Smart Innovation, Systems and Technologies 270

Roumen Kountchev
Rumen Mironov
Kazumi Nakamatsu *Editors*

New Approaches for Multidimensional Signal Processing

Proceedings of International Workshop, NAMSP 2021

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 270

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 https://link.springer.com/bookseries/8767

Roumen Kountchev · Rumen Mironov ·
Kazumi Nakamatsu
Editors

New Approaches for Multidimensional Signal Processing

Proceedings of International Workshop, NAMSP 2021

Springer

Editors
Roumen Kountchev
Technical University of Sofia
Sofia, Bulgaria

Rumen Mironov
Technical University of Sofia
Sofia, Bulgaria

Kazumi Nakamatsu
University of Hyogo
Kobe, Japan

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-981-16-8557-6 ISBN 978-981-16-8558-3 (eBook)
https://doi.org/10.1007/978-981-16-8558-3

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
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

Organizing Committee

2nd Workshop on New Approaches for Multidimensional Signal Processing NAMSP 2021

Technical University of Sofia, Sofia, Bulgaria, July 08–10, 2021

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 & Technology, Murthal, India

International Advisory Chairs:

Prof. Kun Zhang Hainan Tropical Ocean 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:

SR. 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. B. Khan, Virginia International University, USA
Prof. P. Koprinkova-Hristova, Bulgarian Academy of Sciences, Bulgaria
SR. Dr. R. Kountcheva, T&K Engineering, Bulgaria
Prof. V. Georgieva, Technical University of Sofia, Bulgaria
Prof. Jair Abe, University of Sao Paulo, Brazil
Prof. A. Elmaghraby, University of Louisville, USA
Prof. I. Draganov, Technical University of Sofia, Bulgaria

Prof. O. 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. 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. N. Taleb, Djillali Liabes University of Sidi Bel Abbes, Algeria
Prof. N. Hikal, Mansoura University, Egypt
Prof. S. Nagy, Széchenyi István University, Gyor, Hungary
Prof. Br. Milovanović, University of Niš, Serbia
Prof. Z. Bojković, University of Belgrade, Serbia
Prof. M. Talmaciu, University of Bacau
Prof. E. Nechita, University of Bacau
Prof. A. Saro, Bharathiar University, Coimbatore, India
Prof. T. 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. At. Gotchev, Tampere University of Technology, Finland
Prof. B. Mirkin, Higher School of Economics University, Moscow, Russian Federation

Preface

This book presents the result of the 2nd International Workshop "New Approaches for Multidimensional Signal Processing-NAMSP 2021" which was carried out online, during July 8–10, 2021, at the Technical University of Sofia, Bulgaria. The part of the authors are the team members of the bilateral Bulgarian-Indian project KP-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 2021 are Interscience Research Network (IRNet) International Academy Communication Center, China; and Interscience Institute of Management and Technology-Bhubaneswar, India. In the workshop, the participated authors were from India, Bulgaria, China, Egypt, USA, and Bangladesh. 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 27 chapters, divided into the following 4 mean directions.

The Part I *Multidimensional Signal Processing* includes the Chaps. 1–8:

- Object Motion Detection in Video by Fusion of RPCA and NMF Decompositions;
- Masked Face Detection using Artificial Intelligent Techniques;
- Hierarchical Tensor Decomposition trough Adaptive Branched Inverse Difference Pyramid;

- Multimodal Technique for Human Authentication using Fusion of Palm and Dorsal Hand Veins;
- SIFT-based Feature Matching Algorithm for Cartoon Plagiarism Detection;
- Image Recognition-Based Evaluation Index of Ship Navigation Risk in Bridge Area;
- Equalization of Directional Multidimensional Histograms of Matrix and Tensor Images;
- Small Object Detection of Remote Sensing Images Based on Residual Branch of Feature Fusion.

The Part II *Deep Learning in Multidimensional Neural Networks* includes the Chaps. 9–15:

- Meta-Learning With Logistic Regression for Multi-Classification;
- Measurement for Blade Edge Based on Machine Vision;
- Deep Learning 3D Convolutional Neural Networks for Predicting Alzheimer's Disease;
- Dense Optical Flow and Residual Network-Based Human Activity Recognition;
- Density Calculation of Pseudo-breast MRI Based on Adversarial Generative Network;
- Machine Learning enabled Edge Computing: A Survey and Research Challenges;
- Application of Deep Learning in Maize Image Segmentation.

The Part III *Applications of Multidimensional Signal Processing* includes the Chaps. 16–21:

- Identifying People Wearing Masks in a 3D-Scene;
- Contrast Enhancement and Noise Removal From Medical Images Using a Hybrid Technique;
- Application of Computer Image Recognition Technology in Ship Monitoring Direction;
- Application of Remote Network Technology in Engine Room Communication of the Ship;
- Pi Test for Signal Complexity;
- A Method of Respiratory Monitoring based on Knowledge Graph.

The Part IV *New Approaches in Communications and Computer Technologies* includes the Chaps. 22–27:

- Communication Technology-Based Optimization of Ship Remote Control Data Management Platform;
- A Cognitive Radio Adaptive Communication Platform;
- Billing System and 5G Network Slicing Service;
- Location models for public healthcare facilities in India;
- Natural Language Query for Power Grid Information Model;
- Time Power Law Mapping of Signal Complexity Measure.

The book will be useful both for young researchers and students in higher education institutions who study the problems of multidimensional signal processing, as well as for experts working in this field.

Sofia, Bulgaria	Roumen Kountchev
Kobe, Japan	Rumen Mironov
October 2021	Kazumi Nakamatsu

Acknowledgments

The book editors express their special thanks to book chapter reviewers: Ivo Draganov, Roumiana Kountcheva, Mariofanna Milanova, Kun Zhang, Bin Hu, Silai Zhou, and Tian Zhou for their efforts and goodwill to help for the successful preparation of the book. Special thanks to Lakhmi Jain (Honorary Chair), Roumen Kountchev (General Chair), Srikanta Patnaik (General Chair), Parvinder Singh (Co-chair), and Rumen Mironov (Co-chair), and also to Kazumi Nakamatsu for the good organization of the workshop NAMSP 2021. The editors express their warmest thanks to the excellent Springer team which made this book possible.

Contents

Part I Multidimensional Signal Processing

Part II Deep Learning in Multidimensional Neural Networks

Part III Applications of Multidimensional Signal Processing

Part IV New Approaches in Communications and Computer Technologies

About the Editors

Prof. Roumen Kountchev Ph.D., D.Sc. is a professor at the Faculty of Telecommunications, Department of Radio Communications and Video Technologies, Technical University of Sofia, Bulgaria. Areas of interest include digital signal and image processing, image compression, multimedia watermarking, video communications, pattern recognition, and neural networks. Professor Kountchev has 350 papers published in magazines and proceedings of conferences; 20 books; 47 book chapters; 21 patents. He had been the principle investigator of 38 research projects. At present, he is a member of the Euro Mediterranean Academy of Arts and Sciences and president of Bulgarian Association for Pattern Recognition (member of International Association for Pattern Recognition). He is the editorial board member of International J. of Reasoning-based Intelligent Systems; International J. Broad Research in Artificial Intelligence and Neuroscience; KES Focus Group on Intelligent Decision Technologies; Egyptian Computer Science J.; International J. of Bio-Medical Informatics and e-Health, and International J. 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 Zayatoonah University, Amman, Jordan; WSEAS European Conference of Computer Science 2016, Rome, Italy; The 9th International Conference on Circuits, Systems and Signals, London, UK, 2017; IEEE International Conference on High Technology for Sustainable Development 2018 and 2019, Sofia, Bulgaria; The 8th International Congress of Information and Communication Technology, Xiamen, China, 2018; general chair of the International Workshop New Approaches for Multidimensional Signal Processing, July 2020, Sofia, Bulgaria.

Prof. Rumen Mironov 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 the Faculty of Applied Mathematics and Informatics. He is currently the 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 the 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. Kazumi Nakamatsu University of Hyogo, Kobe, Japan, Kazumi Nakamatsu received the Ms. Eng. and Dr. Sci. from Shizuoka University and Kyushu University, Japan, respectively. His research interests encompass various kinds of logic and their applications to Computer Science, especially paraconsistent annotated logic programs and their applications. He has developed some paraconsistent annotated logic programs called Annotated Logic Program with Strong Negation (), Vector ALPSN (VALPSN), Extended VALPSN (EVALPSN), and before-after EVALPSN (bf-EVALPSN) recently and applied them to various intelligent systems such as a safety verification-based railway interlocking control system and process order control. He is an author of over 150 papers, 20 book chapters, and 10 edited books published by prominent publishers. Kazumi Nakamatsu has chaired various international conferences, workshops, and invited sessions, and he has been a member of numerous international program committees of workshops and conferences in the area of Computer Science. He has served as the editor-in-chief of the International Journal of Reasoning-based Intelligent Systems (IJRIS), and he is now the founding editor of IJRIS and an editorial board member of many international journals. He has contributed numerous invited lectures at international workshops, conferences, and academic organizations. He also is a recipient of numerous research paper awards. He is a member of ACM.

Part I
Multidimensional Signal Processing

Chapter 1
Masked Face Detection Using Artificial Intelligent Techniques

Ehsan Nasiri, Mariofanna Milanova, and Ardalan Nasiri

Abstract Nowadays, wearing a face mask is a vital routine in life, but threats are increasing in public due to the advantage of wearing face masks. Existing works do not perfectly detect the human face and also not possible to apply for different faces detection. To overwhelm this issue, in this paper we proposed real-time face mask detection. The proposed work consists of six steps: video acquisition and keyframes selection, data augmentation, facial parts segmentation, pixel-based feature extraction, Bag of Visual Words (BoVW) generation, and face mask detection. In the first step, a set of keyframes are selected using the histogram of gradient (HoG) algorithm. Secondly, data augmentation is involved with three steps as color normalization, illumination correction (parameterized CLAHE), and pose normalization (Angular Affine Transformation). In the third step, facial parts are segmented using the clustering approach i.e., Expectation Maximization with Gaussian Mixture Model (EM-GMM), in which facial regions are segmented into Eyes, Nose, Mouth, Chin, and Forehead. Then, CapsNet based Feature Extraction is performed using CapsNet approach, which performance is higher and lightweight model than the Yolo Tiny V2 and Yolo Tiny V3, and extracted features are constructed into Codebook by Hassanat Similarity with K-Nearest neighbor (H-M with KNN) algorithm. For mask detection, L2 distance function is used. Experiments conducted using Python IDLE 3.8 for the proposed model and also previous works as GMM with Deep learning (GMM + DL), Convolutional Neural Network (CNN) with VGGF, Yolo Tiny V2, and Yolo Tiny V3 in terms of various performance metrics.

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

A. Nasiri
University of Arkansas at Fayetteville, Fayetteville, AR, USA
e-mail: exnasiri@ualr.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_1

1.1 Introduction

In recent years, masked face detection is a widely researched topic that gives several applications. Face detection from disguised/occluded/any other partially covered faces is a little difficult. Today, everyone wears a mask due to the spread of the COVID-19 pandemic [1–3]. The current state-of-the-art works in this field are designed by deep learning approaches. The crucial attributes that must be a matter for masked face detection can be as follows:

i. Faces Type, need to know Ellipse or Circle
ii. Eyes Location, Mark Eye Centers
iii. Face Orientation, including Left, Left Front, Right, and Right Front
iv. Occlusion Degree, need to Define Four Regions (Eye, Forehead, and Eyebrows) [4–6].

Figure 1.1 illustrates the challenges of classifying face-masked persons under the COVID-19 period [7, 8]. Recognizing masked faces and facial images taken from various sources such as video cameras, smartphones, CCTV surveillance cameras. From these sources, datasets are generated [9–11]. Among the several facial regions, the periocular region is one of the significant parts since it's uncovered by medical masks. Hence, it is expected to have poor image quality [12–14]. To address this issue, data augmentation like preprocessing is adopted. The core problem in masked face recognition is caused by the desired attributes of the publicly available real-world datasets which are listed as follows:

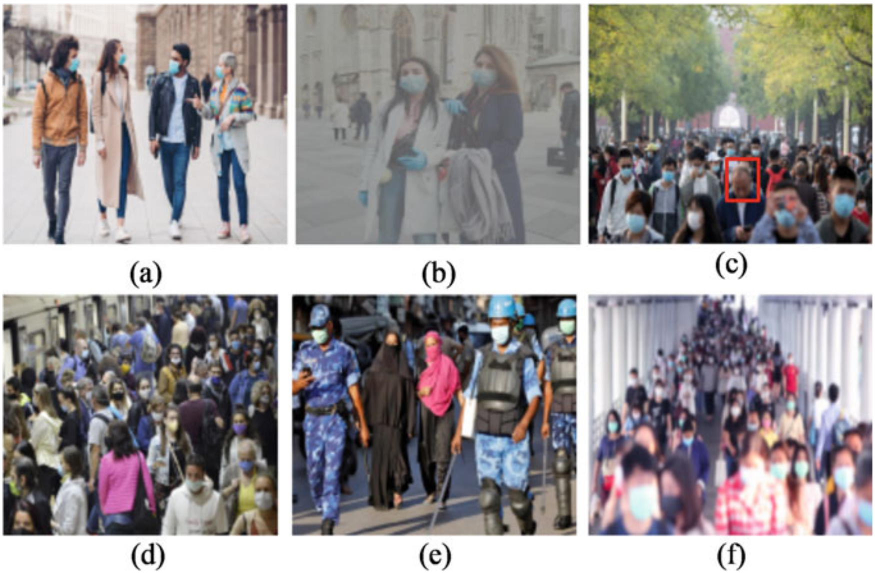


Fig. 1.1 Face Mask Detection Challenges. **a** Rotated faces, **b** foggy environment, **c** un masked faces, **d** crowd area, **e** fully covered faces and **f** blurred faces

i. Insufficient face images to recognize properly;
ii. Most faces have uniform features. In such cases, it's difficult;
iii. Variations in pixel intensity, illumination, occlusion in the captured facial images;
iv. Segmentation of unmasked regions is probably extracted by clustering algorithms;
v. Deep learning with less complexity is required, which reduces the training time.

Wearing face masks helps to slow the spread of this virus and easy to detect the faces who are not wearing the face masks. At the same time, crime rates are increasing these days due to the people wearing masks [15]. Hence, masked face detection is important today to avoid criminal offenses. Currently, a very small number of research works have been focused on this area which results in a lot of limitations. However, masked face recognition suffers from more factors challenged in routines, such as pose variations, illuminations, and uneven intensity distribution [16]. The performance of masked face recognition directly rests on the number of variations observed in the image. Ongoing research issues in this face detection are described as follows:

- Pose Variation: One of the major challenges in face mask detection is pose variation since it directly degrades the performance of image recognition. This is since images with frontal pose only cover maximum information of the face. The images with pose variations cannot be able to cover maximum information of the face that tends to reduce the detection rate. Therefore, pose normalization is required before entering the mask detection process.
- Illumination: The face image may be captured from different illumination conditions such as indoor, outdoor, etc. It has been observed that modification in lighting conditions induces different shading and shadows on the face image. This may deteriorate some of the features in the face image which results in too bright or too dark parts in images. This variation causes a low detection rate, hence illumination influence must be reduced in preprocessing of face mask detection [17, 18].

Therefore, in this paper, we pay great attention to understand key features for covered and uncovered face, i.e., face contour, periocular, nose, mouth, and chin, and so on. Thus, our foremost objective is to design an efficient data augmentation-based masked face detecting model, which is robust under erratic conditions such as illumination, pose changes, pixel intensity distribution. This paper has further objectives as follows:

i. To mitigate the discrepancies in the masked face databases such as lighting conditions, pose variations, and illumination.
ii. To minimize the masked face detection time while processing the masked face image by considering significant regions to the lightweight object detection model.
iii. To improve the recognition rate of the masked face detection using a novel deep learning algorithm that recognizes the face accurately.

1.2 Related Work

Skin texture analysis and illumination conditions are evaluated in [19] for robust face recognition. This paper handles soft and hard shadows and retains color channels and identity details. The presented illumination processing pipeline activates the chromaticity intrinsic image (CII) which highlights the variations of illumination. This paper achieves intrinsic face extraction processing and color tones of the face are recovered by removing shadows. Illumination conditions are different, which must be adaptively changed for each image.

Authors in [20] propose face mask detection by using a new texture feature descriptor, i.e., completed local ternary count (CLTC), which adds threshold value to address the noise issue. This feature descriptor is further improved by using a fast local Laplacian filter in the preprocessing stage. Therefore, the feature descriptor is called FLL-CLTC. Then, texture features are classified using the K-Nearest Neighbor (K-NN) algorithm. A face recognition task is evaluated for different kinds of datasets for JAFFE, Georgia Tech, Caltech, ORL and YALE face imaging datasets. Due to the variety of face mask images, the optimum threshold value is required. K-NN-based face recognition does not give accurate face recognition results since it finds probability values only.

The deep learning approach is proposed in [21], which considers a large volume of data. The input data is a video, in which is processed inter or intra difference between Pose, Occlusion, Scene, Blur, Video Quality, Illumination, etc. The surveillance videos and multiple-shot videos can be processed, even if they are of low quality. Multi-Scale Single Shot Face Detectors efficiently localize faces in videos. In this work, three datasets are applied such as multiple biometric grand challenges (MBGC), face and ocular challenge series (FOCS), IARPA Janus surveillance videos, and benchmark b for multi-shot videos. It takes a long time for handling longer videos and also very huge temporal dependencies must be evaluated in frames before recognizing the faces.

Edge computing-assisted face detection model is presented in [22]. This paper addressed the deep neural network issues such as latency and satisfies the higher accuracy. In this paper, input video is processed using a lightweight neural network, which is performed over edge nodes. In edge nodes, three lightweight NN models are implemented such as O-Net, R-Net, and P-Net for bounding boxes-based face detection. From the cloud server, videos are fetched that are transmitted from surveillance cameras. Video processing is managed in either edge node or cloud-based on the distance between the user and systems. However, human faces are different in color, texture, and shape, which must be analyzed differently. This paper lacks to identify these issues. Authors in [23] have proposed a patch-based attention generative adversarial network for face recognition which is shortly referred to as PA-GAN. This model aggregates raw surveillance frames features into a single model for minimizing the computational cost and maximizing the recognition accuracy. In PA-GAN, an enhanced Center Loss Function was proposed which integrates abundant unlabeled surveillance faces. Experiments validated for two kinds of datasets such as

QMUL-SurvFace Dataset and IJB-A dataset which demonstrates the efficiency of the proposed PA-GAN model. The PA-GAN model reduces the size of feature space by analyzing the interclass and intraclass distance. With this operation, operation time is reduced two times than the previous works.

A warning system is presented in [24], analyzing the suspicious faces from security video cameras. However, earlier security systems determine the suspicious people after the crime occurs. To eliminate it, an early warning system is initiated in this paper, and for training video frames are used convolutional neural networks (CNN). Checkpoint surveillance dataset is used in this paper, in which criminal information is extracted and stored. For face recognition, the threshold was dynamically adjusted in CNN. CNN is an initial deep learning model that does not accurately capture in-depth information of human faces. In particular, CNN does not consider face orientation, position, and poor performance in the max-pool layer.

In [25] authors addressed the problem of multi-faces trajectory in a single video. This is a complex problem and currently emerging issue to be addressed in COVID-19 pandemic situations, since most people wear masks. To address this problem, incremental learning algorithm with Euclidean distance-based greedy algorithm for recognition accuracy improvement is used. For features extraction, a local binary pattern histogram (LBPH) was used which extracts facial features from facial images. Euclidean distance is sensitive to noise and produces huge distortions when time-series data is processed. This is only suited for linear videos since it only aligns linear points.

Race recognition is implemented [26] using deep convolutional neural networks. The proposed architecture for race identification consists of three components: an information collector, face detection and pre-processor, and race recognition component. To train and test the dataset images, deep convolutional neural networks are applied. Finally, the race dataset was categorized into multiple Chinese, Japanese, and Brazilian based on their facial feature's variation. The overall D-CNN accuracy reached 90%. Bounding box detection is a very crucial issue in D-CNN, which must be detected for accurate detection. However, D-CNN is hard to train for high-dimensional data and also takes more time for training. It is a poor understanding of background objects.

1.3 Problem Statement

Masked Face detection accuracy is a crucial element that still suffers from the absence of primary tasks. Certain most important challenges are given below.

- Lack of Data Augmentation—Data augmentation is required before masked face detection. For example, pose normalization is one of the significant steps in data augmentation since head poses may be variant for humans. Pose normalization results in a front face that covers most face regions. Similarly, it requires other data augmentation tasks such as color normalization and illumination correction. Existing works failed to perform these tasks.

- Skin Texture Analysis Nonexistence—Human skin textures must be unique to one another. It has unique lines, pores, and spots appearance. In masked faces, we can't analyze skin texture, but it's possible to extract partial face images when it's covered.
- Absence of Accurate Loss Function – In literature, CNN is a widely used deep learning algorithm for masked face detection. In particular, a softmax loss is higher in CNN and also it suffers from inter and intra class separation problems.

The specific problems that are considered in this paper are as follows: Authors in [27] have presented face detection in partially covered faces. Hybrid CNN and VGGF algorithms are presented for feature extraction and matching. For training and validation, masked faces, occluded faces, zoomed-out faces, and disguised faces are focused. The most several problems in partially covered face recognition are as follows:

- If the face image is rotated and oriented, CNN-based classifiers have obtained poor performance. That is to say, the Position and Orientation of a given input image are ignored in max-pooling layers of CNN algorithm. This information is highly needed for feature extraction. Hence, it has less accuracy in classification.
- SVM is slower in processing which tends to cause higher processing time, especially has higher training time hence this is not applicable for real-time face recognition systems. Furthermore, parameter tuning, and kernel selection must be optimum.
- Softmax loss does not have discriminant power for class separation (Inter and Intra Classes) which does not suit deep face recognition.
- Head poses are different in the dataset, which is not sufficient to mask face recognition. And skin texture was not analyzed which reduces the recognition accuracy.

Human skin is varied [28] for masked and unmasked images and also regions. Hence, color conversion is implemented to predict the color values of the input image. For the surveillance videos, two-color spaces are used including RGB and YCbCr. However, skin patch presence detection merely does not give accurate face recognition.

- Skin texture analysis must turn out the unique lines, spots, appearance. When focusing on skin patches utilizing color tones, it produces poor performance.
- Head pose variations cause changes in face appearance which greatly impact classification results.
- Accurate segmentation is required here for covered and uncovered face detection.

Most of the works in face mask detection have used Yolo, CNN algorithms [29]. In Yolo, a wide variety of algorithms that are Yolo Tiny V2, V3 are used. Though, it has several issues as:

- Yolo's previous versions have several limitations such as (1) The limitation of grid cells since it is hard to detect; (2) It has a lower ability for bounding box detection, which does not optimally change according to human pose.
- This work is greatly affected by artificial factors such as pixel intensity, illumination, and variance in head poses.

In masked faces, the periocular region has played a significant role, which gives the accurate recognition result [30]. Texture features only do not increase accuracy in the periocular region. Dual-stream CNN increases complexity for extracting features and fusion and also long training time is needed in this combined work.

Further, the local binary-coded pattern does not work well when in lightning conditions and is also less robust under disguised and partially covered faces. It produces long histograms which slow down the recognition speed, particularly for large training databases.

1.4 Methodology

Architecture Overview

Our proposed work overthrows problems which occur in masked face detection. Our framework is composed of four sequential processes: Data Augmentation, Unmask Region Segmentation, and Horizontal Slicing, Multi-Feature Extraction, and Classification. Figure 1.2 illustrates the overall proposed model.

A detailed description for the proposed work is given as follows,

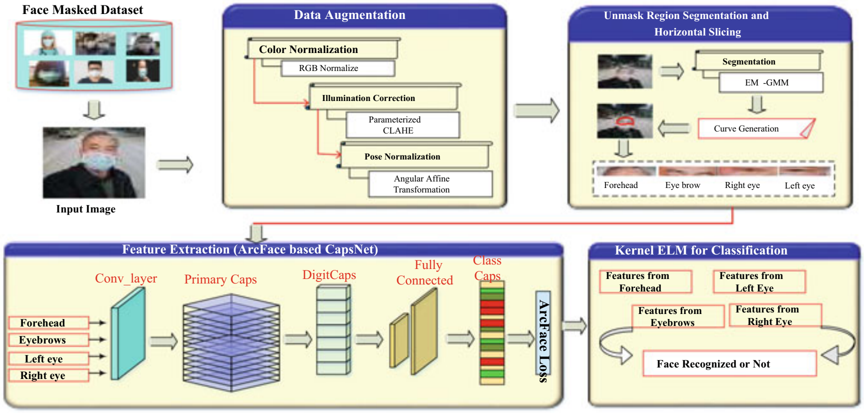


Fig. 1.2 System architecture

1.4.1 Data Augmentation

Our data augmentation step contains three processes that are color normalization, illumination correction, and pose normalization.

i. Color normalization

In the color normalization step, the pixel intensity distribution is performed which results normalized for R, G, and B channels.

RGB is a color model that consists of three-color components as RED, GREEN, and BLUE. It is represented as additive primitives and the color combination function is derived by follows,

$$\varsigma_p = R_p i + G_p j + B_p k \tag{1.1}$$

From the above equation, RGB color values are combined into the single-color value and this combined value plays a vital role in feature extraction for accurate mask face detection results.

ii. Illumination Correction

We adopt Parameterized Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm for illumination normalization that removes illumination in each input image.

The proposed Parameterized CLAHE model uses the luminance and contrast parameters adaptively for each frame. The Gamma correction method is used to establish the dark areas of the given image. It improves the whole luminance of the given image block. A dynamic range of gamma correction for each block is represented as follows:

$$\beta = \frac{p}{d_r}(1 + \tau\frac{g_{max}}{R} + \frac{\alpha}{100}(\frac{\sigma}{A_v + c})) \tag{1.2}$$

Here, p demonstrates the number of pixels in each block, d_r means dynamic range of this block; τ and α represent the stable parameters that are used to control the weight of dynamic range and entropies. σ is mentioned as the standard deviation of the block, A_v points out the mean value, and c is the small value to avoid division by 0. R is the dynamic value of luminance for a whole image. $g_{\max}$ means the maximum pixel value of the image. The gamma corrections are introduced to adjust the contrast value based on the current luminance value.

iii. Pose Normalization

Image databases contain different poses in an image, hence without performing proper pose normalization tends to have low accuracy in classification. Pose normalization was carried out using the Angular Affine Transformation algorithm. We initially estimate the pose angle of a given image using Angle (Yaw, Pitch, and Roll). Then, an estimated angle is provided to the Affine Transformation to get a frontal view of the given image. Image cropping is performed after completion of pose normalization to maintain the same size for all input images.

1.4.2 Unmask Region Segmentation and Horizontal Slicing

Unmasked face regions segmentation and horizontal striping are introduced to segment facial images. Here, the clustering-based segmentation is performed using the Expectation–Maximization based Gaussian Mixture Model (EM-GMM) algorithm. With EM-GMM, a curve is generated. Each time a new curve is generated. It overwhelms the problems of the fuzzy c means algorithm while segmenting facial parts. This way of segmentation tends to ease the process of classification.

In EM-based GMM, similar pixel values are gathered and then we integrate the two clustering approaches such as EM and GMM. In GMM, Gaussian mixture represents the linear superposition of Gaussians which are as follows,

$$p(x) = \sum_{k=1}^{K} \pi_k M\left(x|\mu_{k,}\Sigma_k\right) \tag{1.3}$$

Here K is the total number of Gaussians, k is the mixing coefficient and weightage for each Gaussian distribution. In this work, EM's purpose is to perform the iterative optimization which is performed locally. In the clustering technique, two processes are considered as,

- Expectation: For the input parameters set, we compute the latent variable expected values;
- Maximization: According to the latent variables, the values of parameters are updated.

In the GMM algorithm, likelihood function maximization is a significant part in which mean, and covariance components are measured. EM-GMM based clustering procedure is as follows:

i. Initialize the mean, covariance, and mixing components μ_j, Σ_j, and π_j respectively, and then compute the initial log-likelihood value.
ii. Implement the estimation step in which we calculate the tasks using the current metrics as follows.

$$\gamma_j(x) = \frac{\pi_k M\left(x|\mu_{k,}\Sigma_k\right)}{\sum_{k=1}^{K} \pi_j M\left(x|\mu_{j,}\Sigma_j\right)} \tag{1.4}$$

Then perform the maximization step in which compute the current parameters as follows:

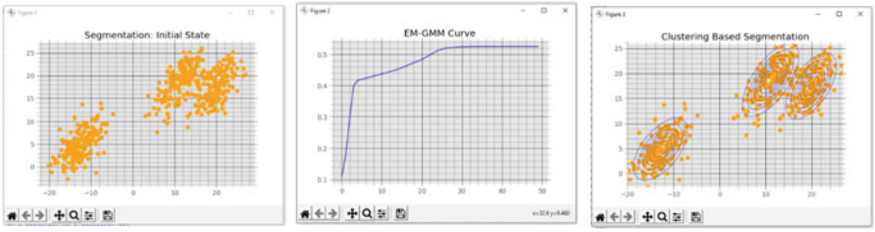


Fig. 1.3 EM-GMM algorithm

$$\mu_j = \frac{\sum_{n=1}^{N} \gamma_j(x_n)x_n}{\sum_{n=1}^{N} \gamma_j(x_n)} \tag{1.5}$$

$$\Sigma_j = \frac{\sum_{n=1}^{N} \gamma_j(x_n)\left(x_n - \mu_j\right)\left(x_n - \mu_j\right)^T}{\sum_{n=1}^{N} \gamma_j(x_n)} \tag{1.6}$$

$$\pi_j = \frac{1}{n}\sum_{n=1}^{N} \gamma_j(x_n) \tag{1.7}$$

iii. Next, compute the Log-Likelihood:

$$\ln p(X\,|\mu,\,\Sigma,\,\pi) = \sum_{n=1}^{N} ln\left\{\sum_{k=1}^{K} \pi_k M\,(x_n|\mu_{k,}\,\Sigma_k)\right\} \tag{1.8}$$

iv. Segments of face parts are implemented using a similar set of clusters. Figure 1.3 demonstrates the performance clustering for EM-based GMM clustering.

1.4.3 CapsNet Based Multi-Feature Extraction

Our feature extraction process extracts three types of features from the image to improve the classification performance and hence it's called multi-feature extraction. Here, features are extracted from four regions that are forehead, left eye, right eye, and eyebrows. For this purpose, we propose Capsule Network to extract features from the given segmented parts. The prime advantage of the Capsule network over CNN is its robustness in extracting the position and orientation information of the given image which is required to classify the masked image accurately. Loss computed by using ArcFace (Additive Angular Margin Loss). Figure 1.4 illustrates Feature Extraction Architecture.

High-level features

i. Motion (Any Feature or Object Changes over a Time);

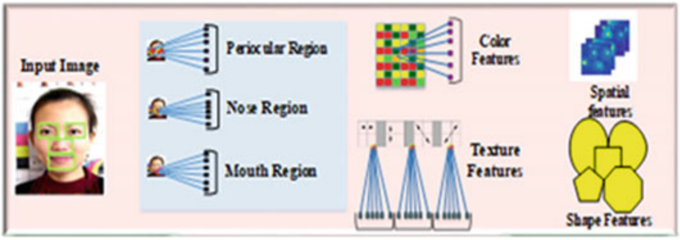


Fig. 1.4 Feature extraction architecture

ii. Spatial (Position, Angle, Orientation).

Low-level features

i. Color (Color Channel Values – HSV);
ii. Shape (Facial Parts Shape Values);
iii. Texture (Skin Surface and Appearance Type).

The extracted features are listed below,

- **Mean**: It is represented as the pixel intensity distribution in the whole region.

$$M_i = \frac{1}{n}\sum_{j=1}^{n} P_{ij} \tag{1.9}$$

- **Variance**: It is determined how each pixel can be varied from the center or neighboring pixels.

$$\Sigma_i = \sqrt{\left(\frac{1}{n}\sum_{j=1}^{n}\left(P_{ij} - \mu_i\right)^2\right)} \tag{1.10}$$

- **Skewness**: It represents the symmetry measure for the given face image, and it defines when the pixel values occur at the regular interval.

$$SW_i = \sqrt{\left(\frac{1}{n}\sum_{j=1}^{n}\left(P_{ij} - \mu_i\right)^3\right)} \tag{1.11}$$

- **Area**: It is calculated by the sum of pixels in the specific region, which are multiplied by the pixel's resolution.

$$A = I(x, y) \times \Delta A \tag{1.12}$$

- **Aspect Ratio**: It is calculated by the length and width of the frame.

$$AR = \frac{Le}{We} \tag{1.13}$$

- **Roundness**: It is calculated by the similarity of the frame in the circular shape.

$$RD = \frac{4\pi \mathrm{A}}{\mathrm{L}} \tag{1.14}$$

- **Perimeter**: It is defined as the structural property of the list of coordinates and also the sum of distance from each coordinate *x* and *y*.

$$PR = \sqrt{(\mathrm{X_i} - \mathrm{X_{i-1}})^2 - (\mathrm{Y_i} - \mathrm{Y_{i-1}})^2} \tag{1.15}$$

- **Circularity**: It is calculated by the largest pixels in each scene of frame region.

$$CR = \frac{4\pi \mathrm{A}}{\mathrm{Q}^2} \tag{1.16}$$

where ΔA is the area of one pixel in the shape of p_i(x, y), $\mathrm{X_I}$ and $\mathrm{Y_I}$ is the *i*th pixel coordinates; A is the object area, and L is the object region boundary length.

- **Uniformity**: It is computed by the uniformity in histogram intensities. This can be formulated as follows:

$$U = \sum\nolimits_{\mathrm{i=0}}^{\mathrm{l-1}} \mathrm{H^2(R_I)} \tag{1.17}$$

- **Mean**: It is the average intensity value computed for the number of pixels in the region:

$$\mu = \sum\nolimits_{\mathrm{i=0}}^{\mathrm{l-1}} \mathrm{p_i} \times \mathrm{H(p_I)} \tag{1.18}$$

- **Standard Variance**: This is the second-moment average value computed from the number of pixels.

$$\sigma = \sum\nolimits_{\mathrm{i=0}}^{\mathrm{l-1}} (\mathrm{p_i} - \mu)^3 \times \mathrm{H(p_I)} \tag{1.19}$$

Texture features are defined as the surface or appearance measurement for a given object, and it predicts the intensity, edges, and direction of pixels. Some of the texture features are described in the following.

- *Energy*: It is defined as the sum of square values of the pixels, which is also known as Uniformity or angular second moment. In mathematically, it is expressed as follows.

$$E_p = \sum_i \sum_j p^2(i,j) \tag{1.20}$$

- *Correlation*: This measures the color values dependency between the neighboring pixels.

$$C_p = \sum_i \sum_j p(i,j) logp(i,j) \tag{1.21}$$

- *Contrast*: The intensity contrast is measured between the current and neighboring pixel values. It is computed by follows,

$$CT_p = \sum_i \sum_j (i-j)^2 p(i,j) \tag{1.22}$$

- *Homogeneity*: It is inversely proportional to the contrast value, and it represents the equivalent distribution of pixels over the region.

$$H_p = \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \tag{1.23}$$

Capsule Network

A Capsule Neural Network is a machine learning system that is a type of artificial neural network that can be used to better model hierarchical relationships. Capsule Networks do not forward individual neuron activations from one layer to the next layer, but instead, each capsule represents a small nested neural network that outputs a whole vector. The total length of a capsule's output vector encodes the probability that a certain feature has been detected. The direction of the vector lengths helps to represent the state of the detected feature (e.g., location, pose, scale). When a feature moves across the image, the length of the vector should stay the same as the feature will be detected, but the vector's direction will change.

Structure of Capsule Network

A capsule *sj* then does not forward its output vector *Yj* blindly to every capsule in the next layer. Instead, a capsule predicts the output of all capsules in the next layer given its output vector *Yj* and the respective coupling coefficient *cij* and forwards its output only to that capsule whose predicted output results in the largest vector. This "next-layer output prediction" capsule *sj* ensures that it selects the most appropriate capsule for a given higher-level feature. Depending on the resulting next-layer output vector, the coupling coefficient *cij* be updated. Each capsule's state *sj* is calculated as the weighted sum of the matrix multiplication of output/prediction vectors of the capsules from the lower layer with the coupling coefficient *cij* between *sj* and the respective lower level capsule *si*.

$$Y\mathrm{j} = \sum\nolimits_{i}^{\infty}\left(C_{ij}\hat{u}_{i|j}, \hat{u}_{ij} = W_{ij}U_i\right) \tag{1.24}$$

Capsules in the first capsule layer of a capsule network calculate their activation based on the input from the previous convolution layer. In this case, no coupling coefficient c_{ij} exists.

As the capsule's output vector indicates the probability of having detected a certain feature, capsule *sj* output vector *Yj* is "squashed", so that long vectors sum up to 1 max and short vectors are close to zero.

$$Y_j = \frac{||S\mathrm{j}||^2}{1 + ||S\mathrm{j}||^2}\frac{S\mathrm{j}}{||S\mathrm{j}||} \tag{1.25}$$

The coupling coefficients *cij* define the "activation routing" between a capsule and all potential parent capsules in the next layer and sum to 1. The softmax-like calculation ensures that the most likely "parent" capsule gets the "most" of capsule *sj*'s output.

$$C_{ij} = \frac{e(b_{ij})}{\sum_{m} e(b_{ij})} \tag{1.26}$$

By following the presented calculations, the routing preferences between capsules, and the prediction of next layer activations, Capsule Network claim to address the CNN limitations listed above, especially modeling stronger feature relationships than CNN could represent which is a very strong tool to boost image segmentation.

Instance Normalization

In this step, the contrast of the image is improved, and content of low-level features is preserved such as textures and strokes. However, instance normalization is represented as follows,

$$Y_{ncij} = \frac{X_{ncij} - \mu_{nc}}{\sqrt{\sigma_{nc}^2 + \in}} \tag{1.27}$$

$$\mu_{nc} = \frac{1}{hw}\sum_{l=1}^{w}\sum_{m=1}^{h} X_{NCLM} \tag{1.28}$$

$$\sigma_{nc}^2 = \frac{1}{hw}\sum_{l=1}^{w}\sum_{m=1}^{h}\left(X_{NCLM} - \mu_{NC}\right)^2 \tag{1.29}$$

where, X_{ncij} represents the *ncij* (the layer feature map), *i* and *j* are the spatial dimensions, *c* is the channel feature, *n* represent the *n*th image in the batch. $\in$ is the small integer number that is involved to eliminate more number of computations. μ_{nc} is the mean value of feature image in the i the image and σ_{nc}^2 is the variance value of the feature image in the *ith* image.

1.4.4 BoVW Model

For extracted features, BoVW is implemented which constructs visual words dictionary and reduces feature space by clustering similar features using HS with KNN algorithm. KNN is a nearest neighbor prediction algorithm that can find the adjacent neighbor based on the high probability value. For visual words generation and codebook generation, in this paper, we presented KNN algorithm. In KNN, the distance between one feature to another is computed. The traditional KNN uses Euclidean distance that produces more noise. And it does not produce low-distance precision. It must be trained properly to eliminate the noise and low precision issues. Therefore, in this paper, we selected the most adopted distance formula in KNN i.e., Hassanat distance. It provides better performance in codebook generation.

Algorithm for HM-KNN

```
Input: Set of features
Begin
State the feature vector sets for a training set $t_S$
For each frame $F_i$ feature set $F_S$ do
Visual words generation……
    (a). Initialize $K$  // $K =$ Small integer value
    (b). Compute the distance between $F_i$ and $t_S$
    (c). Choose $K$ in $t_S$ close to $F_i$
    (d). Assign the most similar class close to the distance to $F_i$
Compute the class label for all frames of features
End for
Output: Assign the class label for all frames in a video $V$
```

Algorithm description is as follows: the training set of videos is denoted as t_s and the class label for each input frame F_I is stored in the database. Then, Hassanat similarity function is used to assign the exact class of the input frame. Based on that, the nearest visual words are constructed into a single group. The HM distance function is calculated as follows,

$$HD(X, Y) = \sum_{i=1}^{N} D(X_i, Y_i) \tag{1.30}$$

$$\text{Where} D(X, Y) = \begin{cases} 1 - \frac{1+Min(X_i,Y_i)}{Max(X_i,Y_i)}, Min(X_i, Y_i) \geq 0 \\ 1 - \frac{1+Min(X_i,Y_i)+|Min(X_i,Y_i)|}{Max(X_i,Y_i)+|Min(X_i,Y_i)|} Min(X_i, Y_i) < 0 \end{cases} \tag{1.31}$$

$D(X, Y)$ is bounded by 0 and 1.

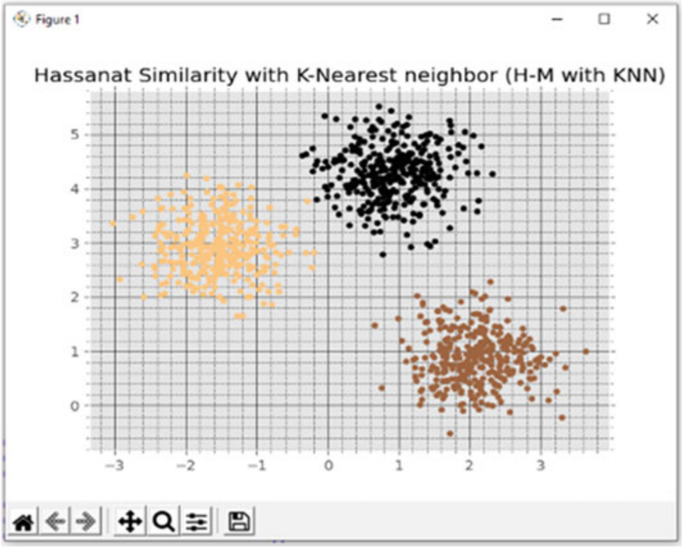


Fig. 1.5 HM- KNN result

In BoVW model, facial features $f_i(1,\ldots N)$ are constructed into a set of local keypoint descriptors as $f_i^p = \{p_{i,1}, p_{i,2}, p_{i,3}, \ldots, p_{1,m}\}$. Thus, the BoVW model is defined as follows:

$$BoVW : R^d \to [1, N] \tag{1.32}$$

$$P_{i,j} \to BoVW(p_{i,j}) \tag{1.33}$$

where, $p_{i,j} \epsilon R^d$ is a mapping descriptor that was used to produce an integer index. HM-KNN result is visualized in Fig. 1.5.

Finally, the masked face is detected (wears a mask or not) by computing the distance using L$_2$ distance function. We fetch the weights from masked values of the testing image and the trained image from HS-KNN. For similarity computation to classify the testing label L_2 distance (D) is used:

$$D = \sqrt{\sum\nolimits_{x=1}^{k} \left(VCt_{xi} - VCt_{xj}\right)^2} \tag{1.34}$$

where, $x = \{1, 2, 3 \ldots .k\}$, and [[$VC$]]_$t$—visual codebook of image.

1.4.5 Kernel ELM for Classification

The classification is a significant process in the masked face detection of the given extracted features. In this, a kernel-based extreme learning machine (ELM) algorithm is used to classify it. Based on the given input, the proposed kernel-ELM classifies the masked face or not. To optimize the performance of Kernel-based ELM, SMO algorithm is used. Based on the given input, the proposed classification algorithm classifies the masked faces person. The classification of the face and non-face region is carried out in the hidden node using the output weight W.FL(x) $= \sum_{n=1}^{L} \alpha_n \mu_n(X)$ where α_n denotes the output weight of the nth hidden node. J(x) $= [j_n(X), .., j_L(X)]$ is the hidden layer output of ELM. Given N the video frame, the hidden layer output matrix J of ELM is given as,

$$\mathrm{J} = \begin{bmatrix} j(X_1) \\ \vdots \\ j(X_N) \end{bmatrix} = \begin{bmatrix} G(p_1, q_1, X_1) & \cdots & G(p_L, q_L, X_L) \\ \vdots & \ddots & \vdots \\ G(p_1, q_1, X_N) & \cdots & G(p_L, q_L, X_N) \end{bmatrix} \tag{1.35}$$

where TM is the training matrix:

$$\mathrm{TM} = \begin{bmatrix} r_1 \\ \vdots \\ r_N \end{bmatrix} \tag{1.36}$$

The objective of ELM is to minimize,

$$||\alpha||_{a}^{\gamma_1} + \mathrm{C}||\mathrm{J}\alpha - R||_{b}^{\gamma_2} \tag{1.37}$$

where $\gamma 1, \gamma 2 > 0, a, b = 0, 12, 1, 2, \ldots, +\infty$.

1.5 Experimental Results and Discussion

In this section, we discuss the experimental and environmental settings for the proposed model implementation. Firstly, environment settings are presented. Secondly, a comparative study with the brief analysis of the proposed work with the existing works is presented for various performance metrics as segmentation accuracy, classification accuracy, precision, recall, f-score, ROC curve, and computational time. Finally, the results and discussion are given which highlights the novelty and significance of the proposed model.

1.5.1 Environment Settings

For implementation, the proposed model uses the Python IDLE 3.8 environment. The implementation settings involved in both hardware and software configuration are illustrated in Table 1.1. Table 1.2 describes the list of algorithms used in the proposed model and the choice of parameters and the values of each parameter are illustrated.

Figure 1.6 describes the main file executed in Python.

Dataset consists of 5000 images for 525 persons with masks and 90,000 images for the same 525 persons without wearing medical masks.

After environment configuration, initially, we load the dataset into the system for processing masked face detection. Our proposed work has the following processes Key Frame Selection, Data Augmentation, Face Regions Segmentation

Table 1.1 Environment settings

Hardware settings	Processor	3.00 GHz
	CPU	Dual core
	RAM	4 GB
	Hard disk	1 TB
Software settings	OS	Windows 10 (64bit)
	Python IDLE	3.8
	Library used	0.5.3
	(a) imutil	1.4.0
	(b) argparse	1.19.2
	(c) numpy	19.21.0
	(d) dlib	3.18.2. post1
	(e) cmake	20.2.3
	(f) pip	4.4.0.44
	(g) opencv-python	0.35.1
	(h) wheel	7.20
	(i) pillow	3.3.2-cp38
	(j) matplotlib	1.5.2-cp38
	(k) scipy	2.3.1-cp38
	(l) tensorflow	2.4.3
	(m) keras	2.8.1
	(n) pygad	0.2.1
	(o) resource	0.0
	(p) sklearn	0.17.2
	(q) scikit-image	0.1.3
	(r) elm	2.0.1
	(s) nano-python	0.5
	(t) yolo-v4	2.1.1
	(u) image-slicer	
	Command used	Pip install package_name

Table 1.2 Algorithm settings

KNN	Number of neighbors	10–100,000
	Distance metric	Hassanat
CapsNet	Epochs	120
	Batch size	8
	Learning rate	2.5e−4
	Fix up	True
	Learning policy	Cosine
ELM	Number of neurons	5 (input), hidden (10…150), output (1)
	Activation function	Tangent sigmoid
	Learning rule	The ELM of SLFN
	Sum squared error	0.0001
SMO	Number of iterations	10–1000
	Functions used	10–15

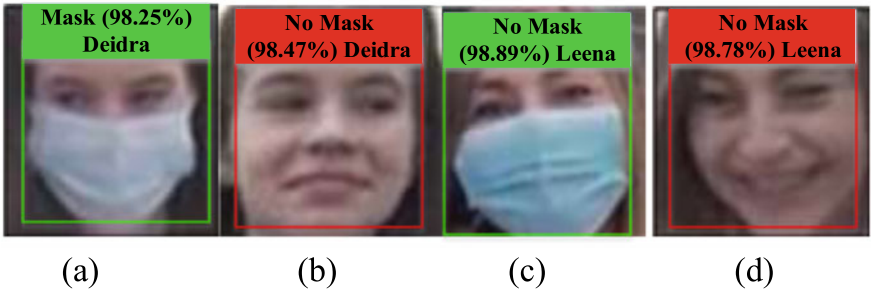


Fig. 1.6 **a** and **c** are the masked faces, **b** and **d** are the non-masked faces

and Horizontal Slicing, CapsNet based Multi-Feature Extraction & Masked Face Detection.

1.5.2 Dataset Description

To conduct experiments for masked face recognition, we use the “Real-time Masked Face Recognition Dataset”, which is shortly referred to as RMFRD. This is a massive real-world dataset.

1.5.3 Comparative Study

In this section, we describe the performance of the proposed model with comparison to the GMM + DL [31], CNN with VGGF [32], Yolo Tiny V2, and Yolo Tiny V3. Further, the performance metrics are listed as follows.

1.5.3.1 Segmentation Accuracy

This metric evaluates the segmentation performance for different facial parts segmentation. For instance, pixel-wise segmentation produces higher accuracy than region-based accuracy. Ground truth is marked in the training stage and pixel values are computed in the validation stage and similarity is computed for segmentation of face regions. Figures 1.7 and 1.8 represent the performance of segmentation accuracy with respect to the cluster scale. We compare the segmentation accuracy for face mask detection.

As a result of histogram analysis of the image to determine the similar pixels using dynamic threshold with EM-GMM model, the proposed work obtained peak

Fig. 1.7 Segmentation accuracy versus cluster scale for masked faces

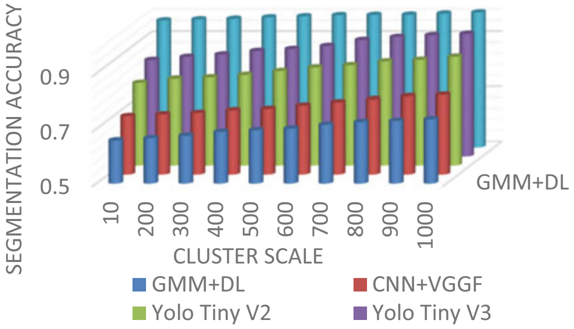


Fig. 1.8 Segmentation accuracy versus cluster scale for non-masked faces

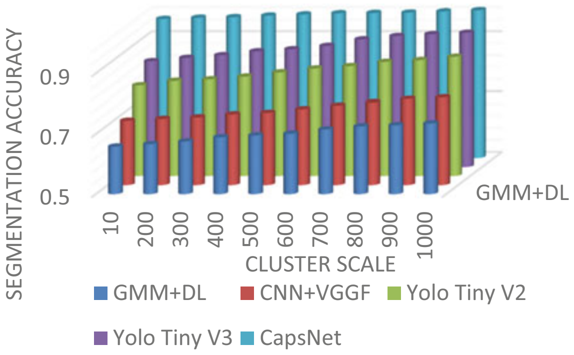


Table 1.3 Segmentation accuracy

Methods	Segmentation accuracy	
	Masked face	Unmasked face
GMM + DL	0.61% ± 0.05%	0.62% ± 0.04%
CNN + VGGF	0.71% ± 0.02%	0.72% ± 0.03%
Yolo Tiny V2	0.84% ± 0.03%	0.85% ± 0.04%
Yolo Tiny V3	0.92% ± 0.02%	0.921% ± 0.1%
CapsNet	0.95% ± 0.2%	0.95% ± 0.2%

segmentation accuracy. In previous works, many details (edge, boundary pixels) get omitted and threshold errors occur. Further, segments obtained in the proposed model are stable, and detected boundaries are distinct. In GMM, the global value of threshold is not good at all the conditions where the frame consists of different actionable characteristics. Hence, we propose an adaptive approach that can change the threshold value dynamically for various frames and different sets of faces. In this, the algorithm considers semantic-based pixel grouping, and based on semantic values pixels of small portions are groped and computed the threshold value for that portion.

Further, by dynamic curve generation by the EM-GMM, we obtained higher performance. To summarize that the proposed work achieved 0.97% ± 0.02% and 0.98% ± 0.01% for masked and unmasked face detection, respectively. Table 1.3 shows the performance achievement for segmentation accuracy.

1.5.3.2 Classification Accuracy

Classification accuracy is a significant metric that is most suitable for denoting the classified result. It is computed by four terms as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) and accuracy is computed by follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{1.38}$$

The evaluation of accuracy for both face mask and non-face mask images is not relevant. When compared to non-masked faces, masked face detection is a critical task, and improving the accuracy level for masked faces shows greater achievement of the presented model. In this study, we can expect higher classification accuracy owing to the CapsNet model, and also segmentation is performed accurately. The graphical plots of classification accuracy can be seen in Figs. 1.9 and 1.10 for masked and unmasked face detection.

For each image, nearly 1020 features are extracted and KNN with HS function is applied to construct the BoVW model, which produces higher classification accuracy. Under this, very few faces are incorrectly classified. Through dynamic thresholding,

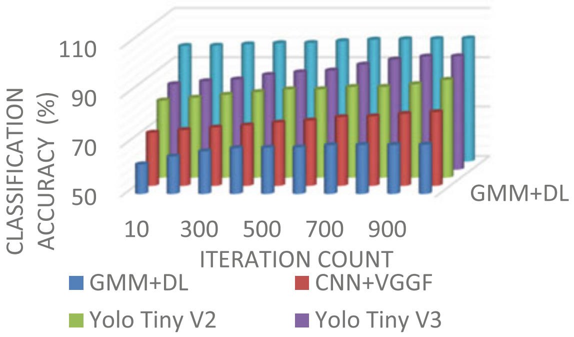


Fig. 1.9 Classification accuracy versus iteration count for masked faces

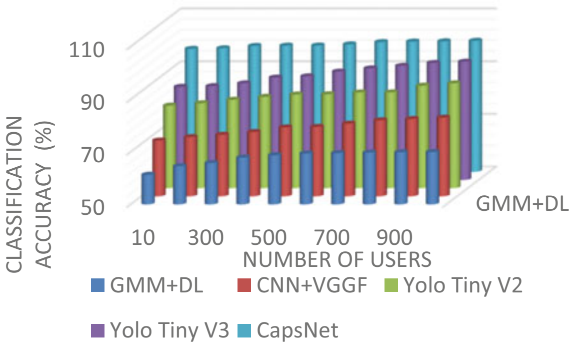


Fig. 1.10 Classification accuracy versus iteration count for non-masked faces

lightweight object detection models and ignorance of redundant features increase the classification accuracy. The highest classification accuracy is achieved by the proposed model at the 95% ± 0.2% and 96.2% ± 0.5% for masked face detection and non-masked face detection. However, training large volumes of input samples requires high computational cost and considers either region or boundary-based classification; lack of pixel-based feature extraction causes lower classification accuracy. Similarly, SVM classifiers do not perform well when the image contains more noise. In this case, target classes overlap with each other. Yolo V2 and V3 are adapted for large sizes of objects and classification requires high computations. Table 1.4 depicts

Table 1.4 Classification accuracy

Methods	Classification accuracy (%)	
	Masked face	Unmasked face
GMM + DL	61% ± 0.05%	62% ± 0.04%
CNN + VGGF	71% ± 0.02%	72.6% ± 0.03%
Yolo Tiny V2	84% ± 0.03%	85.5% ± 0.04%
Yolo Tiny V3	92% ± 0.02%	92.5% ± 0.1%
CapsNet	95% ± 0.2%	96.2% ± 0.5%

the performance of classification accuracy for various metrics.

1.5.3.3 Precision

In this metric, the performance of accurately classified face masks and non-masks is measured. It is computed by

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{1.39}$$

The results illustrated in Figs. 1.11 and 1.12 confirm that the superiority of the proposed model increases the accuracy for the iteration count as compared with the GMM + DL, CNN + VGGF, Yolo Tiny V2, and Yolo Tiny V3. The search strategy of SMO in face mask recognition is maintained by the utilization of kernels with ELM information. The connection between the CapsNet with ArcFace softmax loss function increases precision and decreases the false positives. As compared to the previous works, the proposed CapsNet model is better than ∼30%, ∼15%, ∼10%,

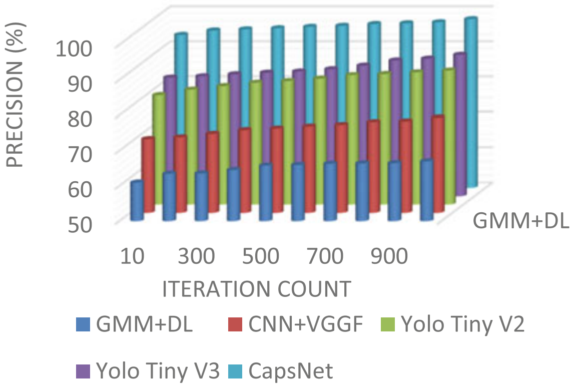


Fig. 1.11 Precision versus iteration count for masked faces

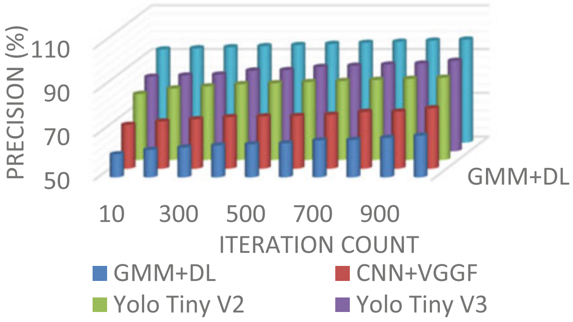


Fig. 1.12 Precision versus iteration count for non-masked faces

Table 1.5 Precision

Methods	Precision (%)	
	Masked face	Unmasked face
GMM + DL	60% ± 0.05%	62% ± 0.04%
CNN + VGGF	71% ± 0.02%	72.6% ± 0.03%
Yolo Tiny V2	84% ± 0.03%	85.5% ± 0.04%
Yolo Tiny V3	92% ± 0.02%	92.5% ± 0.1%
CapsNet	95% ± 0.2%	96.2% ± 0.5%

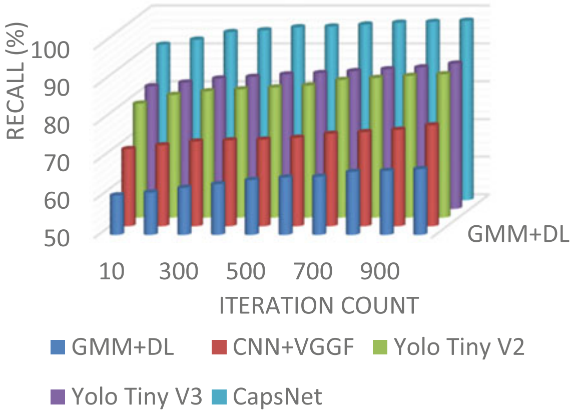


Fig. 1.13 Recall versus iteration count for masked faces

~8%, for GMM + DL, CNN + VGGF, Yolo Tiny V2, and Yolo Tiny V3, respectively. One of the vital advantages of the CapsNet model is in real-time working on any disguised images. In this model, more parameters are learned in the training period, which results in high performance on detecting the disguised images at any resolution. Further, data augmentation step gives increased accuracy in segmentation, face mask detection and thus it results better.

Overall results of the methods proposed and the existing works for precision are discussed in Table 1.5. It infers that the proposed CapsNet classifier provides higher precision values of 95% and 96.2% for masked and unmasked face images, respectively. Figure 1.13 describes the performance of the precision.

