FT · 10 ⁻⁷ , s/px	12.08	10.37	5.14	7.26	3.37	9.51

Table 4.4 HoRPCA-S results

Video	1	2	3	4	5	6
DR	0.5216	0.3101	0.2310	0.4895	0.2217	0.2556
Precision	0.9081	0.9183	0.9782	0.9750	0.7468	0.9575
F-measure	0.6626	0.4636	0.3738	0.6518	0.3419	0.4035
DT · 10 ⁻⁴ , s/px	2.48	3.31	0.83	2.78	0.37	1.45
FT · 10 ⁻⁴ , s/px	2.49	3.32	0.84	2.79	0.38	1.46

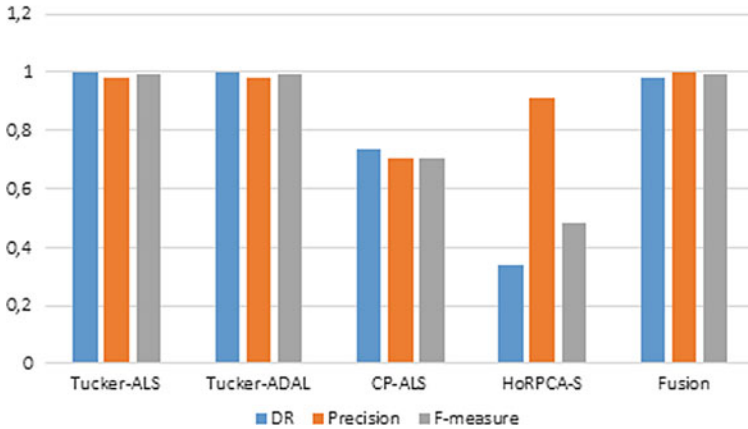


Fig. 4.1 Average DR, Precision and F-measure for all tested algorithms

The average values of DR, Precision, and F-measure from the 4 decompositions and the newly proposed score fusion are compared in Fig. 4.1, while the average DT and FT—in Fig. 4.2.

4.4 Discussion

The average detection rate, precision, and F-measure are virtually equal for the results obtained from the Tucker-ALS and Tucker-ADAL decompositions. They equal 0.9998, 0.9830, and 0.9913, respectively (Fig. 4.1), and indicate very accurate animal detection on a frame basis. The accuracy results yielded from the CP-ALS and HoRPCA-S differ considerably with an absolute offset of 0.2667, 0.2811, 0.2847, respectively, for the first compared to Tucker-ALS, and 0.6615, 0.0690, and 0.5084 for the latter. HoRPCA-S Detection Rate is significantly low, close to 34%, although the precision is high—more than 91%. In its present form, without additional pre- and post-processing of the video, this algorithm seems not applicable for the purpose at hand. CP-ALS has DR around 73% and both the precision and F-measure are just

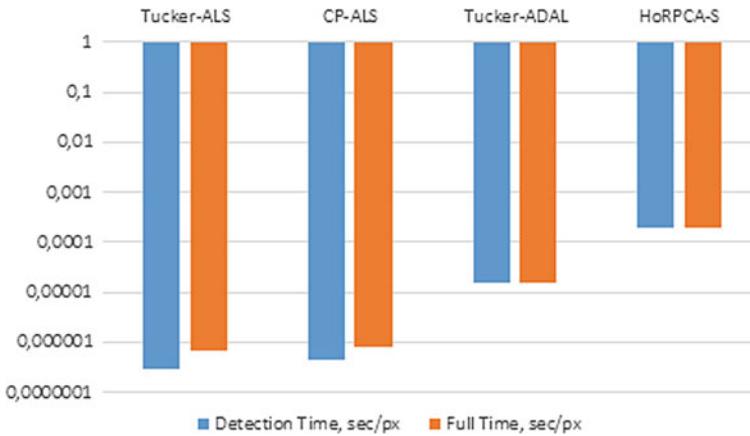


Fig. 4.2 Average DT and FT for all evaluated algorithms

over 70%. After applying the score fusion algorithm with fair voting, the resulting DR is slightly lower with comparison to Tucker-ALS—0.9831 but the precision becomes highest—0.9998 and the F-measure the same as that of Tucker-ALS and Tucker-ADAL—0.9913.

Average detection time is lowest for the Tucker-ALS algorithm—around 0.3 $\mu\text{s}/\text{px}$, very close to CP-ALS with approximately 0.4 $\mu\text{s}/\text{px}$. Two orders of a magnitude higher time are needed for the Tucker-ADAL algorithm to complete the decomposition and three orders—for the HoRPCA-S (Fig. 4.2). With the considerable increase of DT, the FT naturally is coming close to it, which makes input–output operations irrelevant to the overall time consumption of the complete application over the supporting hardware.

At visual inspection of the output videos with detected regions, partially corresponding to the animals contained in the originals, a good agreement with the numerical results from Fig. 4.1 could be observed. In Fig. 4.3.a, a frame from one of the input videos containing a single rabbit is shown.

Pixels from the detected areas, corresponding to false positives and false negatives, are being colored together in red when finding difference images with the groundtruth (Fig. 4.3b–e). Zoomed spot of the rabbit detection area with least number of falsely classified pixels comes from the Tucker-ALS algorithm (Fig. 4.3b). Non-negligible growth of the red zone around all the contour of the rabbit is registered in Fig. 4.3c, d—for CP-ALS and Tucker-ADAL, respectively. Corresponding to the increased precision, the score fusion algorithm reduces all inaccurately detected pixels along the boundary of the object of interest almost homogenously. It could turn an important quality in implementations related to animal counting in dense populations. There, distances among exemplars are small and undesirable merging could take place, leading to inaccurate results.

In Table 4.5, a comparison in terms of DR, precision and F-measure is given between the proposed in this paper score fusion approach with the approach, proposed

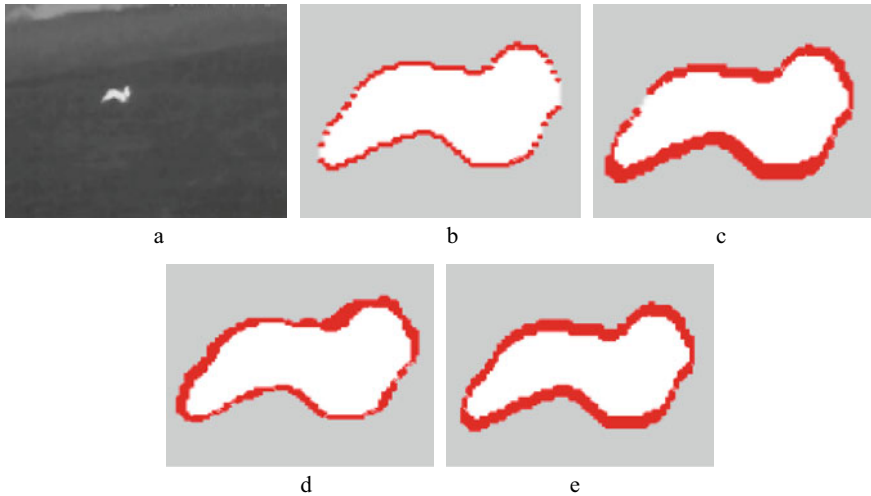


Fig. 4.3 Visual evaluation of detection results: **a**—input frame, and zoomed differences between groundtruth and the result from **b**—Tucker-ALS, **c**—Tucker-ADAL, **d**—CP-ALS, **e**—Score fusion algorithm

Table 4.5 Comparison between the proposed score fusion and the method of Corcoran et al. [3]

Method	DR	Precision	F-measure
Corcoran et al. [3]	0.8809	0.5600	0.6847
Proposed score fusion	0.9831	0.9998	0.9913

by Corcoran et al. [3] for koala detection. As the results clearly show, the proposed score fusion leads to 10% higher DR and considerably higher difference of about 43% for the Precision and 31% for the F-measure.

4.5 Conclusion

Tucker-ALS and Tucker-ADAL look promising as decomposition techniques in domestic animals tracking applications, considering the accuracy. Processing time places Tucker-ALS most suitable as a standalone algorithm of the type. Although CP-ALS is very close to Tucker-ALS in computational time, the decrease in detection rate of about 30% does not make it attractive for the purpose without any additional processing, which could make it higher. The increased precision of the score fusion algorithm, proposed within this study, indicate an important possibility for future work—the joint application of independent decompositions which will reduce false positives and false negatives and more accurate segmentation of once detected areas.

Acknowledgements This work was supported by the National Science Fund of Bulgaria: KP-06-H27/16 *Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas.*

References

1. Ward, S., Hensler, J., Alsalam, B., Gonzalez, L.: Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision. In: Proceedings of the IEEE Aerospace Conference, pp. 1–8. IEEE, Big Sky, MT, USA (2016)
2. Matzner, S., Cullinan, V., Duberstein, C.: Two-dimensional thermal video analysis of offshore bird and bat flight. *Ecol. Inf.* **30**, 20–28 (2015)
3. Corcoran, E., Denman, S., Hanger, J., Wilson, B., Hamilton, G.: Automated detection of koalas using low-level aerial surveillance and machine learning. *Sci. Rep.* **9**(1), 1–9 (2019)
4. Oishi, Y., Oguma, H., Tamura, A., Nakamura, R., Matsunaga, T.: Animal detection using thermal images and its required observation conditions. *Remote Sens* **10**(7), 1050 (2018)
5. Bondi, E., Fang, F., Hamilton, M., Kar, D., Dmello, D., Choi, J., Hannaford, R., Iyer A., Joppa L., Tambe M., Nevatia, R.: Spot poachers in action: augmenting conservation drones with automatic detection in near real time. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, pp. 7741–77460. AAAI, New Orleans, LA, USA (2018)
6. Lhoest, S., Linchant, J., Quevauvillers, S., Vermeulen, C., Lejeune, P.: How many hippos (HOMHIP): algorithm for automatic counts of animals with infra-red thermal imagery from UAV. *Int. Arch. Photogrammetry Remote Sens. Spat. Inf. Sci.* **40**(3), 355–362 (2015)
7. Jorquera-Chavez, M., Fuentes, S., Dunshea, F., Warner, R., Poblete, T., Jongman, E.: Modelling and validation of computer vision techniques to assess heart rate, eye temperature, ear-base temperature and respiration rate in cattle. *Animals* **9**(12), 1089 (2019)
8. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented Lagrangian methods for semidefinite programming. *Math. Program. Comput.* **2**(3–4), 203–230 (2010)
9. Goldfarb, D., Qin, Z.: Robust low-rank tensor recovery: models and algorithms. *SIAM J. Matrix Anal. Appl.* **35**(1), 225–253 (2014)
10. Sobral, A., Bouwmans, T., Zahzah, E.-H.: Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos. In: Bouwmans, T., Aybat, N., Zahzah, E.-H. (eds.) *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press, Cleveland (2016)
11. Kasai, H.: Online low-rank tensor subspace tracking from incomplete data by CP decomposition using recursive least squares. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2519–2523. IEEE, Shanghai, China (2016)
12. Biswas, S., Milanfar, P.: Linear support tensor machine with SLK channels: pedestrian detection in thermal infrared images. *IEEE Trans. Image Process.* **26**(9), 4229–4242 (2017)
13. Pang, Y., Shi, X., Jia, B., Blasch, E., Sheaff, C., Pham, K., Chen, G., Ling, H.: Multiway histogram intersection for multi-target tracking. In: Proceedings of the IEEE International Conference on Information Fusion, pp. 1938–1945. IEEE, Washington, DC, USA (2015)
14. Javed, S., Dias, J., Werghi, N.: Low-rank tensor tracking. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 605–614. IEEE, Seoul, Korea (2019)
15. Fanaee-T, H., Gama, J.: Tensor-based anomaly detection: an interdisciplinary survey. *Knowl.-Based Syst.* **98**, 130–147 (2016)
16. Comon, P., Luciani, X., De Almedia, A.: Tensor decompositions, alternating least squares and other tales. *J. Chemometr. J. Chemometr. Soc.* **23**(7–8), 393–405 (2009)
17. Li, N.: Variants of ALS on Tensor Decompositions and Applications. Ph.D. thesis. Clarkson University, Potsdam, NY, USA (2013)

18. Kaya, O., Ucar, B.: High performance parallel algorithms for the Tucker decomposition of sparse tensors. In: Proceedings of the 45th International conference on parallel processing, ICPP 2016, vol. 1, pp. 103–112. IEEE, Philadelphia, PA, USA (2016)
19. Scoley, G., Gordon, A., Morrison, S.: Use of thermal imaging in dairy calves: exploring the repeatability and accuracy of measures taken from different anatomical regions. *Transl. Animal Sci.* **3**(1), 564–576 (2019)
20. Jorquera-Chavez, M., Fuentes, S., Dunshea, F., Warner, R., Poblete, T., Morrison, R., Jongman, E.: Remotely sensed imagery for early detection of respiratory disease in pigs: a pilot study. *Animals* **10**(3), 451 (2020)
21. Kolda, T., Bader, B.: Tensor decompositions and applications. *SIAM Rev.* **51**(3), 455–500 (2009)
22. Battaglino, C., Ballard, G., Kolda, T.: A practical randomized CP tensor decomposition. *SIAM J. Matrix Anal. Appl.* **39**(2), 876–901 (2018)
23. Sofuoglu, S., Selin, A.: A two-stage approach to robust tensor decomposition. In: Proceedings of the 2018 IEEE/SP Workshop on Statistical Signal Processing, SPP, vol. 1, pp. 831–835. IEEE, Freiburg, Germany (2018)
24. Huang, B., Mu, C., Goldfarb, D., Wright, J.: Provable models for robust low-rank tensor completion. *Pac. J. Optim.* **11**(2), 339–364 (2015)
25. Guyon, C., Bouwmans, T., Zahzah, E.-H.: Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis. In: Sanguansat, P. (eds.) *Principal Component Analysis*, vol. 10, pp. 223–238. IntechOpen (2012)

Chapter 5

Partial Contour Matching Based on Affine Curvature Scale Space Descriptors



Sinda Elghoul and Faouzi Ghorbel

Abstract In real applications, the same object may have been presented by different shapes due to the moment and the angles of image acquisition, which does not guarantee a complete contour extraction without being disturbed by the noise or the distortions. In this paper, we propose a new method to match partially occluded shape based on affine curvature scale space. Firstly, an affine curve re-parameterization is defined, inspired by the properties of affine curvature scale space (ACSS) shape descriptor. Then, the different parts will be matched in order to minimize the L_2 distance by the calculation of the pseudo-inverse matrix to estimate the translation and the linear transformation based on the affine curve matching (ACM) algorithm. Finally, a matching curve algorithm is obtained according to any planar affine transformation and in any partial occluded case. Experiments are conducted on multi-view curve dataset.

5.1 Introduction

The shape matching of planar curves that are subjected to certain occlusion and viewing transformations is motivated by a board of problems arising from different applications in many fields such as robot navigation [1, 2], medical image matching [3, 4], face recognition [5, 6], and object tracking [7, 8].

The research on the shape matching is recognized as a classical field and well-studied in the literature. However, there are other challenges that still remain open until now such as noise, distortion, and deformation. They are caused by two main reasons: First, images taken at different moments and from different angles of the

S. Elghoul (✉) · F. Ghorbel
CRISTAL Laboratory, GRIFT Research Group, National School of Computer Science (ENSI),
Manouba University, Manouba 2010, Tunisia
e-mail: Sinda.elghoul@gmail.com

F. Ghorbel
e-mail: Faouzi.ghorbel@ensi.rnu.tn

same object suffer from perspective distortions [9]. This transformation can be approximated by a two-dimensional affine transformation [10] if the object is planar and far enough from the image plane. The second reason, the curve sometimes can present a little part of the shape [11].

However, it is not obvious to define an approach that satisfies these different criteria, so an appropriate trade-off is necessary. In the last decades, various methods for shape matching have been developed. The best-known researches include affine-invariant Fourier descriptors (AIFD) method [12, 13], shape context (SC) methods [14], the inner distance shape context (IDSC) [15], the height function (HF) descriptor [16], and so on. However, they suppose that the complete shape can be extracted from the images [17]. While in reality, the extracted shape may be occluded by many distortions [18] like noise, articulations and missed contour portions, as well as affine transformation. In [19], Chen et al. introduce a partial shape matching method based on the Smith–Waterman algorithm. Also, Latecki et al. [20] define an elastic partial shape matching to model distortion and occlusion of shapes. In [21], the author proposes a matching algorithm based on dynamic time warping. Therefore, only some methods treat both affine and open curve matching [9, 22]. In [17], Mai and al. define an affine-invariant partial shape matching approach where each contour is segmented into affine-invariant segments by the application the local maxima of curvature scale space (CSS). Then, the different parts are matched using the Smith–Waterman algorithm. However, the Smith–Waterman (SW) algorithm is sensitive to position jitter of a point sequence as indicated in [9]. Moreover, Huijing Fu and al [9] present an affine planar shape matching and exploit it for partial object recognition where an affine-invariant curve descriptor (AICD) using the affine-invariant signature is defined. Then, a partial curve matching algorithm is developed by combining AICD with a curve segmentation strategy based on inflexion points. Yet, their method has limited accuracy in the noise condition case caused by the high number of derivation [23]. In this paper, we are interested in part-to-part affine shape matching. We can present the partial shape matching problem by giving as an input two partial shape, often called source and target, the goal is to recover the affine transformation that optimized the pairing between the source and the target. So to achieve this pairing, we should define the same parameterization for each curve. The underlying idea here is to divide each shape into ordered affine-invariant segments inspired by the expression of affine curvature scale space [24]. Then to handle the matching problem, we estimate the affine transformation parameters using affine curve matching (ACM) [25] algorithm. This algorithm minimizes the approximation of L_2 distance between pairs by the computation of the pseudo-inverse matrix. In experimentation, we will evaluate the performance of proposed approach in the context of shape retrieval and we will compare it with different affine curves matching methods presented in the state of the art.

The remainder of the paper is organized as follows: In Sect. 5.2, the detailed descriptions of suggested curve matching algorithm will be presented. Section 5.3 will investigate the effectiveness of the proposed approach through experiments and analyses. Finally, the last section submits the conclusion.

5.2 Affine Curvature Scale Space Curve Matching Algorithm

Here, we define our new partial contour matching based on curvature scale space matching algorithm. Therefore, this section will be devoted firstly to recall the affine re-parametrization based on the affine curvature scale space (ACSS) [24]. After that, we recall the newly developed affine curve matching (ACM) algorithm [25] and apply the pseudo-inverse matrix to estimate the translation vector B and the special linear transformation A existing between the two curves up to a special affine transformation.

5.2.1 Affine Curvature Scale Space Descriptor (ACSS)

As given Ω_1 and Ω_2 two planar shapes represented by closed or open continuous curves. We extract two contours of each shape which are represented by two parameterizations, respectively, as: $\Omega_1 = [f(t) = (f^x(t)f^y(t))] (t = 1; 2; \dots; N)$ and $\Omega_2 = [h(t') = (h^x(t')h^y(t'))] (t' = 1; 2; \dots; N)$, where their relation is defined by:

$$h(t') = Af(t) + B \quad (1)$$

With B is a translation vector and A is a special linear transformation. It is obvious that each curve can be represented by different parameterizations. Therefore, we cannot consider that the two contours have the same parameterization and we compare the different viewpoints of planar shape. To handle this problem, we must assure that the parameterization is independent from transformations and distortions. As a result, we need to re-parameterize the points of the contour. In [24], it is proved that the locations of the contour local maxima in the affine curvature scale space (ACSS) image is invariant under an affine transformation. Moreover, they are robust to noise as indicated in [17]. The underlying idea is to do an affine curves re-parameterization by applying an ACSS descriptor. Then, we describe the following main steps of the ACSS method as indicated in [24]. Firstly, we re-parameterize the contour using the affine length function $l(t)$ defined by

$$l = \frac{1}{L} \int_0^l |\det(\dot{f}(u), \ddot{f}(u))|^{\frac{1}{3}} \quad (2)$$

where the total affine length L of the considered curve is presented by:

$$L = \int_0^T |\det(\dot{f}(u), \ddot{f}(u))|^{\frac{1}{3}} du \quad (3)$$

With \dot{f} and \ddot{f} denote, respectively, first and second derivatives of f and T are a positive real. As a result of re-parameterization by this affine length, the relation between the two curves becomes:

$$h^*(l) = Af^*(l) + B \quad (4)$$

With h^* and f^* denotes a re-parameterization by this affine length. In the rest of paper, we will replace f^* by f and h^* by h to simplify the notation. Then, we compute the curvature function $k(l)$ expressed by:

$$k(l) = \frac{\dot{f}^x(l)\ddot{f}^y(l) - \ddot{f}^x(l)\dot{f}^y(l)}{(\dot{f}^{x^2}(l) + \dot{f}^{y^2}(l))^{\frac{3}{2}}} \quad (5)$$

where $\dot{f}^x(l)$, $\dot{f}^y(l)$ and $\ddot{f}^y(l)$, $\ddot{f}^x(l)$ are the first and second derivatives. If $g(l, \sigma)$, a 1-D Gaussian kernel of width σ , is convolved with each component of the curve, then $f_\sigma^x(l, \sigma)$ and $f_\sigma^y(l, \sigma)$ represent the components of the resulting curve, f_σ :

$$f_\sigma^x(l, \sigma) = f^x \otimes g(l, \sigma) \quad f_\sigma^y(l, \sigma) = f^y \otimes g(l, \sigma)$$

where \otimes is a convolutional operator. The curvature of f_σ is given by:

$$k(l, \sigma) = \frac{\dot{f}^x(l, \sigma)\ddot{f}^y(l, \sigma) - \ddot{f}^x(l, \sigma)\dot{f}^y(l, \sigma)}{(\dot{f}^{x^2}(l, \sigma) + \dot{f}^{y^2}(l, \sigma))^{\frac{3}{2}}}$$

5.2.2 Affine Curve Matching (ACM) Algorithm

To solve the matching problem, we must find A and B to estimate the relation and motion between the different contours as indicated in [25]. The re-parametrization of two curves by the ACSS gives the following rectangular system formed by $2*N$ equations and 6 unknown variables:

$$\left\{ \begin{array}{l} h_\sigma^x(l_1) = f_\sigma^x(l_1)a_{11} + f_\sigma^y(l_1)a_{12} + B^x \\ h_\sigma^y(l_1) = f_\sigma^x(l_1)a_{21} + f_\sigma^y(l_1)a_{22} + B^y \\ \vdots \\ h_\sigma^x(l_N) = f_\sigma^x(l_N)a_{11} + f_\sigma^y(l_N)a_{12} + B^x \\ h_\sigma^y(l_N) = f_\sigma^x(l_N)a_{21} + f_\sigma^y(l_N)a_{22} + B^y \end{array} \right.$$

with $f(l)$ and $h(l)$ as the re-parametrization, respectively, of two contours $f(t)$ and $h(t')$, $B = (B^x; B^y)$ and $A = (a_{ij})_{1 \leq (i,j) \leq 2}$. Our goal is to minimize the error e between the two contours by the estimation of A and B which will be defined by:

$$\min_{(A,B)} \|Af_{\sigma}(l) + B - h_{\sigma}(l)\|^2 \cong e$$

This system can be written in matrix notation:

$$H = DU$$

With $U = [a_{11} \ a_{12} \ a_{21} \ a_{22} \ B^x \ B^y]^t$; $H = [h_{\sigma}^x(l_1), h_{\sigma}^y(l_1) \dots h_{\sigma}^x(l_N), h_{\sigma}^y(l_N)]$ and

$$D = \begin{pmatrix} f_{\sigma}^x(l_1) & f_{\sigma}^y(l_1) & 0 & 0 & 1 & 0 \\ 0 & 0 & f_{\sigma}^x(l_1) & f_{\sigma}^y(l_1) & 0 & 1 \\ f_{\sigma}^x(l_2) & f_{\sigma}^y(l_2) & 0 & 0 & 1 & 0 \\ 0 & 0 & f_{\sigma}^x(l_2) & f_{\sigma}^y(l_2) & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{\sigma}^x(l_N) & f_{\sigma}^y(l_N) & 0 & 0 & 1 & 0 \\ 0 & 0 & f_{\sigma}^x(l_N) & f_{\sigma}^y(l_N) & 0 & 1 \end{pmatrix}$$

In [25], the author applies the least squares method to solve the over determined system of linear equations when the numbers of equations are more than unknown variables. Thus, the resolution of this rectangular system can be done by minimizing the error via inverting the system by using pseudo-inverse of the matrix D as indicated in [25].

$$U = (D^t D)^{-1} D^t H$$

Then, we calculate the normal matrix $(D^t D)$ which has the following expression:

$$D^t D = N^6 \begin{pmatrix} \overline{X^2} & \overline{XY} & 0 & 0 & \overline{X} & 0 \\ \overline{XY} & \overline{Y^2} & 0 & 0 & \overline{Y} & 0 \\ 0 & 0 & \overline{X^2} & \overline{XY} & 0 & \overline{X} \\ 0 & 0 & \overline{XY} & \overline{Y^2} & 0 & \overline{Y} \\ \overline{X} & \overline{Y} & 0 & 0 & 1 & 0 \\ 0 & 0 & \overline{X} & \overline{Y} & 0 & 1 \end{pmatrix}$$

$$\overline{X} = \frac{1}{N} \sum_{k=1}^N (f_{\sigma}^x(l_k)) \text{ and } \overline{Y} = \frac{1}{N} \sum_{k=1}^N (f_{\sigma}^y(l_k))$$

The reader can find more details in [24, 25].

5.3 Experiments

In this section, we provide the recognition rates of the suggested approach and compare it with the exiting shape matching methods. The experiments are carried on multi-view curve dataset (MCD). First, we evaluate the performance of our approach in the shape retrieval. Then, we prove the performance of the proposed algorithm on shape registration. Finally, we will analyze the algorithm complexity of our approach.

5.3.1 Retrieval Accuracy

To perform our algorithm in shape retrieval task, we calculate the bulls-eye score as defined in [26, 27] on multi-view curve dataset (MCD) [28], since it contains forms that undergo an affine transformation. So to calculate the bulls-eye score, we compare each curve to the whole MCD dataset curve (including itself). Then, the contour number of the same class that are midst the $2 * N_c$ most similar is recovered, with N_c is the sample number per class. The bulls-eye score is the ratio of the number of correct results and the highest possible number of correct results [9, 27].

The MCD contains 40 shape classes taken from MPEG-7 datasets. Each class presents 14 curve samples that correspond to different perspective distortions of the same shape. Samples of contours from MCD datasets are shown in Fig. 5.1.

Table 5.1 compares the retrieval bulls-eye score to the first 10 contours of the MCD database using our approach with some existing methods. In terms of the average rate performance, the proposed approach performs reasonably well as compared to many other techniques such as affine curve matching (ACM) [25] algorithm with 94% of rate and especially the curvature scale space Smith–Waterman (CSS-SW) [17] approach which prove that our algorithm is more efficient in terms of registration since we are based on ACSS in the re-parameterization step.



Fig. 5.1 Different contours images from the MCD database, two images from each class [28]

Table 5.1 Retrieval results on the entire MCD Dataset

Algorithm	Bulls eye score (%)
Arbter [3]	41
Shape Context [14]	56.29
Huang [25]	71
Chaker's invariant [30]	76
Rube [31]	79
CSS-SW [17]	89
ACM [25]	94
Proposed algorithm	95.98

5.3.2 Shape Registration

Shape registration is a crucial applications of the proposed algorithm [29]. Most shapes of MCD datasets are represented by closed curves. So to evaluate the proposed algorithm in partial occlusion and deformable registration, we take off some parts of the contour to make it open as indicated in [17]. Figure 5.2 shows our method in full-to-full and part-to-part registration case.

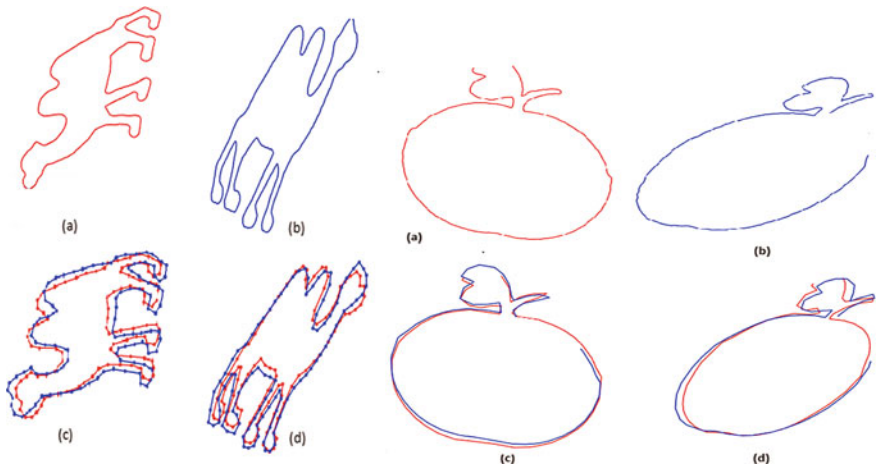


Fig. 5.2 Proposed method: **a** and **b** are the initial shape part; **c** and **d** show the original shape overlaid with the registered shape

5.3.3 Algorithm Complexity

Here, we compare our algorithm complexity with the CSS-SW. The calculation of the shape descriptors and shape matching are two independent steps for the proposed approach. Therefore, the complexity of the proposed method will be evaluated separately here. We consider that N is the number of sample points of the contour. For each curve, we should start by the re-parametrization step based on ACSS descriptor which is a common step between the two methods. In the matching step, the proposed algorithm can do the matching with $O(N)$ complexity which underlines the speed of pseudo-inverse. However, Smith–Waterman complexity is $O(N^3)$ [9], since it requires SVD calculation.

5.4 Conclusion

In this paper, we propose a new matching method which can deal with both occlusion and affine transformations. Firstly, the contour is divided into affine-invariant segment by applying the ACSS descriptor. Then, we estimate the affine transformation using the ACM algorithm. As a result, an affine curve matching is achieved. Experiment results show that the proposed algorithm is simple and can cope partially occlusion and affine transformation. In our future work, we will apply our algorithm in different application domains as remote sensing and robotic recognition.

References

1. Diedrich, W., Latecki, L.J.: Shape matching for robot mapping. In: Pacific Rim International Conference on Artificial Intelligence. Springer, Berlin, Heidelberg (2004)
2. Marius, M., Lowe, D.: Fast matching of binary features. In: Ninth Conference on Computer and Robot Vision (CRV). IEEE (2012)
3. Hava, L., Arridge, S.: A survey of hierarchical non-linear medical image registration. *Pattern Recogn.* **32**(1), 129–149 (1999)
4. Hemamalini, G., Prakash, J.: Medical image analysis of image segmentation and registration techniques. *Int. J. Eng. Technol. (IJET)* **8**(5):2234–2241 (2016)
5. Ehsan Fazl, E., Zelek, J.: Local feature matching for face recognition. In: The 3rd Canadian Conference on Computer and Robot Vision. IEEE (2006)
6. Jilin, T., Huan, T., Tao, H.: Face as mouse through visual face tracking. In: Proceedings of the 2nd Canadian Conference on Computer and Robot Vision. IEEE (2005)
7. Yixin, C., Das, M., Bajpai, D.: Vehicle tracking and distance estimation based on multiple image features. In: Fourth Canadian Conference on Computer and Robot Vision. CRV'07. IEEE (2007)
8. Marc, L.: Real-time eye blink detection with GPU-based SIFT tracking. In: Fourth Canadian Conference on Computer and Robot Vision, CRV'07. IEEE (2007)
9. Huijing, F.: Novel affine-invariant curve descriptor for curve matching and occluded object recognition. *IET Comput. Vis.* **7**(4), 279–292 (2013)

10. Forsyth, D.: Invariant descriptors for 3D object recognition and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **10**, 971–991 (1991)
11. Turney, J.L., Trevor Mudge N., Richard A.V.: Recognizing partially occluded parts. *IEEE Trans. Pattern Anal. Mach. Intell.* **4**, 410–421 (1985)
12. Ghorbel, F.: Towards a unitary formulation for invariant image description: application to image coding. *Annales des Telecommun.* **53**, 242–260. Springer (1992)
13. Arbter, K.: Application of affine-invariant Fourier descriptors to recognition of 3-D objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 640–647 (1990)
14. Mori, G., Serge, B., Jitendra, M.: Efficient shape matching using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(11), 1832–1837 (2005)
15. Ling, H., David, W.: Shape classification using the innerdistance. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(2), 286–299 (2007)
16. Wang, J.: Shape matching and classification using height functions. *Pattern Recogn. Lett.* **33**(2), 134–143 (2012)
17. Mai, F., Chang, C.Q., Hung, Y.S.: Affine-invariant shape matching and recognition under partial occlusion. In: 17th IEEE International Conference on Image Processing (ICIP). IEEE (2010)
18. Yang, C., Hui, W., Qian, Y.: A novel method for 2D nonrigid partial shape matching. *Neurocomputing* **275**, 1160–1176 (2018)
19. Chen, L., Rogerio, F., Turk, M.: Efficient partial shape matching using smith-waterman algorithm. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE (2008)
20. Latecki, L., et al.: An elastic partial shape matching technique. *Pattern Recogn.* **40**(11), 3069–3080 (2007)
21. Bouagar, S., Slimane, L.: Efficient descriptor for full and partial shape matching. *Multimedia Tools Appl.* **75**(6), 2989–3011 (2016)
22. Zhang, G., JiYuan Xu X., JianXin. L.: A new method for recognition partially occluded curved objects under affine transformation. In: 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE (2015)
23. Arulmozhi, P., Abirami, S.: Shape based image retrieval: a review. *Int. J. Comput. Sci. Eng.* **6**(4), 147 (2014)
24. Mokhtarian, F., Sadegh, A.: Affine curvature scale space with affine length parametrisation. *Pattern Anal. Appl.* **4**(1), 1–8 (2001)
25. Elghoul, S., Ghorbel, F.: An efficient 2D curve matching algorithm under affine transformations. In: VISIGRAPP (4: VISAPP) (2018)
26. Yang, C., Wei, H., Yu, Q.: Multiscale triangular centroid distance for shape-based plant leaf recognition. In: ECAI (2016)
27. Yang, C., Wei, H., Yu, Q.: A novel method for 2D nonrigid partial shape matching. *Neurocomputing* **275** (2018)
28. Zuliani, M.: Affine-invariant curve matching. In: International Conference on Image Processing, ICIP'04, vol. 5. IEEE (2004)
29. Mai, F., Chang, C.Q., Hung, Y.S.: A subspace approach for matching 2D shapes under affine distortions. *Pattern Recogn.* **44**(2), 210–221 (2011)
30. Hanbyul, J.: Graph-based robust shape matching for robotic application. In: IEEE International Conference on Robotics and Automation, ICRA'09. IEEE (2009)
31. Mark, G., Bekaert, P.: Local stereo matching with segmentation-based outlier rejection. In: The 3rd Canadian Conference on Computer and Robot Vision. IEEE (2006)

Chapter 6

Vision-Based Line Tracking Control and Stability Analysis of Unicycle Mobile Robots



Plamen Petrov and Veska Georgieva

Abstract This paper addresses the problem of vision-based line tracking control of unicycle mobile robots. First, a robot-camera model suitable for path following applications is derived. Using a look-ahead approach, a feedback controller is proposed for tracking curved paths on the ground using information from an onboard down-looking camera using distance-only measurements. Stability properties of the closed-loop system are analyzed, and asymptotic stability of the resulting closed-loop control system is proved using Lyapunov stability theory. Simulation and experimental results are presented to illustrate the effectiveness of the proposed control scheme.

6.1 Introduction

In the last decades, the wheeled mobile robots (WMRs) have been increasingly used in wide range of applications, such as in factories [1], as service robots [2], autonomous vehicles [3], exploration robots [4], for military operations [5], and research [6]. At control level, important results have been established concerning three fundamental motion control tasks, namely point stabilization (the parking problem), trajectory tracking, and path following [7]. In what concerns the path following problem, in contrast of trajectory tracking, the mobile robot has to follow and to converge to a reference path, which is given without temporal specification. The path following control of mobile robot has been intensively studied during the years, and different solutions have been proposed in the literature, such as nonlinear [8] and linear controllers [9]. When designing controllers for path tracking applications, often the curvature of the reference path is considered a priori known, which

P. Petrov (✉) · V. Georgieva
Technical University of Sofia, Sofia 1000, Bulgaria
e-mail: ppetrov@tu-sofia.bg

V. Georgieva
e-mail: vesg@tu-sofia.bg

permits to introduce it directly in the control law, simplifying the design of the controller [10]. In this case, it is also considered that both measurements for the lateral and orientation errors are available for feedback control design [11]. However, depending of the assigned mobile robot guiding point, using look-ahead approach [12], it is possible to use only distance measurements for the lateral offset from the reference line without using data for the orientation error, which considerably simplify the controller design. The advantage of using such approach is considerable in the case, when the mobile robot has to follow a path with unknown curvature and avoids its direct calculation, which involves higher-order derivative computation. In addition, calculation of the orientation error between the mobile robot and the reference path avoided, which considerably simplifies the structure of the controller.

The use of visual information in the feedback loop has been an attractive solution for the motion control of mobile robots and different visual servo algorithms have been developed over the last decades. The control based of visual measurements is termed visual servoing or vision-based control and two main approaches are distinguished, namely position-based visual servoing (PBVS) and image-based visual servoing (IBVS) [13]. In the PBVS approach, three-dimensional scene information is used and the feedback is based on the pose estimation of the observed object with respect to the camera in order to regulate the motion of the onboard camera to a desired pose [14–16]. For path following applications, a specific feature is that the error coordinates with respect to the desired path to follow are computed in the task space. In the IBVS approach, the pose estimation is omitted, image features are used as the state in the control, such that the error coordinates are measured in the image, and the control law is directly expressed in the image plane and mapped to actuator commands [17–19].

In this paper, we deal with the problem of path tracking control for nonholonomic unicycle WMRs, which use monocular vision guided system for line tracking. The reference path is assumed to be a sequence of circular and/or straight-line segments, where the curvatures of the circular segments are not known. The look-ahead approach [12, 15] used in this paper consists of tracking a reference path with a guiding reference point in front of robot at a given distance ahead from the wheel axle. For this end, first, a robot-camera model of the robot using the look-ahead reference point is derived. A linear controller is designed using lateral error-only measurements without involving values of the path curvature (which is unknown) or the orientation error with respect to the reference path. Based on Lyapunov stability theory, the stability property of the synthesized system is analyzed. Some simulation and experimental results are given, in order to demonstrate the validity of the designed controller.

6.2 Model Development

6.2.1 Robot Kinematic Model

A plane view of a unicycle mobile robot, considered in this paper, which is moving on a horizontal plane, is shown in Fig. 6.1. A monocular camera is placed on the robot pointing downward perpendicular to the ground. For simplicity of exposition, for path following applications, we will use two-dimensional (2D) Cartesian coordinate systems instead 3D systems, where it is possible. With reference to Fig. 6.1, in order to describe the path tracking kinematics, the following coordinate systems are defined: an inertial coordinate system Fxy ; a robot coordinate system Px_Py_P located at the mid-point of the wheel axle with x_P axis directed along the longitudinal base of the robot; a robot coordinate system Rx_Ry_R attached firmly to the robot at a distance h from the ground and at a distance d ahead from point P , in such way that the x_R -axis is aligned with x_P -axis of Px_Py_P ; a camera frame Cx_Cy_C which center is placed in the optical center of the camera and coincides with the center of the robot frame Rx_Ry_R ; and a moving virtual reference frame Lx_Ly_L associated with the reference line to follow, with a center L assigned in the intersection between the y_R -axis of Rx_Ry_R and the reference path, such that the x_L -axis is tangent to the path and oriented in the direction of motion. The look-ahead point R is defined as a robot guided (reference) point for the line tracking scenario.

Let us denote the coordinates of point P in the fixed frame Fxy as

$${}^F P_P = [x_P \ y_P]^T \tag{6.1}$$

where ${}^F P_P \in \mathbb{R}^2$, and the orientation of the robot in Fxy as $\psi \in S^1$. The kinematic equations of motion of the unicycle mobile robot under the nonholonomic constraints

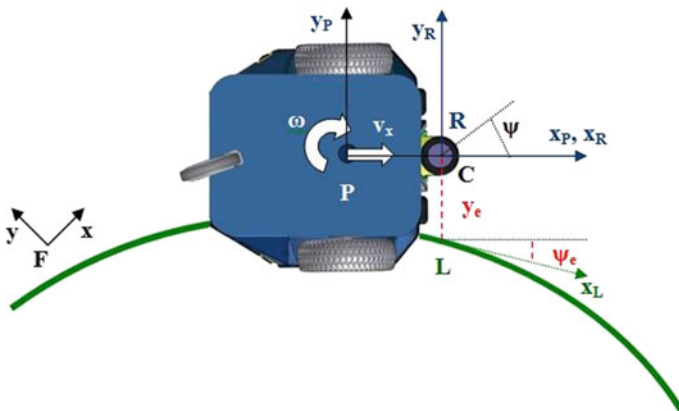


Fig. 6.1 Geometry of the line tracking scenario

of pure rolling and nonslipping can be written as follows [12]

$$\dot{\widehat{p}}_P = B_P \eta, \quad (6.2)$$

where $\dot{\widehat{p}}_P$ represents the time derivative of $\widehat{p}_P \in R^2 \times S^1$, defined as

$$\widehat{p}_P = [x_P \ y_P \ \psi]^T, \quad (6.3)$$

which is a vector of the posture coordinates of the robot using the reference point P , the transformation matrix $B_P \in R^{3 \times 2}$ is given by

$$B_P = \begin{bmatrix} \cos \psi & 0 \\ \sin \psi & 0 \\ 0 & 1 \end{bmatrix},$$

and the vector $\eta \in \mathfrak{R}^2$

$$\eta = [v_{Px} \ \omega]^T \quad (6.4)$$

is composed of the robot linear and angular velocities denoted by $v_{Px} \in \mathfrak{R}$ and $\omega \in \mathfrak{R}$, respectively.

Denoting with ${}^F p_R \in \mathfrak{R}^2$ a vector representing the coordinates of point R with respect to Fxy

$${}^F p_R = [x_R \ y_R]^T, \quad (6.5)$$

and with ${}^P d = [{}^P d_x \ 0]^T \in \mathfrak{R}^2$ a vector from point P to point R expressed in the coordinate frame $Px_P y_P$, using (6.1) the coordinates of points R and P are related by

$${}^F p_R = {}^F p_P + R(\psi) {}^P d, \quad (6.6)$$

where $R(\psi) \in SO(2)$ is an orthogonal rotation matrix of angle ψ , (ψ is the orientation angle of the robot with respect to fixed frame Fxy), given by

$$R(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix}.$$

Differentiating (6.6) with respect to time, one obtains

$$\begin{aligned} {}^F \dot{p}_R &= R(\psi)({}^P \dot{p}_P + \omega S(\pi/2) {}^P d) \\ &= G\eta, \end{aligned} \quad (6.7)$$

where $G = R(\psi)D \in \mathfrak{R}^{2 \times 2}$ with $D \in \mathfrak{R}^{2 \times 2}$ given by

$$D = \begin{bmatrix} 1 & 0 \\ 0 & {}^P d_x \end{bmatrix},$$

$S(\pi/2) \in SS(2)$ is a skew symmetric matrix and ${}^P \dot{p}_P = [v_{Px}, v_{Py}]^T \in R^2$ is a vector of the projections of the velocity of point P relative to the fixed coordinate system Fxy on the axes of the moving robot coordinate system Px_Py_P .

In order to express the robot kinematic model using the coordinates (6.5) of point R in an inertial frame Fxy , we define a vector of the posture coordinates ${}^F \widehat{p}_R \in R^2 \times S^1$, as follows

$${}^F \widehat{p}_R = [x_R \ y_R \ \psi]^T. \quad (6.8)$$

Then, using (6.8), one can write

$${}^F \dot{\widehat{p}}_R = B_R \eta, \quad (6.9)$$

where B_R is a block matrix of the form

$$B_R = \begin{bmatrix} G \\ \text{---} \\ j \end{bmatrix} \in \mathfrak{R}^{3 \times 2},$$

and the row vector j is given by $j = [0 \ 1] \in \mathfrak{R}^{1 \times 2}$.

The path following geometry considered in this paper is illustrated in Fig. 6.1. It is assumed that the path is a smooth planar curve. The coordinate systems Rx_Ry_R and Lx_Ly_L are defined to describe the error kinematics during the path tracking process. The moving reference system Lx_Ly_L is defined such that the x_L -axis is tangent to the path and oriented in the direction of robot motion. The y_R -axis of the robot coordinate system Rx_Ry_R passes through the reference point L associated with the path to follow.

The coordinates and orientation of the frame Lx_Ly_L in the coordinate frame Rx_Ry_R can be expressed in the form

$$e_{\text{pos}} = T(p_L - p_R), \quad (6.10)$$

where $e_{\text{pos}} = [x_e \ y_e \ \psi_e]^T \in \mathfrak{N}^3$ is the error posture, x_e is the longitudinal error, y_e is the lateral error, and ψ_e is the orientation error; the posture vector $p_L \in \mathfrak{N}^2 \times S^1$ associated with the reference path is defined as

$$p_L = [x_L \ y_L \ \psi_L]^T, \quad (6.11)$$

where ψ_L is the orientation of the coordinate frame Lx_Ly_L with respect to the fixed frame F_{xy} , the posture vector p_R is given by (6.8) and the orthogonal matrix $T \in SO(3)$ is given by

$$T = \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Differentiating (6.10) with respect to time and taking into account the nonholonomic constraints $v_{Py} = v_{Ly} = 0$, (v_{Py} and v_{Ly} are the projections of the velocities of points R and L on the y_R and y_L axes, respectively), and using the fact that $x_e(t) = \dot{x}_e(t) = 0$ after some work, the error kinematics for path following applications is obtained in the form

$$\dot{e} = \zeta_1(e) + \zeta_2(e)\omega, \quad (6.12)$$

where $\dot{e} \in \mathfrak{N}^2$ represents the time derivative of $e \in \mathfrak{N}^2$ defined as

$$e = [y_e \ \psi_e]^T, \quad (6.13)$$

$\zeta_1(e)$ and $\zeta_2(e)$ are C^1 vector functions given by

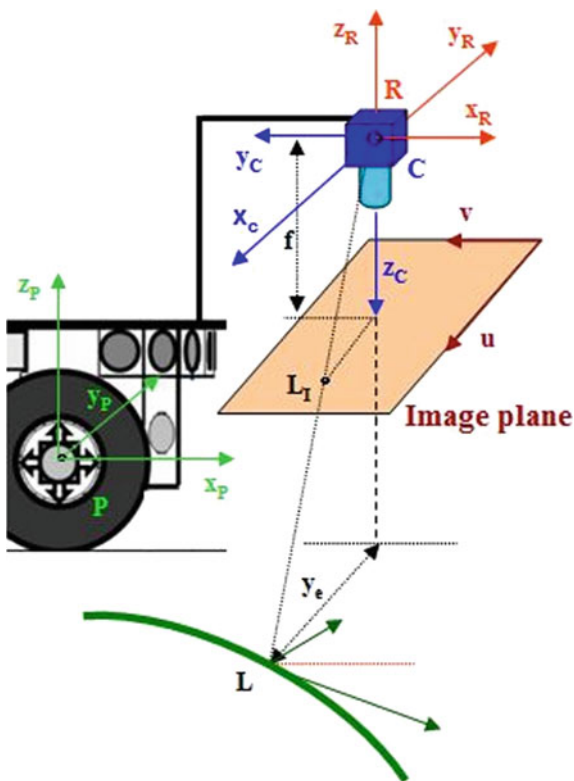
$$\begin{aligned} \zeta_1(e) &= [v \tan \psi_e \ v c_r / \cos \psi_e]^T, \\ \zeta_2(e) &= [-(y_e \tan \psi_e + {}^P d_x) - (1 + y_e c_r / \cos \psi_e)]^T, \end{aligned}$$

v_{Px} is the velocity of mid-point P of the wheel axle, ω is the angular velocity of the robot considered as control variable (input) of the system, and c_r is the curvature of the reference path.

6.2.2 Robot-Camera Model

A monocular camera is placed in front of the mobile robot, where the origin of the camera frame $Cx_Cy_Cz_C$ coincides with the center of the coordinate system $Rx_Ry_Rz_R$,

Fig. 6.2 Robot-camera geometry



at distance h from the ground. The optical axis of the camera is perpendicular to the surface of motion, as shown in Fig. 6.2. The focal length of the camera is denoted by f .

The geometric relationships between the onboard camera and a feature point from the reference line are shown in Fig. 6.2. Let us denote the position of a feature point L on the reference line with respect to the camera frame $Cx_Cy_Cz_C$ by ${}^C p_L \in \mathfrak{R}^3$, as follows

$${}^C p_L = \begin{bmatrix} {}^C x_L \\ {}^C y_L \\ {}^C z_L \end{bmatrix}. \tag{6.14}$$

The corresponding pixel coordinates ${}^I p_L$ in the pixel coordinate system Iuv fixed to the image plane (Fig. 6.2) are obtained as follows

$${}^I p_L = \begin{bmatrix} u_L \\ v_L \\ 1 \end{bmatrix} = \frac{1}{{}^C z_L} T_{\text{int}} {}^C p_L, \tag{6.15}$$

where ${}^C z_L = h \in \mathfrak{R}$, $T_{\text{int}} \in \mathfrak{R}^{3 \times 3}$ is the intrinsic camera calibration matrix given by

$$T_{\text{int}} = \begin{bmatrix} f s_u & 0 & u_0 \\ 0 & f s_v & v_0 \\ 0 & 0 & 1 \end{bmatrix},$$

(s_u , s_v) are the camera scaling factors and f is the focal length of the camera.

From (6.15), given the pixel coordinates (u_L , v_L), one can determine the coordinates of point L in the camera frame $Cx_Cy_Cz_C$, as follows

$${}^C p_L = {}^C z_L T_{\text{int}}^{-1} p_L. \quad (6.16)$$

On the other hand, given the coordinates ${}^C p_L$ of point L in the camera frame $Cx_Cy_Cz_C$, the coordinates of point L in the robot coordinate frame $Rx_Ry_Rz_R$ are obtained as follows

$${}^R p_L = {}^C z_L {}^{C_1} R_{z_{C_1}, \pi/2}^{-1} R_{x_{C_1}, \pi}^{-1} T_{\text{int}}^{-1} p_L, \quad (6.17)$$

where the two consecutive rotation matrices $R_{x_{C_1}, \pi} \in SO(3)$ and $R_{z_{C_1}, \pi/2} \in SO(3)$ are given by

$$R_{x_{C_1}, \pi} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$R_{z_{C_1}, \pi/2} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The lateral error y_e is calculated from the image based on the pixel coordinates in the image plane of the image feature point L_I (Fig. 6.3), corresponding to the feature point L from the reference line on the ground. Knowing the camera intrinsic parameters and the height h of the camera from the ground, the coordinates of a feature point L belonging to the reference line in the robot reference system $Rx_Ry_Rz_R$ can be recovered from its pixel coordinates in the image plane. This error distance is obtained as a difference between the image plane coordinates of the principal point C and the image point L_I , respectively, and is a function of the intrinsic camera parameters and the known distance h from the camera to the ground.

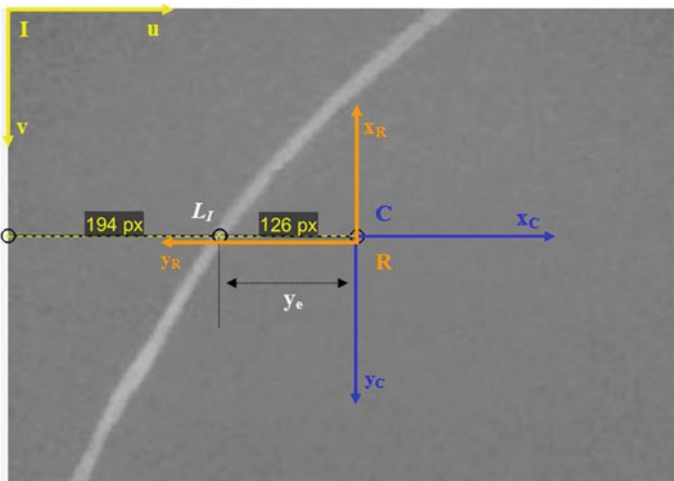


Fig. 6.3 Extraction of a feature point from the line

6.3 Feedback Control Design and Stability Analysis

6.3.1 Circular Line Tracking

Assuming that the reference path is a sequence of circular segments or/and straight lines, and robot velocity v_{Px} is constant and positive (forward robot motion), the path tracking problem consists of finding a feedback control law for the system (6.12) with control input ω , such that the state vector $e = [y_e, \psi_e]^T$ tends to $e_s = [y_{es}, \psi_{es}]^T$ as $t \rightarrow \infty$, where y_{es} and ψ_{es} have constant values.

The following feedback control law is proposed

$$\omega = v_{Px}ky_e, \tag{6.18}$$

where k is a positive constant and v_{Px} , ($v_{Px} = cte > 0$), is the robot speed.

Applying (6.18) to (6.12), the resulting nonlinear closed-loop system has the form

$$\dot{e} = v_{Px}\eta_c(e), \tag{6.19}$$

where the vector function η_c given by

$$\eta_c = \begin{bmatrix} -^P d_x ky_e(1 - ky_e^2) \tan \psi_e \\ -ky_e + ((1 - ky_e^2) / \cos \psi_e) c_r \end{bmatrix}$$

is continuously differentiable.

The equilibrium point e_s of system (6.19) is at

$$\begin{aligned} e_s &= \begin{bmatrix} y_{es} \\ \psi_{\theta_s} \end{bmatrix} \\ &= \begin{bmatrix} (-\text{sqr}t(1 - {}^P d_x^2 c_r^2) + \text{sqr}t(1 - {}^P d_x^2 c_r^2 + 4c_r/k))/2c_r \\ a \sin({}^P d_x c_r) \end{bmatrix}. \end{aligned} \quad (6.20)$$

The stability of the equilibrium point (6.20) for the nonlinear system (6.19) is analyzed by investigating the stability of the linearized system using Lyapunov stability theory [20].