1.5.3.4 Recall

The recall is the proportion of positive cases that are determined accurately. In other words, it is the fraction of relevant images that are successfully determined. It is additionally referred to as a true positive. The recall is computed by the following function:

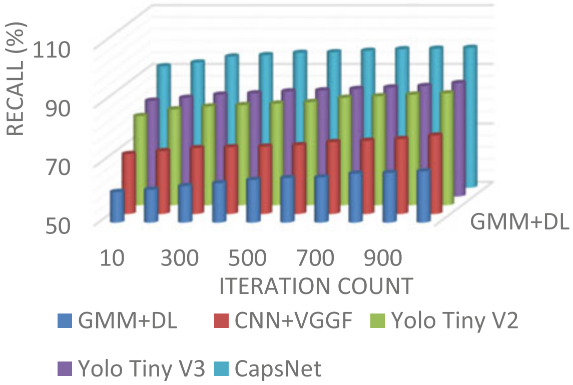


Fig. 1.14 Recall versus iteration count for non-masked faces

Table 1.6 Recall

Methods	Recall (%)	
	Masked face	Unmasked face
GMM + DL	59.6% ± 0.05%	61.5% ± 0.04%
CNN + VGGF	69.8% ± 0.02%	71.5% ± 0.03%
Yolo Tiny V2	84% ± 0.03%	83.5% ± 0.04%
Yolo Tiny V3	91.5% ± 0.02%	92% ± 0.1%
CapsNet	94.5% ± 0.2%	95.9% ± 0.5%

$$\text{Recall} = \frac{No\ of\ true\ positive}{No\ of\ relevant\ pattern} \text{(or)} \frac{TP}{TP + FN} \tag{1.40}$$

The performance of recall is illustrated in Fig. 1.14 for masked and unmasked face images. To compare the results of the CapsNet model, we implemented the proposed model for Yolo Tiny V2 and Yolo Tiny V3. After the comparison, it is proved that CapsNet produces better recall values than the proposed model. In other approaches such as CNN + VGGF, GMM with DL achieved performance due to the use of color, and texture features, and also to the facial landmarks are integrated to provide a good improvement in recall results. When compared with previous methods and algorithms, the proposed work fully addresses illumination variation, race issues, low/high-resolution images, noise, pose in just one shot, scale, and sharp, which are not able to predict in the training stage of the previous works and thus CapsNet model achieved better detection results (can see in Table 1.6).

1.5.3.5 F-Measure

It is the measure of precision and recalls value combination and it is also known as F1-measure and F-score. In particular, it is the mean value of precision and recall. It is denoted as follows,

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1.41}$$

$$= \frac{2 \times TP}{(2 \times TP + FP + FN)} \tag{1.42}$$

The performance of the F-measure is illustrated in Figs. 1.15 and 1.16. The CapsNet model consists of the whole procedure that performs better in data augmentation, feature extraction, codebook construction, and classification of masked and unmasked faces. Poses are determined in more than one shot of a video that can easily determine and correlate the human faces in an accurate manner.

Table 1.7 discusses the overall results of the proposed methods and previous methods (GMM + DL, CNN with VGGF, Yolo Tiny V2, and Yolo Tiny V3). Among the existing methods, the CapsNet model gives higher F-measure results. HoG features in color, texture features, and shape features improve F-measure, and also data augmentation steps increase F-measure than the previous works.

Fig. 1.15 F-measure versus iteration count for masked faces

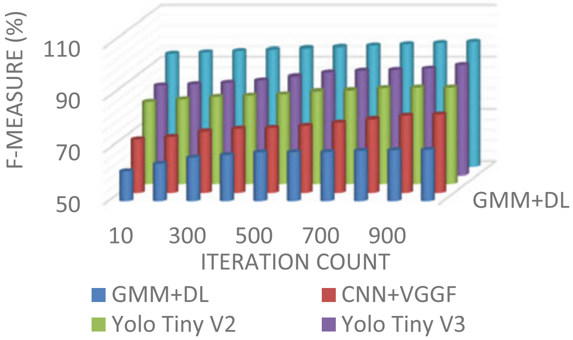


Fig. 1.16 F-measure versus iteration count for non-masked faces

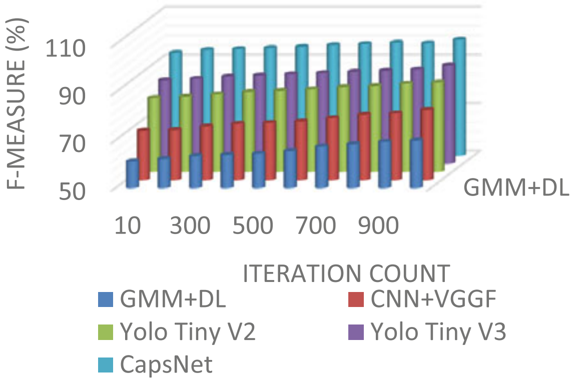


Table 1.7 F-measure

Methods	F-measure (%)	
	Masked face	Unmasked face
GMM + DL	60.2% ± 0.05%	62.3% ± 0.04%
CNN + VGGF	71.2% ± 0.02%	72% ± 0.03%
Yolo Tiny V2	84.5% ± 0.03%	84.2% ± 0.04%
Yolo Tiny V3	92.2% ± 0.02%	92.5% ± 0.1%
CapsNet	95.2% ± 0.2%	96.7% ± 0.5%

1.5.3.6 ROC Curve

A Receiver Operating Curve (ROC) is a graph that is used for system organization and visualization. It is a distinct option that is used for recall and precision curves. ROC curves are normally used in medical diagnosis decision-making and current years it is used for COVID-19 detection in more.

The graphical representation displays the transition between false positive rate (FPR) and true positive rate (TPR). The TPR denotes correctly classified or total positive values and plotted in the Y-axis and FPR denotes incorrectly classified or total negative values plotted over the x-axis. The points plotted over the top left of ROC have high TPR and low FPR represents the smart classification. Similarly, the TPR of the proposed classifier is very high and the previous method has produced very little TPR. Figure 1.17 describes the performance of the ROC curve.

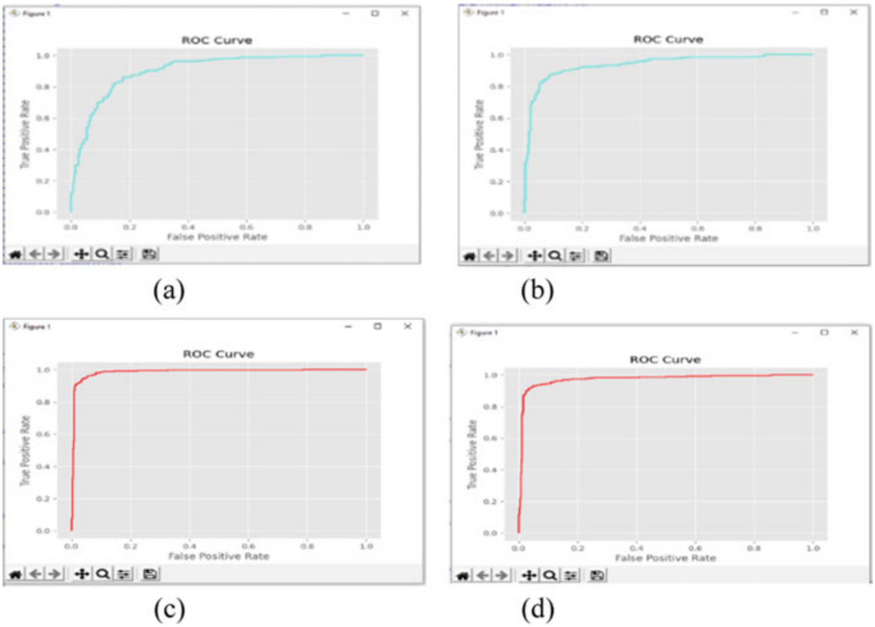
1.5.3.7 Computational Time

It is the sum of time taken to process the inputs for specific processes to reach the expected outcome. Previous methods consume more timing for training and testing, which results in a very high sum of time taken to process the image. We found that the CapsNet model produces the highest accuracy, precision, recall, and F-measure in more consistent computation time. The number of layers in CapsNet is very less and sequential operations do not produce high processing and computational time. Figures 1.18 and 1.19 represent the computational time for the number of iterations.

Table 1.8 denotes the performance of computational time for processing sequential operations. CapsNet model is an ultra-low light face detector that performs speed in providing high accuracy. For sequential operations implementation, CapsNet requires $O(1)$ amount of time and also complexity per layer takes $O(n)$.

1.5.3.8 Confusion Matrix

It is the matrix to deal with the face mask detection evaluation that revolved with the ground truth results with the obtained results. The ROC curve reveals the correlation between the TPR and FPR and differentiates the face and non-face classes in the



(a)

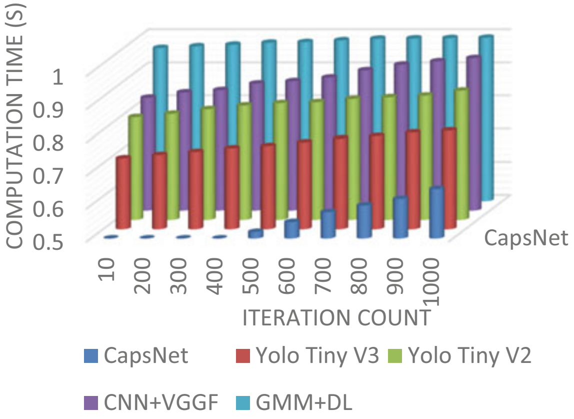
(b)

(c)

(d)

Fig. 1.17 ROC curve. **a** CapsNet (Masked Face), **b** CapsNet (Non-masked Face), **c** CNN + VGGF (Masked Face) and **d** CNN + VGGF (Non-masked Face)

Fig. 1.18 Computational time versus iteration count for masked faces



dataset. In other words, confusion matrix C is a square matrix where $C(ij)$ denotes the number of faces that are known in the Dataset i (True Label) and predicted to be in group j (Predicted Label).

Figures 1.20 and 1.21 represent the performance of the confusion matrix for predicted and true labels to the masked and unmasked face detection.

Fig. 1.19 Computational time versus iteration count for non-masked faces

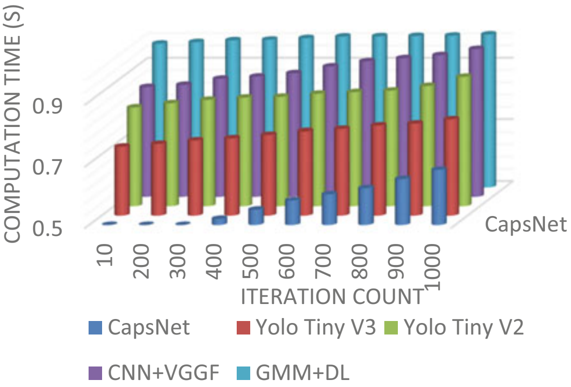
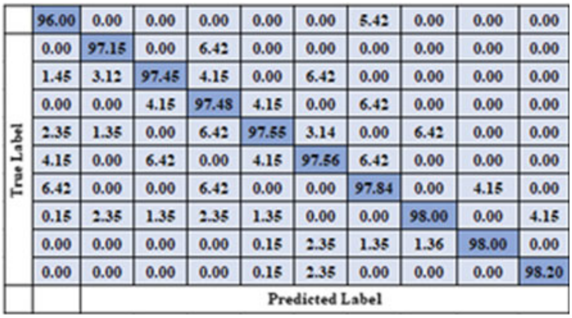


Table 1.8 Computation time (Seconds)

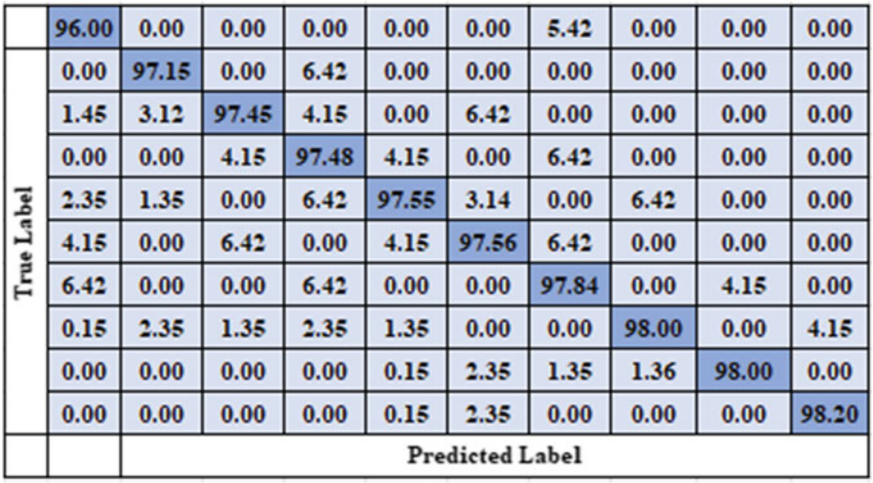
Methods	Computation time	
	Masked face	Unmasked face
GMM + DL	0.9	62.3
CNN + VGGF	0.84	72
Yolo Tiny V2	0.78	84.2
Yolo Tiny V3	0.65	92.5
CapsNet	0.4	96.7

Fig. 1.20 Confusion matrix for masked faces detection



	96.00	0.00	0.00	0.00	0.00	0.00	5.42	0.00	0.00	0.00
True Label	0.00	97.15	0.00	6.42	0.00	0.00	0.00	0.00	0.00	0.00
	1.45	3.12	97.45	4.15	0.00	6.42	0.00	0.00	0.00	0.00
	0.00	0.00	4.15	97.48	4.15	0.00	6.42	0.00	0.00	0.00
	2.35	1.35	0.00	6.42	97.55	3.14	0.00	6.42	0.00	0.00
	4.15	0.00	6.42	0.00	4.15	97.56	6.42	0.00	0.00	0.00
	6.42	0.00	0.00	6.42	0.00	0.00	97.84	0.00	4.15	0.00
	0.15	2.35	1.35	2.35	1.35	0.00	0.00	98.00	0.00	4.15
	0.00	0.00	0.00	0.00	0.15	2.35	1.35	1.36	98.00	0.00
	0.00	0.00	0.00	0.00	0.15	2.35	0.00	0.00	0.00	98.20
		Predicted Label								

Above figures show that the proposed model has higher accuracy in binary classification. It reaches up to 98.2% of the prediction rate in classification.



	96.00	0.00	0.00	0.00	0.00	0.00	5.42	0.00	0.00	0.00
True Label	0.00	97.15	0.00	6.42	0.00	0.00	0.00	0.00	0.00	0.00
	1.45	3.12	97.45	4.15	0.00	6.42	0.00	0.00	0.00	0.00
	0.00	0.00	4.15	97.48	4.15	0.00	6.42	0.00	0.00	0.00
	2.35	1.35	0.00	6.42	97.55	3.14	0.00	6.42	0.00	0.00
	4.15	0.00	6.42	0.00	4.15	97.56	6.42	0.00	0.00	0.00
	6.42	0.00	0.00	6.42	0.00	0.00	97.84	0.00	4.15	0.00
	0.15	2.35	1.35	2.35	1.35	0.00	0.00	98.00	0.00	4.15
	0.00	0.00	0.00	0.00	0.15	2.35	1.35	1.36	98.00	0.00
	0.00	0.00	0.00	0.00	0.15	2.35	0.00	0.00	0.00	98.20
		Predicted Label								

Fig. 1.21 Confusion matrix for un-masked faces detection

1.5.4 Novelty & Significance Analysis

The novelty and significance of the proposed work are listed as follows:

i. The highlights of the proposed work are listed as follows:
ii. We presented a Kernel-based ELM algorithm, which improves mask detection accuracy by greater than 4% of SVM since its, learning is better and also speeds fast.
iii. Arcface based CapsNet is proposed which uses additive marginal loss, which performance is higher when performing deep face recognition
iv. Multiple features are extracted such as color tones, texture, and shape for detection of masked faces.
v. Illumination and non-uniform pixel intensity are the most important artifacts that are removed completely in the data augmentation stage.

1.6 Conclusion

In this paper, the CapsNet objection model is used for face mask detection. For that, data augmentation, feature extraction, codebook generation, face mask detection. We evaluated the performance of the CapsNet model according to the accuracy, precision, recall, f-measure, and computational time on the Dataset. In this paper, we explore the issues of disguised images such as illumination variation, noise, scale, pose in a single frame, and more patches variation. Based on the research gaps, algorithm comparison, and multi-views of performance are analyzed for the proposed and previous methods. The desired face detection model has higher results in terms of accuracy and precision in less computational time. This is a very interesting part of the proposed work and also, we think that due to algorithm selection, processes considered and more facial features of different views of human frontal faces in any angular degrees improve the performance of face mask detection.

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, National Science Foundation under Award No. OIA-1946391(DART) and NSA grant NCAE-C Cyber Curriculum and Research 2020 Program, NCAEC-003-3030

References

1. Cheng, V.C.C., Wong, S.-C., Chuang, V.W.M., So, S.Y.C., Chen, J.H.K., Sridhar, S., Yuen, K.-Y., et al.: The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. J. Infect. (2020)
2. Cabani, A., Hammoudi, K., Benhabiles, H., Melkemi, M.: MaskedFace-Net—a dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health 100144 (2020)
3. Razavi, M., Alikhani, H., Janfaza, V., Sadeghi, B., Alikhani, E.: An automatic system to monitor the physical distance and face mask wearing of construction workers in COVID-19 pandemic (2021)
4. Ejaz, M.S., Islam, M.R.: Masked face recognition using convolutional neural network. In: 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI) (2019)
5. Meenpal, T., Balakrishnan, A., Verma, A.: Facial mask detection using semantic segmentation. In: 2019 4th International Conference on Computing, Communications and Security (ICCCS) (2019)
6. Bhuiyan, M.R., Khushbu, S.A., Islam, M.S.: A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3. In: 2020 11th international conference on computing, communication and networking technologies (ICCCNT) (2020)
7. Bu, W., Xiao, J., Zhou, C., Yang, M., Peng, C.: A cascade framework for masked face detection. In: 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM) (2017)
8. Joshi, A.S., Joshi, S.S., Kanahasabai, G., Kapil, R., Gupta, S.: Deep learning framework to detect face masks from video footage. In: 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 435–440 (2020)
9. Draughon, G., Sun, P., Lynch, J.: Implementation of a computer vision framework for tracking and visualizing face mask usage in urban environments. In: 2020 IEEE International Smart Cities Conference (ISC2), 1–8 (2020)
10. Kose, N., Dugelay, J.-L.: Mask spoofing in face recognition and countermeasures. Image Vis. Comput. **32**(10), 779–789 (2014)
11. Qezavati, H., Majidi, B., Manzuri, M.T.: Partially covered face detection in presence of headscarf for surveillance applications. In: 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), pp. 195–199 (2019)
12. Yuan, C., Yang, Q.: A dynamic face recognition deploy and control system based on deep learning. J. Residuals Sci. Technol. **13** (2016)
13. Engoor, S., Selvaraju, S., Christopher, H.S., Suryanarayanan, M.G., Ranganathan, B.: Effective emotion recognition from partially occluded facial images using deep learning (2020)
14. Salari, S.R., Rostami, H.: Pgu-face: a dataset of partially covered facial images. Data Brief **9**, 288–291 (2016)

15. Song, L., Gong, D., Li, Z., Liu, C., Liu, W.: Occlusion robust face recognition based on mask learning W ith pairwise differential Siamese network. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 773–782 (2019)
16. Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Miao, Y., Huang, Z., Liang, J.: Masked face recognition dataset and application. ArXiv abs/2003.09093 (2020)
17. Nair, A., Potgantwar, A.: Masked face detection using the Viola Jones algorithm: a progressive approach for less time consumption. Int. J. Recent Contrib. Eng. Sci. IT **6**, 4–14 (2018)
18. Ejaz, M.S., Islam, M.N., Sifatullah, M., Sarker, A.: Implementation of principal component analysis on masked and non-masked face recognition. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–5 (2019)
19. Hariri, W.: Efficient masked face recognition method during the COVID-19 pandemic (2020)
20. Dey, S.K., Howlader, A., Deb, C.: MobileNet mask: a multi-phase face mask detection model to prevent person-to-person transmission of SARS-CoV-2 (2021)
21. Loey, M., Manogaran, G., Taha, M., Khalifa, N.E.: Fighting against COVID-19: a novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustain. Cities Soc. **65**, 102600 (2020)
22. Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., Hemanth, J.D.: SSDMNV2: a real-time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustain. Cities Soc. (2020)
23. Chowdary, G.J., Punn, N.S., Sonbhadra, S.K., Agarwal, S.: Face mask detection using transfer learning of inceptionV3. ArXiv abs/2009.08369 (2020)
24. Sikandar, T., Samsudin, W.N.A.W., Rabbi, M.F., Ghazali, K.H.: An efficient method for detecting covered face scenarios in ATM surveillance camera. SN Comput. Sci. **1**(3) (2020)
25. Loey, M., Manogaran, G., Taha, M., & Khalifa, N.: A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement **167**, 108288 (2021)
26. Chen, Q., Sang, L.: Face-mask recognition for fraud prevention using gaussian mixture model. J. Vis. Commun. Image Representation **55** (2018)
27. Kim, M., Koo, J., Cho, S., Baek, N., Park, K.: Convolutional neural network-based periocular recognition in surveillance environments. IEEE Access 1–1 (2018)
28. Liu, D., Bellotto, N., Yue, S.: Deep spiking neural network for video-based disguise face recognition based on dynamic facial movements. IEEE Trans. Neural Netw. Learn. Syst. 1–10 (2019)
29. Ud Din, N., Javed, K., Bae, S., Yi, J.: A novel GAN-based network for unmasking of masked face. IEEE Access **8**, 44276–44287 (2020)
30. Zhao, Z., Kumar, A.: Improving periocular recognition by explicit attention to critical regions in deep neural network. IEEE Trans. Inf. Forensics Secur. **13**(12), 2937–2952 (2018)
31. Zhang, W., Zhao, X., Morvan, J.-M., Chen, L.: Improving shadow suppression for illumination robust face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 1–1 (2018)
32. l-Shaibani, B.: A new fast local Laplacian completed local ternary count (FLL-CLTC) for facial image classification. IEEE Access **8**, 98244–98254 (2020)

Chapter 2
Object Motion Detection in Video by Fusion of RPCA and NMF Decompositions

Ivo Draganov and Rumen Mironov

Abstract In this paper two new schemes are proposed for fusion of the results from video decomposition by Robust Principal Component Analysis and Non-negative Matrix Factorization with the aim of detecting moving objects over stationary background. The schemes use the logical OR and AND operators on a pixel basis over the binary outputs of the base decomposition algorithms. Experimental results from testing with videos, containing natural scenes with humans and vehicles, reveal the applicability of both schemes with higher Detection Rate for the OR operator and considerably higher Precision for the AND operator. The latter gets the highest F-measure of 0.8168 and is considered applicable in various systems where higher reliability is sought. Execution times for all tested implementations are practical, although allowing further optimization, which renders the proposed algorithms applicable in a wide set of applications.

2.1 Introduction

Detection of moving objects in video has important role in numerous applications, such as security surveillance, traffic control, automation of industrial processes and in a lot of other areas. The Robust Principal Component Analysis (RPCA) and the Non-negative Matrix Factorization (NMF) have been two of the most popular approaches for decomposing a noisy input, where lots of factors, such as unpredictable change of the scene illumination, camera sensor noise, artifacts from video compression and others contribute to the complexity of locating a moving object over the background.

Javed et al. [1] introduce regularization of spatial and temporal nature to the sparse component from RPCA decomposition of video. Thus, they achieve a decrease in the effect of the mutual dependence among some of the elements in the sparse matrix

I. Draganov (✉) · R. Mironov
Technical University of Sofia, 8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria
e-mail: idraganov@tu-sofia.bg

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

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_2

from the input one. Graph Laplacians are the tool, which allows for the complete algorithm to take place. It has been shown that algorithm of this type could be executed in real-time. Another challenging factor into detecting moving objects in video is the irregularity of the movement. For that challenge to be solved, Cao et al. [2] propose the total variation regularized RPCA, which comes also as a solution to the problem of dynamic background. Spatial continuity of the foreground and temporal continuity, observable for lingering objects, are the two properties, embedded into this solution. The Lagrangian multiplier with augmentation lays in the foundation of the solver, where the minimization process happens due to the alternating direction method. Aiming to solve the same problem, Javed et al. [3] rely on various manifold regularizations, processing the sparse component and looking for local and global stationaries. Spectral graph structure, based on Laplacian regularization, along with superpixel formations help into getting the appropriate representations. Li et al. [4] undertake another strategy of getting slowly moving objects over complex backgrounds properly detected, entitled Segmentation and Saliency constrained RPCA (SSC-RPCA). They use superpixels, formed by analyzing the priors of the spatial and temporal continuity. Further segmentation is performed by clustering superpixels to subregions and finally merging some of them based on continuity of content. Graph regularization is applied in [5] for application of the online spatiotemporal RPCA over RGB-D videos. In that study, some of the aims are limiting the effect of color saturation for both foreground and background pixels, as well as solving the problem of restructuring of multiway video to matrices and speeding up the whole process to real-time execution.

NMF is another major approach, widely used for moving object detection in videos. Chen et al. [6] look at the problem with ever growing amounts of video, generated by surveillance systems, especially those operated over clouds, and the need for robust detection of moving objects at various scenarios. A new model, related to the sparse level and the low-rank matrix representation of videos, employing contextual regularization, is being proposed. The moving objects themselves are treated as contiguous outliers, while the background's model is rebuilt in a dedicated fashion, further added to a dictionary for later use in other, similar videos. NMF has been also used with other techniques, such as fusing it with vector similarity analysis [7]. This approach relies on the background reconstruction based on the continuity of the image sequences from the video. Later on, similarity between new candidate from an object pixels and already modeled backgrounds is found and decision is being taken about the affiliation of the current area of interest. Kernel density estimation supports the execution of this algorithm in acceptable periods of time. Detection of objects in video, performing non-salient movements is primary objective of the study, described in [8]. The assumption is that the background could be modeled in a subspace of low dimensions. The robust matrix completion turns out to be the effective solution in this case, employing the Frank-Wolfe optimizer and the Fast Principal Component Pursuit (FPCP), which is also known to be efficient as a combination with NMF. The resulting algorithm is computationally effective, while detection rate is 10% higher than other base implementations. The negative influence of shadows on the detection rate of moving objects in video has been addressed in various implementations. Yang et al. [9] use the NMF and its variant of Block-NMF (BNMF), together with selection of key points, to better spot the areas with cast shadows.

All methods, described above, rely on a single type of decomposition of the video, either RPCA or NMF, most often with a combination of additional extension of the objective function, minimizing the difference between the input and representing low-rank and sparse matrices, as well as other techniques. In this study, the main aim is to combine two base decomposition—one of RPCA type and one of NMF with simple logical operators on a pixel level over segmented videos after background subtraction, without significant increase of computational complexity or losing detection accuracy. In Sect. 2.2, description of the 4 basic decompositions, employed in this research is given and 2.2 new schemes for fusion of segmented video frames, followed by experimental results in Sect. 2.3 and discussion in Sect. 2.4. Section 2.5 represents a brief conclusion on the most important qualities of the newly presented implementations.

2.2 Algorithms Description

2.2.1 Robust Principal Component Analysis Semi-Soft Go Decomposition

The Robust Principal Component Analysis Semi-Soft Go Decomposition (RPCA SSGoDec) [10] is useful into accomplishing the task of background subtraction in videos, represented as 3-dimensional entities. It starts from the idea of getting low-rank $\boldsymbol{L}$ and sparse $\boldsymbol{S}$ parts, as matrices, of the input video $\mathcal{X}$, which could be defined as 3rd order tensor, where the following error should be minimized [10]:

$$\min_{\boldsymbol{L},\boldsymbol{S}} \|\mathcal{X} - \boldsymbol{L} - \boldsymbol{S}\|_F^2, rank(\boldsymbol{L}) \leq r, card(\boldsymbol{S}) \leq k, \tag{2.1}$$

where F indicates the Frobenius type of a norm; r—the upper bound of the rank of $\boldsymbol{L}$ and k—the upper bound of the cardinality of $\boldsymbol{S}$. The operators rank(.) gives the rank of an element and card(.)—its cardinality. The general form of the optimization problem (2.1) could be divided into 2 parts, which could be solved subsequently, according to [10]:

$$\left| \begin{array}{l} \boldsymbol{L}_t = arg \min\limits_{rank(\boldsymbol{L}) \leq r} \|\mathcal{X} - \boldsymbol{L} - \boldsymbol{S}_{t-1}\|_F^2 \\ \boldsymbol{S}_t = arg \min\limits_{card(\boldsymbol{S}) \leq k} \|\mathcal{X} - \boldsymbol{L}_t - \boldsymbol{S}\|_F^2 \end{array} \right. . \tag{2.2}$$

Both updates $\boldsymbol{L}_t$ and $\boldsymbol{S}_t$ could be found according to [10]:

$$\left|\begin{array}{c} \boldsymbol{L}_t = \sum_{i=1}^{r} \lambda_i \boldsymbol{U}_i \boldsymbol{V}_i^T, svd(\mathcal{X} - \boldsymbol{S}_{t-1}) = \boldsymbol{U} \wedge \boldsymbol{V}^T \\ \boldsymbol{S}_T = P_\Omega(\mathcal{X} - \boldsymbol{L}_t), \Omega : \left|(\mathcal{X} - \boldsymbol{L}_T)_{i,j\in\Omega}\right| \neq 0 and \geq \left|(\mathcal{X} - \boldsymbol{L}_t)_{i,j\in\overline{\Omega}}\right|, |\Omega| \leq k, \end{array}\right. \tag{2.3}$$

where λ_i is the *i*-th singular value in descending order of magnitude, $\boldsymbol{U}$ and $\boldsymbol{V}$—the components from Singular Value Decomposition (SVD) of $\mathcal{X}$, P_Ω—projection of $\mathcal{X}$ to a set Ω. Since SVD requires considerable time, rising with the dimensions of $\mathcal{X}$, it is preferred to use bilateral random projections (BRP) in order to accomplish the GoDec. The update of $\mathbf{L}$, then, comes to $\boldsymbol{L}_t = \boldsymbol{Y}_1(\boldsymbol{A}_2{}^T\boldsymbol{Y}_1)^{-1}\boldsymbol{Y}_2{}^T$ [10].

If a matrix $\mathcal{X}$ that is known to have low rank, corresponding to rank$(\mathcal{X}) \leq r$, it could be completed precisely from an entity of input values $\boldsymbol{Y} = P_\Omega(\mathcal{X})$. Optimization, according to the following procedure, could lead to the solution of the problem [10]:

$$\min_{X,Z} \|\boldsymbol{Y} - \mathcal{X} - \boldsymbol{Z}\|_F^2, rank(\mathcal{X}) \leq r, supp(\boldsymbol{Z}) = \overline{\Omega}, \tag{2.4}$$

where $\mathbf{Z}$ is the result of solving $P_{\overline{\Omega}}(\mathcal{X})$. The GoDec algorithm is fully applicable to solve (2.4), following proper substitution of the included sets and processed arguments.

2.2.2 Robust Principal Component Analysis Lagrangian Stable Principal Component Pursuit with Quasi-Newton

The Robust Principal Component Analysis Lagrangian Stable Principal Component Pursuit with Quasi-Newton (RPCA Lag-SPCP-QN) method [11] is a variational approach, offering capabilities for processing videos, related to segmentation in particular, when they are represented as complete 3D sets, comprising of all the frames, gathered over time. SPCP treats a noisy input $\boldsymbol{Y}$, a matrix as an instance of dimensions *m*x*n* over $\mathbb{R}$, so it looks for optimum solution to the problem [11]:

$$\min_{\boldsymbol{L},\boldsymbol{S}} \||\boldsymbol{L}|\|_* + \lambda_{sum}\|\boldsymbol{S}\|_1, given\|\boldsymbol{L} + \boldsymbol{S} - \boldsymbol{Y}\|_F \leq \varepsilon, \tag{2.5}$$

where the 1-norm of $\boldsymbol{S}$ is $\|\boldsymbol{S}\|_1 = \sum_{i,j}\left|s_{i,j}\right|$ and the nuclear norm of $\boldsymbol{L}$ is $\||\boldsymbol{L}|\|_* = \sum_i \sigma_i(\boldsymbol{L})$, σ—the singular values of $\boldsymbol{L}$, ordered in a vector. The parameter λ_{sum} plays a role of balancing between the influence of $\boldsymbol{L}$ and $\boldsymbol{S}$ in the process of optimization, while ε takes care of unpredictable deviations within the input data by limiting the whole optimization procedure. Developing further the task in the form of practical solution Aravkin et al. [11] propose the so-called max-SPCP:

$$\underset{L,S}{\min}\max(\||L|\|_*, \lambda_{max}\|S\|_1), given\|L + S - Y\|_F \leq \varepsilon. \tag{2.6}$$

This optimization procedure is thought to be computationally more efficient than (2.5), in which λ_{max} has the same influence as λ_{sum} in (2.5). There are few formulations of this problem, namely flipping within the SPCP_{max} and SPCP_{sum} and the Lagrangian form, given respectively by [11]:

$$\underset{L,S}{\min}\frac{1}{2}\|L + S - Y\|_F^2, givenmax(\||L|\|_*, \lambda_{max}\|S\|_1) \leq \tau_{max}, \tag{2.7}$$

$$\underset{L,S}{\min}\frac{1}{2}\|L + S - Y\|_F^2, given\||L|\|_* + \lambda_{sum}\|S\|_1 \leq \tau_{sum}, \tag{2.8}$$

$$\underset{L,S}{\min}\lambda_L\||L|\|_* + \lambda_S\|S\|_1 + \frac{1}{2}\|L + S - Y\|_F^2. \tag{2.9}$$

Quai-Newton methods are applicable into solving the problems from above, given the representation for a gradient as [11]:

$$\nabla f(\mathcal{X}) = \begin{pmatrix} \nabla_L f(\mathcal{X}) \\ \nabla_S f(\mathcal{X}) \end{pmatrix} = A^T \begin{pmatrix} r(\mathcal{X}) \\ r(\mathcal{X}) \end{pmatrix}, r_k \equiv r(\mathcal{X}_k). \tag{2.10}$$

It has been shown that this approach leads to faster implementation of the SPCP_{max} and it is competitive to other solutions for the RPCA problem [11].

2.2.3 *Manhattan Non-negative Matrix Factorization*

The Manhattan Non-negative Matrix Factorization (ManhNMF) is a technique for finding the low-rank and sparse matrices, representing a non-negative matrix $\mathcal{X}$, according to [12]:

$$\underset{W \geq 0, H \geq 0}{\min} f(W, H) = \|\mathcal{X} - W^T H\|_M, \tag{2.11}$$

where $||.||_M$ represents the norm with the use of Manhattan distance, $\boldsymbol{W}$ and $\boldsymbol{H}$—non-negative low-rank matrices, acting as factors during the approximation of $\mathcal{X}$. Iterative approach could be applied here as well, similar to that from Sect. 2.2.1, where either $\boldsymbol{W}$ or $\boldsymbol{H}$ is being fixed while the other is derived at iteration $t + 1$ from [12]:

$$H_{t+1} = argmin_{H \geq 0}\|\mathcal{X} - W_t^T H\|_M, \tag{2.12}$$

$$W_{t+1} = argmin_{W \geq 0}\|\mathcal{X}^T - H_t^T W\|_M. \tag{2.13}$$

The process continues until the following condition is satisfied [12]:

$$|f(\boldsymbol{W}_t, \boldsymbol{H}_t) - f(\boldsymbol{W}_{t+1}, \boldsymbol{H}_{t+1})| < \xi, \tag{2.14}$$

where the stop parameter, known also as precision reaches preliminary set value, lower than 1. Rank approximation could be employed in order to simplify the procedure of finding the final solution, according to [12]:

$$\begin{aligned} \min_{\boldsymbol{H}(l,j)\geq 0} \|\boldsymbol{Z}^{(j)} - \boldsymbol{W}_{(l)}^{T}\boldsymbol{H}_{(l.j)}\|_1 &= \left|\boldsymbol{W}_{(l,1)}\boldsymbol{H}_{(l.j)} - \boldsymbol{Z}_{(1,j)}\right| + \dots \\ &+ \left|\boldsymbol{W}_{(l,m)}\boldsymbol{H}_{(l.j)} - \boldsymbol{Z}_{(m,j)}\right| \triangleq \zeta_{(l.j)}\big(\boldsymbol{H}_{(l,j)}\big), \end{aligned} \tag{2.15}$$

where ζ represents linear functions, following in multiple segments, which in turn leads to the possibility to transform (2.15) into [12]:

$$\zeta_{(l,j)}(x) = \begin{cases} \Big(-W_{(l,i_{s1})} - \dots - W_{(l,i_{sq})}\Big)x + Z_{(i_{s1},j)} + \dots + Z_{(i_{sq},j)}, x \leq p_{s^1} \\ \Big(W_{(l,i_{s1})} - \dots - W_{(l,i_{sq})}\Big)x - Z_{(i_{s1},j)} + \dots + Z_{(i_{sq},j)}, p_{s^1} \leq x \leq p_{s^2} \\ \Big(W_{(l,i_{s1})} + \dots - W_{(l,i_{sq})}\Big)x - Z_{(i_{s1},j)} - \dots + Z_{(i_{sq},j)}, p_{s^{q-1}} \leq x \leq p_{s^q} \\ \Big(W_{(l,i_{s1})} + \dots + W_{(l,i_{sq})}\Big)x - Z_{(i_{s1},j)} - \dots - Z_{(i_{sq},j)}, p_{s^q} \leq x \end{cases} \tag{2.16}$$

and thus gives the opportunity for finding the final solution to the optimization problem. It is done on a row basis over $\boldsymbol{H}$ with the following stopping criterion [12]:

$$|f(\boldsymbol{W}, \boldsymbol{H}_{k+1}) - f(\boldsymbol{W}, \boldsymbol{H}_k)| \leq \epsilon, \tag{2.17}$$

where ϵ is another, prescaled, precision, again smaller than 1.

2.2.4 Incremental Non-negative Matrix Factorization

The Incremental Non-negative Matrix Factorization (iNMF) is another type of decomposition, successfully applied to the problem of background modeling and subtraction in video analysis [13]. $\boldsymbol{W}$ from (2.11) is also called mixing matrix and let it be of nxr real valued elements, while $\boldsymbol{H}$, also known as encoding matrix, let be of dimensions rxm, again all real numbers. The approximation, which the factorization process leads to, could be represented as [13]:

$$\mathcal{X} \approx \boldsymbol{W}\boldsymbol{H}, \tag{2.18}$$

and in contrast to ManhNMF, let here the second order norm is used as a cost function, which will provide the error from approximating the input matrix [13]:

$$F = \|\mathcal{X} - \boldsymbol{WH}\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(X_{ij} - (WH)_{ij}\right)^2, \tag{2.19}$$

where the current position of an element from the input matrix is denoted by the *i*-th row and the *j*-th column.

One possible solution to the problem, described by (2.19), is the use of multiplicative updates [13]:

$$H_{aj}^{t+1} = H_{aj}^{t}\frac{\left(\boldsymbol{W}^{t^T}\mathcal{X}\right)_{aj}}{\left(\boldsymbol{W}^{t^T}\boldsymbol{W}^{t}\boldsymbol{H}^{t}\right)_{aj}}, W_{ia}^{t+1} = W_{ia}^{t}\frac{\left(\boldsymbol{V}\boldsymbol{H}^{t+1^T}\right)_{ia}}{\left(\boldsymbol{W}^{t}\boldsymbol{H}^{t+1}\boldsymbol{H}^{t+1^T}\right)_{ia}}, \tag{2.20}$$

where *r* is the rank of the decomposition; *t*—the number of the current iteration; *a* = 1, 2, …, *r*; *i* = 1, 2, …, *n*; and *j* = 1, 2, …, *m*.

If *n* frames from a video are being used for background modeling and $\boldsymbol{C}$ is the covariance matrix of size *n*x*n*, then among the eigenvectors matrix **W**, the eigenvalues matrix **Λ** and **C**, the following relation holds [13]:

$$\boldsymbol{CW} = \boldsymbol{W\Lambda}. \tag{2.21}$$

Typically, only $r < n$ columns from $\boldsymbol{W}$ play a role into the consecutive calculations, since they correspond to the highest eigenvalues. The model of the background is represented by a data vector **v**′, which relates to the factorization parameters according to $\boldsymbol{h} = \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{v}' - \boldsymbol{\mu})$ [13]. In the last equation, $\boldsymbol{h}$ is a column from the encoding matrix and $\boldsymbol{\mu}$ is a mean vector. The foreground objects, mainly of interest to this study, are derived from finding the approximation error, given by $F = |\boldsymbol{v}' - \boldsymbol{Wh} + \mu|$ [13].

2.2.5 Fusion Schemes

Two schemes are proposed and tested within this study for fusion of the resulting segmentation results in the form of binary frames. The first one includes logical OR operation over corresponding pixels by spatial and temporal position (Algorithm 1).

```
// Algorithm 1 - Fusion OR
function out.avi = FusionOR (in.avi) {
  int i, j, n, P, Q, N;
  out1.avi = Decompose_by_RPCA();
  out2.avi = Decompose_by_NMF();
     i = 1; j = 1; n = 1;
     out.avi = out1.avi || out2.avi;
     do {
        if (out.avi(i,j,n) >= 1)
          out.avi(i,j,n) = 1;
        } while ((i <= P)&&(j <= Q)&&(n <= N)); }
```

In Algorithm 1 *i*, *j* and *n* represent the spatial coordinates of every pixel over the horizontals and verticals, and the consecutive number of frame, respectively. The input video in.avi is comprised of *N* number of frames, as well as the output video out.avi. The dimensions of frames are *P* by *Q* pixels. In a similar fashion the logical AND operator could be applied in order to get the second fusion scheme, shown as Algorithm 2 below.

Since the OR operator corresponds to union of the binary sets within the output fames, it is expected expansion of the boundaries of detected objects to happen, which will lead to higher detection rate, but also to higher level of the false positives. Contrary, the AND operator, corresponding to intersection, will lead to limiting the area of detected objects, less false detections, more precise location of emphasized moving areas, which in turn will generate higher precision of the detection process. Probability of increasing the number of false negatives in this case is higher.

```
// Algorithm 2 - Fusion AND
function out.avi = FusionAND (in.avi) {
  int i, j, n, P, Q, N;
  out1.avi = Decompose_by_RPCA();
  out2.avi = Decompose_by_NMF();
     i = 1; j = 1; n = 1;
     out.avi = out1.avi && out2.avi;
     do {
        if (out.avi(i,j,n) >= 1)
          out.avi(i,j,n) = 1;
        } while ((i <= P)&&(j <= Q)&&(n <= N)); }
```

2.3 Experimental Results

The experimental setup includes the following hardware platform—a desktop computer with Intel Core i5-3450 CPU, having 4 cores, working in hyper threading mode on a 3.1 GHz frequency, 12 GB of RAM and 2 TB 7200 rpm HDD. The software environment is supported by 64-bit Linux Ubuntu 14.04 LTS operating system, over which the Matlab R2016A simulation application is running and the base functions for decomposing the input videos are obtained from the LRS library [14]. The test video set is comprised of 6 videos with the following parameters: non-compressed RGB representation in AVI format with resolution, equal to 352 × 288 pixels, bit-depth of 8 bpp/component, frame rate of 10 frames/second. They are part of the LASIESTA test set [15] with the following naming convention: I_IL_01 (300 frames), O_CL_01 (250 frames), I_OC_2 (300 frames), I_SI_01 (220 frames), O_RA_02 (370 frames), O_SU_02 (400 frames). The O_CL_01 video contains recording of a car, which turns around into an underpass, while all the other 5 videos contain single or multiple persons, walking in closed or open environment, in some cases with changing illumination or along with the presence of shadows at natural sunlight.

The accuracy of the tested algorithms is evaluated by 3 parameters on a pixel basis: Detection Rate (*DR*), defined as the ratio of the correctly detected pixels as part of moving foreground objects to the sum of the same number and the number of correctly classified background pixels; *Precision*—the ratio of correctly detected pixels from foreground objects to the sum of them with the incorrectly detected pixels as part of objects but in reality, being part of the background; *F-measure*—double the product of *Precision* and *Detection Rate* over the sum of them. All of them are shown in Fig. 2.1. Processing Time (*PT*) reveals the execution time of the various tested decompositions per a pixel (sec/px), and together with the input–output operations for each case—the Full Time (*FT*), again in sec/px, is being measured.

The average (AV) and standard deviation (DEV) values of all accuracy related parameters—DR for *Detection Rate* (DRAV and DRDEV), PREC for *Precision* (PRECAV and PRECDEV), and F—for *F-measure* (FAV and FDEV) are presented in Fig. 2.1.

The average processing and full time (PT and FT) could be seen from Fig. 2.2.

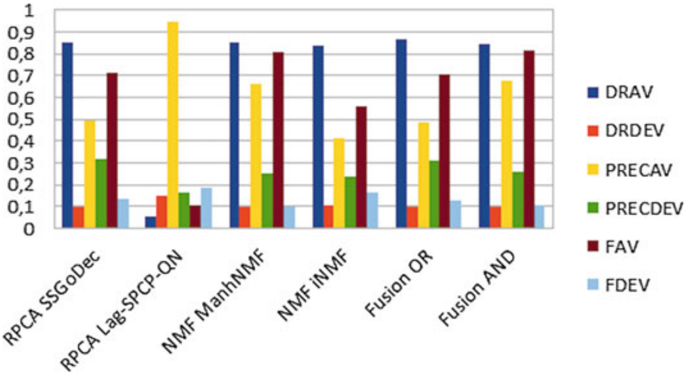
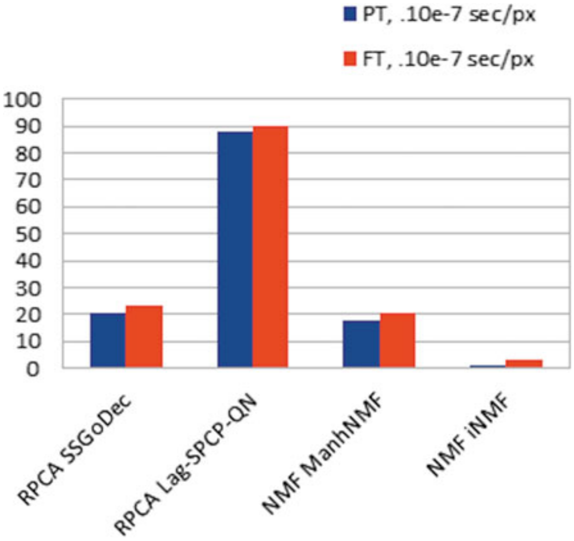


Fig. 2.1 Detection rate, precision and F-measure for all tested algorithms

Fig. 2.2 Processing and full times for the 4 base decompositions



Visual comparison of the segmentation results for a single frame from the 17th second of the video O_CL_01 is given in Fig. 2.3.

2.4 Discussion

The most accurate algorithm in terms of *DR* is ManhNMF with 0.8522, followed by the RPCA SSGoDec with 0.8516 (Fig. 2.1). The RPCA Lag-SPCP-QN has extremely low detection capability and its DR is just 0.0575, while iNMF is significantly accurate with 0.8393. These results are supported by the visual inspection of processed videos—in Fig. 2.3d, almost none of the pixels are marked as part of a moving object and the car is completely missing from the binary output, a product from RPCA Lag-SPCP-QN. Portions of the windows and the roof of the vehicle are missing in the resulting image from iNMF (Fig. 2.3f) and at the same moment these false negatives are smaller in number in Fig. 2.3e—corresponding to ManhNMF. The most of the surface of the moving object is preserved in Fig. 2.3c, a result from RPCA SSGoDec, but in contrast to all other 3 cases vast amount of false positives could be seen – closely positioned vehicles, which are stationary, portions from the nearby parking lot and other non-moving areas. This observation is in agreement with the lower *Precision* for that algorithm—0.4958. Only the iNMF has lower *Precision* of 0.4101. Obviously, the filtering abilities for these two algorithms need to be further enhanced. As an integral measure, embedding both the *DR* and *Precision*, the *F-measure* is highest for the ManhNMF—0.8087, followed by the RPCA SSGoDec with 0.7129. This is the main motivation to select these two algorithms for further fusion with the logical OR and AND operators in order to get additionally refined

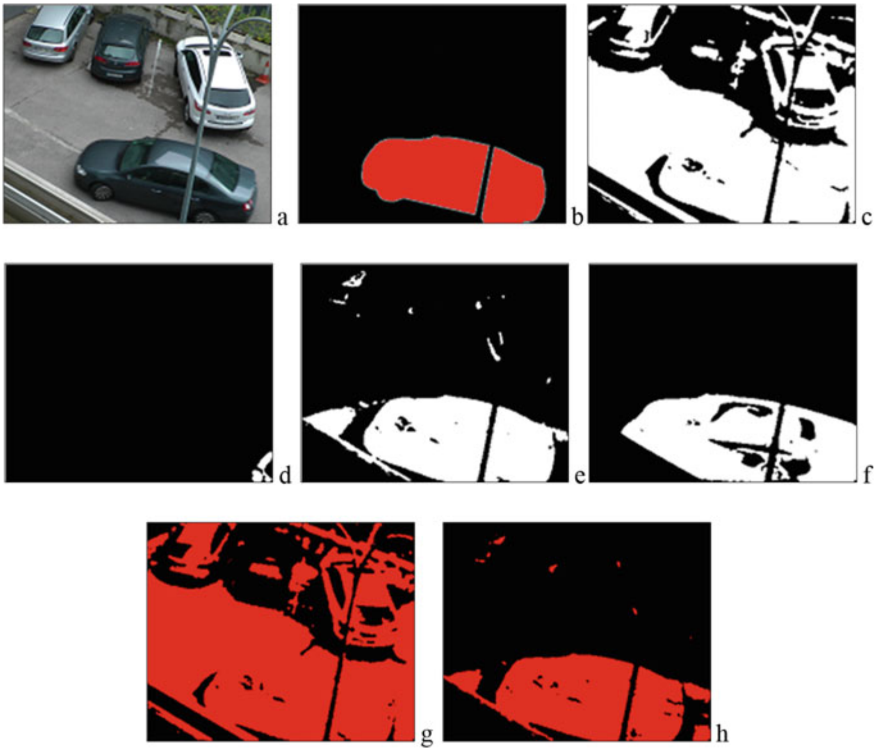


Fig. 2.3 Object detection in a single frame by: **a** original, **b** ground truth, **c** RPCA SSGoDec, **d** RPCA Lag-SPCP-QN, **e** NMF ManhNMF, **f** NMF iNMF, **g** Fusion OR, **h** Fusion AND

segmentation results of processing the videos. Although iNMF appears to be fastest with around 0.96 × 10^{-7} s/px, the selected two algorithms for fusion have moderate execution times of around 20 × 10^{-7} s/px, which is tolerable in terms of practical applications, and it is worth noting the considerably higher processing time of RPCA Lag-SPCP-QN of around 88 × 10^{-7} s/px. Full processing times on a pixel basis are close to the processing times, given the considerable amount of input data (the total amount of number of pixels), which makes the effect of reading and writing values to external drives dispersed as cumulative effect for the whole processing cycle.

The fusion of RPCA SSGoDec and ManhNMF by the logical OR operator leads to higher *DR* than any of the base decompositions—0.8687. The logical AND leads to *DR* of 0.8434, very close to the modestly performing iNMF. The benefit of using the AND is hidden behind the *Precision* it gets—0.6792—highest in comparison to any of the other segmentation techniques and as Fig. 2.3h reveals the false positives are extremely diminished in comparison to Fig. 2.3g, where the result from the logical OR fusion leads to vast incorrectly detected zones, although the moving car itself is almost entirely segmented as one object. It is also worth noting that occluding objects, such as the lamp post, and the edge of the parking lot overpass, are correctly marked as

Table 2.1 Segmentation accuracy comparison

Method	F-measure
Fusion OR	0.7089
Fusion AND	0.8168
RPCA PCP, [16]	0.7500

part of the background, nevertheless of their smaller width. That is a good indication of the selectivity of the tested algorithms, including those with the fusion process. The prevailing value for the F-measure of 0.8168 is observed for the AND fusion, while for the OR operation it is only 0.7089—lower than any of the 4 base algorithms. In terms of confidence and reliability, especially in the case of applications with high liability, e.g. security surveillance, monitoring critical industrial processes and others, it would be preferable to use the AND fusion between RPCA SSGoDec and ManhNMF algorithms. If an explicit aim is posed to get false negatives as low as possible, without considering the false positives, then the logical OR operator will be the obvious selection. Processing times for applying these logical operators are 0.64×10^{-7} s/px for the OR and 0.44×10^{-7} s/px—for the AND, which is not negligible in comparison with the time needed for decomposing the videos by the base algorithms.

Comparison between the proposed Fusion OR and Fusion AND with another implementation, based on RPCA Principal Component Pursuit (PCP) [16], is given in Table 2.1. Although close to RPCA PCP with regard to the *F-measure*, Fusion OR has lower overall performance with around 0.0411. Fusion AND is more accurate in both the detection rate and suppressing false positives than RPCA PCP, which is seen from the higher *F-measure* of around 0.0668. Despite the observed differences, both the Fusion OR and Fusion AND are considered applicable, depending on the particular scenario of their use, e.g. getting the maximum possible detected objects in the first case or getting minimal false positives in the second case.

2.5 Conclusion

In this paper two schemes are proposed for pixel based fusion of segmented videos by the RPCA and NMF decompositions with background subtraction. The logical OR operator in the first scheme leads to higher detection rate of moving objects in the video and in the same time to higher level of false positives. The logical AND operator limits the boundaries of detected objects, yielding higher precision of the segmentation process, but also increase the false negatives. As base decomposition algorithms the RPCA SSGoDec, Lag-SPCP-QN, ManhNMF and iNMF are tested of which most efficient turns out to be the RPCA SSGoDec and the ManhNMF. Processing times are small enough for the overall decomposition scheme to be practically implemented over proper hardware for particular applications, such as road traffic control, security surveillance and many others.

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. Javed, S., Mahmood, A., Al-Maadeed, S., Bouwmans, T., Jung, S.K.: Moving object detection in complex scene using spatiotemporal structured-sparse RPCA. IEEE Trans. Image Process. **28**(2), 1007–1022 (2018)
2. Cao, X., Yang, L., Guo, X.: Total variation regularized RPCA for irregularly moving object detection under dynamic background. IEEE Trans. Cybern. **46**(4), 1014–1027 (2015)
3. Javed, S., Mahmood, A., Bouwmans, T., Soon, K.J.: Superpixels based manifold structured sparse RPCA for moving object detection. In: International Workshop on Activity Monitoring by Multiple Distributed Sensing, BMVC 2017, Londres, United Kingdom, September (2017)
4. Li, Y., Liu, G., Liu, Q., Sun, Y., Chen, S.: Moving object detection via segmentation and saliency constrained RPCA. Neurocomputing **323**, 352–362 (2019)
5. Javed, S., Bouwmans, T., Sultana, M., Jung, S.K.: Moving object detection on RGB-D videos using graph regularized spatiotemporal RPCA. In: International Conference on Image Analysis and Processing, pp. 230–241, Springer, Cham (2017)
6. Chen, B.H., Shi, L.F., Ke, X.: A robust moving object detection in multi-scenario big data for video surveillance. IEEE Trans. Circuits Syst. Video Technol. **29**(4), 982–995 (2018)
7. Fan, X.N., Xue, R.Y., Shi, P.F., Li, M., Ni, J.J.: Moving object detection based on NMF and similarity analysis. Comput. Mod. **04** (2018)
8. Rezaei, B., Ostadabbas, S.: Moving object detection through robust matrix completion augmented with objectness. IEEE J. Sel. Top. Signal Process. **12**(6), 1313–1323 (2018)
9. Yang, X., Liu, D., Zhou, D., Yang, R.: Moving cast shadow detection using block nonnegative matrix factorization. Bull. Pol. Acad. Sci. Tech. Sci. **66**(2), 229–235 (2018)
10. Zhou, T., Tao, D.: Godec: randomized low-rank & sparse matrix decomposition in noisy case. In: Proceedings of the 28th International Conference on Machine Learning, ICML 2011, June 28–July 2, Bellevue, WA, USA, pp. 33–40. ACM, New York, USA (2011)
11. Aravkin, A., Becker, S., Cevher, V., Olsen, P.: A variational approach to stable principal component pursuit. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, July 23–27, Quebec City, Quebec, Canada, pp. 32–41. AUAI Press, Corvallis, Oregon, USA (2014)
12. Guan, N., Tao, D., Luo, Z., Shawe-Taylor, J.: MahNMF: Manhattan non-negative matrix factorization. arXiv preprint arXiv:1207.3438 (2012)
13. Bucak, S., Gunsel, B., Gursoy, O.: Incremental nonnegative matrix factorization for background modeling in surveillance video. In: 2007 IEEE 15th Signal Processing and Communications Applications, pp. 1–4. IEEE (2007)
14. Sobral, A., Bouwmans, T., Zahzah, E.H.: Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos. Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing. CRC Press, Boca Raton, FL, USA (2016)
15. Cuevas, C., Yáñez, E.M., García, N.: Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA. Comput. Vis. Image Underst. **152**, 103–117 (2016)
16. Guyon, C., Bouwmans, T., Zahzah, E.H.: Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis. Princ. Compon. Anal. **10**, 223–238 (2012)

Chapter 3
Hierarchical Decomposition of Third-Order Tensor Through Adaptive Branched Inverse Difference Pyramid Based on 3D-WHT

Roumen Kountchev and Roumiana Kountcheva

Abstract New approach is presented for adaptive decomposition of large-size tensors in the spectrum domain, based on the Three-Dimensional Adaptive Branched Inverse Difference Pyramid (3D-ABIDP). For this, the processed third-order tensor is divided into cubical sub-tensors of size 2^n and then each sub-tensor is transformed through 3D Inverse Difference Pyramid (3D-IDP) of n hierarchical levels, based on the 3D Walsh-Hadamard Transform (3D-WHT). The spectrum coefficients of same spatial frequency, calculated in the same hierarchical level of all pyramids, build new sub-tensors whose size is reduced 2^n times, compared to that of the initial tensor. In the next level, each new tensor is divided again into cubical sub-tensors of size 2^n, which are transformed into n-level 3D-IDP/WHT. The coefficients of same spatial frequency build new sub-tensors of size 2^{2n} times smaller than that of the initial tensor, and the processing continues in a similar way. The division of each sub-tensor stops when the so obtained new sub-tensors have at least one dimension equal to 2, or when all their coefficients are equal to zero. In this case, the initial tensor is represented as a tree-like graph, and the length and the number of its branches depend on the tensor contents. In general, this graph is an incomplete branched tree, whose low-information branches (i.e. the branches, whose coefficients are equal or close to zero), are cut-off. The offered method for hierarchical tensor decomposition has lower computational complexity, compared to well-known orthogonal 3D decompositions: Discrete Fourier Transform, Discrete Cosine Transform, Discrete Wavelet Transform, Contourlet Discrete Transform, etc. The presented decomposition is based on the use of 3D-WHT with frequency-ordered transform matrices, which enhances the concentration of the tensor energy into small number of coefficients and in result are defined faster than the branches, suitable to be retained. These qualities of the new decomposition open many possibilities for its practical application.

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

R. Kountcheva (✉)
TK Engineering, Druzhba 2, Bl. 404/2, Sofia 1582, Bulgaria
e-mail: kountcheva_r@yahoo.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_3

3.1 Introduction

Tensor decompositions became recently the object of numerous research works [1]. The main kinds of the tensor decompositions could be divided into two basic groups: statistical and deterministic. In the group of the *statistical methods* for tensor decomposition are various 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) [2–4]. In [5] is offered a version of HOSVD, namely the multi-way tensor SVD. 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) [3]; the Tensor Train Decomposition (TTD) [6]; the Kruskal decomposition, etc. The statistical methods are implemented through applying various algorithms for calculation of the tensors eigen vectors, which have relatively high computational complexity. The tensor decomposition components are usually calculated by using iterative methods whose iterations stop, when the predefined 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) [7], the Sequentially Truncated HOSVD (ST-HOSVD) [8], the Sequential Unfolding SVD (SUSVD) [9] and Compositional Hierarchical Tensor Factorization [10]. To same group also belongs the non-iterative Hierarchical SVD algorithm for tensor decomposition offered in [11]. 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* tensor decomposition methods are the pyramidal 3D transforms: the 3D Discrete Wavelet Transform (3D-DWT) [12], the 3D Curvelet and the 3D Contourlet Discrete Transform (3D-CDT) [13, 14] and the 3D Shearlet Discrete Transform (SDT) [15]. 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 (CC). The deterministic methods for tensor decomposition are usually executed by using various kinds of 3D orthogonal transforms. In publications [12, 14] 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 multi-level 3D Inverse Difference Pyramid (3D-IDP) and the 3D Walsh-Hadamard Transform (3D-WHT), presented in [16, 17]. The last-mentioned decomposition is not able to ensure sufficient decorrelation degree for its elements in the high hierarchical levels for high number of levels, i.e., for tensors of large size.

In this work is generalized the 3D-IDP tensor decomposition, called 3D Adaptive Branched IDP (3D-ABIDP). It is aimed at the achievement of high efficiency in the decomposition of large-size tensors (for example, sequences of 4K images) without significant increase of its CC.