The linearization of the nonlinear system (6.19) at the equilibrium point (6.20) has the form

$$\Delta \dot{e} = v_{P_x} M_c \Delta e, \quad (6.21)$$

where the vector $\Delta e \in \mathfrak{R}^2$ is given by

$$\Delta e = \begin{bmatrix} \Delta y_e \\ \Delta \psi_e \end{bmatrix} = \begin{bmatrix} y_e - y_{es} \\ \psi_e - \psi_{es} \end{bmatrix},$$

$M_c \in \mathfrak{R}^{2 \times 2}$ is the Jacobian matrix of η_c , ($v_{P_x} = cte > 0$) in the form

$$\begin{aligned} M_c &= \left. \frac{\partial \eta_c}{\partial e} \right|_{e=e_s} \\ &= v_{P_x} \begin{bmatrix} -k({}^P d_x + y_{es} \tan \psi_{\theta_s}) & -k(1 + (c_r y_{es} / \cos \psi_{\theta_s})) \\ (1 - ky_{es}^2) / \cos^2 \psi_{\theta_s} & (1 - ky_{es}^2) / \cos^2 \psi_{\theta_s} c_r \end{bmatrix} \\ &= v_{P_x} \begin{bmatrix} {}^P d_x m_{21} & m_{12} \\ m_{21} & m_{12} {}^P d_x c_r^2 \end{bmatrix} \end{aligned} \quad (6.22)$$

with

$$\begin{aligned} m_{12} &= (1 - ky_{es}^2) / (1 - {}^P d_x^2 c_r^2) \\ m_{21} &= -k[1 + (2y_{es} c_r) / \text{sqr}t(1 - {}^P d_x^2 c_r^2)] \end{aligned} \quad (6.23)$$

To analyze the stability of linear system (6.21), we test the eigenvalues λ_i , ($i=1, 2$) of matrix M_c . The characteristic equation of matrix M_c takes the form

$$\begin{aligned} \lambda^2 - \lambda Tr(M_c) + \det(M_c) \\ &= \lambda^2 - {}^P d_x (m_{12} c_r^2 + m_{21}) \lambda - m_{12} m_{21} (1 - {}^P d_x^2 c_r^2). \\ &= 0 \end{aligned} \quad (6.24)$$

In order that the squared Eq. (6.24) does not have any roots with positive real part, it is necessary and sufficient that all its coefficients be of the same sign, which implies that

$$Tr(M_c) < 0 \quad (6.25)$$

and

$$\det(M_c) > 0 \quad (6.26)$$

To be more concrete, consider the case when the curvature of the reference path $c_r = cte > 0$ (analogous results can be obtained when $c_r = cte < 0$), and also the following inequality holds

$$y_{es}c_r + \text{sqrt}(1 - {}^P d_x^2 c_r^2) > 0, \quad (6.27)$$

irrespective of the sign of y_e .

In order to prove that inequality (6.25) holds, the expression for y_{es} from the first equation of (6.20) is substituted into (6.25). Using (6.23), and taking into account that

$$0 < s < 1, \quad (6.28)$$

where

$$s := 1 - {}^P d_x^2 c_r^2 > 0, \quad (6.29)$$

after some work, the following inequality is obtained

$$(k/2)\{s + \text{sqrt}(s)\text{sqrt}[s + (4c_r^2/k)]\} > 0, \quad (6.30)$$

which means that the inequality (6.25) holds.

In order to prove that inequality (6.26) holds, using the expression for $\det(M_c)$ from the last coefficient of (6.24), and taking into account that $s > 0$, one has to prove that

$$m_{12}m_{21} < 0. \quad (6.31)$$

Using the first equation from (6.20) for y_{es} , and the expressions for m_{12} and m_{21} given in (6.23), after some work, the following inequality is obtained

$$(1/s)[1 + (4c_r^2/k)] - 1 > 0, \quad (6.32)$$

which means that the inequality (6.24) holds.

Based on (6.30) and (6.32), it follows that $Re\lambda_i < 0$, ($i=1, 2$). Application of the Lyapunov's indirect method [13] indicates that the equilibrium point $[y_{es}, \psi_{es}]^T$ is locally asymptotically stable point for the nonlinear system (6.19).

6.3.2 Straight-Line Tracking

In particular, in the case of straight line following ($c_r=0$ in (6.12)), and applying the feedback control law given by (6.18), the resulting time-invariant closed-loop system (6.19), ($v_{Px} = cte > 0$), has the form

$$\dot{e} = \begin{bmatrix} \dot{y}_e \\ \dot{\psi}_e \end{bmatrix} = v_{Px}\eta_c(e), \quad (6.33)$$

where the vector e is given by (6.13) and the vector function η_c is in the form

$$\eta_c = \begin{bmatrix} -{}^P d_x k y_e (1 - k y_e^2) \tan \psi_e \\ -k y_e \end{bmatrix}.$$

The origin (0, 0) becomes an equilibrium point for the nonlinear system (6.33). After linearizing the system (6.33) around the origin, is obtained

$$\dot{e} = M_s e, \quad (6.34)$$

where the matrix $M_s \in \mathfrak{R}^{2 \times 2}$ has the form

$$M_s = \begin{bmatrix} -k {}^P d_x v_{Px} & v_{Px} \\ -k v_{Px} & 0 \end{bmatrix}. \quad (6.35)$$

A Lyapunov function for the linear system (6.34) is found by taking a positive definite matrix $Q \in R^{2 \times 2}$ of the form ($v_{Px} = cte > 0$)

$$Q = \begin{bmatrix} v_{Px} & 0 \\ 0 & v_{Px} \end{bmatrix} \quad (6.36)$$

and solving for $P \in \mathfrak{R}^{2 \times 2}$ the Lyapunov equation

$$P M_s + M_s^T P = -Q. \quad (6.37)$$

The unique solution of the matrix Eq. (6.37) for P is obtained as follows

$$P = \begin{bmatrix} \frac{1+k}{2k^P d_x} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2k^{2P} d_x} (1+k+k^2 P d_x) \end{bmatrix}. \quad (6.38)$$

The symmetric matrix P is positive definite since its leading principal minors are positive ($k = cte > 0$; $^P d_x = cte > 0$)

$$\begin{aligned} \det\left(\frac{1+k}{2k^P d_x}\right) &> 0, \\ \det\left[\begin{array}{cc} \frac{1+k}{2k^P d_x} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2k^{2P} d_x} (1+k+k^2 P d_x) \end{array}\right] \\ &= \frac{1+k}{k} \left(\frac{1}{k^{2P} d_x^2} + \frac{1}{4k^P d_x^2} + \frac{1}{4} \right) - \frac{1}{4} > 0. \end{aligned} \quad (6.39)$$

Hence, M_c is stability matrix (i.e., $Re\lambda_i < 0$ for the eigenvalues of M_c) for the system (6.34), [20]. Since the robot velocity is assumed to be constant and strictly positive, it follows that using control (6.18) local asymptotic stability of the nonlinear system (6.33) is achieved when the control (6.18) is applied.

6.4 Simulation and Experimental Results

Numerical simulation tests using MATLAB and experiments are carried out in order to validate the proposed path tracking control. The look-ahead distance for the robot reference point R (Fig. 6.1) was chosen to be $^P d_x = 0.3$ m. The forward robot velocity was chosen to be $v_{P,x} = 0.3$ m/s and the controller gain in feedback control given by (6.18) was $k_y = 10$.

For the first simulation, a circular reference path of radius 1 m was assigned. In the first test in circular path following, the initial conditions were chosen to be $e = [y_e(0) \ \psi_e(0)]^T = [-0.3 \ -0.3]^T$. The path drawn by the robot guide point R and evolution in time of the path error coordinates are depicted, respectively, in Fig. 6.4a, b.

For the second simulation test in circular path following, the initial conditions were $e = [y_e(0) \ \psi_e(0)]^T = [0.3 \ 0.3]^T$. The path drawn by the robot guide point R and evolution in time of the path error coordinates is depicted in Fig. 6.4c, d, respectively. As seen from Fig. 6.4b, d, the error coordinates $y_e(t)$ and $\psi_e(t)$ asymptotically tend to the steady-state values given by (6.20), which is in conformity with the stability analysis presented in Sect. 6.3. The steady-state lateral error y_{es} in the both cases is equal to 0.095 m.

For the straight-line path tracking simulation test, the initial conditions were chosen to be $e = [y_e(0) \ \psi_e(0)]^T = [0.25 \ 0.1]^T$. The path drawn by the robot guide point R and evolution in time of the path error coordinates are depicted, respectively, in Fig. 6.5a, b, respectively. The results from the simulation confirmed the analyt-

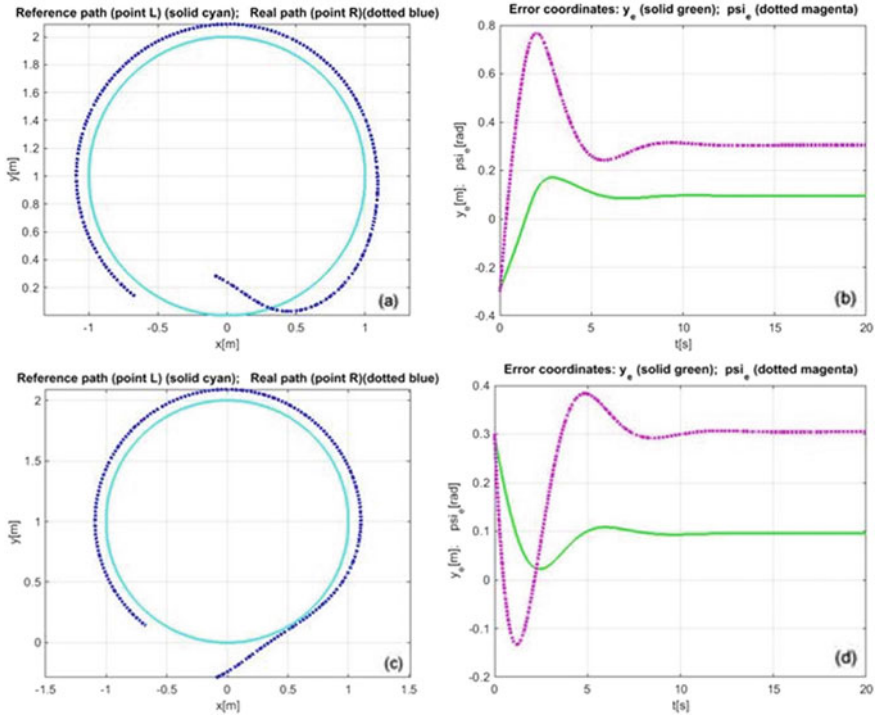


Fig. 6.4 Following a circular path: **a** and **c** planar path drawn by the reference point R of the robot (blue dotted line) and reference circular path (cyan solid line); **b** and **d** evolution in time of the error coordinates y_e (green solid line) and ψ_e (magenta dotted line)

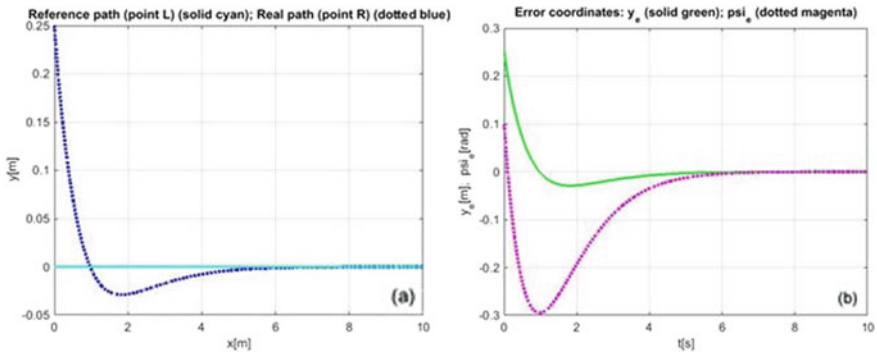


Fig. 6.5 Following a straight-line path: **a** Planar path drawn by the reference point R of the robot (blue dotted line) and reference circular path (cyan solid line); **b** Evolution in time of the error coordinates y_e (green solid line) and ψ_e (magenta dotted line)



Fig. 6.6 Experiments in circular line tracking



Fig. 6.7 Experiments in straight-line tracking

ical result obtained in the previous section for straight-line tracking, that the error coordinates y_e and $\psi_e(t)$ asymptotically tend to zero.

Experiments were carried out for tracking a circular and straight-line path using a differential-drive mobile robot Pioneer-3DX equipped with an onboard low cost camera (640×480 pixels). For the experiments, the velocity of the robot was set to $v_{P,x}=0.3$ m/s. As seen from Fig. 6.6, the robot was able to follow a circular path with small steady-state lateral error according to (6.20) and without error for straight-line tracking (Fig. 6.7).

6.5 Conclusion

In this paper, a vision-based line following controller for unicycle mobile robots was presented. Using a look-ahead approach, a simple and effective feedback control, which achieves local asymptotic stability of the nonlinear closed-loop system for a circular path with unknown curvature, as well and a straight-line path, was designed and analyzed using Lyapunov stability theory. Simulation and experimental results confirm the validity of the designed vision-based control scheme to perform curved path following in clutter environments with a quite good accuracy.

Acknowledgements This work was supported by the National Science Fund at the Ministry of Education and Science, Republic of Bulgaria, within the project KP-06-H27/16 “Development of Efficient Methods and Algorithms for Tensor-based Processing and Analysis of Multidimensional Images with Application in Interdisciplinary Areas.”

References

1. Liaqat, A., Hutabarat, W., Tiwari, D., Tinkler, L.: Autonomous mobile robots in manufacturing: highway code development, simulation, and testing. *Int. J. Adv. Manuf. Technol.* **104**, 4617–4628 (2019)
2. Sinyukov, D., Padir, T.: Adaptive motion control for a differentially driven semi-autonomous wheelchair platform. In: *International Conference on Advanced Robotics*, pp. 288–294. IEEE, Istanbul, Turkey (2015)
3. Petrov, P., Nashashibi, F.: Planning and nonlinear adaptive control for an automated overtaking maneuver. In: *14th International IEEE Conference on Intelligent Transportation Systems*, pp. 662–667. IEEE, Washington, DC, USA (2011)
4. Palacios, A., Sanchez, A., Bedolla, A., Cordero, J.: The random exploration graph for optimal exploration of unknown environments. *Int. J. Adv. Rob. Syst.* **14**(1), 1–11 (2017)
5. Rahmat, M., Hudha, K., Idris, A., Amer, N.: Sliding mode control of target tracking system for two degrees of freedom gun turret model. *Adv. Military Technol.* **11**(1), 13–28 (2016)
6. Calvo, C., Villacorta-Atienza, J., Mironov, V., Gallego, V., Makarov, V.: Waves in isotropic totalistic cellular automata: application to real-time robot navigation. *Adv. Complex Syst.* **19**(4), 1–18 (2016)
7. Petrov, P., Nashashibi, F.: Saturated feedback control for an automated parallel parking assist system. In: *13th International Conference on Control Automation Robotics and Vision*, pp. 577–582. IEEE, Singapore (2014)
8. Ibrahim, F., Abouelsoud, A., Elbab, A., Ogata, T.: Path following algorithm for skid-steering mobile robot based on adaptive discontinuous posture control. *Adv. Robot.* **33**(9), 439–453 (2019)
9. Emam, M., Fakharian, A.: Solving path following problem for car-like robot in the presence of sliding effect via LMI formulation. *J. Comput. Rob.* **10**(2), 11–22 (2017)
10. De Luca, A., Oriolo, G., Vendittelli, M.: Control of wheeled mobile robots: an experimental overview. In: Nicosia, S., Siciliano, B., Bicchi, A., Valigi, P. (eds.), *LNCIS*, vol. 270, pp. 181–226. Springer, Berlin, Heidelberg (2001)
11. Plaskonka, J.: Different kinematic path following controllers for a wheeled mobile robot of (2, 0) type. *J. Intell. Rob. Syst.* **77**, 481–498 (2015)
12. Petrov, P., Kralov, I.: A look-ahead approach to mobile robot path tracking based on distance-only measurements. *AIP Conf. Proc.* **2172**, 110005 (2019)
13. Hutchinson, S., Hager, G., Corke, P.: A tutorial on visual servo control. *IEEE Trans. Rob. Autom.* **12**(6), 651–670 (1996)
14. Cherubini, A., Chaumette, F., Oriolo, G.: A position-based visual servoing scheme for following paths with nonholonomic mobile robots. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1648–1654. IEEE, Nice, France (2008)
15. Petrov, P., Georgieva, V.: Vision-based line tracking control for nonholonomic differential-drive mobile robots. In: *9th National Conference with International Participation, ELECTRONICA 2018*. IEEE, Sofia, Bulgaria (2018)
16. Kim, S., Oh, S.: Hybrid position and image based visual servoing for mobile robots. *J. Intell. Fuzzy Syst.* **18**, 73–82 (2007)
17. Corke, P.: Mobile robot navigation as a planar visual servoing problem. In: Jarvis, R., Zelinsky, A. (eds.), *Robotics Research, STAR* vol. 6, pp. 361–372. Springer, Berlin Heidelberg (2003)
18. Kountchev, R., Mironov, R., Kountcheva, R.: Hierarchical cubical tensor decomposition through low complexity orthogonal transforms. *Symmetry* **12**(5), 1–17 (2020)
19. Mariottini, G., Oriolo, G., Prattichizzo, D.: Image-based visual servoing for nonholonomic mobile robots using epipolar geometry. *IEEE Trans. Rob.* **23**(1), 87–100 (2007)
20. Khalil, H.: *Nonlinear systems*, 3rd edn. Pearson, USA (2002)

Chapter 7

Markerless 3D Virtual Glasses Try-On System



Mariofanna Milanova and Fatima Aldaeif

Abstract This paper presents the implementation of a markerless mobile augmented reality application called a virtual eye glasses try-on system. The system first detects and tracks human face and eyes. Then, the system overlays the 3D virtual glasses over the face in real time. This system helps the consumer to select any style of glasses available on the virtual space saving both time and effort when shopping online. A method based on local-invariant descriptors is implemented to extract image feature points for eyes detection and tracking. A new approach for camera pose estimation is proposed to augment real images with virtual graphics. Experiments are conducted using Haar cascade and speeded up robust features (SURF) cascade. The system is optimized and adapted for a mobile architecture.

7.1 Introduction

Augmented reality (AR) makes use of computer vision and computer graphics techniques to merge virtual content into the real world. Ronald Azuma first outlined the features that a universal AR system should possess: (1) combination of real and virtual world; (2) interactivity and (3) three-dimensional representation of objects [1].

The implementation of AR in physical advertisement has grown rapidly and is being implemented in various industries such as automobile, food, game, engineering and many more. AR applications have the capability to add virtual clothing or apparel onto consumers' reflection, which they seem to "wear". Examples of this approach are "virtual dressing rooms" and "virtual mirrors" of brand names selling accessories such as sunglasses, jewelry or watches. The purpose of these applications is to enrich

M. Milanova (✉) · F. Aldaeif
University of Arkansas at Little Rock, Little Rock, AR, USA
e-mail: mgmilanova@ualr.edu

F. Aldaeif
e-mail: fxaldaeif@ualr.edu

customer shopping experiences, both in real world and online. Shoppers are able to share their choices, or “likes” through social media, and are often able to make their final purchase directly through the AR interface.

In order to strengthen competitiveness, professionals should focus on quality in the form of improving usability and design of the AR system. Using a mobile AR application, the users can get many snapshots with multiple virtual glasses at the same time allowing them to compare different glasses and designs.

7.1.1 Motivation

To meet consumers’ demands, a useful virtual try-on system should be both efficient and effective. Currently, there are three main categories of eyeglasses virtual try-on techniques. The first one is adding 2D glasses image onto user’s 2D facial image. This technique can only deal with frontal view of the face and cannot provide dynamic feedback for user’s action. Another technique is to pre-reconstruct a 3D model of a user’s head/face using the user’s images and then to fit 3D glasses model to this pre-reconstructed 3D head/face model. The advantage of this method is its good fitting result. However, the reconstruction of human face is challenging and requires manual editing to have relisting model. This technique is not working in real time. The third approach is to superimpose virtual glasses 3D model onto a live face image sequence.

The real face image sequence with virtual glasses looks like a mirror using any mobile camera. In this technique, the fitting quality depends on the accuracy of eye detection/tracking algorithm and on the accuracy of a head pose estimation algorithm. The problem with this method is when the user rotates his/her head, and some parts of the virtual glasses are occluded.

7.1.2 Related Work

Huang et al. proposed a cascaded AdaBoost classifier to detect the eyes and then the glasses image fits to the eye area using affine transforms [2]. This method is simple but can only handle a front view. Human-centric design of glasses based on 3D glasses model is presented in [3]. The tracking procedure of the head movement in this system needs a 3D scan of user’s head and after that the user has to wait 5–10 min until a training is done. The drawback of these methods is that the process of reconstructing a realistic 3D head model is challenging and time-consuming.

There are few commercial virtual glasses try-on systems already in use. Ray-Ban has developed a virtual glasses try-on system. Their system requires a frontal snapshot of the user, on which several features points need to be marked manually. The existing systems required manual initialization in order to obtain several pairs

of 2D–3D correspondences between the reference image and a corresponding 3D model.

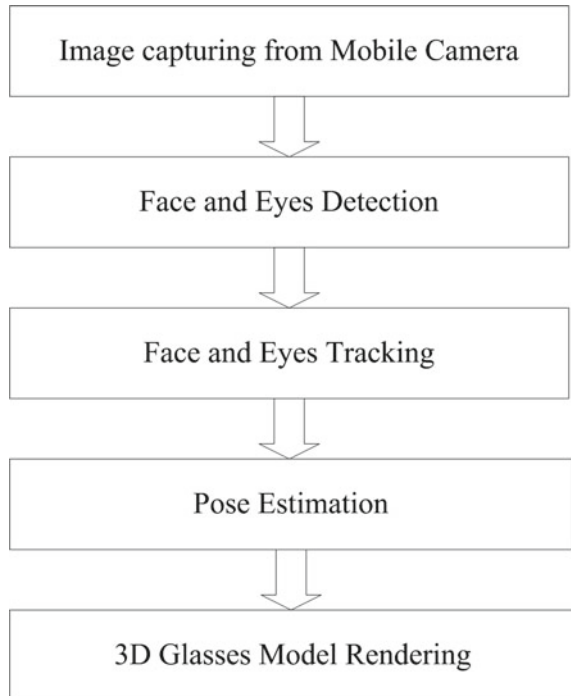
In this paper, we propose the developed mixed reality system for virtual glasses try-on which enables the user to view different virtual models fitted on his/her face in real time. Virtual glasses are automatically scaled and oriented to the size and orientations of the user’s face as far as the user is within the camera’s view. To enhance the efficiency, we tested different algorithms for eyes detection/tracing. We compared the results using Haar cascade and SURF cascade for eyes detection and tracking using database of videos of faces in challenging uncontrolled conditions (“YouTube” database).

The rest of the paper is organized as follows: Sect. 7.2 presents an overview of our virtual mirror glasses try-on system. Section 7.3 presents faces detection methods. In Sect. 7.4, a pose estimation method for building the AR system is presented. Section 7.5 presents experimental results, and Sect. 7.6 is the conclusion.

7.2 Proposed System

The flowchart of the developed mobile AR try-on system is shown in Fig. 7.1.

Fig. 7.1 Try-on AR system flowchart



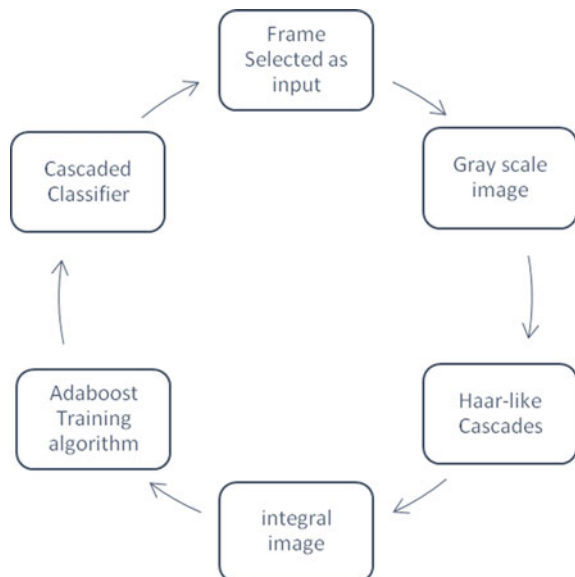
7.3 Face Detection

The field of face detection has made significant progress in the past decade. This minor task for human beings is very challenging for computers. The difficulties associated with face detection are coming from lighting conditions, scale variations, location, orientation, pose, facial expression, occlusions, etc. In [4], a survey of recent advances in face detection is presented. In general, face detection techniques can be divided into four main categories: knowledge-based methods, feature-based approach, template matching methods and appearance-based methods. The appearance-based methods had been showing superior performance to others. Appearance-based methods learn face models from a set of representative training face images to perform detection. There are two key issues in this process, including what features to extract and which learning algorithm to apply. In this section, two techniques for face detection are presented: Haar cascade and SURF cascade.

7.3.1 *Detecting user's Face and Eyes with Viola–Jones Method*

The Viola–Jones (VJ) face detector contains three main ideas that make it possible for real-time object detection: the integral image for efficient Haar feature extraction, the boosting algorithm for ensemble weak classifiers and the attentional cascade structure for fast negative rejection [5] (see Fig. 7.2).

Fig. 7.2 Viola–Jones algorithm



Integral Image

Haar features depend on similar aspects for mankind face: as it is known, all mankind have the similar eyes location, size and intensity value characteristics. The integral image $T(u, v)$ at the location u, v contains the sum of the pixels above and to the left of u', v' defined as follows:

$$T(u, v) = \sum_{u' \leq u, v' \leq v} I(u', v') \tag{7.1}$$

The image can be computed in one pass over the original image using the following pair of relations:

$$S(u, v) = S(u, v - 1) + T(u, v) \tag{7.2}$$

$$T(u, v) = T(u - 1, v) + S(u, v) \tag{7.3}$$

where $S(u, v)$ is the cumulative row sum, $S(u, 1) = 0$ and $T(1, v) = 0$.

Rectangular features can be calculated in steady time, which gives them huge fast features over their more complex relatives. Averages of group of pixels were found to determine patterns like these, as shown in Fig. 7.3.

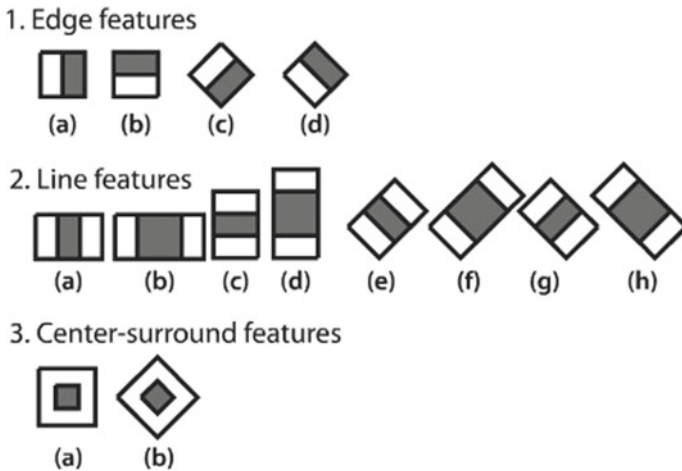


Fig. 7.3 Set of Haar-like features

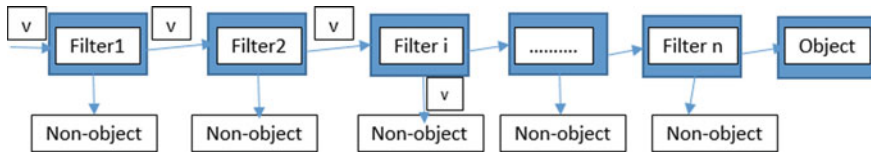


Fig. 7.4 Haar cascade

AdaBoost Training Algorithm

Using a face recognition dataset which includes a set of positive (faces and eyes) and negative (non-faces and eyes) sample images, AdaBoost algorithm will train the classification function.

Haar Cascade

As shown in Fig. 7.4, the filter will refuse non-face windows and will let face windows past to the next layer of the cascade. If an object is considered as “face,” in all the filters layers, then it will be the final result.

The Viola–Jones framework has several limitations. First, the feature pool of Haar-like features is very high which is usually in hundreds of thousands level for a typical 20×20 detection template. This directs to extremely large feature search space for weak classifier learning. Second, Haar features have difficulty to handle variations due to pose and illumination. Some researchers replace Haar feature with local binary patterns (LBP), which are more robust to illumination variations. Third, the attentional cascade is trained based on two conflicted criteria: false-positive rate (FPR) f_i and hit rate for detection rate d_i true-positive rate (TPR). The VJ framework suggests $\min \text{TPR} = 0.995$ and $\max \text{FPR} = 0.5$ during the training procedure. To reach $\text{FPR} = 0.5$ is easy in early stages, but TPR is not converged simultaneously.

SURF Cascade

The SURF-based representation is on the basis of the “speeded up robust features,” which has the properties of scale invariance, rotation invariance and computationally efficiency. SURF features do not only detect points of interest, but also propose a method for creating local-invariant descriptors. These descriptors are working under a variety of disturbing conditions like changes in scale, rotation, illumination, view-points or nose. This invariance is an important criterion for mobile systems where the environment conditions are unpredictable.

In our study, we have implemented an alternative SURF cascade [6]. The proposed approach contains four ingredients: (a) SURF features for local patch description, (b) logistic regression-based weak classifier for each patch, (c) boosting ensemble of weak classifiers for each stage and (d) area under ROC curve (AUC)-based cascade learning algorithm.

(a) SURF features for local patch description.

Features: SURF

- 2×2 cell of patch.
- Each cell is an eight-dimensional vector.
 - Sum of $dx, |dx|$ when $dy \geq 0$.
 - Sum of $dx, |dx|$ when $dy < 0$.
 - Sum of $dy, |dy|$ when $dx \geq 0$.
 - Sum of $dy, |dy|$ when $dx < 0$.
- Total is $2 \times 2 \times 8 = 32$ dim feature vector.
- Eight-channel integral images.

Feature Pool.

- In a 40×40 face detection template.
- Slide the patch (x, y, w, h) with fixed step = four pixels.
- Each cell at least 8×8 pixels, w or h at least 16 pixels.
- With 1:1, 1:2, 2:3... aspect ratio (w/h).
- Totally 396 local SURF patches.

(b) Logistic regression-based weak classifier for each patch

Given SURF feature x over local patch, logistic regression defines a probability model (7.4):

$$P(y = \pm 1|x, w) = \frac{1}{1 + \exp(-y(w^T x + b))} \quad (7.4)$$

where $y = 1$ for face samples, $y = -1$ for non-face samples, w is a weight vector for the model, and b is a bias term.

(c) Boosting ensemble of weak classifiers for each stage

Suppose there are N training samples and K possible local patches represented by d -dimensional (=32) SURF feature x , each stage is a boosting learning procedure with logistic regression as weak classifier. Given weak classifiers $h_t(x)$, the strong classifier is (7.5)

$$H^T(x) = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (7.5)$$

In the t -th boosting round, K logistic regression models are built for each local patch in parallel from the boosting sampled.

(d) Area under ROC curve (AUC)-based cascade learning algorithm

Each model $h_k(x)$ is tested in combination with model of previous $t - 1$ rounds. Each tested model will produce an AUC score $J(H^{t-1}(x) + h_k(x))$. In the final step, the model which produces the highest AUC score is selected (7.6):

$$H^t(x) = \arg_{H^k, k=1, K} \max J(H^k = H^{t-1}(x) + h_k(x)) \tag{7.6}$$

7.4 Pose Estimation for AR Tracking

There are two subsets of camera parameters that are used to determine the relationship between coordinate systems: intrinsic and extrinsic parameters (Fig. 7.5).

The intrinsic parameters are those related to the internal geometry of a physical camera. There are (1) the focal length, (2) the location of the image center in pixel space and (3) the pixel size in the horizontal and vertical directions.

The link between image coordinates (x_{im}, y_{im}) , in pixels, with the respective coordinates (x, y) in the camera coordinate system is

$$x = -(x_{im} - o_x)s_x$$

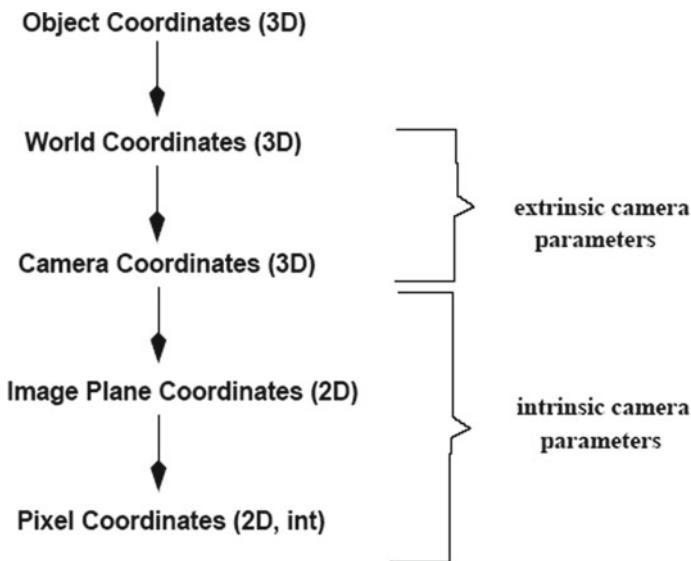


Fig. 7.5: 3D–2D coordinates transformation

$$y = -(y_{\text{im}} - o_y)s_y \quad (7.7)$$

where (o_x, o_y) define the pixel coordinates of the principal point, and (s_x, s_y) define the size of the pixels (in millimeters), in horizontal and vertical directions.

The extrinsic parameters are concerned with external properties of the camera, such as position and orientation information. These parameters are (1) the 3×3 rotation matrix R that brings the coordinate system of the object and camera coordinate system on top of one another and (2) the 3D translation vector T describing the relative positions of the origins of the two coordinate systems.

In other words, if we have a point P_w in world coordinates, then the same point in the camera coordinates p_c would be

$$p_c = RP_w + T \quad (7.8)$$

Using perspective projection and Eqs. (7.7) and (7.8), we get

$$\begin{aligned} -(x_{\text{im}} - o_x)s_x &= f \frac{R_1^T (P_w - T)}{R_3^T (P_w - T)} \\ -(y_{\text{im}} - o_y)s_y &= f \frac{R_2^T (P_w - T)}{R_3^T (P_w - T)} \end{aligned} \quad (7.9)$$

where R_i , $i = 1, 2, 3$, denotes the 3D vector formed by the i -th row of the matrix R .

Inserting the equations in homogeneous matrix form the projection, we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = M_{\text{int}} M_{\text{ext}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (7.10)$$

where $x_1/x_3 = x_{\text{im}}$, $x_2/x_3 = y_{\text{im}}$, M_{int} is intrinsic matrix, and M_{ext} is extrinsic matrix.

In the rendering step, Blender software is used to design the 3D glasses model, which is exported as an OpenGL ES code using tools like mtl2opengl.

7.5 Experimental Results

Comparison of HAAR and SURF in eye detection from a video on android is as follows.

7.5.1 Dataset Collection

We used the 5749 names of subjects included in the LFW dataset to search YouTube for videos of these same individuals [7]. The top six results for each query were downloaded. We minimized the number of duplicate videos by considering two videos' names with edit distance less than three to be duplicates. Downloaded videos were then split to frames at 24 fps. We detected the eyes in these videos using the Haar (Viola–Jones) face detector. Automatic screening was performed to eliminate detections of less than 48 consecutive frames, where detections were considered consecutive if the Euclidean distance between their detected centers was less than ten pixels. This process ensures that the videos contain stable detections and are long enough to provide useful information for various recognition algorithms. Finally, the remaining videos were manually verified to ensure that (a) the videos are correctly labeled by subject, (b) are not semi-static, still-image slide shows and (c) no identical videos are included in the database.

The screening process reduced the original set of videos from the 18,899 originally downloaded (3345 individuals) to 3425 videos of 1595 subjects. An average of 2.15 videos is available for each subject. The shortest clip duration is 48 frames, the longest clip is 6070 frames, and the average length of a video clip is 181.3 frames.

All video frames are encoded using several well-established, face image descriptors. Specifically, we consider the face detector output in each frame. The bounding box around the face is expanded by 2.2 of its original size and cropped from the frame. The result is then resized to standard dimensions of 200×200 pixels. We then crop the image again, leaving 100×100 pixels centered on the face. Images are then converted into grayscale. The eyes are detected from these images of faces (Fig. 7.6).



Fig. 7.6 Images from the YouTube database

Table 7.1

Feature extracted using	Hits	Misses	Speed (on phone) (s)
Haar–Viola–Jones	732	468	132
SURF cascade	936	264	165

7.5.2 Experimental Setup

An android application was developed to test the accuracy of the above classifiers. The dataset for this classification was used from YouTube videos frame dataset. Every second's video frame was converted into an image, and it was tested using Haar and SURF. Haar cascade classifier is already present in OpenCV for eye detection (using Viola-Jones framework). The SURF cascade however extracted 100 keypoints per image and was tested using 20 negative and 80 positive samples. The following results were obtained:

Hardware Setup: Samsung S4 Phone, Android 5.0.1, Quad-core 1.6 GHz Cortex-A15 and quad-core 1.2 GHz Cortex-A7, PowerVR SGX544MP3 GPU, 2 GB RAM (Table 7.1).

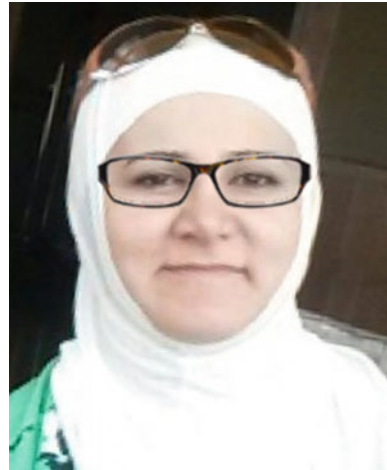
7.5.3 Conclusion

In detection of the eyes from video frames taken from YouTube videos, SURF cascade works better than Haar (Viola–Jones framework) but it takes a slightly longer time. This implementation shows Haar is 23% faster than SURF cascade in face and eye detection. SURF cascade is 22% better at eye detection in the current YouTube faces dataset as compared to Haar–Viola–Jones over the same dataset.

We implemented our application on an android operating system. We used the android SDK version 7 under Ubuntu 9.10 and as IDE: Eclipse and Android Development Tools Plug-in (ADT). Android application program is written by a developer inside a special platform called Eclipse. It emulates the different android devices and can run the application on a real android device. The application is written in Java language and can be run on different versions of android devices.

The virtual try-on algorithms were tested in a mobile AR-prototype application using android device which makes the mobile screen working as a mirror. The developed algorithm was able to calculate the central pose and display a 3D glass model over the human eyes (Fig. 7.7).

Fig. 7.7 Rendering the 3D glasses model in real time



7.6 Conclusions

In this paper, we introduced and implemented a markerless AR application: virtual try-on glasses. The most important part in an augmented reality system is the estimation of the camera poses which use the application of the homography algorithms on the extracted features. Tracking features and camera pose estimation algorithms were then tested in android phones. The SURF features and SURF cascade have proved accuracy of the detection of human eyes in uncontrolled conditions.

Acknowledgements This work was supported by the National Science Fund of Bulgaria: KP-06-H27/16 *Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas* and NOKIA Corporate University Donation number NSN FI (85) 1198342 MCA.

References

1. Azuma R.: A survey of augmented reality. In: *Teleoperators and Virtual Environments*, vol. 6, pp. 355–385 (August 1997)
2. Huang, W., Hsieh, C., Yeh, J.: Vision—based virtual eyeglasses fitting system. *Proc. ISCE* **2013**, 45–46 (2013)
3. Huang, S., Yang, Y., Chu, C.: Human centric design personalization of 3D glasses frame in markerless augmented reality. *Adv. Eng. Inform.* **26**(1), 35–45 (2012)
4. Zang, C., Zhang, Z.: A survey of recent advances in face detection. In: *Technical Report Microsoft Research*, June 2010
5. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. *Proc. CVPR*, 511–518 (2001)

6. Li, J., Zhang, Y.: Learning SURF cascade for fast and accurate object detection. Proc. CVPR, 3468–3475 (2013)
7. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matches background similarity. Proc. CVPR, 529–534 (2011)

Chapter 8

Copy–Move Forgery Detection by Using Key-Point-Based Harris Features and CLA Clustering



Kavita Rathi and Parvinder Singh

Abstract Images can easily be manipulated without any visual marks to the naked human eye with massive improvements in image manipulation software. This tampering is the main propelling force for the need of better image forensics such that field is known as image forgery detection. Any digital image with regions where the image contents are identical is said to have copy–move forgery (CMF). Copy–move forgery is performed to improve the visual features or to cover the underlying truth in the image. Many algorithms have been used for CMF detection, and this work is about improved key-point and clustering-based CMF detection scheme. The proposed scheme combines the efficiency of a key-point-based scheme and clustering of these key points to further improve the results. Modified Harris operator-based key-point detection algorithm with clustering using local gravitation is utilized for key-points selection. The average accuracy, PSNR and SSIM rates are used to evaluate the performance of the proposed algorithm with scale-invariant feature transform (SIFT), which is another state-of-the-art key-point algorithm. The paper concluded with the efficiency of the key-point-based scheme.

8.1 Introduction

Image forensics is a vast field used to verify the images to ascertain credibility and authenticity by using various computation approaches [1, 2]. Image forensics is attracting a lot of attention due to its possible applications in various domains. There are various methods in image forgery detection, which can be categorized as active (copy–move forgery detection) and passive (blind forgery detection) [3]. The copy–move forgery detection algorithms are concerned with revealing the forgery

K. Rathi (✉) · P. Singh
Deenbandhu Chhotu, Ram University of Science and Technology, Murthal 131039, India
e-mail: kavita1217@gmail.com

P. Singh
e-mail: Parvindersingh.cse@dcrustrm.org

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 216,
https://doi.org/10.1007/978-981-33-4676-5_8

113



Fig. 8.1 Copy-move image forgery example [5]

used for hiding the underlying truth or to improve the visual features in the image. However, the process which alters the visual features of the image changes original metadata in the image [4].

A large number of tempered images exist now; one of such, which is example of copy-move forgery from CMFD dataset [5] is presented in Fig. 8.1. A large number of image forgery detection algorithms have been developed over the years, and the efficiency of the existing image forgery detection algorithm is low.

The efficiency of these algorithms is improved with the advent of key-point-based feature detection such as SIFT [6] and speeded-up robust features (SURF) [7] algorithms. The main algorithms for active forgery detection use key-point-based features detection, such as SIFT, algorithms combine with some clustering algorithms, such as K-means. This work is about the improvement of the existing state-of-the-art active image forensics with improved key-point mechanism and improved clustering procedure. For the key points, we have used the Harris key-point detector to improve the computation efficiency and the CLA clustering to improve the clustering procedure.

8.2 Related Works

The existing research in image forensics mainly focuses on improving the feature extraction and clustering stages [6]. Image forensics uses combined methods involving different kinds of key point and feature descriptors; one such example is presented in [7], where Harris corner and SIFT descriptor are used. The key points and feature vectors can be extracted in different color spaces; one of such example of extracting the feature by using the SURF in the opponent color space is presented in [8]. Adaptive non-maximal suppression algorithm is used for smooth tampered regions to select Harris corners and extracted DAISY descriptors in [9]. A two-stage

detection method to detect tempering by using Harris corners and extracted multi-support regions order-based gradient histogram (MROGH) and hue histogram (HH) descriptors is presented in [10]. Both methods using extracted DAISY descriptor and extracted MROGH and HH descriptors have poor owing to the adoption of the Harris corner and poor robustness to scaling. Some methods that integrated block-based and key-point-based methods have been proposed in the last two years. A method based on multi-scale analysis and voting processes determining the range of suspicious feature regions for matching and clustering by using SURF and then block-based method to detect tampered regions in the multi-scale space is presented in [11]. Adaptive over-segmentation and the forgery region extraction algorithm are presented in [12] to detect tampered regions. Feature extraction by DCT, clustering by using k-means and feature matching done by radix sort in [13]. Guaranteed outlier removal is clubbed with key-point-based algorithm to improve the efficiency and robustness in [14]. Techniques for uniformly scattered key-points are adopted with Laplace of Gaussian in [15] for improving the results in smooth areas. A combination of image segmentation and iterative nearest neighborhood methods for detecting suspected tempering and then to gradually improve the detection precision is used in [16]; however, the computational cost of this method is prohibitive for large images.

8.3 SIFT K-Means Algorithm

The SIFT-based image forgery detection approach uses the key-point detector. The scale space S of image $I(x, y)$ is $S(x, y, \sigma)$

$$S(x, y, \sigma) = I(x, y) * G(x, y, \sigma) \quad (8.1)$$

where $G(x, y, \sigma)$ is variable scale Gaussian

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (8.2)$$

where σ is standard deviation of $G(x, y, \sigma)$.

Hessian matrix-based Hessian operator is used for key-point detection. The Hessian matrix $H(x, s)$ for x at scale s for a given image point $x = (x, y)$ in the image $I(x, y)$ is defined by its Laplacian of Gaussian (LoG) as

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (8.3)$$

where $L_{xx}(x, \sigma)$ is the convolution of the Laplacian of Gaussian (LoG) which is the second-order derivative $\frac{\partial^2 G}{\partial x^2}$ in point x for image $I(x, y)$, and similarly for $L_{xy}(x, \sigma)$, and $L_{yy}(x, \sigma)$.

To suppress the noise before using Laplace for edge detection

$$\Delta[G_\sigma(x, y) * f(x, y)] = [\Delta G_\sigma(x, y)] * f(x, y) = LoG * f(x, y) \quad (8.4)$$

The first equal is due to the fact

$$\begin{aligned} \frac{d}{dt}[h(t) * f(t)] &= \frac{d}{dt} \int f(\tau)h(t - \tau)d\tau \\ &= \int f(\tau) \frac{d}{dt}h(t - \tau)d\tau = f(t) * \frac{d}{dt}h(t) \end{aligned} \quad (8.5)$$

First, consider Laplacian of Gaussian (LoG)

$$\begin{aligned} \frac{\partial^2}{\partial x^2} \left(\frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \right) &= \frac{x^2}{\sigma^4} e^{-\frac{(x^2+y^2)}{2\sigma^2}} - \frac{1}{\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \\ &= \frac{x^2 - \sigma^2}{\sigma^4} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \end{aligned} \quad (8.6)$$

For simplicity, normalizing coefficient $1/\sqrt{2\pi\sigma^2}$ is omitted and the convolution of Laplace of Gaussian (LoG) is done by using Eq. 8.7.

$$\begin{aligned} \Delta G_\sigma(x, y) &= \frac{\partial^2}{\partial x^2} G_\sigma(x, y) + \frac{\partial^2}{\partial y^2} G_\sigma(x, y) \\ &= \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \end{aligned} \quad (8.7)$$

All key points are checked for the presence of local extreme, i.e., either lowest or highest. The extreme key points are used to localize the key-point localization edge. Low contrast points and poorly localized points along edges are removed to discard noise and instability of local minima maxima points, which is done by discarding the functional value of $L(x^\wedge)$ above a threshold (usually 0.5). The functional value of $L(x^\wedge)$ is computed by using equation

$$L(x^\wedge) = L + \frac{1}{2} \frac{\partial L^T}{\partial x}(x^\wedge) \quad (8.8)$$

where x^\wedge is the location of extremum and is determined by

$$x^\wedge = \frac{\partial^2 L^{-1}}{\partial x^2} \frac{\partial L}{\partial x} \quad (8.9)$$

With scale-space function, $L(x, y, \sigma)$ can be shifted to compute the origin of the sample point by

$$L(x) = L + \frac{\partial L^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 L}{\partial x^2} x \quad (8.10)$$

where L and its derivatives are evaluated at the same point and $x = (x, y, \sigma)^T$ is offset from this point.

The key points are clustered by using the k-means clustering with minimization problem of two parts.

$$J = \sum_{i=1}^m \sum_{k=1}^K \omega_{ik} \|x^i - \mu_k\|^2 \quad (8.11)$$

where $\omega_{ik} = 1$ for all points belonging to cluster k ; else 0 with μ_k as centroid of cluster.

$$\frac{\partial J}{\partial \omega_{ik}} = \sum_{i=1}^m \sum_{k=1}^K \|x^i - \mu_k\|^2 \quad (8.12)$$

$$\omega_{ik} = \begin{cases} 1 & \text{if } k = \arg \min_j \|x^i - \mu_k\|^2 \\ 0 & \text{otherwise} \end{cases} \quad (8.13)$$

$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^m \omega_{ik} (x^i - \mu_k) = 0 \quad (8.14)$$

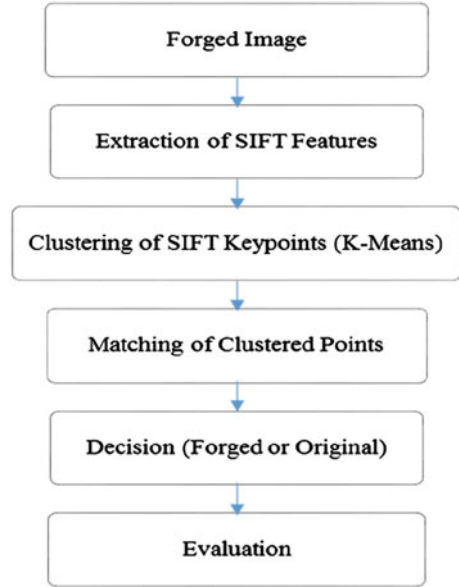
$$\mu_k = \frac{\sum_{i=1}^m \omega_{ik} x^i}{\sum_{i=1}^m \omega_{ik}} \quad (8.15)$$

The SIFT k-means algorithm involves.

- Transformation of RGB image to grayscale.
- Calculation of Hessian matrix of grayscale image.
- Key-point-based feature extraction using SIFT algorithm.
- Select strongest matching key points for matching forgery features.
- Key-point clustering by using k-means algorithm.
- Localize and mark each as clustered region with more than pre-specified group of pixels as forged.
- Show forged regions.

The flowchart of the proposed SIFT k-means algorithm for the detection of tempering is presented in Fig. 8.2.

Fig. 8.2 SIFT k-means forgery detection algorithm



8.4 Proposed Key-Point-Based Harris CLA Clustering Scheme

The Harris key-point detection algorithm works with extracting point features of the images. Harris algorithms utilize the work window ω to search key points (u, v) in any direction of the image. In order to improve the noise capabilities of the Harris algorithm, we have selected the Gaussian window

$$E(u, v) = \sum_{x,y} \omega(x, y) [I(x + u, y + v) - I(x, y)]^2 \quad (8.16)$$

where $\omega(x, y)$ is the window function and u, v are small displacements used to search, $I(x, y)$ is the intensity function of the image, and $I(x + u, y + v)$ is the shifted intensity of the image using Taylor expansions

$$E(u, v) \approx [uv]M \begin{bmatrix} u \\ v \end{bmatrix} \quad (8.17)$$

where M is given by

$$M = \sum_{x,y} \omega(x, y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_y I_x & I_y I_y \end{bmatrix} \quad (8.18)$$

where I_x and I_y are image derivatives.

CLA-based clustering is used to calculate the forged regions. The objective of CLA-based clustering is to group features into homogeneous classes by using a set of conditions. Each key point in the class is very similar in terms appearance, and different features usually belong to other groups. It is inspired by Newton's gravitational theory for reflecting the relation of a data point to its neighbors.

$$\vec{F}_{12} = G \frac{m_1 m_2}{D_{12}^2} \widehat{D}_{12} \quad (8.19)$$

where \vec{F}_{12} is the attractive force between two points of mass m_1 and m_2 located at distance D_{12} , G is gravitational constant, and \widehat{D}_{12} is unit vector specifying the direction of force.

The distance between the neighbors does not vary significantly in a local region, so,

$$\vec{F}_{12} = G m_1 m_2 \widehat{D}_{12} \quad (8.20)$$

Therefore, the total local resultant force at a point i from its k neighbors is

$$\vec{F}_i = G m_i \sum_{j=1}^k m_j \widehat{D}_{ij} \quad (8.21)$$

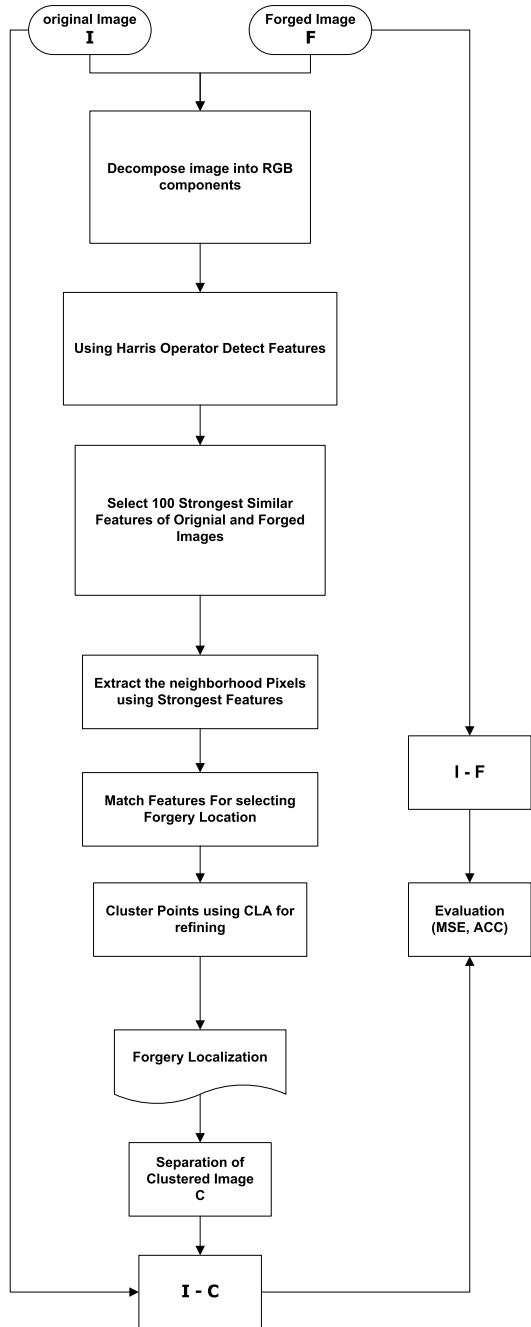
The number of cluster will be dependent on the threshold value of centrality with lower bound 0 and upper bound $m_i \sum_{j=1}^k m_j D_{ij}$.

The proposed key-point-based Harris CLA clustering scheme involves the following steps.

- Initialize the number of clusters to be used usually for image forgery, 8 to 10 points are used.
- Select input features for which clustering of key points will be done.
- Using the Euclidian similarity measure find distances
- Stepwise cluster/group the key points using hierarchical clustering involving a combination and division of the features into a set of clusters.
- If cluster group contains enough neighbors, then validate and select the cluster as a forged region.

The brief flowchart of the Harris CLA clustering scheme is presented in Fig. 8.3.

Fig. 8.3 Proposed key-point based-Harris CLA clustering scheme



8.5 Experimental Analysis

An Intel Core i3 central processing unit having 8 GB random-access memory and operating system raring at 2.4 GHz by using Windows 7 and 64-bit operating system is used for executing the proposed methodology. The proposed methodology is implemented in MATLAB 2016a. The dataset used for analyzing the SIFT k-means and the proposed algorithm is copy–move forgery. The copy–move forgery dataset [5] has four subsets $D0$, $D1$, $D2$, and $D3$ consisting of 1020 image with image size 1000×700 to 700×1000 pixels in JPEG format. It has attacks in the form of plain copy, scaling, and rotation with availability of ground truth. The dataset is freely downloadable [17]. A sample of the results by the proposed algorithm is presented below in Fig. 8.4.

Average accuracy, PSNR, and SSIM are used as metric for performance evaluation on various images. Accuracy (ACC) of the proportions of correctly identified predicted pixels is:

$$\text{Accuracy(ACC)} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (8.22)$$

PSNR is measurement of quality between the original forgery from ground truth and retrieved forgery in image. It is calculated in decibels by using the PSNR block from mathworks, which uses the mean square error (MSE). The mean square error is cumulative squared error between ground truth and retrieved forgery. The higher PSNR value represents the better quality of identified forgery [17]. For $M \times N$ size image I_1 and I_2 .

$$\text{MSE} = \frac{\sum_{M,N} [I_1(m, n) - I_2(m, n)]^2}{M \times N} \quad (8.23)$$

$$\text{PSNR} = 10 \log_{10} \left(\frac{R^2}{\text{MSE}} \right) \quad (8.24)$$



Fig. 8.4 Sample of CMFD by using the proposed Harris corner CLA clustering algorithm (original, forged, and detected from left to right)

R is image dependent, R for image having double-point floating data is 1, and R for unsigned 8-bit image is 255.

SSIM is the weighted combination of contrast, luminance, and structure [18].

$$SSIM(I_1, I_2) = [l(I_1, I_2)^\alpha \cdot c(I_1, I_2)^\beta \cdot s(I_1, I_2)^\gamma] \tag{8.25}$$

The SSIM with α , β , and γ weights as 1 is below.

$$SSIM(I_1, I_2) = \frac{(2\mu_{I_1}\mu_{I_2} + c_1)(2\sigma_{I_1I_2} + c_2)}{(\mu_{I_1}^2 + \mu_{I_2}^2 + c_1)(\sigma_{I_1}^2 + \sigma_{I_2}^2 + c_2)} \tag{8.26}$$

μ , σ^2 , and c are average, variance, and variables for stabilizing denominator.

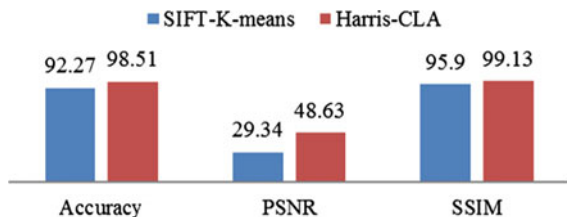
For experimental analysis, various images were considered for the evaluation. The average of each examining parameter was calculated. The average accuracy, PSNR and SSIM rates of SIFT k-means and key-point-based Harris CLA clustering scheme are presented in Table 8.1 and Fig. 8.5.

It is clear from the results that the proposed key-point-based Harris CLA clustering scheme performs better. The proposed key-point-based Harris CLA clustering scheme achieves average accuracy of 98.51%, which is 6.24% higher than SIFT k-means whose average accuracy performance is 92.27%. The average PSNR for the proposed key-point-based Harris CLA clustering scheme is 48.63%, which has still huge potential for improvement, but the proposed scheme outperformed which is about 65% more than SIFT k-means at 29.34. The average structural similarity (SSIM) for the proposed key-point-based Harris CLA clustering scheme is 99.90%, which is again about 4% higher than SIFT k-means whose average SSIM is 95.90%.

Table 8.1 Performance of SIFT k-means and the proposed key-point-based Harris CLA clustering scheme on accuracy, PSNR and SSIM metric

Metric	SIFT k-means	Proposed Harris CLA
Accuracy	92.27	98.51
PSNR	29.34	48.63
SSIM	95.90	99.13

Fig. 8.5 Results of image forgery detection performance of SIFT k-means and the proposed key-point-based Harris CLA clustering scheme on accuracy, PSNR, and SSIM metric



8.6 Conclusion and Future Scope

Copy–move forgery is performed to improve the visual features or to cover the underlying truth in the image. A new key-point-based Harris CLA clustering scheme is proposed which combines the efficiency of a key-point-based scheme and clustering of these key points to further produced better results. Experimental results established that the proposed key-point-based Harris CLA clustering scheme has improved the efficiency of detection process over the key-point detection scheme, namely SIFT k-means. In future work, parallel programming can be tried to improve the detection speed of the proposed scheme. Researchers can also focus on other types of image forgeries.

Acknowledgements This work is part of bilateral Indian-Bulgarian cooperation research project between Technical University of Sofia, Bulgaria, and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonapat, India, under the title “Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications,” financed by the Department of Science and Technology (DST), India, and the Ministry of Education and Science, Bulgaria.

References

1. Farid, H.: Image forgery detection. *IEEE Signal Process. Mag.* **26**(2), 16–25 (2009). <https://doi.org/10.1109/MSP.2008.931079>
2. Birajdar, G.K., Mankar, V.H.: Digital image forgery detection using passive techniques: a survey. *Digit. Investig.* **10**(3), 226–245 (2013). <https://doi.org/10.1016/j.diin.2013.04.007>
3. 3Qureshi, M.A., Deriche, M.: A review on copy move image forgery detection techniques. In: 2014 IEEE 11th International Multi-Conference Systems Signals Devices, SSD 2014, pp. 1–5 (2014). <https://doi.org/10.1109/SSD.2014.6808907>
4. Shivakumar, B.L., Baboo, S.S.: Detecting copy-move forgery in digital images: a survey and analysis of current methods. *Glob. J. Comput. Sci. Technol.* **10**(7), 61–65 (2010)
5. 5E. Ardizzone, Bruno, A., and Mazzola, G., Copy-Move Forgery Detection by Matching Triangles of Keypoints, *IEEE Trans. Inf. Forensics Secur.*, 10(10), 2084–2094, (2015), doi: <https://doi.org/10.1109/TIFS.2015.2445742>.
6. Shivakumar, B.L., Baboo, S.S.: Detection of region duplication forgery in digital images using SURF. *Int. J. Comput. Sci.* **8**(4), 199–205 (2011)
7. Shivakumar, B.L., Baboo, S.S.: Automated forensic method for copy move forgery detection based on Harris interest points and SIFT descriptors. *Int. J. Comput. Appl.* **27**(3), 9–17 (2011). <https://doi.org/10.5120/3283-4472>
8. Gong, J., Guo, J.: Image copy-move forgery detection using SURF in opponent color space. *Trans. Tianjin Univ.* **22**(2), 151–157 (2016). <https://doi.org/10.1007/s12209-016-2705-z>
9. Guo, J.M., Liu, Y.F., Wu, Z.J.: Duplication forgery detection using improved DAISY descriptor. *Expert Syst. Appl.* **40**(2), 707–714 (2013). <https://doi.org/10.1016/j.eswa.2012.08.002>
10. Yu, L., Han, Q., Niu, X.: Feature point-based copy-move forgery detection: covering the non-textured areas. *Multimed. Tools Appl.* **75**(2), 1159–1176 (2014). <https://doi.org/10.1007/s11042-014-2362-y>
11. Silva, E., Carvalho, T., Ferreira, A., Rocha, A.: Going deeper into copy-move forgery detection: exploring image telltales via multi-scale analysis and voting processes. *J. Vis. Commun. Image Represent.* **29**, 16–32 (2015). <https://doi.org/10.1016/j.jvcir.2015.01.016>

12. Pun, C.M., Yuan, X.C., Bi, X.L.: Image forgery detection using adaptive oversegmentation and feature point matching. *IEEE Trans. Inf. Forensics Secur.* **10**(8), 1705–1716 (2015). <https://doi.org/10.1109/TIFS.2015.2423261>
13. Parveen, A., Khan, Z.H., Ahmad, S.N.: Block-based copy–move image forgery detection using DCT. *Iran J. Comput. Sci.* **2**(2), 89–99 (2019). <https://doi.org/10.1007/s42044-019-00029-y>
14. Hegazi, A., Taha, A., Selim, M.M.: An improved copy-move forgery detection based on density-based clustering and guaranteed outlier removal. *J. King Saud Univ. Comput. Inf. Sci.* (2019). <https://doi.org/10.1016/j.jksuci.2019.07.007>
15. Wang, X.Y., Wang, C., Wang, L., Jiao, L.X., Yang, H.Y., Niu, P.P.: A fast and high accurate image copy-move forgery detection approach. *Multidimens. Syst. Signal Process.* **31**(3), 857–883 (2019). <https://doi.org/10.1007/s11045-019-00688-x>
16. Li, J., Li, X., Yang, B., Sun, X.: Segmentation-based image copy-move forgery detection scheme. *IEEE Trans. Inf. Forensics Secur.* **10**(3), 507–518 (2015). <https://doi.org/10.1109/TIFS.2014.2381872>
17. Mathworks, compute peak signal-to-noise ratio (PSNR) between images. MathWorks India (2020). <https://in.mathworks.com/help/vision/ref/psnr.html>. Accessed Jul 04 2020
18. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. *Conf. Record Asilomar Conf. Signals Syst. Comput.* **2**, 1398–1402 (2003). <https://doi.org/10.1109/acssc.2003.1292216>

Chapter 9

Web-Based Virtual Reality for Planning and Simulation of Lifting Operations Performed by a Hydraulic Excavator



Boris Tudjarov, Rosen Mitrev , and Daniela Gotseva

Abstract The paper presents the development of a Web-based virtual reality (VR) environment for planning and simulation of lifting operations performed by a hydraulic excavator. The excavator is represented as a mechanical system with four degrees of freedom (DOF) consisting of a fixed base body, three-link digging manipulator, and a swinging payload. An inverse kinematic model and a dynamic model of the excavator during performing lifting operations with bucket following prescribed vertical straight-line trajectory are developed. Workspace characteristics of the digging manipulator are determined, and the subset of the workspace corresponding to the cutting-edge positions for constant orientation of the bucket is identified. Using modern Web-based technologies, a VR environment for planning, simulation, and 3D animation of the excavator motion is developed. It consists of an HTML5 Web page with an X3D model of the scene, open-source framework X3DOM, and JavaScript functions. The developed VR environment considerably facilitates the design, investigation, planning, and e-learning of the lifting operations, performed by the excavator.

9.1 Introduction

The accelerated rates of economic development over the past three decades have considerably increased the use of the hydraulic excavators in the industry. The significant efforts invested by the engineers to collaboratively develop additional specialized attachments have turned the excavators into multifunctional machines capable

B. Tudjarov · R. Mitrev (✉) · D. Gotseva
Technical University of Sofia, 8 Kliment Ohridski Blvd, Sofia, Bulgaria
e-mail: rosenm@tu-sofia.bg

B. Tudjarov
e-mail: bntv@tu-sofia.bg

D. Gotseva
e-mail: dgoceva@tu-sofia.bg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 216,
https://doi.org/10.1007/978-981-33-4676-5_9

125

of performing multiple technological operations and thus increasing the economic efficiency of the activities performed.

One such frequently performed activity, but not typical for the excavators, is the lifting of various construction elements as pipes, beams, concrete blocks, etc. This allowed the nomenclature of the used machines to be shortened, but the practical use of the excavators as cranes on the mining and construction sites continues to be accompanied by some problems due to their design features. One major problem is that the operators and support staff are not trained to handle swinging payloads, which is a premise for accidents and incidents [1]. Another major reason for the increased number of accidents is that the hydraulic excavators are not equipped with the necessary overload protection devices as the cranes—load and load moment limiters, limit switches, operation radius indicators, etc., leads to a large number of incidents with people. Although there are practical recommendations [2] and standards [3], the peculiarities and potential opportunities for the use of excavators as cranes have not been sufficiently explored. Many authors [4–8] have traditionally focused their efforts on the development of simulation models suitable for the study of the tip-over stability of excavators and similar machines. Insufficient attention is paid to the planning and study of the lifting operations performed by the excavators especially in the case of lifting operations [9] following a pre-set vertical trajectory. The preliminary planning and visualization of the lifting process are indispensable, especially in the case of limited visibility and narrow spaces when the payload swinging is undesirable.

The design and study of the payload lifting require the development of appropriate dedicated research tools, enabling modeling and simulation of the main features of the lifting processes. A previous study [10] presents a VR developed in C++ and Open GL intended for a simulation of a truck-mounted concrete boom pump motion with redundant kinematical structure. The authors of the recent papers [11, 12] used commercial virtual prototyping CAD/CAE systems to simulate and visualize similar machines. The authors of [13] present a 3D physics-based excavator VR simulator intended for training purposes and capable of simulate digging operations in real working conditions. The searching through the literature showed that the interactive simulation and animation directed to the planning of lifting operations performed by hydraulic excavators or machines with similar kinematical structure are not enough studied.

The general aim of the paper is the development of a contemporary Web-based VR combining complex engineering calculations with quality 3D visualization and motion animation, capable of accurately simulating the vertical motion of the digging manipulator with a freely suspended payload taking into account the payload swinging. The easy change of the input data and the availability of an expandable library of industrially manufactured machines will facilitate the planning and study of the lifting operations.

This paper is organized as follows: Sect. 9.1 reports some fundamental issues and studies concerning the use of the hydraulic excavators for performing lifting operations of the freely suspended payload; Sect. 9.2 focuses on the derivation of the kinematical and dynamic models of the excavator; Sect. 9.3 presents the development

and testing of a Web-based virtual reality environment for simulation, 3D visualization, and animation of lifting operations; the paper concludes with a summary in Sect. 9.4.

9.2 Mathematical Model of a Hydraulic Excavator Performing Lifting Operations

A key step in the development of the VR software system is the derivation of a mathematical model of the mechanical system combining a model of the inverse kinematics and a dynamic model including the payload swinging [14].

9.2.1 Kinematical Analysis

Figure 9.1a presents the schematic view of a hydraulic excavator performing vertical lifting operations. It comprises a fixed to the ground base body (pos.1) and a serial link digging manipulator consisting of interconnected by rotational joints boom (pos.2), stick (pos.3), and bucket (pos.4).

The kinematical analysis of the manipulator in the vertical plane is performed for the cutting edge (denoted by O_4) moving along a vertical straight line with an initial point with coordinates $(x_{O_4}, y_{O_4}^s)$ and endpoint with coordinates $(x_{O_4}, y_{O_4}^f)$. The payload -pos.5 is presented as a point mass ($p.O_5$) attached to a lifting eye ($p.P$) by a rope. The motion of the bucket cutting edge along the predefined vertical

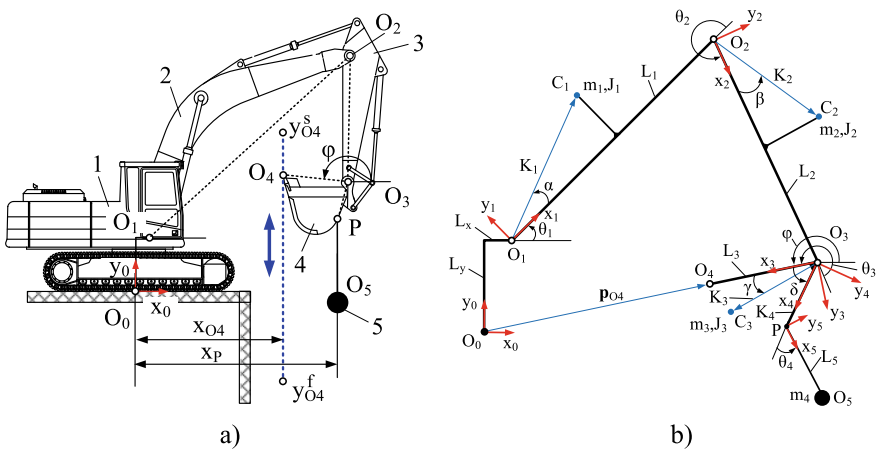


Fig. 9.1 Schematic view of a hydraulic excavator performing lifting operations (a) and geometrical layout of the corresponding kinematical structure (b)

line trajectory is controlled by the lengths of the hydraulic cylinders to maintain a constant angle φ of the bucket relative to the horizontal line. The system is considered with four DOF, three of which are controllable by hydraulic cylinders and one is uncontrolled—the payload swinging.

Figure 9.1b shows the geometrical layout of the excavator kinematical structure considering the oscillatory motion of the payload with mass m_4 . The kinematical model is developed using a fixed global reference frame $\{x_0y_0\}$ attached to the ground and local coordinate systems $\{x_1y_1\}/\{x_5y_5\}$ attached to the links with centers in the rotational joints. The positions of the centers of gravity of the links and specific points in the local coordinate systems are described by vectors transformed to the global reference frame by homogeneous transformation matrices. The desired vertical straight-line trajectory for the motion of p. O_4 is proposed in the form of a fifth-order polynomial [15]. Imposing zero initial and final conditions for the velocity and acceleration of the p. O_4 , the polynomial as a function of time is written as:

$$y_{O_4}^d(t) = y_{O_4}^s + \frac{t^3(y_{O_4}^f - y_{O_4}^s)(6t^2 - 15tt_f + 10t_f^2)}{t_f^5} \quad (9.1)$$

where $t_f > 0$ denotes the duration of the motion along the trajectory.

For the determination of the generalized angular coordinates θ_1 , θ_2 , and θ_3 corresponding to the motion of the p. O_4 along the predefined straight line, the inverse kinematics solution for a three-link planar manipulator with constant orientation φ of the bucket is used [15]. The determination of the values of the angles and their derivatives as a function of the cutting-edge O_4 position and bucket orientation completely ends the inverse kinematics task solution.

9.2.2 Workspace Characteristics of the Digging Manipulator

The digging manipulator total workspace in the plane determines the tasks the excavator is suited for. The horizontal size of the workspace determines the maximum allowable payload weight taking into account the tip-over stability conditions [8].

The workspace is constructed with consideration of the possible rotation intervals of the joints determined by the lower θ_i^{\min} and upper θ_i^{\max} bounds due to the existence of the linear driving mechanisms. The placement of the cutting edge in all points of the desired vertical trajectory maintaining a predefined constant orientation φ of the bucket is only possible in a certain subset of points in the manipulator total workspace. This subset can be identified if the total workspace is discretized, and for every discrete point, the angles θ_i ($i = 1, 2, 3$) are computed according to the inverse kinematics solution with a check of whether they are within the predefined limits $[\theta_i^{\min}, \theta_i^{\max}]$. Figure 9.2 shows the constructed workspaces for different values of the angle φ with values (in degrees) indicated in each picture.

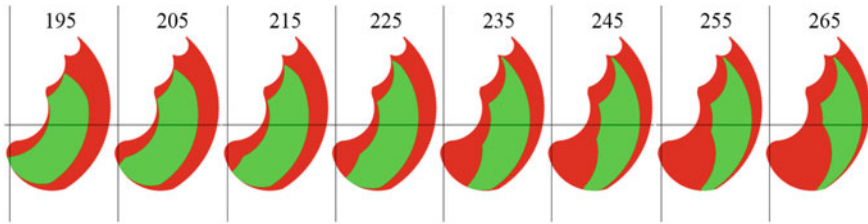


Fig. 9.2 Evolution of the workspace

The red section depicts the total workspace, and the green section depicts the identified subset corresponding to the cutting-edge positions for constant orientation φ of the bucket. As can be seen, the location, shape, and area, as well as the possible maximum height and depth of the cutting edge position, are strongly influenced by the value of φ . A study shows that the ratio of the green and the red sections is greatest for angle φ equal to 225° . Obviously, due to the bucket design features and constant vertical direction of the payload weight force, for the correct execution of the lifting operations, the operator must keep the bucket angle φ within a proper predefined interval of values which fact must be taken into account during the planning of lifting operations.

9.2.3 Dynamic Model

Each digging manipulator link (see Fig. 9.1b) is characterized by its geometrical and inertia parameters and is subjected to forces generated by the hydraulic cylinders and gravity. Four generalized coordinates are used to define the geometrical configuration of the mechanical system—three of them are to describe the rotation of the manipulator links and one—for the rotation of the swinging payload. The derivation of the dynamic model is based on the following assumptions: (1) The manipulator links are considered as rigid bodies; (2) the inertial properties of the closed kinematic chain elements (hydraulic cylinders and linkages) are added to the inertial parameters of the corresponding manipulator links; (3) the friction forces in the rotational joints and hydraulic cylinders are neglected. The differential equations describing the motion of the digging manipulator together with the swinging payload are derived using the Lagrange equations of the second kind [16] and are obtained as:

$$M(q)\ddot{q} + V(q, \dot{q}) + G(q) = Q \tag{9.2}$$

whereby \mathbf{q} and $\dot{\mathbf{q}}$ are denoted are the vectors of generalized coordinates and velocities, respectively.

In (2), the following notations are used:

$\mathbf{M}(\mathbf{q})_{4 \times 4}$ —symmetric and positive definite inertia matrix containing the following inertia terms:

$$M_{11} = A_1 + A_2 + A_3$$

$$A_1 = J_1 + 2K_2L_1m_2c_{\beta 2} + 2L_1(m_4(K_4c_{\delta 23} + L_5c_{\delta 234}) + K_3m_3c_{\gamma 23} + L_2m_3c_2)$$

$$A_2 = m_3(2K_3L_2c_{\gamma 3} + K_3^2 + L_2^2) + m_1K_1^2 + m_2K_2^2 + m_{234}L_1^2$$

$$A_3 = m_4(2K_4(L_2c_{\delta 3} + L_5c_4) + K_4^2 + 2L_2L_5c_{\delta 34} + L_2^2 + L_5^2)$$

$$M_{12} = A_4 + A_5$$

$$A_4 = K_2L_1m_2c_{\beta 2} + K_2^2m_2 + m_3(K_3(2L_2c_{\gamma 3} + L_1c_{\gamma 23}) + K_3^2 + L_2(L_1c_2 + L_2))$$

$$A_5 = m_4(K_4(2L_2c_{\delta 3} + L_1c_{\delta 23} + 2L_5c_4) + K_4^2$$

$$+ L_5(L_1c_{\delta 234} + 2L_2c_{\delta 34}) + L_2(L_1c_2 + L_2) + L_5^2)$$

$$M_{13} = A_6 + A_7$$

$$A_6 = m_4(K_4(L_1c_{\delta 23} + L_2c_{\delta 3} + 2L_5c_4) + L_5(L_1c_{\delta 234} + L_2c_{\delta 34} + L_5))$$

$$A_7 = K_3m_3(K_3 + L_1c_{\gamma 23} + L_2c_{\gamma 3}) + K_4^2m_4$$

$$M_{14} = L_5m_4(K_4c_4 + L_1c_{\delta 234} + L_2c_{\delta 34} + L_5)$$

$$M_{22} = A_8 + A_9$$

$$A_8 = m_4(2K_4L_2c_{\delta 3} + 2L_5(K_4c_4 + L_2c_{\delta 34}) + K_4^2 + L_2^2 + L_5^2)$$

$$A_9 = J_2 + m_3(2K_3L_2c_{\gamma 3} + K_3^2 + L_2^2) + K_2^2m_2$$

$$M_{23} = K_3m_3(K_3 + L_2c_{\gamma 3}) + m_4(K_4(L_2c_{\delta 3} + 2L_5c_4) + K_4^2 + L_5(L_2c_{\delta 34} + L_5))$$

$$M_{24} = L_5m_4(K_4c_4 + L_2c_{\delta 34} + L_5)$$

$$M_{33} = J_3 + m_4(2K_4L_5c_4 + K_4^2 + L_5^2) + K_3^2m_3$$

$$M_{34} = L_5m_4(K_4c_4 + L_5),$$

$$M_{44} = m_4L_5^2;$$

$\mathbf{G}(\mathbf{q})_{4 \times 1}$ —vector containing terms proportional to the weight of the links: $G_{11} = A_{10} + A_{11}$

$$A_{10} = m_3g(K_3c_{\gamma 123} + L_2c_{12}) + m_4g(K_4c_{\delta 123} + L_5c_{\delta 1234} + L_2c_{12})$$

$$A_{11} = m_1gK_1c_{\alpha 1} + m_2gK_2c_{\beta 12} + m_{234}gL_1c_1$$

$$G_{22} = m_4g(K_4c_{\delta 123} + L_5c_{\delta 1234}) + m_2gK_2c_{\beta 12} + m_3gK_3c_{\gamma 123} + m_{34}gL_2c_{12}$$

$$G_{33} = m_4g(K_4c_{\delta 123} + L_5c_{\delta 1234} + K_3m_3c_{\gamma 123})$$

$$G_{44} = m_4gL_5c_{\delta 1234};$$

$\mathbf{Q}_{4 \times 1}$ —vector containing the generalized forces and moments, associated with the respective generalized coordinates:

$$\mathbf{Q} = [\tau_1 \ \tau_2 \ \tau_3 \ 0]^T \quad (9.3)$$

where τ_i denotes the driving torques applied to the corresponding joints;

$\mathbf{V}(\mathbf{q}, \dot{\mathbf{q}})_{4 \times 1}$ —vector containing centrifugal and Coriolis terms with elements not shown here due to their large size;

The following short notations are used: $s_1 \rightarrow \sin \theta_1$, $m_{1234} \rightarrow m_1 + m_2 + m_3 + m_4$, $s_{\gamma 123} \rightarrow \sin(\gamma + \theta_1 + \theta_2 + \theta_3)$, $c_{12} \rightarrow \cos(\theta_1 + \theta_2)$, etc. By g is denoted, by m_1 , m_2 , and m_3 are denoted the masses of the links, by J_1 , J_2 , and J_3 are denoted the mass moments of inertia of the links according to their gravity centers C_1 , C_2 , and C_3 , by L_x , L_y , L_1 , L_2 , L_3 , and L_5 are denoted distances, by K_1 , K_2 , K_3 , and K_4 are denoted vectors to the specific points, defined in the links coordinate systems.

The available from the inverse kinematics solution values for the generalized coordinates $\theta_1, \theta_2,$ and $\theta_3,$ and their time derivatives, allow the system of differential Eq. (9.2) is divided into two sets. The first set consists of one equation, considered as an initial value problem and solved by standard numerical routines for the angle of the payload swinging θ_4 and defined values for the initial angle $\theta_4(0)$ and initial angular velocity $\dot{\theta}_4(0)$. The second set consists of three equations used to compute the necessary driving torques, applied to the corresponding rotational joints of the digging manipulator links.

9.3 Development and Testing of a Web-Based VR for Planning, Simulation, and 3D Animation of Lifting Operations

9.3.1 Description of the Structure of the Developed Web-Based VR Environment

The developed in Sect. 9.2 mathematical model is implemented in a Web-based VR environment intended for planning, simulation, and 3D animation whose framework is shown in Fig. 9.3a. The 3D model of the hydraulic excavator is transferred to X3D format by the use of Vivaty Studio software. The following means are used for the creation of the VR application: (a) X3D language [17] for the representation of 3D models; (b) X3DOM technology for integrating 3D content into Web pages by using WebGL; (c) HTML5 [18] for describing the Web pages; (d) JavaScript [18] for providing the main functionality in the application. Figure 9.3b depicts the general layout of the designed VR environment. The following elements are

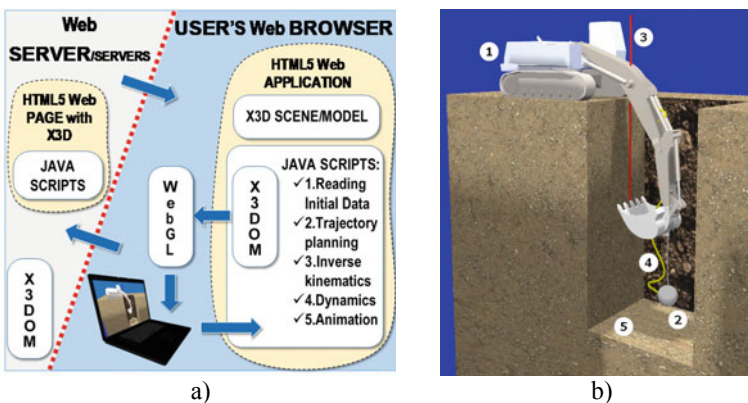


Fig. 9.3 Framework of the developed Web-based VR environment (a) and X3D scene (b)

denoted: the excavator—pos.1; the payload—pos.2; the followed vertical straight-line trajectory—pos.3; the trajectory of the swinging payload—pos.4; a pit or narrow space—pos.5.

The sequence of actions when working with the Web-based VR system is as follows: (1) The user requests a Web page/resource by entering the address of the page; (2) the server (or servers) sends the requested HTML page and open-source framework X3DOM;

(3) By using the graphical user interface, the user enters the values of the following parameters—the coordinates of the initial point and endpoint of the straight-line trajectory to follow, the mass of the payload, the length of the rope, duration of the motion, and the initial values of the payload angle and angular velocity; (4) the user performs the simulation, observes the real-time motion animation, and analyzes the dynamic behavior of the machine and the payload during the operation. JavaScripts are created for the realization of the following tasks: reading of the initial data, trajectory planning, computation of the inverse kinematics (angles, velocities, and accelerations), calculation of the dynamics of payload swinging. For the solution of the differential equation, a JavaScript function implementing the fourth-order fixed-step Runge–Kutta method [15] has been developed.

The easy reuse of the software allows the user to test various combinations of the machines and input parameters and to accept those which comply with the workspace limitations and satisfy the pre-set requirements for the height, depth, and horizontal distance of the required lifting operations. Especially valuable is the ability to accurately assess whether the dynamic behavior of the payload is reasonable in terms of compliance with a requirement for a certain minimum distance between the payload and the pit walls, as each contact between them due to the payload swinging is a prerequisite for an accident.

9.3.2 Correctness of the Mechanical Structure Movements in the VR Environment

Ensuring the correctness of the mechanical structure movements is a key issue during the development of the VR environment. For the models with pure tree structure, the capabilities of the X3D modeling technology for inheritance can be effectively used to inherit the movement along the respective hierarchical relationships “parent”–“child” in the tree structure of the model. When the location of the “parent” element changes, the location of its “children” also changes, as their relative position and orientation with respect to the parent element remain unchanged. Figure 9.4 shows the developed X3D model of the hydraulic excavator with marked subassemblies 1/13, specific points O_1/O_4 , P and O_{1a}/O_{3e} , and main dimensions L_1 , L_2 , and L_3 . Since the scripts compute the absolute angles of rotation, one has to take into account that the model has an initial disposition between its elements (initial angles). The

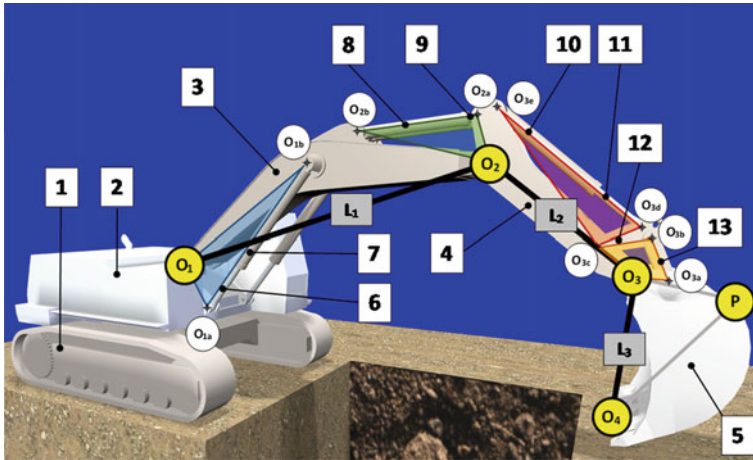


Fig. 9.4 X3D model with marked subassemblies, points, and dimensions

model of the excavator does not possess pure tree structure—due to the presence of closed kinematic chains there are cycles—see Fig. 9.4.

To ensure the correctness of the movements of the elements, the inheritance of the movement of the “child” from the “parent” (for the open kinematic chain consisting of the boom, stick, and bucket) is combined with additional real-time calculations for specific geometric constraints (for the closed kinematic chains, formed by the hydraulic cylinders and four-bar linkage). The initial model with cycles is transformed into a model with pure tree structure via the substitution of some joints with geometric constraints. For the rotations of the elements of the open kinematic chains around the points O_1 , O_2 , and O_3 , X3D capabilities for the inheritance of the movements by a direct use of the calculations are used. The rotations of the elements 6 and 7 (at points O_{1a} and O_{1b}), of the elements 8 and 9 (at points O_{2a} and O_{2b}), of the elements 10 and 11 (at points O_{3e} and O_{3d}), and of the four-bar linkage (at points O_3 , O_{3a} , O_{3b} , and O_{3c}) depend on rotations in the points O_1 , O_2 , and O_3 and have to be calculated at every time step. Additional calculations are required for the triangles $O_1O_{1a}O_{1b}$, $O_2O_{2a}O_{2b}$, $O_{3c}O_{3d}O_{3e}$, and circles intersection between circles with center points O_{3a} and O_{3c} and their corresponding radii $O_{3a}O_{3b}$ and $O_{3c}O_{3b}$ for the four-bar linkage.

9.3.3 Testing of the Developed Web-Based VR Environment

The developed Web-based VR [19] has been tested with different input data and the tests confirmed the possibility such a system to be used as a planning, simulation, and high-quality 3D animation tool for exploring the excavator and payload behavior during performing lifting operations. The additional tests on different mobile devices

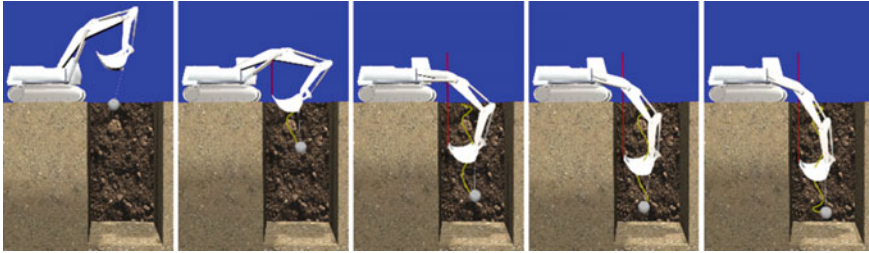


Fig. 9.5 Frames of the motion animation

and operating systems showed great reliability of the system. Figure 9.5 shows frames from the motion simulation and animation, performed for an industrially manufactured excavator for the following values for the tuning parameters: $t_f = 12$ s, $x_{O_4} = 5$ m, $y_{O_4}^s = 4$ m, $y_{O_4}^f = -5$ m, $\theta_4(0) = 15^\circ$, $\dot{\theta}_4(0) = 0$ rad/s. An additional study showed that for the frame rate set to 33.33 fps, the quality of the animation is well enough.

9.4 Conclusions

In this paper, we have presented the development of a Web-based VR environment for planning, simulation, and high-quality 3D animation of lifting operations carried out by a hydraulic excavator. The VR environment is based on the developed dynamic four DOF model of the excavator during performing lifting operations with a bucket following prescribed vertical straight-line trajectory and considering the payload swinging. A crucial factor was the ability to efficiently implement specialized mathematical algorithms in combination with modern tools for visualization.

The developed Web-based VR considerably accelerates the preliminary planning of the lifting operations due to the possibility to virtually test multiple combinations of parameters and to thoroughly understand the excavator and payload dynamic behavior during performing various motions. The Web-based technology used provides an intuitive perception of the information obtained through the simulation, that is why the VR environment offers great educational potential. The numerical and visual information is easily shared in the Web environment, including all kinds of mobile devices. The extendable 3D library of industrially manufactured excavators allows fast experimentation with different machines, followed by a presentation of the obtained results and therefore allowing iterative design and planning.

Acknowledgements The authors would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support.

References

1. Edwards, D., Holt, D.: Case study analysis of risk from using excavators as ‘cranes’. *Autom. Constr.* **19**(2), 127–133 (2010)
2. OPERC Safety alert (ALT-017)—Excavators Used as Cranes. Off-highway Plant and Equipment Research Centre. https://www.operc.com/safetyalert/alt_017_a4.pdf. Last accessed 11 Mar 2020
3. ISO 10567:2007. Earth-moving machinery—Hydraulic excavators—Lift capacity, pp. 1–16 (2007)
4. Lim, T., Kim, Y., Hwan, J., Lee, H., Yang, S.: Development of tipping-over rate computation system for hydraulic excavator having crane function. In: Proceedings of the 8th Russian-Korean International Symposium on Science and Technology KORUS, pp. 76–79. Tomsk, Russia (2004)
5. Yu, B., Park, K., Lee, K., Han, C.: Static compensation ZMP algorithm preventing tips-over of a tele-operation excavator. In: Proceedings of the 29th International Symposium on Automation and Robotics in Construction, pp. 652–658. Eindhoven, The Netherlands (2012)
6. Janosevic, D., Pavlovic, J., Jovanovic, V., Petrovic, G.: A numerical and experimental analysis of the dynamic stability of hydraulic excavators. *Facta Univ. Ser. Mech. Eng.* **16**(2), 157–170 (2018)
7. Abo-Shanab, R., Sepehri, N.: On dynamic stability of manipulators on moveable platforms. In: Proceedings of the 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 479–485. Alberta, Canada (2001)
8. Mitrev, R., Marinkovic, D.: Numerical study of the hydraulic excavator overturning stability during performing lifting operations. *Adv. Mech. Eng.* **11**(5) (2019)
9. Kim, T., Lim, H., Cho, H., Kang, K.: Automated lifting system integrated with construction hoists for table formwork in tall buildings. *J. Constr. Eng. Manag.* **140**(10), 04014049–1–04014049-10 (2014)
10. Zhou, S., Zhang, S.: Co-simulation on automatic pouring of truck-mounted concrete boom pump. In: IEEE International Conference on Automation and Logistics, pp. 928–932. Jinan (2007)
11. Ren, W, Li, Z, Bi, Y., Zhao, S., Peng, B., Zhou, L.: Modeling and analysis of truck mounted concrete pump boom by virtual prototyping. *J. Robot.* **2017** (2017). (Article ID 9182143)
12. Tsirogiannis, E., Vosniakos, G.: Redesign and topology optimization of an industrial robot link for additive manufacturing. *Facta Univ. Ser. Mech. Eng.* **17**(3), 415–424 (2019)
13. Dopico, D., Luaces, A.: A soil model for a hydraulic simulator excavator based on real-time multibody dynamics. In: Proceedings of the 5th Asian Conference on Multibody Dynamics, pp. 28–36. Kyoto, Japan (2010)
14. Mitrev R.: Development of the Theory for Designing, Modelling and Research of Construction and Lifting Manipulators. D.Sc. Thesis, pp. 290–301, Technical University, Sofia (2019)
15. Lynch, K., Park, P.: *Modern Robotics: Mechanics, Planning, and Control*, 1st edn. Cambridge University Press, USA (2017)
16. Rao, A.: *Dynamics of Particles and Rigid Bodies: A Systematic Approach*. Cambridge University Press, Cambridge, UK (2006)
17. Brutzman, D., Daly, L.: *X3D: Extensible 3D Graphics for Web Authors*. Elsevier (2007)
18. Meyer, J.: *HTML5 and JavaScript Projects*. Apress (2018)
19. www.caxsys.com/rmn/. Last accessed 29 June 2020

Chapter 10

On Metrics Used in Colonoscopy Image Processing for Detection of Colorectal Polyps



Raneem Ismail and Szilvia Nagy

Abstract Colorectal cancer is nowadays the fourth cause of cancer death worldwide. Prevention of colorectal cancer by detection and removal of early stage lesions is of essential importance and has become a public health challenge worldwide. As the screening is carried out mainly by some sort of endoscope, and the endoscopic image processing is an important area of research and development, it is essential to know what kind of measures are used in determining whether polyp finding hit rates or miss rates are acceptable. It is rather natural to match the hit rate measures to the method itself; thus, in this contribution, the most typical polyp detecting methods are summarized shortly together with the metrics they use for evaluation of their results. However, in computer-aided diagnostics, the measure that is used by the medical community might differ from the measures typical in image processing researches. Also, the output of such polyp detecting methods is tested as inputs for active contour methods.

10.1 Introduction

Colon cancer is nowadays (depending on the year and the region) the third or fourth most frequent cause of cancer-induced death. In only 2012, almost 750 thousand cases were diagnosed with some type of colorectal cancer (CRC) and over 350 thousand patients died of this disease. The survival rate of CRC depends on the stage it is detected in, going from rates higher than 95% in the case of early stages to rates lower than 35% in latter ones [1]; hence, the prevention of CRC by finding and removing of the so-called preneoplastic lesions (early stage precancerous polyps or colorectal adenomas) is of high importance and one of the worldwide public

R. Ismail · S. Nagy (✉)
Széchenyi István University, Egyetem tér 1, Győr 9026, Hungary
e-mail: nagysz@sze.hu

R. Ismail
e-mail: raneem.ismail@hallgato.sze.hu

health priorities [2], as well as target of multiple computer-aided diagnostics (CAD) researches.

Clinical guidelines based on the medical practice as well as scientific evidence recommend that the screening should be carried out with colonoscopy which allows both the identification and the removal of such polyps, which are at the first step in the many years long process of developing “adenoma” to “high grade dysplasia” to “carcinoma,” i.e., the development of colorectal cancer. According to our present knowledge, the conventional, classical, flexible tube-like colonoscopy, which has a device for removing polyp (perform polypectomy) is the most precise procedure for early colorectal cancer detection and prevention: Several reputed studies with sufficiently large number of cases demonstrate that it reduces the cancer’s incidence by 40–90%. However, colonoscopy has the following drawbacks. Before colonoscopy the patient has to perform a thorough bowel cleansing, and the procedure itself is also rather inconvenient, which makes people less willing to sign up for colonoscopy screening. The procedure requires expensive device and expert medical staff to carry out the examination. It also has a considerable risk of damaging the bowel wall; moreover, it is often carried out in anesthesia, which has its own risk. Besides the costs and risks of the medical process, the most relevant drawback is that the polyp miss rate is usually considered as rather high, 22%. The miss rate increases significantly in the case of smaller sized polyps [1, 2].

Because colonoscopy is considered as a risky operation, mini-invasive techniques such as wireless capsule endoscopy (WCE) have also been developed [3]. As the capsule endoscopy’s methods and devices develop rapidly and the inconvenience caused to the patient is significantly less than in the case of classical endoscopy, it seems to be valid alternative option to conventional colonoscopy for the general screening as well as for patients with contraindication to general anesthesia. A capsule endoscope is a kind of larger pill equipped with one or rather more camera sensors (CCD or CMOS), a processor, batteries, and a radiofrequency transmitter. This makes it possible to identify gastrointestinal phenomena deviating from the normal, like ulcers, blood, and polyps. If the polyp does not need removal, it is a convenient way to perform the screening, without hospitalization or sedation. However, in some cases, conventional endoscopy is needed, not only because of the polyps that are needed to be removed, but also accidents happen when the capsule gets stuck in the bowel at some point [4], so even capsule endoscopy is not a risk-free operation.

The camera of the capsule endoscope takes about 50,000 images during its path in the bowel system and transmits them to a receiver for storage. Offline image processing is used to identify gastrointestinal abnormalities. The complete analysis of the tens of thousands of images is time consuming for physicians and its transmission is energy consuming for the batteries of the capsules. Also, the size of the objects is hard to determine due to the lack of ability to direct the capsule endoscopy device, the lack of reference points, and the lack of stereo vision. These are the main challenges for the capsule endoscopy technique, such as the Cyclope WCE of [3].

A general purpose endoscope is another instrument that can get images from inside the human body: It allows the observation and mostly also the operation inside hollow organs or other cavities in the body through very small invasion cuts

[5]. Diagnosis with general purpose endoscopes can involve both shape detection and the assessment of tissue state [3].

In the past decade exploring the applicability of intelligent systems in computer-aided diagnosis, like in the case of polyp detection, became a trend in image processing. Mostly, polyp characterization methods are based on the calculation of a some feature descriptors over a tile of the image or the complete image: These descriptors support the decision of an intelligent system whether there is a polyp in the image (segment) or not, and often if there is a polyp, the location within the image (segment) is also determined. Two main groups of descriptors exist, first, the shape-based properties, and methods based on them, second, the texture-based ones.

The texture analysis is often used in the case of differential diagnosis [6–9], as based on its surface texture the nature of the polyp can be determined, however, magnified endoscopy images are needed to perform this type of analysis, thus, it is rarely used for detecting only the polyp, and it is used after the polyp is already found. The use of texture and color-based approaches in polyp characterization has been gaining interest only during the last years. Polyps tend to have different patterns than the bowel walls which contain visible blood vessels too. Some works are based on the use of wavelet descriptors to extract this information [10], paying attention on the detail and approximation coefficients of the wavelet transform. Other alternatives include the use of local binary patterns [11] or co-occurrence matrices [12]. One big drawback of this group of methods is that they tend to need of an exhaustive training and they are very sensitive to parameter tuning [13].

Shape-based approaches [1, 2, 4, 5, 13] aim to search in colonoscopy frames for those specific shapes that polyps commonly have, both in the intensity distribution and in the boundary shape. The first subgroup of such methods (i.e., the detection of the intensity distributions typical to some types of polyps) clusters a series of methods mostly consisting of low-level image processing operations such as gradient filters, valley information, or edge detectors. The second subgroup of methods (i.e., the boundary shape-based one) assumes that polyps tend to have elliptical shape. Of course, some other works tend to combine both curvature analysis and shape fitting to detect polyps in the image. Whereas other methods base their localization of polyps in the pure statistical of the tiles of the image, like the structural entropy [14, 15], edge density, combined with mean, standard deviation, and gradient information of both the image and its wavelet transform [16, 17]. Of course, many neural network-based algorithms exist for these purposes as well [12, 13].

As these methods are based on different features of the polyp, of course, their success rate measures are different too. In the following considerations, these measures are compared and studied whether the results are comparable. Also, these methods tend to give circular or rectangular regions of interest, which have to be translated to polyp contours with the help of other type of more refined image segmentation methods.

Image segmentation is considered as one of the most often used branches in image processing which is applied for the segregation of the meaningful regions of interest from the other parts of the image in order to make the subsequent processing steps more efficient. Besides, medical image processing it is appropriate for motion and

stereo tracking tasks. There are various techniques for segmentation of pixels of interest from the image. One of the most successful of such image segmentation methods is the active contour [18, 19].

In the second part of the paper, the outputs of the previously mentioned methods with rectangular regions of interests as outputs are tested as inputs for active contour method with different pre-filtering is studied.

According to the shape of the polyps, there are several groups, from protruding, pedunculated through completely flat till the excavated. These polyp types are officially classified multiple times during the history of the colorectal polyp diagnosis [20–22], and the most recent one is called Paris classification because of the location of the meeting, where it was developed. In most of the studies, only the protruding and elevated polyps are the targets of computer-aided diagnosis, and this paper follows that trend as well: Only protruding polyp databases are used.

The content of the paper looks like as follows. In Sect. 2, a short summary of the methods with a focal point on their evaluation is given. In Sect. 3, the measures corresponding to the units of the decision making are given, and a plausible, medically meaningful compromise is suggested. In the last section, the conclusion is drawn.

10.2 Approaches for the Determination of the Polyp Content on Images

In this section, some interesting methods are summarized.

Window Median Depth of Valleys Accumulation (WM-DOVA) [1] method was developed for polyp localization for colonoscopy videos. It is based on a model how the polyps look, what defines the boundaries of the polyps, based on human medical experts' eye-tracking experiments. The findings of these experiments are that humans search for boundaries that are continuous and have an intensity valley kind of structure. This is the reason why the method concentrates on the valleys around the polyp.

The model of the appearance of the polyps, first, the physical model of a polyp (i.e., a semispherical object with shiny surface) then the way how colonoscopy images are acquired (i.e., the object is enlightened from a point-like light source) are considered. Three different scenarios can appear in the model: (a) zenithal scenario with a complete view of the polyp and the surrounding bowel wall; (b) semi-lateral view of the polyp, when part of the bowel wall is covered by the polyp or hidden in a fold, and (c) lateral view of the polyp, when the polyp is visible from the side in front of a darker background as the polyp is very protruding and the bowel wall segment behind it is hidden in a fold. According to the model, the polyps are surrounded completely by valleys in zenithal views and partially from lateral views. Valleys around the polyps can build a closed or partially closed continuous contour, which is concave from the center of the polyp. The valley intensity depends on the circumstances of the image taking, like the strength and distance of the light source as well as the shape of the

polyp and the movement of the bowel. This means that WM-DOVA can find polyps of type Is, IIa, IIb, and Ip according to Paris classification criteria [20–22]. There are other valley-like objects on the images, like blood vessels, folds in the bowel wall, and they should be treated properly to find the polyps only. Image preprocessing can reduce the effects of the light reflected from the shiny surface.

To find the valleys, the geometrical valley detector is used. The valley detectors, however, give no information about the depth of the valley, hence, the application of the depth of valleys image (DV image): a combination of valley information and gradient information. After acquiring the depth of the valleys, a kind of energy map is introduced, where those pixels have higher energy that are surrounded or partially surrounded by intensity valleys. The model uses four parameters to restrict the possible polyp boundary candidates to real polyp boundaries: completeness, robustness, continuity, and concavity. These parameters can help in distinguishing rather continuous, rather concave and mostly quite complete and robust polyp valleys from random shaped, not concave blood vessels and concave, but less robust valley folds.

For capsule endoscopies another, psychovisual study of medical experts and the thus built model of the polyps is used. As a first step, regions of interest (ROI) are determined. Hough transform [23] for finding round contour segments of appropriate size is performed during this step, thus determining the ROI candidates. As a next step, the homogeneity, granularity, coarseness, and other pattern- and texture-based parameters in the ROIs are calculated by using co-occurrence matrices of learned polyp structures, which automatically scale to the size of the found roundish contour segment, i.e., the size of the ROI. The learning algorithm is Cascade AdaBoost [24, 25].

Another possibility is to use Hessian matrices as filters [5]. With the help of the Hessian filter [26], the convex and concave regions of the image can be determined, and based on this information, the first candidate regions for polyps are selected. By performing Fourier transform [27], those regions can be selected that have larger wavelength eigenvalues of the Hessian. A method based on the calculation of Voronoi diagrams [28] is used to determine the second approach for the polyp regions. Another step, using weighted histogram of gradients of intensity (HOG) method [29] in various directions to determine whether the previously selected regions belong to a polyp or some other blob-like structures. As the number of the dimension arising from the HOG can be large, two different learning algorithms, Real AdaBoost [30] and random forest [31] are used to train the system to perform decisions.

The last method analyzed here consists of cutting the image into tiles of size which can contain either a full polyp or at least a significant part of it. This tiling of the image (similarly to the output of the previous two models) can be a basis of a later, active contour [18, 19]-based precise contour search. The method applies fuzzy [32, 33] rule interpolation [34, 35]-based decision to determine whether the tile contains polyp or not. The method uses simple, statistical parameters of the image segments as antecedents, also based on discussions with gastroenterologists. The input statistical parameters are mean, standard deviation, edge density, and structural information based on Rényi entropies [15–17] of the pixel intensity distribution. These parameters

are calculated to all three color channels, both the original image and its wavelet transformed and gradient filtered version.

Of course, deep neural networks can be trained to detect polyps as well, they usually try to find those pixels that belong to a polyp and distinguish them from the parts without polyps of the images. From this point of view, neural network-based algorithms are somewhat different, and they usually do not need post-processing to find better-fitted contours, unlike the rectangular output methods listed above.

10.3 Measures for Success Rates

In all of the cases, the hit and miss rates are calculated by using well-known true positive (TP), false positive (FP), true negative (TN), and false negative (FN) classification of the results. Using the number of the results in these classes, various metrics can be introduced for characterizing the goodness of the decision-making system. These quantities for a set of K samples are the accuracy,

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{K}, \quad (10.1)$$

precision, or positive predictive value,

$$\text{Pre} = \frac{\text{TP}}{\text{TN} + \text{FP}}, \quad (10.2)$$

the specificity or selectivity

$$\text{Spe} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (10.3)$$

the sensitivity or true positive rate

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (10.4)$$

and the false positive rate

$$\text{FPR} = \frac{\text{FP}}{\text{TP} + \text{FN}}. \quad (10.5)$$

In the application of these metrics, the methods agree, though on what the basis of the TP, TN, FP, FN numbers, they differ significantly, as it can be seen in the following list.

- The maxima of the WM-DOVA energy map are inside polyps or very much polyp-like objects. The result of the method is a center of a polyp, which might or might

not be within the perimeters of the polyp itself. The measure of the correctness of their method is whether the detected dot is within the perimeters of a real polyp or only some polyp-like object. This would imply that if the center somehow gets out of the polyp perimeters due to some weird shape of the polyp—like being over a fold of the bowel wall—would result in a false classification, even though the polyp is found.

- The method has well over 70% accuracy.
- In the Hough transform-based ROI finding method, the metrics of a ROI being correct or not is based on the polyp content of the ROI. If it is at a site with polyp, the ROI is classified as correct, if it finds some other parts of the bowel, it is classified as incorrect. Their sensitivity, or true positive rate is given, it is 91%, but also the specificity, which is 95%, and the false detection rate, which is 4.8%.
- In the Hessian-based method has two types of decision-making systems, the Real AdaBoost and the random forest technique, they are treated separately. In this case, the sensitivity is extremely high on their database, close to 100%, and the selectivity is also rather high, well above 95% for both methods.
- In the case of the fuzzy inference based on statistical parameters, the result is a yes or no answer for all the tiles of the image. As the tiles are not overlapping, and not focusing onto regions of interest, they may contain partial polyps as well, unlike the previous two methods. The tests were carried out on three databases, and the true positive rate varied from database to database, the low quality video images had practically no valuable results, while the higher quality databases had above 80% true positive rate, and even if the area coverage of the polyp is not calculated into the results.
- As in this method, the percentage of the polyp content in the tiles can vary due to the size and location of the polyp, and it is important to note that here an area-based hit and miss rate was also introduced, but it still remained tile-wise. Instead of having a totally wrongly classified tile and totally correctly classified tile, if the polyp content of the tile was small, the error measure became also less, corresponding to the area covered by the polyp.
- In the case of the neural network-based solutions, the results are usually measured pixel-wise, so the 98% accuracy can mean that it is 98% of the pixels corresponding to the polyp and its surroundings are classified correctly, or also, that 98% of the samples are classified correctly, so it is often hard to compare the results.