3.2 Hierarchical Tensor Decomposition Through 3D Adaptive Branched IDP

The building unit in the offered decomposition is the n-level 3D-IDP, based on the 3D-WHT. The 3D-IDP/WHT pyramid is explained through an example for the hierarchical decomposition of the tensor X of size 8 × 8 × 8, for n = 3.

3.2.1 Hierarchical Decomposition for a Tensor of Size 8 × 8 × 8, Through 3D-IDP/WHT

The tensor X with elements x(i, j, k) and of size 8 × 8 × 8 could be represented through the 3-level 3D-IDP based on the Truncated 3D-WHT (3D-TWHT) for levels p = 0, 1, 2, as shown in Fig. 3.1.

In this case, the decomposed tensor **X** is presented as a sum of three tensors, $\tilde{\mathbf{X}}$, $\tilde{\mathbf{E}}_0$, $\mathbf{E}_1$, each of size 8 × 8 × 8 [15]:

$$\mathbf{X} = \tilde{\mathbf{X}} + \tilde{\mathbf{E}}_\mathbf{0} + \mathbf{E}_\mathbf{1} \tag{3.1}$$

where:

$\tilde{\mathbf{X}} = (1/8^3)\sum_{u=0}^{1}\sum_{v=0}^{1}\sum_{l=0}^{1} s(u,v,l)\mathbf{W}_{\mathbf{u},\mathbf{v},\mathbf{l}}$ is the tensor which is the first approximation of the input tensor;

$\mathbf{E}_0 = \mathbf{X} - \tilde{\mathbf{X}}$—the difference tensor, which represents the error of the first approximation;

$\tilde{\mathbf{E}}_0 = \bigcup_{t=1}^{8} \tilde{\mathbf{E}}_0^t$—tensor, which is the first approximation of the tensor $\mathbf{E}_\mathbf{0}$, after uniting the sub-tensors $\tilde{\mathbf{E}}_0^t$ each of size 4 × 4 × 4, for t = 1, 2, …, 8. All they are obtained through dividing the difference sub-tensor $\mathbf{E}_0$ into 8 sub-tensors. Here, each tensor $\tilde{\mathbf{E}}_0^t$ is defined by the relation:

$$\tilde{\mathbf{E}}_0^\mathbf{t} = \left(1/4^3\right)\sum_{u=0}^{1}\sum_{v=0}^{1}\sum_{l=0}^{1} s_0^t(u,v,l)\mathbf{W}_{\mathbf{u},\mathbf{v},\mathbf{l}}^t \quad \text{for } t = 1, 2, \ldots 8 \tag{3.2}$$

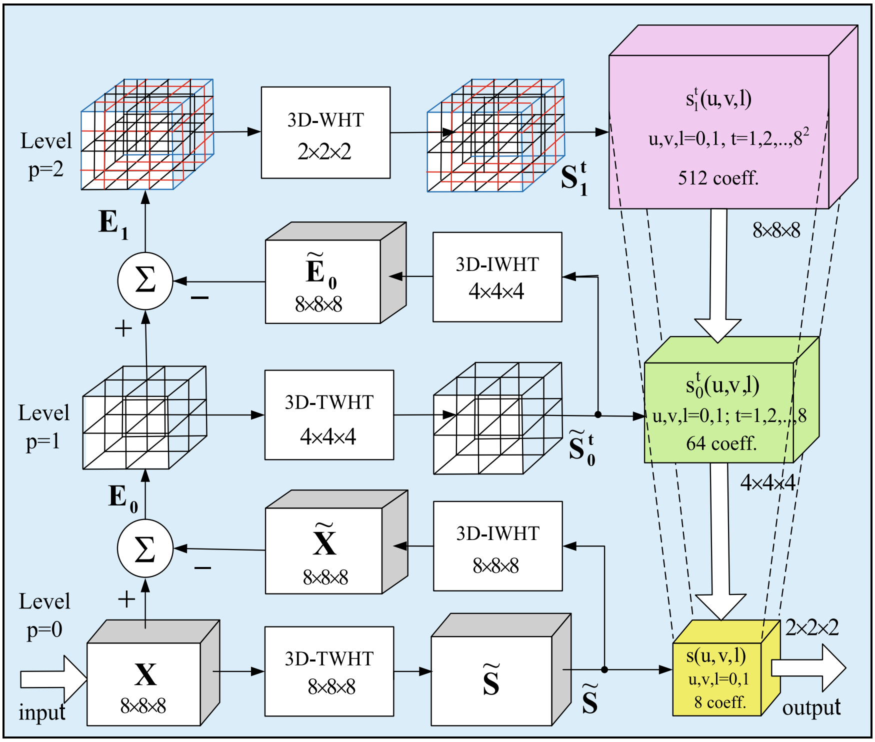


Fig. 3.1 3D-IDP for a tensor X of size 8 × 8 × 8

Here $s_0^t(u,v,l)$ are the coefficients of the direct 3D-WHT, applied on the elements of the tensor $\tilde{\mathbf{E}}_0^t$.

$$\mathbf{E_1} = \bigcup_{t=1}^{8} \mathbf{E_1^t}. \tag{3.3}$$

where

$$\mathbf{E}_1^t = \mathbf{E}_0^t - \tilde{\mathbf{E}}_0^t \quad \text{for}\, t = 1, 2, \ldots 8^2 \tag{3.4}$$

In the relations above, $\mathbf{W_{u,v,l}}$ is the basic tensor with frequency (u, v, l), which could be represented as the outer product of the vectors $\vec{w}_u$, $\vec{w}_v$, $\vec{w}_l$:

$$\mathbf{W_{u,v,l}} = \vec{w}_u \circ \vec{w}_v \circ \vec{w}_l. \tag{3.5}$$

Here the vectors $\vec{w}_u$, $\vec{w}_v$, $\vec{w}_l$, which represent the tensor $\mathbf{W_{u,v,l}}$, are defined by the relations below (Fig. 3.2):

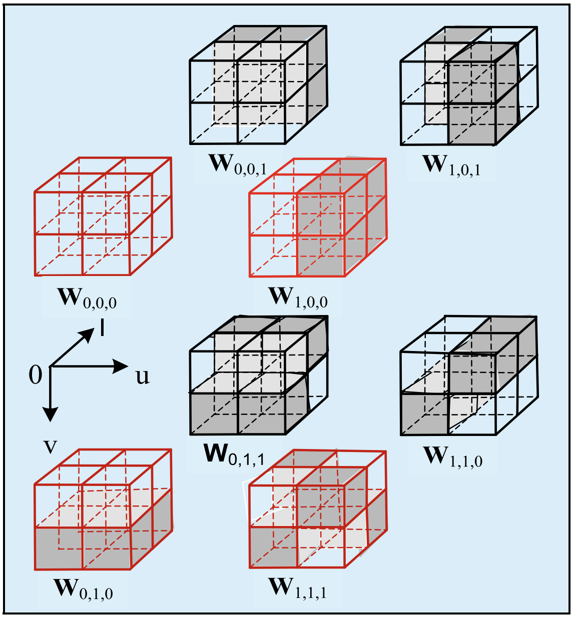


Fig. 3.2 Examples for the basic tensors $\mathbf{W}_{u,v,l}$ of size 4 × 4 × 4

$$\vec{w}_u = [(-1)^{\sum_{r=0}^{2} q_r(0)u_r}, (-1)^{\sum_{r=0}^{2} q_r(1)u_r}, \ldots, (-1)^{\sum_{r=0}^{2} q_r(7)u_r}]^{\mathrm{T}}; \tag{3.6}$$

$$\vec{w}_v = [(-1)^{\sum_{r=0}^{2} q_r(0)v_r}, (-1)^{\sum_{r=0}^{2} q_r(1)v_r}, \ldots, (-1)^{\sum_{r=0}^{2} q_r(7)v_r}]^{\mathrm{T}}; \tag{3.7}$$

$$\vec{w}_l = [(-1)^{\sum_{r=0}^{21} q_r(0)l_r}, (-1)^{\sum_{r=0}^{2} q_r(1)l_r}, \ldots, (-1)^{\sum_{r=0}^{2} q_r(7)l_r}]^{\mathrm{T}}. \tag{3.8}$$

The size of the basic tensors $\mathbf{W}_{\mathbf{u},\mathbf{v},\mathbf{l}}$ in Eq. (3.5) is 8 × 8 × 8, and of tensors $\mathbf{W}^{t}_{\mathbf{u},\mathbf{v},\mathbf{l}}$ in Eq. (3.2) when t = 1, 2, …, 8, it is 4 × 4 × 4, respectively. The 3D-WHT coefficients in levels p = 0, 1, 2 of 3D-IDP/WHT are defined by the relations:

$$\mathrm{s(u,v,l)} = \sum_{\mathrm{i=0}}^{7}\sum_{\mathrm{j=0}}^{7}\sum_{\mathrm{k=0}}^{7} \mathrm{x(i,j,k)wal(i,u,\ 8)wal(j,v,\ 8)wal(k,l,\ 8)} \text{ for } p = 0; \tag{3.9}$$

$$\mathrm{s}^{\mathrm{t}}_{\mathrm{p}}\mathrm{(u,v,l)} = \sum_{\mathrm{i=0}}^{2^{3-\mathrm{p}}-1}\ \sum_{\mathrm{j=0}}^{2^{3-\mathrm{p}}-1}\ \sum_{\mathrm{k=0}}^{2^{3-\mathrm{p}}-1} \tilde{\mathrm{e}}^{\mathrm{t}}_{\mathrm{p}}\mathrm{(i,j,k)\ wal(i,u,}2^{3-\mathrm{p}})\mathrm{wal(j,v,}\ 2^{3-\mathrm{p}})\mathrm{wal(k,l,}2^{3-\mathrm{p}}) \tag{3.10}$$

for t = 1, 2, …, $8^{\mathrm{p+1}}$ and p = 1, 2, where $\tilde{\mathrm{e}}_{\mathrm{p}}^{\mathrm{t}}(\mathrm{i,j,k})$ are the elements of the sub-tensors $\tilde{\mathbf{E}}_0^{\mathrm{t}}$. The Walsh-Hadamard (WH) functions which correspond to the frequency-ordered transform WH matrices, are defined in accordance with the relations:

$$\mathrm{wal}(\mathrm{i,u},8)\mathrm{wal}(\mathrm{j,v},8)\mathrm{wal}(\mathrm{k,l},8)=(-1)^{\sum\limits_{r=0}^{2}[q_r(i)u_r+q_r(j)v_r+q_r(k)l_r]}, \tag{3.11}$$

where

$\mathrm{i}=\sum\limits_{r=0}^{2}\mathrm{i}_r2^r,\ \mathrm{u}=\sum\limits_{r=0}^{2}\mathrm{u}_r2^r;\ \mathrm{j}=\sum\limits_{r=0}^{2}\mathrm{j}_r2^r,\ \mathrm{v}=\sum\limits_{r=0}^{2}\mathrm{v}_r2^r;\ \mathrm{k}=\sum\limits_{r=0}^{2}\mathrm{k}_r2^r,\ \mathrm{l}=\sum\limits_{r=0}^{2}\mathrm{l}_r2^r;$
$q_0(i)=i_2,\ q_1(i)=i_2\oplus i_1,\ q_2(i)=i_1\oplus i_0;$
$q_0(j)=j_2,\ q_1(j)=j_2\oplus j_1,\ q_2(j)=j_1\oplus j_0;$
$q_0(k)=k_2,\ q_1(k)=k_2\oplus k_1,\ q_2(k)=k_1\oplus k_0.$

In result of the 3-level 3D-IDP/WHT transform, the tensor **X** is represented as an inverse pyramid in the spectral domain. In the lowest pyramid level (p = 0), it is represented by the spectrum tensor $\tilde{\mathbf{S}}$ of size 2 × 2 × 2; in the next level (p = 1)—by the spectrum tensor $\tilde{\mathbf{S}}_{\mathbf{0}}$ of size 4 × 4 × 4, and in the last level (p = 2)—by the spectrum tensor $\mathbf{S_1}$, of size 8 × 8 × 8. The tensors $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{S}}_{\mathbf{0}}$ are the spectrum approximations of tensors $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{E}}_0=\bigcup\limits_{t=1}^{8}\tilde{\mathbf{E}}_0^{\mathrm{t}}$, and the tensor $\mathbf{S_1}$ is the spectrum transform of the tensor $\mathbf{E}_1$. The elements of the spectrum tensors are the coefficients s(u,v,l) and $\mathrm{s}_{\mathrm{p}}^{\mathrm{t}}(\mathrm{u,v,l})$, calculated in accordance with Eqs. (3.9) and (3.10). Each spectrum tensor $\tilde{\mathbf{S}}$, $\tilde{\mathbf{S}}_{\mathbf{0}}$, $\mathbf{S_1}$ is of size 2 × 2 × 2, $2^2\times2^2\times2^2$, and $2^3\times2^3\times2^3$ respectively.

In the general case, the number of coefficients in the levels p = 1, 2, …, n −1 of the n-level 3D-IDP/WHT could be reduced on the basis of the next relation [16, 17]:

$$\mathrm{s}_p^1(0,0,0)=-\sum_{t=2}^{8}\mathrm{s}_p^t(0,0,0)\quad\text{for t} - 1,2,\ldots 8^{p+1}. \tag{3.12}$$

Then, the number of the retained coefficients for the pyramid shown in Fig. 3.1 in the level p = 1 is 56, and in the level p = 2 it is 448, respectively. The so obtained pyramid with reduced coefficients $s_{\mathrm{p}}^1(0,0,0)$ is called 3D Reduced IDP/WHT (3D-RIDP/WHT) [16]. The number of operations, O (additions and multiplications), needed for the calculation, is defined by the relation [17]:

$$\mathrm{O}_{3\mathrm{D}-R\mathrm{IDP/WHT}}(\mathrm{n})\approx8^{\mathrm{n}}\times2.5\mathrm{n}. \tag{3.13}$$

The detailed comparison of the CC of 3D-RIDP/WHT with these of the orthogonal 3D transforms DFT, DCT, DWT and CDT, given in [17], proves its lower value. For the same value of n, the CC of 3D-FFT is minimum and is defined by the relation $\mathrm{O}_{3\mathrm{D-FFT}}(\mathrm{n})=8^{\mathrm{n}}\times7.5\,\mathrm{n}$. Hence, the CC of 3D-RIDP/WHT is three times lower than that of 3D-FFT.

3.2.2 Hierarchical Decomposition of a Tensor of Size M × N × P Through 3D Adaptive Branched IDP

The principle of the hierarchical decomposition of a 3D tensor of size M × N × P through branched IDP is shown in Fig. 3.3 for the case, when M = N = P = 16, and the 3D-IDP is built for sub-tensors of size 8 × 8 × 8 (n = 3). For the building of the 3D branched spectrum pyramid (3D-BIDP) should be defined the way used to calculate its branches. The process is illustrated for the tensor X of size 16 × 16 × 16. This tensor is divided into 8 sub-tensors of size 8 × 8 × 8, and each is represented as a 3D-IDP pyramid of n = 3 levels. In the initial level (p = 0) of each pyramid are calculated 8 spectral coefficients with frequencies (0, 0, 0) to (1, 1, 1), whose basic functions are shown in Fig. 3.2. From the group of coefficients of same frequency (u, v, l) for u, v, l = 0, 1, are created 8 tensors of size 2 × 2 × 2.

In Fig. 3.4 is shown the graph of the full tree for the level p = 0 in all 3D-IDP, which represents the tensor of size 16 × 16 × 16. This graph corresponds to the hierarchical tensor decomposition, shown in Fig. 3.3. From the tree branches 1, ..., 8, which correspond to sub-tensors of size 8 × 8 × 8, are obtained 64 branches (tensors), each of size 2 × 2 × 2. Then, the total number of branches of the full 3D-BIDP tree for the level p = 0 is $S_0 = 64 \times 8$; for the level p = 1 it is $S_1 = 64 \times 8^2$; and for the level p = 2, it is $S_2 = 64 \times 8^3$. Then, the total number of the branches in the full tree is:

$$N_B = 64(8 + 8^2 + 8^3) = 37376. \qquad (3.14)$$

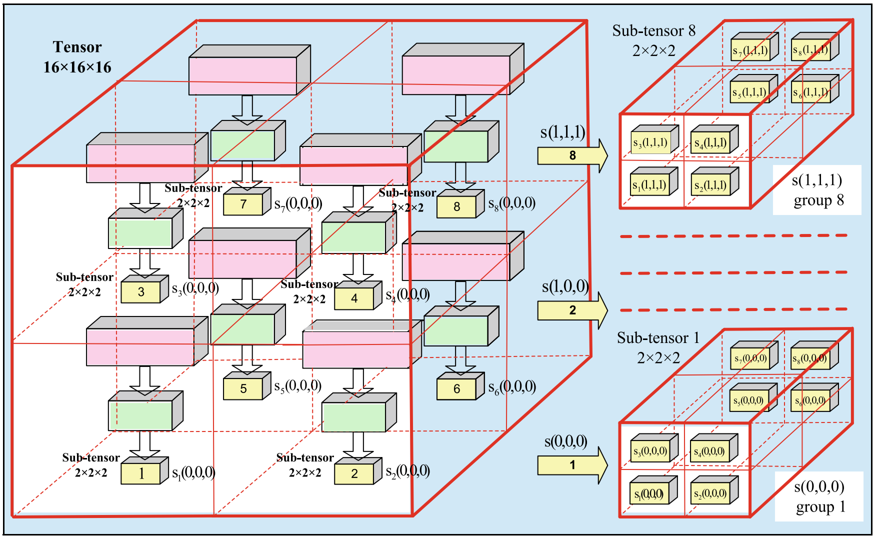


Fig. 3.3 Hierarchical decomposition for a tensor of size 16 × 16 × 16 through 3D branched IDP, based on the spectral coefficients in IDP level p = 0, using sub-tensors of size 8 × 8 × 8

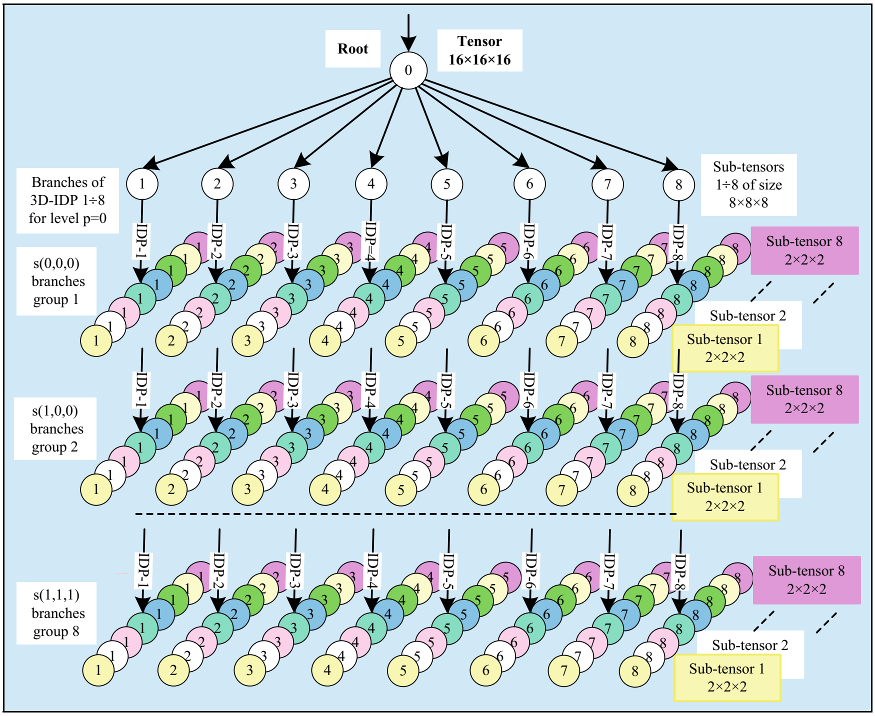

Root
Tensor 16×16×16
0
1
2
3
4
5
6
7
8
Branches of 3D-IDP 1÷8 for level p=0
Sub-tensors 1÷8 of size 8×8×8
s(0,0,0) branches group 1
IDP-1
IDP-2
IDP-3
IDP-4
IDP-5
IDP-6
IDP-7
IDP-8
Sub-tensor 8 2×2×2
Sub-tensor 2
Sub-tensor 1 2×2×2
s(1,0,0) branches group 2
s(1,1,1) branches group 8

Fig. 3.4 The 3D-BIDP graph of the full tree-like representation for a tensor of size 16 × 16 × 16

In the general case, for a tensor of size M × N × P, divided into sub-tensors of size $2^n \times 2^n \times 2^n$, the number of the branches of the full tree in the level k of the hierarchical tensor decomposition, taking into account the sum $S = \sum_{k=1}^{n} 8^k = (8/7)(8^n - 1)$, is:

$$\mathrm{N_{Bk}} = \lfloor 2^{-kn}\mathrm{M}\rfloor \lfloor 2^{-kn}\mathrm{N}\rfloor \lfloor 2^{-kn}\mathrm{P}\rfloor \times 8 \times \mathrm{S} \approx \left(8^2/7\right)\lfloor \mathrm{MNP} \times 8^{-\mathrm{nk}}\rfloor(8\mathrm{n} - 1) \tag{3.15}$$

for k = 1, 2, …, s; s—the number of the levels in the branched tensor decomposition.

Hence, the number of branches in the full tree for a tensor of size M × N × P built on the basis of the s-level 3D-BIDP taking into account that $\sum_{k=1}^{s} 8^{-kn} = \frac{1-8^{-sn}}{8^n-1}$, is defined by the relation:

$$N_B = \sum_{k=1}^{s} N_{B_k} = \sum_{k=1}^{s} \lfloor MNP \times 8^{-kn} \rfloor \times 8 \times S \approx [(8^2/7)(1 - 8^{-sn})]\,MNP \tag{3.16}$$

For example, for M = N = P = 16 and n = s = 3, from Eq. (3.16) it follows that for the full tree $N_B \approx 37{,}449$.

The branches of the full tree which could be cut-off, should satisfy at least one of the following requirements:

- to correspond to spectral coefficients equal to zero;
- the modules of the spectral coefficients in a given branch must be smaller than a predefined threshold value;
- the coefficients in a given branch should correspond to the spatial frequencies in which is concentrated the energy of the noise, contained in the decomposed tensor.

After applying the above criteria, the number of tree branches could be significantly reduced. The further reduction depends on the requirements imposed by the tensor decomposition application.

3.3 Algorithm for Third-Order Tensor Decomposition Through 3D ABIDP/WHT

The main steps of the 3D-ABIDP/WHT algorithm for tensor decomposition, are:

Start: input tensor **X** of size M × N × P, with non-negative elements x(i,j,k) and defined thresholds for decomposition branches truncation in accordance with coefficients energy and signal-to-noise relation.

Step 1. Divide the tensor **X** into sub-tensors $\mathbf{X}_k$ of size $2^n \times 2^n \times 2^n$, for k = 1, 2, …, K_1.

($K_1 = \lfloor MNP/8^n \rfloor$—total number of the sub-tensors, $\mathbf{X}_k$);

Step 2. Transform each sub-tensor $\mathbf{X}_k$ into the corresponding n-level 3D-IDP/WHT pyramid which comprises a sequence of spectral sub-tensors, as follows:

- in the level p = 0—the sub-tensor $\tilde{\mathbf{S}}_k$ of size 2 × 2 × 2, which comprises 8 coefficients $s_k(u, v, l)$, for u, v, l = 0, 1;
- in the level p = 1—the sub-tensor $\tilde{\mathbf{S}}_{\mathbf{0},k}$ of size $2^2 \times 2^2 \times 2^2$, which comprises 64 coefficients $s^t_{0,k}(u,v,l)$, for u, v, l = 0, 1 and t = 1, 2, …, 8;
- in the level p = n-1—the sub-tensor $\mathbf{S}_{n-2,k}$ of size $2^n \times 2^n \times 2^n$, which comprises 8^n coefficients $s^t_{n-2,k}(u,v,l)$ for u, v, l = 0,1 and t = 1, 2, …, 8^{n-1};

Step 3. Unite coefficients of same frequency (u, v, l) from each level p = 0, 1, …, n − 1 from all pyramids into corresponding spectrum sub-tensors $\mathbf{S}_k(1)$ of size $\lfloor M/2^n \rfloor \times \lfloor N/2^n \rfloor \times \lfloor P/2^n \rfloor$, for k = 1, 2, …, K_1;

Step 4. Transform each sub-tensor $\mathbf{S}_k(1)$ into the corresponding n-level 3D-IDP/WHT pyramid comprising a sequence of spectral sub-tensors in levels p = 0, 1, …, n − 1, in accordance with Step 2;

Step 5. Unite the coefficients of same frequency (u, v, l) from level p of all pyramids for the corresponding spectral sub-tensors $\mathbf{S}_k(2)$ of size $\lfloor M/2^{2n}\rfloor \times \lfloor N/2^{2n}\rfloor \times \lfloor P/2^{2n}\rfloor$, for k = 1, 2, …, $K_2 = \lfloor MNP/8^{2n}\rfloor$;

Step 6. Divide the sub-tensor $\mathbf{S}_k(m)$ which corresponds to a certain branch m, into sub-tensors $\mathbf{S}_k(m+1)$ of size $\lfloor M/2^{(m+1)n}\rfloor \times \lfloor N/2^{(m+1)n}\rfloor \times \lfloor P/2^{(m+1)n}\rfloor$ for k = 1, 2, …, $K_{m+1} = \lfloor MNP/8^{(m+1)n}\rfloor$;

Step 7. Stop the dividing of sub-tensors $\mathbf{S}_k(m+1)$ and cut-off the corresponding decomposition branches for the cases, when at least one of the following 4 conditions is satisfied:

- at least one of their dimensions is equal to 2;
- they are built of spectral coefficients, equal to zero;
- they are built of spectral coefficients, whose modules are lower than a predefined threshold value;
- they are built of coefficients with spectral frequencies, in which is concentrated the main part of the noise energy, in correspondence with the accepted statistical model for their distribution (respectively—the value of the relation signal/noise).

Step 8. Go to step 8 after the division of all decomposition branches is finished. Else, each branch corresponding to the sub-tensor $\mathbf{S}_k(m+1)$ which does not satisfy at least one of the conditions from Step 7, is not cut-off and for it are executed sequentially steps 1–5 by analogy with the processing of the input tensor, **X**;

Step 9. End.

As a result of the execution of the algorithm, presented above, is obtained a truncated tensor decomposition in the spectrum domain of 3D-WHT, from which the input tensor **X** could be restored with a predefined accuracy. The number of the 3D-AIDP/WHT levels is chosen in the range 2–4 depending on the data, contained in the input tensor (for example, a video sequence, a group of multispectral images, X-ray images, etc.). The so described algorithm could be also applied in the cases when instead of 3D-WHT is used some other famous deterministic 3D transform, for example, DCT, DST, DHT, etc.

3.4 Analysis of 3D ABIDP/WHT Properties

The main objective at which the 3D-ABIDP is aimed, is to enhance the efficiency of the use of the correlation between tensor elements, so that to achieve maximum energy concentration into minimum number of spectrum coefficients. The branched tree-like decomposition based on the 3D-ABIDP/WHT, differs from 3D-IDP/WHT in the following:

- it ensures better decomposition efficiency for tensors of large size, due to the low CC of 3D-RIDP/WHT pyramid for small number of levels (n = 2/4), used to transform each sub-tensor (decomposition branch);

- it ensures higher energy concentration in the spectrum sub-tensors of reduced size (branches of the tree-like structure), built by the spectrum coefficients of same spatial frequency (u, v, l).

The properties of 3D-ABIDP/WHT depend on the configuration of the truncated tree, which represents the input tensor. The number of the cut-off tree branches grows together with the energy concentration in the retained branches (respectively—the retained spectrum coefficients). The energy concentration is highest in the group of retained spectrum coefficients with spatial frequency (0, 0, 0) in level p = 0 of each 3D-IDP/WHT which represent the corresponding sub-tensor. The sub-tensors built of such coefficients, are copies of the input tensor (the root of the tree), but their size is reduced.

The use of the new 3D-WHT decomposition with frequency-ordered transform matrices leads to additional concentration of the tensor energy into small number of spectrum coefficients in the initial levels of each inverse pyramid. The degree of the tensor energy concentration in selected spectrum coefficients depends on the correlation between its elements.

One important property of the non-negative tensor decomposition through adaptive 3D-BIDP/WHT, is the low computational complexity. The detailed analysis given in [17] proves that the CC of 3D-RIDP/WHT (which is the basic building unit of the branched adaptive pyramidal decomposition) is lower than that of the famous 3D orthogonal transforms: DFT, DHT, DCT, DWT, CDT and SDT.

The CC of 3D-BIDP/WHT is defined by the product of the CC for 3D-RIDP/WHT in accordance with Eq. (3.13), and the number of the branches in the full tree-like structure from Eq. (3.16), i.e.:

$$\mathrm{O}_{3\mathrm{D}-B\mathrm{IDP/WHT}}(\mathrm{n}) \approx 8^{\mathrm{n}} \times 2.5\mathrm{n} \times \mathrm{N}_{\mathrm{B}} = 8^{n} \times 22.8\,n \times MNP. \tag{3.17}$$

The value of $\mathrm{O}_{3\mathrm{D}-B\mathrm{IDP/WHT}}$ for the case when the 3D-BIDP/WHT is based on the 3D-RIDP/WHT (whose CC is the lowest, compared to the famous deterministic 3D orthogonal transforms), is minimum too.

The adaptation of 3D-ABIDP/WHT toward the values of the tensor elements and the reduction of the part of the coefficients of the basic 3D-RIDP/WHT permits significant reduction of the branches in the tree-like structure after the "truncation". In result is achieved lower CC of the decomposition, compared to that of the full tree, defined by Eq. (3.17).

3.5 Conclusions

In this work is offered a new approach for adaptive hierarchical decomposition of third-order tensors in the spectrum area of 3D-WHT, which has higher efficiency in respect of the tensor energy concentration into a small number of spectrum coefficients, and together with this, ensures low CC. These advantages of the offered tensor decomposition open many possibilities for its application in various areas, such as: compression of sequences of correlated images, improvement of their quality through filtration and contrast enhancement, analysis and pattern recognition, accelerated search in databases of tensor images, analysis of multidimensional tensor signals, etc.

The future development of the offered decomposition is aimed at the investigation of the possibilities for its application in the statistical orthogonal transforms (the tensor KLT, and SVD), and also, in the neural networks with deep learning, for the optimization of the tree-like structure truncation.

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

1. Hao, N., Horesh, L., Kilmer, M.: Nonnegative tensor decomposition. In: Carmi, A., Mihaylova, L., Godsill, S. (eds.) Compressed Sensing & Sparse Filtering, pp. 123–148. Springer, Heidelberg (2014)
2. Bergqvist, G., Larsson, E.: The higher-order singular value decomposition: theory and an application. IEEE Signal Process. Mag. **27**(3), 151–154 (2010)
3. Kolda, T., Bader, B.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
4. 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)
5. Kilmer, M., Horesh, L., Avron, H., Newman E.: Tensor-tensor products for optimal representation and compression. arXiv:2001.00046v1 [math.NA] (2019)
6. Oseledets, I.: Tensor-train decomposition. SIAM J. Sci. Comput. **33**(5), 2295–2317 (2011)
7. Grasedyck, L.: Hierarchical singular value decomposition of tensors. SIAM J Matrix Anal Appl **31**(4), 2029–2054 (2010)
8. 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)
9. Salmi, R., Richter, A., Koivunen, V.: Sequential unfolding SVD for tensors with applications in array signal processing. IEEE Trans SP **57**(12), 4719–4733 (2009)
10. Vasilescu, M., Kim, E.: Compositional hierarchical tensor factorization: representing hierarchical intrinsic and extrinsic causal factors. In: The 25th ACM SIGKDD Conf. on KNOWLEDGE DISCOVERY and DATA MINING (KDD'19): Tensor Methods for Emerging Data Science Challenges, Anchorage, USA (2019)
11. Kountchev, R., Kountcheva, R.: 3D image representation through hierarchical tensor decomposition based on SVD for elementary tensor of size 2×2×2. Int. J. WSEAS Trans. SP **12**, 199–207 (2016)
12. Sayood, K.: Mathematical preliminaries for transforms, subbands, and wavelets. In: Dierna, A., Kaufmann, M. (eds.) Introduction to Data Compression (2012)
13. Candes, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. Multiscale Model. Simul. **5**(3), 861–899 (2006)
14. Rao, K., Kim, D., Hwang, J.: Fast Fourier Transform: Algorithms and Applications. Springer, Heidelberg (2010)

15. Goossens, B., Luong, H., Aelterman, J., Pizurica, A., Philips, W.: Efficient multiscale and multidirectional 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)
16. 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)
17. Kountchev, R., Kountcheva, R.: Comparative analysis of the hierarchical 3D-SVD and reduced inverse tensor pyramid in regard to famous 3D orthogonal transforms. In: Kountchev, R., Mironov, R., Li, S. (eds.) Proceedings of NAMSP 2020. Springer Nature Singapore, pp. 35–56 (2021)

Chapter 4
Multimodal Technique for Human Authentication Using Fusion of Palm and Dorsal Hand Veins

Mona A. Ahmed, Mohamed Roushdy, and Abdel-Badeeh M. Salem

Abstract Multimodal biometric systems have been widely used to achieve high recognition accuracy. This paper presents a new multimodal biometric system using an intelligent technique to authenticate human by fusion of palm and dorsal hand veins pattern. We developed an image analysis technique to extract region of interest (ROI) from palm and dorsal hand veins image. After extracting ROI we design a sequence of preprocessing steps to improve palm and dorsal hand veins images using Homomorphic, Median filter, Wiener filter and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance vein image. Our smart technique is based on the following intelligent algorithms, namely; principal component analysis (PCA) algorithm for feature extraction and k-Nearest Neighbors (K-NN) classifier for matching operation. The database chosen was the CASIA Multi-Spectral Palmprint Image Database V1.0 and Bosphorus Hand Vein Database. The achieved result for the fusion of both biometric traits was Correct Recognition Rate (CRR) is 97.6% with FAlse Reject Rate (FRR) 2.4%.

4.1 Introduction

Recently, authentication is considered the most important objective to be satisfied whether it is physical world or the internet of things (IoT) world. Different approaches and techniques exist to authenticate the user such as passwords, smart cards, and pins. Modern approaches to authentication include biometrics like voice, finger

M. A. Ahmed (✉) · A.-B. M. Salem
Faculty of Computer and Information Sciences, Computer Science Department, Ain Shams University, Cairo, Egypt

A.-B. M. Salem
e-mail: absalem@cis.asu.edu.eg

M. Roushdy
Faculty of Computer and Information Technology, Computer Science Department, Future University, New Cairo, Egypt
e-mail: mohamed.roushdy@fue.edu.eg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_4

prints, retina, iris, facial expressions, signatures, face and vein patterns. Among all the authentication techniques present, biometrics is considered as the most reliable authenticator since they are unique to every individual and hard to get [1].

Biometric authentication is a process of identifying a person using physiological or behavioral features. Physiological features are Iris, DNA, hand, finger print and face behavioral features are voice, signature, password, keystroke, etc. Among all the authentication techniques present, biometrics is considered as the most reliable authenticator since they are unique to every individual and hard to get. The technology of Vein Patterns (VP) as a type of biometric authentication was first intended in 1992.VP is the network of blood carriers below a person's skin layers. VP structure distinct and distinguishable patterns across various people and they remain the same irrespective of age. The patterns of blood veins are unique to each person, even among twins. There are internal and external biometric systems. External include face, iris, finger print based systems. Palm vein, finger vein, dorsal veins structure the internal biometric frameworks. Veins are intra-skin elements, consequently this feature makes the frameworks exceptionally secure, and they are not influenced by state of the external skin [2]. Generally, biometric system works in two modes namely: (i) verification mode in which biometrics can be used to verify a person's identity and (ii) identification mode in which biometrics can be used to determine a person's identity, even without that individual's information [3].

Hand vein technology works by identifying the vein patterns (palm, dorsal hand and finger veins) in an individual's hand. When a user's hand is held over a scanner, a near-infrared light maps the location of the veins. The red blood cells present in the veins absorb the rays and show up on the map as black lines, whereas the remaining hand structure shows up as white. This vein pattern is then verified against a preregistered pattern to authenticate the individual [4].

Biometric authentication can be classified into unimodal and multimodal biometric systems. Unimodal systems that use single biometric trait for recognition purposes; and suffers several practical problems like non-universality, noisy sensor data, intra-class variation, restricted degree of freedom, unacceptable error rate, failure-to-enroll and spoof attacks. So, the performance of single biometric system needs to be improved. The techniques of multimodal biometric system can offer a feasible method to solve the problems coming from unimodal biometric system. Multimodal biometric system makes use of different biometric traits simultaneously to authenticate a person's identity. Robustness and high security of authentication can be achieved by using the multimodal biometric systems [5].

The rest of the paper is organized as follows. Section 4.2 presents a comprehensive analysis for the multimodal recognition techniques and systems. We presented an overview of various levels of fusion in multimodal biometrics systems in Sect. 4.3. We briefly explain our methodology of the fusion of palm and hand vein system in Sect. 4.4. Section 4.5 presents the explanation of the process of biometric system. The discussion of results is introduced in Sect. 4.8. Finally; conclusions and future work are presented in Sect. 4.9.

4.2 Analysis of Multimodals Recognition Techniques and Systems

There are many researches in the last years in the field of multimodal biometric recognition system. The table below combines the most researches and compares preprocessing, feature extraction, matching methods, database size and the recognition percentage (Table 4.1).

From the analysis of the above research, it can report the following important results.

1. Most of the models have a very high rate of acceptability (accept to use) in hand (palm, dorsal hand and finger) and the low rate be in possession of the iris model. Hand veins have high level of accuracy and security because it prevented inside the body so there is no influence by state of the external skin. Hand veins are the best model to use because of high rate of acceptability, accuracy and security.
2. There are many algorithms used in researches based on the model used and the quality of images such as Gabor filter, LBP, median filter, PCA and SIFT algorithms. The most algorithms used in pattern veins are PCA, LBP and SIFT because they generate a vector of features that represent the highest detailed variant information.
3. There are many algorithms used in researches based on the model used and features extracted from images such as Euclidean distance, SVM, Hamming Distance, Naive Bayes and K-NN classifier. The most algorithms used in pattern veins are K-NN and Euclidean distance classifier. K-NN classifier works better because it performs much better if all of the data have the same scale, works well with a small number of input variables and when the number of inputs is very large, makes no assumptions about the functional form of the problem being solved, calculation time is very small and it has high accuracy.
4. There are various levels of fusion feature, score and decision level. Fusion at feature extraction level is most effective and hardest to perform simultaneously because features collected from various identifiers must be independent and in same measurement scale which would represent an identity in more discriminating feature space. Matching score level fusion is preferred as it is easy to obtain and combine matching scores of different biometrics but it is more of complexity. The complexity that comes from matching scores cannot be used or combined directly; because these scores are from different modalities and based on different scaling methods. Score normalization is required, by converting the scores into common similar domain or scale. Decision level fusion is very easy to implement and has high accuracy but it needs more time than other levels of fusion.
5. Only few multimodal databases are available publicly. BANCA and XM2VTS include face and voice biometrics. BIOMET which includes face, voice, fingerprint, hand and signature. BIOSEC includes fingerprint, face, iris and voice. SDUMLAHMT is a homologous database which includes face images from 7

Table 4.1 Analysis of multimodal recognition techniques

Ref No	Modalities fused	Method of feature extraction	Method of matching	Fusion levels	Database size	Recognition percentage
[4]	Palm vein + dorsal hand vein	Proposed filter	Euclidean distance	Feature level	2400 images 250 images	High FMR
[5]	Finger vein + fingerprint	Gabor filters	Hamming distance	score level	6264 images	98.78%
[6]	Face + iris	LBP Dauman's	LBPH Hamming distance	Decision level	400 images 200 images	77%
[7]	Palm veins + hand geometry	proposed algorithm HOG	SVM	Feature level	7200 images	98.7%
[8]	Finger vein + finger print	SMR	Weighted sum rule	score level	1500 images	99.22%
[9]	Finger vein + hand vein	MLBP	IGMF	score level	3916 images 4846 images	98%
[10]	Palm print + fingerprint	Gabor filters	Euclidean distance	Feature level	250 images	87%
[11]	Palm print + palm vein	Gabor filter	Euclidean distance	Feature level	28 images	100%
[12]	Hand Vein + hand geometry	median filter	Euclidean distance	Score level	300 images	99.94%
[13]	Palm Print + palm vein	Wavelet packet tree	K-NN Naive Bayes	Feature level	2400 images	95.95%
[14]	Finger vein + iris	Current tracking point	Hamming distance	score level	120 images 140 images	92.40%
[15]	face + finger veins	PCA and LDA	Euclidean distance	score level	210 images 105 images	91.4%
[16]	face + fingerprint	Gabor filter	PCA	Feature level	400 images 136 images	98.11%

(continued)

Table 4.1 (continued)

Ref No	Modalities fused	Method of feature extraction	Method of matching	Fusion levels	Database size	Recognition percentage
[17]	Iris + finger vein	SIFT	SVM SIFT	Score level	756 images 756 images	98%

angles, finger print images, gait videos, iris images. But these databases have some limitations. Homologous multi-biometrics dataset should be complete (contains all the biometrics for large population) for future research testing and multi-biometric system evaluation.

6. Percentage of recognition in multi-biometric systems that contain one or more trait veins has higher rate compared with other multi-biometric systems because of advantages of veins model accuracy and security.

4.3 Intelligent Fusion Levels and Techniques in Biometric Systems

H. S. Ali and M. I. Abdalla [18] have presented an overview of multimodal biometrics and have proposed various levels of fusion, various possible scenarios, the different modes of operation, integration strategies and design issues. The fusion levels proposed for multimodal systems are shown in Fig. 4.1 and described below.

A. Fusion at the Feature Extraction Level

The data obtained from each sensor is used to compute a feature vector. As the features extracted from one biometric trait are independent of those extracted from the other, it is reasonable to concatenate the two vectors into a single new vector. The primary benefit of feature level fusion is the detection of correlated feature

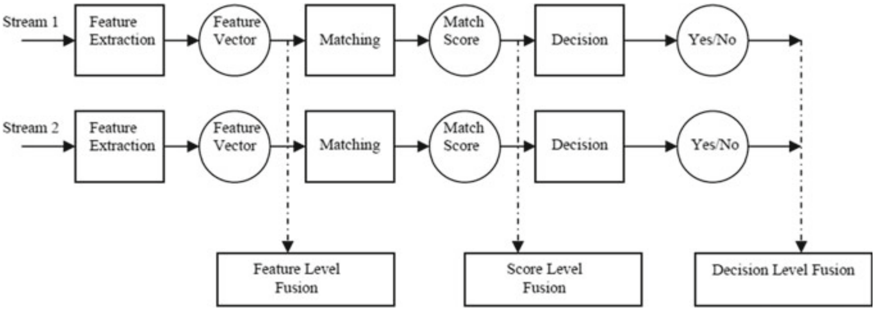


Fig. 4.1 Fusion levels in multimodal biometric system

values generated by different feature extraction algorithms and improved recognition accuracy. The new vector has a higher dimension and represents the identity of the person in a different hyperspace. Eliciting this feature set typically requires the use of dimensionality reduction/selection methods and, therefore, feature level fusion assumes the availability of a large number of training data.

B. **Fusion at the Matching Score Level**

Feature vectors are created independently for each sensor and are then compared to the enrollment templates which are stored separately for each biometric trait. Each system provides a matching sore indicating the proximity of the feature vector with the template vector. These individual scores are finally combined into a total score (using maximum rule, minimum rule, sum rule, etc.) which is passed to the decision module to assert the veracity of the claimed identity. Score level fusion is often used because matcher scores are frequently available from each vendor matcher system and, when multiple scores are fused, the resulting performance may be evaluated in the same manner as a single biometric system. The matching scores of the individual matchers may not be homogeneous. For example, one matcher may output a similarity measure while another may output a dissimilarity measure. Further, the scores of individual matchers need not be on the numerical scale. For these reasons, score normalization is essential to transform the scores of the individual matchers into a common domain before combining them. Common theoretical frameworks for combining classifiers using sum rule, maximum and minimum rules are analyzed, and have observed that sum rule outperforms other classifiers combination schemes.

C. **Fusion at the Decision Level**

A separate identification decision is made for each biometric trait. These decisions are then combined into a final vote. The fusion process is performed by a combination algorithm such as AND, OR, etc. Also a majority voting scheme can be used to make the final decision.

4.4 Proposed Methodology

In our study, we present a proposed intelligent paradigm to authenticate personal based on fusion of palm and dorsal hand veins. This paradigm is used to enhance the accuracy of vein authentication. Figure 4.2 shows the methodology of the authentication model using fusion of palm and dorsal hand veins biometrics. The proposed multimodal biometric system consists of several different submodules, each of them providing its own functionality. There are two sensor modules for palm and dorsal hand veins acquisition, which capture the biometric data. In the feature extraction modules, the acquired data is processed to extract a set of features. In the matcher modules, the extracted features are compared against the stored templates, providing a matching score. These last modules encapsulate the decision-making modules, which can operate either in verification or identification mode. Moreover, there is

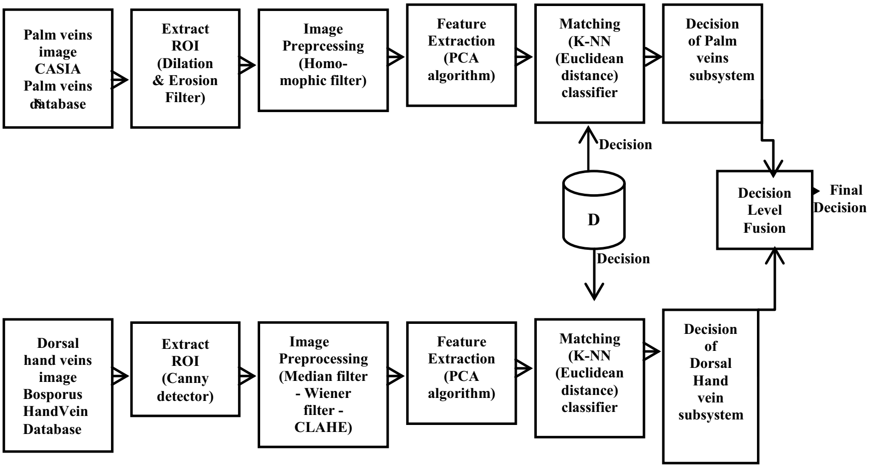

Palm veins image CASIA Palm veins database
Extract ROI (Dilation & Erosion Filter)
Image Preprcessing (Homo-mophic filter)
Feature Extraction (PCA algorithm)
Matching (K-NN (Euclidean distance) classifier
Decision of Palm veins subsystem
Decision
D
Decision
Decision Level Fusion
Final Decision
Dorsal hand veins image Bosporus HandVein Database
Extract ROI (Canny detector)
Image Preprocessing (Median filter - Wiener filter - CLAHE)
Feature Extraction (PCA algorithm)
Matching (K-NN (Euclidean distance) classifier
Decision of Dorsal Hand vein subsystem

Fig. 4.2 The methodology of the authentication model using palm and hand veins biometrics

the system database module, which stores the biometric templates of the enrolled users.

4.5 Process of Biometric System

In this section we describe the recognition process of palm and hand veins characteristics. The process of biometric system involves: image acquisition, extract ROI, preprocessing, feature extraction, matching, decision of each trait and fusion decision.

4.5.1 Image Acquisition System

(A) Palm Veins Database

The experiment reported in this paper for the palm vein authentication is CASIA Multi-Spectral Palmprint Image Database V1.0 (CASIA database) [19]. This CASIA database has been acquired using a contactless imaging device and has images from 100 users. Six images were acquired from each user and these images were acquired in two different data acquisition sessions (three images in each session) with a minimum interval of one month. Since our work is focused on palm vein identification and the vascular details are typically observed in the NIR illumination, only the images that were acquired under 850 and 950 nm wavelength illuminations from CASIA database were utilized in the following experiments.

(B) **Dorsal Hand Veins Database**

The experiments reported in this paper for the hand vein authentication of the Bosphorus Hand Vein Database [20] is designed for research on biometry based dorsal vein patterns of the hand. The hand vein data is captured using NIR imaging innovation with a monochrome NIR CCD camera (WAT-902H2 ULTIMATE) equipped with an infrared lens. The back of the hand is illuminated by two IR light sources. The images have 300 × 240 pixel size with a gray-scale resolution of 8-bit. Every subject experienced four imaging sessions that comprised of the left hand under normal condition (N), after having carried a bag weighing 3 kg for one minute (B), after having squeezed an elastic ball repetitively for one minute (A), after having cooled the hand by holding an ice pack on the surface of the back of the hand (I). We used the images taken under normal conditions (N: Normal). There are overall 600 images of 100 subjects distributed as: Three right-hand images and three left-hand images subject under normal conditions (N).

4.5.2 Extract ROI

(A) **ROI of Palm Veins**

To detect ROI we used morphological operations to extract useful structural information from palm veins images. Morphological operations are applied on binary images and affect the form, structure or shape of an object. They are used in pre or post processing (filtering, thinning, and pruning) or used for smoothing, edge detection or extraction of other features. Morphological operations offer a variety of image transformation to eliminate dark (bright) regions from binary images. The two principal morphological operations are dilation and erosion [21]. Dilation allows objects to expand, thus potentially filling in small holes and connecting disjoint objects. Erosion shrinks objects by etching away (eroding) their boundaries. These operations can be customized for an application by the proper selection of the structuring element, which determines exactly how the objects will be dilated or eroded [22]. The proposed algorithm of ROI extraction of hand vein image includes 5 tasks, as shown in Fig. 4.3.

1. Convert image to binary
2. Estimates the area of the palm in binary image then applies a 201*201 square mask that could perfectly cover the whole region of palm.
3. After then apply the dilatation filter again to get one point that is the middle point of the hand.
4. Then apply the erosion filter on the same square mask, this time to get exact square placed at same point where the region of interest is placed in actual image
5. Then find xmin, ymin, length, and width of this square to crop ROI from original image.

(B) **ROI of Dorsal Hand Veins**

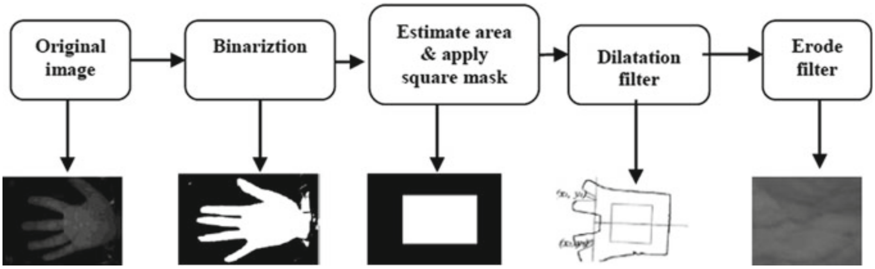


Fig. 4.3 The steps to detect ROI of palm veins

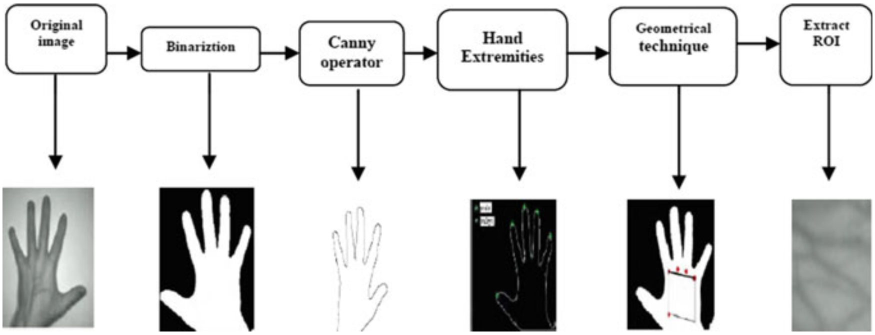


Fig. 4.4 The steps to detect ROI of dorsal hand veins

To detect ROI we used canny edge detector (CED) to extract useful structural information from hand veins images. The edge detection is an important process in many of the image processing algorithms. Significant property of the edge detection is the detection of the specific edges along with the great orientation of the object in the image [23]. The proposed algorithm of ROI extraction of hand vein image includes 5 tasks, as shown in Fig. 4.4.

1. Convert image to binary.
2. Boundaries from the binary image are detected by canny operator.
3. Valleys of hand between index and middle fingers and between little and ring fingers are detected.
4. A geometrical technique is investigated to draw the line connecting the two key points determined in the previous step and the line perpendicular to it.
5. A sub-image is detected and extracted as the ROI of hand vein image.

4.5.3 *Image Preprocessing*

(A) **Preprocessing of Palm Veins**

Homomorphic filtering is a generalized technique for image enhancement and/or correction. It simultaneously normalizes the brightness across an image and increases contrast. Homomorphic filter is a nonlinear enhancement method. Homomorphic filter simultaneously normalizes the brightness across an image and increases contrast. The function of homomorphic filter is likely to decrease the low frequency and increase the high frequency. In general, an image can be regarded as a two-dimensional function of the form $I(x, y)$, whose value at spatial coordinates (x, y) is a positive scalar quantity whose physical meaning is determined by the source of the image. The Homomorphic filtering can be summarized in steps show following:

1. An image $I\ (x, y)$ can be expressed as the product of illumination and reflectance components:

$$I(x,\ y) = L(x,\ y)R(x,\ y) \tag{4.1}$$

where $L(x, y)$ and $R(x, y)$ stand for the illumination and reflectance components.

2. Because the Fourier transform of the product of two functions is not separable, we define

$$Z(x,\ y) = \ln I(x,\ y) = \ln L(x,\ y) + \ln R(x,\ y) \tag{4.2}$$

Or

$$Z\,(u, v) = F\{Z(x, y)\},\ I(u, v) = F\{I(x, y)\},\ R\,(u, v) = \ F\{R(x, y)\} \tag{4.3}$$

where $F\{.\}$ is the operator for the 2D discrete Fourier transform.

3. Doing the Fourier transform, as

$$S(u,\ v) = \ H(u,\ v)Z(u,\ v)$$

$$= \ H\,(u,\ v)\,I\,(u,\ v)\ +\ H\,(u,\ v)\,R\,(u,\ v) \tag{4.4}$$

where H is a high-pass filter given by:

$$H(u, v) = \frac{1}{1 + \left[\left[\frac{D_0}{D(u,v)}\right]\right]^{2n}} \tag{4.5}$$

where D_0 is the cutoff amplitude in wavelet domain, n is the order of filter and $D(u, v)$ is the amplitude at location (u, v):

$$\mathrm{D(u, v)} = \sqrt{\left(u + \frac{M}{2}\right)^2 + \left(v + \frac{N}{2}\right)^2} \tag{4.6}$$

where $M * N$ is the size of image.

4. Taking inverse Fourier transform of S (u, v) brings the result back into natural log domain

$$\mathrm{S(x, y) = F^{-1}\{S(u, v)\}}$$

$$\mathrm{\{H(u, v)I(u, v)\} + F^{-1}\{H(u, v)R(u, v)\}} \tag{4.7}$$

5. So the output image can be expressed by the function [24]

$$g(x, y) = \text{antilog}[S(x, y)] = e^{s}(s(x, y)) \tag{4.8}$$

Figure 4.5 shows the result of applied algorithm (a) the original image (b) the result of extract region of interest (ROI) then (c) the result of applied preprocessing step to enhance the image quality.

(B) **Preprocessing of Dorsal Hand Veins**

In this process, a number of preprocessing techniques are typically required for the purpose of reducing the effect of noise and enhancing the targeted hand veins. The proposed algorithm of preprocessing hand vein image includes the following steps done as shown in Fig. 4.6.

1. The median filter 5 * 5 is applied to the original hand vein image for denoising.

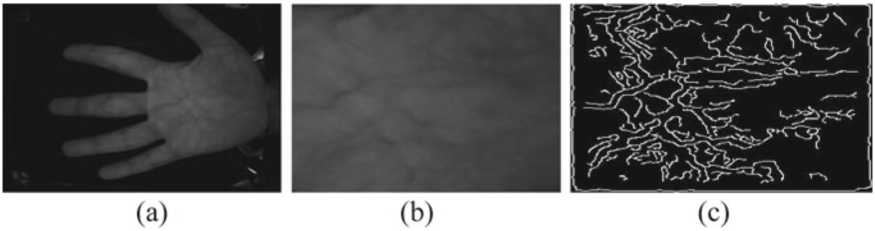


Fig. 4.5 Illustration of image enhancement: **a** the original image, **b** extraction of ROI, **c** extraction of palm vein pattern

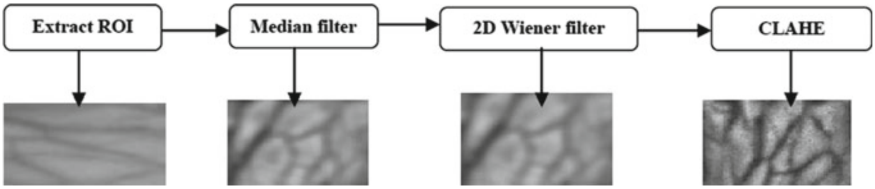


Fig. 4.6 The steps of preprocessing of dorsal hand veins

2. 2D Wiener filter “Gaussian white noise” 3 * 3 is applied to remove the effect of high-level frequency noise.
3. Applied Contrast Limited Adaptive Histogram Equalization (CLAHE) filter to enhance hand vein image.

4.5.4 Feature Extraction

Feature extraction plays an important role in palm vein recognition because the performance of feature matching is greatly influenced by its output. We use principal component analysis (PCA) algorithm to extract features from image [25]. This algorithm is used for extracting features from palm vein images. PCA is applied to generate a vector of features that represent the highest detailed variant information. A matching process is then applied to find the best match from the database to recognize and authenticate the person. It is one of the most widely implemented tools for dimensionality reduction or data exploration used in a variety of scientific and engineering disciplines. It transforms a number of possibly correlated variables into a smaller number of new variables, known as principal components. Since a digital image can be regarded as a two—or more—dimensional function of pixel values and represented as a 2D or 3D data array, PCA can be performed on such an m × n matrix [26].

The algorithm

1. Assume data matrix is B of size m × n. Compute mean μ_i for each dimension.
2. Subtract the mean from each column to get A
3. Compute covariance matrix C of size n × n which C= A^T A
4. Calculate the eigenvalues and eigenvectors (E, V) of the covariance matrix C
5. Project the data step by step onto the principal components $\vec{v_1}$,$\vec{v_2}$,……, etc.
6. Select n eigenvectors that correspond to the largest n eigenvalues to be the new basis.

4.6 Matching

In our technique, we use the K-NN classifier. The nearest neighbor classifier works depending on a simple nonparametric decision. Every query image Iq is inspected depending on the distance of its features from the features of other images in the database. The nearest neighbor is the image which has the minimum distance from the query image in the feature space. The distance between two features can be computed depending on one of the distance functions such as city block distance d_1, and Euclidean distance d_2 or cosine distance d_{cos} [27].

$$d_1(x, y) = \sum_{i=1}^{N} |x_i - y_i| \tag{4.9}$$

$$d_2(x, y) = \sqrt{\sum_{\mathrm{v}_1^{\rightarrow}, \mathrm{v}_2^{\rightarrow} i=1}^{N} \left|(x_i - y_i)^2\right|} \tag{4.10}$$

$$d_{cos}(x, y) = 1 - \frac{\overrightarrow{x}\,.\,\overrightarrow{y}}{|x|.|y|} \tag{4.11}$$

K nearest neighbor algorithm utilizes K nearest samples to the query image. Every one of these samples belongs to a known class Ci. The query image Iq is arranged to the class CM which has the most of events among the K samples. The presentation of the K-NN classifiers is highly related to value of the k, the number of the samples and their topological distribution over the feature space.

4.7 Fusion Decision

The proposed multimodal biometric system relies on two different modules: the module for Palm veins and the module for dorsal hand veins authentication. The fusion methodology adopted at the decision level is a post-classification method, and it follows the AND rule; i.e., it is sufficient that all biometric traits are recognized as genuine to lead to a positive final decision. This serial matching approach gives the possibility of acquiring all the traits to determine if a user is genuine or an impostor. From a numeric value (generally normalized between 0 and 1) that represents the confidence of the matching, each decision module is given two possible different outputs {YES, NO}, depending on the comparison of that value with some predefined thresholds that divide the interval [0, 1]. A decision module outputs the YES value if the obtained score is the interval [1] and the user is recognized as one of the enrolled users (in identification mode) or their claimed identity has been confirmed (in verification mode). The output value NO is produced if the obtained score is one of intervals [0, 1] or [1, 0] or [0, 0] and the user is rejected as if they were impostors.

4.8 Results and Discussion

In this section we describe the result of each system independently and the result of fusion of two traits. Palm and dorsal hand vein recognition includes training and recognition phases. In training phase, features of the training samples are calculated and stored in a database template. In the recognition phase, features of the input vein are determined and then matched by using K-NN matching classifier. After this, these features are compared with the stored template to obtain the recognition result. We do our experiment by dividing the database to 5 Cases as Table 4.2 shows.

By applying the PCA algorithm with K-NN (Euclidean distance) the results are 100, 98.5 and 100% for all training cases in palm, dorsal hand veins and fusion

Table 4.2 Database for 5 cases

Case No	Training	Testing
1	One image for every person (100 images)	Five images for every person (500 images)
2	Two images for every person (200 images)	Four images for every person (400 images)
3	Three images for every person (300 images)	Three images for every person (300 images)
4	Four images for every person (400 images)	Two images for every person (200 images)
5	Five images for every person (500 images)	One image for every person (100 images)

Table 4.3 The testing results for each case

Case No	CRR		
	Palm	Dorsal hand	Fusion palm + dorsal hand
1	94	89	96.5
2	94	90	96.6
3	95	92	97.8
4	96	92	98.3
5	97	93	99

of palm and dorsal hand veins. Testing results of every case are shown in Table 4.3 and Fig. 4.7. We have two potential results, the first result is where the user is unauthorized which means that his/her template is not found in the database, and the other result is the user is authorized, i.e. a template similar to his/her is found in the database. The experimental results show that the results of Correct Recognition Rate (CRR) are 95.20, 91.2 and 97.6% with False Rejected Rate (FRR) 2.4%. Based

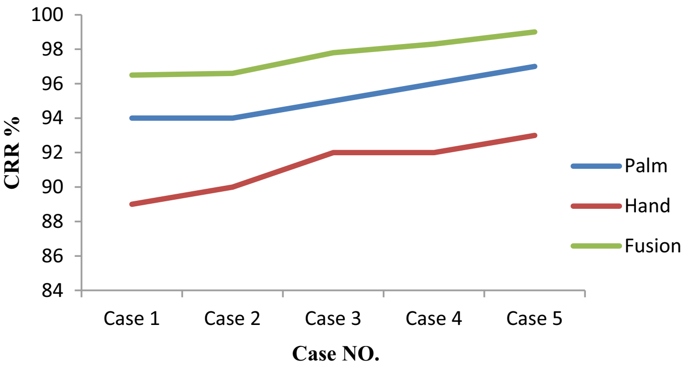


Fig. 4.7 Result of cases

on this experiment, it was suggested that recognition based on authentication by fusing the palm and dorsal hand veins performs better than conventional recognition technique. Hence this method can be successfully used for recognition.

4.9 Conclusion and Future Work

In this paper, we have developed a new practical and intelligent technique for biometric recognition based on fusion of palm and dorsal hand veins. The technique consists of the following steps: Image acquisition, determining the region of interest and preprocessing, extracting the finger vein pattern features and recognition. We proposed an original method based on the principal component analysis (PCA) algorithm to extract features and using K-NN (Euclidean distance) matching classifier in matching. In addition, this smart technique has many advantages and characteristics of flexibility of the former approaches; such as it can overcome the problem of rotation and shift, accurate, simple, practical and fast. In this paper, a complete biometric system based fusion of palm and dorsal hand veins has been developed. We proposed an original method based on the PCA algorithm to extract features and using K-NN classifier in matching. The experimental results show that the result of recognition CRR is 97.6%. Hence this method can be successfully used for recognition. The vein pattern identification can proceed in a perfect way using the method proposed in this paper which is accurate, simple, practical and fast.

In our opinion, this developed improvement increases the usefulness and usability of this efficient technique, especially as regards its application in all security tasks and domains. Future work may involve applying additional/ alternative pattern recognition algorithms or turning it into a multimodal system where other additional biometrics traits are considered and making the system more invariant to illumination conditions.

References

1. Mubeen, F.: Securing the biometric template: a survey. Int. J. Comput. Appl. **160**(2), 1–8 (2017)
2. Sumit, K., Manali, P.: Biometric recognition system based on dorsal hand veins. Int. J. Innov. Res. Sci. Eng. Technol. (IJIRSET) **5**(9), 18899–18905 (2016)
3. Laxmi, M., Kalpana, J.: A survey on biometric template protection. Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. (IJSRCSEIT) **2**(2), 995–999 (2017)
4. Shruthi, B.M., Pooja, M., Mallinath, Ashwin, R.G.: Multimodal biometric authentication combining finger vein and finger print. Int. J. Eng. Res. Dev. **7**(10), 43–54 (2013)
5. Bharathi, S., Sudhakar, R.: Biometric recognition using dorsal and palm vein images. Int. J. Adv. Eng. Technol. **7**(2), 415–419 (2016)
6. Stefani, E., Ferrari, C.: Design and implementation of a multi-modal biometric system for company access control. In: 2nd International Conference on Data Compression, Communication, Processing and Security (CCPS), pp.1–10 (2016)

7. Christo, L.E., Zimmer, A.: Multimodal biometric system for identity verification based on hand geometry and hand palm's veins. In: Communication Papers of the Federated Conference on Computer Science and Information Systems, vol. 13, pp. 207–212 (2017)
8. Raghavendra, R., Kiran Raja, B., Surbiryala, J., Busch, C.: A low-cost multimodal biometric sensor to capture finger vein and fingerprint. In: IEEE International Joint Conference on Biometrics, December (2014)
9. Trabelsi, B.R., Masmoudi, D.A., Masmoudi, S.D.: A new multimodal biometric system based on finger vein and hand vein recognition. Int. J. Eng. Technol. (IJET) **5**(4), 3175–3183 (2013)
10. Dhameliya, D.M., Chaudhari, P.J.: A multimodal biometric recognition system based on fusion of palmprint and fingerprint. Int. J. Eng. Trends Technol. (IJETT) **4**(5), 1908–1911 (2013)
11. Siddharth, A.J., Prabha A.H., Srinivasan, J.T., Lalithamani, N.: Palm print and palm vein biometric authentication system. Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing book (2017)
12. Park G.T., Soowon, K.: Hand biometric recognition based on fused hand geometry and vascular patterns. US National Library of Medicine National Institutes of Health, pp. 2895–2910 (2013)
13. Usharani, V., Saravanan, V.S.: Multi modal biometrics using palmprint and palmvein. J. Theor. Appl. Inf. Technol. **67**(1), 177–185 (2014)
14. Mohammed, E.F., ALdaidamony, M.E., Raid, M.A.: Multi model biometric identification system: finger vein and iris. **4**(4), 50–55 (2014)
15. Razzak, I.M., Yusof, R., Khalid, M.: Multimodal face and finger veins biometric authentication. Sci. Res. Essays **5**(17), 2529–2534 (2010)
16. Ayodeji Makinde, S., Yaw, N.-G., Loserian Laizer, S.: Enhancing the accuracy of biometric feature extraction fusion using Gabor filter and Mahalanobis distance algorithm. Int. J. Comput. Sci. Inf. Secur. (IJCSIS) **12**(7), 41–48 (2014)
17. Mohammed, E.F., ALdaidamony, M.E., Raid, M.A.: Iris and finger vein multi model recognition system based on sift features. Int. J. Intell. Comput. Inf. Sci. (IJICIS) **15**(1), 15–24 (2015)
18. Ali, S.H., Abdalla, I.M.: Score-level fusion for efficient multimodal person identification using face and speech. Int. J. Comput. Sci. Inf. Secur. (IJCSIS) **9**(4), 48–53 (2011)
19. http://www.cbsr.ia.ac.cn/MS_Palmprint
20. http://bosphorus.ee.boun.edu.tr/hand/Home.aspx
21. Jankowski, M.: Erosion, dilation and related operators. In: 8th International Mathematica Symposium, pp. 1–10, Avignon (2006)
22. Van Droogenbroeck, M., Buckley, M.: Morphological erosions and openings: fast algorithms based on anchors. J. Math. Imaging Vis. 1–35 (2005)
23. Amruta Kabade, L., Sangam, G.V.: Canny edge detection algorithm. Int. J. Adv. Res. Electron. Commun. Eng. (IJARECE) **5**(5), 1292–1295 (2016)
24. Liu, W., He, P., Li, H., Yu, H.: Improvement on the algorithm of homomorphic filtering. In: International Conference on Electrical and Computer Engineering Advances in Biomedical Engineering, vol. 11, pp. 120–124 (2012)
25. Halko, N., Martinsson, P.-G., Shkolnisky, Y., Tygert, M.: An algorithm for the principal component analysis of large data sets. SIAM J. Sci. Comput. **33**(5), 1–12 (2011)
26. Hladnik, A.: Image compression and face recognition: two image processing applications of principal component analysis. Int. Circ. Graph. Educ. Res. 6 (2013)
27. Muhammad Mansor, N., Yaacob, S., Muthusamy, H., Shafriza, N., Hi-fi, S.J., Mohd, M.L., Rejab, M.N., Ibrahim, K.Y.K., Syam, H.A., Jamaluddin, A., Ahmad, K.J.: PCA-based feature extraction and k-NN algorithm for early jaundice detection. Int. J. Soft Comput. Softw. Eng. (JSCSE) **1**(1), 25–29 (2011)

Chapter 5
SIFT Based Feature Matching Algorithm for Cartoon Plagiarism Detection

Dongxing Li, Jiazheng Gong, and De Li

Abstract The rise of digital technology has injected new vitality into the development of the animation industry. However, the problem of copyright infringement of cartoon images has also become a major obstacle to its development. The theoretical defects of the current law, the concealment of infringement forms, and the low cost of infringement are the main reasons for this dilemma. With the rapid development of Internet information and digital image processing technology, the use, acquisition, transmission and exchange of image information has become more and more convenient. Large-scale digital images will appear on the Internet and in human life all the time. This topic intends to study the extraction process and matching process based on SIFT feature descriptors, and analyze the advantages and problems of the algorithm at the same time, and finally to propose an improvement method for the lack of color information in the SIFT algorithm. Applying the image grayscale algorithm to the first step of the SIFT algorithm, the image is first converted from the RGB color space to the HSV color space, then is calculated the chromaticity difference between adjacent pixels, and finally is performed the chromaticity difference iterative optimization to obtain the final grayscale image.

5.1 Related Knowledge Introduction

5.1.1 Overview of Image Classification

The basic concept of image classification is to use a computer to divide the image into its corresponding semantic categories based on the content contained in the image. This classification technology can automatically understand the content of the image to a certain extent, transform the digital image into a conceptual model that people can understand, and is an important way to automatically extract the semantic

D. Li · J. Gong · D. Li (✉)
Department of Computer Science, Yanbian University, Yanji, China
e-mail: leader1223@ybu.edu.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_5

content of the image. Image classification is an intersecting research direction applied to multiple fields, including computer vision, image processing, machine learning algorithms, and data mining [1].

The image features represent the essential attributes or original features of the image. Every image has its corresponding characteristics, such as brightness, shape, edge, color or texture. These features are all natural features that can be directly felt by vision, and some image features can be obtained only through measurement calculation or conversion, such as feature histograms and spectra. The basis of image classification lies in the extraction and representation of image features [2]. The basic task of feature extraction and selection is to select the feature with the best classification effect as the classification feature. The selected feature should have the following characteristics: first, it can completely express the semantic information of the image, and secondly, it should have a certain degree of stability and robustness to interference factors such as noise. Therefore, the selection of features is very critical. Improper feature selection will cause inaccurate classification and even result in failure to classify.

5.1.2 SIFT Algorithm Overview

Over the years, Lowe et al. proposed the classic SIFT algorithm. This method is used to extract local features of an image. The local features include the following characteristics: high stability, high adaptability, strong distinguishability, and strong resistance to attack [3]. The characteristics of the above SIFT guarantee the effect of this local feature on image classification.

The process of SIFT descriptor formation mainly includes the following five steps: Color-to-grayscale conversion; Scale-space extreme point detection; Key point localization; Key point Orientation assignment and Key point descriptor. The key step is to use this local feature point to describe the information of its surrounding area, which can reduce the impact of key points on viewing angle, rotation, lighting, etc. By assigning the orientation of the key points, we have been able to get the main orientation of the key point. Then we rotate the area to the main direction within a certain radius with the key point as the center, so that the key point has rotation invariance (Fig. 5.1).

5.1.3 Limitations of the SIFT Algorithm

Although the SIFT algorithm is resistant to scale, rotation, and brightness transformation and has high robustness, it has many limitations.

(1) To simplify the computation of the SIFT algorithm, the first step is to convert the input color image into a grayscale image [4]. Such conversion will cause

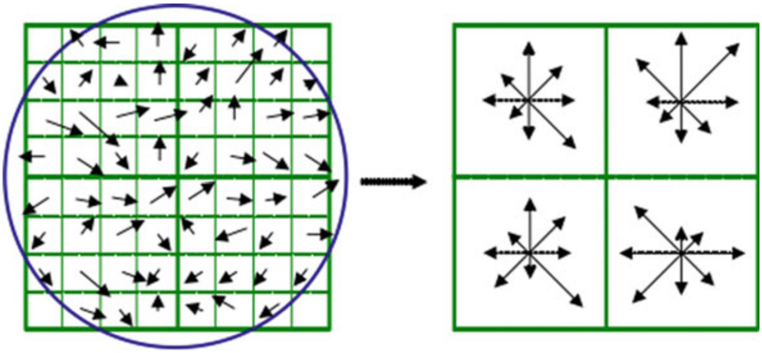


Fig. 5.1 Key point description diagram

the color information in the color image to be lost. At the same time, there are different colors in the image but corresponding to grayscale. Regions with the same degree value cannot be distinguished, and feature points cannot be extracted from them. Therefore, the correct matching rate will also decrease [5]. The SIFT feature calculation method has certain shortcomings. The description part in SIFT is based on gray gradient, single description of image features, statistical feature local area information on a scale, and the feature points detected by SIFT still have some redundancy.

(2) The SIFT algorithm has very few correct matches for two images with a large difference in viewing angle, that is, its anti-affine transformation performance is not very good.

5.2 Improved SIFT Image Matching Algorithm

Moments, histograms, and SIFT algorithms can all get color descriptors, and which generation method to use needs to be determined according to specific application scenarios. The distribution of the local luminosity information and the color space information of the image can be reflected by the color moment, and the color histogram does not contain the color space information of the image [6]. The SIFT color descriptor loses the local luminosity information of the image, and only contains the local spatial information. The SIFT color feature point descriptor is generated in each color space through the SIFT algorithm. The generation process is the same as that of the SIFT gray feature point descriptor. Therefore, it is completely invariant to image rotation and scale transformation. But for color images, lighting changes have a great impact on them. After adding color information, the SIFT algorithm can no longer avoid the impact of lighting changes on the matching results. In order to improve the precision of the SIFT algorithm for matching color images and make the extracted color feature point descriptors have complete color invariance, the SIFT algorithm based on color information has been developed. J Li in the image

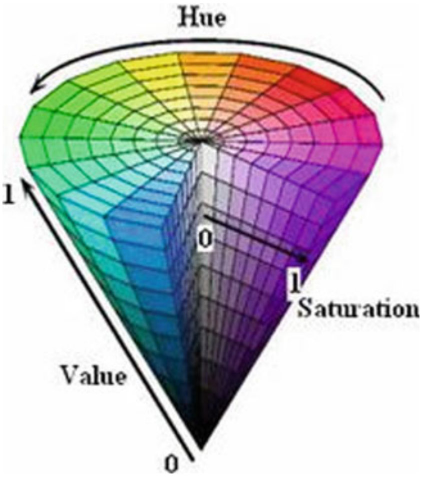


Fig. 5.2 HSV color space model

matching algorithm, explained that the PCA-SIFT algorithm uses principal component analysis [7, 8] for the feature descriptors in the image; this algorithm can play the role of dimensionality reduction and reduce the amount of computation, which can significantly improve matching efficiency [9].

5.2.1 Color SIFT Descriptor Method

HSV-SIFT: Bosch et al. proposed a method of extracting color SIFT descriptors in HSV space: similar to extracting SIFT feature points in gray space, they are extracted from the three channels of H, S, and V in HSV color space 128-dimensional descriptor, and then connect these three 128-dimensional descriptors to form a 3 × 128-dimensional color descriptor. Bosch A et al. proposed that scene classification by pLSA can greatly improve the matching accuracy [10]. The method of extracting SIFT feature points in each channel is consistent with the method of extracting SIFT in grayscale images. Song X et al. proposed an affine transformation between image space and color space for invariant local descriptors [11]. The experimental results prove that the invariance of the SIFT descriptor inherited by the color descriptor can only be partially invariant to the brightness, and it can only produce better results when the brightness changes slightly (Fig. 5.2).

5.2.2 Overall Process

The main idea of the SIFT algorithm is to transform matching between images into matching between feature vectors. The following diagram shows the flow chart of

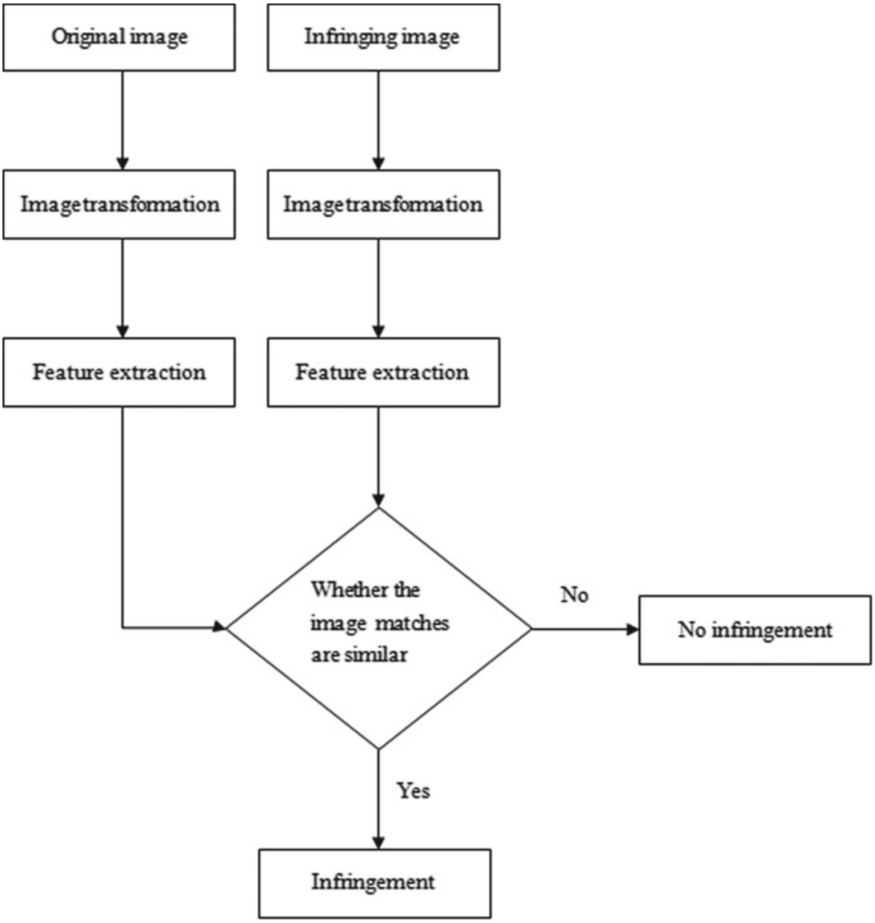


Fig. 5.3 Image matching flowchart

the SIFT feature matching algorithm. SIFT feature matching algorithm is mainly divided into two parts: the extraction of SIFT feature points, and then the matching of the extracted feature points. The detailed flowcharts of these two parts will be introduced below (Fig. 5.3).

5.3 Experimental Results and Analysis

In response to the lack of color information of color images in the SIFT algorithm, which results in the inability to extract feature points in certain areas and the low correct matching rate, we propose a SIFT image matching algorithm based on

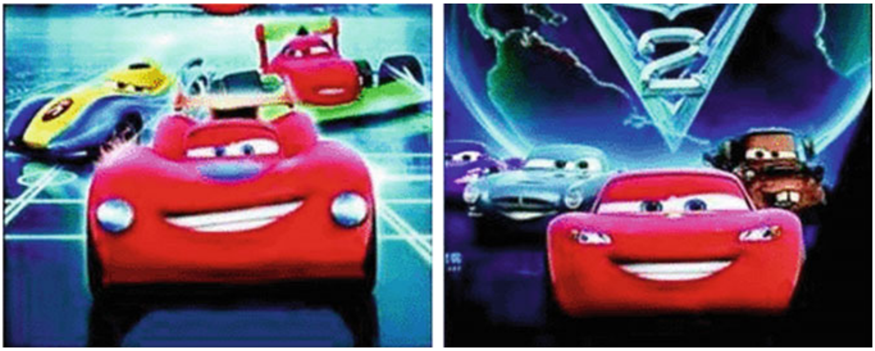


Fig. 5.4 Experimental original image. Feature points extracted from the original image separated by color channels:

image grayscale, color information and exposure information SIFT image matching method. In order to be able to verify the superiority of the proposed algorithm, we compare the improved algorithm with the original algorithm from feature point extraction and correct matching rate.

5.3.1 Experiment

See Figs. 5.4, 5.5, 5.6 and 5.7.

The matching effect of the traditional SIFT algorithm.

SIFT matching after bringing in the color descriptor, and the feature points of similar parts.

5.3.2 Experimental Screenshot Analysis

The images used in the experiments are all from the images of cartoons that are debated to be plagiarized on the Internet and the images of similar cartoons intercepted by ourselves, as well as some pictures of landscape photos taken under different angles on the Internet, and each image is matched with its similar image separately. To verify the effectiveness of the algorithm, different kinds of image data were selected for experimental comparison, and a total of 180 groups of different types of images were experimented. The experimental images had 250 * 200 pixels and 250 * 280 pixels, and the size of each image ranged from 20 to 100 KB, and the initial color space was all RBG color images. In this paper, experiments were conducted using SIFT, RGB-SIFT and HSV-SIFT, and a comparison experiment was conducted using PCA-SIFT and DSP-SIFT (Tables 5.1 and 5.2).

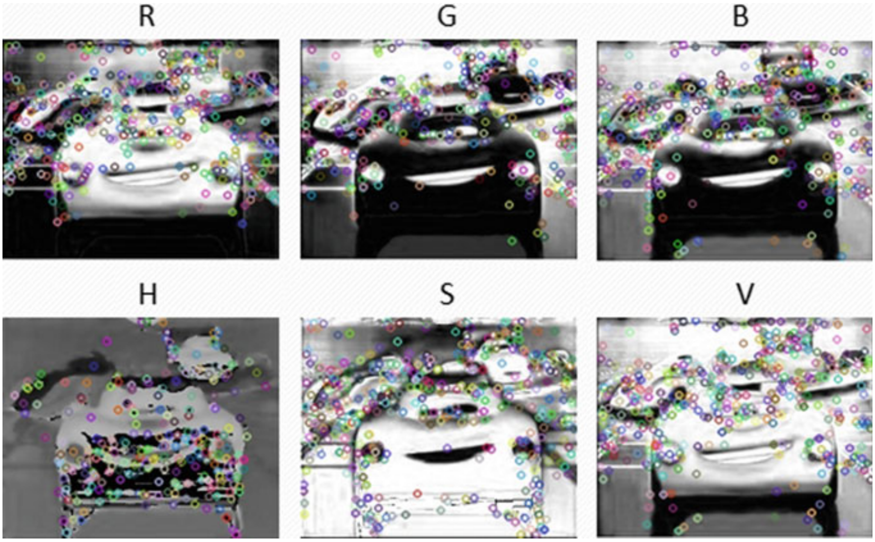


Fig. 5.5 Feature points extracted by SIFT after separating color channels

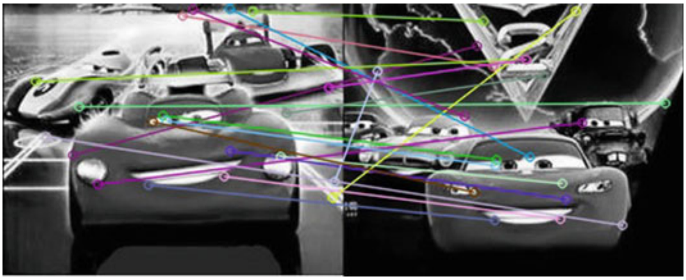


Fig. 5.6 SIFT image matching

5.4 Conclusion

This article focuses on the study of the SIFT-based image matching algorithm, and analyzes the advantages and disadvantages of the SIFT algorithm. Aiming at the problems of SIFT, corresponding improvement strategies are proposed, and through many experiments, the improved algorithm is compared with the original SIFT algorithm, and the result is that the improved algorithm is more superior. This paper proposes an improved method for the lack of color information in the SIFT algorithm: the SIFT image matching algorithm based on image gray-scale and the SIFT image matching algorithm based on color information. The improved method is simulated and implemented, and compared and analyzed with the original algorithm. The experimental

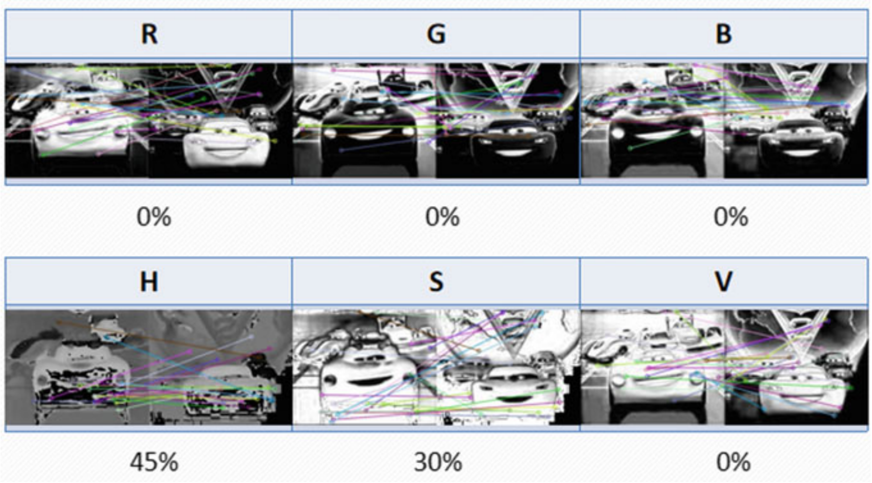


Fig. 5.7 Improved SIFT image matching

Table 5.1 Comparison of the matching performance of different image algorithms

Algorithm	SIFT	RGB-SIFT	HSV-SIFT
Average matching accuracy of general images (%)	83.51	90.21	93.41
Average matching accuracy between animation images (%)	21.27	17.25	35.81
Average matching time (s)	1.75	3.43	4.16

Table 5.2 Comparison of the matching performance of different image algorithms

Algorithm	PCA-SIFT	DSP-SIFT	HSV-SIFT
Average matching accuracy of general images (%)	78.65	85.33	93.41
Average matching accuracy between animation images (%)	19.36	21.23	35.81
Average matching time (s)	1.21	2.69	4.16

results show that the performance of the improved algorithm is improved compared to the original algorithm.

In view of the current development trend and hot issues of image processing technology, the matching speed of the algorithm can be improved on the basis of improving the matching accuracy of the algorithm in the future. The difference between genuine comics and pirated comics is too large, and the feature extraction algorithm is difficult to match images with too large differences. This is one of the contents to be studied in this article.

Acknowledgements This research project was supported by the National Natural Science Foundation of China (Grant No. 62062064).

References

1. Ping, Z., Luo, X.: A robust feature matching algorithm based on CSIFT descriptors. In: 2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE (2011)
2. Li, Q., Peng, Q., Chen, J., et al.: Improving image classification accuracy with ELM and CSIFT. Comput. Sci. Eng. **21**(5), 26–34 (2019)
3. Jian, W., Cui, Z., Sheng, V.S., et al.: A Comparative study of SIFT and its variants. Meas. Sci. Rev. **13**(3), 122–131 (2013)
4. Wu, T., Toet, A.: Color-to-grayscale conversion through weighted multiresolution channel fusion. J. Electron. Imag. **23**(4) (2014)
5. Dong, J., Soatto, S.: Domain-size pooling in local descriptors, DSP-SIFT. IEEE (2014)
6. Li, Y., Wang, Q., Chen, J., et al.: K-means algorithm based on particle swarm optimization for the identification of rock discontinuity sets. Rock Mech. Rock Eng. **48**(1), 375–385 (2015)
7. Li, J., Wang, H., Zhang, L., et al.: The research of random sample consensus matching algorithm in PCA-SIFT stereo matching method. In: 2019 Chinese Control and Decision Conference (CCDC) (2019)
8. Wachs-Lopes, G.A., et al.: A strategy based on non-extensive statistics to improve frame-matching algorithms under large viewpoint changes. Signal Process. Image Commun. (2019)
9. Qiu, J., et al.: Hierarchical resource allocation framework for hyper-dense small cell networks. IEEE Access **4**, 8657–8669 (2017)
10. Bosch, A., Zisserman, A., Muoz, X.: Scene classification via pLSA. In: European Conference on Computer Vision. Springer, Berlin, Heidelberg (2006)
11. Song, X., Muselet, D., Trémeau, A., et al.: Affine transforms between image space and color space for invariant local descriptors. Pattern Recogn. **46**(8), 2376–2389 (2013)

Chapter 6
Image Recognition Technology Based Evaluation Index of Ship Navigation Risk in Bridge Area

Dawei Chen, Renqiang Wang, Yongqian Yang, and Jingdong Li

Abstract China has a vast territory, numerous inland river systems, with abundant water transport resources. More and more ships are traveling in inland waterways. Therefore, the risk of ship accidents in inland waterways is increasing year by year. This paper mainly studies the research and application of the navigation safety risk evaluation index system in bridge area based on image recognition technology. The convolutional neural network based detection model for inland river ships is proposed. Firstly, several common target detection algorithms are compared and analyzed in this paper, and a single-stage target detection algorithm with the best performance is selected, which is combined with the target detection algorithm according to the navigable environment characteristics of the bridge area. On the basis of ship track prediction, this paper studies the quantification of collision, grounding, hitting reef and collision risk and establishes the ship collision risk evaluation model.

6.1 Introduction

The volume of inland river transportation is increasing, and the density of traffic flow is getting higher and higher. Inland waterway is increasingly busy at the same time, waterway traffic safety problems are becoming more and more prominent. Especially with the erection of a large number of Bridges in the area of ship heading safety problems gradually appeared and aroused the attention of all walks of life. In recent years, the country has stepped up efforts to build infrastructure, including roads, and the number of Bridges across major inland rivers such as the Yangtze has been increasing. On the one hand, the construction and opening of the bridge have provided great convenience for land transportation, convenient people's life and promoted the regional economic development. On the other hand, the bridge affects the navigable environment of the bridge area to some extent, which Narrows the original waterway, and also affects the flow velocity and flow direction of the bridge

D. Chen · R. Wang (✉) · Y. Yang · J. Li
Navigation College, Jiangsu Maritime Institute, Nanjing 211170, China
e-mail: wangrenqiang2009@126.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_6

area to varying degrees, thus leading to the occurrence of a variety of accidents [1]. On the one hand, it is of great practical significance to study the navigation safety assurance technology of ships in bridge area. On the other hand, the emergence of new technologies has also created new conditions for research in this field.

In terms of the navigation safety of ships in bridge areas, domestic and foreign scholars' researches on the navigation safety of Bridges mainly focus on the ship-to-bridge collision mechanics research, ship-to-bridge collision accident research, bridge pier anti-collision device research, bridge navigation safety management mechanism research, ship-to-bridge collision monitoring technology research, ship-to-bridge collision risk assessment research, and so on. Owczarzak processes the real-time bridge scene video collected by computer vision technology to extract the dynamic motion parameters of ships, so as to realize the monitoring and early warning of ship-to-bridge collision [2]. Zaman puts forward the concept of comprehensive security assessment (Formal Safety Assessment, FSA) combined with fuzzy theory to evaluate traffic safety in bridge waters [3].

Based on the convolutional neural network and the monitoring video of inland waterway, this paper conducts an in-depth study on the identification and detection tasks of passing ships in the video surveillance. The main influencing factors of ship navigation are clarified, and the main characteristics of ship traffic flow in bridge area are analyzed. According to the related early warning theory, the navigation risk assessment model of ships in bridge area is designed.

6.2 Risk Assessment of Ship Navigation in Bridge Area Based on Image Recognition Technology

6.2.1 Ship Target Monitoring Based on Convolutional Neural Network

At present, the target detection algorithm has two branches: single-stage target detection algorithm and two-stage target detection algorithm based on region suggestion.

(1) *Two-stage Target Detection Algorithm*

The idea of the two-stage target detection algorithm is to search the region of interest (ROI) in the input image, send the ROI into the subsequent classification network to complete the ROI classification, and at the same time to complete the fine tuning of the ROI coordinates, so it is also called the detection algorithm based on the region suggestion. The two-stage target detection algorithm is formed into ROI, and then the ROI is classified and fine-tuned, so the detection accuracy is higher, but it consumes more computation time and the detection speed is slower, and the representative algorithm is Faster R-CNN target detection algorithm [4].

Selective Search (SS) algorithm is used to generate ROI in R-CNN and FAST R-CNN, which has a large amount of computation and cannot complete end-to-end training. Therefore, RPN network is used to replace SS algorithm to generate ROI in Faster R-CNN. RPN network further clarifies the output results of feature extraction layer, and generates 9 prediction boxes with different scales and length-to-width ratios at each position of the feature map, and each prediction box carries its own category information and coordinate information [5]. Through the category information, the effective prediction box is selected and combined with the pre-arranged anchor points. The coordinate information of the prediction box is used to fine-tune the coordinate information of the anchor points, and the final ROI is obtained.

After the ROI is obtained, it is sent to the ROI pooling layer together with the feature map obtained from the feature extraction layer for ROI pooling operation, and the result is sent to the subsequent FAST R-CNN network to complete the detection and recognition of the target.

(2) *Single-stage Target Detection Algorithm*

The idea of single-stage target detection algorithm is to divide the whole input image into several regions, and directly regression and classification of the position of the boundary box of the image in the region, which is called the detection algorithm based on regression. The single-stage target detection algorithm eliminates the regional suggestion link in the two-stage target detection algorithm, so the calculation speed is faster than that of the two-stage target detection algorithm, and it can carry out real-time detection better [6]. YOLO algorithm and SSD algorithm are the representative algorithms of single-stage target detection.

YOLO algorithm is a typical single-stage target detection algorithm, which directly uses the idea of regression to solve the position, size, and category of the target bounding box. Therefore, the detection speed of YOLO algorithm is faster than the two-stage target detection algorithm [7].

In YOLO algorithm, the input image is divided into $s * s$ size of the same area, each area is responsible for the prediction center in fall within the area of the object, in each region to generate two boxes, regression prediction frame's size, at the same time category forecast box, and then through the maximum mechanism to filter forecast box and produce the final target detection.

The work done in this paper is the ship target detection and recognition based on inland river surveillance video, so as to improve the accuracy as far as possible under the condition of meeting the real-time requirements. The accuracy and real-time performance of the above target detection algorithms are compared in PASCAL data set. As shown in Table 6.1, the Faster R-CNN algorithm has a high detection accuracy, but its real-time performance is poor due to the need to generate the suggested region. Single-stage target detection algorithm has high real-time performance, and compared with YOLO algorithm, SSD algorithm draws on the idea of anchor points and has higher accuracy. According to the task requirements, it is necessary to ensure the real-time performance first, so the single-stage target detection algorithm is selected in this paper for subsequent research.

Table 6.1 Target detection algorithm performance

Target detection algorithm	mAP (%)	FPS
Faster R-CNN	73.2	7
YOLO	65.9	21
SSD300	76.5	45

6.2.2 Risk Assessment of Ship Navigation in Bridge Area

In fact, track prediction is to predict the future position of ships. The paper discusses and quantifies the possible dangers in the navigation process of ships in the future. In this paper, four types of accidents including collision, grounding, reef, and contact loss are studied, which can be classified as collision between ships and obstacles, that is, when ships collide with objects such as piers, lights, and reefs, they can be considered as collision with obstacles with zero speed, while when two ships collide, they can be considered collision with obstacles with non-zero speed. For the determination of ship collision risk, the more common method is to measure the collision risk.

Collision Risk Index (CRI) is a measure of the likelihood of Collision between ships. The risk of ship collision can provide basis for crew to take collision avoidance measures [8]. Specifically, when the collision risk value is too large, it indicates that the possibility of collision is high, so the crew should be vigilant, make quick response and take timely measures to avoid collision. When the collision risk is very small, it means that the ship is basically in the safe navigation zone in the future and can maintain the current navigation state. In this sense, collision risk is suitable for early warning judgment of ship navigation risk.

There are three main models [9] commonly used for mathematical calculation of ship collision risk: weighting method, fuzzy mathematical method, and artificial neural network method. In view of the fact that ship pilots mainly rely on the real-time to distance of the closest point of approach (DCPA) and time to the closest point of approach (TCPA) of the target ship in the practice of collision avoidance at sea. The weighting method fully takes into account the real-time DCPA and TCPA of the target ship, and the calculation model of this method is simple. For this reason, this article uses the first method for calculation, and its mathematical expression [10] is:

$$\rho = (a \cdot S_{DCPA})^2 + (b \cdot t_{TCPA})^2 \tag{6.1}$$

In the above formula, S_{DCPA} represents the safety meeting and distance, t_{TCPA} represents the nearest meeting and time. a and b are revision coefficients.S_{DCPA} and t_{TCPA} adopt their own unit measurement without considering the dimension, so the size of the result cannot fully reflect the risk degree of ship collision.

6.3 Assessment of Ship Collision Hazard

Artificial neural network approach simulates human brain behavior and function by connecting artificial neurons. This approach incorporates more indicators into the model. In practice, the minimum encounter distance, minimum encounter time and whether the incoming ship is on the port side or starboard side are generally adopted. The reliability of the method depends on the rationality of the model. If the collision risk model established is not suitable, the results will lack credibility.

In this paper, two factors, space collision risk and time collision risk, are taken into account when considering collision risk of ships. However, the concepts and related models of DCPA, TCPA, spatial collision risk, and temporal collision risk are mostly used in the wide water area. Inland waterways are relatively narrow and navigation areas for ships are limited. Therefore, certain adjustments are needed when these concepts are applied to inland waterways.

6.4 Risk Assessment of Collisions in Bridge Areas

6.4.1 Space Collision Risk

In terms of the risk of space collision, the minimum safety distance between the ship and the navigational obstacle is assumed to be D, and D represents the minimum encounter distance obtained from the predicted point. Then the risk of space collision is expressed as:

$$\rho_d = \frac{D - d}{D} \tag{6.2}$$

When D $\leq$ d, $\rho_d = 0$, that is, the ship is safe. Among them, the minimum safe distance D is obtained through the investigation of the crew.

In order to determine the parameter D in the model, the empirical value of D is obtained by conducting a questionnaire survey among the crew of ships that often pass through. Due to the obvious difference between the upstream and downstream navigation of the ship, the investigation is carried out, respectively, for the different upstream and downstream conditions and obstacles.

As shown in Figs. 6.1 and 6.2, the median value of each section was taken according to the above survey results, and then weighted calculation was carried out. As a result, the safe distance between up-links and navigation marks is 21.63 m (take 22 m), the safe distance between down-links and navigation marks is 28.52 m (take 29 m), the safe distance between up-links and bridge pier is 27.15 m (take 28 m), and the safe distance between down-links and bridge pier is 34.75 m (take 35 m).

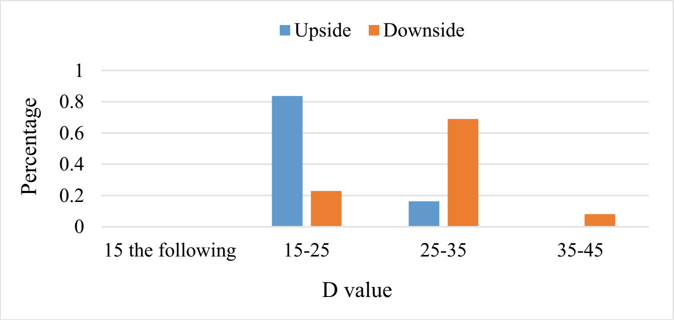


Fig. 6.1 The nearest safe distance from the navigation mark laterally

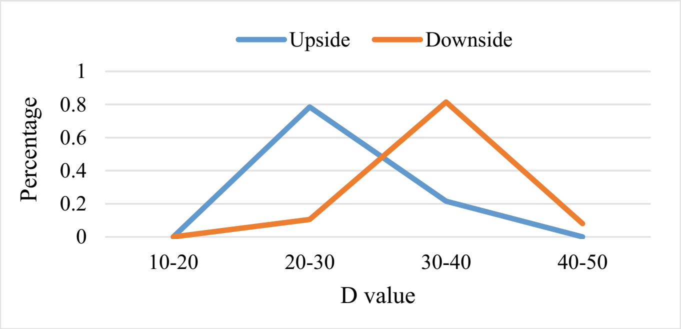


Fig. 6.2 The nearest safe distance from the pier lateral to the ship

6.4.2 Time Collision Risk

In terms of time collision risk, suppose that the time required for the ship to turn 90° is t, and the minimum encounter time is T, then the time collision risk is expressed as:

$$\rho_t = \frac{t}{T} \tag{6.3}$$

when T $\leq$ t, Pt = 1, that is, the time for the ship to adjust the course is quite urgent. The time required for most ships to turn 90° 1.5–3 min, where t is 3 min.

6.4.3 Model of Comprehensive Evaluation Indicators

By integrating the air collision risk and time collision risk, the collision risk of the ship can be expressed as:

$$\rho = \rho_d \cdot \rho_t \cdot k \tag{6.4}$$

where, k is the accident type qualifying parameter, and its value can be 0 or 1. For collision and contact loss accidents, k = 1; For stranding and reef accidents, it is necessary to judge whether they are stranded or struck according to the depth of the shoal and reef, the current water level and the draft of the ship. If they are, then k = 1, otherwise k = 0.

6.5 Conclusions

The research objective of this paper is to solve the practical problems faced by China's inland river bridge area. Through image recognition technology, a set of bridge navigation risk assessment model is established to avoid bridge accidents caused by high elevation, grounding, yaw, etc. Based on the analysis of the advantages and disadvantages of the two-stage target detection algorithm and the single-stage target detection algorithm in the ship target detection, a single-stage target detection with a convolutional neural network as the image recognition technology that combines the DCPA and TCPA weighted risk model is selected algorithm. In our future work we intend to integrate the space collision risk and time collision risk to establish a comprehensive collision risk evaluation index model for ship navigation in the bridge area.

References

1. Lee, B.K., Kim, D.H., Lee, S.D., et al.: A study on advanced seafarers' training for improving abilities of officers in charge of a navigational watch who handle navigational equipment: to focus on the ECDIS. J. Fish. Mar. Educ. **28**(2), 323–335 (2016)
2. Owczarzak, W., Mocek, A., Kaczmarek, Z., et al.: Changes of soil water regime types in the area adjacent to the Tomis^{3}awice open-cast lignite mine (central Poland). Soil Sci. Annu. **68**(1), 39–45 (2017)
3. Zaman, M.B., Pitana, T., Iswantoro, A., et al.: Risk Analysis on ship wreck and container cargo to ship navigation. Trans. Nav. Int. J. Mar. Navig. Saf. Sea Transp. **11**(1), 71–77 (2017)
4. Jin, R., Owais, H.M., Lin, D., et al.: Ellipse proposal and convolutional neural network discriminant for autonomous landing marker detection. J. Field Robot. **36**(1), 6–16 (2019)
5. Yin, S., Liu, J., Teng, L.: Strategic target classification with transfer learning. Int. J. Electron. Inf. Eng. **9**(1), 22–28 (2018)
6. Li, S., Dou, Y., Niu, X., et al.: A fast and memory saved GPU acceleration algorithm of convolutional neural networks for target detection. Neurocomputing **230**(22), 48–59 (2016)

7. Liu, Y., Cao, X., Li, Z., et al.: Video object detection based on correlation feature and convolutional neural network. Huanan Ligong Daxue Xuebao/J. South China Univ. Technol. (Nat. Sci.) **46**(12), 26–33 (2018)
8. Zhang, W., Liu, S., Luo, W., et al.: A new approach for probabilistic risk assessment of ship collision with riverside bridges. Adv. Civ. Eng. **2**, 1–12 (2020)
9. Youssef, S.A.M.: Risk control options against ship collision and grounding accidents: a survey of the state-of the-art. Nav. Eng. J. **129**(1), 99–110 (2017)
10. Cheng, Z., Li, Y., Wu, B.: Early warning method and model of inland ship collision risk based on coordinated collision-avoidance actions. J. Adv. Transp. **2**, 1–14 (2020)

Chapter 7
Equalization of Directional Multidimensional Histograms of Matrix and Tensor Images

Roumiana Kountcheva and Roumen Kountchev

Abstract New approaches are proposed for the equalization of directional multidimensional histograms of 2D-matrix and 3D-tensor images, obtained from CTI or MRI sequences, video, etc. Such equalization opens new possibilities for quality improvement of the tensor images in a selected direction of the 3D space. The directional contrast enhancement is of high importance for example, for the detection of objects oriented in the same direction. In the paper the algorithms for the calculation of the multidimensional directional histograms of 2D and 3D images are defined: in the first case, in the 4 directions of the 2D plane, and in the second—in 13 directions of the 3D space. The conditions for the directional equalization 2D and 3D histograms are defined, on the basis of which is enhanced the contrast of the processed images. In the paper are also defined the criteria for the evaluation of contrast enhancement in 3D images. The new method is illustrated through a digital example. The future application areas of the proposed approach are in the processing of underground images, 3D medical and dental images, etc. In the future the method will be aimed at the double transform of the grey levels of neighboring triples of voxels in correspondence with the selected approximation 3D model of the directional histogram, the local directional equalization of monochrome images, etc.

7.1 Introduction

In the last years, many methods for analysis of sequences of images of various kinds (medical, seismic, multispectral, etc.) were presented in large number of publications. Significant number of research works is already known and aimed at the contrast enhancement of matrix and tensor images. The tensor representation of these images

R. Kountcheva (✉)
TK Engineering, Sofia 1582, Bulgaria
e-mail: kountcheva_r@yahoo.com

R. Kountchev
Technical University of Sofia, Bul. Kl. Ohridsky 8, Sofia 1756, Bulgaria
e-mail: rkountch@tu-sofia.bg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_7

permits the three-dimensional (3D) Gray Level Co-occurrence Matrix (3D-GLCM) to be calculated in 13 independent spatial directions, for each of which to be extracted the features of Haralick, LBP descriptors, or CNN, used for texture and 3D objects classification [1–9]. The obtained results show that compared to the two-dimensional, the 3D-GLCM offers much more possibilities for deeper analysis of tensor images of various kinds. In [10] is offered a method for color image enhancement through 3D histogram equalization, in result of which is achieved simultaneous histogram extension in 3 directions: R, G, B. In [11], the color 3D histogram is used as a tool to achieve exact coincidence between histograms of a couple of images. Methods and algorithms for contrast enhancement of 3D images are presented in papers [12–14]. In [12] is offered new 3D adaptive histogram equalization method for improving the contrast of medical images. This method is a 3D generalization of the famous algorithm CLAHE [13]. In accordance with CLAHE, the tensor image is divided into 3D blocks, and for each is calculated the 1D histogram; then on it is applied the CLAHE algorithm. To avoid inter-block distortions, three-linear interpolation is used. However, in this case the 3D histogram of each block is not used, which represents to highest degree the correlation between its voxels. In [14] is offered a method which is a multidimensional extension of CLAHE (MCLAHE). It could be used for 3D, and for 4D (3D + time) tensor images as well. Unlike the approach given in [12], in MCLAHE is first executed multidimensional padding and after that the tensor is divided into blocks. For these blocks is applied CLAHE, followed by multilinear interpolation between the sides of the nearest-neighbor blocks. Similar to the approach from [12], in MCLAHE is used equalization of the 1D histogram of each block of the tensor image.

The main objective of this work is to present new method for equalization of 3D images, which is based on the use of 3D-GLCM calculated in 13 spatial directions. First, here is analyzed the equalization of matrix (2D) images, based on the 2D-GLCM, and then follows the directional equalization of tensor (3D) images.

7.2 Equalization of the Directional 2D Histograms of the Matrix Image

7.2.1 Definition of the Directional 2D Histograms

The processing of matrix images through equalization of their brightness histograms, aimed at their contrast enhancement, could be generalized for the case of the 2D histogram equalization. Let the original image be halftone of size $N \times N$ pixels. The image is represented as the matrix [B] with elements $b(i, j)$ for $i, j = 0, 1, 2, \ldots, N - 1$ and $b = 0, 1, 2, \ldots, L - 1$ (L—the number of grey levels). The 2D histogram $h(u, v)$ of the monochrome image, also known as 2D co-occurrence matrix $[h(u, v)]$, represents the frequency of appearance of the grey levels for each couple of pixels $b(i, j)$ and $b(q, w)$ placed at a distance d:

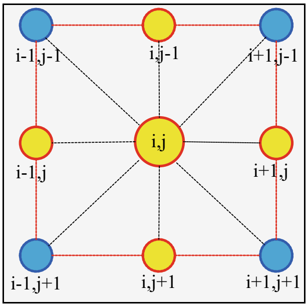


Fig. 7.1 Positioning of the closest 8 neighbor pixels for the pixel $b(i, j)$ of the 2D image, for the case of orthogonal discretization

$$d\ [(i, j), (q, w)] = \ |i - q| + |j - w| \text{ for } i, j, q, w = 0, 1, 2, .., N - 1. \tag{7.1}$$

Depending on the spatial position of the couple of pixels $b(i, j)$ and $b(q, w)$, the corresponding *directional 2D histograms of the image* could be defined.

In particular, if the pixels $b(i, j)$ and $b(q, w)$ are neighbors, the distance between them in horizontal and vertical direction is $d = 1$, and for both diagonals, it is $d = \sqrt{2}$, respectively.

In correspondence with Fig. 7.1, the mutual position of a couple of pixels could be defined in one of the ways, shown below:

- horizontal positioning of the pixels: $b(i - 1, j)$, $b(i, j)$ or $b(i, j)$, $b(i + 1, j)$, where the corresponding 2D histogram is defined by the relation:

$$h_x(u, v) = N_x(u, v)/N_x, \tag{7.2}$$

where $N_x(u, v)$ is the number of the couples of neighbor pixels in horizontal direction; the grey level of the first pixel is u, and of the second it is v, for $u, v = 0, 1, \ldots, L - 1$; $N_x = 2N(N - 1)$—total number of the couples of neighbor pixels in horizontal direction;

- vertical positioning of the pixels: $b(i, j - 1)$, $b(i, j)$ or $b(i, j)$, $b(i, j + 1)$. Their 2D histogram is defined as:

$$h_y(u, v) = N_y(u, v)/N_y, \tag{7.3}$$

where $N_y(u, v)$ is the number of the couples of neighbors in vertical direction; the level of the first is u, and of the second it is v, for $u, v = 0, 1, \ldots, L - 1$. $N_y = 2N(N - 1)$—total number of the couples of neighbor pixels in vertical direction;

- diagonal positioning of the couple of pixels: $b(i + 1, j - 1)$, $b(i, j)$; $b(i, j)$, $b(i - 1, j + 1)$; $b(i - 1, j - 1)$, $b(i, j)$; $b(i, j)$, $b(i + 1, j + 1)$. Their 2D histograms for both diagonals (right and left) with orientation $\pi/4$ or $5\pi/4$ and $3\pi/4$ or $7\pi/4$, respectively, are defined by the relation:

$$h_{d_i}(u, v) = N_{d_i}(u, v)/N_{d_i} \text{ for } i = 3, 4, \tag{7.4}$$

where $N_{d_i}(u, v)$ is the number of the couples of neighbor pixels for the diagonal i; the level of the first is u, and of the second it is v, for $u, v = 0, 1, \ldots, L - 1$; N_{d_i} is the number of the couples of neighbor pixels for the diagonal i.

7.2.2 Equalization of the 2D Histogram of the Matrix Image

The histogram equalization is a basic approach in image contrast enhancement. To apply such operation on the selected directional 2D histogram $h(m, n)$ of the processed image, the condition for the equality of the volume of $h(m, n)$ for a couple of pixels with levels (u, v), and the volume of the equalized 2D histogram $h(u, v) = 1/L^2$ of the so obtained image for the corresponding couple of pixels, with levels (r, s) should be satisfied:

$$\sum_{m=0}^{r}\sum_{n=0}^{s}\frac{1}{L^2} = \frac{(r+1)(s+1)}{L^2} = \sum_{m=0}^{u}\sum_{n=0}^{v} h(m, n) \text{ for } r, s, u, v = 0, 1, 2, .., L - 1. \tag{7.5}$$

In this case, u and v are the grey levels of the couple of neighbor pixels in the original image, and r and s—the levels calculated for the corresponding pixels after the 2D equalization.

Equation (7.5) could be represented here in the following way:

$$(r + 1)(s + 1) = L^2 H(u, v) \text{for } u, v = 0, 1, 2, .., L - 1, \tag{7.6}$$

where $H(u, v) = \sum_{m=0}^{u}\sum_{n=0}^{v} h(m, n)$ is the cumulative 2D histogram of the original image. For each couple of values (u, v) is defined the one with the smallest values (r, s), for which is executed the relation, obtained from Eq. (7.6):

$$(r + 1)(s + 1) - L^2 H(u, v) \geq 0. \tag{7.7}$$

To define the unknown r and s in the above relation could be used an additional requirement for the achievement of maximum contrast in the processed matrix image. If we assume that $A(u, v) = L^2 H(u, v)$, is obtained the following system of equations for r and s:

$$||(r + 1)(s + 1) - A(u, v)|| = \min, \tag{7.8}$$

$$|r - s| = \max. \tag{7.9}$$

The solution of this system of equations gives the values of r and s.

Depending on the selected kind of the directional 2D histogram $h_x(u, v)$, $h_y(u, v)$ or $h_d(u, v)$, after the execution of the operation in accordance with Eqs. (7.8), (7.9), it is equalized in horizontal, vertical or diagonal (left or right) direction, respectively.

The choice of the direction of the directional 2D histogram equalization could be based on one of the following criteria:

- detection of the maximum contrast pixel coefficient (CPC) for each of the 4 images obtained after equalization in each direction, compared to that of the original image.

  The coefficient CPC for the image of size N × N is defined by the relation:

$$CPC = 1/(N^2)\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{(1/8)\sum_{m=-1}^{1}\sum_{n=-1}^{1}|b(i,j) - b(i+m, j+n)|\right\} \quad (7.10)$$

  where $b(i, j)$ is the central pixel in the window of size 3 × 3 pixels. The term in the braces represents the local contrast of the pixel $b(i, j)$ in the window.
- calculation of the maximum variance $\sigma_{t_0}^2$ of the directional histogram $h_{t_0}(m, n)$, depending on its direction, t_0:

$$\sigma_{t_0}^2 = \frac{1}{L^2}\sum_{m=0}^{L-1}\sum_{n=0}^{L-1}[h_{t_0}(m, n) - \mu_{t_0}]^2 = \max \ \text{for } t_0 = 1, 2, 3, 4. \quad (7.11)$$

The parameter t_0 defines one of the 4 possible directions in which the variance of the corresponding 2D histogram is maximum. The parameter μ_{t_0} is the mean value of the histogram $h_{t_0}(m, n)$, defined by the relation:

$$\mu_{t_0} = \frac{1}{L^2}\sum_{m=0}^{L-1}\sum_{n=0}^{L-1} h_{t_0}(m, n) \quad (7.12)$$

To accelerate the calculation of the cumulative histogram $H(u, v)$ for $u, v, = 0, 1, 2, \ldots, L - 1$, in the consecutive processing of the matrix elements $[h(u, v)]$ row by row from left to right and from top to bottom, the following 2D recursive relation could be applied:

$$H(u, v) = h(u, v) + H(u - 1, v) + H(u, v - 1) - H(u - 1, v - 1). \quad (7.13)$$

This relation does not refer the elements $H(u, v)$, placed on the first row and the first column of the matrix $[H(u, v)]$, of size $L \times L$.

The advantages of the method for directional 2D equalization, presented above, are that it permits:

- equalization of the 2D histogram in the cases when its irregularity is mostly in the 3D area;
- equalization of the 2D histogram in 4 directions of the image;

- consecutive recursive equalization of the 2D histograms in the 4 directions (horizontal, vertical, left diagonal and right diagonal).

The presented method for directional 2D equalization of grayscale images could be also generalized for color *R*, *G*, *B* images. In this case is necessary the color components *R*, *G*, *B* to be transformed in advance into *Y*, *Cr*, *Cb* (or into the KLT components L_1, L_2, L_3) and after that on the first component (Y, or L_1) to be applied the 2D equalization. As a result, after the inverse color transform of the components *Y*, *Cr*, *Cb* (respectively, L_1, L_2, L_3) into *R*, *G*, *B* the restored color image is obtained, whose contrast is enhanced.

The method could be also used for the image 2D histogram modification through double transform of the grey levels of the neighbor couples of pixels in accordance with the pre-selected approximation 2D model of the histogram, through division and equalization of selected parts of the original histogram, etc. With this, the possible application areas of the presented contrasting method are expanded.

7.3 Equalization of the Directional 3D Histograms of the Tensor Image

7.3.1 Definition of the Directional 3D Histograms

By analogy with Eqs. (7.2)–(7.4), each tensor image could be defined as directional 3D histograms with various orientation. For the case when in the 3D image space a triad of pixels are neighbors and are placed on a same straight line, then the distance between the neighbor voxels in each direction *x*, *y*, *z*, is $d = 1$, and in the directions of the diagonals it is $d = \sqrt{2}$ or $\sqrt{3}$, respectively.

The mutual position of the three neighbor voxels on a same straight line is defined in one of the 13 ways, shown on Fig. 7.2:

- horizontally positioned three of voxels $b(i+1, j, k)$, $b(i, j, k)$, $b(i-1, j, k)$, when the corresponding 3D histogram is defined by the relation:

$$h_x(u, v, p) = N_x(u, v, p)/N_x, \tag{7.14}$$

where $N_x(u, v, p)$ is the number of the triads of neighbor voxels in horizontal direction in the tensor 3D image of size $N \times N \times N$. Besides, the first voxel is with grey level *u*, the second—with level *v*, and the third—with level *p*, for $u, v, p = 0, 1, .., L-1$;

$N_x = 6N^2(N-2)$—total number of the triads of neighbor voxels in horizontal direction;

- vertically positioned three of voxels $b(i, j-1, k)$, $b(i, j, k)$, $b(i, j+1, k)$. Their 3D histogram is defined by the relation:

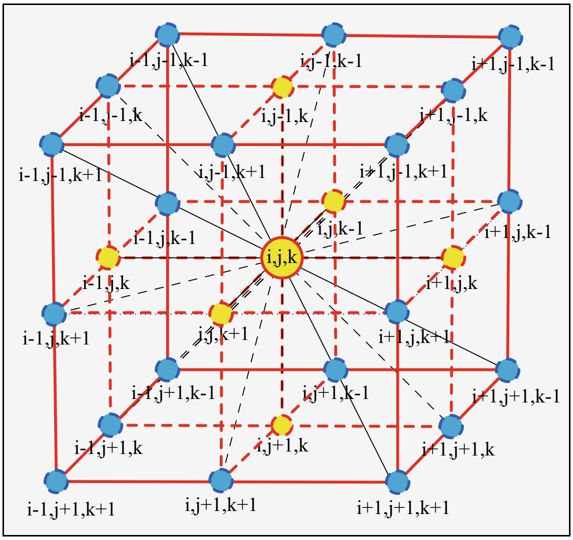


Fig. 7.2 Spatial position of the closest 24 neighbor voxels towards the voxel *b*(*i*, *j*, *k*) of the 3D-tensor image, for the case when orthogonal discretization is used

$$h_y(u, v, p) = N_y(u, v, p)/N_y, \tag{7.15}$$

where $N_y(u, v, p)$ is the number of the triads of neighbor voxels in vertical direction. Here the first voxel is with greylevel *u*, the second—with level *v*, and the third—with level *p*, for $u, v, p = 0, 1,.., L - 1$; $N_y = 6N^2(N - 2)$.total number of the triads of neighbor voxels in horizontal direction;

- laterally positioned three of voxels (in direction z) $b(i, j, k + 1)$, $b(i, j, k)$, $b(i, j, k - 1)$. Their 3D histogram is defined by the relation:

$$h_z(u, v, p) = N_z(u, v, p)/N_z, \tag{7.16}$$

where $N_z(u, v, p)$ is the number of the triads of neighbor voxels in direction *z*. In this case, the first voxel is with greylevel *u*, the second—with level *v*, and the third—with level p, for $u, v, p = 0, 1,.., L - 1$; $N_z = 6N^2(N - 2)$—the number of the triads of neighbor voxels in direction *z*.