10.4 Using Outputs of the Different Approaches as an Input to an Active Contour Method

Active contour is one of the effective models in segmentation techniques and its concept was first introduced in the paper “Snakes: Active Contour Models” [18] which was published in 1987 and since then it has been further developed by many researchers [36–38]. Active contour makes use of some kinds of energy constraints

and forces in the image for separation of the region of interest. In plausible terms, an active contour is an energy minimizing spline that finds out given features within an image. These features are converted into forces and energies that form a kind of energy map, where the active contour sits into a closed series of valleys, where it minimizes the energy and the force that tries to shrink (or expand) it to minimal length and minimal number of sharp bendings. At the end, it becomes a curve (or surface in 3D) that can adapt dynamically and with great flexibility to fit the shape of the specific objects within the image.

As a first step, the active contour is defined by an initial mask, which is often a set of points and straight lines, but it can be more sophisticated one too. Then, if the energies according to the desired properties are set, the contour will iterate into a more optimal place by making its energy decrease in each consecutive step, and at the same time, trying to fulfill the demands of the forces that prescribe the penalties for sharp curves, or extra length in the contour line, or whatever properties the developers find undesirable. These quantities are summarized in the deformable contour's force equations that can be either parametric or geometric. The method then usually converts all forces to energies that can be minimized by iteration.

The total energy function consists of internal and external energies from which the internal energy gives smoothness to the contour, the external forces pull the contour toward the desired local minimum, as if we would have a heavy but loose elastic rope around a weirdly shaped series of mountains (surrounded by other mountains), the internal forces stretch the rope to make it as short and curveless as possible, and the external gravity pulls the rope toward the valley.

The most commonly used active contour methods are the Chan-Vese method [36] and the geodesic active contour [37]. In this paper, both methods were tested with rather common parameter settings. Instead of fine-tuning the parameters, the images were prefiltered, this was the way the energy function was influenced. An example of the series of the filtered images can be seen in Figs. 1, 2, 3. The first set of filters consisted of mean, median, and standard deviation, which are the simplest statistical parameters. As the difference of the mean and the median gives information about rapidly varying parts of the image, it was kept as the fourth option.

The second set of filtered images is the gradient filtered ones, where the most promising one for energy and contour finding is the gradient magnitude.

The third set of filters is somewhat unique, it is based on the Rényi entropies [39] of the image, which is a generalization of the well-known Shannon entropy S_1 [40] by

$$S_n = \frac{1}{1-n} \ln \left(\sum_{i=1}^N I_i^n \right),$$

where I_i denotes the i th pixel intensity normalized to be a probability distribution component. The structural entropy $S_{\text{str}} = S_1 - S_2$, and the spatial filling factor $\ln q = S_0 - S_2$ are proven [14, 40–43] to characterize the shape of probability distributions, so they are used as 5×5 filters as well.

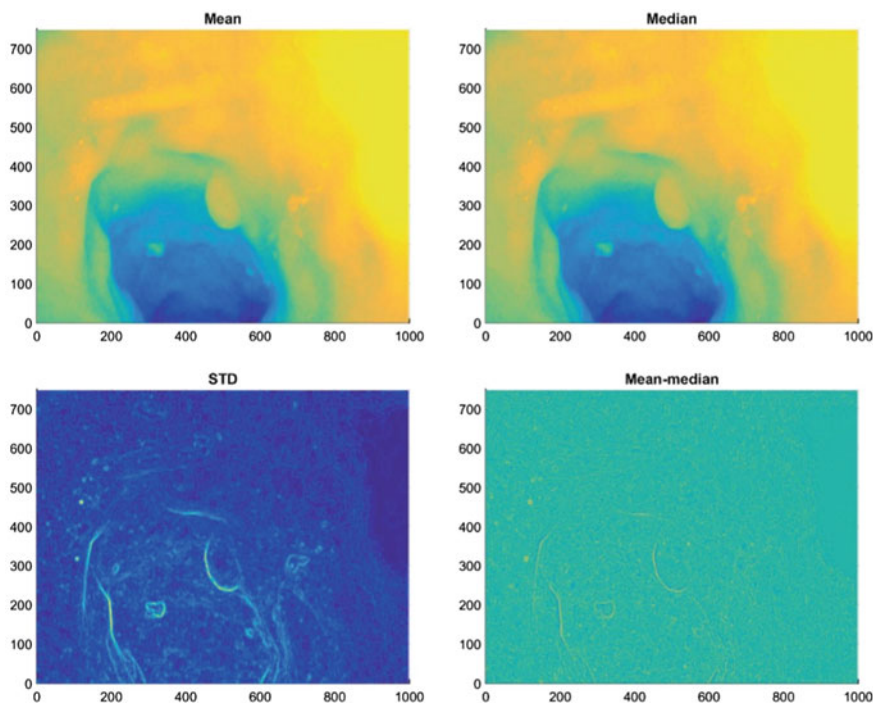


Fig. 10.1 Mean, median, and standard deviation (STD) filtered images of the red color channel from image 99 in database ETIS LARIB [3]. The last image is the difference between the mean and the median filtered images. The filter size was 5×5

Figures 4 and 5 give the resulting masked image for the same picture component as the previous plots. The parameters of the active contour methods were set as follows. The iteration number was 100, smoothness value is 0 and 1, the contraction bias 0 and 0.3 for the Chan-Vese and the geodesic active contour methods, respectively. It can be seen that in most of the cases, the Chan-Vese algorithm fail to find the contour, it extends to a nice regular-shaped mask, in most cases. The geodesic methods give better results, but still, the contour finds irregular nearby valleys in most of the cases. The best candidates seem to be the standard deviation and the difference of the mean and median filtered images in the case of the geodesic active contour method.

10.5 Conclusion

Four methods were selected to demonstrate the comparability problems in medical image processing accuracy measures. All examples used colonoscopy image databases. The first, valley information and contour convexity-based method resulted in a central point of the polyp, which could be highlighted in the image. The second

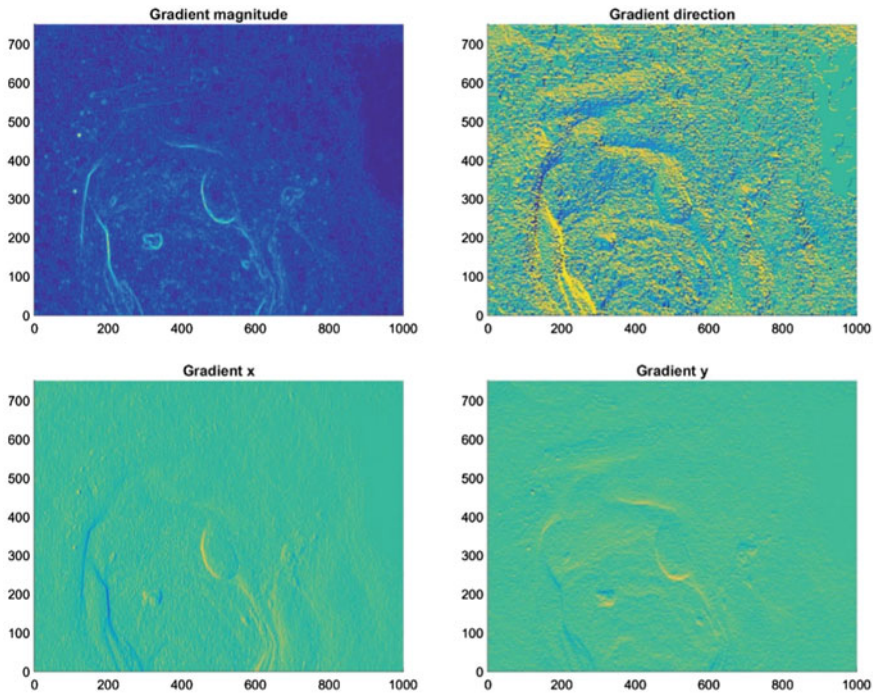


Fig. 10.2 Gradient filtered images of the red color channel from image 99 in database ETIS LARIB [3]. The first image is the gradient magnitude, the second the direction, the second row pictures show the x and y direction components of the gradient. The filter size was 5×5

method using Hough transform is based on regions of interest, and the metric for determining the effectiveness of the method is based on whether the ROI is correct or not.

The third method based on convexity of the pixel intensity distribution and some elliptical, edge detecting methods, and different types of learning algorithms also determines rectangular areas, where the polyps can be located, and it is also depending on the position of the polyp. The last method uses fixed grid to cut the images into tiles and determines whether these tiles contain polyp segments or not. This method, besides using the classical, crisp evaluation of the tile polyp content result, i.e., whether a tile classified as polyp containing really contains polyps, introduced a fuzzy-like correctness measure based on the proportion of the polyp within the tile area.

As a next step, these contours as basis of an active contour method were tested. Despite the fact that the rectangular contours were very far from ideal starting contour, the geodesic active contour method found the contour of the polyp in very few iterations, to rather good precision, if not the original, but filtered images was used, especially in the case of edge enhancing filters like standard deviation or gradient or the difference of the mean and median filter.

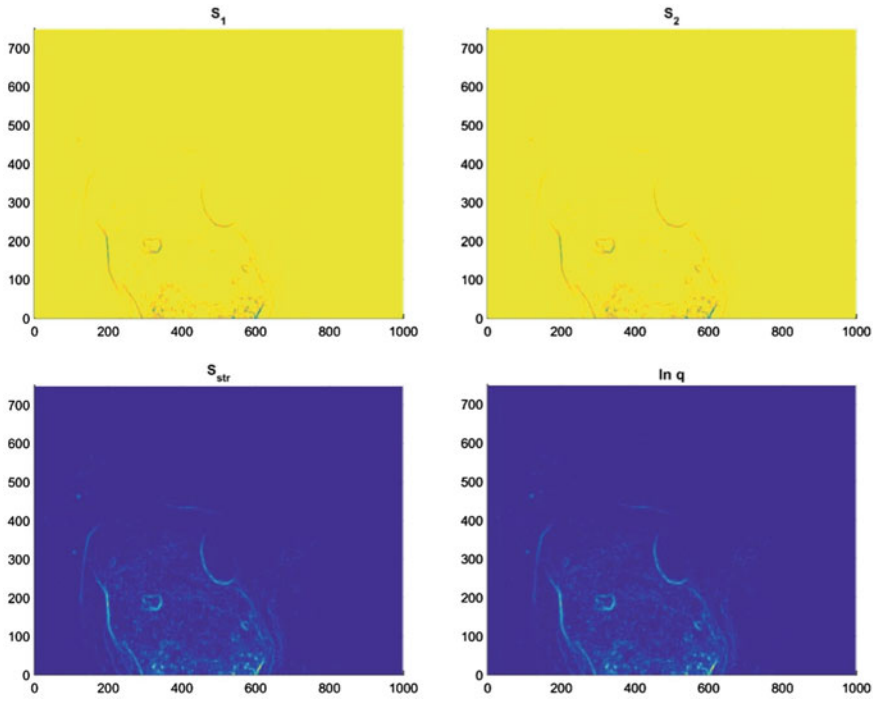


Fig. 10.3 Rényi entropy, structural entropy, and spatial filling factor filtered images of the red color channel from image 99 in database ETIS LARIB [3]. The filter size was 5×5

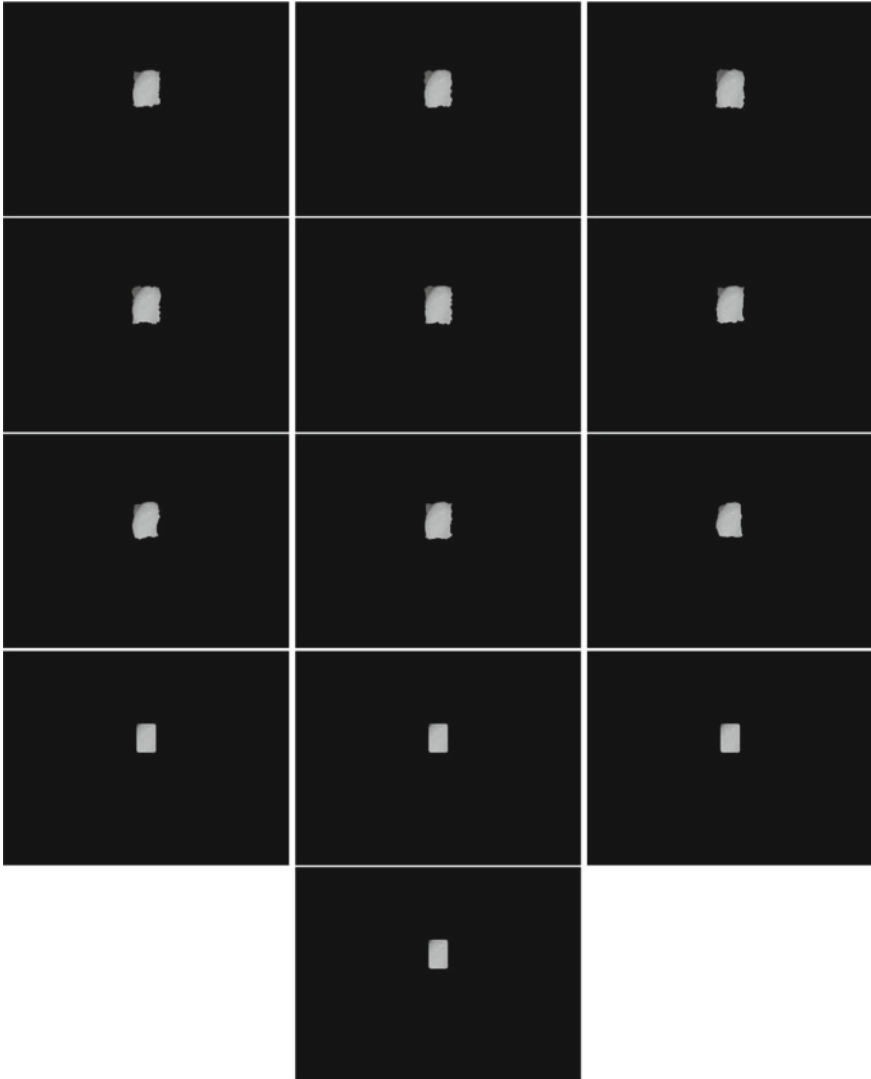


Fig. 10.4 Masks resulting from the active contour method for the same images as above, using the geodesic method. The original picture is the red color channel from image 99 in database ETIS LARIB [3]. The order of appearance is (1) original, unfiltered image, (2–5) gradient filters, (6) mean filtered, (7) standard deviation filtered, (8) median filtered images, (9) difference of mean and median filtered images, (10) S_1 , (11) S_2 , (12) S_{str} , and (13) lnq filtered images used as basis of the active contour

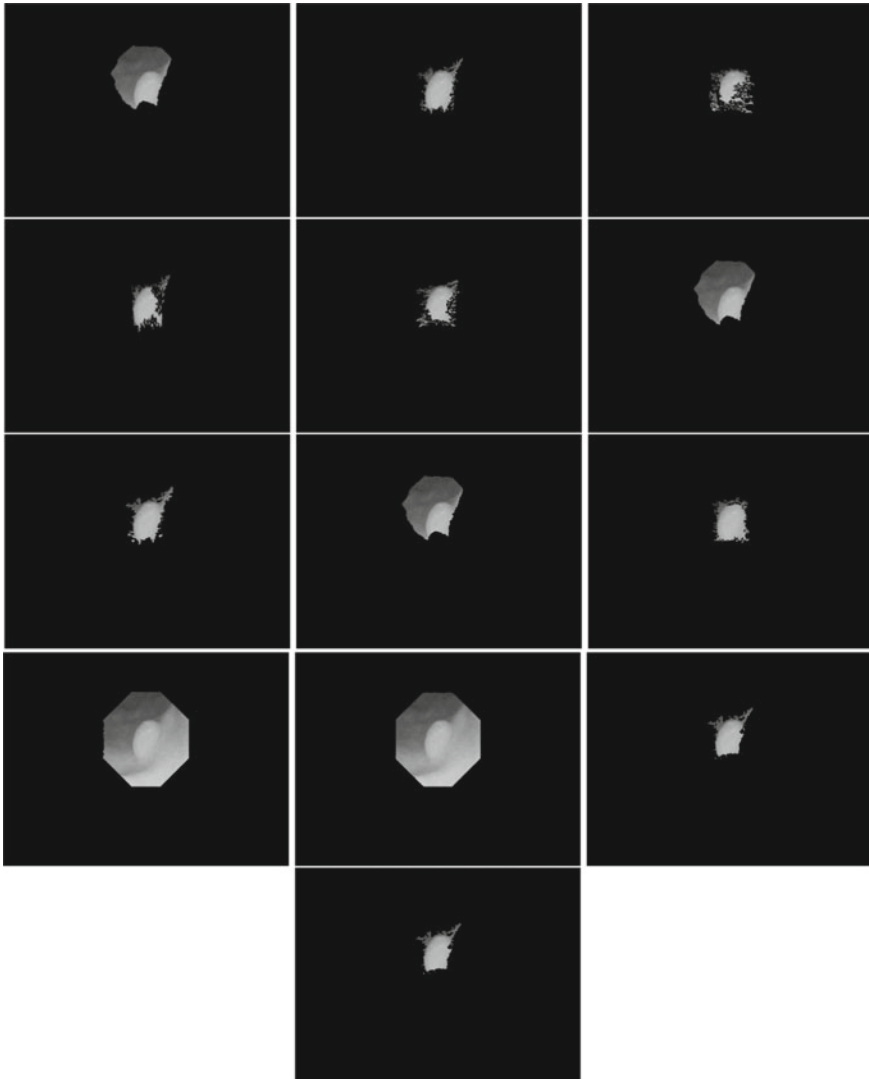


Fig. 10.5 Masks resulting from the active contour method for the same images as above, using the Chan-Vese method. The original picture is the red color channel from image 99 in database ETIS LARIB [3]. The order of appearance is (1) original, unfiltered image, (2–5) gradient filters, (6) mean filtered, (7) standard deviation filtered, (8) median filtered images, (9) difference of mean and median filtered images, (10) S_1 , (11) S_2 , (12) S_{str} , and (13) $\ln|q|$ filtered images used as basis of the active contour

References

1. Bernal, J., Sánchez, F., Esparrach, G., Gil, D., Rodríguez, C., and Vilariño, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Comput. Med. Imag. Graph* **43**(6), 99–111 (2015)
2. Bernal, J., Sanchez, F.J., Vilariño, F.: Towards automatic polyp detection with a polyp appearance model. *Pattern Recog.* **45**, 3166–3182 (2012)
3. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. *Int. J. Comput. Assist. Radiol. Surg.* **9**(2), 283–293 (2014)
4. Rácz, I., Jánoki, M., Saleh, H.: Colon cancer detection by rendezvous colonoscopy: successful removal of stuck colon capsule by conventional colonoscopy. *Case Rep. Gastroenterol.* **4**(1), 19–24 (2010)
5. Yuji, I., Akira, H., Yoshinori, A., Bhuyan, M., Robert, J. and Kunio, K.: Automatic detection of polyp using hessian filter and HOG features. In: 19th International Conference in Knowledge Based and Intelligent Information and Engineering Systems—KES2015, *Procedia Computer Science* vol. 60, pp. 730–739 (2015)
6. Kudo, S., Hirota S., T. Nakajima, et al.: Colorectal tumours and pit pattern. *J. Clin. Pathol.* **47**, 880–885 (1994)
7. Horváth, A., Spindler, S., Szalai, M., Rácz, I.: Preprocessing endoscopic images of colorectal polyps. *Acta Technica Jaurinensis* **9**(1), 65–82 (2016)
8. Rácz, I., Horváth, A., Szalai, M., Spindler, S., Kiss, G., Regőczy, H., Horváth, Z.: Digital image processing software for predicting the histology of small colorectal polyps by using narrow-band imaging magnifying colonoscopy. *Gastroint. Endoscopy* **81**(5), 259 (2015)
9. Georgieva, V.M., Nagy, S., Kamenova, E., Horváth, A.: An Approach for pit pattern recognition in colonoscopy images. *Egypt. Comput. Sci. J.* **39**, 72–82 (2015)
10. Karkanis, S., Iakodivis, D., Karras, D., Maroulis, D.: Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures. In: *Proceedings of the IEEE International Conference in Image Processing*, pp. 833–863 (2001)
11. Häfner, M., Gangl, A., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Classification of endoscopic images using delaunay triangulation-based edge features. In *Proceedings of the 7th International Conference, ICIAR 2010, Póvoa de Varzim, Portugal, 21–23 June 2010*
12. Tan, J.X., Gao, Y.F., Liang, Z.R., Cao, W.G., Pomeroy, M.J., Huo, Y.M., Li, L.H., Barish, M.A., Abbasi, A.F., Pickhardt, P.J.: 3D-GLCM CNN: A 3-dimensional gray-level co-occurrence matrix based CNN model for polyp classification via CT colonography. *IEEE Trans. Med. Imaging* **39**(6), 2013–2024 (2020)
13. Bernal, J., et al.: Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. *IEEE Trans. Med. Imaging* **36**, 1231–1249 (2017)
14. Pipek, J., Varga, I.: Universal classification scheme for the spatial localization properties of one-particle states in finite d-dimensional systems. *Phys. Rev. A* **46**, 3148–3164 (1992)
15. Nagy, S., Sziová, B., Pipek, J.: On structural entropy and spatial filling factor analysis of colonoscopy pictures. *Entropy* **21**, 256 (32 pages) (2019)
16. Nagy, S., Lilik, F., Kóczy, L.T.: Entropy based fuzzy classification and detection aid for colorectal polyps. *IEEE Africon 2017, Cape Town, South Africa, 15–17 September 2017*
17. Nagy, S., Lilik, F., Kóczy, L.T.: The effect of image feature qualifiers on fuzzy colorectal polyp detection schemes using KH interpolation—towards hierarchical fuzzy classification of coloscopic still images. *WCCI-FuzzIEEE 2018, Rio de Janeiro, Brasil, 8–13 July 2018*
18. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. *Int. J. Comput. Vis.* **1**, 321–331 (1988)
19. Georgieva, V.M., Vassilev, S.G.: Kidney Segmentation in ultrasound images via active contours. *11th International Conference on Communications, Electromagnetics and Medical Applications, Athens, Greece (2016, October)*

20. Schlemper, R.J., Riddell, R.H., Kato, Y., et al.: The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* **47**, 251–255 (2000)
21. Inoue, H., Kashida, H., Kudo, S., Sasako, M., Shimoda, T., Watanabe, H., Yoshida, S., Guelrud, M., Lightdale, C., Wang, K., et al.: The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest. Endosc.* **58**, 343 (2003)
22. Schlemper, R.J., Hirata, I., Dixon, M.F.: The macroscopic classification of early neoplasia of the digestive tract. *Endoscopy* **34**, 163–168 (2002)
23. Hough, P.V.C.: Machine Analysis of Bubble Chamber Pictures. In: Proceedings of the 2nd International Conference on High Energy Accelerators and Instrumentation, HEACC 1959, CERN, Geneva, Switzerland, 14–19 September 1959
24. Freund, Y., Schapire, R.E.: A short introduction to boosting. *J. Jpn. Soc. Artif. Intell.* **14**, 771–780 (1999). (In Japanese)
25. Viola, P.; Jones, R.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. I-511–I-518, IEEE, 8–14 December 2001, Kauai, HI, USA
26. Iwhori, Y., Shinohara, T., et al.: Automatic polyp detection in endoscope images using a hessian filter. In: Proceedings of MVA2013, IAPR International Conference on Machine Vision Applications, pp. 1–4. Paper 3-1. Kyoto, Japan, 20–23 May 2013
27. Fourier, J-B-J.: *Théorie Analytique de la Chaleur*. Didot, Paris (1822)
28. Cheddad, A., et al.: Exploiting Voronoi diagram properties in face segmentation and feature extraction. *Pattern Recogn.* **41**, 3842–3859 (2008)
29. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886–893 (2005)
30. Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *Int. J. Comput. Vision* **75**, 247–266 (2007)
31. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001)
32. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
33. Zadeh, L.A.: Fuzzy algorithms. *Inf. Control* **12**, 94–102 (1968)
34. Kóczy, L.T., Hirota, K.: Approximate reasoning by linear rule interpolation and general approximation. *Int. J. Approximate Reasoning* **9**, 197–225 (1993)
35. Kóczy, L.T., Hirota, K.: Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases. *Inf. Sci.* **71**, 169–201 (1993)
36. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* **10**, 266–277 (2001)
37. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *Int. J. Comput. Vision* **22**, 61–79 (1997)
38. Whitaker, R.T.: A level-set approach to 3d reconstruction from range data. *Int. J. Comput. Vision* **29**, 203–231 (1998)
39. Rényi, A.: On measures of information and entropy. In: Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; pp. 547–561
40. Shannon, C.E.: A mathematic theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948)
41. Varga, I., Pipek, J.: Rényi entropies characterizing the shape and the extension of the phase space representation of quantum wave functions in disordered systems. *Phys. Rev. E* **68**, 026202 (2003)
42. Molnár, L.M., Nagy, S., Mojzes, I.: Structural entropy in detecting background patterns of AFM images. *Vacuum* **84**, 179–183 (2010)
43. Bonyár, A., Molnár, L.M., Harsányi, G.: Localization factor: a new parameter for the quantitative characterization of surface structure with atomic force microscopy (AFM). *MICRON* **43**, 305–310 (2012)

Chapter 11

MedSecureChain: Applying Blockchain for Delegated Access in Health Care



Tripti Rathee and Parvinder Singh

Abstract Blockchain in today's standard lacks delegated access and proper identity and access management (IAM) support. In the medical ecosystem, there are different parties, which have access to patient's data. However, every third party must not get all the data about the patient to preserve the user's privacy. In this paper, we present **MedSecureChain**, which is implemented on a private blockchain-based OAuth type authentication to protect and give the respective user control over their data. Delegated access to different categories of users is provided thus giving the user total control over his data. Asymmetric key cryptography has been used to achieve secure delegated access, wherein each different node sharing a common data uses different private keys for accessing the data that has been encrypted using the same public key.

11.1 Introduction

Digitalization has become popular in almost every business and therefore the data related to every individual is now becoming a part of economic assets. Nowadays, personalized services such as e-government, e-commerce and e-health applications involve and process significant amount of personal information [1]. The medical records of a patient generated by the hospitals are becoming honeypots of data for the intruders. Although there has been a lot of work going on to protect the privacy and security of patient's data, there have been continuous data breaches [2]. The major problem lies in the fact that the private data is put into central repositories and probably is under the control of third party authorities. The reliance on the central repository servers may prove to be highly ruinous. The assimilation of digital

T. Rathee (✉)

Maharaja Surajmal Institute of Technology, New Delhi, India

e-mail: rathee.tripti@msit.in

T. Rathee · P. Singh

Deenbandhu Chhotu Ram University of Science and Technology, Murthal, India

technologies and acquisition of tremendous amount of data needs some mechanisms which are able to determine absolutely who are the intended users [3] and authorize their basic characteristics like name, address and personality [1, 4–6]. Therefore, securing and providing access to user’s personal data has become a necessary security measure. There have been many algorithms in the literature [7–9] which focus on the problems arising from sharing of medical information in the healthcare industry. But they have been proved to be not very sufficient. With the availability of public/private key cryptography and distributed ledger technology named blockchain, the above-mentioned problem can be solved. The blockchain technology can be used to deliver the trust infrastructure to everyone who is a part of the network [10–12].

11.1.1 Our Contribution

The main contributions of this paper are described as follows: Firstly, we are creating a secure medical blockchain with OAuth like delegated access for doctors, hospitals, therapists, insurance companies, chemist, etc., to allow limited access to user’s data and allow accessing when authorization tokens expire. Delegated access to different categories of users is given at different levels so that a particular user can view only that data which has been authorized to him. The rest of the paper is organized as follows: Sect. 11.2 describes the background of blockchain technology and the need for the new type of delegated access. Section 11.3 describes the proposed MedSecureChain and Sect. 11.4 describes the user interface created with one practical example in health care. Finally, Sect. 11.5 concludes this paper.

11.2 Preliminaries

11.2.1 Blockchains

Blockchain technology was introduced in 2008 by “Satoshi Nakamoto” in his groundbreaking paper [13], wherein he claimed blockchain as a trustless technology. “Trustless” means that without involvement of the central authorities, the transaction which takes place over a computer network can be made public can be monitored and verified. Blockchain is considered to be a set of interconnected technologies which provides the functionality to the framework, as illustrated in Fig. 11.1.

Blockchain follows *distributed ledger technology* (DLT), which is fabricated, on the principles of *public key cryptography* and *point-to-point network* consisting of the participants called as *nodes*. The ledger is encrypted so as to keep the details of the people involved into it as anonymous. It is a data store of controlled records called as *blocks*. Every block in the blockchain is timestamped and is linked to the previous block by the hash of the previous block. The hash is calculated using a very

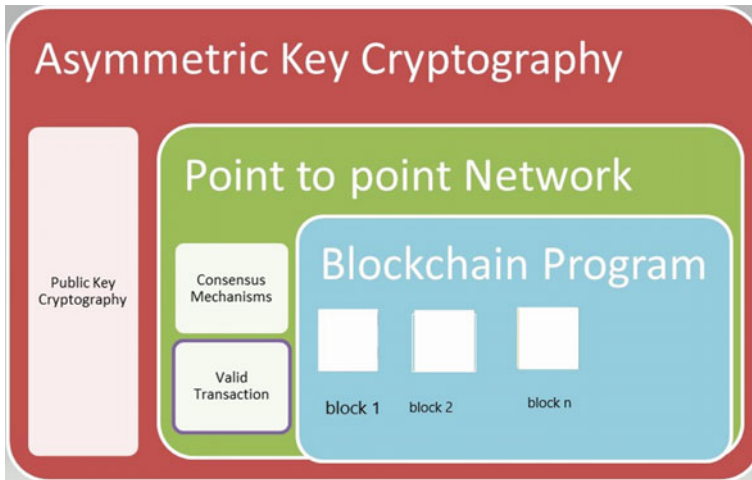


Fig. 11.1 Blockchain structure

complex mechanism or algorithm and is used to validate a transaction. The *consensus* mechanism is used to achieve this particular task. There are various consensus mechanisms available in the literature [13–15]. Once the transaction is verified, the block becomes a permanent part of the blockchain and the chain keeps on growing. Since the chain is linked with immutable blocks, it is named as *blockchain*. The process of creating a new block is called as *mining* and the people who mine the blocks are called as miners. The distributed ledger thus holds all the transactions that have been done by the nodes in the network and all the nodes hold the copy of the transaction involved which is kept in sync with other copies. The existing literature identifies blockchain into various categories [16–19]. We can categorize blockchain into three types: *public*, *private* and *consortium*. In *public* blockchain anybody is allowed to participate in the existing network and perform the transaction, whereas in *private* blockchain, the participating users are selected in advance and only those users are granted access to the network. The well-known examples of public and private blockchain are given in [20–22]. *Consortium* blockchain is semi-private blockchain which means they are *permissioned* blockchain.

11.2.2 Related Work Summary and Demand for New Type of Delegated Access Control

A cloud-based security problem has been addressed in [23]. The authors have tried to minimize the leakage of private information in health care. The authors in [24] have proposed an architecture and algorithm for electronic health record (EHR) system to grant access to participants and to achieve privacy and security of the data provided

by the patients. The authors in [25] have constructed the gateway for healthcare data so as to preserve the privacy of the users. The authors in [26] have provided an access control mechanism based on blockchain for sharing data over cloud. IAM has always been as emerging topic of research in the literature [27–30]. Identity management consists of two things; processes and the technology needed to determine whether a user has access to system or not, as well as to set the level of access a user has on a given system. Therefore, it is a necessity to build strong IAM models. Through our proposed system, we are able to secure user privacy and authorization is provided only to the people who are supposed to be liable for that particular record.

11.3 Proposed MedSecureChain in Health Care

11.3.1 Overview

MedSecureChain gives users access to information limited to their authority. This is achieved through three different Rivest–Shamir–Adleman (RSA) key pairs for the three expected user parties for each customer. The user registers with the application at the register endpoint. This triggers the key creation process and three pairs of RSA keys are created. The public keys for all the three pairs are kept at the server and fed to a JSON creation function, storing the respective user metadata for the keys. This public key JSON data is stored in the blocks of the blockchain. The block is mined using the proof-of-work (PoW) algorithm which confirms this block transaction (Fig. 11.2).

To get the data the user must enter their private key, which is then used to sign a message block from the client side. This signed message is sent to the endpoint, which authenticates the user and provides data according to the authorization table. Figure 11.3 shows the data retrieval flow.

11.3.2 Procedures

The procedures which are used are described in detail in this section

- (1) **Register_frontend (Client Side):** Take registration info from user.

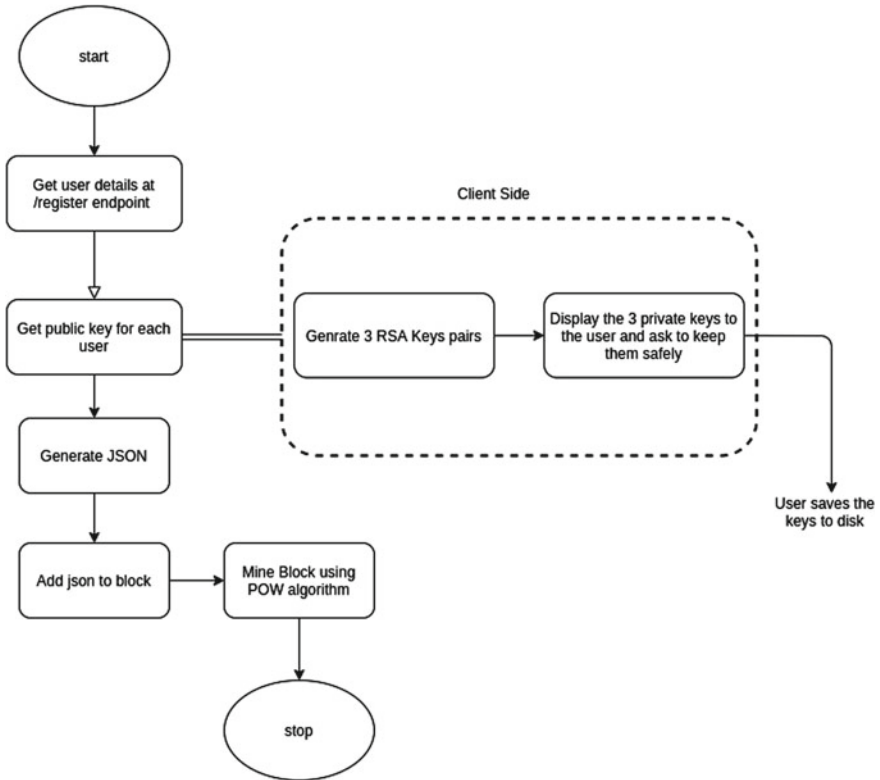


Fig. 11.2 Registration data flow

```
IF Username_taken(User_details[User_id]) then
Return "Choose a different Username"
Else
    Generate key_pair for Personal, Doctor & Chemist
    FOR EACH key_pairs as key_pairs
do
    Private_keys[] ← key_pair.private()
    Public_keys[] ← key_pair.public()
    END FOR
Return Register_backend (User_details, Public_keys)
END IF
```

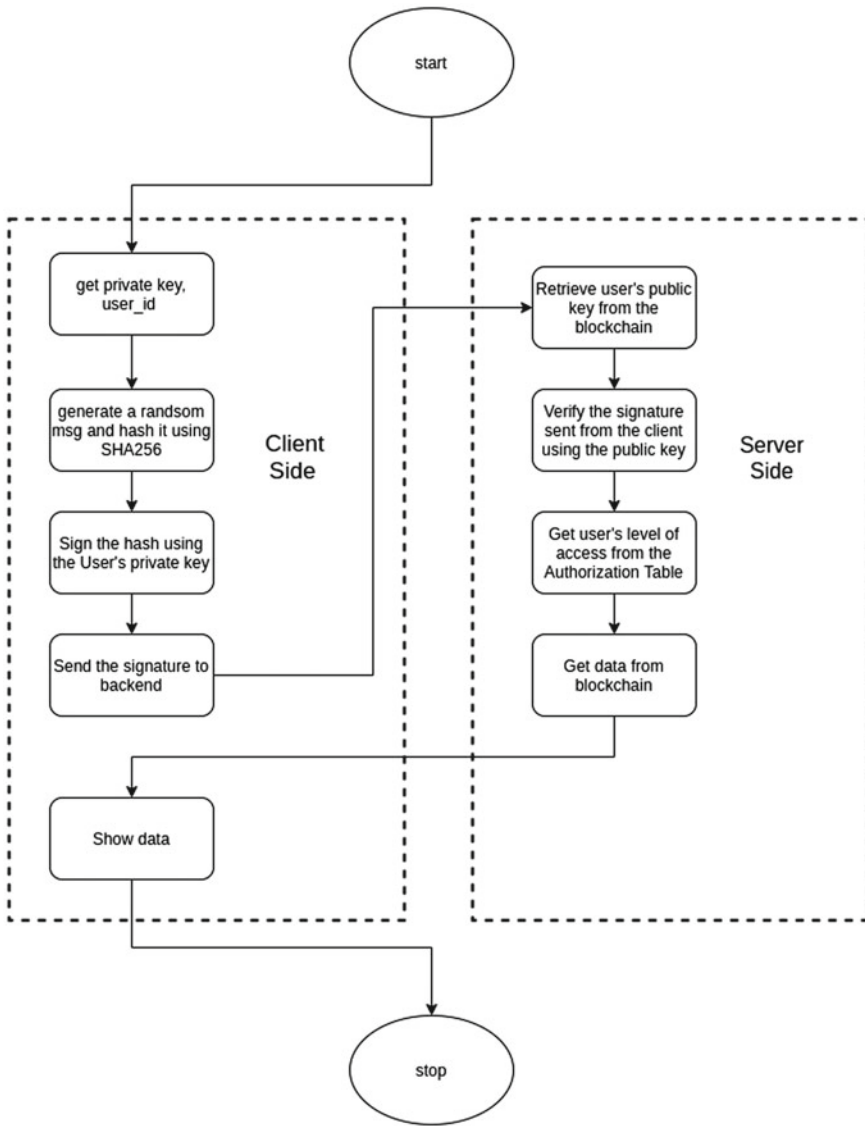


Fig. 11.3 Data retrieval flow

(2) **Write_data_backend:** Writing the data at blockchain

```

Public_key ←Get_public_key_from_blockchain (UserID, Form_data [Ac-
cess_level])
hasher ←SHA256.new(Challenge_msg)
verifier ←PKCS1_v1_5.new(key)
IF verifier.verify(hasher, Form_data[Dig_Signature]) then
    BlockChain.write_block (Data_json)
ENDIF

```

(3) **Get_data_backend:** Getting the data from blockchain

```

Public_key ←Get_public_key_from_blockchain (UserID, Form_Data [Ac-
cess_level])
hasher ←SHA256.new(Challenge_msg)

```

```

verifier ←PKCS1_v1_5.new(key)

```

```

IF verifier.verify(hasher, Form_Data[Dig_Signature]) then
IF Access_check (Access_table,Form_Data[Requested_data],
Form_Data [Access_level])
THEN
    Return Blockchain.read_data (User_id, Form_Data [Access_level])
ELSE
    Return "Not authorized for given data"

```

```

END IF
ELSE

```

```

    Return "Authentication Failed"

```

```

ENDIF

```

11.4 Implementation

The whole implementation is done in Python with desired requirements. It consists of a blockchain backend, user registration, user authentication, data retrieval and user-specific pages. There are basically three parties who are involved in working on our platform, the patient, doctor and chemist. Some relevant information about the patient's diagnosis is shared among them; while some information is private to the individual. Like, in case of patient, the doctor may want to know the patient's diagnosis history, but the insurance details should not be visible to the doctor. As soon as the patient registers, a 1024 bit RSA keys are generated which are used for authentication. The generated public key will be used as the identity. The initial form contains only private details about the user. After getting the details, the backend generates user ID and 3 RSA 1024 bit key pairs. The private keys are sent to the client side and are not stored by the server. The registration page is shown in Fig. 11.4.

The screenshot shows a web application interface for 'MegSecure Chain'. The header is teal and contains the logo on the left and a navigation menu with 'User', 'Doctor', and 'Chemist' on the right. The main content area is white and features a 'Register' form. The form includes the following fields and options:

- First Name:** Text input field containing 'John'.
- Last Name:** Text input field containing 'Doe'.
- Gender:** Radio button options for 'Male' (selected), 'Female', and 'Other'.
- Blood Group:** Text input field containing 'A+'.
- Age:** Text input field containing '82'.
- Height(cm):** Text input field containing '160'.
- Weight(kg):** Text input field containing '160'.
- Submit:** A blue button labeled 'SUBMIT'.

Fig. 11.4 Registration page

The diagnostic information is added to the blockchain, which takes the diagnostic information from the doctor and doctor's private key and use it to add the block to the blockchain. Adding the doctor's diagnostic data to blockchain is shown in Fig. 11.5.

The authorization table is used to define access levels to various users and third party plugins. Here, numeric access codes are given to each level of information and to each type of user. Getting the data with delegated access is shown in Fig. 11.6 .

11.5 Conclusion

This paper has presented a blockchain-based ecosystem with enhanced encryption algorithms and delegated access of different users at different levels which is a new type of identity and access management for any healthcare industry. The procedure for each operation has been described in detail. In addition to this, a practical example with user interface has been also described which clearly depicts the feasibility of the proposed work. Using blockchain for delegated access, it definitely has

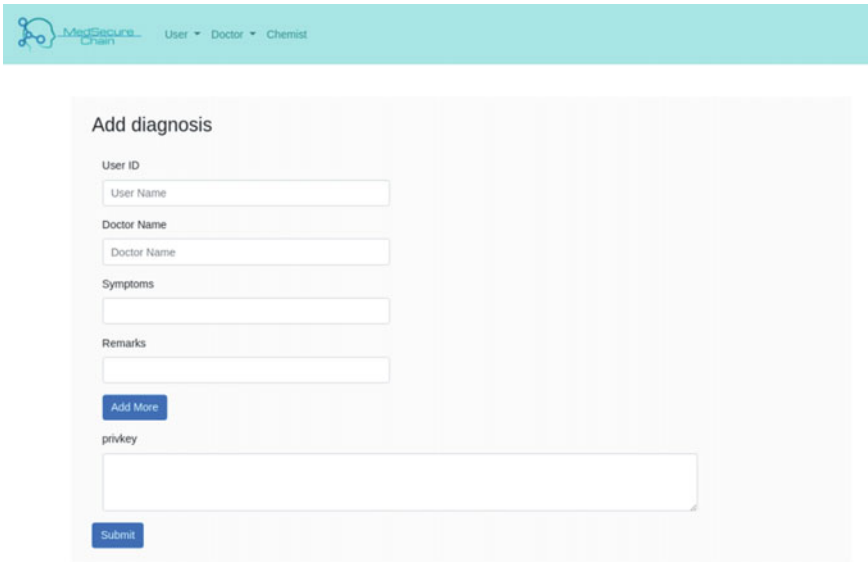


Fig. 11.5 Adding doctor’s diagnostic data to blockchain

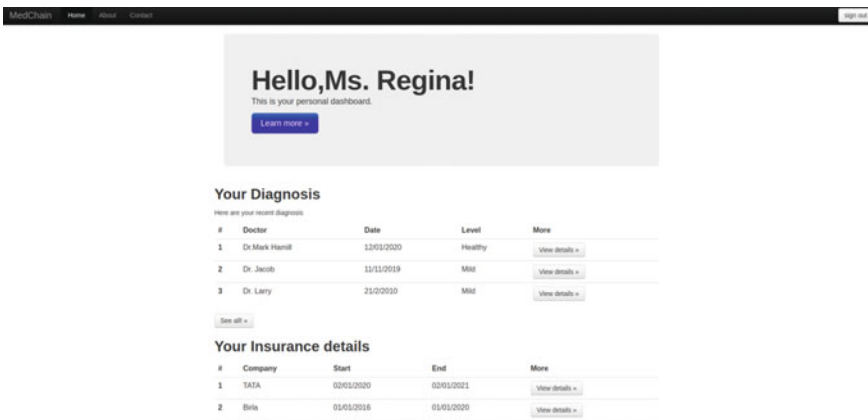


Fig. 11.6 Getting data with delegated access

room for improvement. First of all, it can be applied to large set of users to check the performance and interoperability. Secondly, the private key can be stored on a hardware-based wallet so as to improve the security of the whole system.

Acknowledgements This work is part of the bilateral Indian-Bulgarian cooperation research project between Technical University of Sofia, Bulgaria and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonapat, India financed by the Department of Science and Technology (DST), India and the Ministry of Education and Science, Bulgaria.

References

1. Buchmann, N., Rathgeb, C., Baier, H., Busch, C., Margraf, M.: Enhancing breeder document long-term security using blockchain technology. In: 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), pp. 744–748 (2017)
2. Dagher, G.G., Mohler, J., Milojkovic, M., Marella, P.B.: Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustain Cities Soc.* **39**, 283–297 (2018)
3. Lee, J.-H.: BIDaaS: Blockchain based ID as a service. *IEEE Access* **6**, 2274–2278 (2017)
4. Lemieux, V.L.: Trusting records: is Blockchain technology the answer? *Rec. Manag. J.* (2016)
5. Augot, D., Chabanne, H., Chenevier, T., George, W., Lambert, L.: A user-centric system for verified identities on the bitcoin blockchain. In: *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pp. 390–407. Springer, Berlin (2017)
6. Leiding, B., Norta, A.: Mapping requirements specifications into a formalized blockchain-enabled authentication protocol for secured personal identity assurance. In: *International Conference on Future Data and Security Engineering*, pp. 181–196 (2017)
7. Thilakanathan, D., Chen, S., Nepal, S., Calvo, R.A., Liu, D., Zic, J.: Secure multiparty data sharing in the cloud using hardware-based TPM devices. In: *2014 IEEE 7th International Conference on Cloud Computing*, pp. 224–231 (2014)
8. Khan, A.N., Kiah, M.L.M., Ali, M., Madani, S.A., Shamshirband, S.: others: BSS: block-based sharing scheme for secure data storage services in mobile cloud environment. *J. Supercomputing*. **70**, 946–976 (2014)
9. Dong, X., Yu, J., Luo, Y., Chen, Y., Xue, G., Li, M.: Achieving an effective, scalable and privacy-preserving data sharing service in cloud computing. *Comput. Secur.* **42**, 151–164 (2014)
10. Peterson, K., Deeduvanu, R., Kanjamala, P., Boles, K.: A blockchain-based approach to health information exchange networks. In: *Proceedings of NIST Workshop Blockchain Healthcare*, pp. 1–10 (2016)
11. Pilkington, M.: Blockchain technology: principles and applications. In: *Research Handbook on Digital Transformations*. Edward Elgar Publishing (2016)
12. Liu, P.T.S.: Medical record system using blockchain, big data and tokenization. In: *International Conference on Information and Communications Security*, pp. 254–261 (2016)
13. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. [bitcoin.org. https://bitcoin.org/bitcoin.pdf](https://bitcoin.org/bitcoin.pdf) (2008). Accessed 24 Feb 2020
14. Castro, M., Liskov, B.: Practical Byzantine fault tolerance and proactive recovery. *ACM Trans. Comput. Syst. (TOCS)* **20**, 398–461 (2002)
15. Cachin, C., et al.: Architecture of the hyperledger blockchain fabric. In: *Workshop on Distributed Cryptocurrencies and Consensus Ledgers* (2016)
16. Zheng, Z., Xie, S., Dai, H.-N., Chen, X., Wang, H.: Blockchain challenges and opportunities: a survey. *Int. J. Web Grid Serv.* **14**, 352–375 (2018)
17. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. *Ieee Access* **4**, 2292–2303 (2016)
18. Buterin, V.: On public and private blockchains.[online] Ethereum Blog (2015)
19. Kravchenko, P.: Ok, I need a blockchain, but which one? Medium. <https://medium.com/@pavlekravchenko/ca75c1e2100> (2016)

20. Sasson, E. Ben, Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., Virza, M.: Zerocash: decentralized anonymous payments from bitcoin. In: 2014 IEEE Symposium on Security and Privacy, pp. 459–474 (2014)
21. Wood, G.: Ethereum: A secure decentralized generalized distributed ledger. Ethereum & Ethcore, London, UK, Tech. Rep. (2018)
22. Brown, R.G., Carlyle, J., Grigg, I., Hearn, M.: Corda: an introduction. *R3 CEV* **1**, 15 (2016, August)
23. Fernando, R., Ranchal, R., An, B., Othman, L. Ben, Bhargava, B.: Consumer oriented privacy preserving access control for electronic health records in the cloud. In: 2016 IEEE 9th International Conference on Cloud Computing (CLOUD), pp. 608–615 (2016)
24. Tanwar, S., Parekh, K., Evans, R.: Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *J. Inf. Secur. Appl.* **50**, 102407 (2020)
25. Yue, X., Wang, H., Jin, D., Li, M., Jiang, W.: Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *J. Med. Syst.* **40**, 218 (2016)
26. Xia, Q.I., Sifah, E.B., Asamoah, K.O., Gao, J., Du, X., Guizani, M.: MeDShare: trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access* **5**, 14757–14767 (2017)
27. Halperin, R., Backhouse, J.: A roadmap for research on identity in the information society. *Identity Inf. Soc.* **1**, 71–87 (2008)
28. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management technical report, TU Dresden and ULD Kiel. https://dud.inf.tu-dresden.de/Anon_Terminology.shtml (2010)
29. Jensen, J.: Federated identity management challenges. In: 2012 Seventh International Conference on Availability, Reliability and Security, pp. 230–235 (2012)
30. Jensen, J., Jaatun, M.G.: Federated identity management—we built it; why won't they come? *IEEE Secur. Priv.* **11**, 34–41 (2012)

Chapter 12

Cosine and Soft Cosine Similarity-Based Anti-Phishing Model



Bhawna Sharma, Parvinder Singh, and Jasvinder Kaur

Abstract Phishing attack has posed a greater threat to user information over network. In addition to the existence of various disguise illegal URL's, instances had been seen when users are redirected to phishing URL's that challenges their privacy concern. In the current work, author tried to develop an effective anti-phishing method based on hybrid similarity approach combining cosine and soft cosine similarity that measures the resemblance between user query and database. The strength of the proposed hybrid approach is further enhanced with the incorporation of feed forward backpropagation neural network (FFBPNN) so as to validate the similarity-based predictions. The model evaluated against 3000 sample files demonstrated to effectively detect phishing attacks with positive predictive value, true positive rate and F-measure of 71.9%, 72.6% and 72.23%, respectively.

12.1 Introduction

Modern advances in the technological sector have raised the popularity of Internet technology. Presently, none of the field exists that remains untouched by the network frame work of Internet. The rising popularity social media including Twitter, Instagram and Facebook further adds up to the popularity of Internet technology as an indispensable daily need [1]. It is observed that the number of people using social media has been nearly doubled since 2014. As shown in Fig. 12.1, 9% rise in the

B. Sharma · P. Singh (✉)

Department of Computer Science and Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonapat, Haryana, India
e-mail: parvindarsingh.cse@dcrustm.org

B. Sharma

e-mail: bhawnash024@gmail.com

J. Kaur

Computer Science and Engineering, PDM University, Bahadurgarh, Haryana, India
e-mail: jasvinder.kaur@pdm.ac.in

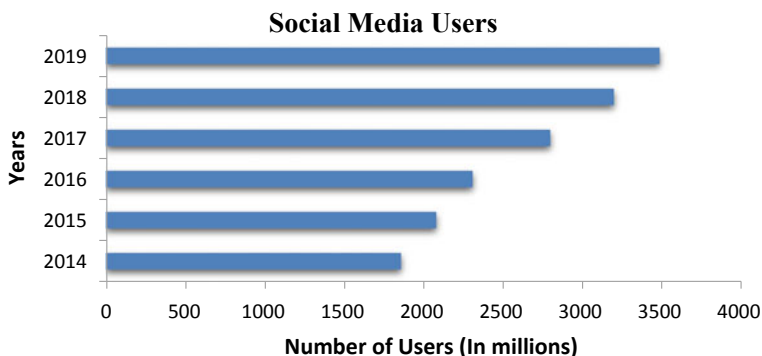


Fig. 12.1 Rising popularity of social media

number of users has been observed in last year [2]. Such an enormous rise in the network traffic has significantly raised twofold challenges. One is malicious attacks and other is requirement of advanced hardware. This had further led to the discovery of big data computing technologies with more sophisticated hardware configuration. Internet has become a means that connects numerous users sharing information in the form of media, images, videos, files, etc., that also require attention for the risk of privacy leakage in one way or the other.

Phishing is one of the cyber-attacks that frequently appear in personal computers and mobile platforms. It is a criminal mechanism taking advantage of both social engineering and technical subterfuge that advertises and sends illegal links to users to deception their private information and financial credentials. Spoofed emails are employed as legitimate business tricking recipients to accidentally share their login and password details. In other words, phishing is defined as stealing someone private information by befooling them to be genuine [3–5]. Majority of the Internet users get fascinated by these illusions and get trapped. Software-as-a-Service (SaaS) and webmail services are the industry sectors, which are the main targets of phishing [6].

Researchers around the world have been constantly involved to tackle phishing attacks and deploy various anti-phishing protocols. It is found that one-third of the anti-phishing mechanisms are rule-based prevention methods implemented in recent past [7]. The major limitation adjoining this mechanism is the adaptability. This means that there is requirement of incorporation of new data of rule set with every new phishing attack. To deal with adaptability issues, authors have used swarm intelligence with machine learning [8]. Some similarity-based approaches have also been proposed that compares the phishing with legitimate websites [9]. In the current research, authors used rule-based architecture to develop an anti-phishing protocol based on the cosine and soft cosine similarity index powered by the machine learning design. Cosine similarity measures the resemblance between two documents independent of the document size. In Fig. 12.2, cosine angle is reflected by a set of vectors representing query and repository vectors projected in the space. The plus point of this similarity aspect is the fact that larger documents can still be expected

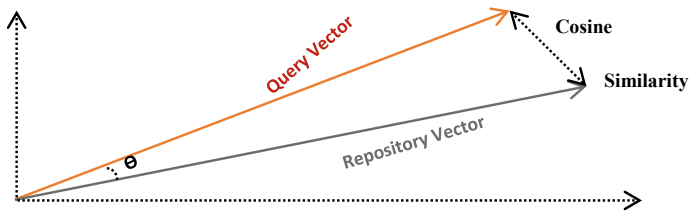


Fig. 12.2 Cosine similarity

to get oriented in the environs. Thus, smaller cosine angle reflects higher similarity, for instance, cosine 0° reflects complete similarity as 1. More detailed mechanism is discussed in the later part of the article.

Author had introduced phishing attacks and related statistics in the current section. Section 12.2 summarizes the research revolving around anti-phishing protocols deployed by researchers, and Sect. 12.3 describes the proposed methodology based on cosine and soft cosine followed by Sect. 12.4 dedicated for observations and discusses the accomplish result. The paper is concluded in Sect. 12.5.