- triads of neighbor voxels positioned along one of the diagonals and presented as a three-component vector d_i for $i = 1, 2,.., 10$—one vector for each one of the 10 diagonals shown on Fig. 7.2 with black lines:
  d_1—$b(i + 1, j - 1, k + 1)$, $b(i, j, k)$, $b(i - 1, j + 1, k - 1)$;
  d_2—$b(i, j - 1, k + 1)$, $b(i, j, k)$, $b(i, j + 1, k - 1)$;
  d_3—$b(i - 1, j - 1, k)$, $b(i, j, k)$, $b(i + 1, j + 1, k)$;
  d_4—$b(i - 1, j - 1, k - 1)$, $b(i, j, k)$, $b(i + 1, j + 1, k + 1)$;
  d_5—$b(i - 1, j - 1, k + 1)$, $b(i, j, k)$, $b(i + 1, j + 1, k - 1)$;
  d_6—$b(i, j - 1, k - 1)$, $b(i, j, k)$, $b(i, j + 1, k + 1)$;
  d_7—$b(i + 1, j - 1, k)$, $b(i, j, k)$, $b(i - 1, j + 1, k)$;

d_8—$b(i-1,j,k+1)$, $b(i,j,k)$, $b(i+1,j,k-1)$;
d_9—$b(i+1,j,k+1)$, $b(i,j,k)$, $b(i-1,j,k-1)$;
d_{10}—$b(i+1,j-1,k-1)$, $b(i,j,k)$, $b(i-1,j+1,k+1)$.

The corresponding directional 3D histograms are 13 in total, respectively: 3, corresponding to axes x, y, z, and 10 which follow the diagonals. They are defined by the relations below:

$$h_s(u,v,p) = N_s(u,v,p)/N_s \text{ for } s = 1, 2, .., 13. \tag{7.17}$$

where N_s is the number of the triads of neighbor voxels in direction s.

7.3.2 Equalization of the 3D Histograms of the Tensor Image

The methods for 2D equalization could be generalized for global and local directional equalization of 3D images, for 3D contrast limited adaptive histogram equalization (CLAHE), and for double histogram transform in correspondence with the selected 3D directional histogram.

To equalize the directional 3D histogram of the image, presented as a third-order tensor, it should satisfy the condition, similar to that from Eq. (7.5):

$$\frac{(r+1)(s+1)(t+1)}{L^3} = \sum_{m=0}^{u}\sum_{n=0}^{v}\sum_{l=0}^{p} h(m,n,l)$$
$$\text{for } r, s, t, u, v, p = 0, 1, 2, .., L-1, \tag{7.18}$$

where u, v, p are the grey levels of arbitrary chosen triad of neighbor voxels from the original image, and r, s, t are the corresponding levels for the same triad, from the processed image, obtained after 3D equalization.

From Eq. (7.18) it follows that each three of voxels from the original image with grey levels u, v, p are transformed into corresponding three of voxels with grey levels r, s, t. These voxels could be represented as a vector of 3 components, defined by the relation:

$$(r+1)(s+1)(t+1) = L^3 H(u,v,p), \tag{7.19}$$

where $H(u,v,p) = \sum_{m=0}^{u}\sum_{n=0}^{v}\sum_{l=0}^{p} h(m,n,l)$ is the cumulative 3D histogram of the original image. From each triad from the original image with values (u, v, p) the smallest set of values (r, s, t) is searched, for which is satisfied the relation:

$$(r+1)(s+1)(t+1) - L^3 H(u,v,p) \geq 0 \tag{7.20}$$

To define the unknown r, s, t in the relation above, the additional requirement for maximum contrast of the transformed tensor image could be used. If assumed that $L^3H(u, v, p) = A(u, v, p)$, the system of equations with unknown values r, s, t is obtained:

$$|(r+1)(s+1)(t+1) - A(u, v, p)| = \min; \tag{7.21}$$

$$dif_1 = |r-s| = \max \ \vee dif_2 = |r-t| = \max \ \vee dif_3 = |s-t| = \max; \tag{7.22}$$

$$d_k(\alpha, \beta)|_{\max} = 1 \mathrm{for} k = 1, 2, 3 \text{ and } \alpha, \beta = r, s, t. \tag{7.23}$$

In Eq. (7.22), dif_k for $k = 1, 2, 3$ is the maximum of the module for the difference between the couple of pixels with levels of grey (r, s), (r, t) and (s, t), respectively. In Eq. (7.23), $d_k(\alpha, \beta)|_{\max}$ is the distance between a couple of pixels with grey levels α and β, whose difference is the maximum value. The solution of the system of Eqs. (7.21)–(7.23) permits to calculate the values of r, s, t.

Depending on the orientation of the selected 3D directional histogram $h_s(u, v, p)$ for $s = 1, 2, .., 12$, and after applying the transform from Eq. (7.18), it is equalized in the corresponding direction.

The choice of the kind of the directional histogram, which is used for the equalization, could be done on the basis of the following criteria:

- through calculation of the maximum contrast pixel coefficient (CPC) for each of the obtained through directional equalization 13 output images, after comparison with the original. The CPC for an image of size $N \times N \times N$ is defined by the relation:

$$CPC = (1/N^3)\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\left\{(1/24)\sum_{m=-1}^{1}\sum_{n=-1}^{1}\sum_{r=-1}^{1}|b(i,j,k) - b(i+m, j+n, k+r)|\right\} \tag{7.24}$$

  where $b(i, j)$ is the central voxel in a window of size $3 \times 3 \times 3$ voxels. The relation in the braces defines the local contrast of the voxel $b(i, j)$ in the window.
- through definition of the maximum variance of the histogram, depending on the chosen direction:

$$\sigma_{t_0}^2 = \frac{1}{L^3}\sum_{m=0}^{L-1}\sum_{n=0}^{L-1}\sum_{l=0}^{L-1}[h_{t_0}(m, n, l) - \mu_{t_0}]^2 = \max \ \text{for } t_0 = 1, 2, .., 13. \tag{7.25}$$

The parameter t_0 defines one of the 13 possible directions, where the corresponding directional 2D histogram has maximum variance, and the parameter

$$\mu_{t_0} = \frac{1}{L^3}\sum_{m=0}^{L-1}\sum_{n=0}^{L-1}\sum_{l=0}^{L-1} h_{t_0}(m, n, l) \tag{7.26}$$

is the mean value of the directional histogram $h_{t_0}(u, v, l)$.

To enhance the calculation of the cumulative histogram in accordance with Eq. (7.25) 3D recursion could be used in accordance with the rule [10]:

$$\begin{aligned} H(u, v, p) = h(u, v, p) + H(u-1, v-1, p-1) + H(u-1, v, p) \\ + H(u, v-1, p) + H(u, v, p-1) - H(u-1, v-1, p) \\ - H(u-1, v, p-1) - H(u, v-1, p-1) \end{aligned} \tag{7.27}$$

The advantages of the new method presented in this work are in the opportunities which it offers:

- equalization of the directional 3D histogram for the cases when its irregularity is revealed mainly in the 3D area;
- ability to equalize the directional 3D histogram in 13 directions;
- ability to perform recursive equalization of the directional 3D histograms in the 13 possible directions consecutively.

The presented method for directional 3D equalization of halftone tensor images could be generalized also for color *R*, *G*, *B* tensor images, in the way, similar to that used for matrix color images. The method could also be used for the modification of the 3D histogram of the original image by setting an approximation 3D model of the directional histogram; dividing of the original histogram and equalization of its parts, etc. Besides, the method could be used for the equalization of the global directional histograms, as well as for the local histograms, calculated for a sliding 3D window placed around each voxel of the tensor image.

7.4 Example for the Equalization of the Directional Histogram of the Tensor Image

Let the 3D image be represented as a third-order tensor of size $3 \times 3 \times 3$ (for $N = 3$) and voxels $b(i, j, k)$, which have 4 levels of grey, $L = 4$. On Fig. 7.3 one example distribution of the grey levels on the voxels of the 3D image is shown. For this image the following directional 3D histograms are calculated:

- 3D histograms for the triads of pixels placed in horizontal/vertical direction, calculated in accordance with the relation:

$$h_{x/y}(u, v, p) = N_{x/y}(u, v, p)/54 \text{ for } u, v, p = 0, 1, 2, \ldots, 7 and Nx/y = 54. \tag{7.28}$$

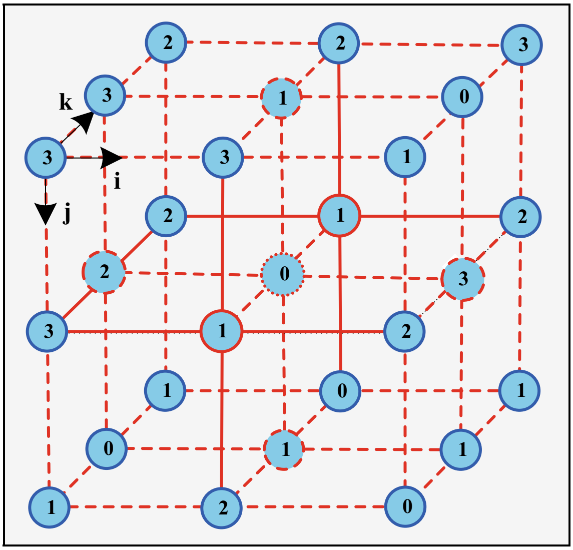


Fig. 7.3 Example for a 3D image of size 3 × 3 × 3 and 4 grey levels ($L = 4$)

Each of the so calculated 3D histograms is represented as the tensor $h_{x/y}$ of size 4 × 4 × 4, shown on Fig. 7.4. The digital values of the 3D non-normalized histograms N_x/N_y, calculated in accordance with Eq. (7.23), are given in Tables 7.1, 7.2, 7.3, 7.4 for each grey level $p = 0, 1, 2, 3$, respectively. On the basis of the data from these tables and Eqs. (7.16)–(7.18) the histograms H_x/H_y of the image from Fig. 7.3 are equalized. The cumulative histogram of the tensor image is defined by the relation:

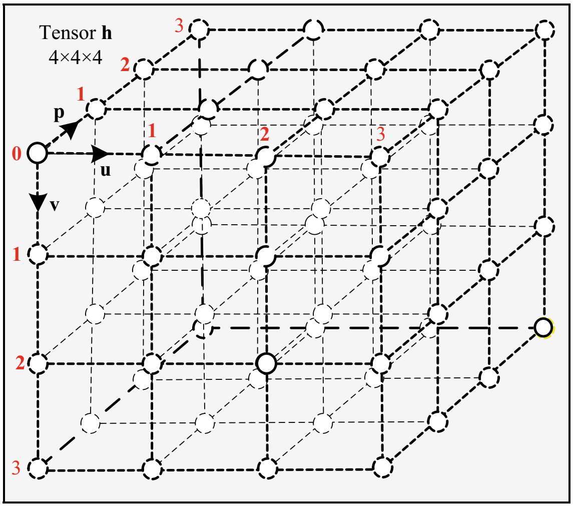


Fig. 7.4 3D histogram $h(u, v, p)$ of the triads of voxels for the example from Fig. 7.3

Table 7.1 3D-N_x/N_y for $p = 0$

↓v/u →	0	1	2	3
0	0/0	0/0	0/0	0/0
1	0/0	0/0	0/1	1/0
2	0/0	1/1	0/0	0/1
3	0/0	0/0	0/0	0/0

Table 7.2 3D-N_x/N_y for $p = 1$

↓v/u →	0	1	2	3
0	0/0	1/1	0/0	0/0
1	1/0	0/0	0/0	0/0
2	0/0	0/0	0/1	0/1
3	0/1	0/0	0/0	1/1

Table 7.3 3D-N_x/N_y for $p = 2$

↓v/u →	0	1	2	3
0	0/0	0/0	0/0	0/0
1	0/0	0/0	1/0	1/1
2	0/0	0/0	0/0	0/0
3	0/0	0/0	0/0	0/0

Table 7.4 3D-N_x/N_y for $p = 3$

↓v/u →	0	1	2	3
0	0/0	0/0	1/0	0/0
1	0/0	0/0	0/0	0/0
2	0/0	0/0	1/0	0/0
3	0/0	0/0	0/0	0/0

$$H(u,v,p) = \ (1/54)\sum_{m=0}^{u}\sum_{n=0}^{v}\ \sum_{l=0}^{p} N(m,n,l) \text{ for } u,v,p = 0, 1, 2, 3. \qquad (7.29)$$

The digital values of the 3D cumulative histograms H_x/H_y calculated in accordance with Eq. (7.20), are given in Tables 7.5, 7.6, 7.7, 7.8 for each value of $p = 0$, 1, 2, 3. To calculate the equalized 3D image is used the following relation:

$$(r+1)(s+1)(t+1) - 64\,H(u,v,p) \geq 0 \text{ for } r,s,t,u,v,p = 0, 1, 2, 3. \quad (7.30)$$

Table 7.5 3D-54H_x/H_y for $p = 0$

$\downarrow v/u \rightarrow$	0	1	2	3
0	0/0	0/0	0/0	0/0
1	0/0	0/0	0/1	1/1
2	1/1	2/2	2/2	2/3
3	2/3	2/3	2/3	2/3

Table 7.6 3D-54H_x/H_y for $p = 1$

$\downarrow v/u \rightarrow$	0	1	2	3
0	2/3	3/4	3/4	3/4
1	4/4	4/4	4/4	4/4
2	4/4	4/4	4/5	4/6
3	4/7	4/7	4/7	5/8

Table 7.7 3D-54H_x/H_y for $p = 2$

$\downarrow v/u \rightarrow$	0	1	2	3
0	5/8	5/8	5/8	5/8
1	5/8	5/8	6/8	7/9
2	7/9	7/9	7/9	7/9
3	7/9	7/9	7/9	7/9

Table 7.8 3D-54H_x/H_y for $p = 3$

$\downarrow v/u \rightarrow$	0	1	2	3
0	7/9	7/9	8/9	8/9
1	8/9	8/9	8/9	8/9
2	8/9	8/9	9/9	9/9
3	9/9	9/9	9/9	9/9

On the basis of the data in Tables 7.5, 7.6, 7.7, 7.8 and by using Eq. (7.30) the 3D histograms in directions x, y, z are equalized. For each triad of values (u, v, p) the smallest (r, s, t) is searched, which satisfies Eq. (7.30).

For example,

- for the horizontally directed vector with components $u = 3$, $v = 3$, $p = 1$, shown on Fig. 7.3, from Eq. (7.29) it follows that $64H_x(3, 3, 1) \approx 6$. In this case $(r + 1)(s + 1)(t + 1) - 6 \geq 0$, which is satisfied for the triads with minimum values $r = 3$, $s = 1$, $t = 0$ and $r = 3$, $s = 0$, $t = 1$. However, taking into account Eqs. (7.22) and (7.23), the final solution is obtained: $r = 3$, $s = 0$, $t = 1$.

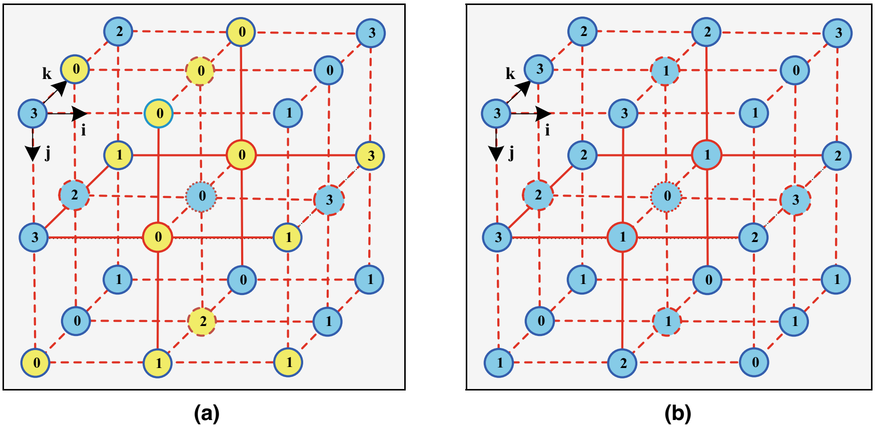


Fig. 7.5 Result of the equalization in horizontal direction

- for the horizontally directed vector with components $u = 2$, $v = 0$, $p = 3$ shown on Fig. 7.3 follows that $64H_x(2, 0, 3) \approx 9$. Then $(r + 1)(s + 1)(t + 1) - 9 \geq 0$ for $r = 2$, $s = 0$, $t = 3$.

By analogy are calculated the 3 components of the remaining 7 horizontally directed vectors. The result of the directional 3D equalization in horizontal direction for the 9 three-component vectors which compose the transformed tensor, is shown on Fig. 7.5. For the evaluation of the mean change of the contrast in the result image from Fig. 7.5 compared to the original from Fig. 7.3, could be used the criterion below:

$$\Delta = (1/N^3)\sum_{i=0}^{2}\sum_{j=0}^{2}\sum_{k=0}^{2}|b_0(i,j,k) - b_1(i,j,k)| \tag{7.31}$$

Here $b_0(i, j, k)$ and $b_1(i, j, k)$ are the voxels of the 3D images before and after the equalization. For the example images from Figs. 7.3 and 7.5 and Eq. (7.31) it follows that $\Delta = 18/27 = 0.66$, i.e., the mean change of the contrast is about 66%.

7.5 Conclusions

In this work, new approaches are proposed for the equalization of directional multidimensional histograms. This equalization opens new possibilities for quality improvement of tensor images in the selected direction of the 3D space. The algorithms used for the calculation of the directional 3D histograms are presented, which could be used for the global and local equalization of the corresponding tensor images. Also, in this work the criteria are offered for the evaluation of the 3D images contrast enhancement and the possible new applications are shown through double transform of the grey levels for each triad of voxels in correspondence with the pre-selected approximation model of the 3D directional histogram; the quality improvement of color tensor images; the local adaptive equalization 3D-CLAHE based on the 3D-GLCM, etc.

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

1. Ghoneim, D., Constans, J. M., Certaines, J.: Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas. Magn. Resonan. Imaging (Elsevier) **21**, 983–987 (2003)
2. Chen, W., Giger, M., Li, H., Bick, U., Newstead, G.: Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. Magn. Reson. Med. **58**, 562–571 (2007)
3. Eichkitz, C., Amtmann, J., Schreilechner, M.: Calculation of grey level co-occurrence matrix-based seismic attributes in three dimensions. Comput. Geosci. (Elsevier) **60**, 176–183 (2013)
4. Kim, T., Cho, N., Jeong, G., Bengtsson, E., Choi, H.: 3D texture analysis in renal cell carcinoma tissue image grading. Comput. Math. Methods Med. (Hindawi Publishing Corporation) **2014,** 12 (2014)
5. Cao, W., Pomeroy, M., Gao, Y., Barish, M., Abbasi, A., Pickhardt, P., Liang, Z.: Multi-scale characterizations of colon polyps via computed tomographic colonography. Vis. Comput. Ind. Biomed. Art (Springer, Open Access) **2**, 25 (2019)
6. Moyaa, L., Zakerib, H., Yamazakic, F., Liuc, W., Masa, E., Koshimuraa, S.: 3D gray level co-occurrence matrix and its application to identifying collapsed buildings. J. Photogramm. Remote Sens. (Elsevier) **149**, 14–28 (2019)
7. Yan, L., Xia, W.: A modified three-dimensional gray-level co-occurrence matrix for image classification with digital surface model. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2/W13, 2019, ISPRS Geospatial Week 2019, Enschede, Netherlands, 10–14 June (2019)
8. Barburiceanu, S., Terebes, R., Meza, S.: 3D Texture feature extraction and classification using GLCM and LBP-based descriptors. Appl. Sci. MDPI **11**, 2332, 5, 1–25 (2021)
9. Tan, J., Gao, Y., Liang, Z., Cao, W., Pomeroy, M., Huo, Y., Li, L., Barish, M., Abbasi, A., Pickhardt, P.: 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)
10. Trahanias, P., Venetsanopoulos, A.: Color image enhancement through 3-D histogram equalization. In: Proceedings of the 11th International Conference on IMAGE, SPEECH and SIGNAL ANALYSIS, vol. III, 1992, pp. 545–548, IEEE Xplore, 0-81862920-7/9, January (1992)
11. Morovic, J., Sun, P.: Accurate 3D image color histogram transformation. Pattern Recognit. Lett. (Elsevier) **24**, 1725–1735 (2003)
12. Amorim, P., Moraes, T., Silva, J., Pedrini, H.: 3D adaptive histogram equalization method for medical volumes. In: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), vol. 4, VISAPP, pp. 363–370. Elsevier (2018)

13. Zuiderveld, K.: Contrast Limited Adaptive Histogram Equalization, Graphics Gems IV, pp. 474–485. Elsevier, Amsterdam (1994)
14. Stimper, V., Bauer, S., Ernstorfer, R., Scholkopf, B., Xian, R.: Multidimensional Contrast Limited Adaptive Histogram Equalization, vol. 7, pp. 165437–165447 (2019)

Chapter 8
Small Object Detection of Remote Sensing Images Based on Residual Branch of Feature Fusion

Xiaoling Feng

Abstract In recent years, the detection of remote sensing images has been developed widely, and small objects have been paid more and more attention. The existing small object detection methods fuse the multi-scale features of different layers directly when using the feature pyramid network. However, due to the decrease of channels in feature fusion, the top-level feature of pyramid will lose information of the object, which is disadvantageous to detect small object ion. In order to fuse multi-scale features more effectively, we propose an object detection method based on the residual branch of feature fusion (RBFF), which is specially used to detect small objects. Our approach improves the network structure of the feature pyramid. We also recalculated the weights to reduce the semantic gap in feature fusion. In addition, we also introduce sub-pixel convolution to reconstruct the low-frequency information of the feature map accurately, to obtain the feature map with more information. The experimental results show that our method has a good effect.

8.1 Introduction

With the advance of deep learning, object detection can be divided into two groups: two-stage detectors and one-stage detectors. Two-stage detectors such as [1, 2] first generate some RoIs in the first stage and make an object classification and RoI-wise bounding box regression next. One-stage detectors, e.g., YOLO [3] and SSD [4], do not generate the RoIs and directly detect objects. Owing to extreme imbalance of foreground–background class, the performance of two-stage detectors is usually better than one-stage detectors. Anchor-free detectors are used to address this problem, such as [2, 5, 6]. It alternatively transforms object detection into a points detection problem to avoid complex computations of anchors and run faster.

To recognize and locate objects in remote sensing images more effectively, the research of remote sensors detection is urgent. In recent years, the research on object

X. Feng (✉)
Tiangong University, Tianjin 300387, China
e-mail: 1930081292@tiangong.edu.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_8

detection is mostly based on Convolution Neural Network (CNN). For example, Region-based Convolutional Neural Networks [7] (R-CNN), known as a pioneering method, first generated region proposals using selective search and then refined them by extracting regional features from a convolution network. A region proposal network and an end-to-end trainable detector have been proposed to improve performance, which is named Faster R-CNN [8]. The Feature Pyramid Networks [9] (FPN) constructed a feature pyramid and predicted different objects at different pyramid feature maps by the scales of the region proposal. RetinaNet [10] chose a feature pyramid network likely FPN as its backbone and introduced a new focal loss to alleviate the imbalance between easy and hard examples. In aerial images, however, since the objects are mostly very small, these methods do not have good results in detecting them. This presents us with great challenges.

In recent years, many methods based on feature pyramid have been proposed. This is because FPN can combine low-level high-resolution information with higher-level strong semantic information, and simultaneously predict at different levels using lower-level features and higher-level features. As a result, targets in remote sensing images are not too small to be ignored by the detectors. Mou et al. [11] proposed a method to establish a feature pyramid network at all scales with strong semantic feature maps, which use a top-down pathway and horizontal connection. The feature map of different layers was responsible for detecting objects of different sizes. A dense feature pyramid network (DFPN) has been proposed by Yang et al. [12] to achieve automatic detection of ships: each feature map was closely linked and combined by concatenation.

With the improvement of the above methods, the ability of FPN network to recognize small objects has been improved, but some problems still exist. FPN proposes different features at each layer of the image pyramids, and then makes corresponding predictions. The shallow networks in the feature pyramid are more concerned with details and location information, while the upper layers focus more on semantics, which helps locate objects. First, feature maps of higher levels contributed to enhance the semantic information of lower levels. Second, the topmost convolution layer losses some information due to a few feature channels and is not compatible with other feature levels since it only has single-scale context information. So, the feature map on the top layer is very important to detect. To improve this shortcoming, we propose a method to enrich the top-level feature information. We use a five-layer feature pyramid network (C_1-C_5), and our method uses residual branch to get a new convolution layer C_6. Residual branch is used to indoctrinate the original branches with different spatial background information. Generation of a new convolution layer C_6 is used to alleviate the loss of information due to reduced channel convergence.

In addition to the above method, we also introduce super-resolution (SR) technology to enrich some detailed information of feature maps. Image super-resolution refers to make recovery in images or image sequences from low-resolution (LR) to high-resolution (HR). In general, the higher the resolution of an image the more detail and information it contains. However, the resolution is not the same as the pixel size. For example, an image that is multiplied by five by an interpolation does not tell you how much detail it contains. Image super-resolution is concerned with recovering the

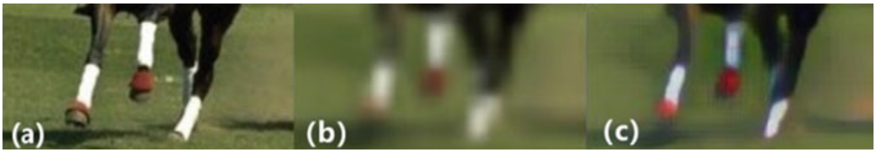


Fig. 8.1 The figure is an example of the SR technique, **a** is the ground truth, **b** is the low-resolution image, and **c** is the recovered high-resolution image

missing details in the image, that is high-frequency information. Figure 8.1 shows an example of SR technology, where a is the clear image, b is an image that needs to be restored to high resolution, and c is the result of the restoration. As you can see from the image, the restored image with SR contains more details and information. We use sub-pixel convolution to enrich the detail in the case of high-level details so that C_5 has more information. We hope this method can reduce the information loss and improve the performance of generated feature pyramids.

In order to realize the above method, we first improve the network structure of the traditional feature pyramid and propose a module to add a convolution layer before multi-scale feature fusion. The module also recalculates the fusion weight to fuse the extracted multi-scale feature layers more effectively. Finally, we introduce sub-pixel convolution to improve the semantic richness of the feature map to reduce the loss of detail.

8.2 Methods

Previous methods cannot solve the problem of incompatibility between high-level feature map and other level feature map. We propose a new RBFF network consisting of residual branches and sub-pixel convolution which is to detect small objects in aerial images. Figure 8.2 shows the framework of our method. The module we designed performs several operations on the tensor in order to fuse feature maps more efficiently. In addition, we use the sub-pixel convolution to enrich the high-frequency information of the feature map. Our method is described in detail below.

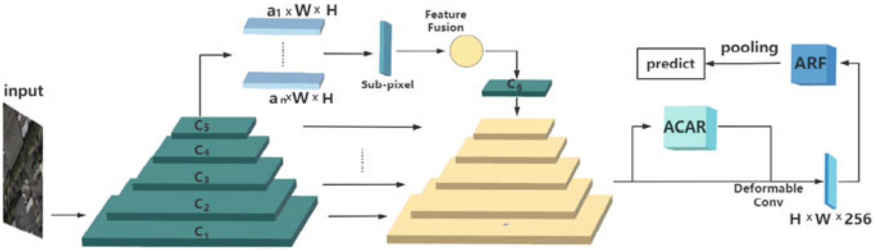


Fig. 8.2 The figure shows the RBFF network architecture

Our method adds a residual branch to generate a new feature map C_6 and recalculate weights. These features are then fused with recalculated weights. The ACAR module consists of the anchor classification branch and the anchor regression branch. Then we sent the anchor box and input feature maps into the deformable convolution [6] to extract aligned features. Finally, the active rotating filter [13] (ARF) is used to extract invariant directional features and produce the final detection results.

8.2.1 Sub-pixel Convolution

Most remote sensing images are very large. For example, the size of images in the DOTA dataset is about 4000 × 4000, and small objects like vehicles have very little information in the image. In addition, when the image is extracted by the feature pyramid network, there is less detail left, making it impossible to fully identify small objects in the image. The appearance of image super-resolution technology solves this problem.

In general, both I^{LR} and I^{HR} can have C color channels, thus they are represented as real-valued tensors of size rH × rW × C and rH × rW × C, respectively. There is a way to realize image super resolution is convolution that uses fractional stride of $\frac{1}{r}$ in the LR space. But this way will increase the computational cost because that process happens in the HR space. So, we use a convolution with stride of $\frac{1}{r}$ in LR space filters W_a of size k_a with weight spacing $\frac{1}{r}$, which do not active all W_a convolution. And we do not need to activate weights and do not need to calculate the weights which are between pixels. The activated pattern has activated at most $\lceil \frac{k_a}{r} \rceil^2$ weights. These patterns are activated periodically throughout the convolution, relying on the different sub-pixel positions: mod (a, r), mod (b, r) where a, b is the coordinates of output pixel in HR space. In this paper, we use a more effective way called sub-pixel convolution to achieve the above process when mod $(k_a, r) = 0$:

$$U^{SR} = t^K\left(U^{LR}\right) = VB(S_K \times t^{K-1}\left(U^{LR}\right) + c_K) \tag{8.1}$$

where VB is a periodic shuffle operator that ranges the elements of the $H \times W \times C \cdot r^2$ tensor again into a tensor of the size $rH \times rW \times C$. This operation can mathematically be described as follows:

$$PS(T)x, y, c = T_{\lfloor x/r \rfloor, \lfloor y/r \rfloor}, c \cdot r \cdot mod(y, r) + c \cdot mod(x, r) \tag{8.2}$$

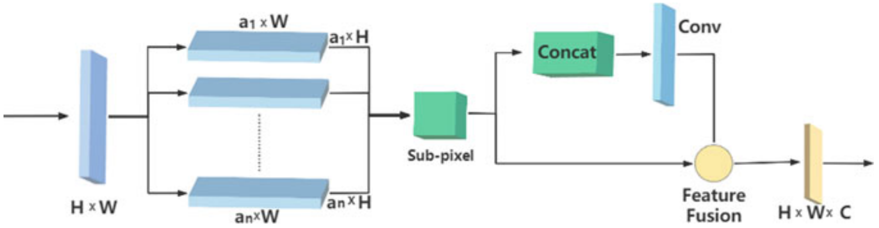


Fig. 8.3 The diagram shows the detailed structure of the residual branch that we propose. First of all, the topmost feature map has to go through three scales of adaptive pooling. Then the feature is amplified by sub-pixel convolution and then horizontally concatenated

8.2.2 Residual Branches

In the feature pyramid network, the top-down feature fusion process in the pyramids loses information at the top level due to fewer channels. To this end, we use a ratio-invariant adaptive pooling on the topmost layer of the feature pyramid to produce feature pyramid with different scales ($a_1 \times S$, $a_2 \times S$, ., $a_n \times S$) of multiple contextual features. To avoid the aliasing effects caused by interpolation, we have set three different scales to fit these contextual functions rather than simply summarizing them. Next sub-pixel convolution is used to scale up to the scale of S for subsequent fusion. Each context feature then independently passes through a 1 × 1 convolution layer, to reduce the channel dimension to 256 of the feature maps. Finally, in order to construct a feature pyramid, we use a 3 × 3 convolution layer at each feature map, as shown in Fig. 8.3.

8.3 Methods

8.3.1 Data Set

Our experiments were running primarily on the DOTA [14] dataset, which contains 2,806 aerial images of approximately 4000 × 4000 in size and 188,282 instances. And the dataset has 15 categories: plane (PL), ship (SH), storage tank (ST), baseball diamond (BD), tennis court (TC), basketball court (BC), ground track field (GTF), harbor (HA), bridge (BR), large vehicle (LV), small vehicle (SV), helicopter (HC), roundabout (RA), soccer ball field (SBF), and swimming pool (SP). It is marked as a quadrilateral with an arbitrary shape and orientation determined by four points rather than a traditional horizontal box. Specifically, first mark an initial point (x_1, y_1) and then mark 2, 3, and 4 in clockwise order. The initial point is usually selected at the head of the object. If it is an object such as a port with no obvious visual shape, choose the upper-left corner as the first point, as shown in Fig. 8.4.

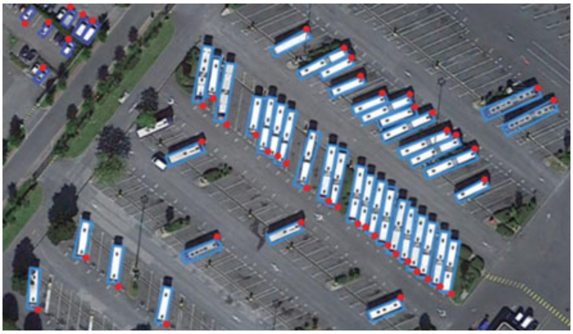


Fig. 8.4 The figure shows how the dataset labels are defined

Function of Loss. The loss function of our method consists of two parts. The loss function is defined as follows:

$$L = \frac{1}{N_R}\left(\sum_i L_c\left(c_i^R, l_i^*\right) + \sum_i 1_{l_i^* \geq 1} L_r\left(x_i^R, g_i^*\right)\right) + \frac{\lambda}{N_M}\left(\sum_i L_c\left(c_i^F, l_i^*\right) + \sum_i 1_{l_i^* \geq 1} l_r\left(x_i^F, g_i^*\right)\right), \tag{8.3}$$

where λ is a loss balance parameter, **1** is an indicator function, N_R and N_M are the numbers of positive samples in the ACAR and ARF, respectively, i is the index of a sample in a minibatch. c_i^R and x_i^R are the predicted category and refined locations of the anchor i in ACAR. c_i^F and x_i^F are the predicted object category and locations of the bounding box in ARF. l_i^* and g_i^* are the ground-truth category and locations of the anchor i. The Focal loss [10] and smooth $L1$ loss are adopted as the classification loss L_C and the regression loss L_R, respectively. The hyperparameters of Focal loss Lc are set to $\alpha = 0.25$ and $\gamma = 2.0$. We use the same training procedure as in Detectron [15].

8.3.2 Ablation Study

Residual Branches. In our approach, the network is enhanced by changing its structure and adding a new branch. To compare with another method, we use ResNet-50 as the backbone of the two methods. S^{2}A-Net [16] was chosen for comparison with our method. The result of using and not using residual branch are shown in Table 8.1. We use S^{2}A-Net to represent the S^{2}A-Net method and RBFF to show our method. Our method provides better detection results for small objects on the DOTA validation dataset.

Sub-pixel convolution. To test the impact of adding sub-pixel convolution on improving the accuracy of small target detection, we work on two tests with our

Table 8.1 Experimental results with different networks

Network	PL	BR	SV	LV	SH	TC	BC	ST
S^2A-Net	89.64	47.01	66.87	83.26	88.41	90.69	63.09	87.39
RBFF	**89.74**	**47.42**	**67.91**	**83.34**	**88.72**	**90.72**	**65.26**	**88.21**

Table 8.2 Comparison of the results of the experiment

Network	PL	BR	SV	LV	SH	TC	BC	ST
S^2A-Net	89.64	47.01	66.87	83.26	88.41	90.69	63.09	87.39
RBFF	**89.89**	**47.42**	**69.85**	**83.49**	**88.82**	**90.69**	**65.62**	**88.29**

network, one using sub-pixel convolution and the other not. Here we use sub-pixel to denote the network using sub-pixel convolution and S^2A-Net to denote the method we did not use. The result of adding sub-pixel convolution or not is shown in Table 8.2. The table shows that the use of sub-pixel convolution has a positive impact on the detection of small objects in general.

8.3.3 Comparison of Experimental Results

The RBFF method was compared with other popular methods in the DOTA dataset. The results of the experiment are shown in Table 8.3. In contrast to many previous works [13, 17] was designed to detect large scale targets, our experimental results presented in the table show detection results for nine types of objects which is aimed at evaluating the small objects. The mAP in the last row of the table is also the

Table 8.3 Comparison with other methods on DOTA dataset. FFA-3(M) implies the use of the multi-stage detector of FFA-3 for experiments

Method	Back	PL	GTF	SV	LV	SH	TC	ST	SBF	HA	mAP
RetinaNet [10]	R101	88.82	65.72	67.11	55.82	72.77	90.55	76.30	54.19	63.71	70.05
FFA-3 [18]	R101	88.80	57.90	63.60	75.90	79.60	**90.80**	82.90	54.30	66.90	71.49
FFA-3(M) [18]	R101	89.60	58.90	67.20	76.50	81.40	90.01	83.40	55.70	73.20	75.11
R^3Det [19]	R101	89.54	62.52	70.84	74.29	77.54	**90.80**	83.54	61.97	65.44	75.12
S^2A-Net [16]	R101	89.64	74.13	66.87	83.26	88.41	90.69	87.39	73.53	73.58	80.83
RBFF	R50	**90.05**	67.30	67.83	**83.33**	**88.62**	90.61	87.64	70.07	73.34	79.87
RBFF	R101	89.91	**75.82**	**70.49**	82.99	88.50	90.73	**87.92**	**74.65**	**75.25**	**81.81**

average of the detection of these 9 types of objects. From the result, it is clear that our method outperforms some previous detection methods. With the default input size, e.g., 1024 × 1024, RBFF can run at 399 ms per image on the RTX2080. A single-scale test can run at 66 ms per image. Finally, some visualization of detection results can be seen in Figs. 8.5 and 8.6.

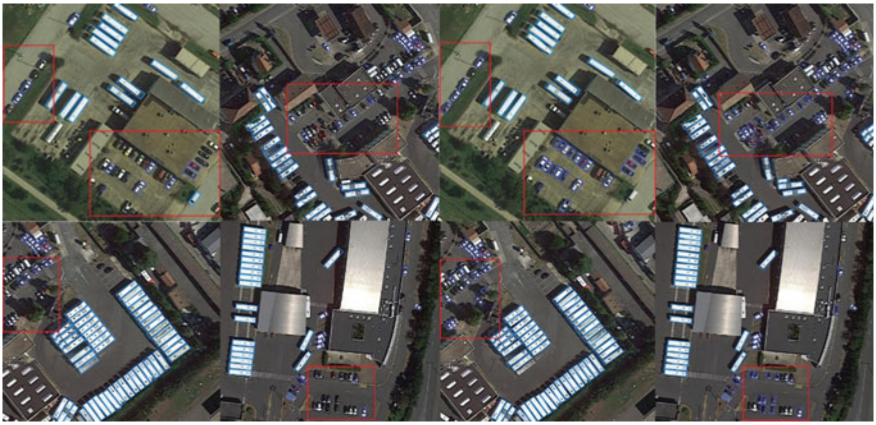


Fig. 8.5 The figure shows visualization results of our method. In the figure, the four pictures on the left are detection results of the S^{2}A-Net, and the four pictures on the right are the detection results of our method. Significantly more objects are identified in the red boxes in the four pictures on the right than on the left

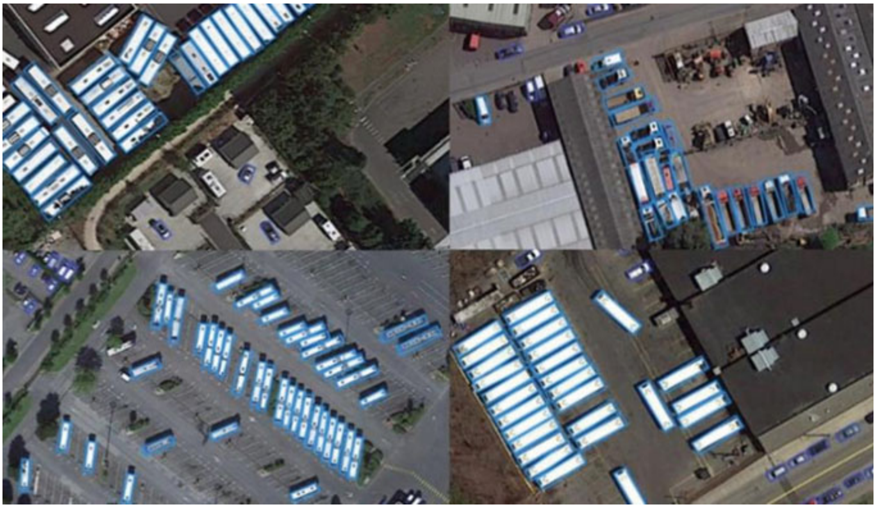


Fig. 8.6 This figure shows part of detection results obtained by our method

8.4 Conclusion

In this paper, a novel method for remote sensing detection has been proposed based on the feature pyramid network. Our method uses the residual branch to improve the network structure and reduce the feature loss that occurs during feature fusion. The features are then scaled by sub-pixel convolution. Our method uses the focal loss to better rebalance the variant scales of the bounding box. Multi-scale testing can significantly improve detection performance. Our RBFF was trained using ResNet-50-FPN and ResNet-101-FPN, both achieved good performance on DOTA dataset. I hope that our approach will be useful in the field of remote sensing object detection or data statistics.

References

1. Girshick, R.: Fast R-CNN. In: ICCV, pp. 1440–1448 (2015)
2. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points (2019). arXiv:1904.07850
3. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: CVPR, pp. 779–788 (2016)
4. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot multibox detector. In: ECCV, pp. 21–37 (2016)
5. Yang, Z., Liu, S., Hu, H., Wang, L., Lin, S.: Reppoints: point set representation for object detection. In: ICCV, pp. 9656–9665 (2019)
6. Law, H., Deng, J.: Cornernet: detecting objects as paired keypoints. In: ECCV (2018)
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Columbus, OH, United states, pp. 580–587 (2014)
8. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
9. Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection, vol. 2017-January, Honolulu, HI, United States, pp. 936–944 (2017)
10. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. **42**(2), 318–327 (2020)
11. Mou, L., Zhu, X.X.: Vehicle instance segmentation from aerial image and video using a multi-task learning residual fully convolutional network. IEEE Trans. Geosci. Remote Sens. **56**(11), 6699–6711 (2018)
12. Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., Guo, Z.: Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sens. **10**(1) (2018)
13. Zhou, Y., Ye, Q., Qiu, Q., Jiao, J.: Oriented response networks. In: CVPR, pp. 4961–4970 (2017)
14. Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.: DOTA: a large-scale dataset for object detection in aerial images. In: CVPR, pp. 3974–3983 (2018)
15. Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K.: Detectron (2018). https://github.com/facebookresearch/detectron
16. Han, J., Ding, J., Li, J., Xia, G.-S.: Align deep features for oriented object detection. IEEE Trans. Geosci. Remote Sens. (2021)
17. Chen, K., Ouyang, W., Loy, C.C., Lin, D., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J.: Hybrid Task Cascade for Instance Segmentation, pp. 4969–4978 (2019)

18. Fu, K., Chang, Z., Zhang, Y., Xu, G., Zhang, K., Sun, X.: Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. ISPRS J. Photogramm. Remote. Sens. **161**, 294–308 (2020)
19. Yang, X., Liu, Q., Yan, J., Li, A.: R3det: refined single-stage detector with feature refinement for rotating object. CoRR, vol. abs/1908.05612 (2019)

Part II
Deep Learning in Multidimensional Neural Networks

Chapter 9
Meta-learning with Logistic Regression for Multi-classification

Wenfeng Wang, Jingjing Zhang, and Bin Hu

Abstract The current classifiers and basic learners for few-shot meta-learning is based on distance rules and a series of linear classifiers, such as ridge regression, and linear support vector machine. This study introduces a nonlinear basic learner-logistic regression to improve meta-learning through fast convergence in learning downstream tasks and obtaining the global optimal solution. The Woodbury identity is utilized to express our advantages in a small number of samples. This helps to reduce the consumption in the process of matrix operation. The prototype network and residual network are employed as embedding models. The performance on data sets CIFAR-FS, FC100 and MiniImagenet demonstrate the competitiveness of our method.

9.1 Introduction

Meta-learning has been widely used in various fields [1]. Particularly, the model-agnostic meta-learning can be combined into unsupervised learning, few-shot learning and reinforcement learning [2]. These learning systems can adopt tasks to train and test and achieve the objective of meta-learning that minimize the generalization error loss [3–5]. The goal of meta-learning is to learn a function through a set of learning algorithms, as model-agnostic meta-learning which is widely used recently [6]. Maximum likelihood estimation is a method for us to find the maximum value of the log-likelihood function to form an unconstrained optimization problem.

W. Wang (✉) · J. Zhang
Shanghai Institute of Technology, Shanghai 201418, China
e-mail: wangwenfeng@nimte.ac.cn

W. Wang
Interscience Institute of Management and Technology, Bhubaneswar 752054, India

B. Hu
Changsha Normal University, Changsha 410111, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_9

In this paper, we also use the way of task training and we mainly focus on the maximum likelihood estimation of our model [3–5]. We mainly use it to update parameters, so that our objective function can find its global optimal solution, which can greatly reduce the training time of the model and the model achieve better training effect within the allowable range. And we adopt residuals network as our embedding model [7, 8].

The goal of the present study is to achieve the stability of the algorithm, minimize the training error in the training process, and at the same time achieve good generalization ability through test. For the parameter trajectories of logistic regression, we mainly form an unconstraint convex optimization problem, it is unlike SVM which adapts a constraint convex optimization problem [4]. We can use iterative reweighted least square method (IRLS) to get the solver of model [5].

9.2 Proposed Method

9.2.1 Problem Formulation

We have mainly undertaken the experiment on two data set—CIFAR-FS and FC100 and experiment on three forms of K ways N shot (5-way-5-shot, 5-way-1-shot, 5-way-2-shot) for classification. On the one hand, our method is mainly divided into two stages. One is the basic learner stage, which is mainly about learning how to calculate the value of w^i completed by logistic regression differentiation. As shown in Fig. 9.1, w^i are the weights of the linear classifier. The second is the meta-learning stage, which needs to improve the learning ability through back propagation error.

We mainly use meta-learning for few-shot learning gradient-based methods, using gradient descent methods to adapt new tasks [9, 10]. Meta-learning enables a few steps of gradient descent to obtain good parameters in parameter space. In logistic regression, the maximum likelihood estimation can be transformed into a minimum unconstrained optimization problem [11]. Meanwhile, logical regression has closed solution like ridge regression [5]. Our method requires a large amount of computation, which requires GPU to calculate the gradient and the solution of the model. As shown in the following Fig. 9.1, we have depicted the overview of our method; it illustrates 1-way 3-shot classification tasks and we adapt logistic regression method as our classifier. The embedding features of the training samples can be learned and obtaining the corresponding weights and testing examples are same. A task is a tuple for fewshot. Finally, the errors are minimized by the meta-learner.

We have traced back to the previous work of the meta-learning framework, explored the convex base learner again, and proposed the base learner [12] of logistic regression. And we compare it with other convex base learners, such as linear SVM and ridge regression.

According to the two components of the previous meta-learning algorithm, namely the base learner and the meta-learner [12], meta-learning is learning to learn, and it

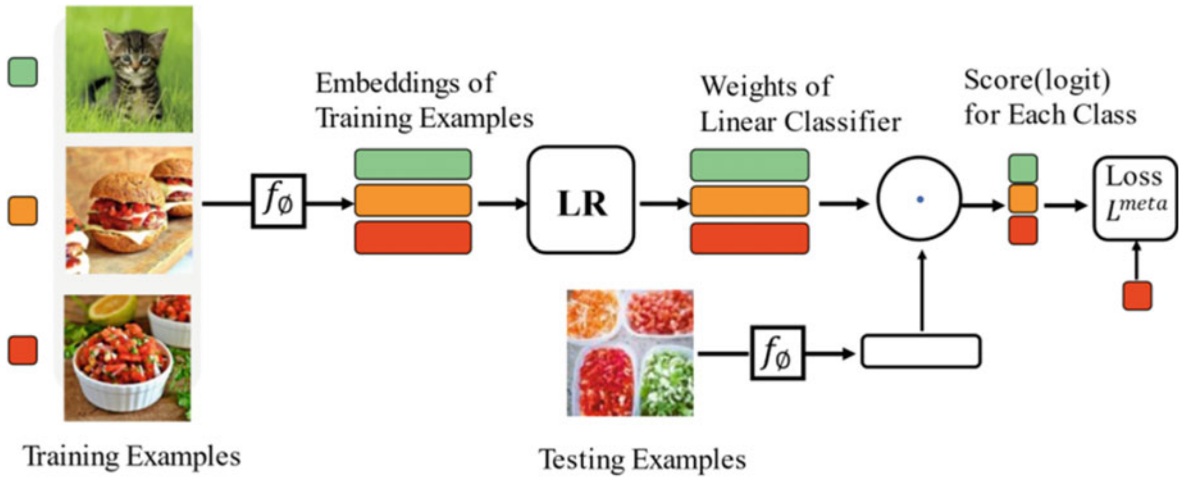

Training Examples
f_ϕ
Embeddings of Training Examples
LR
Weights of Linear Classifier
Testing Examples
f_ϕ
Score(logit) for Each Class
Loss L^{meta}

Fig. 9.1 A general overview of our method

is a good way to improve learning skills [13]. The goal of meta-learning is to make the base learning algorithm adapt well to new episodes.

Given a data set $S = \{x_i, y_i\}_{i=1}^{n}$, which includes a meta-training set and a meta-test set, the meta-training set and a meta-test set also include a training set and a test set, but we named it support set and query set. The support set is used for training, and the query set is used for testing so that they construct a task for training. In this paper, there are a group of tasks that is used as a meta-training set $I = \{(D_i^{train}, D_i^{test})\}_{i=1}^{I}$, $D_i^{train} \cap D_i^{test} = \varnothing$. The embedded model is parameterized mainly through $\varnothing$ that mainly uses the support set of the meta-training set. Given J tasks for meta-test J = $\{(D_i^{train}, D_i^{test})\}_{j=1}^{J}$. As we have shown that Fig. 9.2 explains the partition process of data set. The data set is mainly composed of two parts, one is the test set, the other is the training set. At the same time, the test and training set includes support set and query set.

In this paper, the base learner is to estimate the parameter θ of $f(x;\theta)$, here we use the method of university function approximation [14] $y = f(x;\theta)$, and base learner $\mathcal{B}$ is used to achieve better generalization ability. We write it as:

$$\theta = \mathcal{B}\left(D^{train};\varnothing\right) = \underset{\theta}{argmin}\mathcal{L}^{base}\left(D^{train};\theta,\varnothing\right) + R(\theta) \quad (9.1)$$

where $\mathcal{L}^{base}$ is the loss function which is computed by the base learner, such as the negative log-likelihood function. As we all know, $R(\theta)$ is a regularization of a function which plays a great important to generalize the loss [15]. As with most meta-learning methods, we regard the training program as episodes, so each episode

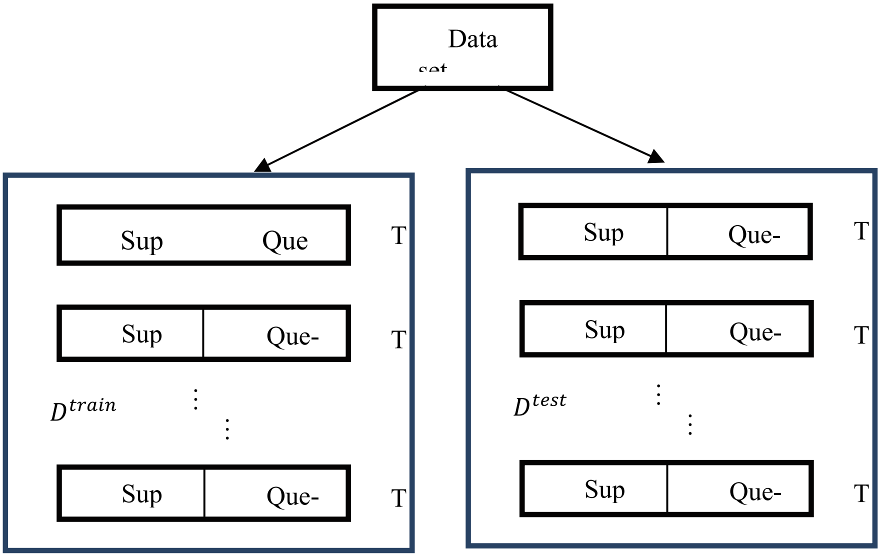


Fig. 9.2 The partition of data set

can be regarded as a small sample classification problem. Usually, the classification of small samples adopts the classification method of K-way and n-shot [16]. Here, we need to consider the values of K and N. Generally, $N = \{1, \ldots, n\}$. In the above, we have described the tasks, a task (or episode) $\daleth_i = (D_i^{train}, D_i^{test})$. Simultaneously, $D_i^{train} \cap D_i^{test} = \varnothing$ and D_i^{val} also disjoint with them.

9.2.2 *Efficient Logistic Regression Convex Optimization*

The base learner is mainly based on the principle of logistic regression, which is an unconstrained optimization problem. Therefore, we need to discuss the first-order and second-order optimality condition [17, 18], and we first give the unconstrained optimization problem:

$$\theta = \mathcal{B}\big(D^{train}; \varnothing\big) = argmin - \sum_{i=1}^{N} lnp(Y_i | X_i, w_1, \ldots, w_M) + \frac{\lambda}{2} w^T w \tag{9.2}$$

where λ is the regularization and $D^{train} = \{(x_n, y_n)\}$, Y_i is the labels of dataset, $\theta = \{w_k\}_{k=1}^{K}$. Because our objective function is differentiable and convex and there is the quality that if the objective function is continuously differentiable, a practical optimality judgment condition can be obtained by virtue of the property of continuous differentiable function.

Theorem 9.1 (The necessary condition of first order) *If x^* is the local optimal solution of the unconstrained optimization problem* [19], *then* $\nabla f(x^*) = 0$.

Theorem 9.2 (The sufficient condition of second order) *When you suppose that point x^* is the local optimal solution of the unconstrained optimization problem, and if f(x) is continuously differentiable for second order in the neighborhood of point x^*, then*

$$\nabla f\big(x^*\big) = 0 \; and \; \nabla^2 f\big(x^*\big) > 0 \tag{9.3}$$

where $\nabla^2 f(x^*)$ represents Hessian matrix is positive defined, then x^* is a strictly local optimal solution of f(x).

Now we consider the logistic regression multi-class classification problem. Given data have a total of M classes, and each sample x_i corresponds to a vector (or one-hot label) $y_i = [y_{i1}, \ldots, y_{iM}]^T$ of M dimension. Each element of y_i is 0 or 1: If x_i belongs to m-th class, then $y_{im} = 1$, and all other elements are 0. The multinomial logistic regression model uses the following soft-max function as the sample x of the conditional probability belongs to the m class [20].

$$p(y_m = 1|x) = \frac{exp(w_m^T x)}{\sum_{j=1}^{M} exp((w_j^T x))} \tag{9.4}$$

where $w_1, \ldots., w_M$ are the parameters of our model.

We use the following distribution:

$$p(y_m = 1|x) = \sigma\left(w_m^T x\right) = \frac{exp(w_m^T x)}{1 + \sum_{j=1}^{M-1} exp(w_j^T x)}, m = 1, \ldots, M - 1 \tag{9.5}$$

$$p(y_M = 1|x) = 1 - \sigma\left(w_m^T x\right) = \frac{1}{1 + \sum_{j=1}^{M-1} exp(w_j^T x)} \tag{9.6}$$

The likelihood function of a single sample is:

$$p(Y_i|X_i, w_1, \ldots, w_M) = \prod_{m=1}^{M} p(y_{im} = 1|x_i)^{y_{im}} \tag{9.7}$$

Therefore, the likelihood function for the meta-training set is:

$$p(Y_i|X_i, w_1, \ldots, w_M) = \prod_{i=1}^{N} \prod_{m=1}^{M} p(y_{im} = 1|x_i)^{y_{im}} \tag{9.8}$$

And we can get the log-likelihood function:

$$lnp(Y_i|X_i, w_1, \ldots, w_M) = \sum_{i=1}^{N} \sum_{m=1}^{M} y_{im} lnp(y_{im} = 1|x_i) \tag{9.9}$$

Newton's-Method and Solving Unconstrained Optimization Problems

Newton's method is a descent method. The difference between Newton's method and gradient descent method lies in the choice of descent direction [21, 22]. For unconstrained optimization problem:

$$min f(x) \tag{9.10}$$

Assuming that f is a convex function and second-order differentiable (the domain is an open set), then the second-order Taylor approximation of f(x) near x is:

$$\widehat{f}(x + v) = f(x) + g(x)^T v + \frac{1}{2} v^T H(x) v \tag{9.11}$$

where $g(x) = \nabla f(x)$ is a gradient, $H(x) = \nabla^2 f(x)$ is a Hessian matrix. Must be noted that the above is only a quadratic approximation, not a complete Taylor expansion.

If x is regarded as a constant, then the above expression is a quadratic function of v, minimized with respect to v, making the gradient zero:

$$g + Hv = 0 \rightarrow v = -H^{-1}g \tag{9.12}$$

It is the Newton step. Since H is positive definite, its inverse is also positive definite,

$$g^T \Delta x_{nt} = -gH^{-1}g \tag{9.13}$$

Unless g = 0, Δx_{nt} is the descent direction. When f is a quadratic function, $x + \Delta x_{nt}$ is its minimum point; As f approaches quadratic, $x + \Delta x_{nt}$ is a good estimate of its minimum point [23]; Since f is quadratic differentiable, the quadratic approximation is very accurate around the minimum value, and $x + \Delta x_{nt}$ is a good estimate of the minimum point [24]. The steps of Newton's method are similar to those of gradient descent, except that the direction of descent is $\Delta x_{nt} = -H^{-1}g$.

There's an objective function (9.2). We should judge whether our goal function is positive definite or not. So let's calculate the gradient:

$$\lambda w + \sum_{i=1}^{N} \frac{-y_i x_i exp(-y_i w^T x_i)}{1 + exp(-y_i w^T x_i)} = \lambda w + \sum_{i=1}^{N} -y_i x_i [1 - \sigma(y_i w^T x)] \tag{9.14}$$

$$g_k = \lambda w_k + \sum_{i=1}^{N} -y_i x_{ik} [1 - \sigma(y_i w^T x_i)] \tag{9.15}$$

where w_l is the lth element of w, and x_{ik} is the kth element of sample x_i, $\sigma(y_i w^T x)$ is sigmoid function. To calculate the Hessian matrix, we need:

$$\frac{\partial \sigma(y_i w^T x_i)}{\partial w_l} = \frac{exp\left(-y_i w^T x_i\right)}{\left[1 + exp\left(-y_i w^T x_i\right)\right]^2}(y_i x_{il}) = \sigma\left(y_i w^T x_i\right)[1 - y_i w^T x_i](y_i x_{il}) \tag{9.16}$$

Let's calculate the elements in k row of the Hessian matrix, k, l = 0, 1..., K. When $k \neq l$,

$$\begin{aligned} H_{kl} = \frac{\partial g_k}{\partial w_l} &= \sum_{i=1}^{N} y_i x_{il} \frac{\sigma\left(y_i w^T x_i\right)}{\partial w_l} \\ &= \sum_{i=1}^{N} \sigma\left(y_i w^T x_i\right)[1 - \sigma\left(y_i w^T x_i\right)(y_i x_{il})(y_i x_{il})] \end{aligned}$$

$$= \sum_{i=1}^{N} \sigma\left(w^T x_i\right)[1 - \sigma\left(w^T x_i\right)]x_{il}x_{ik} \tag{9.17}$$

When $k = l$,

$$H_{kl} = \frac{\partial g_k}{\partial w_l} = \lambda + \sum_{i=1}^{N}\sum_{i=1}^{N} \sigma\left(w^T x_i\right)[1 - \sigma\left(w^T x_i\right)]x_{il}x_{ik} \tag{9.18}$$

Noting the matrix $X = [x_1, x_2, \ldots, x_N]$, $A_{ii} = \sigma\left(w^T x_i\right)\left[1 - \sigma\left(w^T x_i\right)\right]$, the Hessian matrix of (9.2) is

$$H = \lambda I + \sum_{i=1}^{N} \sigma\left(y_i w^T x_i\right)\left[1 - \sigma\left(y_i w^T x_i\right)\right]x_i x_i^T = \lambda I + \sum_{i=1}^{N} A_{ii} x_i x_i^T = \lambda I + XAX^T \tag{9.19}$$

where A is a diagonal matrix of order N, whose elements in i row and i column are A_{ii}, $A_{ii} > 0$.

Because $u^T Hu = \lambda u^T u + (X^T u)^T A\left(X^T u\right) > 0, \forall u \neq 0$, so H is positive definite, function (9.2) is a convex function, problem $min - \sum_{i=1}^{N} \ln[1 + \exp(-y_i w^T x_i)] + \frac{\lambda}{2} w^T w$ for unconstrained convex optimization problem.

9.2.3 Approach to the Objective of Meta-learning

When we want to solve unconstrained optimization problems [25], before we do that, we must determine this is a convex optimization problem. The convex function is determined by the Hessian matrix of the objective function $\mathcal{L}^{base}$, for which the Hessian matrix $H = \frac{\partial^2 \theta(w)}{\partial w \partial w^T}$ is positive defined.

$$\begin{aligned}\theta = \mathcal{B}\left(D^{train}; \varnothing\right) &= \underset{\theta}{argmin} \mathcal{L}^{base}\left(D^{train}; \theta, \varnothing\right) + R(\theta) \\ &= argmin - \sum_{i=1}^{N} lnp(Y_i | X_i, w_1, \ldots, w_M) + \frac{\lambda}{2} w^T w\end{aligned} \tag{9.20}$$

We can confirm that the Hessian matrix of our objective function satisfies the condition of the theorem.

And in order to obtain a closed solution, we must consider using an iterative method to solve it. In there we adopt iteratively reweighted least squares (IRLS) method to optimize the problem, the following iteration [26]:

$$w^i = w^{i-1} - H^{-1} g \tag{9.21}$$

H is the Hessian matrix of objective function. The number of Newton steps related to the Hessian matrix can be obtained by the second-order Taylor approximation of the objective function. Among them, the ith iteration updates the parameters

$$H_i = \lambda I + XAX^T, g_i = \lambda w - XAt \tag{9.22}$$

$t_i = \frac{y_i[1-\sigma(y_i w^T x_i)]}{A_i}$, $A = \sigma(w^T X)[1 - \sigma(w^T X)]$, σ is the sigmoid function, g_i is the gradient. So the formula can be obtained by substituting (9.22) into (9.21) that we can compute:

$$w^i = (XAX^T + \lambda I)^{-1} XAz \tag{9.23}$$

where

$$z = (X^T w^{i-1} + t) \tag{9.24}$$

$$z_i = X^T w^{i-1} + t_{i-1} = X^T w^{i-1} + \frac{y_i[1 - \sigma(y_i w^T x_i)]}{A_i} \tag{9.25}$$

$$A_i = \sigma(w^T x_i)[1 - \sigma(w^T x_i)] \tag{9.26}$$

$min - \sum_{i=1}^{N} lnp(Y_i|X_i, w_1, \ldots, w_M)$ also called the cross-entropy error function of logistic regression multi-classification [27].

Although there are many options for measuring losses, here we use a negative log-likelihood function to measure losses, which are same as in the paper of prototype network [28, 29]. The negative log-likelihood function can measure the performance of the meta-test sample, and we think it is very effective way to adopt this function.

$$L^{meta}(D^{test}; \theta, \varnothing, \alpha) = \sum_{(x,y)\in D^{test}} [-\alpha w^i f_{\varnothing}(x) + log \sum_k exp(\alpha w^j f_{\varnothing}(x))] \tag{9.27}$$

where $\theta = \mathcal{B}(D^{train}; \varnothing) = \{w^j\}_{j=1}^{K}$ and α is a parameter which can be learned from the process.

9.3 Results and Discussions

In this paper, we mainly use Resnet and prototypical networks as our embedding model. When experiment on the CIFAR and FC100 data set, the network architecture: R64-MP-DB(0.9,1)-R160-MP-DB(0.9,1)-R320-MP-DB(0.9,2)-R640-MP-DB(0.9,2). We initially set the learning rate to 0.1 and change to 0.006 at epoch

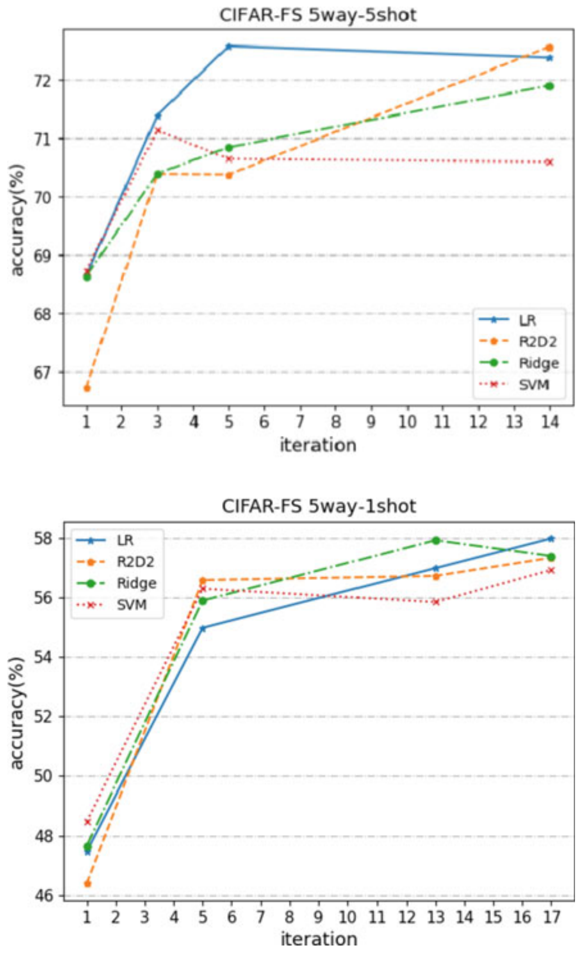
Table 9.1 Comparison of other algorithms on CIFAE-FS and FC100. Average few-shot classification accuracies (%) which on the backbone Resnet12. 'R2D2' and Ridge stand for ridge regression but for two different forms. 'LR' stands for the logistic regression

		CIFAR-FS		FC100	
Model	Backbone	1-shot (%)	5-shot (%)	1-shot (%)	2-shot (%)
R2D2	Resnet12	55.52	71.81	31.18	36.91
Ridge	Resnet12	55.19	71.56	31.24	37.03
SVM	Resnet12	55.53	71.33	31.35	37.41
LR	Resnet12	55.60	71.91	31.14	36.78

20. The use of such parameters here is in full compliance with the criteria of gradient descent. We referred to the corresponding parameter settings in the Meta-learning of different- able convex optimization [2]. In order to make full use of the device's availability and available memory space, we tried to set epochs as 20 for many times, which was a wise choice because the GPU often needed to carry out a lot of calculations in the case of many tasks, which would cost a lot of time. The minibatch consists of 8 episodes and every epoch consists of 1000 episodes. And Table 9.1 shows the result of our method and make a comparison to other base learners.

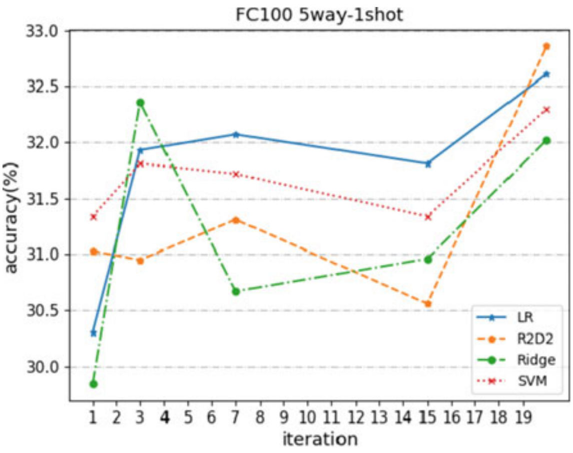
As shown in Table 9.1, LR as our base learner can achieve better performance and be more stable when we use CIFAR-FS data set. As shown in Figs. 9.3 and9.4, we compare four base learners with the same k-way n-shot(5-way 1-shot; 5-way 2-shot; 5-way 5-shot) on CIFAR-FS data set and FC100 data set, MiniImagent data set, it depicts our method can stably get the results. But when we use data set FC100, we find that SVM method will be more efficient to test tasks. In this way, although logistic regression method in FC100 data set doesn't get enough good results but it can confirm that it can be stable for classification. At the same time, it also reflects the authenticity of experiments, the whole operation process is you don't know FC100 data gathering in the effect of the LR algorithm accuracy is lower than the other. It is believed that LR meta-learning has better stability than the other three kinds of algorithms, so it can be as our further exploration work, we can explore that the logistic regression meta-learning algorithm better adapts to all of the downstream tasks. However, when we use MiniImagenet data set to achieve our method, the base learner of SVM becomes the lowest of accuracy in Table 9.2. And LR as the base learner will get 62.48% accuracy with 5-way 5-shot. As shown in Table 9.1, the more samples there are, the higher the accuracy will be. 5-shot means there gives five samples, and 2-shot means there gives only two samples. Therefore, these two samples and five samples will be more accurate than one sample; either a 5-way 10-shot or a 5-way 15-shot (Table 9.2).

Fig. 9.3 Comparison for four base learners with the same k-way n-shot on CIFAR-FS data set



9.4 Conclusion

In this paper, we mainly show that the performance of logistic regression as the base learner and compare it to other base learners. Our method principally considers the unconstrained optimization problem, and the closed-form solution can be obtained through the iterative method. Moreover, experiments have been carried out on all three data sets, which are fully reflected in the figure above. Finally, we make the conclusion that logistic regression method can stably run than other base learners when there are less epochs as you can see in Figs. 9.3, 9.4, and 9.5. And we just adopt 3 ways to experiment with our convex base learner, it can be seen, our method performs well in CIFAR-FS. At the running level, we further save the time to run our process, because data set is great and the process will be long and complex. It is also an effective way to classification as a base learner after embedding features.



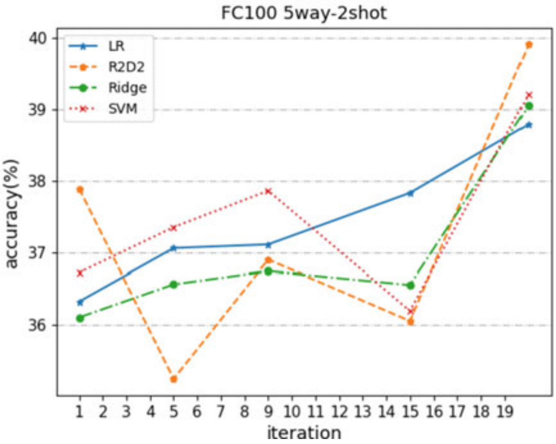


Fig. 9.4 Comparison for four base learners with the same k-way n-shot on FC100 data set

Table 9.2 Comparison of other algorithms on MiniImagenet dataset. Average few-shot classification accuracies (%) which on the backbone 64-64-64-64. 'R2D2' and Ridge stand for ridge regression but for two different forms. 'LR' stands for the logistic regression

	MiniImagenet	
Model	Backbone	5-way 5-shot (%)
R2D2	64-64-64-64	62.38
Ridge	64-64-64-64	62.18
SVM	64-64-64-64	60.59
LR	64-64-64-64	62.48

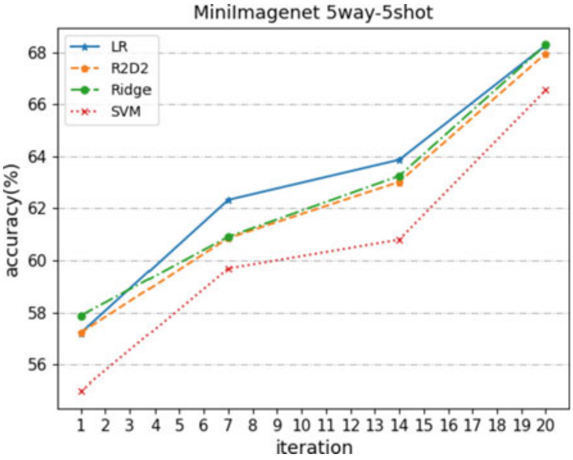


Fig. 9.5 Comparison for four base learners with the same 5-way 5-shot on MiniImagenet data set

Acknowledgements This research was supported by the Shanghai High-Level Base-Building Project for Industrial Technology Innovation (1021GN204005-A06).

References

1. Abramson, N., Braverman, D.J., Sebestyen, G.S.: Pattern recognition and machine learning. Publ. Am. Stat. Assoc. **103**(4), 886–887 (2006)
2. Lee, K., et al.: Meta-learning with differentiable convex optimization. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2019)
3. Menard, S.: Logistic regression. Am. Stat. **58**(4), 364 (2004)
4. Hsu, C.W., Lin, C.J.: A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. **13**(2), 415–425 (2002)
5. Cottle, R.W., Olkin, I.: Closed-form solution of a maximization problem. J. Global Optim. **42**(4), 609–617 (2008)
6. Yuan, K., Ling, Q., Yin, W.: On the convergence of decentralized gradient descent. SIAM J. Optim. **26**(3) (2013)
7. Demiralp, C., Scheidegger, C.E., Kindlmann, G.L., et al.: Visual embedding: a model for visualization. IEEE Comput. Graph. Appl. **34**(1) (2014)
8. Xu, Z., Chen, X., Tang, W., et al.: Meta weight learning via model-agnostic meta-learning. Neurocomputing **432**(7587) (2020)
9. Li, Z., Zhou, F., Fei, C., et al.: Meta-SGD: Learning to Learn Quickly for Few-Shot Learning (2017)
10. Rich, C.: Multitask learning. Mach. Learn. (1997)
11. Silvestre, L.: On the differentiability of the solution to the Hamilton-Jacobi equation with critical fractional diffusion. Adv. Math.-N. Y. **226**(2), 2020–2039 (2009)
12. Bertinetto, L., et al.: Meta-learning with Differentiable Closed-Form Solvers (2018)
13. Chen, Y., Guan, C., Wei, Z., et al.: MetaDelta: A Meta-Learning System for Few-Shot Image Classification (2021)
14. Vilalta, R., Giraud-Carrier, C., Brazdil, P.: Meta-learning. In: Data Mining & Knowledge Discovery Handbook (2005)

15. Bartlett, P.L., Helmbold, D.P., Long, P.M.: Gradient descent with identity initialization efficiently learns positive-definite linear transformations by deep residual networks. Neural Comput. (2019)
16. Dan, L.I., Gao, H.Y., Chen, S., et al.: A proximal gradient method for solving a class of bilevel programming problem. J. Dalian Univ. (2019)
17. Nichol, A., Achiam, J., Schulman, J.: On First-Order Meta-learning Algorithms (2018)
18. Guo-Xun, et al.: A comparison of optimization methods and software for large-scale L1-regularized linear classification. J. Mach. Learn. Res. **11**(11), 3183–3234 (2010)
19. Dontchev, A.L., Rockafellar, R.T.: Solution Mappings for Variational Problems. Springer, New York (2014)
20. Song, T., Song, Y., Wang, Y., et al.: Residual network with dense block. J. Electron. Imaging **27**(PT.2), 053036.1–053036.9 (2018)
21. Stachowiak, M.K.: Cross-entropy method in application to the sirc model. Algorithms **13**(11), 281 (2020)
22. Boyd, S., Crusius, C., Hansson, A.: Advances in convex optimization: theory, algorithms, and applications. IFAC Proc. Vol. **30**(9), 365–393 (1997)
23. Hosmer, D.W., Hosmer, T., Le, C.S., et al.: A comparison of goodness-of-fit tests for the logistic regression model. Stat. Med. **16**(9), 965–980 (2015)
24. Wu, J., Chen, S.P., Liu, X.Y.: Efficient hyperparameter optimization through model-based reinforcement learning. Neurocomputing (2020)
25. Lai, N., Kan, M., Han, C., et al.: Learning to learn adaptive classifier-predictor for few-shot learning. IEEE Trans. Neural Netw. Learn. Syst. (99), 1–13 (2020)
26. Jian, G., Zhang, L., Xiao, X.: Log-Sigmoid nonlinear Lagrange method for nonlinear optimization problems over second-order cones. J. Comput. Appl. Math. **229**(1), 129–144 (2009)
27. Rafi, R., Tang, B., Du, Q., et al.: Attention-based domain adaptation for hyperspectral image classification. In: IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE (2019)
28. Pahde, F., Puscas, M., Klein, T., et al.: Multimodal prototypical networks for few-shot learning. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE (2021)
29. Rokach, L.: Ensemble-based classifiers. Artif. Intell. Rev. **33**(1–2), 1–3 (2010)

Chapter 10
Measurement for Blade Edge Based on Machine Vision

Ke Zhang, Yunwen Zhu, Wenfeng Wang, and Jingjing Zhang

Abstract In view of problems that measurements of grinding blade edge mainly depend on manual detection, which is difficult to detect and has low measuring efficiency, a measurement method based on machine vision is designed. Through series of processing of blade edge image with software, the measurement results are displayed in real time, and the dynamic real-time monitoring of the grinding quality of blade is realized. Through the test, the system can successfully realize the measurement of blade edge size with high efficiency and accuracy, and realize the measurement automation, which has great application value.

10.1 Introduction

10.1.1 Advantage of Machine Vision on Blade

With the rapid development of machine vision technology, visual size measurements are applied to more and more manufacturing industry [1–3]. In the traditional blade manufacturing industry, blade detection has the following disadvantages: (1) The blade width inspection is at low efficiency; (2) The blade is very narrow with a certain angle, Generally, the traditional measuring tools, such as caliper, are not convenient to operate the measurement; (3) Cannot measure the blade width and grinding surface quality dynamically, in real time. According to the above problems, this experiment has designed a set of machine vision with the following advantages:

K. Zhang · Y. Zhu
School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai 201418, China

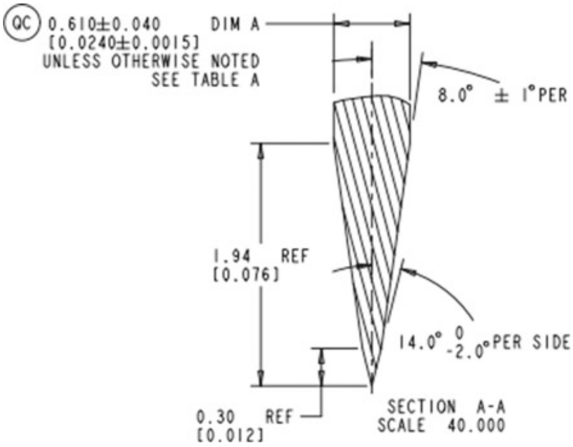
Y. Zhu
Shangyai Yeerui Imaging Technology Co. Ltd, Shanghai, China

W. Wang (✉) · J. Zhang
School of Science, Shanghai Institute of Technology, Shanghai 201418, China
e-mail: wangwenfeng@nimte.ac.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_10

Fig. 10.1 Fine grinding edge drawing



(1) Onsite dynamic inspection, the size beyond the tolerance range can be immediately detected and correct the sand wheel position; (2) Non-contact measurement, camera lens with high precision, measurement accuracy is higher; (3) The grinding surface quality can be intuitively seen in the screen. This experiment mainly has the following equipment: size measuring projector and visual measuring device (Fig. 10.1).

10.1.2 Noise and Denoising

In our life, noise is a factor that hinders people's sensory organs from understanding the received source information. There are also various factors in the image that prevent people from accepting their own information, that is, the familiar image noise. Noise is a very important factor affecting the quality of digital images. Its source is mainly in the process of image acquisition and transmission. In the process of image acquisition, the performance of camera or imaging sensor will be affected by many factors, such as external environmental conditions, sensor element quality and so on. Image noise will affect our reception of information expressed in images, which is our most intuitive understanding. In addition, image noise will also affect the scientific research and various applications of image recognition processing, such as face recognition, character recognition, flower recognition, target recognition, unmanned driving technology and other research and applications based on computer vision [4–6].

Image denoising technology is a kind of image restoration technology, which is a classic problem in the field of image processing. Up to now, scientists have put forward many different types of denoising methods [7–9], such as methods based on probability theory and methods based on statistical theory. Methods based on partial differential equations, linear and nonlinear filtering methods, frequency spectrum analysis, multi-resolution analysis and so on. Therefore, this paper mainly studies the non-local mean filtering denoising algorithm.

10.2 Research Method

Specific Research Steps of This Topic

Sampling 30 pcs blades to measure the real width of blade (see Fig. 10.2).

Below pictures show 30 pcs blade samples' real dimensions (Fig. 10.3).

10.2.1 The Selection of Camera, Lens, Lighting, Computer and Hardware Parts

According to the needs of the experimental system, below items were chosen to set up the system.

High-speed industrial camera:

Camera selection mainly depends on the chip.

At present, the mainstream chip manufacturers in the market are Sony semiconductor and onsime semiconductor. The chips produced by these two companies are trustworthy, but onsime is out of stock in China. Sony can guarantee normal delivery. The camera brand is Hikvision which is the first brand in China's domestic industrial

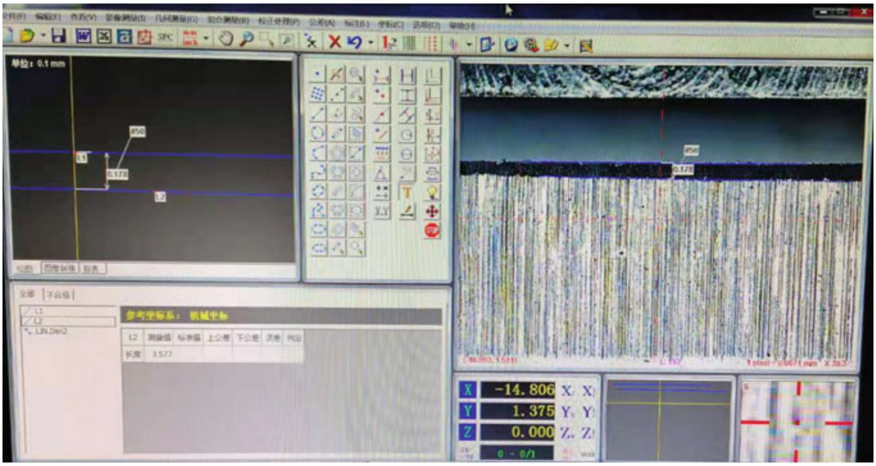


Fig. 10.2 Measuring fine grinding edge by projector

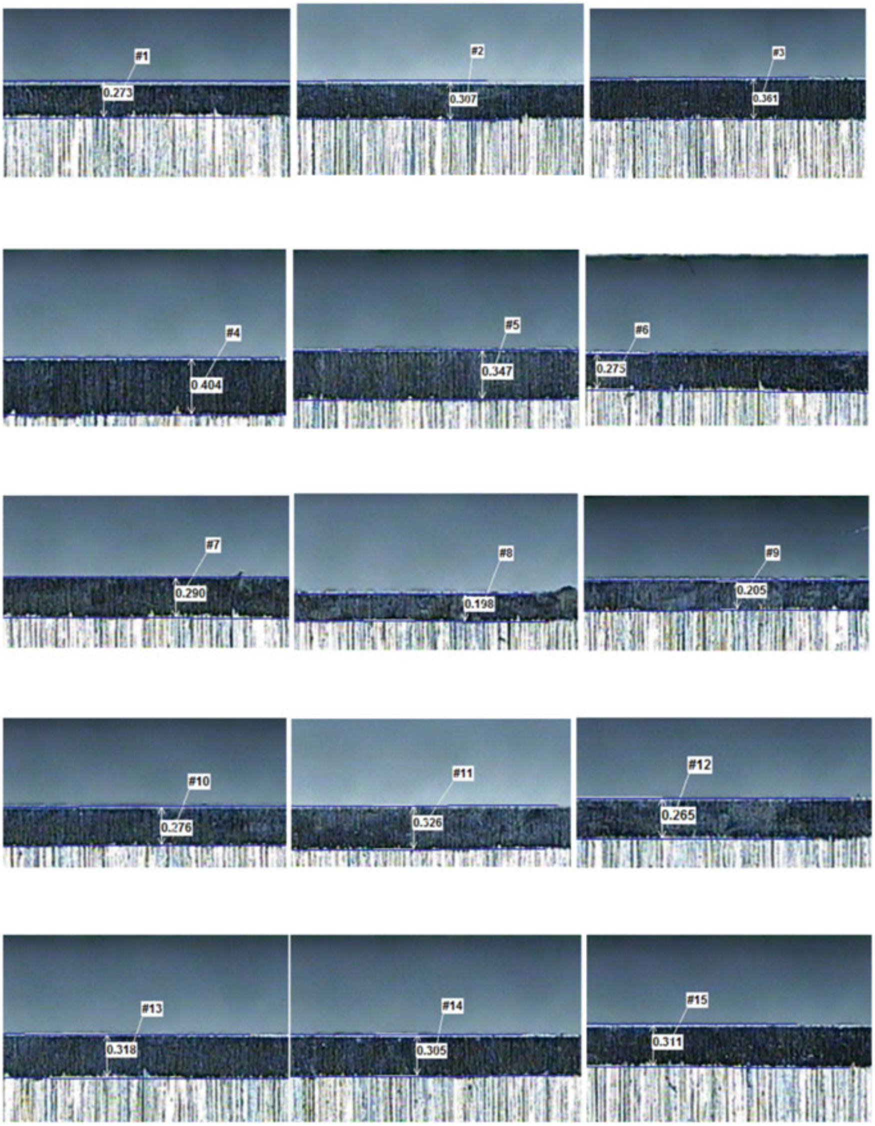


Fig. 10.3 Width of fine grinding edge measured by projector

camera market share. Hikvision is the best choice for cost performance at present. The selected Sony imx264 chip is a CMOS, global chip, which can collect pictures during movement and ensure that the images are not deformed. 5 million pixels can ensure that we not only have higher accuracy, but also have a larger field of vision and can be compatible with more products.

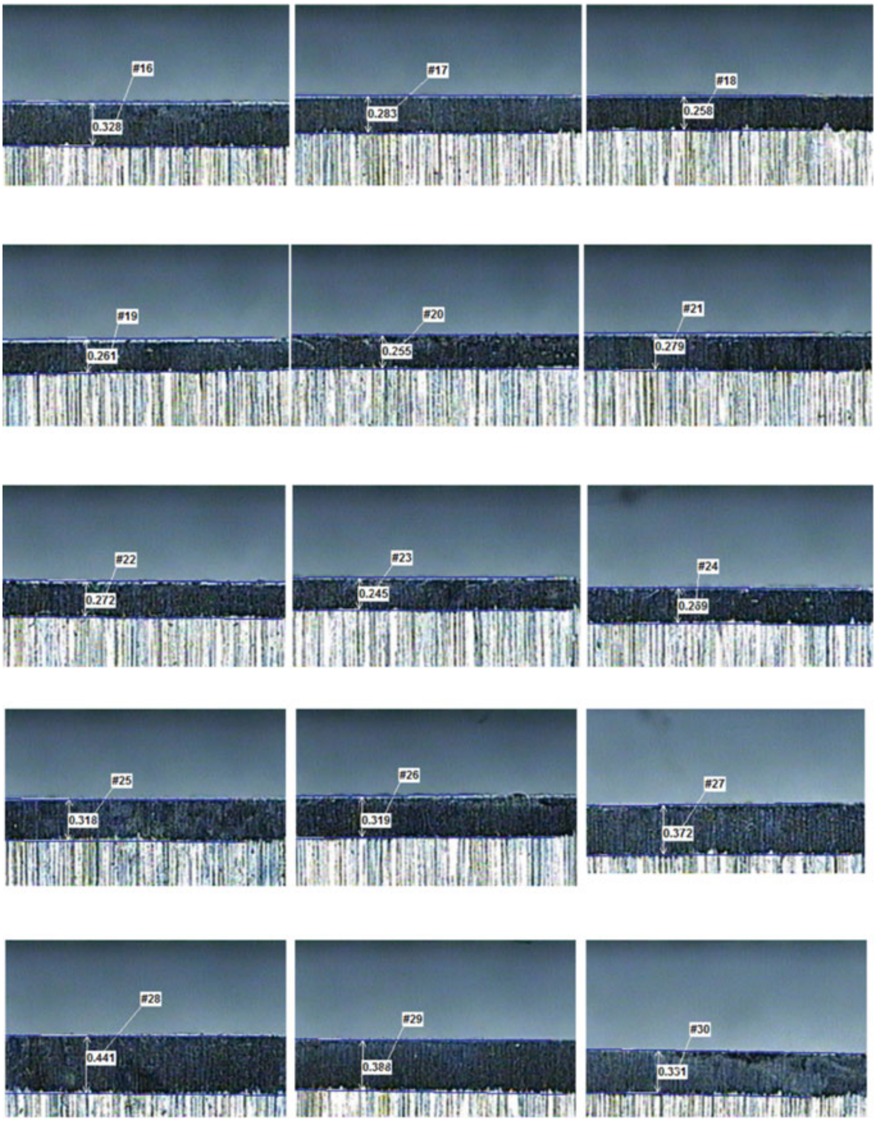


Fig. 10.3 (continued)

Resolution Ratio: 2448 × 2048;
Pixel Dimension: 3.45 µm × 3.45 µm;
Max.Frame Rate: 24.1 fps;
Dynamic Range: 72dB;
Signal Noise Ratio: 40.2 dB;
Gain: 0~20 dB;

Exposure Time: Min. Exposure mode: 1~14 μs; Normal Exposure mode: 15 μs~10 sec;

High-precision telecentric lens:

The resolution of the lens we selected matches the pixel size of the camera. The selected lens needs to have good definition and depth of field to ensure that there will be no difference in the collected images when the product shakes slightly during movement. The lens manufacturer is Italian opto, an internationally renowned industrial telecentric lens manufacturer. After comparing and testing the products of several manufacturers, the current lens is finally selected.

Apocenter Altitude: 0.06°;
Maximum Distortion: 0.20%;
Depth of Focus: 2.3 mm;
CTF @ 35 lp/mm >58%;
Working F value: 12.

Light source:

A light source with good condensing property to ensure that the features we need can be accurately reflected.

Lighting illuminance: 25,000~50,000 lx;

Vision system:

The accuracy of dimension measurement is required to be 0.02 mm/20 μm. The actual size of a pixel is 3.45 μm. After magnifying by the lens, each pixel can represent about 5.5 μm, this means the accuracy can meet the requirements.

Accuracy: 0.0055 mm/pixel;

Speed: 100 mm/s;

Acceptable edge Tolerance: ± 0.02 mm.

Select an appropriate calibration method to complete the calibration of the camera.

Set a suitable image processing process; optimize the efficiency of image processing, eliminate the influence of image noise on edge extraction, quickly and accurately measure the dimension of the edge.

10.2.2 Principle of Non-local Mean Filtering Algorithm

We say that noise is an obstacle to people's acceptance of image information and a stumbling block to computer vision research [10]. For example, in the process of shooting a video, the camera we use will be affected by light, and the shooting picture will also be affected by various large and small objects in the environment. In the process of transmission, the image quality will also decrease, and some random, discrete and isolated pixels will appear in the image due to factors such as channel transmission error. Therefore, it is difficult to avoid noise in the image, which will adversely affect people's image information reception, subsequent image processing and image visual effects.

Image denoising has been a classic problem in the field of image processing. In this part, we can divide image denoising methods into transform domain-based denoising methods and spatial domain-based denoising methods. The development status and existing problems of these two methods are discussed, respectively. On the basis of analyzing noise and signal characteristics, researchers have proposed a series of noise reduction algorithms in spatial domain and transform domain. Spatial domain denoising algorithm is to smooth pixels in the image directly to remove noise. Mean filtering, median filtering and Wiener filtering are all traditional filtering denoising algorithms. Transform domain denoising algorithm is to transform the image matrix into another space which is easier to identify. Then, the features of this space are used to denoise the image. Traditional denoising algorithms based on transform domain mainly include wavelet transform, Fourier transform and sparse representation of over-complete dictionary, which is a method based on multi-scale geometric analysis.

Mean filtering algorithm is one of the most typical linear filtering algorithms. In the image space, the gray value of the original image is replaced by the average value of adjacent pixels, so as to smooth the image and achieve the effect of image noise reduction. The mean filtering algorithm is simple to implement and fast to run. The algorithm is widely used in industry, but because only the neighborhood average is used to restore pixel values, it is difficult to distinguish the noisy parts of the image from the edge details, and all pixel values will be smoothed by the neighborhood average. If the neighborhood is large, the filtered image will be more blurred and the edge information will be lost. That's why we often see that the edge of an image will become very blurred after the average filtering and denoising.

However, the non-local mean filtering algorithm has excellent denoising effect, so we first analyze its principle [11]. In fact, every pixel in the image is related, and they are not isolated. Each pixel and its surrounding pixels constitute the geometric structure in the image. Considering the complex spatia interrelationships in images, Pixel-centered window neighborhood, that is, image block, can well reflect the structural characteristics of pixels. The set of a group of image blocks compared by each pixel can be regarded as the complete representation of the image. Generally speaking, the image is self-similar, that is to say, the pixels at different positions in the image usually have strong correlation. Just like texture image, this is a typical example. In a small window randomly taken out of the image, the natural image usually contains various redundant or redundant additional information, and many similar window structures can be found in this image. Enough repetitive structures are also included in the natural image. For example, there are a large number of similar pixels in the flat area of the image, and pixels on the same line or curve boundary also have similar neighborhood patterns. This conclusion is obviously correct for the windows with close spatial distance in the image, and this is the assumption of local regularity. Therefore, if the image slice which can describe the structural features of the image is used to measure the similarity between pixels, it will be more accurate than the measurement of a single pixel, so it can better protect the structural information of our image. Efros and Leung [12] were the first to notice that images have this characteristic. They make use of the similarity between image slices for texture synthesis and fill the small holes in the image. The algorithm looks for pixels similar to those to be processed in a large area of the image. Buades et al. [13] proposed a non-local mean filtering denoising algorithm in 2005. The concept of non-local filtering is introduced for the first time. A denoising method based on transform domain filtering and a local smoothing method based on structural similarity are adopted to remove noise. The main purpose of these methods is to restore the main geometric structure of the image. These methods are all based on the regular assumption of the original image. Because noise and the fine structure and detail information of the image have very similar characteristics at the same time, they can smoothly define the fine structure and detail information of the image. Only then can we effectively protect the structural information of the image and achieve the most advanced and excellent denoising effect at present. Non-local mean filtering image denoising algorithm is an important improvement to the traditional neighborhood filtering method. Firstly, the algorithm breaks through the limitation that neighborhood filtering is only used for local filtering. Self-similarity of images is considered. Because the distance of similar pixels in spatial position is not necessarily very close, the periodic image is one kind, so it is more advantageous to find similar pixels in a larger range. Secondly, similar pixels are defined as pixels with the same neighborhood pattern; compared with the similarity obtained only by using the information of a single pixel. It is more reliable and stable to express the characteristics of pixels by using the information in the window with fixed size around the pixels.

For a given pixel point *i*, an image block is an area with a size of *n***n* centered on *i*. For two image blocks in the neighborhood, the Gaussian weighted Euclidean distance between the pixels is used to calculate the similarity between pixel *i* and the pixel *j*. After the image is blocked, the distance between two pixels becomes smaller and the weight given by pixel point *j* becomes greater during accumulative recovery. The specific calculation of the non-local mean filtering algorithm is as follows:

$$\hat{f}\ (i) = \frac{\sum_{i,j} w(i, j) f(j)}{\sum_{i,j} w(i, j)} \tag{10.1}$$

$$w(i, j) = \exp(-\frac{d(i, j)}{h^2}) \tag{10.2}$$

$$d(i, j) = \left|\left|N(i) - N(j)\right|\right|_{2,\alpha}^2 \tag{10.3}$$

where α is the standard deviation of Gaussian kernel function, and convolution processing of image block with Gaussian kernel can reduce the influence of noise on distance calculation and highlight the role of image block center in pixels. *d* represents the weighted Euclidean distance between two image blocks; *h* is a filter parameter for controlling the smoothness. But this method will make the algorithm too complex, so the search window will usually be reduced to a certain size.

Grayscale image

In the process of image acquisition and transmission, it will be polluted by noise, which is not conducive to image processing and analysis. The median filter operator and Gaussian filter operator are used to eliminate the salt and pepper noise and Gaussian noise of the considered part.

Art region feature extraction

In order to obtain the features of the part and separate the part area from the background area, it is necessary to segment the considered part. The image segmentation adopts the threshold segmentation method. The threshold segmentation is divided into different levels according to the gray range of image pixels, and the part area is extracted by setting the gray threshold.

Image area confirmation

In order to measure parts quickly and efficiently and ensure the measurement accuracy, it is necessary to affine transform the collected part image to the horizontal position. Affine transformation mainly includes translation transformation and rotation transformation. The translation transformation makes all pixels of the image move horizontally and vertically according to the required offset; the rotation transformation makes the image rotate a certain angle around a certain point.

Image edge extraction

In order to ensure the accuracy of part measurement results, it is necessary to obtain the sub-pixel data of part image, that is, the edge of part image. Edge is a very important feature of an image. The pixels around the image whose gray value changes are the points on the image edge, that is, the area with the largest gray value derivative. Edge detection is a process of extracting edge and line features with large spatial gradient of gray value in the image by calculating the derivative of gray value change.

Line fitting

The clutch operator is called, the feature histogram is used, and the feature range of width is used to eliminate the aperture edge, and only the outer contour of the part is retained. If one edge of the part is extracted, the contour of the part needs to be divided into straight line segments.

10.2.3 Results From Machine Vision System

30 pcs blade samples measured by projector are measured by machine vision system again, below are shown the results (Fig. 10.4).

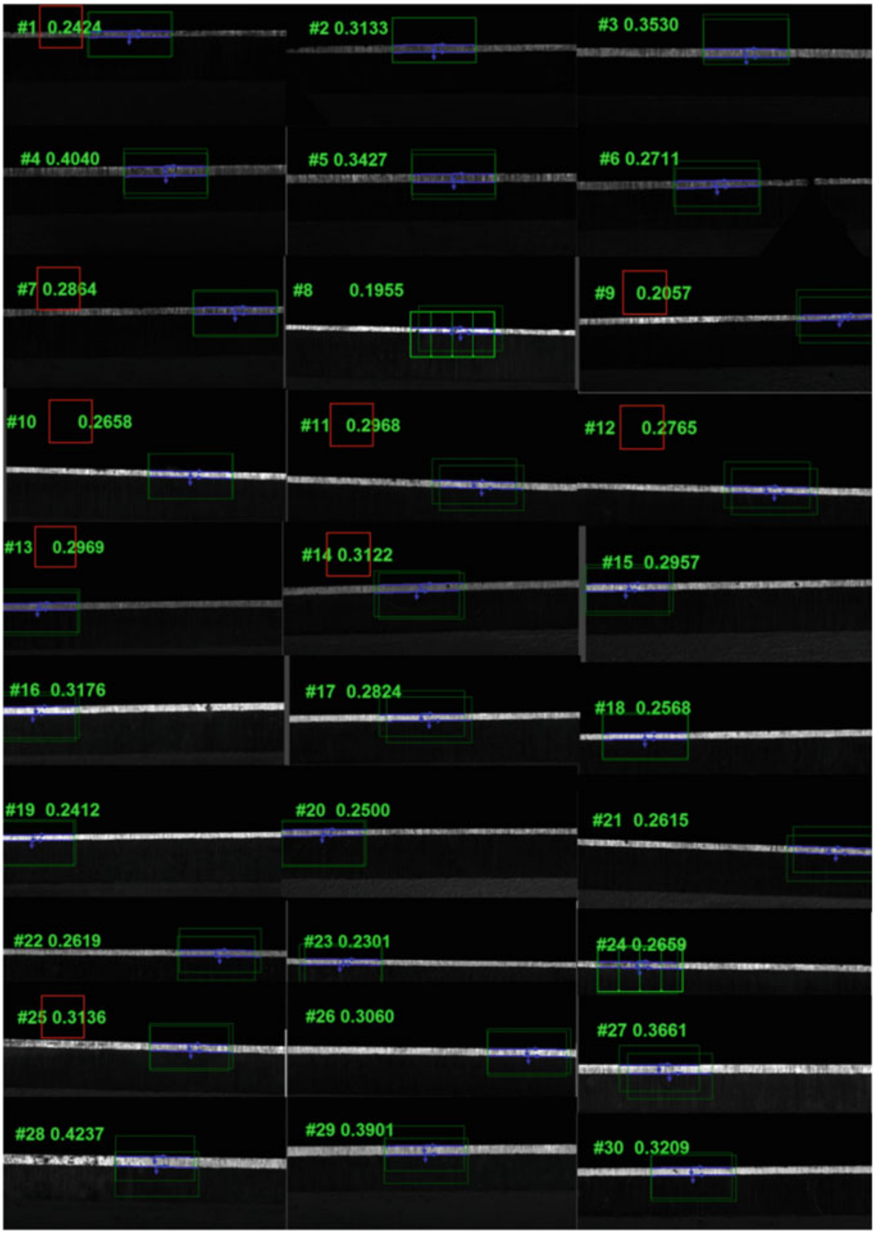


Fig. 10.4 Measuring results graphic

10.2.4 Result Analysis

Below distribution curve and trend graphic, respectively, show those results' position and trend of those 30 pcs blades measured by two different system-projector and machine vision system (Fig. 10.5). The results from machine vision system closely follow the results from the projector, most of deflections are within 0.02 mm, which is acceptable in real manufacturing process.

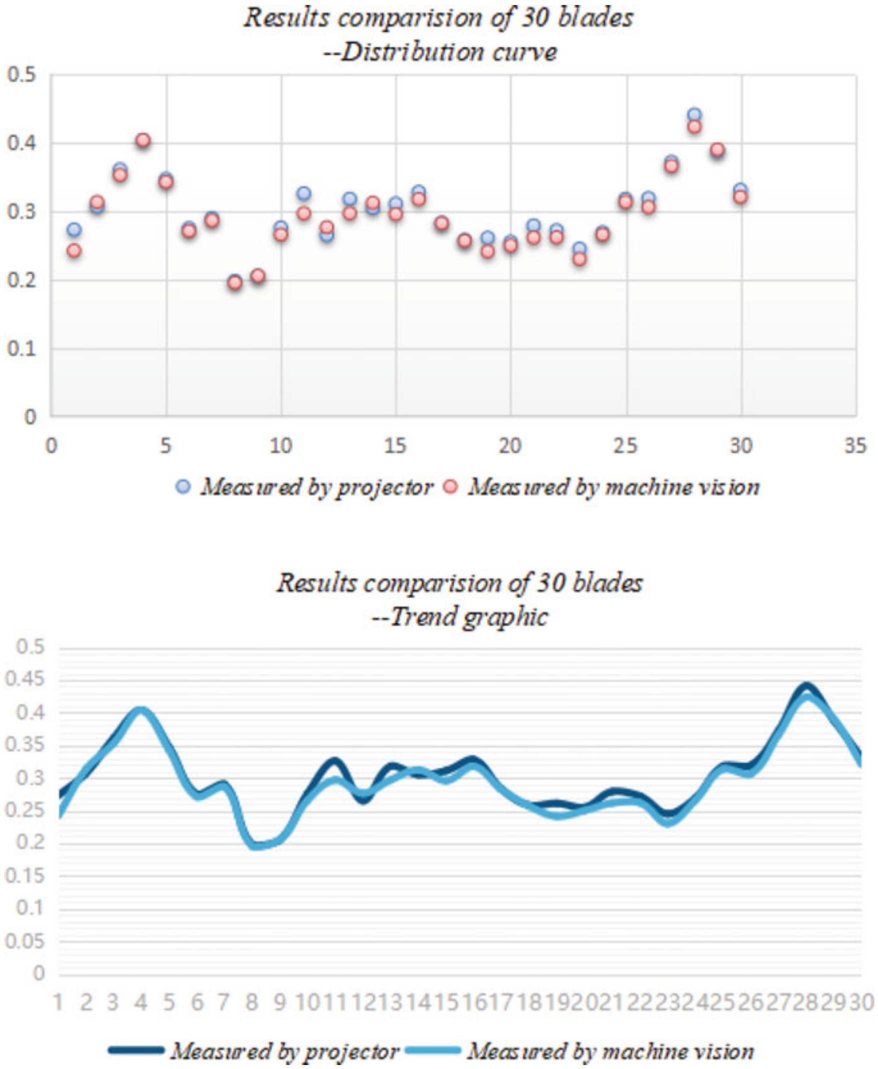


Fig. 10.5 Measuring result analysis

10.3 Conclusions

Based on above experiment's result and analysis, the designed machine vision system is able to measure correctly blade edge width, this system not only measures the blade edge dimension, but also can monitor the quality of grinding surface, which is needed for manufacturer as well. From above analysis graphic (Fig. 10.5), the blade edge width change between 0.2 and 0.4 mm, that means the blade width is not stable in production, manufacturer needs an equipment to monitor real production situation timely, and our research helps to solve this problem. We will continue to optimize algorithms, aiming to reduce equipment costs by improving design, combining with PLC control system, and finally realize an automatic blade grinder.

References

1. Guofu, Z., Hongyan, S.: The application of machine vision technology in industrial testing. Electron. Technol. Softw. Eng. (2013)
2. Chen C.: Design and application of blade angle measurement, modern manufacturing technology and equipment (2018)
3. Jiangping, Q., Min, B.: Tool angle measurement system based on image processing. Electromech. Eng. (2010)
4. Ying, C., Wei, Z., Zhiwei, L., Xuyang, Z., Hui, C.: Linear edge tool angle measurement based on image processing. Tool Technol. (2007)
5. Guozheng, Y.: D. Marr and its theory of visual computing. Foreign Autom. (1984)
6. Guoyang, L.: Size Measurement Techniques Based on Machine Vision. Harbin University of Technology (2014)
7. Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.T.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif. Intell. 87–119 (1995)
8. Nemer Pelliza K.A., Pucheta M.A.: Analysis of the efficiency of the adaptive canny method for the detection of icebergs at open sea. XLII-3/W12-2020, 459–464 (2020)
9. Multimedia: Data on Multimedia Reported by Researchers at Nanchang University: Robust and Imperceptible Watermarking Scheme Based on Canny Edge Detection and SVD in the Contourlet Domain, pp 328–330 (2020)
10. Loderer, M., Beitelschmidt, M.: Improved edge detection based on fractional derivatives for real-time measurement systems. PAMM **19**(1) (2019)
11. Liu, B., Liu, J.: Non-local mean filtering algorithm based on deep learning. In: MATEC Web of Conferences, vol. 232, p. 03025 (2018)
12. Efros, A., Leung, T.: Texture synthesis by nonparametric sampling. In: Proceedings of the IEEE ICCV, pp. 1033–1038 (1999)
13. Pang, C., et al.: A fast NL-Means method in image denoising based on the similarity of spatially sampled pixels. In: Multimedia Signal Processing, 2009 MMSP '09. IEEE International Workshop on IEEE (2009)

Chapter 11
Deep Learning 3D Convolutional Neural Networks for Predicting Alzheimer's Disease (ALD)

Sarah A. Soliman, El-Sayed A. El-Dahshan, and Abdel-Badeeh M. Salem

Abstract Deep learning is a power machine learning algorithm in classification while extracting high-level features. This paper is proposed to predict Alzheimer's disease (AlD) with a deep 3D convolutional neural network (3D-CNN), which can learn generic features capturing AlD biomarkers, classify Alzheimer's brain from normal healthy brain based on Magnetic Resonance Imaging (MRI) scans of the brain. In this work, we successfully classified MRI data of Alzheimer's subjects from normal controls using the Alzheimer's disease Neuroimaging Initiative (ADNI) data set using 3,013 scans where the accuracy for training data reached 96.5% and for test data reached 80.6%. This experiment suggests us the shift- and scale-invariant features extracted by 3D-CNN followed by deep learning classification which is the most powerful method to distinguish clinical data from healthy data in MRI. This approach also enables us to expand our methodology to predict more complicated systems.

11.1 Introduction

11.1.1 Alzheimer's Disease

Alzheimer's disease (AlD) is a neurological, chronic, and progressive brain disease that causes memory loss, reduced cognitive capacity, and inability to execute basic tasks [1]. The annual number of new cases of Alzheimer's disease is expected to double by 2050, according to the statistics [2]. Consequently, it is crucial to diagnose

S. A. Soliman (✉)
Computer Science Department, Higher Technological Institute, Cairo, Egypt

E.-S. A. El-Dahshan
Egyptian E-Learning University, Cairo, Egypt

A.-B. M. Salem
Ain Shams University, Cairo, Egypt
e-mail: absalem@cis.asu.edu.eg

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_11

people with Alzheimer's disease in an early stage. This will be beneficial. This will aid in determining its progression and attempting various clinical treatments in order to improve a patient's quality of life.

Alzheimer's disease causes a significant loss in the hippocampal region of the brain compared to a healthy person, as well as a reduction in the cerebral cortex and an enlargement of the brain's ventricles. Memory, planning, thinking, and judgment are all processes performed by these regions [3]. The amplitude of alterations in distinct parts of the brain is determined by the phases of illness progression. The volume loss of the hippocampus region and cerebral cortex, as well as the enlargement of the ventricles, can be detected by Magnetic Resonance Imaging (MRI).

Differentiating between Mild Cognitive Impairment (MCI), Alzheimer's disease (AD), and Normal control (CN) has become one of the most challenging for doctors. Medical imaging technology, including neuroimaging approaches, can be exploited to solve such challenges. Magnetic Resonance Imaging (MRI) [4], Functional Magnetic Resonance Imaging (fMRI), FDG-Positron Emission Tomography (PET) [5], and Computer Tomography (CT) play an energetic role in determining the anatomical and functional changes in the brain. MRI is the best biomarker to analyze structural changes caused by Alzheimer's disease.

11.1.2 Deep Learning

Deep learning is a pioneering field of machine learning inspired by the human brain. This methodology was developed using complicated algorithms that describe high-level properties and extract those abstractions from data using a similar but more intricate neural network design. The "neocortex," which is a portion of the cerebral cortex that deals with sight and hearing in animals, is identified by neuroscientists to process sensory information by propagating them through a complicated hierarchy over time [6]. That was the driving force behind the development of deep machine learning, which focuses on computational models for information representation that are analogous to those of the neocortex [7–9].

3D Convolutional Neural Networks are specially built on the basis of the explicit assumption that raw data are two-dimensional (images) enabling us to encode those properties and also to decrease the amount of hyperparameters [10]. The 3D-CNN topology uses spatial structures to minimize the number of parameters which must be learned and thus develops upon general feed-forward back-propagation training.

3D-CNNs transform the input data from the input layer through all connected layers into a set of class scores given by the output layer [11]. There are many variations of the 3D-CNN architecture, but they are based on the pattern of layers, as demonstrated in Fig. 11.1. The input layer accepts three-dimensional input generally in the form spatially of the size (width × height) of the image and has a depth representing the color channels [12]. A typical 3D-CNN architecture generally comprises alternate layers of convolution and pooling followed by one or more fully connected layers at the end. In some cases, the fully connected layer is replaced with global

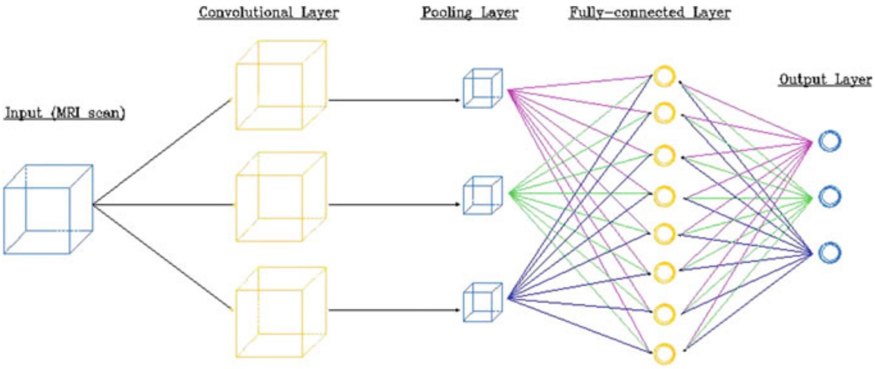


Fig. 11.1 3D convolutional neural network architecture

average pooling layer. In addition to different mapping functions, different regulatory units such as batch normalization and dropout are also incorporated to optimize 3D-CNN performance [13]. The arrangement of 3D-CNN components plays a fundamental role in designing new architectures and thus achieving enhanced performance. The following table will discuss briefly the role of these components in a 3D-CNN architecture (Table 11.1).

Activation function [13, 14] serves as a decision function and helps in learning a complex pattern. Selection of an appropriate activation function can accelerate the learning process. Activation function for a convolved feature map is defined in Eq. 11.1

$$cT_i^k = g_a(cF_i^k) \tag{11.1}$$

In the above equation, cF_i^k is an output of a convolution, which is assigned to activation function g_a that adds non-linearity and returns a transformed output cT_i^k for lth layer.

These layers find a number of features in the images and progressively construct higher-order features. This corresponds directly to the ongoing theme in deep learning by which features are automatically learned as opposed to traditionally hand-engineered. Finally, we have the classification layers in which we have one or more fully connected layers to take the higher-order features and produce class probabilities or scores. These layers are fully connected to all of the neurons in the previous layer, as their name implies. The output of these layers produces typically a two-dimensional output of the dimensions $[b \times N]$, where b is the number of examples in the mini-batch and N is the number of classes we're interested in scoring [13–15].

11.2 Methods

The main purpose of this study is to develop a 3D-CNN model able to discriminate AD from MCI and NC and to build a CAD system and study its performance. In this work, we intended to develop an effective classification system for AD by using the 3D Convolutional Neural Network (3D-CNN).

Classification of Alzheimer's disease images and normal, healthy images required several steps, from preprocessing to recognition (see Fig. 11.2). Three main components formed this recognition pipeline: (a) data acquisition; (b) preprocessing; and (c) classification, respectively. First, the dataset was obtained from ADNI (will discuss in the next section). Two different methods were used in preprocessing phase: reshaping and 3D-scaling. After preprocessing phase, the 3D-CNN-based architecture receiving images in its input layer was trained and tested (validated) using 70% and 30% of the dataset, respectively.

11.2.1 Data Acquisition

We used the structural brain MRI scans from the ADNI dataset [14]. The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see WWW.ADNI-INFO.ORG. A total of 3,013 subjects (955 patients with probable AD, 835 patients with MCI, and 1,223 healthy controls) were considered in this study (Table 11.2). Standard 3 T baseline T1-weighted images were included from the ADNI dataset.

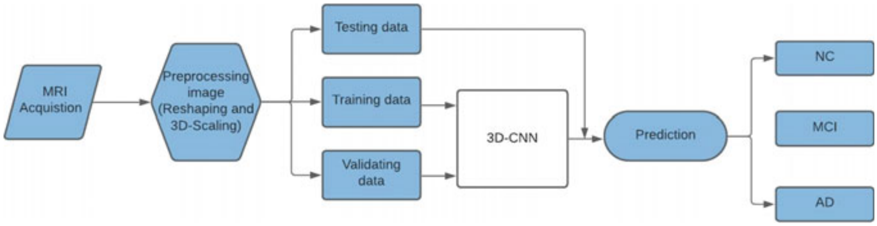


Fig. 11.2 The methodology of the proposed work

Table 11.1 3D-CNN architecture layers

3D-CNN architecture	Description
Convolution layer	Convolutional layers are considered the core building blocks of CNN architectures. Convolutional layers transform the input data by using a patch of locally connecting neurons from the previous layer. The layer will compute a dot product between the region of the neurons in the input layer and the weights to which they are locally connected in the output layer [14]
Pooling layer	Pooling layers [13, 14] provide an approach to downsampling feature maps by summarizing the presence of features on patches of the feature map. Pooling layers reduce the dimensions of the data by combing the outputs of neurons clusters at one layer into a single neuron in the next layer. The pooling layer uses the max operation to resize the input data spatially (width, height). This operation is referred to as *max pooling*
Flatten layer	Flattening involves transforming the entire pooled feature map matrix into a single column which is then fed to the neural network for processing
Fully connected layer	Is mostly used at the end of the network for classification purpose. Unlike pooling and convolution, it is a global operation. It takes input from feature extraction stages and globally analyses output of all the preceding layers [13]

Table 11.2 Demographic clinical features of AD and MCI patients and healthy controls from the ADNI dataset

Modality	Total subj	Group	Subj	Female	Mean of age	SD	Male	Mean of age	SD
MRI	3,013	HC	1,223	615	77.866	4.4	608	78.022	5.5
		MCI	835	381	75.246	7.5	454	76.807	7.8
		AD	955	441	74.860	7.8	514	76.759	7.3

Abbreviations: AD = Alzheimer's disease, HC = healthy controls, MCI = Mild Cognitive Impairment, SD = Standard Deviation

11.2.2 Image Preprocessing

The preprocessed MRI data included were loaded into memory using a similar approach to any other format, but the data was used as a Numpy Array, hence the usage of the library Numpy and OpenCV for the manipulation of the used images such as resizing and reshaping, and then the data is loaded to the variables and split into train and test and validation for each of the images and the labels, image size is 96 × 96 each of 62 channels as the third dimension of the image of total 2,109 subjects for training (130,758 images) and 435 for validation (26,970 images) and 469 for testing (29,078 images) these are assigned to the appropriate variables in the memory, then fed to the training process of our model, with number of epoch = 55, base learning rate = 0.01 and batch size of 16 per batch, the training process was done on Google Collaboratory servers to insure that the hardware is sufficient for the training process; 70% of the data was assigned for the training and validation processes while 30% is for the testing process and the results for 25 epoch gave us 91% of training accuracy, 71% of validation accuracy, and 70.70% of total accuracy (testing accuracy), but after increasing the epoch number to the 55 epochs, training accuracy reached to 96.6%, 79% for validation, and 80.60% of total accuracy (testing accuracy).

11.2.3 Classification Using 3D-CNN

Our proposed model is a deep convolutional neural network that was written and evaluated in Python 3 using both tensor flow and Keras [16] packages as shown in Fig. 11.3. The model has several layers performing four basic operations: convolution, max pooling, flatten, and dense. The layers in the model follow a particular connection pattern known as dense connectivity, where each layer is connected to every other layer. For final classification, there is a softmax layer with three different output classes: Normal Control (CN), Mild Cognitive Impairment (MCI), and AD.

The input layer size is (96 × 96 × 62 × 1) saved from the data processing step. First, a (3 × 3 × 3) 3D-convolution layer with ReLU activation function was used to create 162 feature maps. This was followed by (2 × 2 × 2) maximum pooling. This process was repeated with two (3 × 3 × 3) 3D-convolution layers with kernel size 128, 256 consequentially and the ReLU activation function and another max pooling layer. After that, there are two (3 × 3 × 3) 3D-convolutional layers with kernel size 324, (2 × 2 × 2) max pooling, and (3 × 3 × 3) 3D-convolutional layer with 512 kernel size and (2 × 2 × 2) max pooling. The last 162 feature maps were attended and were fully connected to neural nodes. This was tracked by flatten layer which is responsible for transforming the data into a one-dimensional array for inputting it to the dense layer to feed the output of the pervious layer to all its neurons. Finally, a layer with softmax activation was employed to yield probabilities for each zone. For this CNN, the maximum probability in this vector was used to determine the class of the image (See Table 11.3). We trained the proposed model on 62 images for each patient.