12.2 Literature Review

This section covers various types of protocols proposed to deal and prevent phishing attacks. Ramanathan et al. in [7] had proposed PhishGILLNET as multi-layered anti-phishing model. The technique demonstrated effective results with a prerequisite that webpage should be in HTML and MIME formats [7]. Li et al. [10] proposed a novel anti-phishing method based on ball-support vector machine (BVM) to distinguish a malicious URL from a genuine. In the process they extracted various topological features of the website and analysis, 12 out of them followed by the BVM-based vector analysis. Evaluation against SVM proved BVM to be highly effective in detecting phishing websites with a relatively slower speed for big data [10]. Kaur and Kalra [11] had proposed a five-tier anti-phishing design to protect. The hybrid approach analyses the URL and reflects the page status as secure or phishing website. [11]. In 2016, Nguyen et al. proposed a novel neuro-fuzzy ideal for detecting phishing attacks. The technique had employed a dataset of legitimate (10,000) and phishing websites (11,660) that was trained using neural network with adaptive learning rates. The results of the study showed its effectiveness in identifying the phishing websites [12]. Sonowal and Kuppusamy [13] developed PhiDMA as a multi-layered anti-phishing architecture divided into five layers corresponding to whitelist, URL feature, signature, string matching and score layer. The model was developed to offer easy access to even visually impaired individuals with 92.72% phishing detection accuracy [13]. Ugochi [14] focused his research towards depth analysis of IP address and URL cosine similarity in order to identify phishing URLs.

Experimental evaluation proved to be highly effective against 100 phishing URL used in the study with least memory requirement [14]. Makki et al. [15] postulated a cost-sensitive K-nearest neighbour (KNN) approach enhanced with cosine similarity to identify cheat instances affecting financial market, money transactions, telecommunication and credit card. The model proved to outperform the approach based on traditional KNN alone [15]. Jain et al. in 2018 had presented an innovative approach that was based on outstanding hyperlink features that aids in the recognitions of phishing attacks. It offered an effective client side solution and on logistic regression classifiers and it exhibited an accuracy of even higher than 98.4% [16]. Krodestani and Shajari in [9] proposed a novel textual similarity-based method to identify phishing sites. The method was evaluated against real website and demonstrated optimal phishing detection accuracy while discriminating between phishing and legitimate website and guides user towards genuine website [9]. Azeez in [17] had developed PhishDetect technique aimed to identify phishing websites by evaluating URL features and web contents. The technique proved to be highly efficient in identifying the phishing URLs [17]. Morovati et al. [18] dedicated their research study towards the identification of phishing attacks spread through phishing emails. To achieve distinguishable results, they had incorporated email forensic analysis with machine learning methodology [18]. In the same year, Zhu et al. had proposed OFS-NN as a phishing website detector. This model was based on optimal feature selection approach followed by neural network (NN) technique. They had developed an optimal classifier that could accurately detect different types of phishing websites [19]. Li in [20] had developed an architecture that mediates a cascade of Kaizen events. In this approach, the cosine similarity of data objects is used to classify protection levels. Experimental evaluation had shown that the designed fuzzy architecture proved to be competent as compared to other methods to support Kaizen architecture to identify susceptibility of web applications [20].

12.3 Proposed Methodology

12.3.1 Dataset

Current research work employs the dataset obtained from *PhishTank* [21]. The downloadable data attributes consist of Phishing Id, URL information, type of target, online status, etc. It also presents a community-based evaluation platform where users query is classified to be phishing or legitimate-based votes. The site could be accessed at <https://www.phishtank.com/>.

The proposed anti-phishing design is largely a two-layered architecture. First layer is calculating the website similarity based on hybrid cosine and soft cosine between the input user query and the database repository. The second layer functions as a validation layer to check the effectiveness of the prediction performed by the first layer. This layer classifies the phishing and non-phishing sites based on a multiclass

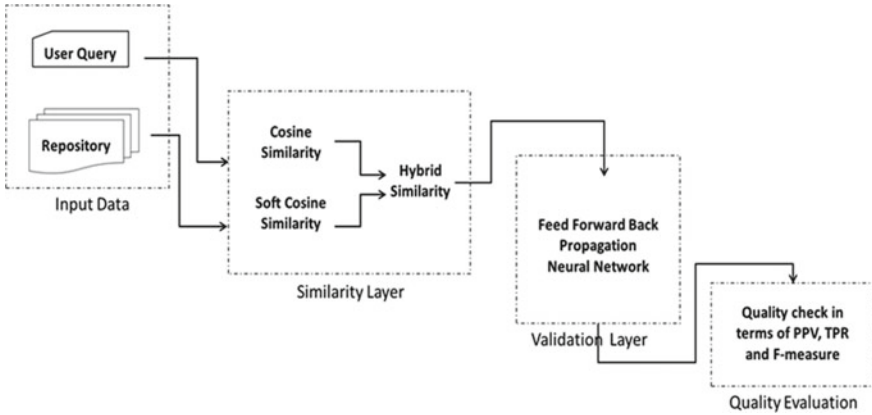


Fig. 12.3 Proposed multi-layered anti-phishing architecture

neural network (NN). The quality of the proposed architecture is evaluated in terms of positive predictive value (PPV), true positive rate (TPR) and F-measure. The overview of the steps is shown in Fig. 12.3.

12.3.2 Similarity Layer (SL)

The layer is dedicated to perform similarity predictions between the query or the test data and the repository. To achieve this, author had proposed a hybrid cosine and soft cosine-based similarity predictions that take advantage of angular co-relation established between the query and the repository vector as shown in Fig. 12.3, where “ θ ” defines the angle stretched by the two vectors. Mathematically, similarity is calculated as follows:

$$\text{Cosine}_{\text{Sim}} = \sum_{i=1}^n \frac{Q_{\text{Vect}(i)} R_{\text{Vect}}}{\sum_{i=1}^n (Q_{\text{Vect}})^2 \sum_{j=1}^n (R_{\text{Vect}})^2} \quad (12.1)$$

where, $\text{Cosine}_{\text{Sim}}$ represents, cosine similarity observed between user query represented by Q_{Vect} and repository represented by R_{Vect} . The pseudocode to compute the similarity is summarized in Algorithm 1.

Algorithm 1: Pseudocode for Cosine Similarity.

1. Input: R_{Value} // Repository data values
2. For each I_{Val} in R_{Value} // Scan Every data value present in the repository
3. Vectorization of repository data:
 $R_{\text{Vect}} = \text{conversion}(R_{\text{Value}})$ // prune out stop words from the list.

4. Isolate words from repository data
 $w_{\text{data}} = \text{Segregate}(R_{\text{Vect}})$ // segregate words from repository files.
5. Eliminate Stop Words
 $w_{\text{stop}} = \text{Remove}(w_{\text{data}})$ // removal of stop words.
6. ASCII code generation
 $w_{\text{data_ASCII}} = \text{ASCII}(w_{\text{stop}})$ // generate ASCII Code for each word.
7. Calculate Cosine Similarity:

$$\text{Cosine}_{\text{Sim}} = \sum_{i=1}^n \frac{Q_{\text{Vect}(i)} R_{\text{Vect}}}{\sum_{i=1}^n (Q_{\text{Vect}})^2 \sum_{j=1}^n (R_{\text{Vect}})^2}$$

Store to List

9. Output: $\text{Cosine}_{\text{Sim}}$ // Cosine similarity between query and repository.
10. End for

The above algorithm calculates the cosine similarity to predict the phishing sites or URLs. In the process, it first converts the data values to vectors and evaluates the stop words present in the repository data and the query data. Another similarity aspect, i.e., soft cosine is also calculated that takes the advantage of the same algorithmic flow except the similarity calculation made in step 7. The mathematically soft cosine similarity is calculated as follows:

$$\text{SCosine}_{\text{Sim}} = \frac{\sum_{i,j} Q_{\text{Vect}_i} R_{\text{Vect}_j}}{\sum_{i,j} Q_{\text{Vect}_i} Q_{\text{Vect}_j} \sum_{i,j} R_{\text{Vect}_i} R_{\text{Vect}_j}} \quad (12.2)$$

where, $\text{SCosine}_{\text{Sim}}$ represents the soft cosine similarity observed between user query represented by Q_{Vect} and the repository data represented by R_{Vect} .

Further, the hybrid similarity prediction is performed by combining the observed similarity predicted by individual similarity calculations, i.e., using cosine similarity and soft cosine similarity as follows:

$$H_{\text{Sim}} = \text{Cosine}_{\text{Sim}} + \text{SCosine}_{\text{Sim}} \quad (12.3)$$

where hybrid similarity prediction is represented by H_{Sim} based on the similarity predictions made by cosine and soft cosine similarity calculators represented by $\text{Cosine}_{\text{Sim}}$ and $\text{SCosine}_{\text{Sim}}$, respectively. The similarity calculations made in this layer is sent to validation layer.

12.3.3 Validation Layer (VL)

The current layer evaluates the similarity predications of the similarity layer based on neural network (NN). It consists of input layer that inputs the user query, hidden

layer that act as a processing framework based on the weights and the output layer that returns the classification results. It is important to understand that raw data from input layer is passed to the hidden layer in the form of sigmoid function. Algorithm 2 summarizes the steps employed in the validation of the similarity predictions using neural network.

Algorithm 2: Pseudocode for Validation using NN.

1. Input: H_{Sim} // hybrid similarity value
2. Initialize Variables:
 T_{Val} // training value.
 G_{Val} // group value.
3. Assign training value:
 $T_{Val} = H_{Sim}$ // assign respective training value.
4. Assign group value:
 $G_{Val} = G_{num}$ // group value is represented by respective group number.
5. Assign NN parameters:
 $D_{Ratio} = 0.7$ // distribution ratio.
 $C_{V_{ratio}} = 0.15$ // cross validation ratio.
 $T_{ratio} = 0.15$ // test ratio.
 $N_{num} = 20$ neurons //number of neurons.
6. Initialize Neural Network for Training
 $Train_{NN}(T_{Val}, G_{Val}, N_{num})$ // initialize Neural Network.
7. Start Training Neural Network
8. For each X in $F_{classified}$ //for every value in classified frame
9. If $F_{Val} == T_{Val}$ // classified result matches with the training value
10. $T_{class} ++$ //auto increment True classified class label
11. Else
12. $F_{class} ++$ //auto increment False classified class label
End if
End for

The similarity predications made in the previous layer are arranged in groups. The blocks in a group represent the similarity calculations corresponding to each repository. However, similarity predictions corresponding to a repository are represented by a single group tagged by repository name. The obtained sets then undergo supervised learning using neural network (NN). This multiclass classifier is used to distinguish phishing sites from legitimate sites and URLs. The group value and classified value are compared for each query value. If the group value and the classified values match, NN classifies the query to the true else false.

12.4 Result and Discussions

The phishing predications made by the proposed framework are evaluated in terms of demonstrated quality parameters, namely positive predictive value (precision), true positive rate (recall) and F-measure.

- a. Positive predictive value (PPV) is represented by the number of true detections made by the prediction model in comparison with the total number of detections. Mathematically, it can be calculated as follows:

$$PPV = \frac{T_{\text{positive}}}{(T_{\text{positive}} + F_{\text{positive}})} \quad (12.4)$$

where T_{positive} and F_{positive} represent true positive and false positive detections, respectively.

- b. True positive rate (TPR) defines a number of positive results obtained that are actually correct or positive. It can be calculated as follows:

$$TPR = \frac{T_{\text{positive}}}{(T_{\text{positive}} + F_{\text{negative}})} \quad (12.5)$$

where F_{negative} represents the false negative detections.

- c. F-measure represents the harmonic mean of above two parameters. It is calculated by product and arithmetic summation as follows:

$$F_{\text{measure}} = 2 * \left(\frac{PPV * TPR}{PPV + TPR} \right) \quad (12.6)$$

The values obtained using above relationships are summarized in Table 12.1. The column 2, 3 and 4 represents the average value of PPV, TPR and F-measure observed for respective number of files mentioned in column 1.

It is generalized from Table 12.1 that increase in the number of test files increases the average value of each considered parameter. Figure 12.4 shows the average value of PPV, TPR and F-measure for different number of test files. It concludes that

Table 12.1 Average value of PPV, TPR and F-measure

Number of test files	PPV	TPR	F-measure
100	0.709	0.718	0.713
200	0.714	0.721	0.717
500	0.715	0.722	0.718
1000	0.719	0.725	0.722
2000	0.725	0.729	0.727
3000	0.732	0.738	0.735

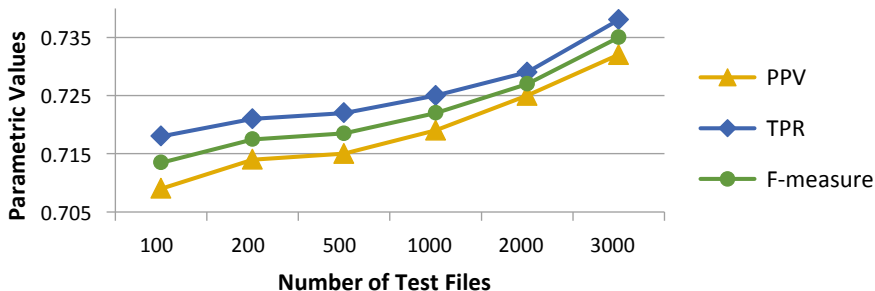


Fig. 12.4 Average value of PPV, TPR and F -measure

best performance (maximum value of each parameter) is obtained by considering maximum number of test files.

12.5 Conclusion

In the current work, author has proposed an anti-phishing framework based on new rule-based architecture. The first layer of anti-phishing model performs phishing detection based on cosine and soft cosine similarity followed by neural network machine learning for performing cross-validation of the prediction results. The quality of results is evaluated in terms of PPV, TPR and F -measure. It is observed that the prediction model showed average enhanced PPV of 2.3%, TPR of 2% and F -measure of 2.15% over 3000 test files. However, an average value of PPV of 0.719, TPR of 0.726 and F -measure of 0.722 is observed that proved the effectiveness of the proposed anti-phishing design.

Acknowledgements This work is part of bilateral Indian-Bulgarian cooperation research project between Technical University of Sofia, Bulgaria and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonapat, India under the title “Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications”, financed by the Department of Science and Technology (DST), India and the Ministry of Education and Science, Bulgaria.

References

- Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., Nunge, E.: Anti-phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In: Proceedings of the 3rd Symposium on Usable Privacy and Security, pp. 88–99. ACM, Pittsburgh, Pennsylvania, USA (2007)
- Published in Global Digital Report 2019. <https://datareportal.com/>. Last accessed 2020/02/21

3. Cao, Y., Han, W., Le, Y.: Anti-phishing based on automated individual white-list. In: Proceedings of the 4th ACM workshop on Digital identity management, pp. 51–60. ACM (2008)
4. Aksu, D., Turgut, Z., Üstebay, S., Aydin, M.A.: Phishing analysis of websites using classification techniques. In: International Telecommunications Conference, pp. 251–258. Springer, Singapore (2019)
5. Lam, T., Kettani, H.: PhAttApp: A phishing attack detection application. In: Proceedings of the 2019 3rd International Conference on Information System and Data Mining, pp. 154–158. Houston, TX, USA (2019)
6. Anti-Phishing Trend Report (Q2 2019). <https://apwg.org/trendsreports/>
7. Ramanathan, V., Wechsler, H.: phishGILLNET—phishing detection methodology using probabilistic latent semantic analysis, AdaBoost, and co-training. *EURASIP J. Inf. Secur.* **1**(1), 1–22 (2012)
8. Mensah, P., Blanc, G., Okada, K., Miyamoto, D., Kadobayashi, Y.: AJNA: Anti-phishing JS-based visual analysis, to mitigate users' excessive trust in SSL/TLS. In: 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), pp. 74–84. IEEE, Kyoto, Japan (2015)
9. Kordestani, H., Shajari, M.: A similarity-based framework for detecting phishing websites. *Int. J. Adv. Res. Comput. Sci.* **9**(1), 792–796 (2018)
10. Li, Y., Yang, L., Ding, J.: A minimum enclosing ball-based support vector machine approach for detection of phishing websites. *Optik* **127**(1), 345–351 (2016)
11. Kaur, D., Kalra, S.: Five-tier barrier anti-phishing scheme using hybrid approach. *Inf. Secur. J. Glob. Perspect.* **25**(4–6), 247–260 (2016)
12. Nguyen, L.A.T., Nguyen, H.K., To, B.L.: An efficient approach based on neuro-fuzzy for phishing detection. *J. Autom. Control Eng.* **4**(2), 159–165 (2016)
13. Sonowal, G., Kuppasamy, K.S.: PhiDMA—a phishing detection model with multi-filter approach. *J. King Saud Univ. Comput. Inf. Sci.* **32**(1), 99–112 (2020)
14. Ugochi, O. C.: A novel web page anti-phishing approach using URL cosine similarity and IP address comparison. In: International Conferences on WWW/Internet, ICWI 2018 and Applied Computing pp. 321–328. IADIS Press (2018)
15. Makki, S., Haque, R., Taher, Y., Assaghir, Z., Hacid, M. S., Zeineddine, H.: A cost-sensitive cosine similarity K-nearest neighbor for credit card fraud detection. *Big Data Cyber-S Secur. Intell.* 42–47 (2018)
16. Jain, A.K., Gupta, B.B.: A machine learning based approach for phishing detection using hyperlinks information. *J. Ambient Intell. Human. Comput.* **10**(5), 2015–2028 (2019)
17. Azeez, N., Salaudeen, B., Misra, S., Damaševičius, R., Maskeliūnas, R.: Identifying phishing attacks in communication networks using URL consistency features. *Int. J. Electron. Secur. Digit. Forensics* **12**(2), 200–213 (2020)
18. Morovati, K., Kadam, S.S.: Detection of phishing emails with email forensic analysis and machine learning techniques. *Int. J. Cyber-S Secur. Digit. Forensics* **8**(2), 98–108 (2019)
19. Zhu, E., Chen, Y., Ye, C., Li, X., Liu, F.: OFS-NN: an effective phishing websites detection model based on optimal feature selection and neural network. *IEEE Access* **7**, 73271–73284 (2019)
20. Lin, K.S.: New attack potential measurement method to kaizen event for web application security vulnerabilities. *Int. J. Electron. Commerce Stud.* **10**(2), 89–112 (2019)
21. PhishTank. <https://www.phishtank.com/>, last accessed 2020/02/01

Chapter 13

Enhanced Image Steganography Technique Using Cryptography for Data Hiding



Jasvinder Kaur and Shivani Sharma

Abstract The fast improvement of the Web has expanded the simplicity of sharing data to individuals around the world. In any case, this headway additionally raises a trouble about information control when the data is transferred by the sender to the beneficiary. Along these lines, data security is a significant issue in information correspondence. Steganography and cryptography assume significant jobs in the field of data security. Steganography can be applied, however, a different computerized media, for example, video, pictures, and sound to cover data in such a manner that nobody else realizes that there is a secret data. Cryptography alludes to the specialty of changing over a plaintext (message) into an ambiguous organization. Both steganography and cryptography strategies are strong. The contents obtained after doing so is secret and its existence is also hidden. This method is tested and it is observed that it prevents steganalysis too as well as parameters like PSNR and MSE are also tested which gave good results.

13.1 Introduction

Over past decades, the use of Internet has helped in devising different mechanisms for transmitting data from sender to the recipient. It has opened the new entryway for the aggressor to assault on that data and effectively takes the data of the clients [1]. There is a lot of classified data like military's privileged insights, law requirement's insider facts, and so forth. Fear-based oppressors can likewise utilize information covering up just as cryptography strategies for making sure about their information. Cryptography is the study of utilizing science to encode and decode the information. Steganography is a workmanship and science of concealing the mystery information into another

J. Kaur (✉) · S. Sharma

Department of Computer Science and Engineering, PDM University, Bahadurgarh, Haryana, India
e-mail: jasvinder.kaur@pdm.ac.in

S. Sharma

e-mail: shivanibitto@gmail.com

media. This paper presents the combination of cryptography and steganography so as to ensure effective delivery of message to the recipients.

13.1.1 Steganography

Steganography is characterized as the “concealed composition” that assists with concealing the nearness of data. By doing this, it becomes very hard for gate crashers to retrieve the information as it is hard to track down the contrast between the images [2]. The sender is communicating something specific which has been implanted by a mystery key and a stego-picture is framed. This procedure is known as steganalysis, after this the picture is additionally handled and the collector will separate the picture with utilization of the key. Thus, the message will be effectively sent to the beneficiary. Steganography assists with concealing the record in the different structure, for example, in picture, sound, video, or text. Also, the target to do this is concealing the accessibility of information in the spread picture that is mixed up by people. Steganography involves essential three segments which are transporter, message, and key [3]. The bearer can be a picture, media player, or a TCP/IP parcel. Furthermore, a key is used to encode or unscramble a message and the secret phrase can be anything, a design, or a video. The thought driving steganography is that if an individual needs to send a message to other person, the correspondence between them is compelled by a switch or server. We can watch that one gathering needs to move a message to beneficiary, to do so it inserts message into a spread picture and gets a stego-picture. In a standard definition, this methodology of exemplifying message is not referred to and is kept as a secret between the two [4]. In any case, it is seen that the calculation being utilized is not a riddle yet the key is secret between the two, and it is otherwise called Kerchoff’s rule. It is a method to make sure about the unstable information [5] (Fig. 13.1).

13.1.2 Cryptography

We can define cryptography as the process in which it encrypts the actual message and converts it into secret message [6]. This message can be retrieved by receiver by using either private key or public key. Also, the message can be encapsulated by using a mathematical equation or algorithm and converts it into a non-readable form which can be any mystery information. Cryptography helps to encapsulate the data in order to protect it from an attacker, whereas steganography helps to conceal the presence of that data from the attacker. So, the combination of these two techniques gives superior communication. The aim of this paper is to combine the technologies and have better experience. Some basic terms are as follows:

- Load: data that requires to be concealed.

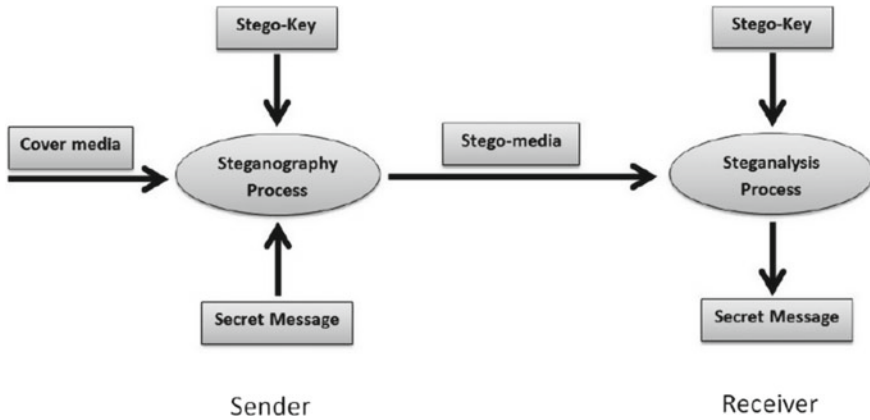


Fig. 13.1 Generalized steganographic technique

- Carrier file: medium in which the load has to be covered up.
- Stego-medium: place where the data has to be covered up [7].
- Redundant-bits: the data inside a record, which can be adjusted without harming the document.
- Steganalysis: The way toward identifying the hidden data which is put away inside a document [8] (Fig. 13.2).

This is the figure for cryptography:

So, now, the paper contains Sect. 2 which has the review and analysis briefly described that are about our proposed technique. Section 3 describes the presented technique which is followed by results and discussions in Sect. 4. The conclusion and future scope are present in Sect. 5.

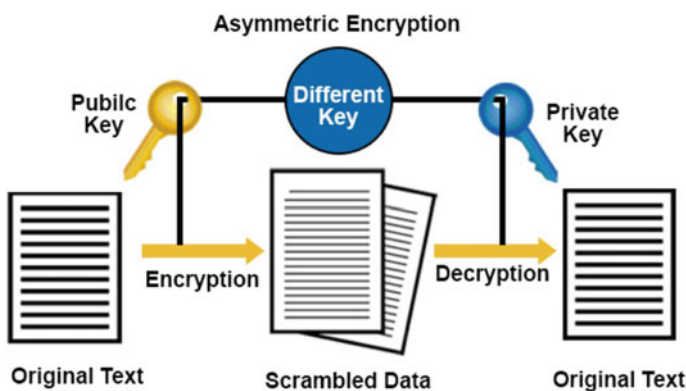


Fig. 13.2 Generalized cryptographic technique

13.2 Related Work

We have many algorithms present today for securing the data and sending it efficiently. But each of them has their own advantage and disadvantage. Let us have a look on some of them:

- Ammad Ul Islam, Faiza Khalid et al. suggested a picture steganographic strategy dependent on MSB. The procedure utilizes the distinction of two pixel's bits of the spread picture bit No. 5 and 6 of pixels are focused for installing. The contrast between bit 5 and 6 is set by the approaching mystery data bit.
- Ian McAteer, Ahmed Ibrahim et al. has recognized two essential applications consolidating biometrics and steganography, which are get to control and the transmission of touchy eHealth/biometric information. In any case, neither of these applications have made the effective change from the laboratory to this present reality setting. Proposed models for e-casting a ballot and e-shopping are remembered for this audit, yet neither of these or comparative frameworks have been executed so far.
- Ying Zou, Ge Zhang, Leian Liu show that the proposed technique dependent on multi-task has preferable execution over the immediate cross-breed preparing strategy.
- N Wu, P Shang, J Fan, Z Yang understood the steganography to secure data dependent on Markov model which was frequently utilized in regular language handling. For different reasons, past-related calculations had not completely used or dismissed the significant ideas, progress likelihood, in Markov chains.

So, the methods discussed above have one or other drawback which we can easily fix by applying a two-layer protection on the data, i.e., using both of the technologies together. Results of above method have low-quality stego-image and also easily detectable image. This present research work solves this problem and proposes a new scheme which gives good quality stego-image as well as increasing security of data while transmission.

13.3 Proposed Methodology

The presented work is categorized into two steps, firstly data is encrypted and converted into cipher then secondly conceal the information using LSB technique given by the algorithm. The message is hidden into two layers such as first layer conceals it using cryptography and second conceals the message using the new algorithm. This new method is combined with existing technology to build up the security and have smooth communication. Firstly, we encode the data then the message is hidden using algorithm. Simply LSB was not efficient as it was easy to discover so we used combined layer of steganography and cryptography which automatically increased the security and efficiency of data [9]. Even after detecting, the

steganography attacker needs to decode the new algorithm which is quite difficult to access.

13.3.1 AES Algorithm

Here, we show the design of our technique, which can use any encryption algorithm. The encryption method which we used here to scramble the information is AES algorithm. The steps are as follows:

- i. In this technique, we use bytes instead of bits. There are 128 bits of plaintext taken as 16 bytes.
- ii. These bytes are masterminded in 4 lines and 4 segments to fill in as a grid.
- iii. The no. of rounds is changeable and relies upon the measurement of our key.
- iv. There are 10 rounds for 128-piece keys, 12 rounds for 192-piece keys and 14 rounds for 256 keys.
- v. Most of them has distinctive key that is assessed by unique AES key.

Now, we have the diagram showing the functionality of this algorithm. There are two steps in this algorithm which are substitution and permutation. The 128-bit plaintext goes to pre-round transformation which goes on from 1, 2, ... N and gives us 128-bit cipher text [10] (Fig. 13.3).

Now, we begin with the embedding algorithm.

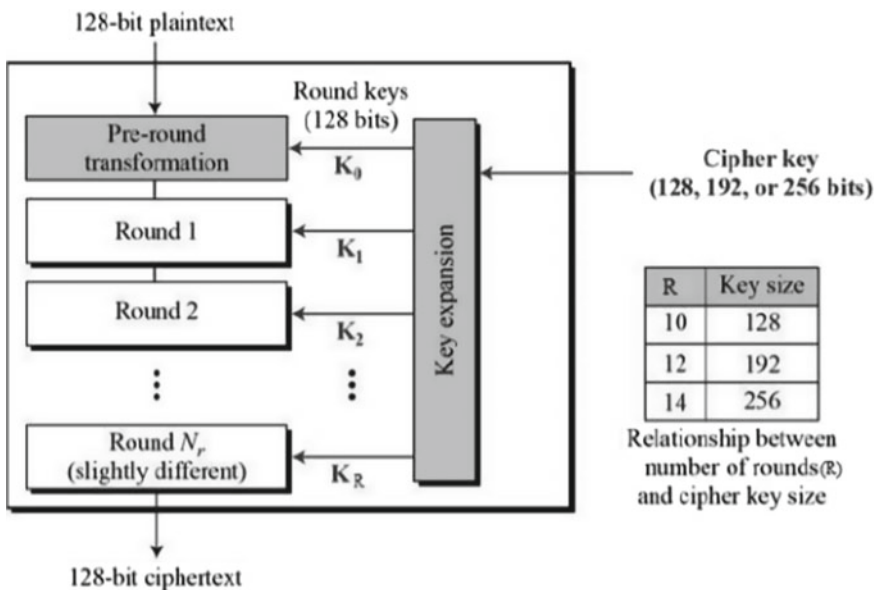
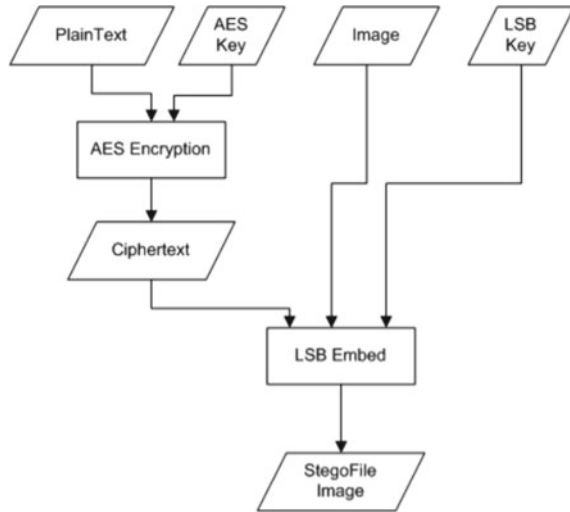


Fig. 13.3 AES algorithm

Fig. 13.4 Embedding process



13.3.2 Encryption Module

Encoding image file to give output as a text.

The algorithm is as follows:

Begin

- i. Take input as any audio, video, image, or text. Here, we are taking an image.
- ii. Select the image where you want to insert another image. And run the program using AES algorithm.
- iii. Do the encryption process using AES which will create a cipher text.
- iv. Now take the cover image and key and perform embedding using the algorithm.
- v. Lastly, we obtain result as our stego-image.

Stop (Fig. 13.4).

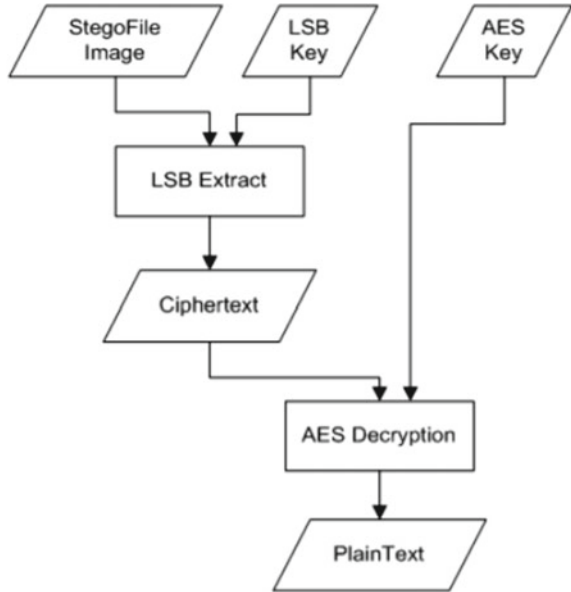
13.3.3 Extraction Module

Extracting image file to give output as plaintext.

The algorithm is as follows:

Begin

- i. Initially we put the stego bmp file and LSB key to do the decoding part.
- ii. After doing so, we obtain the cipher text.
- iii. Lastly, we use the AES key to do the extraction part. This process goes until we obtain plaintext.

Fig. 13.5 Extracting process

Stop (Fig. 13.5).

13.4 Results and Discussion

The proposed calculation was actualized utilizing MATLAB. This area presents the trial results and shows the estimations of PSNR and MSE determined for stego and carrier picture utilizing formulas are given beneath.

13.4.1 PSNR

PSNR is the peak signal to noise ratio which helps in getting to the nature of stego-picture as for the first picture. It ascertains the subtlety of the stego-picture. It helps to compare two pictures and helps to determine the closeness between them [11]. The more is the PSNR of the image the more it will be accurate. The formula for PSNR is depicted beneath (Fig. 13.6).

$$\text{PSNR} = 10 \log_{10} \left[\frac{I^2}{\text{MSE}} \right] \quad (1)$$

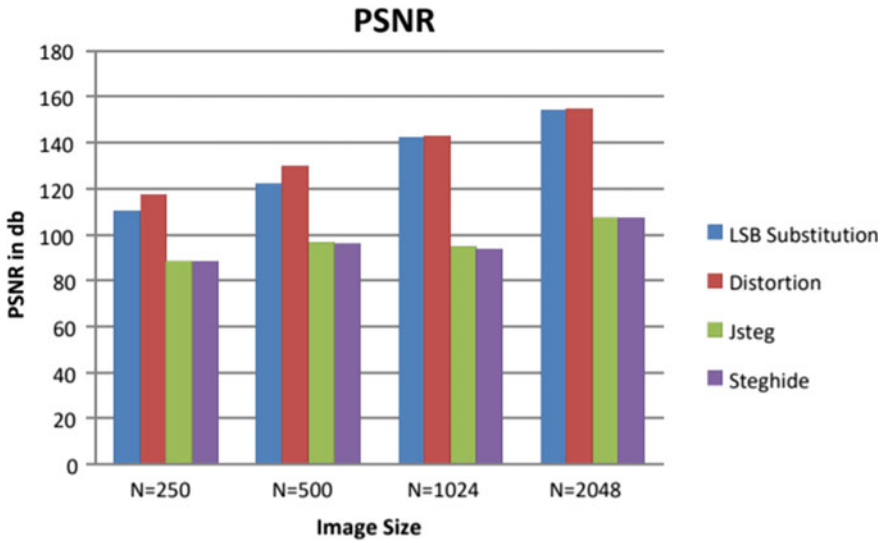


Fig. 13.6 PSNR values of few techniques

13.4.2 MSE

MSE is defined as mean square error that helps in ascertaining the error in the first picture and stego-picture. The contrast among the estimations of unique and stego-picture are multiplied and afterward their normal is determined. It is utilized essentially if there should arise an occurrence of huge mistakes since it gives generally high weight to these blunders. In this way, RMSE is extremely requesting when huge blunders are unfortunate in the image. The little is the estimation of RMSE, the more will be the nature of framework. Formula to compute MSE is as follows:

$$\text{MSE} = \frac{1}{(R * C)^2} \sum_{i=1}^N * \sum_{j=1}^M (X_{ij} - Y_{ij})^2 \quad (2)$$

where:

I = max estimation of the pixel. The maximum incentive for gray scale picture is 255.

R and C are the no. of lines and segment in the spread picture (Fig. 13.7).

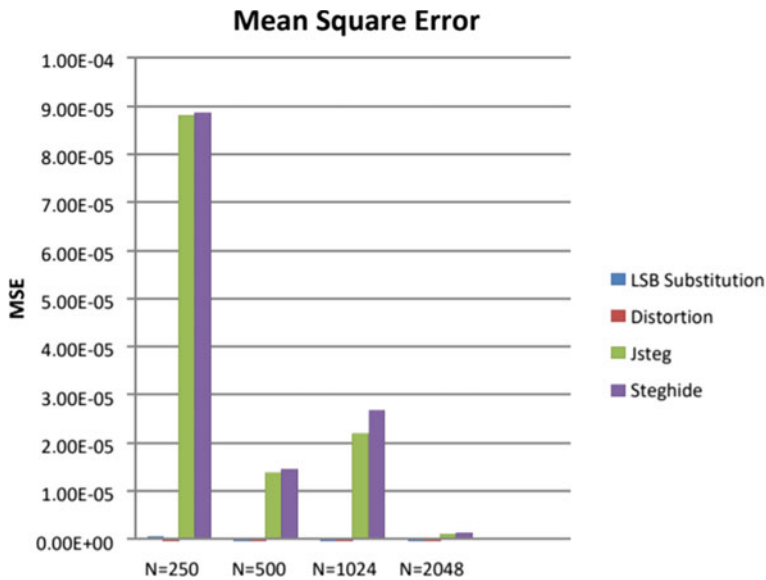


Fig. 13.7 MSE of few techniques

13.4.3 Embedding Capacity

It is the size of the mystery information that can be embedded in spread picture without crumbling the honesty of the spread picture. It very well may be spoken to in bytes. It relies on the qualities of spread picture and the inserting calculation utilized for steganography (Fig. 13.8).

Now in Fig. 13.9, we are showing the comparison graph of few techniques discussed here, in which the output of proposed method is much better than others.

13.5 Conclusion

Here, we intended combined approach for image steganography which overcome the limitation of existing methods. This approach used them as a combination of layers and is implemented in MATLAB. The results obtained provide us a better security and privacy of data and enhance communication. This also overcome the problem of steganalysis. This method improved the quality of stego-image as well as gave a good PSNR and MSE values. This method creates multiple barriers in front of the attacker so it is impossible for intruder to extract the data.

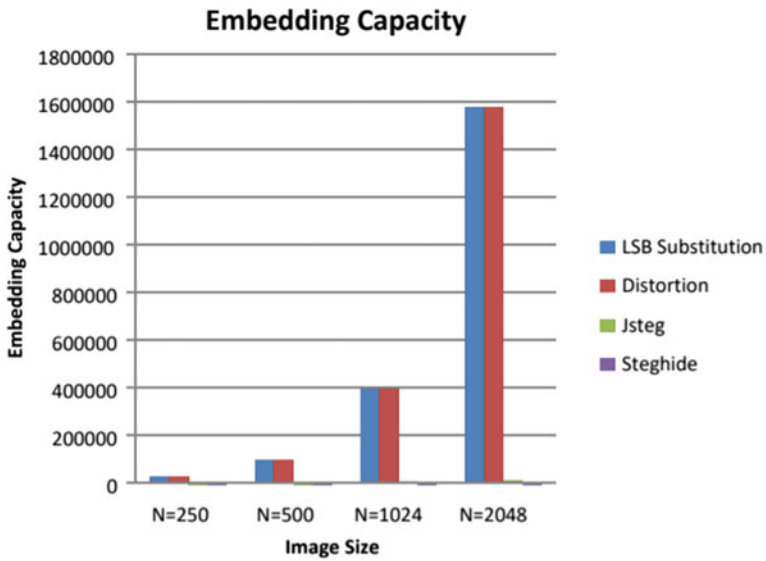


Fig. 13.8 Embedding capacity of few techniques

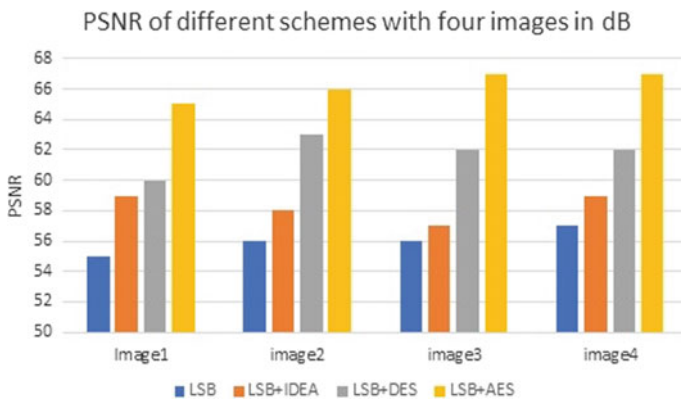


Fig. 13.9 Comparison graph of PSNR values

Acknowledgements This work is part of bilateral Indian-Bulgarian cooperation research project between Technical University of Sofia, Bulgaria and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, India under the title “Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications,” financed by the Department of Science and Technology (DST), India and the Ministry of Educations and Science, Bulgaria.

References

1. Islam, A., Khalid, F., Shah, M., Khan, Z., Mahmood, T., Khan, A., Ali, U., Naeem, M.: An improved image steganography technique based on MSB using bit differencing. In: *The Sixth International Conference on Innovative Computing Technology*, pp. 265–269. IEEE, Dublin (2016)
2. McAteer, I., Ibrahim, A., Zheng, G., Yang, W., Valli, C.: Integration of biometrics and steganography: a comprehensive review. *Technologies* **7**, 34 (2019)
3. Zou, Y., Zhang, G., Liu, L.: Research on image steganography analysis based on deep learning. *J. Vis. Commun. Image Represent.* **60**, 266–275 (2019)
4. Muhajjar, A.R., Badr, F.A.: Secure data communications using cryptography and IPv6 steganography. *Int. J. Eng. Technol.* **7**(4.19), 624–628 (2018)
5. Wu, N., Fan, P., Yang, Z., Ma, W., Liu, H.: Research on coverless text steganography based on single bit rules. *J. Phys. Conf. Series* **1237** (2019)
6. Ali, U.A., Sohrawordi, M., Uddin, M.: A robust and secured image steganography using LSB and random bit substitution. *Am. J. Eng. Res.* **8**(2), 39–44 (2019)
7. Thabit, R.: Improved steganography techniques for different types of secret data. *Int. J. Adv. Syst. Sci. Appl.* **6**, 38–51 (2019)
8. Verma, M., Saini, H.: Analysis of various techniques for audio steganography in data security. *Int. J. Sci. Res. Netw. Secur. Commun.* **7**(2), 1–5 (2019)
9. Kalamkar, P., Gaikwad, M., Gore, S., Sonule, D., Bodhe, V.: A review on implementation visual cryptography and steganography. *Int. J. Sci. Res. Sci. Technol.* **6**(2), 420–428 (2019)
10. Sharma, A., Poriye, M., Kumar, V.: A secure steganography technique using MSB. *Int. J. Emerg. Res. Manag. Technol.* **6**(6) (2017)
11. Miramirkhani, F., Narmanlioglu, O., Uysal, M., Panayirci, E.: A mobile channel model for VLC and application to adaptive system design. *IEEE Commun. Lett.* **21**(5), 1035–1038 (2017)

Chapter 14

Comparative Analysis of Various Recommendation Systems



Ekta Dalal and Parvinder Singh

Abstract In recent years, there is a rise in platforms like YouTube, Amazon, Netflix and many other Web services. Thus, there is a need of recommender systems in our lives. Recommender systems have become unavoidable in online journeys. Also, people are largely depending on online shopping platforms. When people buy from online shopping platforms they search and browse various products for their purchase. And even if they only browse, they can see the products everywhere on the Internet, i.e., on their social media platforms, blogs, etc. So, online shopping platforms like Flipkart and Amazon use recommendation systems to recommend products related to their customers. These platforms collect information about their customer's pattern of shopping and purchasing behavior to give accurate and needful recommendation to the customers. In this paper, various recommendation systems are discussed. Comparative analysis is also performed for these recommendation systems. It also gives the idea of how these recommendation systems help to improve and manage audience for online shopping platforms. It also discusses that how the similarity can be measured among the items for content-based filtering and similarity among users for collaborative filtering. It also describes that hybrid-based filtering can be used to overcome the disadvantages of content-based and collaborative filtering to give item of interest and other items recommendation to users.

E. Dalal (✉) · P. Singh
Department of Computer Science and Engineering, Deenbandhu Chhotu Ram University of
Science and Technology, Murthal, Sonipat 131039, India
e-mail: Ekta.dalal@gmail.com

P. Singh
e-mail: Parvindarsingh.cse@dcrustm.org

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 216,
https://doi.org/10.1007/978-981-33-4676-5_14

14.1 Introduction

Till recently, people for the most part would in general purchase items prescribed to them by their companions or the individuals they trust. This used to be the essential strategy for buy when there was any uncertainty about the item. Yet, with the approach of the computerized age, that circle has extended to online sites that utilize some sort of recommendation engine. There are so many examples that can be as a use case of recommendation system. Suppose we have some family function at our home, and we need to get the best function planner. We ask our relatives, neighbors and other sources which will be the best one. This is the example of a basic recommendation system which will tell you about the best function planner in layman terms. Now, according to the technology perspective, recommendation systems can be implemented in various domains according to the need, i.e., music recommendation system and Twitter sentiment analysis to cope up with fake id's and many more.

The most general and basic definition for recommendation systems is “recommendation systems help to discover the items of interest” [1]. For example, if user is going to new city for a vacation, then user needs to know about the best vegetarian food restaurants near the city. So, recommendation systems can help user to find the best vegetarian restaurants in nearby places. Recommendation systems also play an important and big fat value role in online platforms like Facebook, Netflix, Amazon, Flipkart, etc. If take an example of Amazon, which is an online shopping platform which helps to find and get the items at user's place by home delivery. It becomes very essential for Amazon to understand the need of the customers which can help to improve customer service at user level and can provide the best shopping experience for the customer. The user comes to Amazon and requests a few things. Thus, in this procedure, the person likes and waitlists a few things for future buy or might be keen on same sort of thing which is not looked by them. Thus, for this situation, the proposed framework acts as the hero and locates a similar class thing for the customers. Thus, Amazon shows the best things having same comparability to the things and causes the client to purchase an item or thing. This entire procedure improves the shopping experience of the client in the stage and makes it simple for the clients to locate the best results of their advantage [2, 3]. The most significant part of this is when recommendation systems can be actualized. It is quite straight forward on the grounds that when your business is running effectively and making average income from it, however, it is constant consistently. This is the ideal chance to execute a proposed framework to assemble a huge crowd for your business It shows only relative items to the serious buyers, reduces the effort of customer to search appropriate product of interest and improves the customer experience of your business services which leads to generate large amount of revenue from it [4, 5].

In this paper, different kinds of proposal frameworks and their extension are examined and played out a similar investigation of suggestion frameworks and finished up which proposal framework will be progressively viable and in which use case. This paper is sorted out into four areas, Sect. 14.2 contains a presentation about

recommendation systems and the advancement of different kinds of frameworks. Section 14.2 portrays distinctive proposal frameworks. In Sect. 14.3, the near investigation is done to depict utilization, favorable circumstances and inconveniences of proposal frameworks. Section 14.4 finishes up the as of now work done and the future degree to improve the suggestion frameworks for better use.

14.2 Types of Recommendation Systems

There are primarily three categories of recommendation systems are: content-based, collaborative filtering and hybrid recommendation systems.

14.2.1 Content-Based Recommendation Systems

Content-based recommendation systems are based on a description of the product/item (Fig. 14.1). It also considers user’s interest and preferences. It recommends items/products to those also who liked the same kind of items. The basic process of content-based recommendation system includes the comparison of attributes of a user which has interest and preferences with content of the item to give recommendation items of interest [6, 7].

Content-based recommendation systems can be implemented in three following steps:

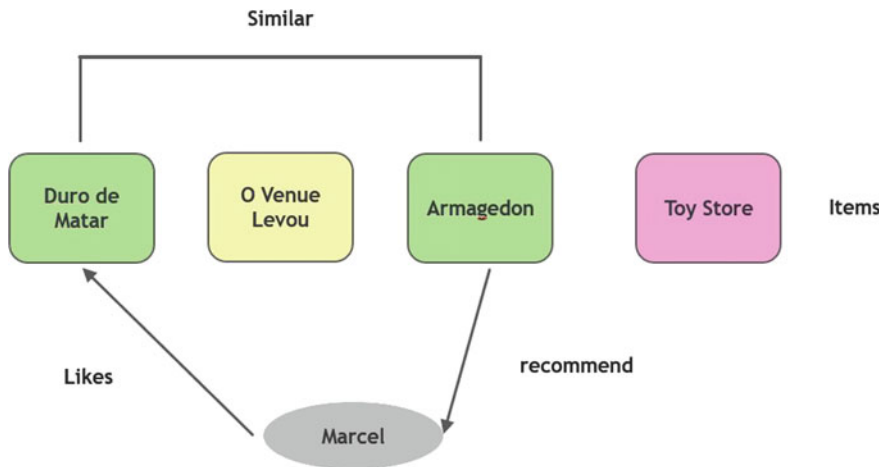


Fig. 14.1 Content-based filtering recommendation system

Content Analyzer. This component takes care of unstructured data for pre-processing to retrieve the useful information from unstructured text. Input data sources for this component can be documents, Web pages, news, product/item descriptions, etc. Input data comes from various data sources and pre-process which extract the relevant information from text and prepare this data for further processing in profile learner module and filtering component. This module mainly helps to prepare the data from unstructured to some relevant and desired format which can be used further.

Profile learner. This component takes the interest and preferences of the users to process using machine learning techniques. This helps to understand about the items/products liked by users which can be used to give the recommendation based on the feedback given by user. Profile learner processed data based on the interest of the user and given relevant recommendation for item of interest.

Filtering component. This component scans the user profile to recommend the items of interest. In this, for similarity measures between the items, cosine similarity is commonly used. Comparison can be done between prototype vector and vectors of items. After comparing with other vectors, most similar vectors are taken and recommend the most matched items to user.

Formula cosine similarity:

$$\sin(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

A can be taken as prototype vector and **B** can be assumed as item vector to calculate the similarity between the items.

This content-based algorithm will help recommend only those items/products which user liked or purchased in same category or same type. This will help to recommend the items those are liked or viewed by customer in past. So, it can be said that recommendation will be very relevant for user according to its interest [8, 9].

14.2.2 Collaborative Filtering-Based Recommendation Systems

Collaborative filtering basically involves the process of collecting feedback of user, ratings, interests and activities. It also analyses the user behavior to give recommendation based on the similarity between them. This type of method helps to develop user-specific recommendation systems based on their patterns, ratings and usage of the items from the businesses [10] (Fig. 14.2).

User *A* and user *B* are two different customers for an online shopping platform. Both users have different behavior of their shopping patterns and they purchase items/products on regular basis from shopping portal. But user *A* and *B* liked one common product/item. So, there are chances that user *A* may be interested in items

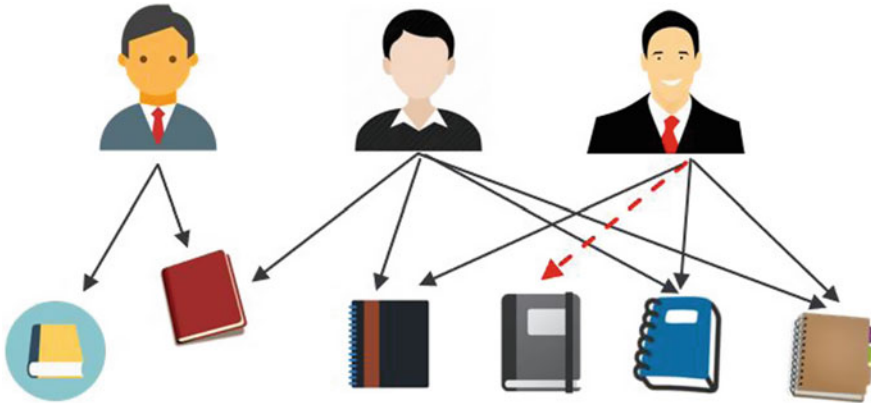


Fig. 14.2 Collaborative filtering-based recommendation system

purchased by user B and vice versa. Therefore, the collaborative recommendation system will recommend items to user A purchased by user B and vice versa.

Collaborative filtering can be categorized into further two categories are:

User-user collaborative filtering. This algorithm calculates the similarity between the users. Similarity may be calculated based on their behaviors, purchasing patterns, etc. After calculating similarity between users, this user-user collaborative filtering takes most similar users and recommend them items/products liked or purchased by their similar users.

Item-item collaborative filtering. This algorithm also works similarly user-user collaborative filtering method, but in this, similarity is used to calculate between items those are liked by other users. Similarity score is calculated between items and user liked an item then similar item can be recommended to user which is liked or purchased by other users [8].

14.2.3 Hybrid Recommendation System

This hybrid recommendation system can be implemented by combining both content-based filtering and collaborative-based filtering methods. This technique can be used to reduce disadvantages of content-based and collaborative-based filtering method [11, 12].

Hybrid recommendation system can be implemented in the following ways:

- One simple way to develop a hybrid recommendation system is to combine the results of both content-based filtering and collaborative-based filtering methods. So, it can help to get most accurate out of the box recommendations for users which can help to improve the shopping experience of users.
- Collaborative features can be used in content-based filtering method.

Table 14.1 Comparison of algorithms

Approaches	Characteristics	Drawbacks
Knowledge-based representation system	<ul style="list-style-type: none"> • Do not depend on the user ratings of items • Can adjust its recommendations easily • Cold start issue is less 	<ul style="list-style-type: none"> • Requires knowledge engineering
KBRS with sentiment analysis	<ul style="list-style-type: none"> • Identify new business opportunities • Better understanding of consumer needs • Better monitoring of brand reputation 	<ul style="list-style-type: none"> • Cannot gauge the context of user • Lack of proper data may lead to inefficient results
Matrix factorization	<ul style="list-style-type: none"> • Accurate • Better correlation • Easy to understand affordable way of solving the system 	<ul style="list-style-type: none"> • Data visualization is not good • Non-optimal results sometimes
Consensus function approach	<ul style="list-style-type: none"> • Good for group recommendation • All the considerations are taken into account 	<ul style="list-style-type: none"> • Could be time consuming • Not optimal sometimes

- Features of content-based algorithm can be combined with collaborative filtering-based algorithms.

After studying various research papers, the various recommendation algorithms are listed in this Table 14.1.

14.3 Comparative Analysis

See the Table 14.2.