The 3D-CNN was trained using 70% of the collected data and was tested using 30% of the data. This translated into using data from seven tags in each zone for training, and three tags in each zone for testing. The loss function was set to minimize the categorical cross-entropy using the Adadelta optimizer [17]. The training was performed with a batch size of 25 and 55 epochs.

11.2.4 Performance Evaluation

In this work, we consider the following performance measures: Precision, Recall, F1 score, and Accuracy to measure the performance. True positives (TP), true negatives

Fig. 11.3 Block diagram of proposed Alzheimer's disease diagnosis framework

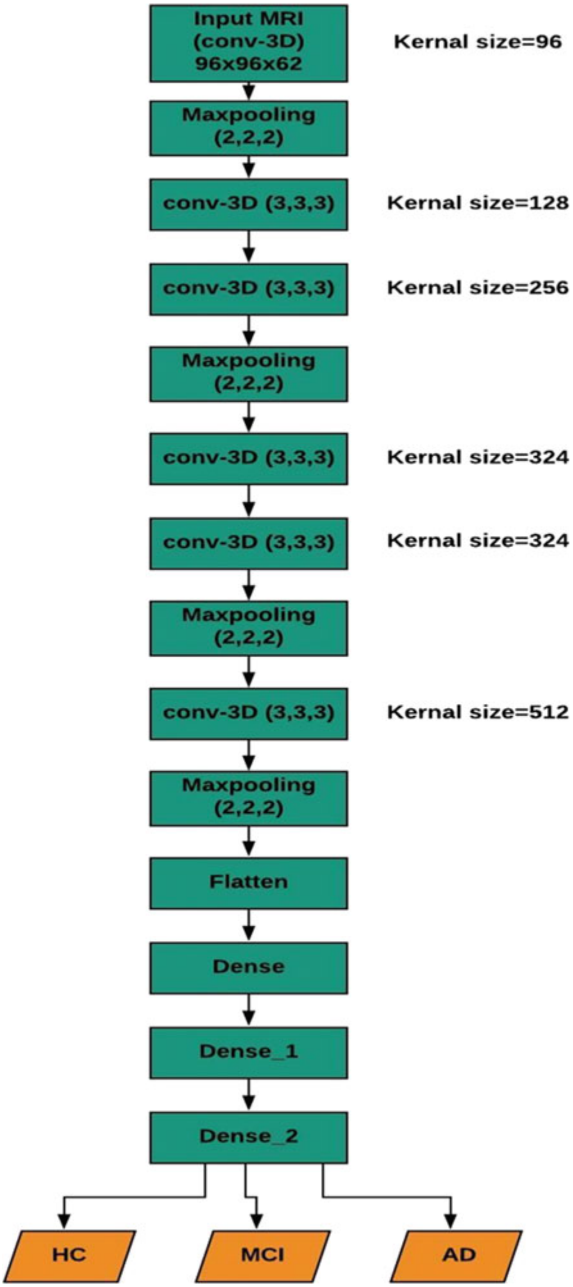


Table 11.3 Detailed parameters of the designed model

Layer (type)	Output shape	Parameters (sum of weights and biases)
conv3d (conv3D)	(None,94,94,60,96)	2688
max_pooling3d (MaxPooling3D)	(None,47,47,30,96)	0
conv3d_1 (conv3D)	(None,45,45,28,128)	331,904
conv3d _2 (conv3D)	(None,43,43,26,256)	884,992
max_pooling3d _1 (MaxPooling3D)	(None,21,21,13,256)	0
conv3d_3 (conv3D)	(None,19,19,11,324)	2,239,812
conv3d _4 (conv3D)	(None,17,17,0,324)	2,834,676
max_pooling3d _2 (MaxPooling3D)	(None,8,8,4,324)	0
conv3d _5 (conv3D)	(None,6,6,2,512)	4,479,488
max_pooling3d _3 (MaxPooling3D)	(None,3,3,1,512)	0
flatten (Flatten)	(None,4608)	0
dense (Dense)	(None,1000)	4,609,000
Dense_1 (Dense)	(None,1000)	1,001,000
dense _2 (Dense)	(None,3)	3003

(TN), false positive (FP), and false negative (FN) are used to calculate accuracy and F1 score. The calculations are as follows:

- Precision also called positive predictive value (PPV), or probability of correct positive prediction. It is given by the following formula [17]

$$\text{precision} = \frac{TP}{TP + FP} \tag{11.2}$$

- Recall also called sensitivity (SN) which refers to the ability of identifying the AD patients. It is given by the following formula [17]

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11.3}$$

- F1 score is a measure of test's accuracy that considers both the precision and recall of the test to compute the score. It is given by the formula [17]

$$F_{1\text{score}} = \frac{{}_2TP}{{}_2TP + FP + FN} \tag{11.4}$$

- Accuracy (ACC), which is the probability of both correct positive and negative predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11.5}$$

where the parameters TP, FP, TN, and FN are defined as follows:

– True positive (TP): the patient has the AD and the classification result is positive (AD).
– False positive (FP): the patient is normal and the classification result is positive.
– True negative (TN): the patient is normal and the classification result is negative (Normal).
– False negative (FN): the patient has the AD but the test is negative.

11.3 Experimental Results

Performance of the 3D-CNN was validated and tested on patients and controls, with three classifications: HC, MCI, and AD. For each classification, the CNN was evaluated on ADNI dataset. Each classification included three steps as shown in Fig. 11.4: (i) training, (ii) validation, and (iii) testing. We describe the performance of our model in confusion matrixes. Confusion matrix is shown in Fig. 11.4. In confusion matrix, each column represents the instances in a predicted class, and each row represents the instances in the actual class. Values on the matrix diagonal indicate correct prediction, and the values outside the matrix diagonal show incorrect prediction. A summary of the experimental results is given in Table 11.4. We obtained an average accuracy of 80%. A classification accuracy of 92% was achieved for the CN, 80% was achieved for MCI, and AD achieved an accuracy of 69%.To improve the performance of our classifier, we used adadelta optimizer in order to adapt learning rates based on a moving window of gradient updates, instead of accumulating all past gradients [18].

Fig. 11.4 Confusion matrix based on the proposed model

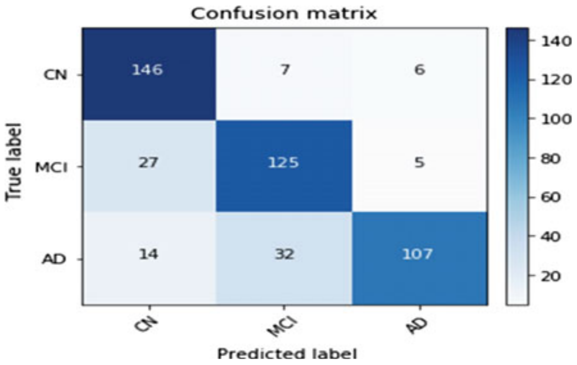


Table 11.4 Performance of the proposed model on ADNI dataset

Precision	Recall	F1-score	Accuracy	Support
0.78	0.92	0.84	0.92	159
0.76	0.80	0.78	0.80	157
0.91	0.70	0.79	0.69	153

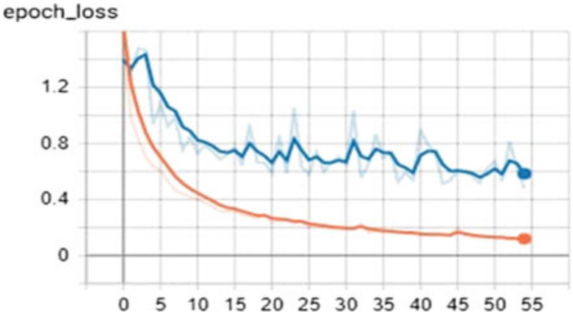


Fig. 11.5 The categorical cross-entropy loss is shown for the 3D-convolutional neural network training as a function of the epoch number

The 3D-CNN training process took approximately 666 s per epoch for a total of nine hours on Google collapse (25.5 Gigabytes RAM, GPU is NVIDIA tesla k80, and 2VCPU @ 2.2GHZ). The training and validation losses are shown in Fig. 11.5.

First, the network was trained using 25 epochs and the training accuracy reached to 91%. After that, the network was trained with 55 epochs and the accuracy achieved 97.8%. It is clear that overfitting is not an issue for the CNN, since both the training and validation losses follow the same decreasing curve. This is further supported by the fact that the testing set had the exact same loss when it was evaluated after the CNN was trained. The network only took 666 s to train on Google collapse, and yielded extraordinary results even after epoch-24 93% accuracy. Figure 11.6 shows the accuracy of some samples of each epoch.

11.4 Conclusion and Future Work

We presented deep learning-based classifiers of MRI data to distinguish brains affected by Alzheimer's disease from normal healthy brains in older adults. Convolutional neural network architecture was used to extract scale- and shift-invariant low- to high-level characteristics from a large amount of whole-brain data, resulting in a highly accurate and repeatable predictive model. Our proposed network can be very positive for diagnosing Alzheimer's disease in early-stage. Though the proposed model has been tested only on Alzheimer's disease dataset, we believe it can be used successfully for other classification problems of medical domain. In future, we aim

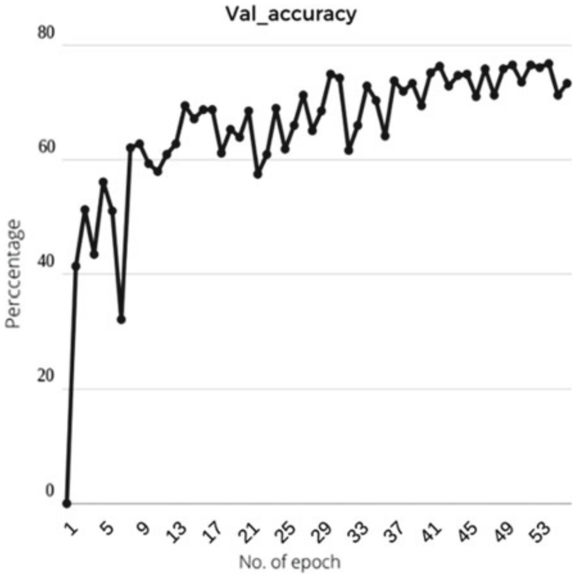


Fig. 11.6 Accuracy of some samples of epochs

to build CNN model based on multimodal data and use segmentation algorithms to reach better results.

References

1. Gupta, A., Ayhan, M., Maida, A.: Natural image bases to represent neuroimaging data. Int. Conf. Mach. Learn. **28**, 987–994 (2013)
2. Sarraf, S., Tofighi, G., DeSouza, D., Anderson, J.: DeepAD: Alzheimer’s disease classification via deep convolutional neural networks using MRI and fMRI. bioRxiv (2016)
3. Farooq, A., Anwar, S., Awais, M., Rehman, S.: A deep CNN based multi-class classification of Alzheimer’s disease using MRI. In: IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1–6 (2017)
4. Frisoni, G., Fox, N., Jack, C., et al.: The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol **6**, 67–77 (2010)
5. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., Umer, A.: Classification of Alzheimer’s disease using machine learning techniques. In: 8th International Conference on Data Science, Technology and Applications (2019)
6. Pallas, S.L., Mao, Y.-T.: The evolution of multisensory neocortex. In: New Handbook of Multisensory Processes, pp. 627–642 (2012)
7. Arel, I., Rose, C.D., Karnowski, T.P.: Deep machine learning-a new frontier in artificial intelligence research [research frontier]. Comput. Intell. Mag. **5**(4), 13–18 (2010)
8. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S.D., Caffe Caffe, T.: Convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia, pp. 675–678. arXiv: 1408.5093v1 (2014)
9. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)

10. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-Gonzalez, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O.: Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. arXiv:1904.07773 (2019)
11. Jain, R., Jain, N., Aggarwal, A., Hemanth, D.J.: Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn. Syst. Res. **57**, 147–159 (2019)
12. Karasawa, H., Liu, C.-L., Ohwada, H.: Deep 3D convolutional neural network architectures for Alzheimer's disease diagnosis. In: Asian Conference on Intelligent Information and Database Systems, pp. 287–296. Springer Link (2018)
13. Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S.: A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. (2019). https://doi.org/10.1007/s10462-020-09825-6
14. Litjens, G., Kooi, T., Bejnordi, E.B., Arindra, A., Setio, A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
15. Phung, V.H., Rhee, E.J.: A deep learning approach for classification of cloud image patches on small datasets. J. Inf. Commun. Converg. Eng. **16**, 173–178 (2018)
16. Chauhan, K., Ram, S.: Image classification with deep learning and comparison between different convolutional neural network structures using tensorflow and keras. Int. J. Adv. Eng. Res. Develop. **5**(2), 533–538 (2018)
17. Rajeswari, K.K.R., Maheshappa, H.D.: Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. Inform. Med. Elsiever **14**, 34–42 (2019)
18. Diederik, P., Kingma, J.B.: Adam: a method for stochastic optimization. In: 3rd International Conference for Learning Representations, vol. 1, p. 15. arXiv: 1412.6980v9 (2015)

Chapter 12
Dense Optical Flow and Residual Network-Based Human Activity Recognition

Negar Sultana, Danastan Tasaouf Mridula, Zahid Sheikh, Fariha Iffath, and Md Shopon

Abstract Nowadays Human activity recognition takes a fascinating part in miscellaneous fields of computer vision like medical care, video surveillance, human-computer interface. Recently optical flow were shown to be an efficient feature for action recognition and attained state-of-the-art accuracy on different datasets. In this work, we have taken a deeper look at the combination of action recognition and optical flow. This paper proposes a novel human activity recognition technique based on the 3D dense optical flow from video sequences. Transfer learning and 3D dense optical flow was used in a two-stream neural network architecture. For transfer learning, ResNet152 pre-trained architecture was used. ResNet152 extracted several fine grained features from the dense optical flow. The proposed method was tested with UCF-101 dataset and outperformed the existing state-of-the art methods in terms of accuracy.

12.1 Introduction

Human activity recognition aims to predict the actions of humans. Human activity recognition is required to detect a set of human activities by training a supervised

N. Sultana (✉) · D. T. Mridula · Z. Sheikh · M. Shopon
Department of Computer Science and Engineering, University of Asia Pacific, Dhaka, Bangladesh
e-mail: 17101054@uap-bd.edu

D. T. Mridula
e-mail: 17101075@uap-bd.edu

Z. Sheikh
e-mail: 17101096@uap-bd.edu

M. Shopon
e-mail: shopon@uap-bd.edu

F. Iffath
Department of Computer Science and Engineering, BGC Trust University Bangladesh,
Chandanaish, Bangladesh
e-mail: fariha@bgctub.ac.bd

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_12

model and displaying the activity result as per the input activity received from the camera input.

Human activity recognition is a broad field of study in pattern recognition and computer vision. It plays a significant part in pattern recognition and computer vision applications like robotics, human-computer interaction, human-to-human interaction, monitoring, etc. Movements are often typical activities such as walking, talking, standing, eating, and sitting. It was a challenging problem earlier because there was no feasible approach to recognize human activity. Now deep learning methods have been deployed in human activity recognition problems with their ability to learn higher-order features automatically.

Recently, there have been developed immense approaches for various problems, including image analysis, artificial intelligence. The main object of human activity recognition is to explore ongoing activities automatically from video sequences or still images. Usually, from an input video, the human action recognition systems aim to classify activity categories correctly. Human activities are categorized into: (i) gestures; (ii) atomic actions; (iii) human-to-object or human-to-human interactions; (iv) group actions; (v) behaviors; and (vi) events, depending on activity complexity. In pattern recognition and computer vision fields, the feature extraction procedure is a significant step. It allows the image and video contents to be more particular space than the direct use of pixels. Most of the previous works in human activity recognition have focused on using videos [1, 2].

In deep learning, complicated features and concept extraction are getting faster and easier with high accuracy and in a short time using their approaches. Convolutional Neural Networks (CNNs) and Optical flow methods have recently shown massive image/video classification success. In this paper, we proposed a new model for the Human Activity Recognition (HAR) implementation by extracting frames using the dense optical flow and a pre-trained CNN called ResNet152 for training our dataset. Our implementation results have shown that the proposed method we have used for the HAR implementation is very high. The goal of our work is to find a model with the best performance for the HAR using these approaches.

This paper is presented as follows. Section 12.1 describes the related work done in this area. Section 12.2 explains the overview of our methodology part and illustrates the classification methods used to perform the activity recognition. Section 12.3 represents the experimentation and results, and finally Sect. 12.4 concludes our paper.

12.1.1 Related Work

In this section, we have discussed and classified the Human Activity Recognition (HAR) related works in two subsections.

12.1.2 Optical Flow-Based HAR

In computer vision, the optical flow has created significant progress by improving accuracy on standard benchmarks. Optical flow shows as input for video segmentation, tracking, depth estimation, and other problems. Cutler and Turk presented a better approach in paper [3] using dense optical flow, and into motion bubbles, they classified those flows. This approach tends to be very time-consuming and costly. Despite being time-consuming and expensive, this method provides a high accuracy [4, 5]. Wang and Schmid in paper [6] proposed a dense sampling method by using a dense trajectory. This technique worked on different features with a particular structure. This method is not optimized but gives high accuracy. Jongwoo Lim and Bohyung Han proposed a method in [7] that maintains two models. Appearance models and motion models are used to define the pixels' belonging to moving objects or static backgrounds. The superpixel region of a given image is used for both model extraction and computation. Based on the color histogram from the corresponding superpixel, the appearance model is defined. On the other hand, based on the dense optical information, the motion models are extracted. Despite its high accuracy, this method is time-consuming because manual labeling is required for the moving object in the first frame. Due to this reason, until the initial label is provided, a newly appeared moving object cannot be detected that makes this method unsuitable for real application.

12.1.3 Deep Learning-Based HAR

In deep learning, more complicated concepts can be extracted from videos by considering frame sequence. Action recognition methods can be grouped in deep learning in two categories: two-stream and space-time networks. Two-stream networks [8] are the most existing deep learning HAR methods. In two-stream, one stream for spatial information and inter-frame motion information for the other one. In the spatial stream, raw frames are fed, and in the temporal stream, optical flow images are provided. The motion information is processed separately from the visual information that is one of the drawbacks of two-stream approaches. Jay et al. in paper [9] introduced 3D CNN of space-time networks. There are various space-time networks such as Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTM) that have been used in the video [10, 11]. Despite high processing volumes, both 3D CNN and recursive methods need a large volume of dataset for training such networks. For such networks, the video dataset is costly and onerous.

These methods require a large training dataset and can not provide optimal temporal information for processing. A strategy can be needed that includes optimal information into the processing and requires less training dataset. The proposed method presented in this paper require less training data than Spatio-temporal net-

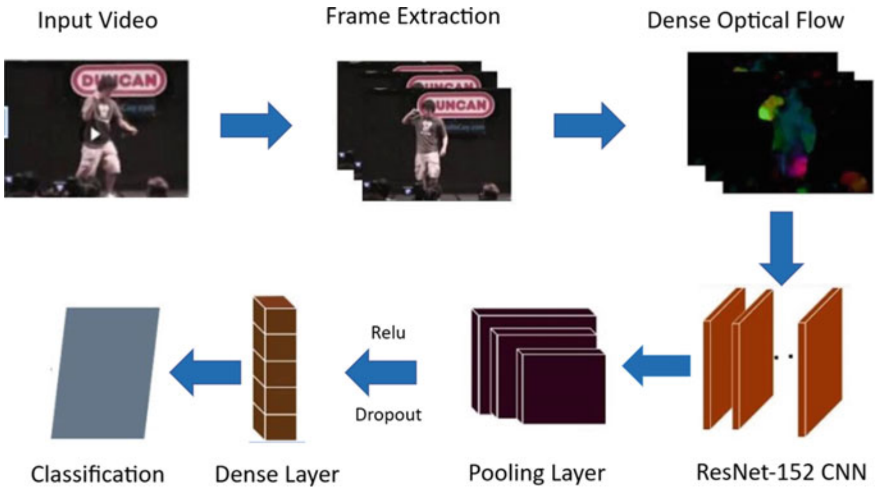


Fig. 12.1 Flow chart of the proposed method

works because of ResNet152. The dense optical flow provides the optimal information about the images that helps to achieve high accuracy in this method.

12.2 Methodology

In this section, we are going to discuss the proposed method for the HAR implementation. Figure 12.1 depicts a basic idea about the proposed method. In this work, we have used dense optical flow method and residual network architecture for human activity recognition. The methodology part is further divided into five main parts: data pre-processing, optical flow, transfer learning, hyperparameters, and implementation.

12.2.1 Data Pre-processing

Data pre-processing is the most fundamental step before using it for further usage. First, the video frames are extracted from the videos. After extracting the video frames from the dataset, we have converted the video frames into grayscale channels. We have extracted the dense optical flow frame from these grayscale frames, and this extraction part is described in the optical flow section. After that, the dense optical flow frames are resized into 224 × 224 pixels. As the action labels of the UCF-101 are non-numerical, we have used one hot-encoding transform the labels into categorical values.

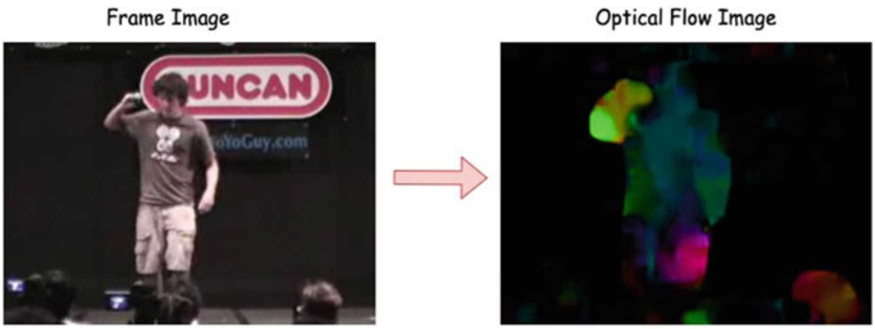


Fig. 12.2 Optical flow image extraction

12.2.2 Optical Flow

In computer vision research, machines are allowed to perform pixel-wise classification for object detection. For the real-time video input, we need to track the motion of objects according to their frames to predict their position and estimate their current velocity in the next frame. The main idea of the optical flow is per-pixel prediction and the pixel brightness that is moved over time across the screen. In many research areas such as image processing, object segmentation, control of navigation, including motion detection, luminance, time to-contact information, motion-compensated encoding, and stereo disparity measurement [12, 13] optical flow is employed. For video classification, it has been used significantly and improves the accuracy of classification. There are two main methods for optical flow implementation, which are sparse and dense. Sparse optical flow attempts to compute the flow vectors of some features, such as a few pixels of an object within the frame. On the other hand, the dense optical flow computes the flow vectors of the entire frame of every pixel. Because of this reason, the dense optical flow performs so well for image classification and demonstrates higher accuracy though it is computationally expensive compared to the sparse optical flow. In this paper, we have chosen the Gunnar Farneback method for computing the dense optical flow.

The dense optical flow computes the optical flow vector for every pixel of each frame. The dense optical flow is intended to find the correspondence between two video frames by temporal variation of the pixels in the sequence of frames. Farneback et al. [14] proposed a two-frame motion estimation method based on polynomial expansion. The Farneback method is a beneficial method to estimate the motion of objects. So using this method, we have calculated the dense optical flow and this optical flow helped us to implement HAR. For the implementation of the dense optical flow here, we have used our pre-processed data frames. We used the Gunnar Farneback proposed method [14] for calculating the dense optical flow. For the visualization, the magnitude and direction of the optical flow from a 2D array of flow vectors are computed. The angle (direction) of flow is visualized by hue. By using the HSV color representation value, the distance (magnitude) of flow is visualized.

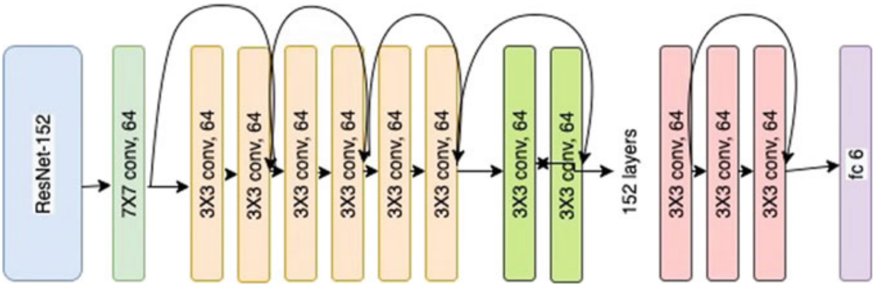


Fig. 12.3 ResNet152 diagram

After that, we have normalized the HSV by maximizing the strength up to 255. Then, we converted the HSV image frame to the RGB frame. This RGB frame is ultimately a dense optical flow for each frame. Figure 12.2 shows the optical flow image frame that we have extracted.

12.2.3 *Transfer Learning*

We will train our optical flow images on Convolutional Neural Networks (CNNs), identifying objects from the optical flow images. We are using the UCF-101 dataset, which has images of 101 categories. To get high accuracy, we need to build up the best neural network model. For that, instead of training a CNN from scratch, we have used a pre-trained and pre-built model. Transfer learning is a pre-trained and pre-build model that can be used to different but related models. In our work, we used ResNet152 transfer learning for our model. In deep learning, ResNet152 is a pre-trained model for image classification of the ImageNet [16] dataset. The ResNet152 model has 152 deep layers, where 150 layers are convolution layers, a pooling layer, and a fully connected layer. It is trained on 15 million images over 1000 plus categories (Fig. 12.3).

It has a variation on deep architecture networks. The learnable parameter numbers are increased with the increasing depth number. For the rising computational overhead, generally, the speed of training and the learning network is reduced. So to avoid this problem, ResNet architecture is using a technique which is called a bottleneck. In ResNet architecture, the input size 224*224 is compressed and reduced to 56*56 after two layers. In each layer, Relu and normalization are used. For using the Relu activation function and normalization of batch size in each layer, the learning time is getting faster, weights are updated, and computational overhead is getting lower. For improving the performance and stability of neural networks, batch normalization is an essential technique. By using the above steps in each layer, the error rate is reducing very fast.

12.2.4 Hyperparameter Tuning

Choosing proper hyperparamters is one of the key things to achieve better result in deep learning architecture. Before passing the dataset into architectures, hyperparameters are needed to be tuned in order to correctly differentiate and identify the human actions in the videos. In this work, we have used categorical cross entropy loss function which is denoted by Eq. 12.1

$$\mathrm{Loss} = -\sum_{\mathrm{i=1}}^{\mathrm{outputsize}} \mathrm{y_i} \cdot \log \widehat{\mathrm{y_i}} \tag{12.1}$$

In Eq. (12.1), $\widehat{y_i}$ is the target value where y_i is the ith scalar value in the model output and output size is the number of scalar values in the model output.

In order to chose the best optimizer for adjusting weights during backpropagation, we experimented with Adam, RMSProp, and SGD optimizer. In our proposed method Adam optimization function has shown better performance and accuracy compared to other optimization functions. The result can be seen in Table 12.1.

12.2.5 Implementation

In our implementation, we have divided our frames and level into training (75%) and testing (25%). Data augmentation is used on our training data. For the training data augmentation we have used rotation range, zoom range width shift range, height shift range, shear range, horizontal flip, and fill mode strategies. Using these techniques, our proposed model has enabled us to learn differentiated and robust data and be able to reach the high accuracy. After using data augmentation, we have passed the training and testing values to the ResNet152 model. The ResNet152 model is a pre-trained model on ImageNet dataset [16]. GeForce GTX 1080 Ti GPU and 8 GB of GPU RAM is used for the image classification in this paper. We have used AveragePooling, ReLU-Layer (for thresholding) Flattening and passed it to the Convolution Neural Network. We have attempted with thick 512 dense layers with "relu" initiation with 0.5 dropouts. Considering the overall memory availability 64 batch size parameter is used for training. Finally, at last we have used dense layers with "softmax" activation to predict the action classes.

12.3 Experiments and Results

This section presents the details about the dataset that the proposed method was evaluated in.

Table 12.1 Shows the accuracy of the proposed method that is experimented with different optimizers

Optimizer	Accuracy (%)
Adam	94.31
RMSprop	90
SGD	85

12.3.1 Dataset

In this work, we have used the UCF-101 dataset for the HAR implementation. The UCF-101 is the most significant action recognition dataset which provides realistic action videos. All of these videos are taken from YouTube. UCF-101 has 101 different action classes and it consists of 13320 videos. The action categories are found in five kinds such as body-motion, human-object interaction, sports, human-human interaction, and playing musical instruments. The number of videos in each classes lies between the range of 100–200. Each image frame is 320*240 dimensions, and the shortest video of this dataset contains 28 frames.

12.3.2 Results and Discussion

In this subsection, we will demonstrate the results achieved using the proposed method. Figure 12.4 depicts the HAR implementation accuracy and loss of the training and testing steps. As we can see, the accuracy increases by the increasing number of implementation steps, and the loss decreases at that same time. By observing the figure, we can say that the method is showing no sign of overfitting.

To show the superiority of our chosen optimizer, we have compared the result achieved using the proposed method. Table 12.1 shows the results of the accuracy that we have attained during our experiment. We have used three different optimizers, and they are Adam [15], SGD [16], and RMSprop [17]. Using Adam optimizer, we have achieved the highest accuracy which is 94.31% whereas 85% accuracy was attained when SGD optimizer was used, and 90% accuracy was acquired when we used RMSprop optimizer.

As the given results in Table 12.2, the accuracy of our proposed method is higher compared to the other methods. When compared to the two-stream methods [8], the accuracy is about 6% higher. The accuracy is also higher than the C3D(3net) [18], EMV+RGB-CNN [19], and DBLSTM+CNN [21] methods. This accuracy achieved within a smaller number of iterations. ResNet152 pre-trained model extracted fine grained features which produced better results.

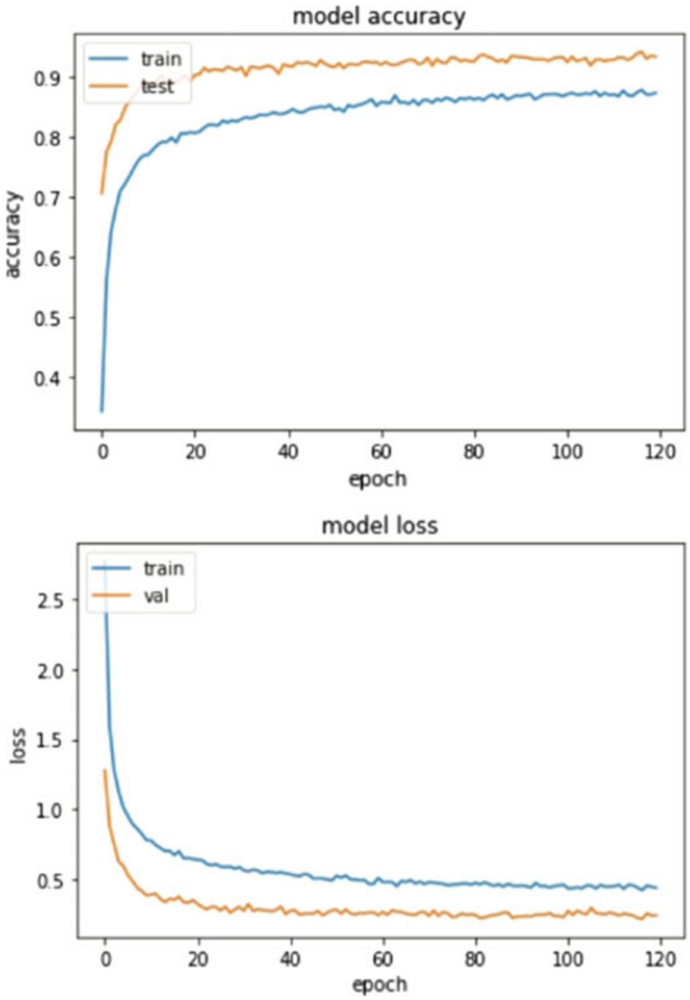


Fig. 12.4 Model accuracy and loss diagram in both training and testing steps of the proposed model

Table 12.2 Comparison of human action recognition score based on the UCF-101 dataset of our proposed method with previous some methods

Method	UCF-101 Accuracy (%)
C3D(3net) [18]	85.2
EMV+RGB-CNN [19]	86.4
Two-stream CNNs [8]	88.0
RLSTM-g3 [20]	86.9
DBLSTM+CNN [21]	91
Our method (dense optical flow + ResNet)	94.31

12.4 Conclusion

This paper proposed a model using the dense optical flow and ResNet152 for the HAR implementation. Our proposed method investigated the ability of the dense optical flow method to extract features from videos and achieved the 94.31% validation accuracy on the UCF-101 dataset. The proposed method outperforms previous state-of-the art methods. ResNet152 extracted fine grained features from the video frames and due to that the proposed model learned the pattern of the data more smoothly. Our future work includes incorporating Graph Neural Networks (GNNs) for better utilizing the kinematic relationship between different body joints for activity recognition.

References

1. Caetano, C., dos Santos, J.A., Schwartz, W.R.: Optical flow co-occurrence matrices: A novel spatiotemporal feature descriptor. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 1947–1952. IEEE (2016)
2. Eleyan, A., Demirel, H.: Co-occurrence matrix and its statistical features as a new approach for face recognition. Turk. J. Electr. Eng. Comput. Sci. **19**(1), 97–107 (2011)
3. Cutler, R., Turk, M.: View-based interpretation of real-time optical flow for gesture recognition. In: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 416–421. IEEE (1998)
4. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: European Conference on Computer Vision, pp. 25–36. Springer (2004)
5. Kantorov, V., Laptev, I.: Efficient feature extraction, encoding and classification for action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2593–2600 (2014)
6. Wang, H., Schmid, C.: Action recognition with improved trajectories. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3551–3558 (2013)
7. Lim, J., Han, B.: Generalized background subtraction using superpixels with label integrated motion estimation. In: European Conference on Computer Vision, pp. 173–187. Springer (2014)
8. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. arXiv preprint arXiv:1406.2199 (2014)
9. Ji, S., Xu, W., Yang, M., Yu, K.: 3d convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 221–231 (2012)
10. Du, Y., Wang, W., Wang, L.: Hierarchical recurrent neural network for skeleton based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1110–1118 (2015)
11. Srivastava, N., Mansimov, E., Salakhudinov, R.: Unsupervised learning of video representations using lstms. In: International Conference on Machine Learning, pp. 843–852. PMLR (2015)
12. Aires, K.R., Santana, A.M., Medeiros, A.A.: Optical flow using color information: preliminary results. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1607–1611 (2008)
13. Beauchemin, S.S., Barron, J.L.: The computation of optical flow. ACM Comput. Surv. (CSUR) **27**(3), 433–466 (1995)
14. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Scandinavian Conference on Image Analysis, pp. 363–370. Springer (2003)

15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv:1412.6980 (2014)
16. Bottou, L., Bousquet, O.: 13 the tradeoffs of large-scale learning. Optim. Mach. Learn. 351 (2011)
17. Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Netw. Mach. Learn. **4**(2), 26–31 (2012)
18. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4489–4497 (2015)
19. Zhang, B., Wang, L., Wang, Z., Qiao, Y., Wang, H.: Real-time action recognition with enhanced motion vector cnns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2718–2726 (2016)
20. Mahasseni, B., Todorovic, S.: Regularizing long short term memory with 3d human-skeleton sequences for action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3054–3062 (2016)
21. Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., Baik, S.W.: Action recognition in video sequences using deep bi-directional ISTM with CNN features. IEEE Access **6**, 1155–1166 (2017)

Chapter 13
Density Calculation of Pseudo Breast MRI Based on Adversarial Generative Network

Yuanzhen Liu, Wei Lin, and Yujia Cheng

Abstract In this paper is proposed a method for mammographic percentage density (PD%) calculation from reconstructed pseudo MRI from real breast MRI. Firstly, the mammography and real breast MRI data were collected from the same women in one year. Then, a direct mapping model was constructed from mammographic to another breast MRI by Gan, and we called the generated MRI pseudo breast MRI. Secondly, a U-Net was used to segment the ROI on the pseudo breast MRI, so that the PD% can be obtained. Finally, DSC was used to evaluate the mapping model and the U-Net, and linear regression and Pearson correlation coefficient were used to evaluate the PD%. The results showed that the average DSC in the breast region and fibro-glandular tissues were 0.937 and 0.853, respectively. In addition, the PD% of the pseudo breast MRI was 2.576%, and the average accuracy between real and pseudo breast MRI images was 0.987.

13.1 Introduction

Mammography is the first major means of breast cancer screening and plays an important role in reducing the mortality of female breast cancer. As early as 1976, Wolfe [1, 2] proposed that Mammographic breast density (MBD) can be used as an independent risk factor for breast cancer. Later, studies by scholars Park and Eriksson [3, 4] showed that gland density was a major risk factor for new and recurrent breast cancer. Gland density is described and recorded as an important sign in ACRBI-RADS (Breast imaging reporting and data system) classification. It is obvious that it is very important to accurately evaluate the density of female breast glands.

In the 2017 China Cancer report, China revealed that the incidence of female breast cancer is also the first among malignant tumors, and the risk of breast cancer among women in big cities is nearly twice as high as that in small cities [5], and the mortality rate is much higher than that in the United States. Although the relationship

Y. Liu · W. Lin (✉) · Y. Cheng
Shanghai Institute of Technology, Shanghai 0086, China
e-mail: linwei_0622@aliyun.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_13

between gland density and the molecular and biological mechanism of breast cancer is still being studied, breast gland density is an important factor affecting breast cancer screening sensitivity and prevention risk prediction. Literature [6] reported that the sensitivity of breast X-ray to detect lesions in fatty and dense glands decreased from 87 to 62.9%. Women with dense glands have a 4.64 times higher risk of breast cancer during their lifetime than women with low gland density due to the increased density of glands and due to their rich glandular matrix [7].

The clinical evaluation of the density of female glands has always been observed by doctors with the naked eye, and the four grades of ABCD of glands have been quantitatively evaluated according to the BI-RADS standard based on experience. In the past, it was difficult to evaluate accurately and quantitatively on the two-dimensional image of mammography. Lehman and Fieselmann [8, 9] have shown that the quantitative evaluation of gland content by computer on breast X-ray images is highly consistent with that by doctors' naked eyes, which indicates that it is reliable to use a computer to quantify glands in a clinic. Quantitative calculation of glands is more accurate. Studies by Ng and Lau et al. [10] confirmed that no matter what equipment was used to examine the glands of the same woman in one year, the physical quantity of the glands was the same. Bonmat et al. [11] used the same female breast X-ray and breast MRI examination to register 2D and 3D gland images obtained by the two examination modes, so as to obtain accurate quantification of glands on conventional breast X-ray images, which makes it possible to quantitatively evaluate glands on breast X-ray images. The advantage of measuring breast density by breast X-ray image lies in its low cost and wide use. Lu [12] comparing breast MRI with breast X-ray images stated that the quantitative evaluation of glands in the same woman has a high consistency within a certain period of time. In order to obtain accurate gland content, segmentation is very important. This study attempts to find the mapping relationship between mammography and breast MRI and to establish a mapping model between mammography and breast MRI. At present, there is little similar work in the literature; firstly, the breast x-ray image is mapped to the pseudo breast MRI image through the mapping model.

13.2 Material and Methods

The proposed method for calculating mammographic PD% consists of the following core steps:

a. Preprocessing of mammographic images about the breast;
b. Reconstruction of mammographic image based on real breast MRI.
c. ROI segmentation and PD% calculation.

MRI data were taken from more than 500 patients at different times in a year, with an average of more than two times per patient. The mammography data used in this experiment were synthesized by breast MRI images mentioned above. Real breast MRI data in this study were all obtained from open source. Breast MRI was performed on 1.5-T scanner (signa, GE Healthcare, Milwaukee, WI) with a bilateral phased array breast coil. Approximately 120 + cases were sorted out from the data collected, each of which contained 60 slices.

13.2.1 Preprocessing of Images About the Breast

During preprocessing of medical images, low-quality images were manually eliminated, so that the adverse impact brought by the original image on the mapping model could be evaded. The breast MRI images were denoised by CBDNet, designed by Kai Zhang's team for real photograph denoising in 2019 [13–15]. It is mainly composed of two parts, one is the noise generation network, which generates an estimated noise image from the input; and the other is a U-Net denoising network (Fig. 13.1). In CBDNet, the original image and noise image predicted by the noise generation network were sent into the second network for denoising.

In order to achieve CBDnet denoising, it was necessary to denoise the original breast MRI manually to create a dataset. In our study, a professional software to view and process medical images (Radiant DICOM Viewer) was used to denoise the original breast MRI images. The prepared dataset was then substituted into CBDNet inputting, and the denoising outcome was shown in Fig. 13.2.

There were 60 slices in the breast MRI dataset, containing information about the entire breast. The mapping model between mammography and breast MRI was implemented by Generative Adversarial Networks (GAN). Only 12 breast MRI slices

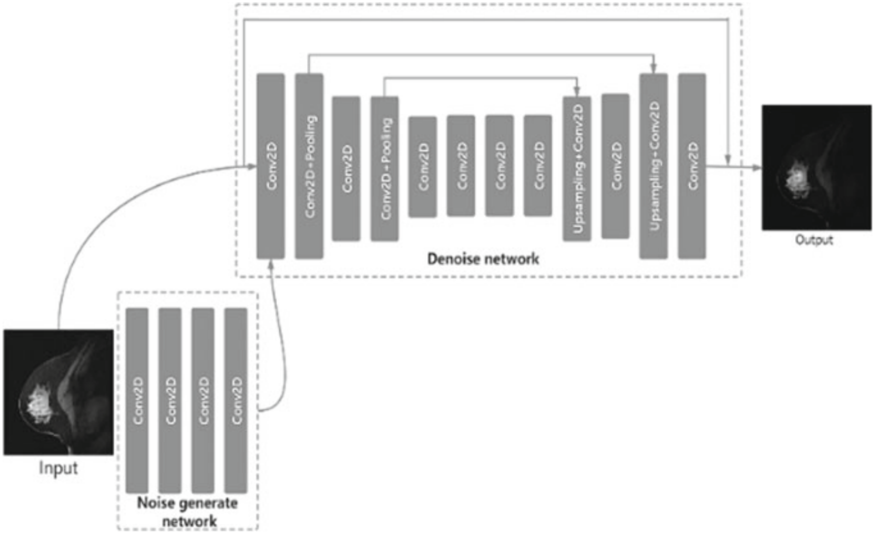


Fig. 13.1 The structure of a CBDNet

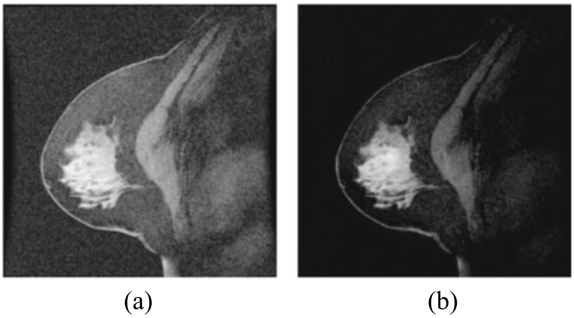


Fig. 13.2 Denoising outcome of CBDNet: **a** original breast MRI image and **b** the same MRI images denoised by CBDNet

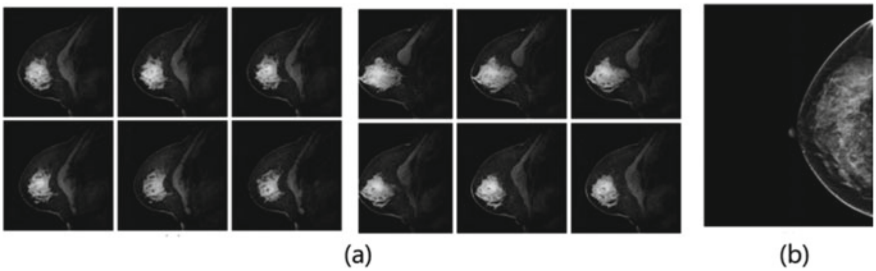


Fig. 13.3 An example of mammography image constructed with breast MRI: **a** 12 slices of breast MRI and **b** the constructed mammography image

were selected for one MRI sample, and they must be physically continuous and as close to the breast center as possible, which was set to exhibit a linear decrease along with its slice number, so as to imitate the attenuation of X-ray energy inside human body during medical X-ray imaging. Breast MRI slices and a sample mammography image are shown in Fig. 13.3.

13.2.2 Construction of the Mapping Model from Mammography to MRI

In this study, we try to generate several breast MRI slices from one mammography image by a mapping model, and GAN was a good tool to realize that [16]. In our model, we planned to set mammography images as inputs, and the output (target) was designed to be 12 breast MRI slices. For GAN, pix-to-pix was a typical one-to-one correlation in previous researches [17], which successfully tackles the problem of realized one-to-one image transformations.

Parameters to generate our GAN went as follows: input dimension was one (for mammography), the output dimension was 12 (for 12 breast MRI slices). When generating DCGan [18], the input was a small dimension of the noise data, and the

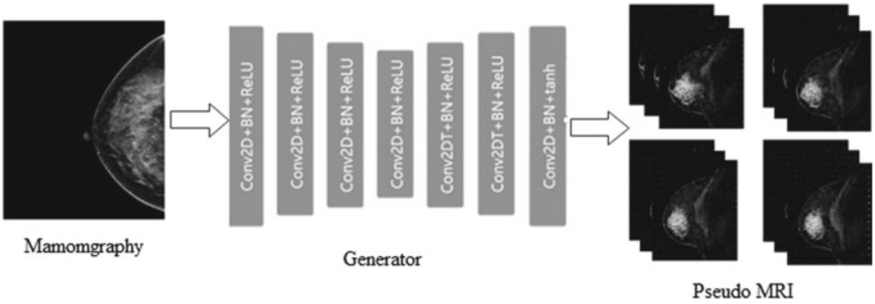


Fig. 13.4 Generator network

target image was output by sampling and increasing dimensions on the convolution layer. Procedures of dimensional reduction and increase correspond to encoding and decoding in the encoder [19], and the generator network is demonstrated in Fig. 13.4.

For the discriminator network, its parameters went as follows: input dimension was 24 (12 real or pseudo breast MRI plus 12 real MRI images), and the output dimension was one, which represented the probability of all the current input. The structure of our discriminator network is shown in Fig. 13.5.

Based on the GAN structures, binary cross-entropy loss function was selected to describe the generator and discriminator networks. The prepared mammography and breast MRI images were put into the GAN network for training, and Fig. 13.6 illustrates the outcome after 20,000 times of alternating training between the two networks.

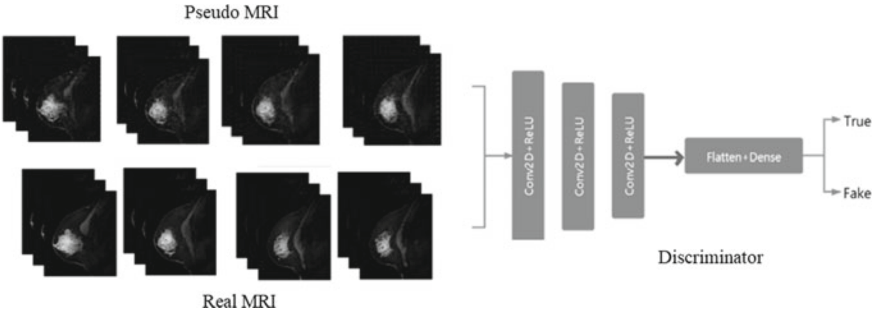


Fig. 13.5 Discriminator network

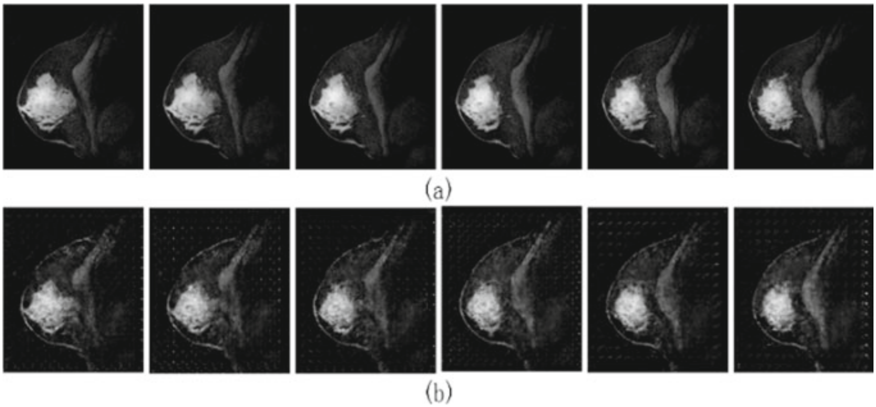


Fig. 13.6 Outcome of applying mapping model on breast MRI: **a** real breast MRI and **b** pseudo breast MRI generated by mapping model

13.2.3 ROI Segmentation and PD% Calculation

In this study, U-Net was used to segment breast regions and fibro-glandular tissue from breast MRI for further PD% calculations. Mask data was needed for training U-Net to automatically achieve such operations, and we used a software called Labelme to create mask datasets (Fig. 13.7). The K-means algorithm was used to realize binarization for the fibro-glandular region, which were implemented as follows [20]:

a. K cluster centers were randomly initialized.
b. While the centers of K clusters moved:

    I. Calculate their distances to reach all samples.
    II. Separate the samples by their nearest clusters.
    III. Recalculate cluster centers.

   Repeat this process until centers didn't move anymore.

c. And the results were output.

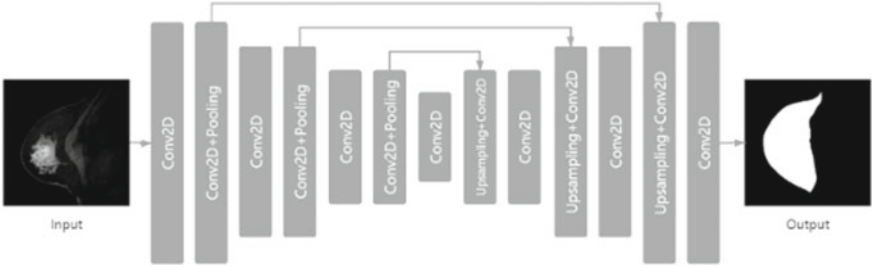


Fig. 13.7 Structure of the U-Net

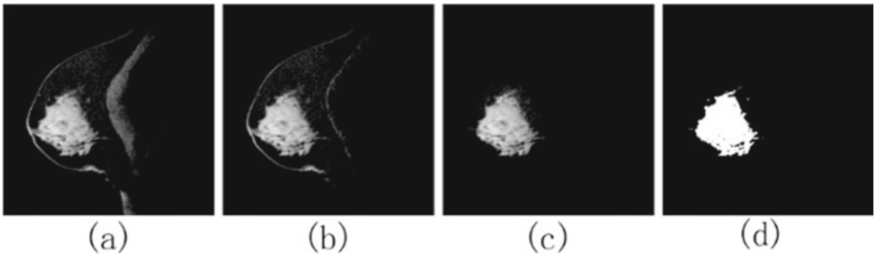


Fig. 13.8 Outcome of U-net segmentation: **a** original breast MRI, **b** breast region, **c** fibro-glandular tissue region, and **d** binarization result from

The region reflecting the fibro-glandular tissue was separated from the background with K-means algorithm, and K = 2. The segmentation outcome is shown in Fig. 13.8.

FGT% is defined as the relative volume percentage of fibro-glandular tissue within the breast, and it is calculated as:

$$\text{PD\%} = FGT\% = \frac{|FGT|}{|\text{Breast}|} \times 100 \tag{13.1}$$

13.2.4 Validation

The effect of U-Net segmentation on breast and fibro-glandular tissue was evaluated by the DSC difference between regions obtained by manual definitions and U-Net, which was a common parameter for this application. It will directly affect the PD% and the evaluation results of our mapping model, taking U-Net segmentation results on breast region as an example.

$$\text{s} = \frac{2|A \cap B|}{|A| + |B|} \tag{13.2}$$

where A was the manually defined breast region and B was the breast region after U-Net segmentation. The closer the DSC is to 1, the better effect U-Net segmentation produces.

PD% values were calculated for both pseudo and real breast MRI images to evaluate the pseudo's effectiveness. Linear regression and Pearson's correlation coefficients were used to evaluate the correlation of PD%, and the calculation was carried out by the following equation:

$$\text{r} = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N\sum x_i^2 - (\sum x_i)^2}\sqrt{N\sum y_i^2 - (\sum y_i)^2}} \tag{13.3}$$

where x is the PD content of real breast and Y is the gland content calculated by the algorithm.

In order to evaluate the stability of our method of calculation, identical procedures were carried out on the same patient at different times in a year. The smaller the numerical changes of breast PD% obtained were, the more stable the calculation method was.

13.3 Experiments and Results

13.3.1 Evaluation of the Mapping Model from Mammography to Breast MRI Images

DSC was used to evaluate the spatial consistency between real and pseudo breast MRI images as a reflection of the mapping model's performance. In one experiment, breast and fibro-glandular tissue regions were segmented from 30 random real images of their corresponding pseudo ones, followed by subsequent calculations for mean DSC. A total of ten experiments were conducted, and the results are shown in Table 13.1. The overall mean DSC was 0.937 with a mean standard deviation of 0.0069 (mean standard deviation) for the breast region in real and pseudo MRI images. For real and pseudo fibro-glandular tissue regions, the overall mean DSC was 0.8509, and the mean standard deviation of 0.039. Therefore, we could conclude that our mapping model is significantly less effective for the fibro-glandular tissue.

Table 13.1 DSC calculations of breast and fibro-glandular tissue in real and pseudo MRI images

No	Breast region				Breast fibro-glandular tissue			
	Average	Max	Min	Standard Deviation	Average	Max	Min	Standard Deviation
1	0.938	0.956	0.929	0.008	0.836	0.904	0.761	0.049
2	0.938	0.950	0.928	0.007	0.839	0.883	0.773	0.032
3	0.936	0.956	0.921	0.010	0.871	0.924	0.797	0.035
4	0.937	0.948	0.928	0.006	0.851	0.911	0.755	0.047
5	0.936	0.954	0.919	0.012	0.839	0.885	0.773	0.037
6	0.941	0.956	0.931	0.006	0.844	0.884	0.791	0.030
7	0.935	0.950	0.924	0.009	0.857	0.911	0.777	0.044
8	0.938	0.953	0.925	0.008	0.857	0.899	0.796	0.035
9	0.938	0.955	0.924	0.008	0.870	0.924	0.815	0.036
10	0.936	0.953	0.924	0.008	0.845	0.911	0.764	

Table 13.2 PD% calculations of breast and fibro-glandular tissue in real and pseudo MRI

No	Real PD%	Pseudo PD%	|Error| (%)	Max error (%)	Min error (%)	Average error (%)
1	57.01	53.64	3.37	5.97	0.28	2.576
2	66.49	66.21	0.28			
3	20.10	17.37	2.73			
4	36.36	35.39	0.97			
5	31.65	27.72	3.93			
6	10.45	9.87	0.58			
7	25.80	23.41	2.39			
8	68.38	67.17	1.21			
9	13.70	9.37	4.33			
10	36.91	31.12	5.97			

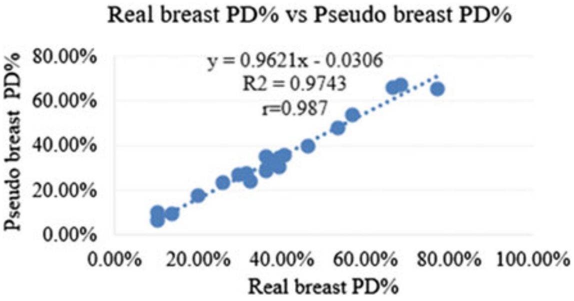


Fig. 13.9 Linear regression results for PD% between real and pseudo MRI

13.3.2 *Evaluation of the Breast PD%*

Table 13.2 showed the breast PD% of ten data in the test dataset, and the maximum, minimum, and average errors of PD% were 5.97%, 0.28%, and 2.576% between real and pseudo MRI images, respectively. In order to further explore the stability for PD% calculation, 20 patients were randomly selected from the data set, and their breast PD% values were calculated based on the method mentioned with linear regression and the results are shown in Fig. 13.9.

13.4 Discussion

In this study, a GAN-based mapping model between mammography and breast MRI was constructed, providing a new way to calculate PD%. It can directly transform mammography images into pseudo breast MRI, and then segment the breast and fibro-landular tissue regions, enabling easy calculations for the breast PD%.

Calculation results for DSC verify the effectiveness of our model. In terms of breast PD% calculation, the breast density and error analysis show that real and pseudo breast MRI images are highly consistent and correlated, suggesting actual feasibility for 3D reconstruction results from mammography to be applied as MRI images in actual practice.

Further efforts could focus on building a more promising mapping model to generate pseudo breast MRI images with higher quality, so that a more accurate fibro-glandular tissue segmentation could be achieved.

References

1. Wolfe, J.N., Chief, M.D.: Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer **37**(5), 2486–2492 (1976)
2. Wolfe, J.N.: Breast patterns as an index of risk for developing breast cancer. Am. J. Roentgenol **126**, 1130–1137 (1976)
3. Eriksson, L., Czene, K., Rosenberg, L.U., Törnberg, S., Humphreys, K., Hall, P.: Mammographic density and survival in interval breast cancers. Breast Cancer Res. 15(3), 48 (2013)
4. Park, C.C., Rembert, J., Chew, K., Moore, D., Kerlikowske, K.: High mammographic breast density is independent predictor of local for invasive breast cancer. Int. J. Radiat. Oncol. Biol. Phys. **73**(1), 5–79 (2009)
5. Chen, W.Q., Zheng, R.S., Zhang, S.W.: Cancer incidence and mortality in China in 2013: an analysis based on urbanization level. Chin. J. Cancer Res. **29**(1), 1–10 (2017)
6. Carney, P.A., Miglioretti, D.L., Yankaskas, B.C., Kerlikowske, K.: Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann. Intern. Med. **138**(3), 168–175 (2003)
7. McCormack, V.A., dos Santos Silva, I.: Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol. Biomarkers **15**(6), 1159–1169 (2006)
8. Lehman, C.D., Yala, A., Schuster, T., Dontchos, B., Barzilay, R.: Mammographic breast density assessment using deep learning: clinical implementation. Radiology **290**(1), 52–58 (2018)
9. Fieselmann, A., Frnvik, D., Lng, K.: Volumetric breast density measurement for personalized screening: accuracy, reproducibility, consistency, and agreement with visual assessment. J. Med. Imaging **6**(3), 031406 (2019)
10. Ng, K.H., Lau, S.: Vision 20/20: Mammographic breast density and its clinical applications. Med. Phys. **43**(12), 7059–7077 (2015)
11. Marcos, E.G.: Glandular tissue pattern analysis through multimodal MRI-mammography registration. M.S. thesis, Univ (2018)
12. Lu, L.J.W., Nishino, T.K.: Comparison of breast tissue measurements using magnetic resonance imaging, digital mammography and a mathematical algorithm. Phys. Med. Biol. 57(21), 6903–6927 (2012)
13. Zhang, K., Zuo, W.M., Chen, Y.J., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process **26**(7), 3142–3155 (2017)
14. Zhang, K., Zuo, W.M., Zhang, L.: FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans. Image Process. **27**(4), 4608–4622 (2018)
15. Guo, S., Yan, Z., Zhang, K., Zuo, W.M., Zhang, L.: Toward convolutional blind denoising of real photographs. In: CVPR 2019, pp. 1712–1722 (2019)
16. Goodfellow, I.J., Abadie, J.P., Mirza, M., Xu, B.: Generative adversarial nets. In: NIPS 2014, Montreal, QC, Canada, pp. 2672–2680 (2014)

17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.: Image-to-image translation with conditional adversarial networks. In: CVPR 2017, Honolulu, HI, pp. 1125–1134 (2017)
18. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv 1511, 06434 (2015)
19. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: 25th ICML 2008, Helsinki, pp. 1096–1103, Finland (2008)
20. Kanungo, T., Mount, D.M., Netanyahu, N.S., Christine, C.D., Siverman, R., Angela, Y.: An efficient k-means clustering algorithm: analysis and implementation. IEEE T. Pattern. Anal **24**(7), 881–892 (2002)

Chapter 14
Machine Learning Enabled Edge Computing: A Survey and Research Challenges

Shilong Xu, Zhuozhi Yu, Kai Fu, Qingmin Jia, and Renchao Xie

Abstract Edge Computing has been regarded as a significant technology in 5G/B5G communication networks, which can improve the performance of network by deploying storage as well as computing resources on the edge of the network. At the same time, machine learning as a significant optimization approach has attracted wide attention in industry and academia. Hence, it is a natural trend to use machine learning to optimize the performance of edge computing. In addition, many excellent works on machine learning and edge computing have been done. Hence, it is necessary to survey these works. In this article, we survey the latest development for machine learning enabled edge computing. We first introduce the edge computing and machine learning separately and present the motivation for machine learning enabled edge computing. And then the research issues of the machine learning enabled edge computing from the perspective of user side and network side are presented. Finally, the research challenges and future directions are presented.