14.4 Conclusion

With the growth of online shopping businesses, user experience and services problem are rising day by day. When more and more users are increasing, they are generating large volume of data and it becomes important that user can search and find useful items/products on platform. Therefore, businesses are using different approaches for recommendation systems to give best and delightful shopping experience for their customers. In this paper, distinctive suggestion frameworks are talked about with focal points and detriments. To give an exact and customized level cautions,

Table 14.2 Comparative analysis

Content-based filtering	Collaborative-based filtering	Hybrid-based filtering
<ul style="list-style-type: none"> Analyze content of item to give recommendation 	<ul style="list-style-type: none"> Analyze user behavior to give recommendation 	<ul style="list-style-type: none"> Combines both analysis of user behavior and content of the items
<ul style="list-style-type: none"> Only give recommendation of similar items those are liked or purchased by the user 	<ul style="list-style-type: none"> Give recommendation of items those are not purchased by user 	<ul style="list-style-type: none"> Can help to give both types of recommendations
<ul style="list-style-type: none"> Easy to implement 	<ul style="list-style-type: none"> Easy to implement 	<ul style="list-style-type: none"> Complex to implement
<ul style="list-style-type: none"> Less time consume to implement 	<ul style="list-style-type: none"> Takes more time then content-based but less than hybrid-based filtering 	<ul style="list-style-type: none"> Takes lots of time to implement
<ul style="list-style-type: none"> Similarity can be calculated between items 	<ul style="list-style-type: none"> Similarity can be measured between users 	–
<ul style="list-style-type: none"> Algorithms used: TF-IDF classifiers (Naive Bayes) 	<ul style="list-style-type: none"> Algorithms used: nearest neighbors matrix factorization clustering association rules 	–
<ul style="list-style-type: none"> When to use: when business has lots of item of similar kinds and has no connect among users then content-based filtering can be very effective 	<ul style="list-style-type: none"> When to use: when business is of social kind and has some relation among its customers. This collaborative filtering can be very effective to give recommendations 	<ul style="list-style-type: none"> When to use: when business has user relations and similar kind of items then this hybrid approach can be very useful to improve the customer engagement

content-based suggestion framework assumes a significant job. This substance-based proposal can be accomplished utilizing register cosine likeness, vectorization and framework factorization. In future work, to improve and actualize the novel approach/method cosmology-based learning, charts will be utilized to keep up and streamline the intensity of systems. By utilizing the knowledge diagram connection between information focuses and clients can be investigated which will sort out and retain the alarms for suggestions. Furthermore, the inference motor will be conveyed which will contain rules mapper to separate data from information and offer alarms to clients.

Acknowledgements This work is part of bilateral Indian-Bulgarian cooperation research project between Technical University of Sofia, Bulgaria and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonapat, India under the title “Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications”, financed by the Department of Science and Technology (DST), India and the Ministry of Education and Science, Bulgaria.

References

1. Yi, P., Li, C., Yang, C., Chen, M.: An optimization method for recommendation system based on user implicit behavior. In: 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), pp. 1537–1540. IEEE, Qinhuangdao, China (2015)
2. Sharma, R., Singh, R.: Evolution of recommender systems from ancient times to modern era: a survey. *Indian J. Sci. Technol.* **9**(20), 1–12 (2016)
3. Haruna, K., Akmar, M., Damiasih, D., Sutopo, J., Herawan, T.: A collaborative approach for research paper recommender system. *Int. J. Digit. Libr.* **12**(10), 1–17 (2017)
4. Beel, J., Gipp, B., Langer, S., Brietinger, C.: Research-paper recommender systems: a literature survey. *Int. J. Digit. Libr.* **17**(4), 305–338 (2015)
5. Zhang, L., Luo, T., Zhang, F., Wu, Y.: A recommendation model based on deep neural network. *IEEE Access* **6**, 9454–9463 (2018)
6. Patel, A., Dharwa, J.: An integrated hybrid recommendation model using graph database. In: 2016 International Conference on ICT in Business Industry & Government (ICTBIG), pp. 1–5. IEEE, Indore, India (2016)
7. Christakopoulou, E., Karypis, G.: Local item-item models for top-N recommendation. In: RecSys'16- Proceedings of the 10th ACM conference on Recommender Systems, pp. 67–74. ACM, Boston, U.S. (2016)
8. Macedo, A., Marinho, L., Santos, R.: Context aware recommendation in event based social networks. In: RecSys'15: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 123–130. ACM, Vienna, Austria (2015)
9. Shaikh, S., Rathi, S., Janrao, P.: Recommendation system in E-commerce websites: a graph based approach. In: 2017 IEEE 7th International Advance Computing Conference, IACC, pp. 931–934. IEEE, Hyderabad, India (2017)
10. Massa, P., Avesani, P.: Trust-aware recommender systems. In: RecSys'07: Proceedings of the 2007 ACM conference on Recommender Systems, pp. 17–24. ACM, NY, US. (2007)
11. Wang, J., Feng, Y., Naghizade, E., Rashidi, L., Lim, K., Lee, K.: Happiness is a choice: sentiment and activity-aware location recommendation. In: Proceedings of the 2018 Web Conference Companion, WWW'18, pp. 1–10. Lyon, France (2018)
12. Abdollahpouri, H.: Popularity bias in ranking and recommendation. In: AAAI/ACM Conference on AI, Ethic and Society, pp. 529–530. ACM, Honolulu, USA (2019)

Chapter 15

Finger Knuckle Print Feature Extraction Using Artificial Intelligence Algorithm



Chander Kant, Sheetal Chaudhary, Sukhdev Singh, and Parvinder Singh

Abstract Recently, the smart use of biometric traits, i.e., fingerprints, face, finger knuckle print, etc., in user authentication system seems to be integral part because of their user-friendly and robust behavior. All of these traits present different degrees of uniqueness, permanence, durability, performance, user acceptance, and robustness and are valuable according to their need in respective application. The proposed approach illustrates finger knuckle print biometric (FKP) for the authentication of user as it avoids latent FKP and criminal investigation stigma associated with printing the surface of the knuckles. The proposed approach used public database and the preprocessing carried out on the images collected, in order to separate the index finger, middle, and ring fingers of the hand. Bayesian network is used to extract the feature of FKP for the authentication and identification of user. The image processing is carried out using MATLAB R2014 software.

15.1 Introduction

Automated personal identification systems have become a very important component of information systems with the wide range of security applications. The benefits of personal identification are far too great, as they improve the reliability and security

C. Kant (✉) · S. Chaudhary
Kurukshetra University, Kurukshetra, Haryana, India
e-mail: ckverma@rediffmail.com

S. Chaudhary
e-mail: sheetalkuk@rediffmail.com

S. Singh
D.A.V. College (Lahore), Ambala City, Haryana, India
e-mail: sukhdev_kuk@rediffmail.com

P. Singh
Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Haryana, India
e-mail: parvindarsingh.cse@dcrustm.org

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 216,
https://doi.org/10.1007/978-981-33-4676-5_15

195

of electronic commerce transactions that often overcome privacy-related concerns in the use or implementation of these technologies [1]. In order to identify a person, the biometric authentication can be a promising alternative with convenient personal identification as it is free from something to carry or remember [2].

The purpose of this paper is to investigate the biometric feature of the knuckle surface for the identification of user, because of its various abilities over other traits. The benefits of this paper have obtained better result for the identification of user by applying it in security systems, in addition to having a high social acceptance, another advantage is that it does not use invasive capture systems for the user, unlike others biometric systems.

Biometric systems are planned for consistent and precise identification/authentication. In recent years, the knuckle surface image pattern (FKP) has been found to be unique and can be used in personal authentication systems [3]. Due to the intrinsic skin patterns of the outer surface about the finger phalanx joint, the knuckle surface is very distinctive due to the amount of skin folds; it has high ability to discriminate different persons; in addition, it is not easily destroyed and is therefore considered a typical biometric verification system. FKP biometric has a high acceptance by users, since it does not contain criminal investigation as in the fingerprint. Therefore, the function of the fingerprint knuckles has a high probability of being identifier in general terms as a biometric authentication system [4].

This paper reflects the use of Bayesian networks, that mock-up a phenomenon using a set of variables and the dependency associations between them. These models can have various functions, for organization, forecast, diagnosis, etc. [5]. Furthermore, they can give interesting information regarding how domain variables are related, that can occasionally be interpreted as cause–effect associations.

The extraction of biometric characteristics is carried out in the digital image processing of the upper part of the hand, using Bayesian networks, which will identify and extract from the region of interest (knuckle surface) the biometric characteristics. It later on will be compared in a reference database in order to be able to make a decision and know if the type to be identified is known or unknown and then carry out a control action.

15.2 Literature Review

The literature has been reviewed to study the existing feature extraction methods for FKP proposed by various authors. Zhang et al. [6] discussed a novel feature set extraction approach to combine fisher-criterion with manifold-criterion which is known as (WLE) weighted linear embedding. Gaussian weights were employed to merge various types of information. The proposed approach intends to locate a map vector, that the ratio among weight class scatter to the weight within class scatter is maximized. It produces a recognition rate of 78.2% when applied on index finger of different persons. Yang et al. [7] used Gabor wavelets for image analysis and pattern

recognition of FKP feature set extraction. The orthogonal linear discriminant analysis (LDA) transformation methodology in PCA is used and classified by nearest neighbor classifier thereby increasing the efficiency up to 98.14%. Jing et al. [8] used distances and angles concurrently between image data vectors to compute data similarities. To eliminate unnecessary information, orthogonal complex preserving locality projections method has been used. They achieved recognition rate of 88% for the left index finger. Zhun Lei Quin [9] devise a robust FKP feature extraction cum matching methods based on speed up features that is considered a good improvement over scale invariant transform. It defines a system that is based on local convex direction map of FKP for alignment of images and a ROI is cropped for feature set extraction. For matching process, the relative distance of the closest neighbor with second closest neighbor is compared and distance ratio with less than 0.6 unit. It produces accuracy of 90.63% for verification and 96.91% for identification. Shen et al. [10] proposed an approach based upon results of neurophysiology studies which show that both local and global features are very important for image perception. It is suitable for images with plentiful line like structures and has the features like high accuracy, robustness to illumination variations, and fast matching. Rui Zhao et al. [11] proposed a novel approach that reduces the load of large database to train the classifier model where the edges of images are characterized at gray levels. The experimental work proves that FKP is consistent and appropriate trait for recognition rate of 95.68% with 30 threshold value. Shariatmadar and Faez [12] used a collection of Gabor filters to extract the orientation information from FKP images. Five different scales and eight different orientations were selected to keep the remaining parameters constant and PCA is applied to reduce their dimensions. Combination of PCA and LDA provides efficient feature selection method. The proposed algorithm was tested for 4-fingers and concludes that right middle finger provides a better performance with 75.2% accuracy. Kumar and Ravikanth [13] proposed a personal authentication scheme by finger knuckle surface. The feature extraction for finger knuckle surface was performed with both texture and geometrical feature analysis methods. The texture information of knuckle surface was obtained by PCA, ICA, and LDA approaches. Scores were calculated by calculating Euclidean distance (ED) obtained from reference and input vectors.

15.3 Proposed Approach

The objective of this research is to identify and authenticate people through the use of FKP biometrics. An artificial intelligence methodology will be used through Bayesian networks by taking as a reference in database. The reference database used here is from Center for Biometric Research of the Polytechnic University from Hong Kong [14] made up of 165 samples, out of which 125 males and 40 females between the age of 18 and 50.

In the first practical stage of this paper, a model was designed and implemented to capture images of the knuckle surface, which consists of an acrylic box with a

Fig. 15.1 Model for image capture



dimension of 30 cm \times 25 cm \times 20 cm, as shown in Fig. 15.1. The device includes a Sony Cyber-Shot model digital camera with a resolution of 14.1 megapixels; this external camera allows the capture of photographs from the user's hand when the user introduces them to the model.

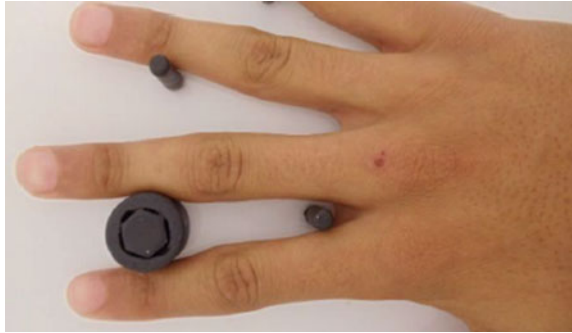
The acrylic used in the model is white to prevent external lighting from affecting the capture of the images. At the bottom of the model, a plate was installed with posts that have the same location as a hand punch, so that the samples captured by the model always have the same shape and position, as shown in Fig. 15.2.

Once the images of the model have been captured, as shown in Fig. 15.3, the following is to carry out a preprocessing, that is, of the hand that was photographed, the index, middle, and ring fingers are separated and obtained the region of interest (ROI) from each of them on which the biometric characteristics will be extracted. It is important to note that in this paper, the image processing is carried out using MATLAB R2014 software.

Fig. 15.2 Posts for hand position



Fig. 15.3 Sample image of the database



15.3.1 Digital Preprocessing

From the image of the database like the one shown in Fig. 15.3, which has a dimension of 1920×1080 pixels, the next step is to cut the index, middle, and ring fingers from the image (as shown in Fig. 15.3) and then each finger separately extracts biometric characteristics from the region of interest.

15.3.1.1 Detection of ROI

To crop the image along, it is necessary to binarize the original image, that is, convert it to black and white, as shown in Fig. 15.4. The next step to crop the image from left to right is to make a summation of each one of the pixels of the binarized image (to carry out the summation it is necessary to transpose the image), in order to derive the ROI. Figure 15.5 shows the graph of the summation of the binarized image.

In the summation graph (Fig. 15.5), the contour of the binarized image is observed (using MATLAB), rotated 90° to detect the cut point up to the tip of the middle finger, just enough to obtain the minimum point of the summation, knowing this point we

Fig. 15.4 Binarized image



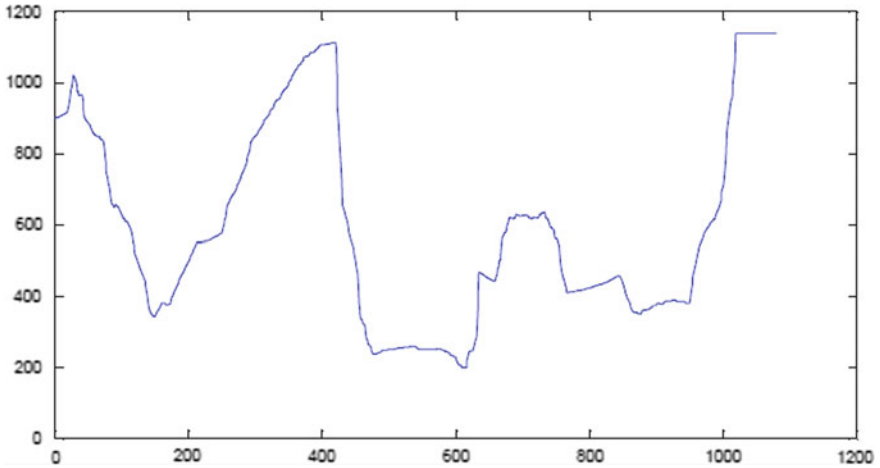


Fig. 15.5 Summation of the binarized image

proceed to cut the image up to the fingertips, but to have only the cut out of the fingers, a new sweep of the image is carried out starting from the center to the end of the same, to detect the pole of the base of the model located between the middle finger and the index, once you have the location of the post on the X-axis, the original and binarized image is cut, as shown in Fig. 15.6.

The width of the resulting image remains the same as that of the original 1080 pixel image; now with the resulting image, the next thing to separate is the index, middle, and ring fingers; this is done by dividing the image into three equal parts, with a 360 pixels wide, thus having an image for each finger. Figure 15.7 shows the fingers separately when cropping the previously cropped image.

Fig. 15.6 Original image cropped with the new coordinates

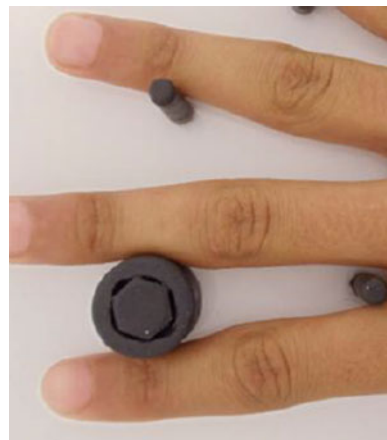
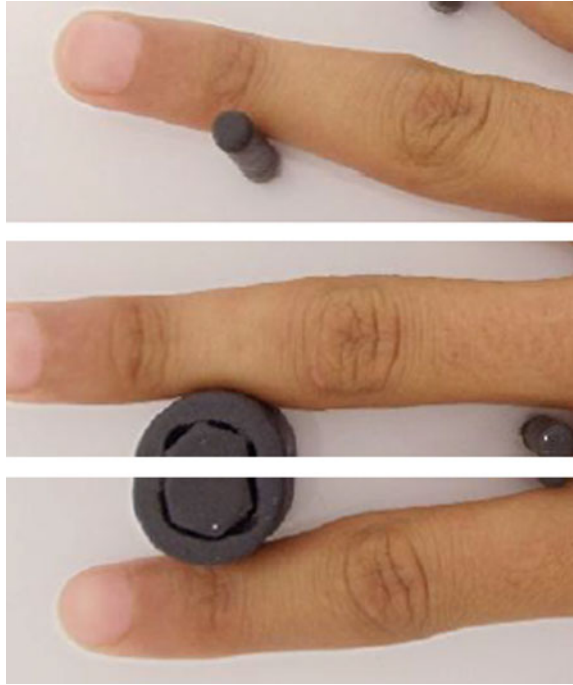


Fig. 15.7 Index, middle, and ring fingers trimmed separately



Due to the posts located at the base of the model, it is possible to know the angle of inclination of the index finger and the ring finger, for which the index finger is rotated with an angle of -10° and the ring finger with an angle of 15° . Figure 15.8 shows the result of rotation to each of the images separately; note that the middle finger is not rotated because its orientation is in an optimal location for feature extraction.

Once the region of interest is completely isolated, techniques will be applied to extract biometric characteristics that can identify distinctive points of each image. By obtaining this information, the artificial intelligence algorithm will carry out a comparison between images, which is located in the database and that will be the new image entered by the user to identify; if the images have a high percentage of coincidence, the algorithm will determine that it is a genuine user, otherwise it will be marked as an impostor.

15.3.2 Proposed Architecture

Figure 15.9 shows the architecture or block diagram of proposed approach. The steps involved are discussed below:

Fig. 15.8 Rotation of the index and annular method

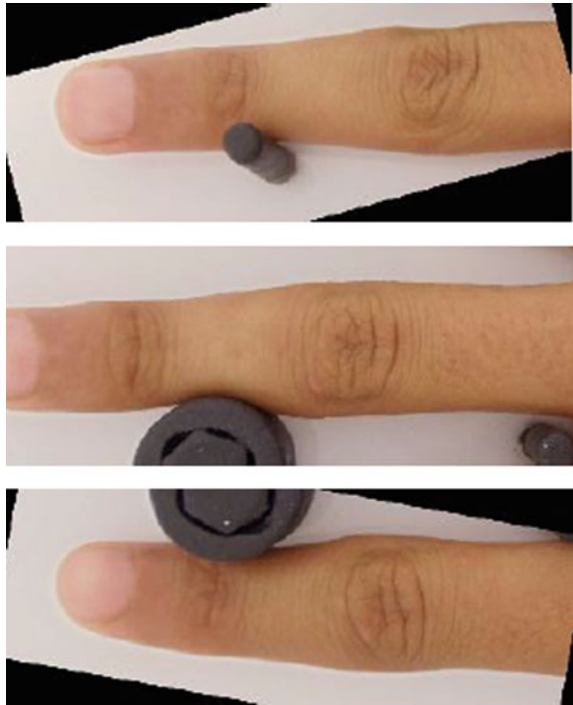
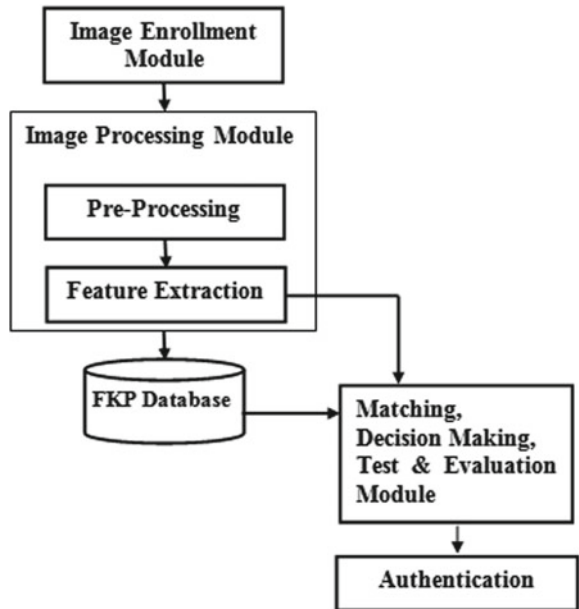


Fig. 15.9 Block diagram for proposed approach



Step 1: In image enrollment module, a digital camera with 14.1 megapixel was used for image acquisition of the index, middle, and ring finger knuckle with dimension of 1920×1080 pixel.

Step 2: In image preprocessing module, an image with dimension of 1920×1080 pixels was used to cut the index, middle, and ring finger from the image, and then each finger separately to extract biometric feature from the region of interest (ROI).

Step 3: This step performs feature extraction from the preprocessed image of index, middle, and ring finger knuckle with the help of the Bayesian network. Bayesian network is a directed acyclic graph (ADG) that has structure of network consisting of nodes and dependency probabilities.

Step 4: This step stores the extracted feature set in FKP database for authentication purposes, i.e., identification or verification.

15.4 Experimental Results

As seen along Figs. 15.3, 15.4, 15.5, 15.6, 15.7 and 15.8, initially having an image without processing, it was possible to separate the fingers of the hand. So far there is a (test) method to cut the knuckle of the index and ring fingers, this is achieved through the location of the resulting contours by rotating the image, that is, sweeps are made in the image in different lines and columns (of the contours) to detect color changes in the image; in Fig. 15.10, the cut points are marked to generate the new image, as seen in the column on the left side the image of the index and ring finger. The crosses made to separate the knuckle from these fingers are marked in red, and the result of the new cut is shown in the right-hand column.

It should be noted that this method is only a test (for the moment) for the extraction of the region of interest from the index and ring fingers, and currently, we are working on optimizing this algorithm so that it can detect the knuckle of the three fingers separately and can generate a unique and exclusive image of the region to which biometric features are desired.

15.5 Conclusions and Future Scope

It has been observed from the above proposed work that the feature set for FKP is obtained effectively using different effective steps and the efficiency of the approach is proven to be better than existing approach. After applying digital image processing to the database, it has also been possible to extract the region of interest and the next stage is to apply a technique for extracting image features. Once having this information, the third stage consists of developing the artificial intelligence algorithm using Bayesian networks, which makes the decision (based on how much similarity there is between one image and another) to accept or reject the user. In future,

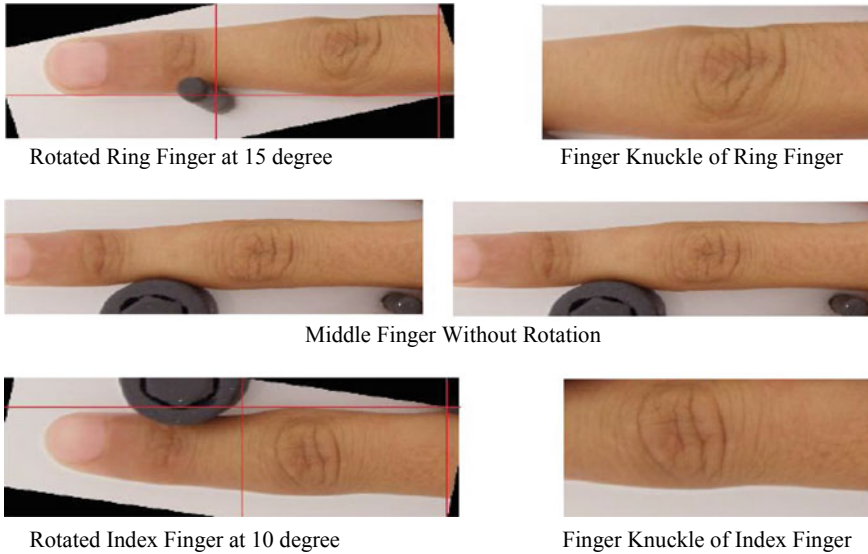


Fig. 15.10 ROI of the knuckle of the index and ring fingers based on the black outlines caused by the rotation of the image

the approach may be used with fused templates with multimodal biometrics rather than individual biometric that of course further increase the efficiency and accuracy parameters of proposed approach.

Acknowledgements This work is part of bilateral Indian–Bulgarian cooperation research project between Technical University of Sofia (Bulgaria), Kurukshetra University Kurukshetra (Haryana) and DCR University of Science and Technology, Murthal, Sonapat, India under the title “Contemporary Approaches for Processing and Analysis of Multidimensional Signals in Telecommunications,” financed by the Department of Science and Technology (DST), India and the Ministry of Education and Science, Bulgaria.

References

1. Kumar, M.N., Premalatha, K.: Finger knuckle-print identification based on local and global feature extraction using sdst. *Am. J. Appl. Sci.* **11**(6), 929–939 (2014)
2. de Luna-Ortega, C.A., Ramirez-Marquez, J.A., Mora-Gonzalez, M., Martínez-Romo, J.C., Lopez-Luevano, C.A.: Fingerprint verification using the center of mass and learning vector quantization. In: 2013 12th Mexican International Conference on Artificial Intelligence, pp. 123–127. IEEE (2013)
3. Liu, M., Tian, Y., Lihua, L.: A new approach for inner-knuckle-print recognition. *J. Vis. Lang. Comput.* **25**(1), 33–42 (2014)
4. Aoyama, S., Ito, K., Aoki, T.: A finger-knuckle-print recognition algorithm using phase-based local block matching. *Inf. Sci.* **268**, 53–64

5. Kulkarni, S.S., Rout, R.D.: Secure biometrics: finger knuckle print. *Int. J. Adv. Res. Comput. Commun. Eng.* **1**(10) (2012)
6. Zhang, L., Zhang, L., Zhang, D.: Monogeniccode: a novel fast feature coding algorithm with applications to finger-knuckle-print recognition. In: 2010 International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics, pp. 1–4. IEEE (2010)
7. Yang, W., Sun, C., Wang, Z.: Finger-knuckle-print recognition using Gabor feature and MMDA. *Front. Electr. Electron. Eng. China* **6**(2), 374 (2011)
8. Jing, X., Li, W., Lan, C., Yao, Y., Cheng, X., Han, L.: Orthogonal complex locality preserving projections based on image space metric for finger-knuckle-print recognition. In: 2011 International Conference on Hand-Based Biometrics, pp. 1–6. IEEE (2011)
9. Zhun Lei Qin: Robust palm print and knuckle print recognition system using a contactless approach. In: 2010 5th IEEE Conference on Industrial Electronics and Applications, pp. 323–329. IEEE (2010)
10. Shen, L., Bai, L., Ji, Z.: Hand-based biometrics fusing palmprint and finger-knuckle-print. In: 2010 International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics, pp. 1–4. IEEE (2010)
11. Zhao, R., Ouyang, W., Li, H., Wang, X.: IEEE international conference on computer vision and pattern recognition (CVPR) (2015)
12. Shariatmadar, Z.S., Faez, K.: A novel approach for Finger-Knuckle-Print recognition based on Gabor feature fusion. In: 2011 4th International Congress on Image and Signal Processing, vol. 3, pp. 1480–1484. IEEE (2011)
13. Kumar, A., Ravikanth, C.: Personal authentication using finger knuckle surface. *IEEE Trans. Inf. Forens. Secur.* **4**(1), 98–110 (2009)
14. PolyU, FKP Database. The Hong Kong Polytechnic University, Hong Kong (2014)

Chapter 16

The Using of Deep Neural Networks and Acoustic Waves Modulated by Triangular Waveform for Extinguishing Fires



Stefan Ivanov, Stanko Stankov, Jacek Wilk-Jakubowski,
and Paweł Stawczyk

Abstract Current article introduces a new approach for detection of fires based on deep neural network (DNN) and their extinguishing using an acoustic fire extinguisher. Finding fires on video stream is based on low-cost hardware platform containing Movidius stick for hardware acceleration of the DNN used for fire detection. For this purpose, the fire extinguisher uses a sinusoidal acoustic wave modulated by a triangular waveform. The special design of the extinguisher guarantees that the sound pressure level will be sufficient for successful extinguishing of the fire in distance up to 130 cm.

16.1 Introduction

Nowadays, the traditional fire extinguishing methods are based on the effect of gaseous, liquid, or solid extinguishing agents on the selected flames area. The acoustic method turns out to be a good solution, among others, for fighting fires of spilled liquids. Its action is based on the dispersion of fire over a large area with the use of acoustic waves. This results in a gradual reduction of the flames, which are located on a larger area. However, in order to use this method effectively, it is

S. Ivanov (✉) · S. Stankov
Department of Automation, Information and Control Systems, Technical University of Gabrovo,
4, H. Dimitar, 5300 Gabrovo, Bulgaria
e-mail: st_ivanov@abv.bg

J. Wilk-Jakubowski
Department of Information Systems, Kielce University of Technology, 7 Tysiąclecia Państwa
Polskiego. Ave., 25-314 Kielce, Poland
e-mail: j.wilk@tu.kielce.pl

P. Stawczyk
Department of Industrial Electrical Engineering, Kielce University of Technology, 7 Tysiąclecia
Państwa Polskiego. Ave., 25-314 Kielce, Poland

necessary to emit acoustic waves at appropriate frequencies. The first documented attempts to extinguish flames with the use of acoustic waves were made in Poland (in 90 s) in the last century [1, 2], as well as in the last years in the USA (especially since 10 years ago) [3–9]. Further research has been carried out in Poland since 2017, which resulted in, among others, applications to the Patent Office [10–14]. Similar researches have become the subject of scientific efforts in agency Defense Advanced Research Projects Agency (DARPA), which in 2008 launched the Instant Fire Suppression Program (IFS program) aimed at searching for effective firefighting methods [3]. In the video presented by DARPA, it can be seen that the flames placed in the acoustic field (generated by two loudspeakers positioned on both sides of the fuel tank) were quickly extinguished. We are therefore dealing with two dynamics. First, the acoustic field increases the air velocity. As air velocity increases, the flame boundary layer of the combustion flame appears, facilitating the process of flame surface disturbance. Secondly, disturbing the flame surface leads to higher evaporation of the fuel, which expands the flame, but also reduces the total flame temperature. The combustion process will be disrupted by the distribution of heat over a larger area. Consequently, it is possible not only to significantly reduce the flames, but also to extinguish them completely. Similarly, a fire extinguisher designed by American students Seth Robertson and Viet Tran is suitable for extinguishing flames [4]. The device developed after a year of experimenting is equipped with an amplifier, a power supply, and a tube from which acoustic waves are emitted. The process of extinguishing with acoustic waves assumed that oxygen is the main fuel of fire. In turn the acoustic waves of appropriate frequency are able to push oxygen further out of the air. Still another solution presented in the “Myth Busters” program which proves that it is possible to extinguish a fire with the amplified, modulated human voice (with the use of computer techniques) [5]. Because the acoustic amplification limit reached by the firefighters exceeds the pain threshold of the human ear (due to excessive sound pressure), this solution is dangerous for human health.

Acoustic flames extinguishers using high and very high-power loudspeakers can be effectively supported by an innovative fire detection method using deep neural networks, based on learning from the images or video streams [16–23]. Dedicated sets can be used for image recognition [24, 25]. In order to replace traditional fire protection methods by the new technologies, numerous studies are needed, including the modulated acoustic waves produced by acoustic extinguishers. The benefit of the use of the acoustic waves is non-invasive character of operation (significant reduction of the equipment damage placed in the fire zone, thus reducing the costs and time needed for repairs), no need to give pressure tests of the tank with extinguishing agent and unlimited time of use in relation to traditional extinguishing methods. This time is limited with only by power supply delivered to extinguisher, electronics, and many others. For this reason, it is also important to understand the operation of conventional fire protection methods [26–28], the new developments in signal processing [29, 30], as well as the use of the Internet possibilities and electronics in practice. However, more effective extinguishing requires further research in the direction of the impact of low frequency and high-power acoustic waves on human health. Knowledge of

the practical extinguishing properties of simultaneously operation of several acoustic extinguishers (or acoustic extinguishers composed of a set of sound sources in the form of subwoofers) is also needed.

16.2 Structure of the Acoustic Fire Extinguisher

In the scientific literature, there are known researches concerning the possibility of using acoustic waves of very low power to extinguish flames in a short distance from the output of the device. This distance was only a few centimeters from the extinguisher output. During these tests, electrical power supplied to the loudspeaker was very low (up to 20 W). The authors noticed that in order to find out how to use the technology of extinguishing flames by means of acoustic waves, it is necessary to conduct research on acoustic extinguisher with much higher power. The use of high and very high acoustic power makes it possible to increase the distance between the flame sources and the extinguisher output. The paper is a response to this call, allowing to determine the effect of acoustic waves on the extinguishing process at a much longer distance from the fire extinguisher output. It was unexpectedly noticed that the modulation of the extinguishing wave frequency has a positive effect on extinguishing process (scientific novelty). In addition, it has been noted that better concentration of the acoustic stream has a positive effect on the range and effectiveness of extinguishing effects, especially in the case of small areas. The research clearly confirmed the extinguishing effect of the acoustic waves.

The elements of the measuring station are: Rigol DG4102 generator with AM modulator; Proel HPX2800 power amplifier; SVAN 979 sound level meter with instrumentation and analog meters for measuring electrical quantities. The construction of the acoustic extinguisher is shown in Fig. 16.1.

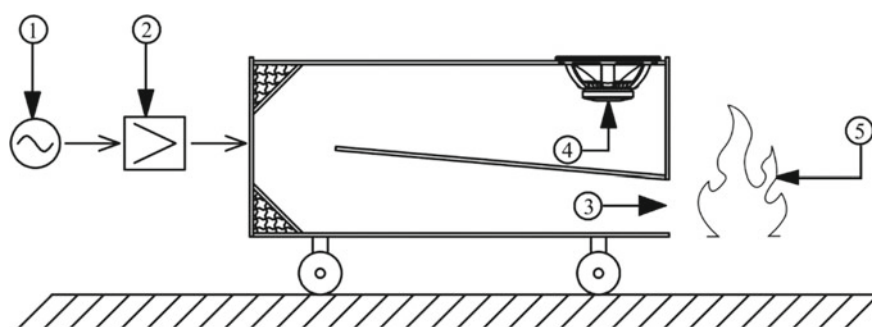


Fig. 16.1 Block diagram of the measurement station for testing extinguishing flames with the use of acoustic waves: (1) signal generator with AM modulator, (2) high and very high power amplifier, (3) waveguide (extinguisher) output, (4) loudspeaker, and (5) source of flames

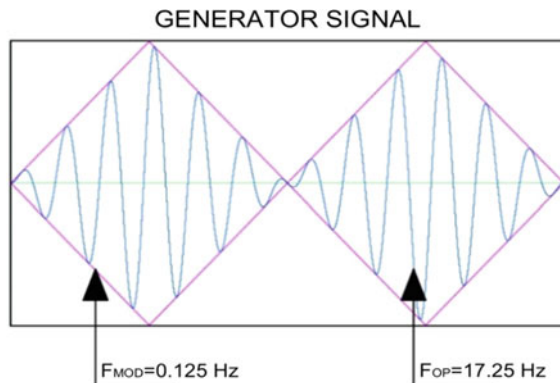
B&C 21DS115 loudspeaker with a nominal power of 1700 W was installed at the beginning of the waveguide. The required waveguide length is two times smaller in the closed end tube in relation to the open tunnel [15]. For this reason, acoustic extinguisher was designed as the waveguide with tunnel bend with a rectangular cross-section of 428 cm in length. 1 V RMS was applied to the power amplifier input. The background noise level during measurements was equal to 64.7 dB. A professional trainer (fire mock-up) was used as a source of flames. This type of equipment is used by professional services investigating the possibilities of traditional fire extinguishers. The trainer was fuelled by propane–butane gas. Thus, during the experiments, the flames obtained as a result of igniting the gas were extinguished.

Apart from the application of high and very high acoustic power, a scientific novelty of the research is the presentation of the measurements results showing the possibilities of extinguishing flames using modulated waveforms. In this paper, AM modulation was used. The results show the sound pressure level (SPL) at which the flames were extinguished successfully as a function of the distance from the extinguisher output in the range from 50 to 130 cm, with a step of 10 cm, for a sinusoidal wave with a operational frequency $F_{OP} = 17.25$ Hz modulated by triangular waveform (AM modulation with frequency $F_{MOD} = F_{MFreq} = 0.125$ Hz). During the measurement, the peak value of the power supplied to the extinguisher was 1000 W. The waveform given from the generator to the power amplifier is presented in Fig. 16.2.

Measurements of sound pressure at which the flames were completely extinguished were carried out for both longitudinal and transverse positioning of the trainer. Measurement diagram for the longitudinal position of the trainer is presented in Fig. 16.3.

The following results of the experimental studies show the minimum sound pressure level (SPL) causing the successful extinguishing effect in the function of the distance (L) from the extinguisher output. There is an inversely proportional relationship between the decrease in sound pressure level and the increase in the distance from the waveguide (extinguisher) output. As the distance from the fire extinguisher

Fig. 16.2 Sine wave AM-modulated by triangular waveform



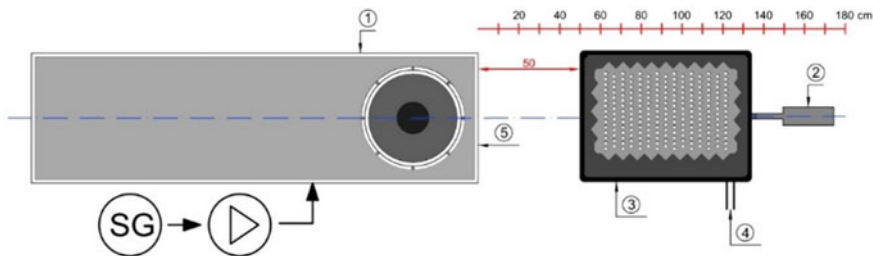


Fig. 16.3 Measurement diagram for the longitudinal position of the trainer: (1) acoustic extinguisher, (2) sound pressure level meter, (3) trainer, (4) propane inlet, and (5) extinguisher output

output increases, sound pressure level which is necessary to extinguish the flames decreases. The results showing the dependence of the extinguishing acoustic pressure as a function of the distance from the extinguisher output are shown in Fig. 16.4.

The measurement diagram for the transverse position of the trainer is presented in Fig. 16.5.

The minimum value of the sound pressure level (SPL) causing the successful extinguishing affect as a function of the distance from the extinguisher output is shown in Fig. 16.6.

As you can see, the minimum SPL causing the successful extinguishing effect as a function of the distance from the extinguisher output may be determined by the regression function (Figs. 16.4 and 16.6). On this basis it is possible to obtain the trend functions of the SPL from the extinguisher output within the analyzed distance range (the curves called “trendline”) and their character. Thus, it is possible to determine the approximate values of the sound pressure level causing extinguishing effects for other distances. Trends can be distinguished using the method of mechanical time series

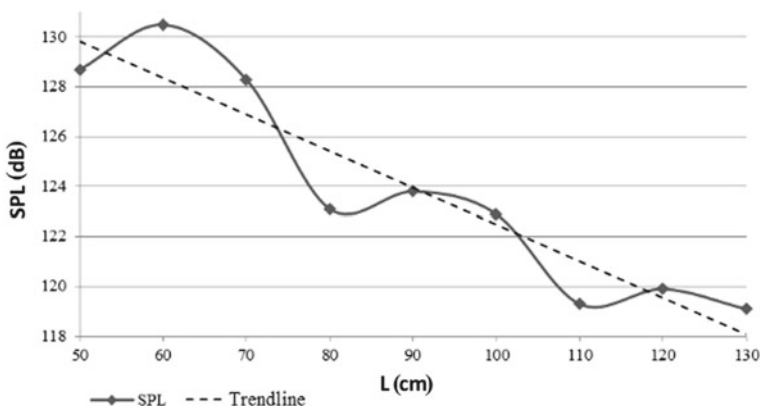


Fig. 16.4 Dependence of the sound pressure level as a function of the distance from the extinguisher output for the longitudinal position of the trainer

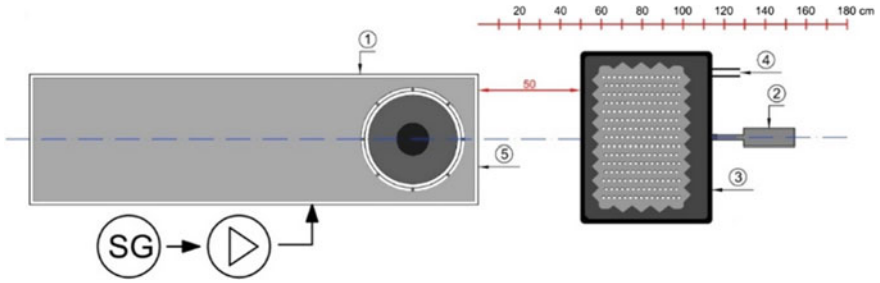


Fig. 16.5 Measurement diagram for the transverse position of the trainer: (1) acoustic extinguisher, (2) sound pressure level meter, (3) trainer, (4) propane Inlet, and (5) extinguisher output

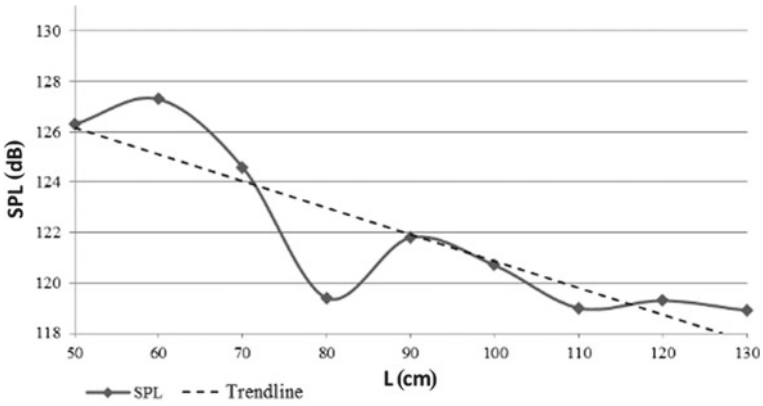


Fig. 16.6 Dependence of the sound pressure as a function of the distance from the extinguisher output for the transverse position of the trainer

alignment or the analytical method based on empirical data. For this purpose, the authors used the analytical method of extracting the developmental trend by adjusting the specific mathematical function to the time series. The linear trend function was used as a function of the development trend and if it is necessary the parameters can be estimated using the method of least squares.

16.3 Using Deep Neural Networks for Fire Detection

16.3.1 Hardware System for Fire Detection and Control of Fire Extinguisher

The proposed system consists of low-cost computer module, USB camera, and Movidius stick for implementation of deep neural network for fire detection (Fig. 16.7). The system is responsible for flames source detection and can be connected to the acoustic fire extinguisher and its activating.

The hardware is based on a Raspberry Pi board which has a quad Core 1.2 GHz Broadcom BCM2837 64 bit CPU, 1 GB RAM, plenty of communication capabilities, camera slot, and display port. The board can support various operating systems (OS) like: Ubuntu, Raspbian, Windows 10 IoT. The video controller can support modern resolution standards such as HD and Full HD. It can also generate 576i and 480i composite video signals for PAL-BGHID, PAL-M, PAL-N, NTSC, and NTSC-J.

An USB camera Logitech C310 is connected to the Raspberry Pi providing 1280 × 720 pixels resolution. It guarantees a 30fps video signal and has a fixed focus and 60° field of view. This USB camera is selected because of its price and the good quality of the output video signal.

The Movidius stick is vision processing unit (VPU) that uses specialized processor with high computing capacity to perform complex operations on static and dynamic data using artificial neural networks. The direct communication between the Raspberry Pi and the Movidius VPU module allows to be achieved a significant acceleration of processing performance. This acceleration is due to the lack of complex computational operations inherent from deep neural network which has to be processed in the Raspberry Pi. The Movidius USB stick increases the performance by integrating the neural network that performs the math computations using the so-called vision processing unit—Myriad 2. This type of processor is specifically designed for tasks related to machine vision and it is also very energy efficient. The proposed system incorporates LCD display for visualization of video stream from



Fig. 16.7 Hardware system for fire detection **a** structure and **b** final system

camera and for drawing of contours of the fire sources and flames that are found in the video signal. To control the extinguisher, two relay modules are used. They send 24 V control signals to the acoustic extinguisher depending on presence of fire in the video stream.

Several software technologies are used in the process of fire detection as: OpenCV, NumPy, Matplotlib, Imutils, and TensorFlow.

16.3.2 Training of the DNN

The neural network used in developed system is based on MobileNet architecture, which has a high speed of object detection. The MobileNet is an architecture that is suitable for mobile and embedded applications. It was developed by Google.

The training requires a preliminary database of pictures from which the DNN will retrieve the characteristics of the fire objects which should be found afterwards. The neural network is initially capable of recognizing multiple objects and by replacing some layers it can be trained to recognize specifically which part in the image is fire. The mentioned neural network is trained using the TensorFlow Machine Learning Library using more than 250 images of fire, which are freely available in the Internet. Initially, the fires (and their position in the image) have to be labeled manually before the training process. A Python scripts generates 1000 shifted and scaled images from the mentioned above also with supporting information for the coordinates of fires in the images. These images are saved in “png” format and used for training of the network.

Figure 16.8 presents some of the images used in the training process.



Fig. 16.8 Some of the images used for training of the deep neural network

The neural network expects in its input layer images with format $300 \times 300 \times 3$ (300×300 pixels \times 3 colors). Therefore, all training images are scaled to have such resolution. The training process continues 200,000 epochs with batch size of 24 images. The training is done with GPU device GTX 1080 Ti and it takes 8 h to complete.

16.4 Experimental Results

After the successful training, the well-trained neural network is implemented in the Movidius USB stick. The arrays of fire of burning objects are further processed in the Raspberry Pi and their positions are marked on the images. The trained network is able to detect fire under different conditions of the background. The neural network was tested with 100 test images which were not included in training process. The precision is 93%. Figure 16.9 presents some of the test images and the fire detected in it.

To be sure that the neural network is properly trained, it is necessary to test it with images captured by the Web camera connected to Raspberry Pi. The color video stream from the camera with resolution 1280×720 pixels is scaled to have resolution 300×300 pixels and after that the video frames are supplied to the input of the neural network. The fire sources in the video stream are recognized successfully by the trained deep neural network. Figure 16.10 presents a fire detection process using the real time acquisition of images of fire.



Fig. 16.9 Some of the images used for testing

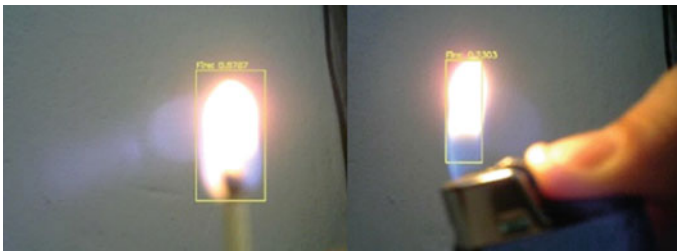


Fig. 16.10 Images from camera for real-time testing

The testing for “false positive” reaction of the system was done in different rooms under different lightning conditions—direct sunlight in the room, artificial light, and low level of light intensity. In all tests, the system does not demonstrate “false positive” detection of fire.

All the experimental results received during the test of the deep neural network prove its suitability for fire detection. The system (based on Movidius and Raspberry Pi) can be used successfully as smart system for fire detection and control of acoustic fire extinguisher. When the fire is detected, the system can generate signal for activation of acoustic extinguisher. In this case, Raspberry Pi board switches on two relay modules which activate the power supplies of signal generator and power amplifier used in the construction of the acoustic fire extinguisher.

16.5 Summary

In the long-term perspective, the technology of extinguishing flames using acoustic waves’ extinguishers controlled by neural networks, proposed by the authors, can be permanently implemented in various devices including buildings or transport (both land, water, and air transport means). Contemporary research on flame detection using neural networks is crucial in this respect. Moreover, the technology of extinguishing flames with acoustic waves may support the fire protection of halls, warehouses, or liquid tanks. The big advantages of this solution are significantly lower costs of extinguishing and operation (what was mentioned earlier) in relation to traditional extinguishers. The results of current paper can be used as basis in future research where the DNN for fire detection will be used for control of autonomous acoustic fire extinguisher which will be able to react independently when fires are detected.

References

1. Węsierski, T., Wilczkowski, S., Radomiak, H.: Wygaszanie procesu spalania przy pomocy fal akustycznych. *Bezpieczeństwo i Technika Pożarnicza* **30**(2), 59–64 (2013)
2. Radomiak, H., Mazur, M., Zajemska, M., Musiał, D.: Gaszenie płomienia dyfuzyjnego przy pomocy fal akustycznych. *Bezpieczeństwo i Technika Pożarnicza* **40**(4), 29–38 (2015)
3. DARPA sound based fire extinguisher, Defense Advanced Research Projects Agency. [Online]. Available: <https://www.extremetech.com/extreme/132859-darpa-creates-sound-based-fire-extinguisher>. Last accessed 14 June 2020
4. Robertson, S., Tran, V., Wave extinguisher. [Online]. Available: https://ece.gmu.edu/~ppach/ECE_Awards/Posters/S-15-I.pdf. Last accessed 18 Nov 2019
5. Myth Busters, Voice Flame Extinguisher, Episode 76. [Online]. Available: <https://mythresults.com/episode76>. Last accessed 14 June 2020
6. Bong-Young, K., Myung-Jin, B., Seong-Geon, B.: A study on suitability of sound fire extinguisher in duct environment. *Int. J. Appl. Eng. Res.* **12**(24), 15796–15800 (2017)
7. Eun-Young, Y., Myung-Jin, B.: A study on the directionality of sound fire extinguisher in electric fire. *Convergence Res. Lett. Multimedia Serv. Convergent Art Humanit. Sociol.* **3**(4), 1449–1452 (2017)

8. Sai, R.T., Sharma, G.: Sonic fire extinguisher. *Pramana Res. J.* **8**, 337–346 (2017)
9. Myung-Sook, K., Myung-Jin, B.: A study on a fire extinguisher with sound focus. *Int. Inf. Inst.* **20**(6), 4055–4062 (2017)
10. Wilk-Jakubowski, J., Urządzenie do gaszenia płomieni falami akustycznymi. Device for flames suppression with acoustic waves (right in force), Exclusive right number: PAT.234266, no application: P.428615, date of application: 18 Jan 2019
11. Wilk-Jakubowski, J., Urządzenie do gaszenia płomieni falami akustycznymi. Device for flames suppression with acoustic waves (right in force). Exclusive right number: PAT.233025, no application: P.427999, date of application: 30 Nov 2018
12. Wilk-Jakubowski, J., Urządzenie do gaszenia płomieni falami akustycznymi. Device for flames suppression with acoustic waves (right in force). Exclusive right number: PAT.233026, no application: P.428002, date of application: 30 Nov 2018
13. Wilk-Jakubowski, J., Urządzenie do gaszenia płomieni falami akustycznymi. System for suppressing flames by acoustic waves (right in force). Exclusive right number: RWU.070441, no application: W.127019, date of application: 13 Feb 2018
14. Stawczyk, P., Wilk-Jakubowski, J.: Non-invasive attempts to extinguish flames with the use of high-power acoustic extinguisher, article submitted for publication [in]: “Open Engineering” (2020/2021)
15. Hausdorf, F.: *Podręcznik budowy zestawów głośnikowych*. VISATON, Poznań (1996)
16. Foley, D., O’Reilly, R.: An evaluation of convolutional neural network models for object detection in images on low-end devices. In: *Proceedings for the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*, Dublin, pp. 350–361 (2018)
17. Janků, P., Komínková Oplatková, Z., Dulík, T.: Fire detection in video stream by using simple artificial neural network. *Mendel* **24**(2), 55–60 (2018)
18. Szegedy, Ch., Toshev, A., Erhan, D.: Deep neural networks for object detection. *Adv. Neural. Inf. Process. Syst.* **26**, 1–9 (2013)
19. Kurup, R.: Vision based fire flame detection system using optical flow features and artificial neural network. *Int. J. Sci. Res.* **3**(10), 2161–2168 (2014)
20. Laganier, R.: *OpenCV 3 Computer Vision Application Programming Cookbook*, 3rd edn. Packt Publishing, Birmingham (2017)
21. Weng, L.: Object Detection for Dummies Part 3: R-CNN Family, 31 Dec., 2017. [Online]. Available: <https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html#roi-pooling>. Last accessed 14 June 2020
22. Zhang, X.: Simple understanding of Mask RCNN, 22 Apr 2018. [Online]. Available: <https://medium.com/@alittlepain833/simple-understanding-of-mask-rcnn-134b5b330e95>. Last accessed 14 June 2020
23. Šerić, L., Stipanicev, D., Krstinić, D.: ML/AI in intelligent forest fire observer network. In: *Third EAI International Conference on Management of Manufacturing Systems*, Dubrovnik (2018)
24. NVIDIA Jetson Nano Developer Kit Detailed Review, 3 Apr 2019. [Online]. Available: <https://www.seeedstudio.com/blog/2019/04/03/nvidia-jetson-nano-developer-kit-detailed-review>. Last accessed 14 June 2020
25. Jetson Nano Developer Kit: User Guide, NVIDIA, DA_09402_002, 8 July, 2019. [Online] Available: <https://developer.nvidia.com/embedded/dlc/jetson-nano-dev-kit-user-guide>. Last accessed 14 June 2020
26. Jensen, G.: *Manual Fire Extinguishing Equipment for Protection of Heritage*. COWI AS, Oslo (2006)
27. Radwan, K., Rakowska, J.: Analiza skuteczności zastosowania wodnych roztworów mieszanin koncentratów pianotwórczych do gaszenia pożarów cieczy palnych. *Przemysł chemiczny* **90**(12), 2118–2121 (2011)
28. Wnęk, W., Kubica, P., Basiak, M.: Standardy projektowania urządzeń gaśniczych tryskaczowych—porównanie głównych parametrów. *Bezpieczeństwo i Technika Pożarnicza* **27**(3), 83–96 (2012)

29. Raghothaman, B., Linebarger, D.A., Begušić, D.: A new method for low-rank transform domain adaptive filtering. *IEEE Trans. Signal Process.* 48(4) (2000)
30. Mihelj, M., Novak, D., Beguš, S.: *Virtual Reality Technology and Applications* (Part of the Intelligent Systems, Control and Automation: Science and Engineering book series). ISCA, vol. 68, Springer (2013)

Chapter 17

Electronic Information Image Processing Technology Based on Convolutional Neural Network



Xiao Min and Guo Mei

Abstract Under the background that modern information technology and the Internet are closely related to people's lives, deep learning algorithm based on convolutional neural network has important application significance in image processing. At present, the development of the top science and technology in various industries in society is often related to artificial intelligence, convolutional neural network algorithm applied to electronic information image processing provides a means and approach for artificial intelligence, and it has been widely used for image analysis in the medical field. The fields of intelligent recognition in computer systems and even image restoration in criminal investigation have contributed to the development of high-quality scientific and technological life in modern society. Based on the above background, this article gives a brief introduction to the related technologies of convolutional neural network applications and electronic information image processing. I hope that those who are interested in research in the related fields will have some knowledge.

17.1 Research Status

In 1994, LeNet was born as the earliest convolutional neural network promoting the development of deep learning. After many successful iterations, Professor Yann LeCun proposed in the paper Gradien-based learning applied to document recognition in 1998. It is the first convolutional neural network successfully applied to digital recognition problems, on the MNIST dataset. LeNet-5 can achieve an accuracy rate of about 99.2% [1].

In 2012, AlexNet implemented by Professor Alex Krizhevsky won the championship in the Image Classification Competition (ILSVRC) organized by ImageNet

X. Min · G. Mei (✉)

College of Software and Communication Engineering, Xiangnan University, Chenzhou, Hunan 423000, China

e-mail: xnxy_gm@xnu.edu.cn

[2]. AlexNet successfully applied tricks such as ReLU, Dropout, and LRN in CNN for the first time.

It avoids the negative effects of increasing the number of network layers in order to obtain better training results. For example, overfit, gradient disappearance, gradient explosion, inception proposed another angle to improve the training effect, the convolutional layers are combined together in a parallel manner to make more efficient use of computing resources and obtain more features under the same amount of calculation [3].

In 2014, the Visual Geometry Group (VGG) model ranked second in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), second only to GoogLeNet, but the VGG model performed better than GoogLeNet in multiple transfer learning tasks [4]. Moreover, VGG model is the preferred algorithm to extract convolutional neural network (CNN) features from images.

In 2015, in order to solve the downgrade problem, Dr. He Kaiming adopted a deep-level residual learning framework to solve the problem of reduced accuracy. Studies have shown that these residual networks are easier to optimize and accuracy can be obtained by increasing depth. This residual network reached an error of 3.57% on the ImageNet test machine and became a champion in the mission of ILSVRC 2015 [5].

17.2 The Principle of Convolutional Neural Network

17.2.1 *Local Receptive Field*

The convolutional layer neurons and the neurons in the preceding layer are not fully connected. In the actual algorithm, neurons connect with a small part of the neurons in the previous layer to communicate data and information. In convolutional neural networks, the area between neurons and neurons is the local receptive field. This model mainly refers to the structure of biological neural networks in nature [6].

Because the biological neural network performs information transmission between neurons, a neuron only needs to be connected to a small part of other neurons, the application of this structure in the neural network is actually considering that the spatial correlation between adjacent pixels in the electronic information image can be fully utilized. It can greatly reduce the number of samples that the system needs to select for image processing, so as to reduce the time required for image processing, reduce the CPU performance requirements for image processing, and achieve better adaptability of image processing technology in different computer systems [7].

17.2.2 Weight Sharing

Each neuron of a convolutional layer in a convolutional neural network has both a partial value and the weight of the local receptive field. In the application of convolutional neural network algorithm, let each neuron have the same bias and weight. The same neuron bias and weights for this entire neural network are called convolution kernels. In using convolutional neural network algorithms, neurons use the same convolution kernel to convolve with neurons in the previous layer. In this way, regardless of the number of neurons, the number of models that need to be trained has always been kept on the same convolution kernel. Therefore, the feature of weight sharing in convolutional neural network is mainly to reduce the number of parameters used in model training, which is beneficial to reduce the memory demand of computer system and speed up CPU processing. And it is an important technical factor to prevent the occurrence of overfitting [6].

17.2.3 Pooling

The main function of pooling operation in the whole convolutional neural network is to keep the model graphics consistent in translation, rotation, scaling, and perspective. In practical applications, pooling is divided into types such as maximum pooling and mean pooling. Among them, the mainstream application method maximum pooling refers to the maximum value of the output pooling area, and the average pooling refers to the average value of the output pooling area.

17.3 Application of Convolutional Neural Network in Image Processing

17.3.1 Image Identification

Image recognition mainly includes three links, namely data preprocessing, feature extraction, and discriminant classification. Data preprocessing is to save the key information in the image and to facilitate the subsequent feature extraction operation. In data preprocessing, there are methods such as image normalization and median filtering. Feature extraction is an abstraction of the original image. The simplified convolutional neural network structure as an extractor can identify the extracted features to the image during the discriminant classification stage. Among them, the methods used in discriminating classification are mainly hidden Markov models, support vector machines based on kernel functions, artificial neural networks, and so on.

17.3.2 Image Semantic Segmentation

Image semantic points can get more information on the collection and analysis of feature sets. It is widely used in medical, criminal investigation, and other fields such as organs or lesions in MR images, cell structures or tumor areas in pathological images. Although the benchmark data has achieved good results, the depth segmentation model has poor generalization ability to unknown datasets due to partial data transfer. This regional migration is more common in histopathological image analysis [8].

The algorithm mainly attempts to align the visual appearance or characteristic distribution between the source domain and the target domain. We propose an adaptive algorithm in the field of histopathology image segmentation, namely the double adaptive pyramid network (DAPNet). The proposed DAPNet reduces the difference between the two domains by combining two domain adaptive components at the image and feature levels. Image-level adaptation considers the overall difference between the source domain and the target domain, such as the color and style of the image, while feature-level adaptation considers the spatial inconsistency between the source domain and the target domain. Specifically, some strategies are used to learn. Firstly, we developed a deep unsupervised domain adaptive algorithm for image segmentation of histopathology. Secondly, we propose two regional adaptive components based on pyramid features to mitigate regional differences in image and feature levels, which provide smarter and more accurate detection methods for modern medical [9]. As shown in Fig. 17.1, both the source image and the target

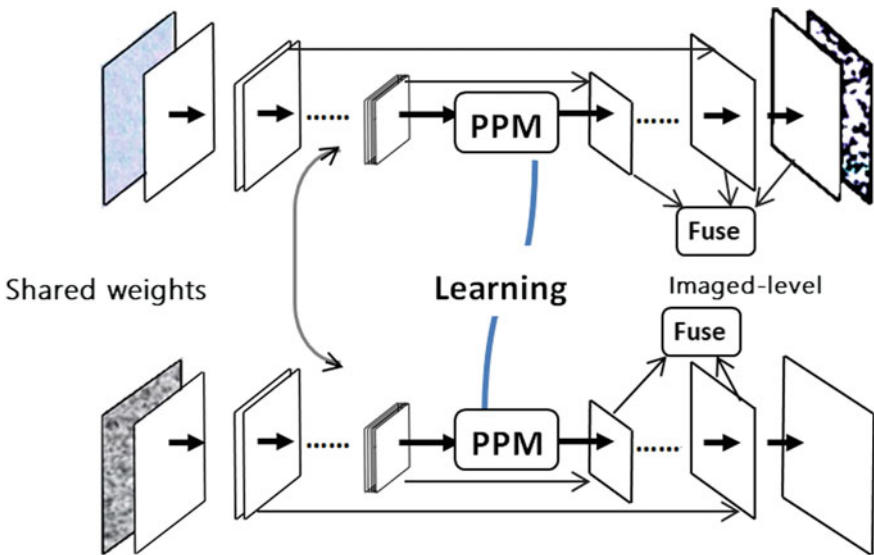


Fig. 17.1 Output image label close to the source domain

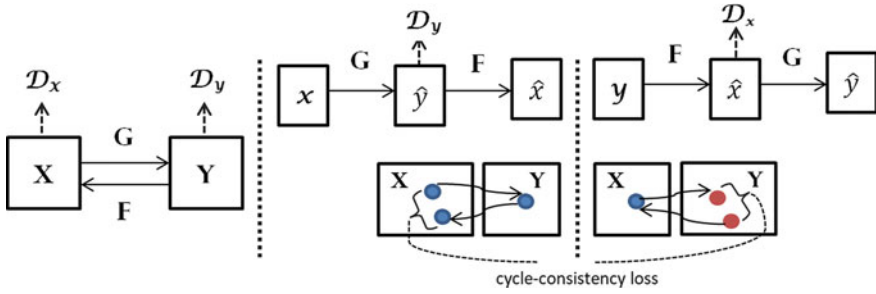


Fig. 17.2 D_x and D_y refer to the discriminator

image are input to the segmentation network, to make it easier to output some images and labels.

17.3.3 Object Detection

The latest methods apply the same idea to conditional image generation applications. The key to the successful application of this image processing technique is that the generated image is in principle indistinguishable from the real one. This leads to a particularly large task of image generation, and this is precisely the goal of many computer graphics optimization. The anti-loss method is used to learn the mapping, which makes the transformed image indistinguishable from the image in the target domain. Using a single input to train a non-parametric texture model on an image pair, the concept of object matching between images dates back at least to earlier image analogies. It is very effective to use the convolutional neural network to learn the parameterized translation function [10]. The method is based on the “pix*pix” framework, which uses a conditional inverse transformation network to learn the mapping from input to output image. Similar functions have been applied to different tasks, such as generating photos from sketches, or generating photos from attributes and semantic layout [11]. Compared with other image processing techniques, the biggest difference of this kind of learning mapping is that there is no paired training example. As shown in Fig. 17.2, the object detection involves matching between the target domain and the source domain and D_x and D_y refer to the discriminator.

17.4 The Conclusion

Deep learning based on convolutional neural networks is a situation where today’s technology is developing faster and faster. As a method of information processing by computer technology simulating human brain thinking, it has important meaning

in the application of artificial intelligence. In deep learning, neural networks have a greater advantage in image processing than traditional image processing methods. They can extract more and more complex features of images. They have applications in various industries in society and have changed people's lifestyles. In this paper, the relevant concepts of image processing technology for electronic information based on convolutional neural network are expounded, and the application of convolutional neural network in image processing is briefly introduced based on the current development of technology, which is expected to be helpful for scholars in related fields to study related problems.

Acknowledgements This paper is funded by Project of:

1. Scientific Research Fund of Hunan Provincial Education Department, Research on Intelligent Medical Logistics Distribution based on Particle swarm optimization (No. 19C1708).
2. Chenzhou Municipal Science and Technology Project, Research on key technologies of intelligent parking in smart city (No. zdyf201913).
3. Chenzhou Municipal Science and Technology Project, Research on AR system based on Chenzhou Mining Fair (No. zdyf201911)
4. Hunan Provincial Department of Education Fund, Computer Application Technology Innovation and Entrepreneurship Education Center, (No. [2018] 380)
5. Hunan Provincial Department of Education Fund, Big Data and Processing Innovation Entrepreneurship Education Base (No. [2019] 333).

References

1. Sun, C., Pan, S.: An artificial target detection method combining a polarimetric feature extractor with deep convolutional neural networks. *Int. J. Rem. Sens.* **41**(13) (2020)
2. Yang G.: Detecting regional dominant movement patterns in trajectory data with a convolutional neural network. *Int. J. Geograph. Inform. Sci.* **34**(5) (2020)
3. Bándi, P., Balkenhol, M., van Ginneken, B., van der Laak, J., Litjens, G.: Resolution-agnostic tissue segmentation in whole-slide histopathology images with convolutional neural networks. *Peer J.* **7** (2019)
4. Szegedy, C., et al.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015)
5. Shibata, N., Tanito, M., Mitsuhashi, K., Fujino, Y., Matsuura, M., Murata, H., Asaoka, R.: Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci. Rep.* **8**(1), 262–267 (2018)
6. Kappeler, A., Yoo, S., Dai, Q., et al.: Video super-resolution with convolutional neural networks. *IEEE Trans. Comput. Imag.* **2**(2), 109–122 (2016)
7. Shuya, G., Yueqing, G.: A radar target classification method based on convolutional neural network. *Inform. Technol.* **44**(01), 91–94 + 100 (2020)
8. Changmei, C., Yanbin, L.: Research on modulation pattern recognition based on convolutional neural network. *Inform. Technol.* **44**(01), 101–106 (2020)
9. Wen, G., Zhongjian, J., Qingnan, W., Hong, Q., Xiangkun, D.: Research progress of automatic organ segmentation based on deep learning. *Med. Health Equip.* **41**(01), 85–94 (2020)
10. Yingying, X., Hongbin, S.: Research status of biomedical image processing based on pattern recognition. *J. Electron. Inform. Technol.* **42**(01), 201–213 (2020)
11. Fengshou, H, You, H, Zhuoduo, L., Congan, X.: Research progress of convolutional neural network in radar automatic target recognition. *J. Electron. Inform. Technol.* **42**(1):119-131 (2020)

Chapter 18

Research on Related Problems of Intelligent Medical Logistics Distribution Based on Particle Swarm Optimization



Guo Mei and Xiao Min

Abstract Pharmaceutical logistics distribution is a kind of physical transportation method that sends drugs from manufacturers to relevant medical units and departments. It is based on the extensive application of traditional logistics systems in the field of medicine. Since logistics distribution is a way to directly contact consumers, it largely determines the service level and operating costs of medical companies. Therefore, logistics distribution is also an important part of the logistics link, and it has been widely valued and studied by relevant parties. In addition, to show a good service attitude in the logistics system how to reduce operating costs is also a key issue. Therefore, the particle swarm intelligence optimization algorithm is used to reasonably plan the distribution path of transportation vehicles in logistics transportation to effectively improve transportation efficiency. Reducing transportation costs to a greater extent has extremely important theoretical significance and use value for the research and application of logistics.

18.1 Introduction

With the rapid development of medical logistics in my country in recent years and the improvement of the medical system by relevant departments, medical logistics distribution has become the focus of social attention. For traditional medical drug distribution, order processing, cargo sorting, and drug delivery routes are all determined by staff based on experience and lack scientificity and rationality [1]. Therefore, the automation and informationization of logistics distribution can not only shorten the inventory time and reduce logistics costs, but also improve the service level and the use efficiency of funds. This paper describes the integration of logistics distribution and particle swarm optimization algorithm, proposes mathematical models and

G. Mei · X. Min (✉)

College of Software and Communication Engineering, Xiangnan University, Chenzhou 423000, Hunan, China

e-mail: xnxy_xm@xnu.edu.cn

application advantages for the application of particle swarm optimization algorithm in logistics distribution path, and makes clear the importance of intelligent optimization algorithm in reducing distribution cost, improving logistics management level and building a new intelligent medical logistics distribution system [2].

18.2 Problems in My Country's Medical Logistics Distribution

Although my country's medical transportation industry has developed rapidly in recent years, compared with other countries, there are still many shortcomings in my country's medical logistics distribution system. In my country, there are many enterprises, high costs, low efficiency, poor order, and rigid management. There is no logistics distribution center that realizes industry collectivization or regionalization. Therefore, there is still much room for improvement on how to reduce transportation costs and increase the economic benefits of enterprises.

1. Delivery is scattered, items are not concentrated: warehouse storage costs are large, industry logistics costs increase.
2. Policy restrictions, the distribution center cannot be established: Restricted management of the medical industry by the state and relevant departments has caused the increase in the cost of goods and the rise in product prices.
3. Procurement is unscientific, human, and material resources are wasted: purchasing drugs is not specified, resulting in waste of resources [3].
4. The distribution route is random, and the punctuality is poor: as the delivery requirements are not strict, the requirements for customers cannot be completed.

18.3 Introduction and Advantages of Particle Swarm Optimization Algorithm

18.3.1 Introduction to Particle Swarm Optimization

Particle swarm optimization is a unique algorithmic approach in optimization algorithms, and it also belongs to a form of evolutionary algorithms. Its design principle is to simulate the predatory behavior of birds, and it is a random search algorithm based on group collaboration [4]. Every problem that needs optimization can be seen as a bird in the search space: They do not know how far they are from the food, so the simple and efficient way is to look around the bird closest to the food. Therefore, the different optimization problems are regarded as many particles, each particle has an optimized function to determine the fitness value, and then follow the optimal particle through continuous iteration to find the optimal solution in the search space. In the face of the extensive search space and time-limited task requirements, the

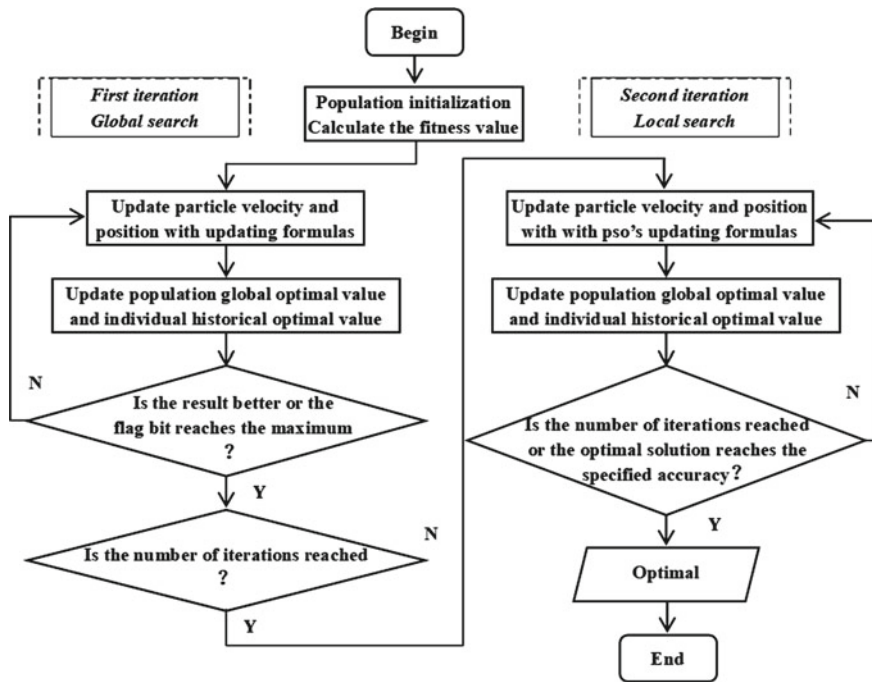


Fig. 18.1 Flow chart of particle swarm algorithm

algorithm can explore the target through the two stages of global search and local search. Its flow chart is shown in Fig. 18.1:

18.3.2 Application Advantages of Particle Swarm Optimization

1. Fast search speed: the particle swarm algorithm does not have crossover and mutation operations. It uses the transmitted information to search for the optimal solution [5].
2. Strong information memory: particle swarm's super algorithm ability can handle many complex problems.
3. The algorithm rules are simple and efficient: by using real code, the adjustable parameters are reduced, which reduces the cost of enterprises to some extent and slows down the accumulation of goods warehouse [6].
4. Fast convergence speed: In the process of searching and optimizing information, problems can be continuously dealt with, through different ways and measures to avoid falling into local optimum, and reduce the error rate.

18.4 Application of Particle Swarm Optimization in Logistics Route Optimization

18.4.1 *Mathematical Model of Distribution Vehicle Routing Problem*

The routing problem of logistics transportation vehicles can be divided into two categories: objective function and constraints. The objective function includes the shortest total mileage of delivery, the smallest tonnage kilometers of delivery vehicles, the lowest comprehensive cost, the highest punctuality, the most reasonable capacity utilization, and the lowest labor consumption. The bundle conditions include meeting the customer's requirements for the quality and quantity of the goods, meeting the customer's limited time range, ensuring the transit time, the customer's requirements are allowed, within the capacity of the distribution center, and the delivery vehicle's load capacity is qualified. In the establishment of the mathematical model, both the objective function and the constraint conditions need to be fully considered. In addition, the mathematical model of the vehicle path needs to establish a reasonable coding method. The enterprise has a total of N car yards, each of which has M vehicles for logistics and distribution. It is responsible for the distribution of drugs to N customers. The demand of customer A is Q (Q is less than N), and each car in the yard can serve customers. However, each vehicle can only be serviced once for each customer, and the vehicle can return to the yard after completing the task, so that each vehicle can serve different customers, reduce transportation costs, and save transportation time. In the mode, you need to pay attention to the actual situation, calculate the shortest path and transportation time, and ensure that the vehicle can complete the task on time.

18.4.2 *Components of Mathematical Model of Distribution Route*

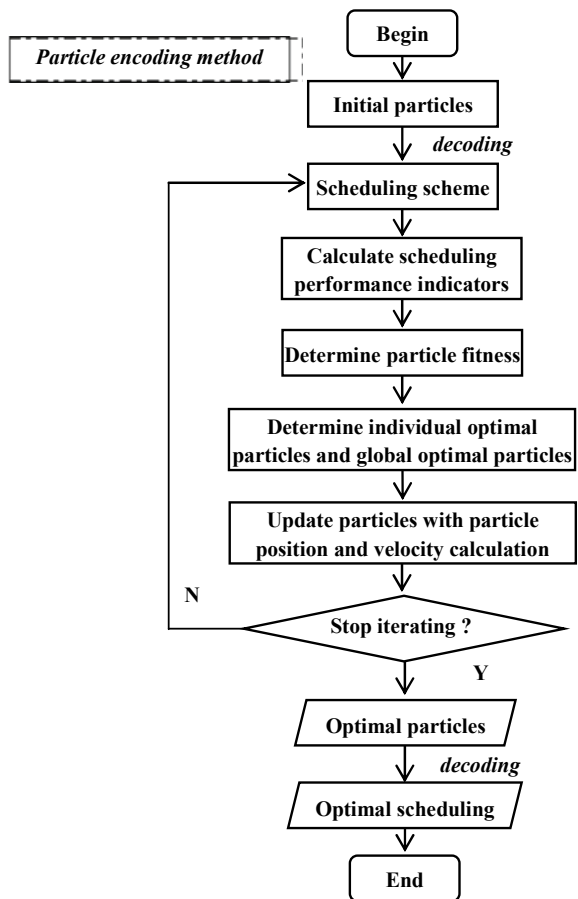
1. Clarify the principle of particle swarm optimization
The principle of the particle swarm optimization algorithm is to select the optimal solution from the candidate solutions through iterations. In the search space, it is determined by moving different particle velocities. Move to the best position.
2. Determine the encoding and decoding method
Particle swarm optimization is an optimization algorithm based on continuous space usually used in order integration of work tasks and distribution of data, and its spatial data is discrete. Therefore, the key to using particle swarm optimization algorithm is to establish a reasonable mapping relationship between the position of the particle and the solution of the problem, and determine the appropriate coding idea according to different optimization problems. In addition, because

the fitness value of particles needs to be guided by the optimal particle in the search space, and the optimal particle is determined by the optimal performance index and the survey plan, it is also very important to obtain the optimal survey plan in the particle swarm. Therefore, determining the encoding and decoding methods often determines the quality of the optimal survey plan. The output process based on the encoding method and decoding to the optimal survey plan in the algorithm process is shown in Fig. 18.2:

3. Algorithm implementation process

First of all, according to the difference of the initial conditions of different “particles”, it is concluded that the behavior should be made in accordance with the rationality and standardization of things. In the process of assigning “particles”, the initial speed and initial position of “particles” should be used as the basis for the behavior after the implementation.

Fig. 18.2 Flow chart of structure based on encoding and decoding in particle swarm optimization



Keeping the state of things stable, the appropriate situation is to make the “particles” have a single stable individual transportation environment and the right group travel speed. Let each “particle” have a subtle relationship to make transportation more efficient. Therefore, the fitness value of a “particle” is calculated by various methods to stabilize the relationship.

In transportation, individuals and groups must maintain a good relationship. While keeping the individual’s speed as fast as possible, the speed of the entire group should also be maintained.

In practical applications, this calculation method is very labor-intensive. High-tech calculation methods should be used reasonably to achieve the most fundamental purpose.

18.5 The Conclusion

In summary, the optimization of the intelligent logistics distribution system based on particle swarm can improve the problems in the logistics system from many aspects. It effectively solves the situation that the order information is not concentrated in the logistics distribution process, the cargo classification is not in place, the waste of manpower and material resources in the logistics distribution process, and the low service level. The application of particle swarm optimization algorithm in the logistics distribution process clarifies the distribution tasks of each vehicle, saves the waste of time in the logistics transportation process, improves the efficiency of logistics distribution, reduces the operating costs of enterprises, and improves the logistics system management mechanism. However, particle swarm optimization also has more areas to be improved. More workers are required to develop and research to fully improve its defects, and it is widely used in different distribution fields to create greater value for smart medical logistics distribution.

Acknowledgements This paper is funded by project of:

1. Scientific Research Fund of Hunan Provincial Education Department, Research on Intelligent Medical Logistics Distribution based on Particle swarm optimization (No. 19C1708).
2. Chenzhou Municipal Science and Technology Project, research on key technologies of intelligent parking in smart city (No. zdyf201913).
3. Chenzhou Municipal Science and Technology Project, Research on AR system based on Chenzhou Mining Fair (No. zdyf201911).
4. Hunan Provincial Department of Education Fund, Computer Application Technology Innovation and Entrepreneurship Education Center (No. [2018]380).
5. Hunan Provincial Department of Education Fund, Big Data and Processing Innovation Entrepreneurship Education Base (No. [2019]333).

References

1. Zhuangkuo, L.: Research on Medical Logistics Distribution Vehicle Routing Problem Based on Particle Swarm Optimization. Guilin University of Electronic Technology (05) (2016)
2. Shuli, Z., Tao, Z., Yan, C., Rengui, I.: Logistics Distribution Path Optimization Method Based on Chaotic Particle Swarm Optimization, Northwestern Polytechnical University Mingde College, Air Traffic Engineering College of Air Force Engineering University, 2019 (08)
3. Guo, S., Qin, G., Zhang, J., Yu, H., Lu, Z., Yu, J.: J. Xi'an Jiaotong Univ. (2016)
4. Yuan, X., Yang, J., Zhang, X.: Logistics Engineering and Management (2017)
5. Ge, L., Evans, G.M., Moreno-Atanasio, R.: CFD-DEM Investigation of the Interaction Between a Particle Swarm and a Stationary Bubble: Particle-Bubble Collision Efficiency. Powder Technol. **366** (2020)
6. Zuazola, I.J.G., Moreno, A.: Telematics system for the intelligent transport and distribution of medicines, IET Intell. Transp. Syst. **7**(1), 131–137 (2013)

Chapter 19

Study of Wool Image Recognition Based on Texture Features



Yao Juan, Xu Wang, Zhang Cheng, and Tian Fang

Abstract Among researches on parameter extraction in sheep body measurements, wool is one of the important influence factors causing sheep body measurement errors. Wool images were analyzed in this study using ridge regression algorithm, KNN algorithm, and SVM to discriminate wool length. The final body measurement parameters were processed according to the discrimination results in order to reduce the measurement errors caused by non-uniform wool length.

19.1 Gray-Level Co-occurrence Matrix

In the early 1970s, R. Haralick [1] and others designed a statistical method of gray-level co-occurrence matrix (GLDM), which can analyze the texture in the space. Gray-level co-occurrence matrix is applied to texture analysis in space, and the precondition for using this method is that spatial distribution relations between pixels in the image reflect the image texture information [2].

Haralick has constructed 14 kinds of statistics extracted from gray-level co-occurrence matrix, including energy, entropy, contrast, variance, and maximum correlation coefficient [3].

In [4], Wang K. J. et al. from Jiangnan University used a four-scale grey-level co-occurrence matrix to extract contrast ratio, correlation, angular second moment, homogeneity, and entropy in order to characterize the change of fabric drape, followed by classification through SVM, and good effect was achieved. In [5], Shi Y. F. et al. made differential diagnosis of glioblastoma and primary central nervous system lymphoma through a texture analysis based on the gray-level co-occurrence matrix. In [6], Xu J. C. et al. used the gray-level co-occurrence matrix to extract image texture features, and then constructed an aphid damage diagnosis model, which could effectively realize identification of aphid-damaged cotton leaves and provide a

Y. Juan · X. Wang · Z. Cheng · T. Fang (✉)
Huazhong Agricultural University, Wuhan, Hubei, China
e-mail: fangzhihai_2003@mail.hzau.edu.cn

technical support for fast recognition of insect pest situation. In [7], Wang Q. T. et al. used the improved gray-level co-occurrence matrix to extract multiple eigenvalues of wood material, which were then trained and classified using a pattern recognition algorithm. The results showed that the classification effect was favorable, so it could be a new wood recognition method [7].

Some mathematical statistical quantities are constructed by the grey-level co-occurrence matrix in various fields to classify texture features. Sheep images were analyzed, and length classification was performed in this study using the gray-level co-occurrence matrix.

19.2 Image Preprocessing

The background region of sheep images acquired from non-contact sheep body measurements was large. To reduce program calculation and accelerate calculating speed, the images should be clipped in consideration of objective situation of wool image recognition. The images used in this experiment were sheep gray-level images with background already removed, their resolution was 800×600 , and the clipping process of wool images is shown in Fig. 19.1.

1. The Center1 of sheep circumscribed matrix was acquired through the minimum circumscribed matrix method;
2. The maximum inscribed matrix was obtained using the center expansion method with Center1 being the starting point, and Center2 of the maximum inscribed matrix was determined;
3. 160×120 (pixels) rectangular areas were clipped from the sheep images by centering on Center2; the 160×120 wool images were saved in the specific folder.

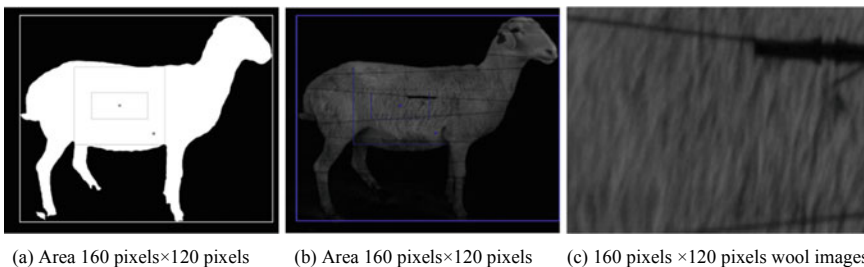


Fig. 19.1 Acquisition process of local wool images

19.3 Texture Features and Parameter Extraction of Wool Image Texture Features

19.3.1 Texture Features

1. Energy

The size of energy can determine whether the gray distribution of the image is uniform and the texture thickness. If the values of each element in the gray-level co-occurrence matrix are similar and the energy value of the image is small, it means that the texture in the image is fine. If the values of the elements are different greatly, the energy value is large, and the texture of the image is uniform and changes regularly.

2. Entropy

The size of entropy reflects the richness of information in the image. When all elements in the gray-level co-occurrence matrix are random and scattered in the numerical distribution, the entropy value will be larger. It represents the degree of texture non-uniformity and the complexity of image gray distribution. The larger the entropy value is, the more complex the image composition is.

3. Contrast

Contrast is a parameter used to measure the distribution rule of gray-level co-occurrence proof median and the degree of local change of the image. It can be used to estimate the clarity of the image and the depth of the gully of the object texture. The greater the contrast value is, the deeper the corresponding texture ravines will be, the greater the pixel contrast is, the clearer the image will look. The greater the value of elements far away from the diagonal in the grayscale common matrix is, the greater the value of contrast will be.

4. Reverse differential distance (consistency)

The inverse variance is related to the homogeneity of image texture, which reflects the characteristics of local change of image texture. If the distribution of object texture in different regions of the image is relatively uniform and the change is slow, the value of inverse variance will be larger. Otherwise, it will be smaller.

5. Correlation

Correlation is used to measure the similarity of gray level of image in the direction of image row or column. It can be used to measure the local correlation of image. When the gray matrix element values are even and equal, the correlation value will be larger; otherwise, if the matrix element values are very different, the correlation value will be smaller. If there are horizontal texture features in the image, the correlation value of the horizontal matrix will be greater than that of the rest of the matrix.

19.3.2 Parameter Extraction of Wool Image Texture Features

The parameters of 393 background-removed images were calculated and recorded via the gray-level co-occurrence matrix. According to the calculated original data, they were directly observed first with naked eyes to construct a length classification dataset of wool images. Through calculation, the entropy of long-wool sheep was generally higher than that of short-wool sheep, and long-wool sheep was different from short-wool sheep to a certain degree also in contrast ratio. However, the consistency change was minor in the 393 groups of data, and then energy, entropy, and contrast ratio were selected to investigate the discrimination method of wool length. The following are the definitions of the three selected features.

19.4 Wool Image Judgment and Recognition

19.4.1 Ridge Regression Method

Ridge regression is a biased estimates-based regression method, and it is actually an improved linear least square method with L2 regularization, which is mainly applied to linear data analysis. This method is used to improve the acquisition efficiency of regression coefficients by abandoning unbiasedness of least square method, ignoring partial data information and degrading the calculation accuracy. As a regression method closer to reality, it has higher better fitting accuracy rate for ill-posed data than least square method. When concentrated data is of colinearity, the ridge regression algorithm will reach a better estimation effect.

Energy, entropy, and contrast ratio in each group of data was set as independent variables, and the corresponding long wool or short wool as dependent variable. In the data analysis, figures 0 and 1 expressed short wool and long wool, respectively. The regression formula was decided by taking 0.5 as threshold value, the predicted result after data substitution was long wool or short wool, and if the value was above 0.5, the predicted result was long wool, and it was short wool when the value was below 0.5.

19.4.2 KNN Algorithm

KNN algorithm is a classification algorithm in supervised learning. Its working principle is to train the sample dataset, complete space partitioning of sample eigenvectors, and then take the partitioning results as the final algorithm model. The sample dataset is also called training sample set, where each data has its own label. When an unlabeled data is processed in the experimental process, its features will be compared with those of already known data in the sample set first, and then data having the

highest similarity with it among the samples is obtained and labeled as its classification label, namely the classification of the data. Under normal conditions, only the first k data which are most similar in the sample dataset are selected, which is the origin of k value in KNN algorithm, and k is generally an integer smaller than or equal to 20. In the end, labels which appear most frequently in the k most similar data are selected as the classification of this data.

Energy, entropy, and contrast ratio in each group was combined and regarded as one point to be investigated using three parameters. Hence, the problem could be transformed into selecting the k -nearest points in 3D space according to Euclidean distance between data, so as to solve the label with the largest proportion in the result as the discrimination result.

19.4.3 SVM Method

This is a classical and commonly used classification method which improves the theoretical weaknesses of traditional neural networks. As a supervised learning method, it only needs to organize and classify data in the training set and test set. The verification result can be obtained by quoting relevant libraries and calling relevant functions in C++ program.

The core of an SVM-specific classification algorithm lies in the kernel function adopted in model training, where common kernel functions include LINEAR, POLY, and RBF. Given this, the three kernel functions were successively used to perform training and the same test as before. It was found that the quantities of images not conforming to practical situation were two and one, respectively. Among the three kernel functions, RBF had more excellent performance with error rate of 2.5%. Based on a comparison with the results previously obtained through ridge regression, it was rightly found that the ridge regression discrimination result of that image with discrimination error was also the opposite to the fact. It was the problem of the image itself through inspection. Although this image seemed presenting the opposite wool length result to the discrimination result, this further proved the necessity of applying multiple images of the same sheep from different angles to discrimination for improving the accuracy rate. However, SVM only performed training once, after which the imported test data should be tested only through a trained model, so thanks to training ahead of time, there was no reason to worry that the enlarged data size in the training set might impact the test time. After the training was completed, it was only necessary to import test data again for testing. The following presents a comparative analysis between different kernel functions (Fig. 19.2).

The first image was trained using RBF, LINEAR, and POLY kernel functions so as to obtain fitted results. As for fitting degree, LINEAR was not much different from RBF under linearly separable situation. However, it was obviously superior to LINEAR under linearly inseparable situation. POLY did not show very good effect under both situations, but POLY could be slightly better under dramatic change. When it comes to speed, LINEAR was the fastest and POLY was the lowest due

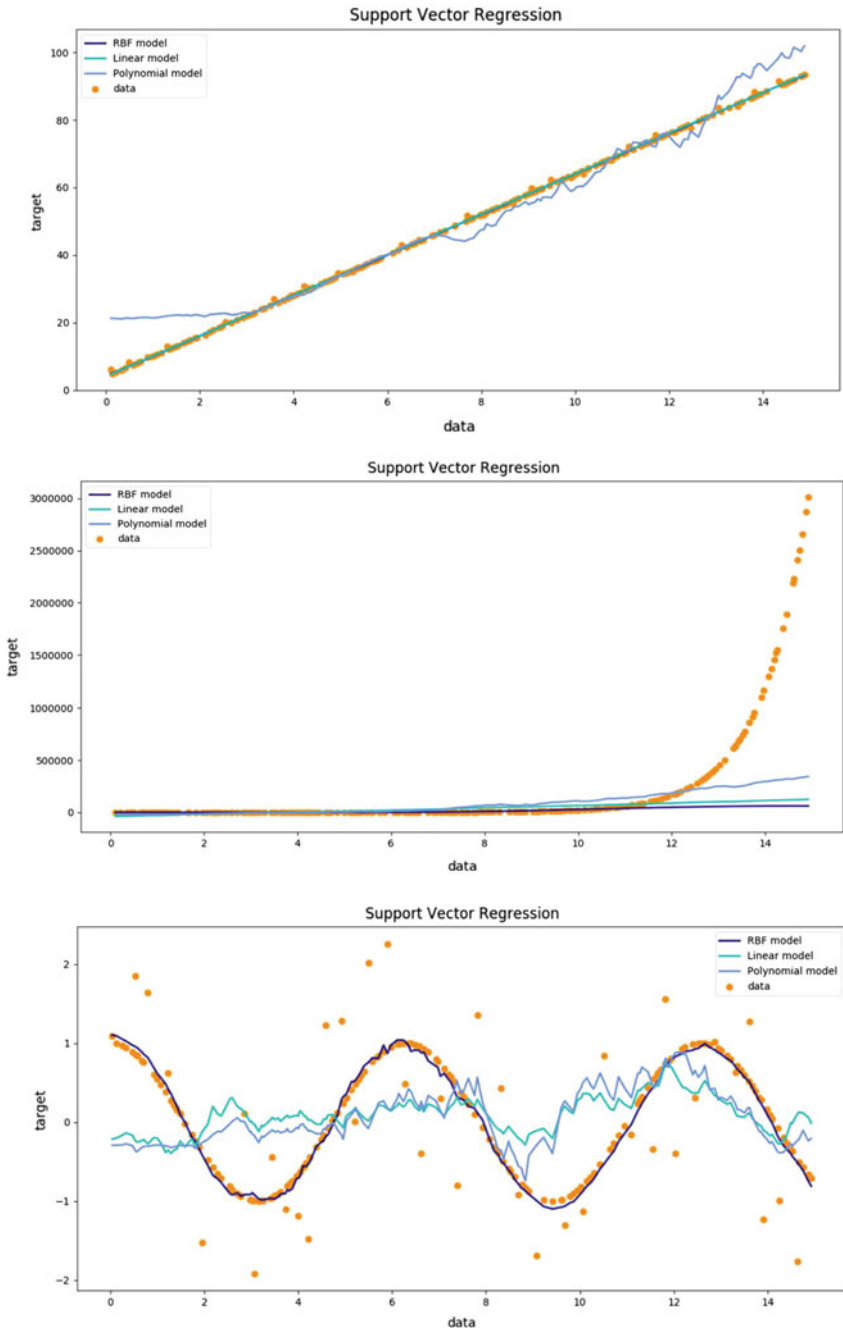


Fig. 19.2 A comparative analysis between different kernel functions

Table 19.1 Error rate of different algorithms

Algorithm	Ridge regression	KNN ($k = 10$)	KNN ($k = 15$)	KNN ($k = 20$)	SVM (RBF as kernel function)
Error rate	2.5%	0%	0%	5.1%	2.5%

to many parameters. In terms of parameters, LINEAR was simple and easy to use. Though having many parameters, RBF and POLY could obtain good results only if the parameters were well adjusted. Hence, on the whole, RBF kernel function could reach favorable practical effect, and then RBF was selected as the final kernel function.

19.5 Result Analysis

19.5.1 Analysis of Error Rate

The same regression or training data and the same verification set were adopted in the three methods used in this paper, where for the selection of verification set, several groups of data of the sheep were randomly selected from the 393 groups of data, a total of 40 groups of data were selected, and the rest data served as regression or training data. The error rates of the three algorithms are compared as seen in Table 19.1.

It could be seen from the results listed in the above table that under appropriate k value, the error rate of KNN algorithm was the lowest, followed by those of ridge regression algorithm and SVM algorithm in succession. The reason for high accuracy of KNN algorithm was that the classification decision rules were considered in the algorithm implementation process, the distance between each group of test data and sample data was calculated, and sample size and data dimensionality were both considered. To acquire regression coefficients more conforming to reality, ridge regression algorithm gave up the unbiasedness of least square method, which led to partial information loss and accuracy degradation.

19.5.2 Analysis of Operation Time

As different algorithms varied in the needed calculated quantity and also in time needed in batch image recognition, the same computer and compiling environment were used to test the same several hundred groups of data, and it was found that the operation time needed by the ridge regression algorithm was the shortest, followed by SVM and KNN successively. In terms of algorithm mechanism, the core step of ridge

regression method was regression analysis via a regression analysis software. Only the obtained formula was input into the discriminant program and then substituted into the test data, and it was obviously the simplest discriminant mechanism. The reason for long time needed by KNN algorithm was that after each group of test data were imported, Euclidean distances between these data and data in the training set needed to be calculated, and moreover, the solved results should be sorted to select k distances as short as possible. Under an increasing number of training sets and test sets, the needed operation time would present geometric growth. However, SVM algorithm had one set of its own operation flow. Through the test, it was found that under the present data size, the needed test time it needed was not much different from that needed by the ridge regression algorithm.

To sum up, three different methods were used to accurately discriminate wool lengths in the images. All of the three methods carried out machine learning based on related parameter data in the gray-level co-occurrence matrix of the images. Ridge regression and SVM were both ideal, KNN algorithm had high accuracy, but it had an obvious defect in its discriminant operation time. Through a comparison of ridge regression and SVM in operation time and error rate, SVM algorithm was selected in this study as discrimination method for wool length by virtue of its advantages in solving problems like small sample size and nonlinearity.

19.6 Conclusion

In this paper, the maximum inscribed rectangle, the minimum circumscribed rectangle, and other mathematical features are used to cut in the original image of sheep, and the local image of wool with the size of 160 pixels \times 120 pixels is cut out, which is divided into training set (70%) and testing set (30%), and the methods of ridge regression, KNN and SVM are, respectively, used to train these images and judge the length of wool image, and the errors of various methods are also analyzed compared with the running time, SVM is chosen as the method of wool image discrimination because of its short running time and relatively good accuracy.

References

1. Haralick, R., Shanmugam, U., Dinstein, I.: Textural features for image classification. *IEEE Trans. Syst. Man Cybern. SMC* **3**(6), 610–621 (1973)
2. Gao, C.C., Hui, X.W.: GLCM-based texture feature extraction. *Comput. Syst. Appl.* **19**(6), 195–198 (2010)
3. Wang, Z. X.: Detection of intensive crowd in City public places. [Master's thesis]. North University of Technology, Beijing, China (2017)
4. Wang, K.J., Wang, J.A., Gao, W.D.: Multi-scale GLCM analysis of fabric wrinkles. *J. Silk* **57**(2), 35–40 (2020)

5. Shi, Y.F., Qian, L.X., Guo, X.Y.: Texture feature analysis based on gray level co-occurrence matrix for differential diagnosis of glioblastoma an primary central nervous system lymphoma. *Chin. J. Interv. Imaging Ther.* **4**, 228–232 (2000)
6. Xu, J.C., Lv, X., Lin, J., Zhang, Z., Yao, Q.S., Fan, X.L., Hong, Y.H.: The diagnostic model of cotton aphids based on leaf textural features. *Cotton Sci.* **2**, 133–142 (2020)
7. Wang, Q.T., Yang, J.: Application of improved gray symbiosis matrix to identify the multiple characteristic values of wood texture. *J. Northwest Forestry Univ.* **3**, 191–195 (2019)

Chapter 20

The Observation and Simulation of Dynamic Diffraction Patterns Caused by a Cylindrical Liquid Diffusion Pool for Diffusivity Measurement



Licun Sun, Yuanfangzhou Wang, Linhai Li, Jie Feng, Ya Liu, and Shuwu Sheng

Abstract The dynamic diffraction patterns caused by an asymmetric liquid-core cylindrical lens filled with diffusion solution are observed using CCD and simulated using MATLAB based on Collins formula in this paper. The diffusion coefficient of ethylene glycol diffusing in pure water is measured at 298.15 K by analyzing the movement of the narrowest position in diffraction patterns base on Fick's second law. The correctness and application conditions of this measurement method are further interpreted by comparing the simulation and experimental diffraction patterns for the first time. The work expressed in this paper deepened the understanding of the scalar diffraction theory, diffusion process and Fick's second law, which paved a way for the further research on diffusion phenomenon.

20.1 Introduction

The particle movement caused by concentration gradient leads to diffusion, which can be expressed as Fick's law [1, 2]. As a significant physical property, the liquid diffusion coefficient (D) plays an important role in many fields, such as pharmaceutical industries, biological reactors, and wastewater treatment and so on [3–6]. In order to determine the liquid D values, many chemical and physical experimental methods have been put forward in recent years, e.g., Taylor dispersion method [7], the fluorescent molecule tracing [8], holographic interferometry [9] and dynamic light scattering [10] and so on, each with its own strengths and limitations, respectively. Long

L. Sun (✉) · Y. Wang · J. Feng · Y. Liu · S. Sheng

Yunnan Key Laboratory of Opto-Electronic Information Technology, School of Physics and Electronic Information Technology, Yunnan Normal University, Kunming 650500, China
e-mail: aliceckczy@126.com

L. Li

School of Cultural Tourism and International Exchange, Yunnan Open University, Kunming 650500, China

time-consuming (>2 h), strict requirements for experimental environment, and unobservable dynamic diffusion processes are several common disadvantages existing in these measurement methods.

In order to overcome these shortcomings, our group recently designed an asymmetric liquid-core cylindrical lens (ALCL) [11], a key imaging element in the liquid D values measurement system, which can be used as a diffusion pool due to its one-dimensional resolution ability for the refractive index (RI) of the filled liquids. Once two diffusion solutions contact in the ALCL, dynamic gradient distribution of the mixed solution concentration, and RI is formed gradually along the diffusion direction due to particle motion and a group of dynamic diffusion patterns will be received by the observation device CCD. Both the shape and the variation of the images reflect the diffusion information, therefore, two ways, namely “instantaneous image analytical method” [12] and “equivalent RI of thin liquid layer method” [13], are brought to determine the liquid D values using Fick’s second law by analyzing the diffusion images.

The “instantaneous image analytical method” introduced in detail in [12] figured out the liquid D values by analyzing a single diffusion image demonstrating our experimental setup equipped with the ALCL has a significant advantage in avoiding time-consuming. The “equivalent RI of thin liquid layer method” elaborated in [13] measured the liquid D values by recording the locations of the “waist” of the diffraction images changing with the diffusion time showing that the equipment is simple and stable and clarifying the validity and visualization of our method.

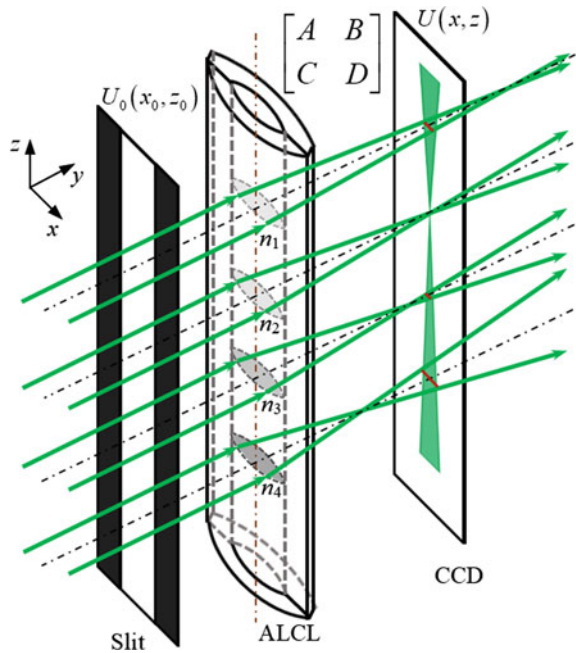
In this paper, the correctness and application conditions of the “equivalent RI of thin liquid layer method” were further interpreted by comparing the simulation and experimental diffraction patterns, and the ethylene glycol (EG) diffusing in pure water at 298.15 K is chosen as an experimental study case. The diffraction patterns are formed by the collimated light passing through the ALCL filled with diffusion sample. As the beam size at the aperture plane and the distance from the aperture plane to the evaluation are large enough compared with the wavelength in this experiment, the scalar diffraction theory (SDT) is accurate enough to model the propagation of light [14]. Collins formula, belonging to SDT, is convenient in the diffraction research on paraxial optical systems [15, 16] and can be completed by using the fast Fourier transform (FFT) [17] usually. FFT has been used widely in signal processing [18] and the application of Fourier optical theory in solving diffraction problems has been becoming more mature [19], but the study of diffraction from cylinder lens is relatively scarce [20]. Based on the Collins formula, using FFT techniques and MATLAB software running in a PC, the dynamic diffraction patterns of aperture of ALCL, in which two different solutions are mutual diffusing, are calculated for the first time in this paper. Both similarities and differences existing between the simulation results and the CCD recorded help us understand the diffusion phenomenon more deeply.

20.2 Principle

20.2.1 The Experimental Setup and Imaging Principle

The expanded and attenuated collimated beam with $\lambda = 589 \text{ nm}$ illuminates an adjustable slit. After passing through the ALCL filled with liquid, the wave converges on the sensitive plane of the CCD with 2448×2058 pixels and $3.45 \times 3.45 \mu\text{m}^2$ per pixel size. The CCD, connected to a PC via a USB, was driven by a translation stage with 0.001 mm minimum scale to adjust the imaging plane position slightly, which made the observation for a diffusion image convenient. The key optical elements of the experimental setup were shown in Fig. 20.1. The ALCL consists of two distinctive K9 glass ($n_0 = 1.5163$ at $\lambda = 589 \text{ nm}$) cylindrical lenses. $R_1 = 20.0 \text{ mm}$, $R_2 = R_3 = 17.0 \text{ mm}$, and $R_4 = 37.6 \text{ mm}$ are the curvature radii of the four refraction surfaces of the ALCL, respectively. And $d_1 = d_2 = d_3 = d = 3 \text{ mm}$ are the wall thicknesses and cavity thicknesses, respectively. Seal the bottom of the ALCL by rubber film firstly and then inject the higher density solution EG into the ALCL, occupying the lower half part of the ALCL. Sequentially, trick the same volume of distilled pure water slowly along the inner wall 10 min later, avoiding convection and turbulence in the two liquids. As soon as the two solutions contact, the diffusion process begins, and the time of the two solutions starting to contact was defined as the onset of diffusion ($t = 0$).

Fig. 20.1 Imaging principle for ALCL filled with diffusion solution. A RI gradient distribution of the filled liquid is formed along Z-axis, $n_1 < n_2 = n_c < n_3 < n_4$



According to the geometrical optics theory, the back focal length f_i of the ALCL filled with liquid is a function of the liquid RI (n_i), which can be written as the following iterative relations,

$$f_i = \frac{R_4 S_4}{n_0 R_4 + (n_0 - 1) S_4}, \quad (20.1a)$$

$$S_4 = \frac{n_0 R_3 S_3}{n_i R_3 + (n_i - n_0) S_3} - d_3, \quad (20.1b)$$

$$S_3 = \frac{n_i R_2 S_2}{n_0 R_2 + (n_i - n_0) S_2} - d_2, \quad (20.1c)$$

$$S_2 = \frac{n_0 R_1}{n_0 - 1} - d_1. \quad (20.1d)$$

where the S_i means the object distance of the i th refracting surface of the ALCL. The CCD is set at the position between the two focal planes along the Y -axis; that is, the ALCL is filled with EG and water respectively. According to Eq. (20.1), the back focal length f_i is defined in terms of the liquid RI (n_i). Therefore, the collimated light beams can only image clearly for only a thin liquid layer with an explicit RI value (say $n = n_c$) if dynamic gradient distribution of concentration and RI is formed gradually along the Z -axis as diffusion of the two liquids. While the ALCL filled with liquid of RI = $n_i > n_c$ (or $n_i < n_c$), the collimated beams shall project a diffuse spot (short red line marked in Fig. 20.1) on the CCD plane. So, a “beam waist” pattern shown as the right part of Fig. 20.1 will appear on the CCD chip. Since the dynamic of diffusion, the “waist” of the diffraction pattern will drift slowly along the Z -axis, and the drift rate reflects the diffusion speed vividly.

20.2.2 The Calculations Theory for Diffusion Coefficient

Assuming that the binary substances involved in diffusion are A and B, the mass fraction of A in B at diffusion time t and position Z expressed as $C(Z, t)$ and the diffusion direction is Z -axis, the diffusion coefficient is D , then the $C(Z, t)$ satisfies the following Fick’s second law:

$$\frac{\partial C(Z, t)}{\partial t} = D \frac{\partial^2 C(Z, t)}{\partial Z^2}. \quad (20.2)$$

Assuming the initial concentration of the lower solution to be C_1 and the upper solution concentration to be C_2 at $t = 0$, according to the initial and boundary conditions, the solution of Eq. (20.2) can be expressed as an Gauss error function

$$C(Z, t) = \frac{C_1 + C_2}{2} + \frac{C_1 - C_2}{2} \operatorname{erf}\left(\frac{Z}{2\sqrt{Dt}}\right). \quad (20.3)$$

Specifically, $\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u \exp(-t^2) dt$. Take use of the inverse error function, Eq. (20.3) can be rewritten as

$$Z = 2\sqrt{D} \operatorname{erfinv}\left\{\left[\frac{g[n(Z, t)] - \frac{C_1 + C_2}{2}}{\left[\frac{C_1 - C_2}{2}\right]}\right]\right\} \cdot \sqrt{t} - \Delta Z_0. \quad (20.4)$$

Here, $g[n(Z, t)] = C(Z, t)$ reflects the relationship of solution concentration and its RI, which can be pre-determined experimentally. ΔZ_0 is a fixed difference between the measured and actual values mainly caused by the undulate of the contact interface due to the intermolecular attraction.

When the distance (f_c) between CCD and ALCL is fixed, the RI of the only one thin liquid layer can converge collimated light into a focal point (the “waist” of the diffraction image) on the CCD is a fixed value n_c based on Eq. (20.1) and the imaging principle. The location (Z_i) of the “waist” varies with diffusion time (t_i), and the drift velocity of the “waist” reflects the diffusion rate intuitively. Fit linearly Z_i and $\sqrt{t_i}$ values and compare it with Eq. (20.4), the D value can be worked out easily as $[g[n(Z, t)] - (C_1 + C_2)/2] / [(C_1 - C_2)/2]$ is a constant. This method is named as “equivalent RI of thin liquid layer method.”

20.2.3 The Simulation Principle

According to the matrix optics theory, the optical property of paraxial optical system can be described as a 2×2 matrix with four elements $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ [21]. Ignoring the diffraction limitation, combining the matrix optics with Fresnel diffraction integral, Collins derived a convenient operation formula to solve diffraction problems of axisymmetric paraxial optical system, known as [15]

$$U(x, z) = \frac{\exp(jky)}{j\lambda B} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(x_0, z_0) \times \exp\left\{\frac{jk}{2B} [A(x_0^2 + z_0^2) + D(x^2 + z^2) - 2(x_0x + z_0z)]\right\} dx_0 dz_0 \quad (20.5)$$

where $j = \sqrt{-1}$, $k = 2\pi/\lambda$, λ is the light wavelength, y is the optical path length along propagation axis, A , B and D are the elements of the ray transfer matrix of the optical system. $U_0(x_0, z_0)$ and $U(x, z)$ are the input complex amplitude and output one respectively, marked as in Fig. 20.1.

Put irrelative terms out of the integral sign, Eq. (20.5) can be rewritten as

$$\begin{aligned}
 U(x, z) &= \frac{\exp(jky)}{j\lambda B} \exp\left\{\frac{jk}{2B}D(x^2 + z^2)\right\} \\
 &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{U_0(x_0, z_0) \exp\left[\frac{jk}{2B}A(x_0^2 + z_0^2)\right]\right\} \\
 &\times \exp\left[-j2\pi\left(x_0\frac{x}{\lambda B} + z_0\frac{z}{\lambda B}\right)\right] dx_0 dz_0 \tag{20.6}
 \end{aligned}$$

Equation (6) implies that the Collins diffraction integral can be reckoned as the product of a quadratic phase factor multiplied by the Fourier transform of the input complex amplitude times another quadratic phase factor. There are very limited functions can get analytical expression of Fourier transform, so FFT is usually used to calculate Eq. (20.6).

Let L_0 and L be the spatial dimensions on input and output planes, respectively, set sampling number be $N \times N$. Thus, the FFT form of Eq. (20.6) can be written as

$$\begin{aligned}
 U(p\Delta x, q\Delta z) &= \frac{\exp(jky)}{j\lambda B} \exp\left\{\frac{jk}{2B}D((p\Delta x)^2 + (q\Delta z)^2)\right\} \\
 &\times \text{FFT}\left\{U_0(mx_0, nz_0) \exp\left[\frac{jk}{2B}A((m\Delta x_0)^2 + (n\Delta z_0)^2)\right]\right\}_{\substack{p\Delta x, q\Delta z \\ \lambda B, \lambda B}}, \\
 &\left(p, q, m, n = -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2} - 1\right). \tag{20.7}
 \end{aligned}$$

where $\Delta x_0 = \Delta z_0 = L_0/N, \Delta x = \Delta z = L/N$ is sampling spacing during discrete calculation. Based on Eq. (20.7), we can work out output complex amplitude $U(p\Delta x, q\Delta z)$ using FFT only one time; therefore, Eq. (20.7) is referred to the S-FFT form of Collins diffraction integral.

For the cylindrical lens, a 4×4 matrix is needed to describe all the information of it, which leads to the form of Collins formula be more complex than Eqs. (5)–(7). In this paper, ALCL is divided into plenty of section along Z-axis. In this case, distance and included angle between light ray and propagation direction two parameters are enough to trace the light, namely 2×2 matrix similar to axisymmetric lens is enough. Thus, the ray-tracing matrix form for our experimental setup will be

$$\begin{aligned}
 \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} 1 & y_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1-n_0}{R_4} & n_0 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{n_0-n}{n_0 R_3} & \frac{n}{n_0} \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \\
 &\cdot \begin{bmatrix} 1 & 0 \\ \frac{n_0-n}{n R_2} & \frac{n}{n_0} \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1-n_0}{n_0 R_1} & \frac{1}{n_0} \end{bmatrix} \begin{bmatrix} 1 & y_1 \\ 0 & 1 \end{bmatrix} \tag{20.8}
 \end{aligned}$$

where $y_2 (= f_c)$ means the distance between CCD and the vertex of the fourth refraction surface of ALCL; y_1 means the distance between the vertex of the first refraction surface of ALCL and the silt; n is the liquid RI. In terms of Eq. (20.8),

numerical solutions of matrix elements A, B, C, D can be obtained easily. Substituting them into Eq. (20.7) and considering the input light as equal-amplitude parallel light, we can simulate out the numerical expression of complex amplitude on CCD planes.

20.3 Observation and Simulation Results

20.3.1 ALCL Filled with Uniform Liquid

All the experiments were carried out in a constant temperature environment with a fixed temperature 298.15 K. Adjust the optical path and set the slit width be 7 mm. Filling ALCL with distilled pure water (RI = 1.3334, measured by Abbe refractometer) and moving the translation stage from far to near gradually, we can observe that the diffraction images recorded by CCD change with the distance (y_2) between the CCD and the vertex of the fourth refraction surface of the ALCL, as shown in Fig. 20.2, and their corresponding normalized one-dimensional light intensity distribution curves are shown in Fig. 20.3.

The RI n is a constant value along Z -axis in this case. Matrix elements A, B, C, D calculated based on Eq. (20.8) are all equal for any thin liquid layer. Combined calculated matrix elements with Eq. (20.7), a series of diffraction images at different observed surfaces with uniform width along Z -axis are simulated using MATLAB software (see Fig. 20.4). Their normalized intensity distributions can be obtained at the same time as shown on Fig. 20.5. The simulation results are similar with the experimental ones, which indicate that Collins formula is applicable to model the propagation of collimation light passing through ALCL filled with liquid.

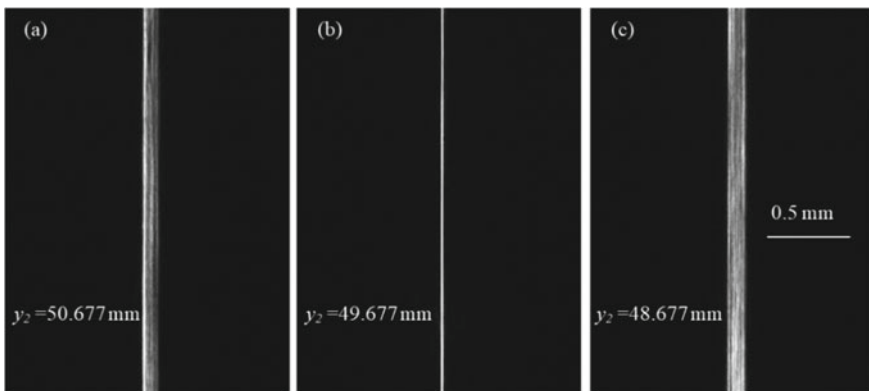


Fig. 20.2 Images recorded by CCD at different distances (y_2) between the CCD and the vertex of the fourth refraction surface of the ALCL filled with distilled pure water. **a** $y_2 = 50.677$ mm; **b** $y_2 = 49.677$ mm, equal with the back focal length calculated by Eq. (20.1); **c** $y_2 = 48.677$ mm

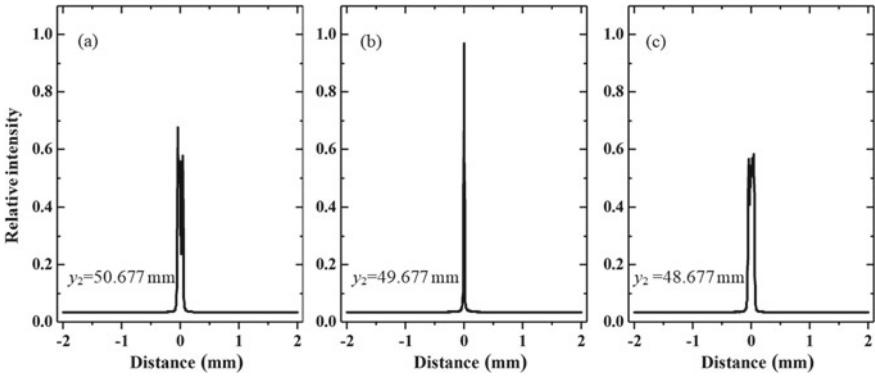


Fig. 20.3 Normalized one-dimensional light intensity distribution curves corresponding to Fig. 20.2

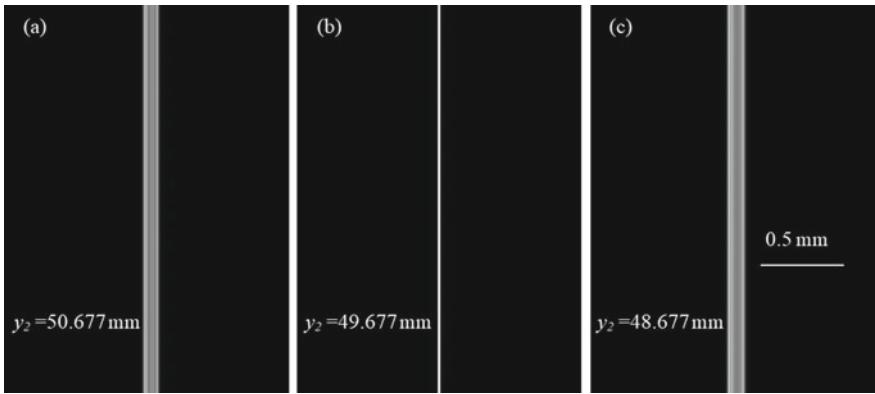


Fig. 20.4 Simulation images at different distances (y_2) between the CCD and the vertex of the fourth refraction surface of the ALCL filled with distilled pure water. **a** $y_2 = 50.677$ mm; **b** $y_2 = 49.677$ mm, equal with the back focal length calculated by Eq. (20.1); **c** $y_2 = 48.677$ mm

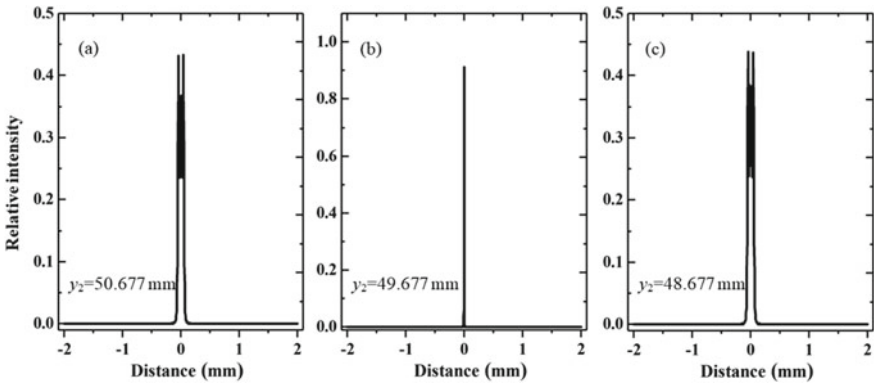


Fig. 20.5 Normalized one-dimensional light intensity distribution curves corresponding to Fig. 20.4

20.3.2 ALCL Filled with Diffusion Liquid

Good linear relationship $C = g(n) = 9.8478n - 13.130$ between the EG aqueous solutions concentrations and its RI values was achieved by measuring the RI of eight groups of EG with different concentrations using an Abbe refractometer. The pure EG and deionized water were injected into the ALCL sequentially as described in Sect. 20.2, constituting a diffusion system. Adjust the translation stage to make the CCD plane overlap with the focal plane of the ALCL filled with the liquid of $RI = n_c = 1.3391$, which means the thin liquid layer with the concentration of 5.72% related to the “waist” of the diffusion images. Keep the experiment device fixing, a series of diffusion images appeared on CCD successively, as shown in Fig. 20.6. The “waist” position (Z_i) of every diffraction image was listed in Table 20.1, and its offsets for every 5 min are marked on Fig. 20.6.

In Table 20.1, N_i means the number of pixels between the diffusion interface and the “waist.” Z_i is equal to the pixel size ($3.45 \mu\text{m}$) multiplying with N_i . Fit linearly Z_i and $\sqrt{t_i}$ using the least square method, thus

$$Z_i = 74.7629\sqrt{t_i} + 85.44 \text{ (}\mu\text{m)}, \tag{20.9}$$

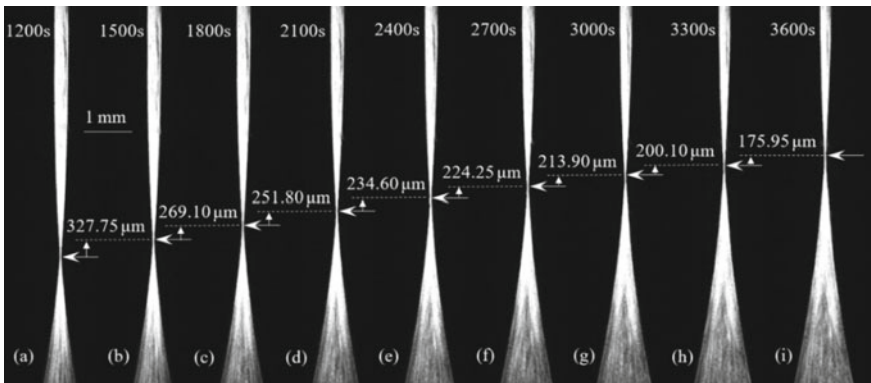


Fig. 20.6 A group of diffraction images recorded by CCD at different diffusion time t_i for EG diffusing in deionized water. The measured offsets of the “waist” for every 300 s are noted between every two images

Table 20.1 Data of positions Z_i varied with diffusion time

t_i/s	1200	1500	1800	2100	2400	2700	3000	3300	3600
$\sqrt{t_i}$	34.6410	38.7298	42.4264	45.8258	48.9898	51.9615	54.7723	57.4456	60.0000
N_i	772	867	945	1018	1086	1151	1213	1271	1322
$Z_i/\mu\text{m}$	2663.40	2991.15	3260.25	3512.10	3746.70	3970.95	4184.85	4384.95	4560.90

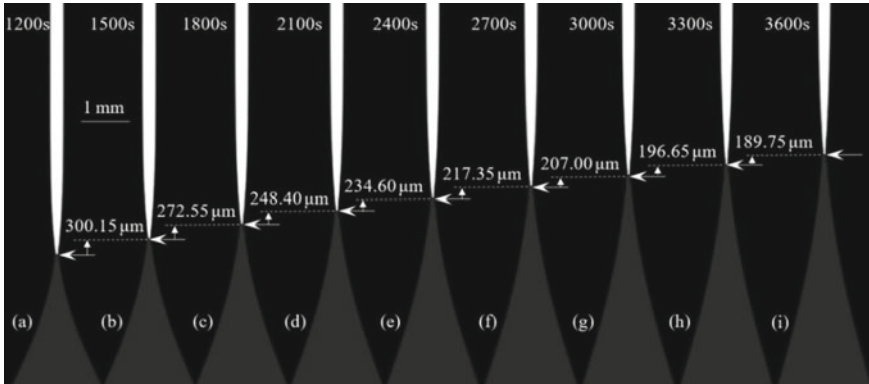


Fig. 20.7 A group of diffraction images simulated by MATLAB at different diffusion time t_i for EG diffusing in deionized water. The offsets of the “waist” for every 300 s are noted between every two images

The linear correlation coefficient in fitting Eq. (20.9) is up to 0.999. Then, the $D = 1.1212 \times 10^{-5} \text{ cm}^2/\text{s}$ is calculated out by comparing Eq. (20.9) with Eq. (20.4).

During simulation, diffusion coefficient of EG in water is considered as a constant $1.1212 \times 10^{-5} \text{ cm}^2/\text{s}$. It is easy to work out the concentration C_i of EG in water relate to any thin liquid layer with position Z_i at any given moment t_i based on Eq. (20.3). A series of n_i can be got based on the function $C = g(n) = 9.8478n - 13.130$. Taking n_i into Eq. (20.8), we figure out a sequence of matrix elements A_i, B_i, C_i, D_i . A group of “beam waist” diffraction patterns along Z -axis are simulated at different diffusion time t_i based on Eq. (20.7), as shown in Fig. 20.7. The “waist” drifts along Z -axis, and the offsets of “waist” every 300 s simulated by MZTAB software are marked on Fig. 20.7.

Comparing synthetic (Fig. 20.7) and real images (Fig. 20.6), the result showed that: (1) the “waists” of those two group images, corresponding to the 1.3391 RI thin liquid layer in the diffusion system, drifted in a fairly close velocity; (2) the images were similar in shape, and the outlines of the part “waist” above (shown as white) were coincident; (3) the outlines of the part below the “waist” (shown as gray) were anamorphic in a large scale.

The first and second points indicate that the thought of simulation on the intensity distribution of the collimated light beam refracted by an ALCL filled with diffusion solution is right. Especially, the first point demonstrates that the way to calculate D values is valid by linear fitting the positions (Z_i) of a specific thin liquid layer and the diffusion time ($\sqrt{t_i}$). Large deviation mentioned in point three is caused by that we consider the D value to be constant all the time. Actually, D value is a function of controlled inputs such as temperature, intensity of pressure, diffusing substance category, concentration, and so on. Only if the concentration gradient is very low, D value can be viewed as constant and Eq. (20.2) is workable, which is the precondition for calculation and simulation. It is obvious that, the closer to the interface of water and EG, the higher concentration gradient of EG aqueous solution,

the further from the application conditions of Eq. (20.2), and the bigger deviation between the simulation results and the real images when we consider the D value as constant during the simulation in our experiments. So, the aberration of simulation images reflected the applicable conditions of Eq. (20.2) from one side.

20.4 Conclusion

This paper analyzed the cylindrical lens layer by layer, which simplified four-dimensional matrix describing cylindrical lens into two-dimensional, making Collins formula appear in its simplest form to study of diffraction from cylinder lens. Based on that, the simulation on the CCD received diffraction images, projected by the collimated light passing through the ALCL filled with mutual diffusing solutions was realized for the first time. The resemblances between the simulation and the experimental images indicate that it is feasible to measure D values by using “equivalent RI of thin liquid layer method” based on ALCL. However, the difference between the two group images implies the application condition of Eq. (20.2) from a side. The work expressed in this paper deepened the understanding of SDT, diffusion process and Fick’s second law, which paved a way for further research on diffusion phenomenon.

Acknowledgements Natural National Science Foundation of China (61705192, 62065019); Yunnan Fundamental Research Projects (2017FD069, 2018FD014); Post-Doctoral Directional Trained Foundation of Yun Nan Province; Doctoral Scientific Research Foundation of YNNU; Core Curriculum Foundation of YNNU (YH2018-C04).

References

1. Crank, J.: *The Mathematics of Diffusion*, 2nd edn. Oxford University Press, Oxford (1975)
2. Cussler, E.L.: *Diffusion: Mass Transfer in Fluid Systems*, 3rd edn. Cambridge University Press, London (2009)
3. Chip, J.S., Sascha, G., Oldamur, H., Durgesh, V.W., Mark, P.H., Gary, A.B.: NMR relaxometric probing of ionic liquid dynamics and diffusion under mesoscopic confinement within bacterial cellulose ionogels. *J. Chem. Phys.* **148**, 193845 (2018)
4. Karol, M., Robert, H., Tomasz, K., Piotr, G., John, F.B.: Diffusion and flow in complex liquids. *Soft Matter* **16**(1), 114–124 (2020)
5. Kong, C.Y., Sugiura, K., Natsume, S., Sakabe, J., Funazukuri, T., Miyake, K., Okajima, I., Badhulika, S., Sako, T.: Measurements and correlation of diffusion coefficients of ibuprofen in both liquid and supercritical fluids. *J. Supercrit. Fluids* **159**, 104776 (2020)
6. Lang, S., Kazdal, T.J., Kühl, F., Hampe, M.J.: Diffusion coefficients and VLE data of aqueous phosphoric acid. *J. Chem. Thermodyn.* **68**, 75–81 (2014)
7. Chan, T.C., Tang, W.K.: Diffusion of aromatic compounds in nonaqueous solvents: a study of solute, solvent, and temperature dependences. *J. Chem. Phys.* **138**, 224503 (2013)
8. Culbertson, C.T., Jacobson, S.C., Ramsey, J.M.: Diffusion coefficient measurements in microfluidic devices. *Talanta* **56**, 365–373 (2002)

9. He, M.G., Zhang, S., Zhang, Y., Peng, S.G.: Development of measuring diffusion coefficients by digital holographic interferometry in transparent liquid mixtures. *Opt. Express* **23**(9), 10884 (2015)
10. Zhang, Y., Zheng, Y., He, M.G.: Improvement of dynamic light scattering method for measurement of particle diameter and liquid viscosity. *Acta Phys. Sinica* **67**(16), 167801 (2018)
11. Sun, L.C., Du, C., Li, Q., Pu, X.Y.: Asymmetric liquid-core cylindrical lens used to measure liquid diffusion coefficient. *Appl. Opt.* **55**(8), 2011–2017 (2016)
12. Sun, L.C., Meng, W.D., Pu, X.Y.: New method to measure liquid diffusivity by analyzing an instantaneous diffusion image. *Opt. Express* **23**(18), 23155–23166 (2015)
13. Sun, L.C., Pu, X.Y.: A novel visualization technique for measuring liquid diffusion coefficient based on asymmetric liquid-core cylindrical lens. *Sci. Rep.* **6**(1), 28264 (2016)
14. Goodman, J.W.: *Introduction to Fourier Optics*, 3rd edn. McGraw-Hill, New York (1996)
15. Collins, S.A.: Laser-system diffraction integral written in terms of matrix optics. *J. Opt. Soc. Am.* **60**(9), 1168–1177 (1970)
16. Chen, L.F., Zhao, D.M., Mao, H.D., Ge, F., Guan, R.X.: The Collins formula applied in optical image encryption. *Chin. Phys. Lett.* **30**(4), 044201 (2013)
17. Sziklas, E.A., Siegman, A.E.: Diffraction calculations using fast Fourier transform methods. *Proc. IEEE* **62**(3), 410–412 (1974)
18. Brigham, E.: *The Fast Fourier Transform and its Applications*, 1st edn. Prentice Hall, Englewood Cliffs, NJ (1988)
19. Shimobaba, T., Kakue, T., Oikawa, M., Okada, N., Endo, Y., Hirayama, R., Ito, T.: Non-uniform sampled scalar diffraction calculation using non-uniform fast Fourier transform. *Opt. Lett.* **38**(23), 5130–5133 (2013)
20. Sando, Y., Barada, D., Jackin, B.J., Yatagai, T.: Fast calculation method for computer-generated cylindrical holograms based on the three-dimensional Fourier spectrum. *Opt. Lett.* **38**(23), 5172–5175 (2013)
21. Wang, S., Zhao, D.: *Matrix Optics*, 1st edn. China Higher Education Press, Beijing (2000)

Chapter 21

Multidimensional Graphic Objects Filtration Using HoSVD Tensor Decomposition



Rumen Mironov and Ivo Draganov

Abstract A new approach for multidimensional graphic objects filtration using HoSVD tensor decomposition is presented. The experimental studies were performed on a set of test 3D images of size of $100 \times 100 \times 100$, containing simple geometric objects—a sphere, a cylinder, a cone, etc. After that, Gaussian noise with different variation is added and the low-frequency part of the decomposition matrices U and S is filtered. The results obtained of the filtered images show that the quality of the restored images in the different planes and in total for the 3D image is excellent. The peak signal-to-noise ratio for different samples ranged from 28 to 50 dB.

21.1 Introduction

Digital image processing has a very rich history dating back to the mid-1960s. It allows the use of many different algorithms to improve quality, reduce information redundancy, image transformation, segmentation and many others in various fields of communications, computer graphics, economics, transport, agriculture, remote monitoring, etc. Image filtering is one of the most important and widely used image processing operations.

Recently, techniques based on image transforms, especially in connection with multidimensional processing, have gained popularity and success in terms of productivity. The HoSVD is a summary of the SVD transform in the multidimensional space. Many innovative and successful applications of HoSVD for digital image processing have been shown in the literature [1–3].

R. Mironov (✉) · I. Draganov
Faculty of Telecommunications, Technical University of Sofia, Boul. Kl. Ohridsky 8, 1000 Sofia,
Bulgaria
e-mail: rmironov@tu-sofia.bg

I. Draganov
e-mail: idraganov@tu-sofia.bg

A wide variety of techniques is already developed for noise reduction from multi-dimensional images: partial differential equations (PDEs), spatially varying convolution and regression, non-local techniques, transform-based techniques and methods based on machine learning [2, 3]. **PDE-based** methods process anisotropically noisy images. This technique extracts and respects the edge geometry, allowing diffusion along but not across the image edges [4, 5]. Some PDEs are obtained from the Euler-Lagrange equations corresponding to functionals that are based on a piecewise constant [6] or piecewise linear [7] model for natural images. In practice, the energy functionals are augmented with prior terms that reduce the error between the noisy and filtered image, as per the assumed noise model [6]. In **spatially varying convolution methods**, an image is convolved with a pointwise-varying, local geometry-driven mask [8] or the local modeling of an image with low-order polynomial function coefficients, computed by a weighted least-squares regression. These weights are chosen based on estimates of local geometry [9] or the difference in intensity/spatial coordinate values between neighboring pixels and the one to be filtered [10, 11]. The most recent advance in the area of local convolutions is presented in the work [12].

In **transform-domain** denoising approaches, the usually small image area is projected onto an orthonormal basis, such as a wavelet [13] or DCT [14] to yield a set of decorrelated coefficients [15]. The smaller coefficients usually correspond to the higher frequency components of the signal which are often dominated by noise. To perform denoising, the smaller coefficients are modified (typically, by “hard thresholding”), and the area is reconstructed by inversion of the transform. If the selected areas are not overlapping, artifacts are observed around their borders and around the edges of the image, which can be reduced by using a sliding window and averaging the multiple hypotheses [13, 15].

In **non-local techniques** (NL) [2, 16], the fact that natural images often contain areas in distant regions that are very similar to each other is used. NL-Means obtains a de-noised image by minimizing a penalty term on the average weighted distance between an image area and all other areas in the image, where the weights are decreasing functions of the squared difference between the intensity values in the areas. This yields an update rule that can be interpreted as a spatially varying convolution with non-locally derived masks [17, 18].

The block matching in three dimensions (BM3D) method [19] is a **combination of non-local** and **transform-domain** approaches which is considered the current state of the art in image denoising. This method operates at the small areas, and for each reference area in the image, it collects a group of similar areas (after a DCT-based pre-filtering step), which are then stacked together to form a 3D array. The entire 3D array is projected onto a 3D transform basis (product of DCT/biorthogonal and Haar bases) to yield a set of coefficients which are hard-thresholded. The filtered areas are then reconstructed by inversion of the transform. This image is then smoothed (heuristically) with a non-local empirical Wiener filter to produce a final filtered image.

There are also methods based on learning the transform basis from the statistics of image features. There has been recent interest in learning over complete bases

(also called dictionaries) [20, 18], whose inherent redundancy leads to sparser representation of natural signals. A multi-scale variant of the KSVD algorithm (known as MS-KSVD) learns dictionaries to represent the areas at two or more scales leading to further redundancy [21]. This algorithm has yielded state-of-the-art performance, on par with the BM3D algorithm [19]. The results obtained from methods, given in [19, 21–23] are very good, but in turn they are quite complex with several selectable parameters such as choice of bases, patch-size, transform thresholds, and similarity measures.

In this paper, we propose a very simple and effective algorithm for 3D graphical image filtration using HoSVD tensor decomposition. The experimental studies were performed on a set of test 3D graphics images, containing simple geometric objects—a sphere, a cylinder, a cone, etc.

21.2 Mathematical Description

A fundamental concept of data decomposition is to analyze, extract, and preserve the most relevant data features and remove redundant and irrelevant information. Data reduction linked to concepts and techniques of data compression, noise reduction as well as feature extraction and recognition. Computed bases, learned from the input data, may provide a better data fit, approximation and fast reconstruction of the input data. These basic decompositions are SVD, PCA, and tensor decomposition [1].

The SVD transformation is well-known for analyzing and filtration 2D matrices and can be used for processing of high dimensional arrays (N —dimensional arrays) by extending the SVD analyses with tensor theory. The tensor decomposition of N th-order tensor $A \in R^{I_1 \times I_2 \times I_3 \times \dots \times I_N}$ having N indices is:

$$A = S \times_1 U^{(1)} \times_2 U^{(2)} \times_3 \dots \times_N U^{(N)} = S \times_{n=1}^N U^{(n)} \quad (21.1)$$

where $I_1 \times I_2 \times I_3 \times \dots \times I_N$ are the dimensions of the tensor A . The tensors can be decomposed into slices and fibers by fixing all indices except one or two. A tensor slice is a 2D matrix obtained by fixing one of the three indices in a 3rd-order tensor.

There are different types of tensor decompositions as canonical decomposition (or parallel factor analysis model—CP), Tucker model, Tensor-Trains, etc. The higher-order SVD (HoSVD) is a special case of Tucker decomposition obtained by adding an orthogonality constraint to the component matrices [1, 2]. In the decomposition Eq. (21.1), the factor matrices, $U^{(i)}$'s, are the left singular vectors of each flattening $A^{(i)}$. Rearranging elements of a tensor into a 2D matrix is known as unfolding or matricization. The HoSVD is formed from the SVDs of all the tensor modal unfoldings.

For the purpose of the presented studies, a 3D tensor A is used, formed by three-dimensional image data, as shown in Fig. 21.1.

Using HoSVD, the 3D tensor A is transformed into 2D matrices. The core tensor S is the equivalent of the diagonal matrix in the 2D matrix SVD, except that it is

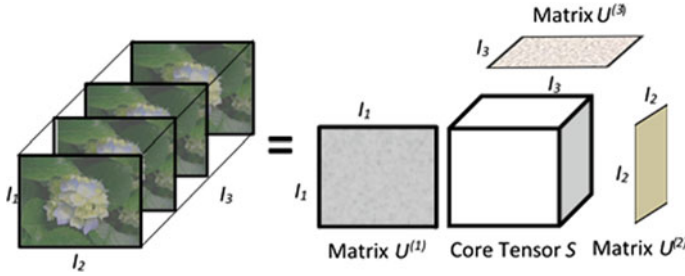


Fig. 21.1 Tensor decomposition of third-order tensor $A \in R^{I_1 \times I_2 \times I_3}$ into a core tensor $S \in R^{I_1 \times I_2 \times I_3}$ and unitary matrices— $U^{(1)} \in R^{I_1 \times I_1}$, $U^{(2)} \in R^{I_2 \times I_2}$, and $U^{(3)} \in R^{I_3 \times I_3}$

pseudo-diagonal. There is no strict equivalent to singular values for the matrices in the case of three-dimensional arrays, but in spite of this, the ordered norms of the submatrices of S can be seen as the equivalent of the singular values; moreover, the energy of the three-dimensional array A is concentrated at the (1,1,1) corner of the core array S .

We use the information from the matrices of the core tensor S to perform the noise analysis in the 3D data. As SVD is used for 2D matrices, the HoSVD computes a core tensor from a 3D signal tensor, and this core tensor can be manipulated for instance through a threshold transforms. Since the HoSVD is easily invertible, the modified core tensor permits to reconstruct an approximate tensor which is the de-noised tensor of the 3D signal.

To perform the tensor decomposition, we used MATLAB tensor tools [24]. The input 3D data for the decomposition are of size $100 \times 100 \times 100$ and include simple geometric objects—a sphere, a cylinder, a cone, etc.—as is shown in Fig. 21.2 in 3D view.

After the decomposition of the input 3D tensor A , the 3 projections of the core tensor S are shown in Fig. 21.3. Here in Fig. 21.3a, b, and c, the decompositions of the core tensor in horizontal, vertical, and depth directions— S^1 , S^2 , and S^3 —are shown, respectively.

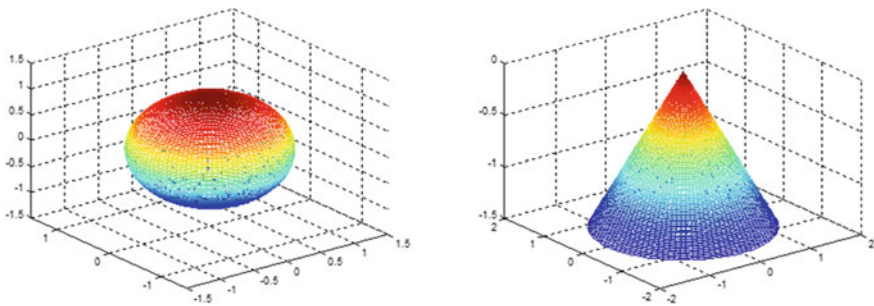


Fig. 21.2 Input 3D sample geometric objects—sphere and cone—of size $100 \times 100 \times 100$

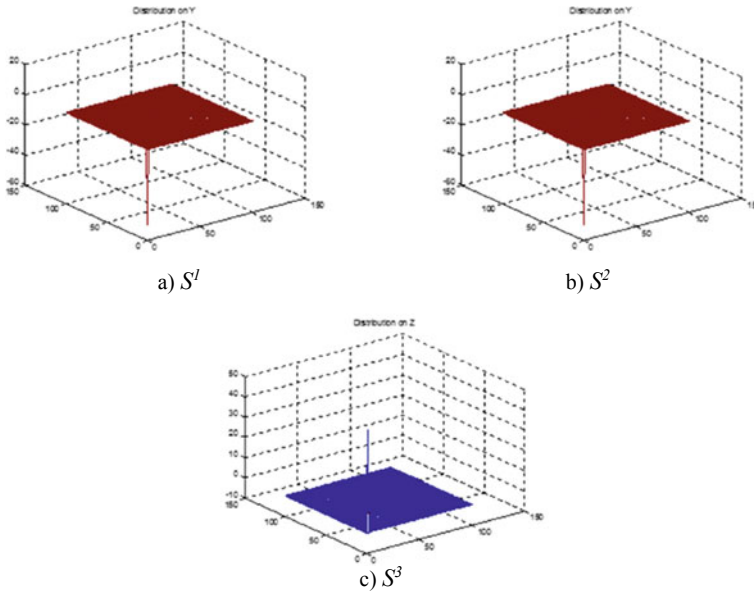


Fig. 21.3 Projections S^1 , S^2 , and S^3 of the core tensor S after HoSVD decomposition. **a** S^1 , **b** S^2 , **c** S^3

Figure 21.4a, c, and e shows the same decompositions as the values of the first coefficients of the transforms in the window 4×4 being suppressed by a factor of 100,000, and in Fig. 21.4b, d, and f, the projections of the corresponding major diagonals of the transforms are shown.

The graphs show that the first 4 coefficients of the decompositions (for all three projections) contain the main part of the energy of the input image data (their order is higher than 5×10^5). From Figs. 21.4b, d, and f, it can be seen that the rest of the energy is concentrated in the next 30 coefficients and then periodic peaks with small amplitude are observed, which are likely to meet the periodic noise distribution in the graphics data.

The central idea of our approach is to reduce the data noises as is done in the 2D SVD analysis. Let the tensor A represents the noisy signal, then compute the HoSVD, and after that discard small singular values of A in the respective projections of the core tensor S . It can be shown that the small singular values mainly represent the noise, and thus, the reconstructed rank- k matrix A^k represents a filtered signal with less noise.

To verify the assumptions, a noise reduction algorithm for multidimensional signals, using HoSVD tensor decomposition was developed. It comprises the following basic steps:

- construct tensor A from the input data planes X, Y, Z ;
- decompose A using HoSVD tensor decomposition;

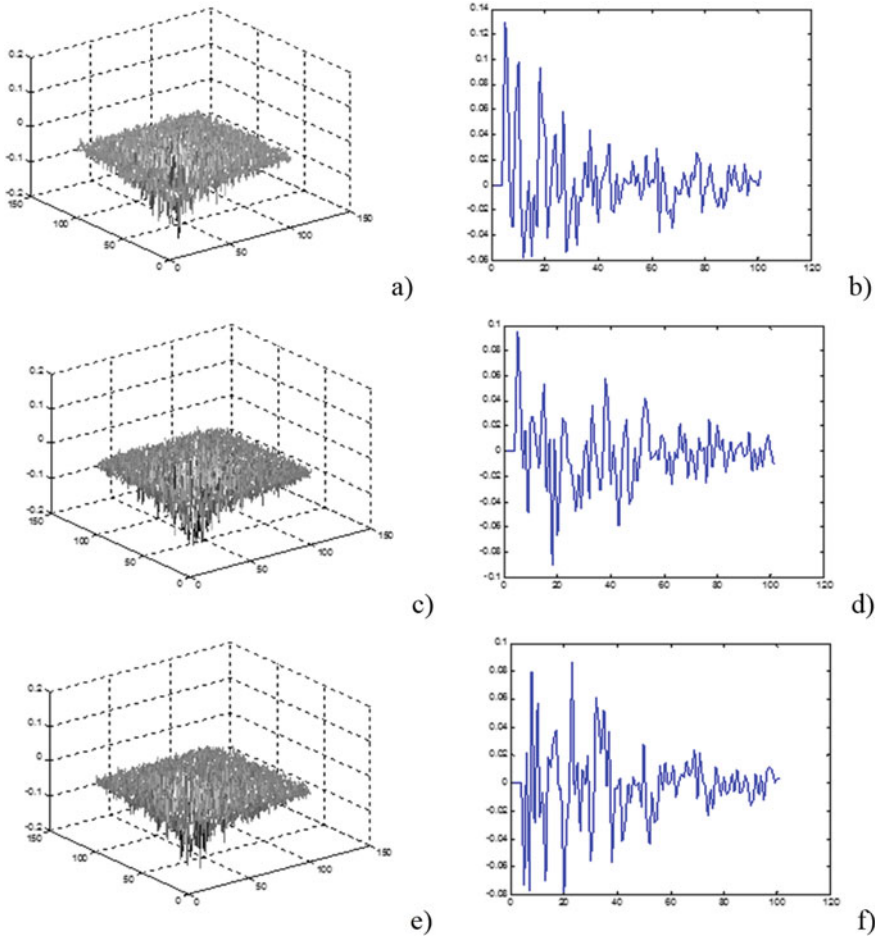


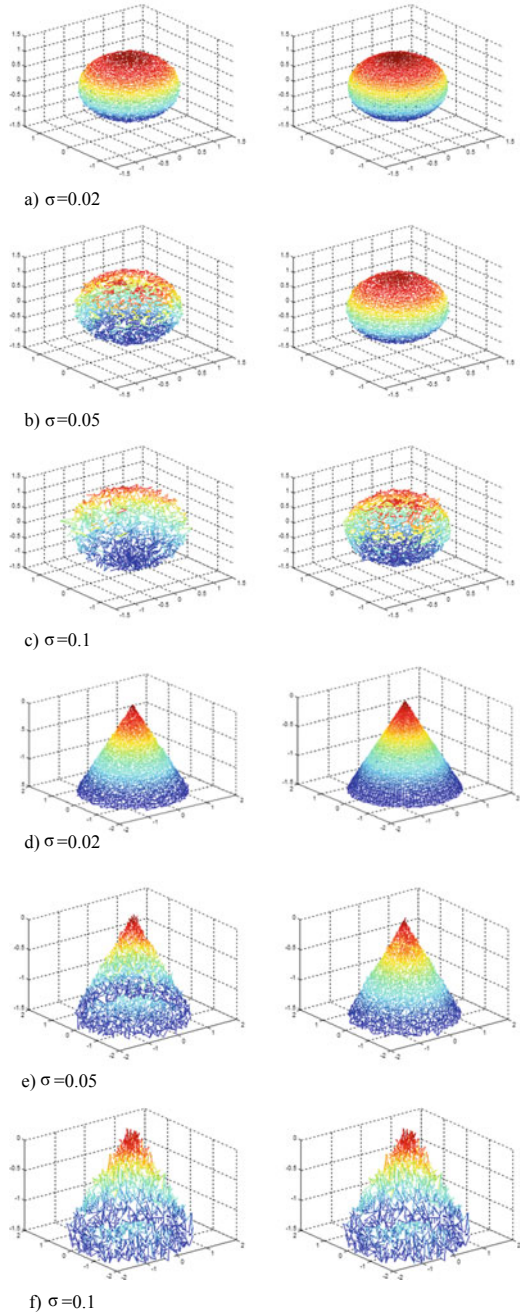
Fig. 21.4 Projections S^1 , S^2 , and S^3 of the core tensor S after suppression of the first coefficients of the transforms

- truncate planes S^1 and S^2 of the core tensor—keeping the first SVDs components in a window of size 30×30 ;
- filtrate plane S^3 with adaptive LMS filter [25] of size 4×4 and preserve the first SVD components in the window 4×4 ;
- compose tensor A using HoSVD.

21.3 Experimental Results

Initial verification was carried out by adding Gaussian noise with zero mean and a small variation of deviation σ from 0.01 to 0.1. Figure 21.5a–f is visual-

Fig. 21.5 Noised and filtrated parts of the surfaces of simple graphic figures (sphere and cone) for different variations of σ (0.02, 0.05, and 0.1). **a** $\sigma = 0.02$, **b** $\sigma = 0.05$, **c** $\sigma = 0.1$, **d** $\sigma = 0.02$, **e** $\sigma = 0.05$, **f** $\sigma = 0.1$



ized, respectively, the incoming noised part of the surfaces of simple graphic figures (sphere and cone), shown in Fig. 21.2, and the same part after the noise reduction for different values of σ . The numerical results of the calculation with the developed MATLAB program are presented in Table 21.1. The first column of the table is given the value of the deviation, in the second are defined the planes for which the error is calculated, and in the third to the fifth columns are given the values of the mean square error (MSE), the normalized mean square error (NMSE), the signal-to-noise ratio (SNR), and the peak signal-to-noise ratio (PSNR). In the last row, the same results for noiseless surfaces are given.

The results obtained show that the quality of the reconstituted signal in the different planes and in a total of the 3D signal are excellent. The average peak signal-to-noise ratio ranged from 26 to 50 dB. The best results are obtained with a 0.01 deviation. For these values, the noise is noticeably low and possibly resembles the noise of the surface scan. For larger values of the deviation, the noise predominates visually over the data and does not correspond to the character of the noise we are exploring. For smaller values of σ , the noise is practically invisible.

The results after processing the same part of the surfaces without adding noise shows that the values for SNR and PSNR are about 113.592 dB and 78.3683 dB, respectively.

21.4 Conclusion

A new approach for multidimensional image filtration using HoSVD tensor decomposition is presented. The experimental studies were performed on a set of test 3D images of size of $100 \times 100 \times 100$, containing simple geometric objects—a sphere, a cylinder, a cone, etc. After that, random noise with different dispersion is added on it and the low-frequency part of the decomposition matrices U and S is filtered. Since the HoSVD is easily invertible, the modified core tensor permits to reconstruct an approximate tensor which is the de-noised tensor of the original 3D image.

The results obtained of filtered images show that the quality of the restored images in the different planes and in a total of the 3D image, are excellent. The average peak signal-to-noise ratio for different samples is ranged from 28 to 50 dB (Tables 21.1 and 21.2).

Table 21.1 Filtration results of graphics figure sphere

	Deviation	Plane	MSE	NMSE	SNR, dB	PSNR, dB
1.	$\sigma = 0.01$	X	0.78856e-004	0.00031547	42.878	48.8989
		Y	3.46413e-006	0.00014135	48.497	54.6041
		Z	0.000027148	0.000151802	45.9894	48.9569
		XYZ	9.68817e-006	0.000190645	45.3664	50.1376
2.	$\sigma = 0.02$	X	1.28856e-004	0.00051547	32.878	38.8989
		Y	3.46413e-005	0.00014135	38.497	44.6041
		Z	0.000127148	0.000251802	35.9894	38.9569
		XYZ	9.68817e-005	0.000290645	35.3664	40.1376
3.	$\sigma = 0.03$	X	0.000276378	0.00110562	29.5639	35.585
		Y	7.16357e-005	0.00029230	35.3417	41.4487
		Z	0.000290176	0.000574662	32.4059	35.3734
		XYZ	0.00021273	0.00063819	31.9505	36.7217
4.	$\sigma = 0.04$	X	0.000485096	0.00194057	27.1207	33.1417
		Y	0.000131961	0.00053845	32.6885	38.7956
		Z	0.00052463	0.00103897	29.834	34.1957
		XYZ	0.000380562	0.00114169	29.4245	34.1957
5.	$\sigma = 0.05$	X	0.000801392	0.00320588	24.9405	30.9616
		Y	0.000196046	0.00079994	30.9694	37.0764
		Z	0.000779686	0.00154408	28.1133	31.0808
		XYZ	0.000592374	0.00177712	27.5028	32.274
6.	$\sigma = 0.06$	X	0.00123166	0.00492711	23.0741	29.0951
		Y	0.000290671	0.00118605	29.259	35.366
		Z	0.00115263	0.00228266	26.4156	29.3831
		XYZ	0.000891653	0.00267496	25.7268	30.498
7.	$\sigma = 0.07$	X	0.00148095	0.00592439	22.2736	28.2946
		Y	0.000423227	0.00172694	27.6272	33.7343
		Z	0.00162103	0.00321028	24.9346	27.9021
		XYZ	0.00117507	0.00352521	24.5282	29.2994
8.	$\sigma = 0.08$	X	0.00211623	0.00846576	20.7233	26.7444
		Y	0.00051796	0.00211348	26.75	32.857
		Z	0.00210109	0.00416099	23.808	26.7755
		XYZ	0.00157843	0.00473528	23.2465	28.0178
9.	$\sigma = 0.09$	X	0.00245682	0.00982826	20.0752	26.0963
		Y	0.000669938	0.00273362	25.6326	31.7397
		Z	0.00264597	0.00524006	22.8066	25.7741
		XYZ	0.00192424	0.00577273	22.3862	27.1574

(continued)

Table 21.1 (continued)

	Deviation	Plane	MSE	NMSE	SNR, dB	PSNR, dB
10.	$\sigma = 0.1$	X	0.00311407	0.0124575	19.0457	25.0667
		Y	0.00087796	0.00358243	24.4582	30.5653
		Z	0.00332735	0.00658946	21.8115	24.779
		XYZ	0.00243979	0.00731938	21.3553	26.1265

Table 21.2 Filtration results for graphics figure cone

	Deviation	Plane	MSE	NMSE	SNR, dB	PSNR, dB
1.	$\sigma = 0.01$	X	3.25552e-005	8.55225e-005	40.6792	46.6342
		Y	7.65009e-006	2.05028e-005	46.8819	52.9237
		Z	3.24116e-005	4.29983e-005	43.6655	46.6534
		XYZ	2.42057e-005	4.8168e-005	43.1724	47.9212
2.	$\sigma = 0.02$	X	0.000127806	0.000335747	34.7399	40.6948
		Y	3.00338e-005	8.04926e-005	40.9424	46.9843
		Z	0.000139163	0.000184619	37.3372	40.3251
		XYZ	9.90012e-005	0.000197007	37.0552	41.804
3.	$\sigma = 0.03$	X	0.000286224	0.00075191	31.2383	37.1933
		Y	7.03314e-005	0.00018849	37.2471	43.2889
		Z	0.000308927	0.000409833	33.8739	36.8618
		XYZ	0.000221828	0.000441426	33.5514	38.3002
4.	$\sigma = 0.04$	X	0.00052866	0.00138879	28.5736	34.5286
		Y	0.000134658	0.00036089	34.4262	40.4681
		Z	0.000535382	0.000710256	31.4859	34.4737
		XYZ	0.000399567	0.000795118	30.9957	35.7445
5.	$\sigma = 0.05$	X	0.000843453	0.00221575	26.5448	32.4998
		Y	0.000191674	0.00051369	32.8929	38.9347
		Z	0.000842021	0.00111705	29.5193	32.5071
		XYZ	0.000625716	0.00124514	29.0478	33.7966
6.	$\sigma = 0.06$	X	0.00119847	0.00314837	25.0191	30.9741
		Y	0.000274066	0.000734517	31.34	37.3818
		Z	0.00113827	0.00151007	28.21	31.1979
		XYZ	0.000870269	0.00173179	27.615	32.3638
7.	$\sigma = 0.07$	X	0.00158135	0.00415419	23.8151	29.7701
		Y	0.000371395	0.000995363	30.0202	36.062

(continued)

Table 21.2 (continued)

	Deviation	Plane	MSE	NMSE	SNR, dB	PSNR, dB
		Z	0.0016655	0.00220951	26.557	29.5449
		XYZ	0.00120608	0.00240004	26.1978	30.9466
8.	$\sigma = 0.08$	X	0.00205176	0.00538998	22.6841	28.6391
		Y	0.000567533	0.00152103	28.1786	34.2205
		Z	0.00221929	0.00294418	25.3104	28.2982
		XYZ	0.00161286	0.00320952	24.9356	29.6844
9.	$\sigma = 0.09$	X	0.00262382	0.00689277	21.6161	27.571
		Y	0.000717359	0.00192257	27.1612	33.203
		Z	0.00266177	0.00353119	24.5208	27.5087
		XYZ	0.00200098	0.00398186	23.9991	28.7479
10.	$\sigma = 0.1$	X	0.00317324	0.00833609	20.7904	26.7453
		Y	0.000899241	0.00241003	26.1798	32.2216
		Z	0.00340352	0.00451523	23.4532	26.4411
		XYZ	0.002492	0.00495896	23.0461	27.7949

Acknowledgements This work was supported by the National Science Fund of Bulgaria: Project No. KP-06-H27/16 “Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas.”

References

- Özdemir, A., Iwen, M.A., Aviyente, S.: A multiscale approach for tensor denoising. In: IEEE Statistical Signal Processing Workshop (SSP), Palma de Mallorca, pp. 1–5. <https://doi.org/10.1109/ssp.2016.7551841>. IEEE (2016)
- Buades, A., Coll, B., Morel, J.-M.: A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **4**(2), 490–530 (2005)
- Rajwade, A., Rangarajan, A., Banerjee, A.: Image denoising using the higher order singular value decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(4), 849–862 (2013)
- Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.* **12**(7), 629–639 (1990)
- Weickert, J.: *Anisotropic diffusion in image processing*. Teubner, Stuttgart, Germany (1998)
- Rudin, L., Osher, S.: Total variation based image restoration with free local constraints. In: *IEEE International Conference Image Processing (ICIP)*, pp. 31–35 (1994)
- You, Y., Kaveh, M.: Fourth order partial differential equations for noise removal. *IEEE Trans. Image Process.* **9**(10), 1723–1730 (2000)
- Tschumperlé, D., Deriche, R.: Vector-valued image regularization with PDEs: a common framework for different applications. *IEEE Trans. Patt. Anal. Mach. Intell.* **27**(4), 506–517 (2005)
- Takeda, H., Farsiu, S., Milanfar, P.: Kernel regression for image processing and reconstruction. *IEEE Trans. Image Process.* **16**(2), 349–366 (2007)
- Buades, A., Coll, B., Morel, J.-M.: Neighborhood filters and PDEs. *Numer. Math.* **105**(1), 1–34 (2006)

11. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(5), 603–619 (2002)
12. Subakan, O., Jian, B., Vemuri, B., Vallejos, E.: Feature preserving image smoothing using a continuous mixture of tensors. *IEEE International Conference Computer Vision (ICCV)*, pp. 1–6 (2007)
13. Coifman, R., Donoho, D.: Translation-invariant denoising. Yale University, Tech. Rep. 1995
14. Yaroslavsky, L., Egiazarian, K., Astola, J.: Transform domain image restoration methods: review, comparison and interpretation. In: *SPIE Proceedings Series, Nonlinear Processing and Pattern Analysis*, pp. 1–15 (2001)
15. Hyvarinen, A., Hoyer, P., Oja, E.: Image denoising by sparse code shrinkage. *Intell. Sign. Process.* 1–6.28 (1999)
16. Zhang, D., Wang, Z.: Image information restoration based on long-range correlation. *IEEE Trans. Circuit Syst. Video Technol.* **12**(5), 331–341 (2002)
17. Awate, S., Whitaker, R.: Unsupervised, informationtheoretic, adaptive image filtering for image restoration. *IEEE Trans. Patt. Anal. Mach. Intell.* **28**(3), 364–376 (2006)
18. Popat, K., Picard, R.: Cluster-based probability model and its application to image and texture processing. *IEEE Trans. Image Process.* **6**(2), 268–284 (1997)
19. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007)
20. Aharon, M., Elad, M., Bruckstein, A.: The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006)
21. Mairal, J., Sapiro, G., Elad, M.: Learning multiscale sparse representations for image and video restoration. *Multiscale Model. Simul.* **7**(1), 214–241 (2008)
22. Zare, A., Ozdemir, A., Iwen, M., Aviyente, S.: Extension of PCA to higher order data structures: an introduction to tensors, tensor decompositions, and tensor PCA. *Proc. IEEE* **106**(8), 1341–1358 (2018)
23. Cui, J., Zhang, K., Ma, T.: An efficient algorithm for the tensor product model transformation. *Int. J. Control Autom. Syst.* **14**(5), 1205–1212 (2016)
24. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM review* **51**(3), 455–500 (2009)
25. Mironov, R.: Local adaptive image processing. New approaches in intelligent image analysis. In: Kountchev, R., Nakamatzu, K. (eds.) *Series intelligent systems reference library*, vol. 108, pp. 295–330. Springer International Publishing, Switzerland (2016)

Author Index

A

Aldaeif, Fatima, [99](#)

C

Chaudhary, Sheetal, [195](#)

Cheng, Zhang, [233](#)

D

Dalal, Ekta, [187](#)

Draganov, Ivo, [57](#), [255](#)

E

Elghoul, Sinda, [73](#)

F

Fang, Tian, [233](#)

Feng, Jie, [243](#)

G

Georgieva, Veska, [83](#)

Ghorbel, Faouzi, [73](#)

Gotseva, Daniela, [125](#)

I

Ismail, Raneem, [137](#)

Ivanov, Stefan, [207](#)

J

Juan, Yao, [233](#)

K

Kant, Chander, [195](#)

Kaur, Jasvinder, [165](#), [175](#)

Kountcheva, Roumiana, [35](#)

Kountchev, Roumen, [35](#)

L

Li, Linhai, [243](#)

Liu, Xinyi, [15](#)

Liu, Ya, [243](#)

M

Mei, Guo, [219](#), [225](#)

Milanova, Mariofanna, [15](#), [99](#)

Min, Xiao, [219](#), [225](#)

Mironov, Rumen, [57](#), [255](#)

Mitrev, Rosen, [125](#)

N

Nagy, Szilvia, [137](#)

O

O’Gorman, Lawrence, [15](#)

P

Petrov, Plamen, [83](#)

R

Rathee, Tripti, [153](#)

Rathi, Kavita, [113](#)

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*, Smart Innovation, Systems and Technologies 216,

<https://doi.org/10.1007/978-981-33-4676-5>

S

Salem, Abdel-Badeeh M., [1](#)
Sarker, Md Imran, [15](#)
Sharma, Bhawna, [165](#)
Sharma, Shivani, [175](#)
Sheng, Shuwu, [243](#)
Singh, Parvinder, [113](#), [153](#), [165](#), [175](#), [187](#),
[195](#)
Singh, Sukhdev, [195](#)
Stankov, Stanko, [207](#)
Stawczyk, Paweł, [207](#)

Sun, Licun, [243](#)

T

Tudjarov, Boris, [125](#)

W

Wang, Xu, [233](#)
Wang, Yuanfangzhou, [243](#)
Wilk-Jakubowski, Jacek, [207](#)