14.1 Introduction

With the continuous development of new network technology, a lot of new Internet applications or services are widely used, such as 4K/8K ultra-high definition video, Augmented Reality, Virtual Reality, etc., which bring new challenges and problems for the mobile network. In addition, as the Cisco VNI report reveals [1], by 2022, the global mobile data traffic will reach 77 EB per month, the annual traffic will reach nearly 1 zettabyte, and the mobile video traffic will account for 79% of the global

S. Xu (✉) · Z. Yu · K. Fu
State Grid Information & Telecommunication Co., Ltd., Beijing Branch, Beijing 100031, China
e-mail: xushilong@sgitg.sgcc.com.cn

Q. Jia · R. Xie
Purple Mountain Laboratories, Nanjing 211111, China

R. Xie
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_14

mobile data traffic. To cope with these challenges, edge computing has been presented to optimize the performance of network as well as the Quality of Experience (QoE) of consumers [2]. The edge computing has storage, network and computing resources, and can provide cloud computing infrastructure as well as cloud computing service on the edge of network. Meanwhile, this environment has the characteristics of ultra-high bandwidth, ultra-low latency and real-time network information access capability [2]. Once the edge computing technology has been proposed, it has attracted wide attention from industry and academia.

At present, many edge computing solutions have been proposed, such as cloudlet, multi-access edge computing (MEC) as well as fog computing. The rapid development of edge computing brings many opportunities for the improvement of network performance and user experience. For example, in the image recognition field, sensor device traditionally transmits the captured images to a cloud computing server to process, which usually leads to a higher delay. However, after deploying the edge computing, the delay will reduce a lot. Meanwhile, the deployment of edge computing also brings many challenges. And many new research issues need to be solved, such as computation offloading, content caching and delivery, resource management and so on.

To cope with these problems in edge computing, many optimization approaches and schemes have been proposed. Meanwhile, machine learning (ML) as a major branch of artificial intelligence (AI) has been regarded as one of the most popular methods to optimize the performance of the network. In addition, edge computing can provide cloud capabilities as well as resources to satisfy the requirements of machine learning. Accordingly, some problems of edge computing field have been solved by machine learning method recently.

In this article, a brief survey of machine learning enabled edge computing is introduced. The remaining of this article is organized as follows. We first introduce the machine learning, and then we present the motivation and the main research issues for machine learning enabled edge computing from the user side and network side perspective. We also outline some research challenges for machine learning enabled edge computing. Finally, we conclude the article.

14.2 The Overview of Edge Computing and Machine Learning

14.2.1 Edge Computing

Since edge computing technology was proposed, it has attracted wide attention from academia and industry. At present, three kinds of computing paradigms are widely recognized, which include fog computing, cloudlet and MEC. In this subsection, these edge computing solutions are introduced. In 2009, the authors in [3] proposed the concept of cloudlet, which represents the middle layer of three-tier structure:

IoT device, cloudlet as well as cloud. Actually, cloudlet is a miniaturized cloud computing data center, which is used to enhance the mobility of network edge. And the main goal of cloudlet is to support computing intensive and interactive services by providing powerful computing power for mobile terminals.

Then, the researcher of Cisco proposed the concept of fog computing in 2012 [4]. Fog computing sinks cloud computing facilities to the edge of the network, and then forms a new network service. In order to promote the development of fog computing, the OpenFog Consortium was created in 2015, which aims at the industrialization and standardization of fog computing.

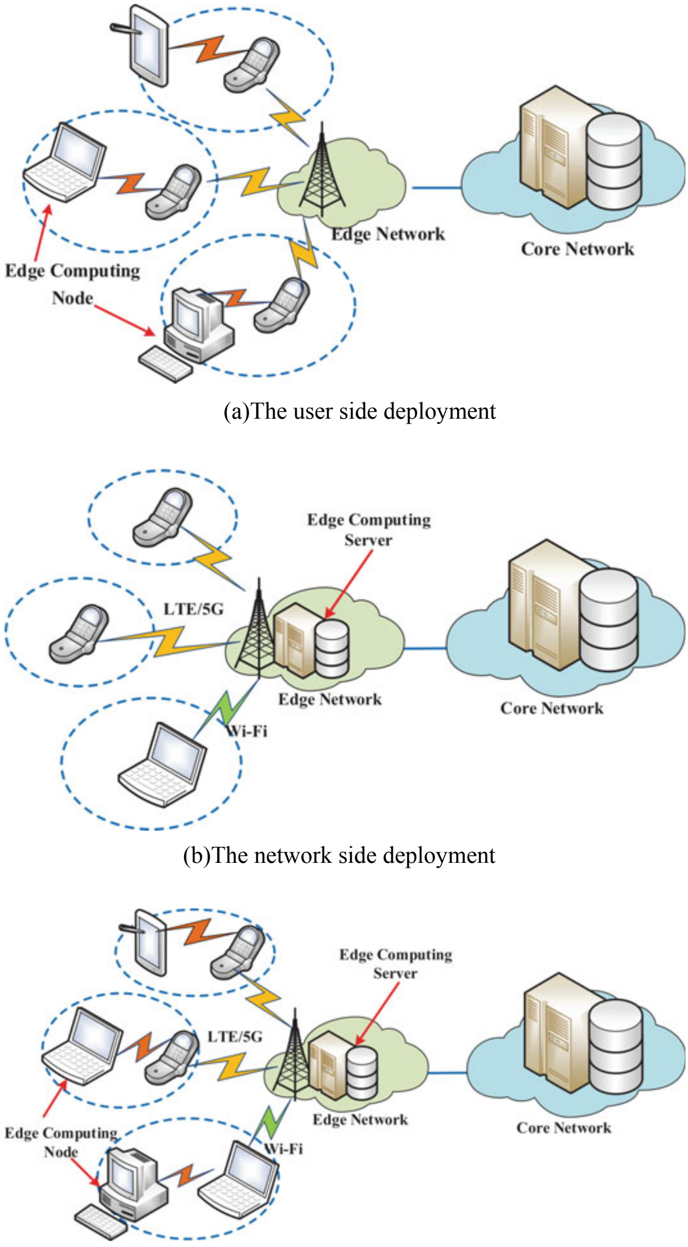
With the continuous evolution of mobile communication technology, the fifth generation mobile communication (5G) has high requirements, namely ultra-high bandwidth, ultra-low delay and ultra large connection [5]. It is necessary to make a deep change in mobile network from the perspective of network architecture. As the idea of deploying cloud computing resources on the edge of mobile network has become a common consensus in the academia and industry, Mobile Edge Computing (MEC) has been gradually nurtured and developed. In 2014, the ETSI first proposed the concept of MEC, and MEC is defined as providing cloud computing capabilities as well as cloud computing service environments on the edge of mobile networks. In addition, ETSI established the MEC Industry Specification Group to promote the development of a unified specification for running third-party applications in MEC multi-tenant environments. In 2016, ETSI further extended the definition of “M” in MEC from the perspective of access mode. Namely, the mobile terminal can access the edge computing server not only through cellular networks but also through fixed networks, such as Wi-Fi. Accordingly, the concept of MEC was extended to multi-access edge computing.

In summary, the main idea of edge computing (EC) is to make cloud computing resource facilities closer to user equipment (UE). Generally, there are three modes of deploying edge computing: user side deployment, network side deployment and hybrid deployment. In user side deployment mode as shown in Fig. 14.1a, many network devices integrating storage as well as computing resource can be considered as the edge, such as PC, mobile terminal and so on. In this case, the end user can use these edge computing devices to deal with data or computation tasks. In network side deployment mode as shown in Fig. 14.1b, the EC server not only can assist to deal with computation tasks for end users but also can cache popular content for content delivery with low delay.

In the hybrid deployment mode as shown in Fig. 14.1c, the EC server or device can be deployed at the user side as well as the network side.

14.2.2 Machine Learning

In this subsection, we first present the main concept of machine learning as well as the main classification of machine learning. And then we introduce the main application for machine learning. Machine learning has a long history. And with



(a)The user side deployment

(b)The network side deployment

(c) The hybrid deployment

Fig. 14.1 The deployment mode for edge computing

the development of big data as well as the improvement of computing power, the machine learning has progressed dramatically recently [6]. Machine learning can be divided into *supervised learning, unsupervised learning, reinforcement learning as well as deep learning*.

In supervised learning, the training data has a pre-defined tag, and a classifier is built and trained to predict the tag of test data. The classifier is properly optimized and adjusted to achieve a suitable level of accuracy. Additionally, supervised learning problems can be further grouped into regression and classification problems. However, in unsupervised learning, training data is not labelled, and a classifier is designed by deducing existing patterns or cluster in the training datasets. And unsupervised learning problems can be further grouped into clustering and dimension reduction problems. As for reinforcement learning, the training algorithm maps the action to the environment to maximize the reward. At the same time, the classifier is trained and tested repeatedly to find the action with maximum reward [7].

14.3 The Motivation for Machine Learning Enabled Edge Computing

On the one hand, machine learning can improve the system performance and service quality for edge computing. Meanwhile, the requirements of end users for service quality and experience quality are also getting higher and higher. Hence, a lot of services are deployed in edge computing, such as video content caching, task computation service and so on. Machine learning can improve the performance of these services. For example, machine learning can be applied to the computation offloading issue. The terminal device can use the machine learning approach to optimize the decision problem of computing task offloading [8].

On the other hand, EC can provide computing power for machine learning. With the explosive growth of network data as well as the development big data technology, the network service provider and content service provider usually need to collect, process and analyze these data, and then make the optimal decision, thus improving the service quality. Generally, these operations need machine learning approach, which requires a lot of computing and storage resource. Edge computing integrates storage as well as computing resources in the edge of network, and can assist service provider to store, process and analyze these data. For example, supervised learning needs a lot of data training to improve system performance.

14.4 The Representative Research Issues for Machine Learning Enabled Edge Computing

In this section, we present the representative research issues for machine learning enabled edge computing. Because the edge computing is deployed at the edge of network, there is a lot of interaction between user terminals and edge computing. Hence, we introduce the research issues from two perspectives: user side and network side (Table 14.1).

14.4.1 From the Perspective of the User Side

(1) *Computation offloading decision in edge computing*

Nowadays, mobile equipment needs to handle more and more computing tasks, thus consuming more computing resources. However, the battery capacity and computing power of mobile terminal are limited, which makes computation tasks difficult to be processed at mobile terminal. Hence, offloading computing tasks to MEC server becomes a significant approach [14].

Table 14.1 Machine learning enabled edge computing from the perspective of the user side

Research issues	Approach	Contributions
Computation offloading decision [9]	Deep supervised learning	Developing a dynamic computation offloading solution using Deep Supervised Learning technology
Computation offloading decision [8]	Deep reinforcement learning (DRL)	Proposing a DRL-based computation offloading solution
Computation offloading decision[10]	Reinforcement learning	Use a game-theoretic machine learning method to make optimal computation offloading decisions
Adaptive video streaming [11]	Deep reinforcement learning	Predicting network throughput by Deep reinforcement learning, and then select the most matching bitrate version
Adaptive video streaming [12]	Supervised learning	Predicting network throughput by Deep reinforcement learning, and then select the most matching bitrate version
Adaptive video streaming [13]	Supervised learning and unsupervised learning	Predicting network throughput by Deep reinforcement learning, and then select the most matching bitrate version

In computation offloading, a key issue is computation task offloading decision. Generally, computation task offloading decision can be divided into three types: *local execution, full offloading as well as partial offloading*. Local execution means the whole computation task is processed locally at the mobile terminal, and full offloading means the whole computation task is offloaded and processed by the edge computing server. These two offloading decisions are relatively simple. In comparison, partial offloading is more complex, which requires that a part of the computation task is processed locally while the rest is offloaded to the edge computing server. In computation offloading decision, many factors need to be considered, such as execution delay, energy consumption, communication link, and edge computing server location. With the development of machine learning, it is necessary to use machine learning approach to optimize the performance of computation offloading decision.

In [9], the authors propose a dynamic computation offloading scheme, considering the mobile device load, communication capacity and computing power. And then the authors use Deep Supervised Learning approach to optimize the computation task offloading performance. In [8], the authors propose a computation offloading solution based on DRL in the scenario of ad-hoc mobile cloud, which can enable the user to make near-optimal task offloading decisions. In [10], the authors investigate the problem of multi-user computation offloading, and use a game-theoretic machine learning approach to make the offloading decisions.

(2) *Adaptive video streaming in edge computing*

In response to the rapid growth of mobile video traffic, caching the video content in MEC server has been considered as a significant approach. Moreover, duo to the randomness of the wireless network condition, video transmission rate needs to match the wireless network condition. Hence, dynamic adaptive streaming over HTTP (DASH) is presented to solve this problem. However, In DASH scheme, the video content is divided into small video clips, and each video clip is encoded into multiple bit rate versions, including several seconds of video. A new problem is introduced, namely, the mobile terminal should select which bitrate version of the video segment. With the development of machine learning, using the machine learning approach to optimize the adaptive video streaming in edge computing becomes a significant method.

In [11], the authors propose a scheme named Pensieve based on deep reinforcement learning. The penseve scheme can train a neural network model, which can select the bit rate for the future video block according to the observation results collected by the client video player. In [12], the authors use supervised ML means to predict one objective QoE metric, video starvation, with the users features, recorded at the beginning of each video session. In [13], the authors propose a ML-based video admission control and resource management approach, and a multi-stage learning system is developed.

14.4.2 From the Perspective of the Network Side

(1) *Content caching and delivery in edge computing:*

Due to the edge computing integrating computing and storage resource, it is very convenient and feasible to process and analyze data at the edge of network. In addition, with the development of social networks, more and more data are generated by end users. In order to provide end users better and more customized services, it is very important to predict and recommend content for end users. Accordingly, using machine learning approach to process and analyze data has attracted wide attention from industry as well as academia. By machine learning methods to analyze a lot of user interests as well as network behaviors, active caching content is an important research issue.

In [15], the authors propose a data-driven strategy to constructing an efficient performance model of CDN cache server group, and they use a deep neural network to forecast the arrival rate. In [16], the authors present a proactive caching approach, based on MEC framework to reduce transmission cost and improve the QoE of end users. In this solution, the authors propose a content popularity estimating scheme based on Transfer Learning (TL) method. In [17], the authors present a Deep Reinforcement Learning-based scheme for base station caching, and aims at maximizing the long-term cache hit rate.

Moreover, the deployment of edge computing is usually distributed, and the resources of each edge computing node are usually limited. Hence, it is valuable to enhance the cooperation in edge computing. For example, the popular video content can be divided into multiple parts, and placed in several adjacent edge computing nodes separately, thus improving the caching resource utilization.

(2) *Service continuity (or service migration) in edge computing:*

The mobility of mobile devices as well as the limited coverage of MEC services will lead to a serious decline in network performance, and even interrupt the continuous service of edge computing. Hence, it is significant to guarantee the service continuity. In addition, when the mobile terminals move at high speed, not only the session connections need to switch quickly between the base stations but also network services need to migrate between different edge computing servers. Hence, an efficient service migration scheme is very important for the service continuity of edge computing. In order to ensure the continuity of end-to-end session and QoS/QoE, it is significant to research edge computing mobility and service continuity.

Mobility of user behavior often can be a “track record,” especially for high-speed mobile scenarios, so that the user’s mobile trajectory can usually adopt the method of artificial intelligence (such as machine learning, etc.) to forecast, mobility switch will often lead to service migration, so the design of collaborative mobile switching service migration mechanism is of great significance.

In [18], the authors present an adaptive ML-based service continuity model, which can accurately predict the key characteristics of live migration. The container technology has made great progress, which has the advantage on resource utilization and agility. In fact, the main service migration includes virtual machines migration and container migration. And the service migration based on container technology is also an important research direction [19].

(3) *Resource management in edge computing*

In order to improve the resource utilization, reduce the cost and guarantee the quality of experience of the end users, it is significant to optimize the resource management and orchestration. In addition, machine learning as a powerful optimization tool can be used for the resource management as well as orchestration in edge computing. In [20], the authors design a system that can learn to manage resources directly from experience using deep reinforcement learning.

On the other hand, the multidimensional resources tradeoff is also an important issue, due to the finiteness of edge computing resources. For example, the video caching as well as transcoding problem for the adaptive video streaming in edge computing. Generally, the edge computing server integrates limited computing and caching resources, thus caching all bitrate versions video segment will consume a large amounts of storage resources. However, only caching the highest bitrate version video segment and transcoding will consume a lot of computing resources. Hence, it is significant to make a tradeoff decision for resource management (See Table 14.2).

Table 14.2 Machine learning enabled edge computing from the perspective of the network side

Research issues	Approach	Contributions
Data-driven content caching and delivery [15]	Sequence Learning	Use a deep neural network to evaluate the performance of cache server groups in CDN
Data-driven content caching and delivery [16]	Transfer Learning	Exploiting a Transfer Learning based method to estimate content popularity, and build the proactive caching model
Mobility and service continuity [18]	Supervised Learning	An adaptive machine learning based model that can predict the key characteristics of live migration
Resource management [20]	Deep Reinforcement Learning	Design a system that can learn to manage resources directly from experience using deep reinforcement learning

14.5 The Research Challenges for Machine Learning Enabled Edge Computing

14.5.1 Information-Centric Networking

Information-Centric Networking (ICN) is a clean-slate network solution, which has become a promising choice for the future networking architecture [21]. And ICN aims to achieve content centric communication instead of host-centric end-to-end communication. In ICN, many research issues can be solved by machine learning, such as data forwarding and routing, congestion control and so on. Through learning the ICN network data, the network node can make an optimized decision so as to improve the performance of ICN networking system.

14.5.2 Blockchain

Recently, blockchain has become a hot research issue and blockchain technology has many advantages, such as full decentralization, true redundancy, complete privacy and so on. In order to ensure the integrity and effectiveness of data in blockchain, a calculation process is needed, including mining, consensus process and so on. Therefore, blockchain needs enough computing power, and edge computing can provide computing power for blockchain. Hence, combing the edge computing and blockchain has become a significant research direction [22]. In blockchain-based edge computing, computing resource management and energy-efficiency optimization are extremely important research direction [23].

14.5.3 Security and Privacy

Security and privacy are the key technologies for the normal operation of the edge computing system. And federated learning plays an important role in security as well as privacy for machine learning enabled edge computing. Federated learning is an important machine learning scheme, in which the shared prediction model can be used for collaborative learning by multiple distributed nodes using their local stored data [24]. Hence, it is significant to build such an edge federated learning system. However, there are still a lot of challenges in edge federated learning, such as communication efficiency, migration and scheduling.

14.5.4 *Intelligent Interconnection and Sharing*

At present, a large number of applications based on machine learning (or artificial intelligence technology) are deployed in edge computing system. Artificial intelligence has shown the trend of ubiquitous intelligence, that is, intelligence is everywhere. This provides convenience for people to make full use of intelligent resources, models, etc. How to make full use of ubiquitous intelligent resources and services has become an important issue for the current academic community. In this context, the concept of Intelligent Networking is proposed [25]. Intelligent Networking aims to interconnect distributed intelligent resources, make full use of and share intelligent resources and intelligent services, improve the utilization rate of intelligent resources and improve the intelligent decision-making level of devices and applications. It is a new network paradigm of deep integration of edge computing technology and artificial intelligence technology in the future. Therefore, the study of Intelligent Interconnection and Sharing is of great significance to break through the technical bottleneck of Internet and artificial intelligence and promote the development of edge computing.

14.6 Conclusion

In this article, we survey the latest development in ML-enabled EC. First, the EC and ML are introduced separately. And then, the motivation for ML-enabled EC is analyzed. Next, the research issues of the ML-enabled EC from the perspective of user side and network side are presented. Finally, we discuss and analyze the research challenges and future directions.

References

1. Cisco: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper (2017)
2. ETSI: Multi-access edge computing. http://www.etsi.org/technologiesclusters/technologies/multi-access-edge-computing.
3. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The case for VM-based cloudlets in mobile computing. IEEE Pervasive Comput. **8**(4), 14–23 (2009)
4. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the internet of things. In: Proceedings of ACM MCC, New York, NY, USA, pp. 13–16 (2012)
5. Jia, Q., Xie, R., Huang, T., Liu, J., Liu, Y.: Caching resource sharing for network slicing in 5g core network: a game theoretic approach. J. Organ. End User Comput. (JOEUC) **31**(4), 1–18 (2019)
6. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)
7. Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A.: Deep reinforcement learning: a brief survey. IEEE Signal Process. Mag. **34**(6), 26–38 (2017)

8. Le, D.V., Tham, C.K.: A deep reinforcement learning based offloading scheme in ad-hoc mobile clouds. In: Proceedings of IEEE INFOCOM WKSHPS, pp. 760–765 (2018)
9. Yu, S., Wang, X., and Langar, R.: Computation offloading for mobile edge computing: a deep learning approach. In: Proceedings of IEEE PIMRC, pp. 1–6 (2017)
10. Cao, H., Cai, J.: Distributed multiuser computation offloading for cloudlet-based mobile cloud computing: a game-theoretic machine learning approach. IEEE Trans. Veh. Tech. **67**(1), 752–764 (2018)
11. Mao, H., Netravali, R., Alizadeh, M.: Neural adaptive video streaming with pensieve. In: Proceedings of ACM SIGCOMM, pp. 197–210 (2017)
12. Lin, Y.T., Oliveira, E.M.R.S., Jemaa, B., Elayoubi, S.E.: Machine learning for predicting QOE of video streaming in mobile networks. In: Proceedings of IEEE ICC, pp. 1–6 (2017)
13. Grazia, M.D.F.D., Zucchetto, D., Testolin, A., Zanella, A., Zorzi, M., Zorzi, M.: Qoe multi-stage machine learning for dynamic video streaming. IEEE Trans. Cognitive Commun. Netw. **4**(1), 146–161 (2018)
14. Jia, Q., Xie, R., Tang, Q., Li, X., Huang, T., Liu, J., Liu, Y.: Energy efficient computation offloading in 5g cellular networks with edge computing and d2d communications. IET Commun. **13**, 1122–1130(8) (2019)
15. Wu, Z., Lu, Z., Zhang, W., Wu, J., Huang, S., Hung, P.C.: A datadriven approach of performance evaluation for cache server groups in content delivery network. J. Parallel Distrib. Comput. **119**, 162–171 (2018)
16. Hou, T., Feng, G., Qin, S., Jiang, W.: Proactive content caching by exploiting transfer learning for mobile edge computing. In: Proceedings of IEEE GLOBECOM, pp. 1–6 (2017)
17. Zhong, C., Gursoy, M.C., Velipasalar, S.: A deep reinforcement learning-based framework for content caching. In: Proceedings of IEEE CISS, pp. 1–6 (2018)
18. Jo, C., Cho, Y., Egger, B.: A machine learning approach to live migration modeling. In: Proceedings of ACM SoCC, pp. 351–364 (2017)
19. Machen, A., Wang, S., Leung, K.K., Ko, B.J., Salonidis, T.: Live service migration in mobile edge clouds. IEEE Wirel. Commun. **25**(1), 140–147 (2018)
20. Mao, H., Alizadeh, M., Menache, I., Kandula, S.: Resource management with deep reinforcement learning. In: Proceedings of ACM HotNets, pp. 50–56 (2016)
21. Jia, Q., Xie, R., Huang, T., Liu, J., Liu, Y.: The collaboration for content delivery and network infrastructures: a survey. IEEE Access **5**, 18088–18106 (2017)
22. Xiong, Z., Zhang, Y., Niyato, D., Wang, P., Han, Z.: When mobile blockchain meets edge computing: challenges and applications, arXiv preprint arXiv:1711.05938 (2017)
23. Luong, N.C., Xiong, Z., Wang, P.: Optimal auction for edge computing resource management in mobile blockchain networks: a deep learning approach. In: Proceeding of IEEE ICC, pp. 1–6 (2018)
24. Xia, Q., Ye, W., Tao, Z., Wu, J., Li, Q.: A survey of federated learning for edge computing: research problems and solutions. High-Confidence Comput. 100008 (2021)
25. Yu, F.R.: From information networking to intelligence networking: motivations, scenarios, and challenges. IEEE Netw. (2021)

Chapter 15
Application of Deep Learning in Maize Image Segmentation

Lijuan Shi, Xingang Xie, and Yang Zhang

Abstract The selection and breeding of maize varieties demand the rapid acquisition of parameters of the maize bald tip. And image segmentation is a crucial step for the automatic acquisition of parameters of the maize bald tip. This study proposes a U-Net model suitable for semantic segmentation of maize ear images. Compared with the traditional machine learning algorithms, the deep learning network doesn't need the complex feature extraction process. In this work, we label all the images and make data augmentation at first. Then to improve the performance of the maize image segmentation model, we discuss the impact of the learning rate, convolution kernel size, and pooling method on the performance of the model. The experimental results show that the best learning rate is 0.0001, the suitable size for convolution kernel is 3 × 3, and the optimal pooling method is max pooling. It is concluded that the maize image segmentation model with these settings can obtain higher segmentation accuracy and better generalization ability of the model effectively. The goal is to improve the accuracy and efficiency for bald tip recognition, provide new tools for automatic measurement of maize phenotype, and further serve maize breeding and cultivation.

15.1 Introduction

Image segmentation is the technology and process of dividing the image into some specific regions on the basis of their unique properties. It is a crucial stage in image analysis [1–5]. Currently, the most popular segmentation method is based on machine learning [6]. In the traditional machine learning methods, the feature design and selection are completed manually. In addition, the pros and cons of the features depend on the expert's domain knowledge background. In recent years, deep learning methods have been introduced to overcome the above drawbacks, allowing the algorithm to learn features from the original image automatically [7].

L. Shi · X. Xie (✉) · Y. Zhang
College of Informatics, Huazhong Agricultural University, Wuhan, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_15

Since deep learning was proposed by Hinton in 2006, it has been widely used in agriculture with its advantages of independent feature selection, shared weights, and local receptive fields [8]. Literature [9] presents a computer vision-based method using spatial and temporal information of nursing behavior under commercial farm conditions for automatic recognition of nursing interactions. In this method, a fully convolutional network was applied to segment sows accurately. Literature [10] performed a test on ten pigs using three techniques used by the human face recognition. Literature [11] presented a method based on deep learning to detect diseases and pests in tomato plants. Literature [12] used the convolution neural network to classify plant diseases and suggest remedies.

During the growth and development of maize, due to factors such as differences between varieties, soil, nutrition, climate, cultivation management, diseases, and insect pests, the male and female ears have poor pollination and fruiting, resulting in different degrees of bald tips and then different degrees of grain shortage in maize ears. Therefore, the selection and breeding of maize varieties demand the rapid acquisition of parameters of the maize bald tip. It is an important way to calculate the parameters of the maize bald tip quickly by using the camera instead of human eyes to obtain images of maize ears and by using machine learning instead of the human brain to recognize the bald tip area.

Therefore, this paper selects the U-Net model, which is suitable for semantic segmentation in the deep learning network model to construct the segmentation model of maize ear images and discusses the impact of the learning rate, convolution kernel size, and pooling method on model performance. The goal is to improve the accuracy and efficiency of bald tip recognition, provide new tools for automatic measurement of maize phenotype, and further serve maize breeding and cultivation.

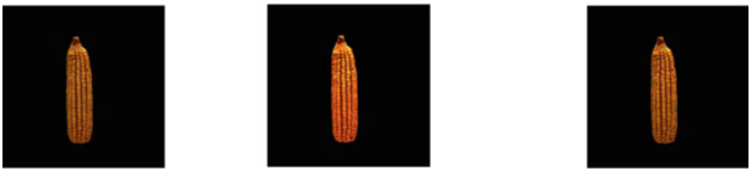
15.2 Methods

15.2.1 Data Preparation

The work of data preparation consists of three stages: data collection, data labeling, and data augment.

In the data collection stage, the color CCD camera manufactured by the Imaging Source Company in Germany is used to collect images of different varieties of maize ear. The collected images are saved in.tif format, and the image size is 1280 pixels × 960 pixels. To meet the requirements of the U-Net network on the input image size, the image size is adjusted to 572 × 572 by supplementing background pixels and scaling. A total of 150 images are collected, and the train set and the test set are divided into 2:1.

As a supervised machine learning method, it is necessary to label the images in the training sample at first, that is, to tell the machine which category each pixel belongs to [13]. We use the Labelme tool software to mark the three object regions



a. Original image b. Contrast increases by 20% c. Contrast decreases

Fig. 15.1 Data augmentation

(background, bald tip, and kernel) of all maize ear images and they are marked as black, red, and green manually. The corresponding RGB values are [0, 0, 0], [255, 0, 0], and [0, 255, 0], respectively.

In this study, we face the problem that the dataset was too small. Data augmentation can solve the over-fitting phenomenon caused by such problems and improve the accuracy of the model. Therefore, we conduct synchronous data augmentation for the original images and labeled images in the training set. The specific operations are as follows:

(1) Rotation transformation: the sample images and the labeled images are rotated clockwise by 30° and are rotated counterclockwise by 30° with the image center as the rotation center.
(2) Flip transformation: the sample images and the labeled images are flipped left and right with the midpoint as the center and are flipped up and down with the midpoint of the y-axis coordinate as the center.
(3) Shift transformation: the sample images and the labeled images are shifted to the left by 20% and are shifted to the right by 20% in the horizontal direction.
(4) Contrast transformation: the contrast of the sample images and labeled images are increased by 20% and decreased by 20%, respectively.
(5) Noise disturbance: two percent salt-and-pepper noise and Gaussian noise are added to each pixel in the images randomly.

In the training set, the number of images increased to 11 times the original images after the data augmentation. Two kinds of contrast transformation are shown in Fig. 15.1.

15.2.2 The U-Net Model

The U-Net network was proposed by Olaf Ronneberger at the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention [14]. This network is widely applied in the field of image segmentation because of its concise segmentation logic and excellent segmentation efficiency [15–18]. Figure 15.2 shows the structure of the U-Net network used in this paper. There are two main characteristics in the structure, the U-shaped symmetrical structure and

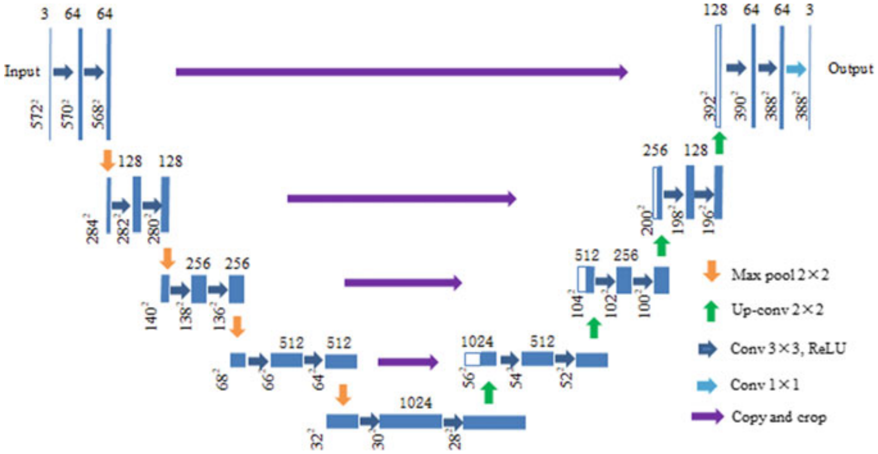


Fig. 15.2 U-Net structure

the skip connection. The U-Net network is symmetrical, an encoding process on the left implements down-sampling from top to bottom to extract features of images at different scales, and a decoding process on the right implements up-sampling from bottom to top to recover feature details acquired by the down-sampling process gradually [19]. And the skip connection including copy and crop operations is introduced between the left and the right sides to reduce the loss of spatial information caused by the down-sampling process. In this research, the input images are 572 × 572 in size and the output images are 388 × 388 in size.

15.3 Results and Analysis

A small graphics workstation is used as the platform for model training and testing. In hardware, the workstation has a CPU with I9 10900X, 32 GB main memory, and GPU with NVIDIA GeForce RTX 2080 Ti. In software, the operating system Ubuntu 18.04.1, the parallel computing environment CUDA 10.2, the deep learning framework Tensorflow 1.12, and the integrated developing environment for Python are installed.

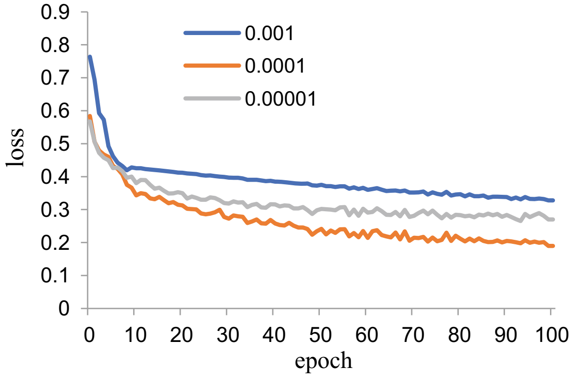


Fig. 15.3 Effects of learning rate on training loss

15.3.1 Learning Rate

Learning rate is an important hyperparameter in the training process of the U-Net network, which determines the moving step on weight in the gradient direction in a mini-batch. It directly affects the convergence speed of the model and the best accuracy the model can reach [20, 21]. In this paper, we compare the influence of three different learning rates on the model performance. The results are shown in Fig. 15.3.

Figure 15.3 shows that the model with a learning rate of 0.001 converges faster than the other two modes. However, its loss value drops down slightly after the tenth epoch because the parameters may fluctuate back and forth on both sides of the local optimal value. Obviously, the model with a learning rate of 0.0001 keeps a stable decline in loss value with the increase of training epochs.

15.3.2 Convolution Kernel Size

The convolution kernel is a weight matrix, which indicates how to deal with the relationship between a single pixel and its neighboring pixels.

We construct three modes with different size of convolution kernels. They are 3 × 3, 5 × 5, and 7 × 7, respectively. And the training results are shown in Table 15.1. From Table15.1, we can see that the model with 3 × 3 kernel size has the highest accuracy and the model with 7 × 7 has the lowest accuracy.

Table 15.1 Accuracies of modes with different convolution kernel sizes

Kernel size	Accuracy (%)	Training time (s)
3 × 3	91.90	72.35
5 × 5	85.35	106.89
7 × 7	78.62	156.23

The effect of using the superposition of multiple small convolution kernels is much better than using one large convolution kernel alone. Also, small convolution kernels can reduce the parameter number and calculation complexity. Therefore, the model with 3 × 3 kernel size runs faster than the other two models.

15.3.3 Pooling Method

After obtaining features through the convolution layers, the next step is to integrate and classify these features. All the features extracted by convolution can be used as the input of the classifier, which will bring a huge amount of computation. The pooling layer plays an important role in reducing the dimension of the feature map to solve the problem. The common pooling methods include max pooling and average pooling. The Max pooling method is to calculate the maximum value of the current element and the elements in its adjacent matrix area and then take this maximum value as the value of the current element. The average pooling method is to calculate the average value of the current element and elements in its adjacent matrix area and then take this value as the value of the current element. Pooling operation will not change the depth of the data matrix but will lead to a decrease in height and width of the data matrix to achieve the goal of dimensionality reduction.

In this paper, max pooling and average pooling are adopted to train two different models (see Fig. 15.4).

Figure 15.4 shows that the accuracy of the model with max pooling is lower than the model with average pooling in the early stage of the training process, while the former surpasses the latter after about ten training epochs, and finally reaches 93.86%.

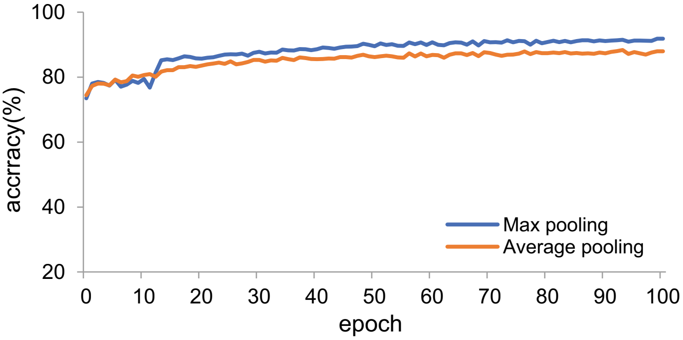


Fig. 15.4 Effects of pooling method on training accuracy

15.4 Conclusion

A segmentation model for the maize image is proposed based on the U-Net deep learning network in this paper.

(1) The loss value change during the training process are compared between three models with different learning rates of 0.001, 0.0001, and 0.00001, and the model with the learning rate of 0.0001 has better performance in terms of convergence speed and loss value.

(2) We discuss the effects of convolution kernels of different sizes (3 × 3, 5 × 5, and 7 × 7) on the performance of the model. The experimental results show that the most suitable convolution kernel size for the maize ear image dataset is 3 × 3.

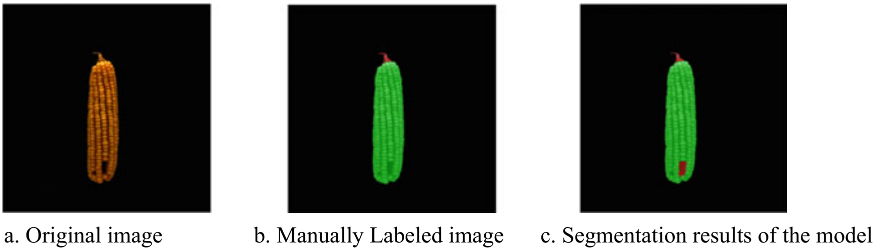
(3) Different pooling approaches including max pooling and average pooling are adopted to train two models, and the model with max pooling obtains a higher accuracy than the model with average pooling.

15.5 Discussion and Future Work

In the process of maize harvesting and transportation, some kernels will fall down from the ear. Figure 15.5a shows a particular sample which has a large part of connected region and a small part of connected region without kernels. As we know, the bald tip is defined as the region which has maldeveloped grains, resulting from poor pollination and fruiting during the growth and development of maize. Therefore, the above parts lacking kernels were labeled as the kernel class when we labeled the region of three classes for this sample image in Labelme software.

However, the model gave different answers for the two parts without kernels. Figure 15.5c shows that the model classified the small part without kernels to the bald tip class, and identified the large part without kernels as bald tip class.

It can be concluded that the model has the context ability, but the context ability is not enough. Therefore, how to improve the context ability will be great challenge in future work.



a. Original image b. Manually Labeled image c. Segmentation results of the model

Fig. 15.5 An incorrect segmentation sample

Acknowledgements This work is supported by the Fundamental Research Funds for the Central Universities under Grant No. 2662020XXPY07.

References

1. Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
2. Liu, Y.Y., Sun, J.H., Zhang, S.J., Yu, H.Y., Wang, Y.Y.: Detection of straw coverage based on multi-threshold and multi-target UAV image segmentation optimization algorithm. Trans. Chin. Soc. Agric. Eng. **36**(20), 134–143 (2020)
3. Ma, J.L., Deng, Y.Y., Ma, Z.P.: Review of deep learning segmentation methods for CT images of liver tumors. J. Image Graph. **25**(10), 2024–2046 (2020)
4. Alia, O.M., Mandava, R., Aziz, M.E.: A hybrid harmony search algorithm for MRI brain segmentation. Evol. Intel. **4**(1), 31–49 (2011)
5. Wu, Q.P., Wu, C.M.: Fast robust kernel-based fuzzy C-means clustering segmentation. J. Image Graph. **23**(12), 1838–1851 (2018)
6. Gu, Z.W., Cheng, J., Fu, H.Z., Zhou, K., Hao, H.Y., Zhao, Y.T., Zhang, T.Y., Gao, S.H., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE Trans. Med. Imaging **38**(10), 2281–2292 (2019)
7. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
8. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
9. Yang, A.Q., Huang, H.S., Zhu, X.M., Yang, X.F., Chen, P.F., Li, S.M., Xue, Y.J.: Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features. Biosys. Eng. **175**, 133–145 (2018)
10. Hansen, M.F., Smith, M.L., Smith, L.N., Salter, M.G., Baxter, E.M., Farish, M., Grieve, B.: Towards on-farm pig face recognition using convolutional neural networks. Comput. Ind. **98**, 145–152 (2018)
11. Fuentes, A., Yoon, S., Kim, S.C., Park, D.S.: A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors **17**(9), 2022 (2017)
12. Mamatha, K.R., Singh, S., Hariprasad, S.A.: Detection and analysis of plant leaf diseases using convolutional neural network. J. Comput. Theor. Nanosci. **17**(9–10), 3899–3903 (2020)
13. Zhou, Q., Zheng, B.Y., Zhu, W.P., Latecki, L.J.: Multi-scale context for scene labeling via flexible segmentation graph. Pattern Recogn. **59**, 312–324 (2016)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241 (2015)
15. Chen, J., Han, M.N., Lian, Y., Zhang, S.: Segmentation of impurity rice grain images based on U-Net model. Trans. Chin. Soc. Agri. Eng. **36**(10), 174–180 (2020)
16. Zhang, R.R., Xia, L., Chen, L.P., Xie, C.C., Chen, M.X., Wang, W.J.: Recognition of wilt wood caused by pine wilt nematode based on U-Net network and unmanned aerial vehicle images. Trans. Chin. Soc. Agri. Eng. **36**(12), 61–68 (2020)
17. Chen,L.C., Zhu,Y., Papandreou,G., Schroff,F. and Adam,H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision, pp. 801–818 (2018)
18. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: Unet 3+: a full-scale connected unet for medical image segmentation. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 1055–1059 (2020)

19. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2017)
20. Xie, W.J., Ding, Z.C., Wang, F.H., Wei, S., Yang, D.Y.: Integrity recognition of camellia oleifera seeds based on convolutional neural network. Trans. Chin. Soc. Agri. Mach. **51**(07), 13–21 (2020)
21. Lu, Y.H., Huang, S.: Application of deep learning in identification of ID card number. Appl. Sci. Technol. **46**(01), 123–128 (2019)

Part III
Applications of Multidimensional Signal Processing

Chapter 16
Identifying People Wearing Masks in a 3D-Scene

Wenfeng Wang, Jingjing Zhang, Xiwei Liu, Bin Hu, and Zerui Meng

Abstract Now people are facing the pandemic COVID-19 and have to wear masks. This brings a problem in face recognition—occlusion problem and particularly, identifying people wearing masks in 3D-scenes is a great challenge. This study aims to develop a system for tackling this challenge. The 3D-scene is constructed with the 2D-3D coordinate transformation. For the convenience of the fusion between the virtual scene and real scene, a 3D model is achieved by Sketchup Pro. The faces and masks data are explored from the video and occluded faces recognition is achieved with the convolutional neural network.

16.1 Introduction

In the past year, the whole world has been affected by the pandemic COVID-19. Under the influences of the epidemic, people will choose to wear masks before they go out. Although wearing a mask is an efficient way to prevent the epidemic in daily life, it also brings a challenge to face recognition. For example, when checking tickets at a train station, there is dilemma—it will affect tickets inspection if people wear masks, but it is difficult for people to take off their masks to face the virus. Especially, identifying people wearing masks in 3D-scenes can be a great challenge [1, 2].

Traditionally, a 3D-scene can be constructed as follows. First, taking a picture of the scene through a sensor and then, obtaining multiple photos and shooting from as many angles as possible when conditions permit. Finally, processing these images to obtain three-dimensional video images. In recent years, with the development of

W. Wang (✉) · J. Zhang · X. Liu · Z. Meng
Shanghai Institute of Technology, Shanghai 201418, China
e-mail: wangwenfeng@nimte.ac.cn

W. Wang
Interscience Institute of Management and Technology, Bhubaneswar 752054, India

B. Hu
Changsha Normal University, Changsha 410111, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_16

Internet technology, the efficiency of building the 3D visualization was improved [3]. At the same time, 3D models have the advantage of displaying huge amounts of information, which can be utilized to identify people wearing masks in 3D-scenes.

Objectives of this study are (1) construct a 3D-scene and manually build a 3D model (2) fusion the 3D data and the 3D model to obtain the 3D visualization scene, and (3), develop a system to recognize the occluded faces in the 3D-scene [4]. In Sect. 16.2, we mainly describe our platform which was constructed in the school office. Then, in Sect. 16.3, we show the 2D face recognition and collect frames, respectively, accordingly whether wear a mask or not. In Sect. 16.3, we will carry out the 3D face recognition in the real scene and finally solve the occluded faces recognition in the real 3D-scene.

16.2 Construction of the Platform

The actual equipment installation of the experimental platform: a five-way ceiling panoramic view, two face capture cameras, one set of VR equipment, computer workstations, laptops, temperature measurement integrated prototypes, and switch sockets. The installation is as follows:

(1) The ceiling panorama is installed on the indoor roof, and the grooves are wired;
(2) VR demonstration in the entrance area, it is easy to demonstrate the area;
(3) Face capture camera is connected to system but not fixed for the convenience of subsequent development.
(4) Computer as our workstation; booting can be normal use;
(5) Integrated temperature measurement prototype, connected to system, temperature measurement as in normal use.

The hardware situation f indoor can be described as follows:

(1) Set up a complete temperature measurement development environment indoors for further development;
(2) Ceiling design can make more space for the interior to be used and prevents a lot of accidental damage;
(3) HTC Vive Pro2.0 VR headset is used for VR equipment, which allows users to use VR in a certain area and feel the VR effect;
(4) Face capture camera, normal operation; HR-IPC2143 intelligent face recognition gun-type network camera is used in face capture machine, which can provide high face recognition accuracy with low power consumption;
(5) Computer workstation for normal use; DELL 5540 mobile workstation was used;
(6) Using an eight inches dual visual temperature measuring living face recognition machine.

Fig. 16.1 Original data: saved face images of every frames

16.3 Identifying People Wearing Masks

16.3.1 Collecting and Classifying Faces Data

At first, we just experiment on a video to recognize whether people wear a mask or not. And we have saved face images of every frame as shown in Fig. 16.1.

Then, we automatically saved faces that were without masks and with masks separately after recognition. As shown in Fig. 16.2, we can see some images especially with paper over the face which didn't save well from some relative images.

We measured the accuracy of the corresponding 2D faces with and without masks as shown in Table 16.1. It mainly reads from our prerecorded video which includes 1321 frames.

16.3.2 2D-3D Face Recognition with Mask

First of all, the camera needs to be calibrated before positioning [4]. In this case, we only need to use the camera's internal parameters and distortion parameters. We can get two-dimensional coordinates through camera identification, and then define a world coordinate system. The three-dimensional coordinates of the target point are defined, so we get the coordination of the camera [5]. Because of the relative position, we can get the coordination of the target point which is relative to the camera. Then the coordinate of the target point in the world coordinate system can be obtained by Euler Angle transformation and TF transformation. Because coordinate translation means matrix addition and subtraction, coordinate rotation means matrix multiplication. The advantage of homogeneous coordinates is through adding a dimension and expressing the addition multiplication in a formula.

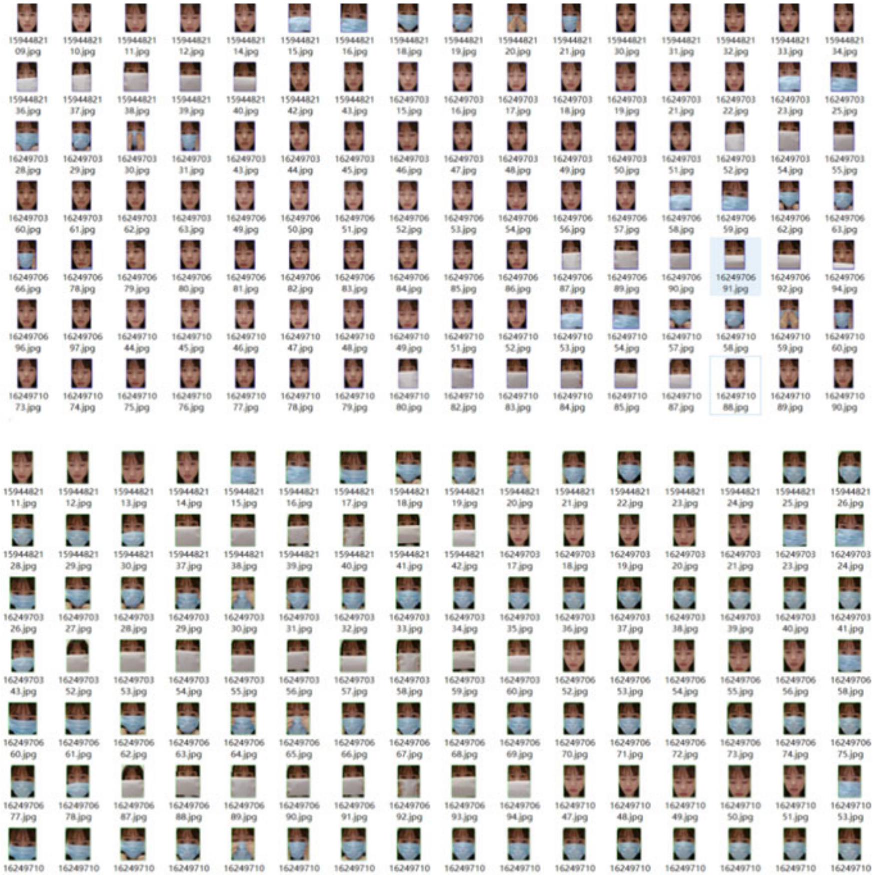


Fig. 16.2 Classified data: automatically saved images with and without masks

Table1 Accuracy of faces data classification

	Wearing mask	No wearing masks
Accuracy	0.906782247	0.973189369
Frames	720	601

$$\begin{pmatrix} \mathrm{x} \\ y \\ z \end{pmatrix} \sim \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{16.1}$$

The above formula is for coordinate transformation, $[R|t]$ is an augmented matrix which includes rotation and translation.

Euler angles define the order of rotation of objects, and the degrees they have rotated about an axis. A lot of people tend to ignore the rotation order, and a lot of books call it a rule, which can be interpreted as a rule of the rotation order of Euler angles [6]. (α, β, γ) in different rotation order will have different results, firstly rotate α about the X-axis, or rotate β about the Y axis, the final result is different. There are many rules for Euler Angle, such as Z-X-Y, X-Y-Z, X-Y-X, and Z-X-Y, which have many permutations and combinations.

In the next formula about $x = r\cos\phi$, $y = r\sin\phi$, $x' = r\cos(\theta + \phi)$, and $y' = r\sin(\theta + \phi)$, putting x' and y' into the x and y, we can get the matrix form:

$$\begin{bmatrix} \mathrm{x}' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ y \end{bmatrix} \tag{16.2}$$

We're going to scale it up, so the final form is as follows.

Rotating about the X-axis:

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{16.3}$$

Rotating about the Y-axis:

$$R_{\mathrm{y}}(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{16.4}$$

Rotating about the Z-axis:

$$R_{\mathrm{z}}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{16.5}$$

We need to find the rotation matrices of each axis, and multiply them in order to get the whole rotation matrix. Since the rotation matrix is left-multiplied, rotation matrix [7] is $R{=}R\mathrm{x}RyRz$ for the Y-X-Z Euler angle. If it's a Z-Y-X regular Euler angle, the corresponding combined rotation matrix is $R{=}R\mathrm{x}RyRz$. R is a 3 × 3 matrix, the rotation matrix for the entire transformation of coordinates. We can figure out the angle by using the inverse trigonometric function:

$$\theta_{\mathrm{x}} = \arctan\frac{r_{32}}{r_{33}},\ \theta_{\mathrm{y}} = \arctan\frac{-r_{31}}{\sqrt{r_{32}^2 + r_{33}^2}},\ \theta_{\mathrm{z}} = \arctan\frac{r_{21}}{r_{11}} \tag{16.6}$$

16.3.3 2D Alignment of the Fundamental Ratio

It is easy to verify the relative translation between frames in the scene, which can be obtained as follows:

$$E_{\mathrm{i,j}} \sim [t_{ij}]_X R_{i,j} \sim E'_{c(i),c(j)} \tag{16.7}$$

where $\sim$ means multiplication which is equal to non-zero scale factor, $[]_x$ is the symbol of skew symmetric matrix and represents the cross product. Therefore, when the camera calibration matrices K and K' of sequence V and V' are the same, the corresponding uncalibrated matrix $F_{\mathrm{i},j}$ with $F'_{c(i),c(j)}$ should be equal and can be used for video synchronization [8]. $R_{i,j}$ is the rotation matrix.

But in the more common case, K and K' are constants, but they are different in sequences.

In the case of a simplified camera model, such as unit aspect ratio and zero deviation, it can be verified that:

$$F^{2\times 2} \sim \begin{bmatrix} \in_{1\mathrm{st}}\ t^s_{i,j} r^t_1 & \in_{1\mathrm{st}}\ t^s_{i,j} r^t_2 \\ \in_{2\mathrm{st}}\ t^s_{i,j} r^t_1 & \in_{2\mathrm{st}}\ t^s_{i,j} r^t_2 \end{bmatrix} \tag{16.8}$$

where $\in_{\mathrm{rst}}$ for r, s, t = 1,2, …,3 means permutation tensor, r_i means columns of $R_{i,j}$. t_j means camera translation. It is worth noting that $F^{2\times 2}$ is the state of observation, its elements are $F_{i,j}$. The ratios have nothing to do with the internal parameters of the camera, and only reflect the self-motion of the camera. In this article, we call these fundamental ratios [9]. Therefore, we can extract an independent four-dimensional feature V_{f}:

$$V_{\mathrm{f}} = sig(F_{11})[F_{11}, F_{12}, F_{21}, F_{22}]/||F^{2\times 2}||_F \tag{16.9}$$

among them $\|\cdot\|_F$ is Frobenius norm. When two cameras are associated with a similar transformation matrix H_{s}, we can prove that two cameras have the same motion trajectory. Under this premise, they can be aligned. Proportional ambiguity H_{s} is defined as $H\begin{pmatrix} 0.8I & 0 \\ 0 & 1 \end{pmatrix}$. The position and posture of the camera are related by the proportional ambiguity, but there exists noise pollution. We calculate the camera pair which corresponds to each different position.

V_f has five freedom degrees: rotation $R_{\mathrm{i,j}}$ and $\mathrm{t}_{i,j}$. There are three freedom degrees, but there is a fuzzy scale. In addition, for the same basic matrix, there are four possible settings for the relative camera position and orientation. In fact, based on the proposed method has not available in some situations, such as pure translation, where the camera center is fixed (that is, there is no change in camera position) or the basic ratio [10] is calculated in a flat scene. However, as shown in this paper, similar camera self-motion will produce the same V_f. This can be used to synchronize video sequences.

In the process of calculating the basic ratio, SIFT features [11] are used to complete the correspondence between the initial frames and the frame, and the MAPSAC algorithm that minimizes reprojection is used to calculate the purely rotated planar homography and the basic matrix error of general camera motion [12]. The eigenvalue decomposition of the matrix is used to calculate the outer pole of a pure translation straight line. In the two frames, i and l may be far apart. Therefore, when there is no correspondence between the calculation of the basic matrix, the observation graph theory is used to calculate the basic matrix between two frames [13]. In other words, three views (i, j, k) with (j, k, l). The basic matrix inside is available.

Finally, in order to improve the robustness of the proposed method, we use a coarse-to-fine framework, because the coarse-level synchronization captures global features, so errors will not propagate to the rest of the regular path in the frame correspondence calculation [14].

The calibration error in the time axis model is used:

$$E(j, j') = \mathrm{dist}(j, j') + \min\{E(j, j' - 1), E(j - 1, j' - 1), E(j - 1, j')\} \tag{16.10}$$

among them, the dist $E(j, j)$, the mean square error between them is calculated.

Therefore, by a set of parameters $c(j)$ the determined synchronization calculation is as follows:

$$\mathrm{c(j)} = \arg \min_{c(1)\leq\cdots\leq c(N)} \sum_{\mathrm{j}=1}^{N} E(\mathrm{j}, c(\mathrm{j})) \tag{16.11}$$

where N means the number of input video frames, and then the dynamic program is used to solve the optimization problem which is defined in the equation.

16.4 3D Face Recognition in the Real Scene

16.4.1 Video Image Format Conversion

Many video images are usually in YUV format [15] obtained from the camera, but we only use process images in RGB format in the PC. Here we need to convert YUV image format to RGB format. But YUV and RGB are two different color decoding schemes. In YUV, Y is brightness, Chroma is represented by U and V, which are used to specify the color of pixels in the acquired image, and described the saturation and color of the image. Therefore, if a picture only has Y channel data, it can still display the complete picture, but the picture is black and white. And we can convert an image in YUV format to an image in RGB format through the following formula, as it has mentioned in Keith Jack's book [16]:

$$\begin{aligned} B &= 1.164(Y - R) + 2.018(U - 128) \\ G &= 1.164(Y - 16) - 0.813(V - 128) - 0.391(U - 128) \\ R &= 1.164(Y - 16) + 1.596(V - 128) \end{aligned} \tag{16.12}$$

Noting in the above formula, the range for RGB is [0, 255], the range for Y is [16, 235], and the range for UV is [16, 239]. If the result is out of this range, the processing is truncated.

It is the simplest and most direct way to convert YUV format into RGB format. By accessing each pixel in the image pixel by pixel, the conversion from YUV to RGB format image can be completed.

16.4.2 The 3D-Visualized Face Recognition

Through panoramic camera to obtain panoramic video, we need to built a three-dimensional model and fuse with it, and then a three-dimensional visualization scene can be established for face recognition. SketchUp is used as a design tool oriented to the creation process of design schemes for 3D architectural design. The program runs, as shown in Fig. 16.3. First, draw a floor plan of the room, which can be done using the Line Tool. Then, using the push-pull tool to build a preliminary

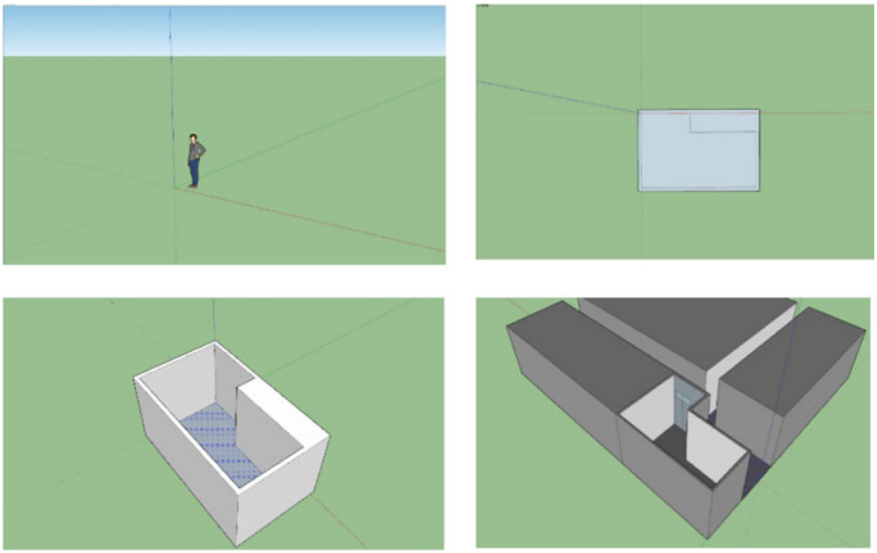


Fig. 16.3 The process for 3D-visualized face recognition

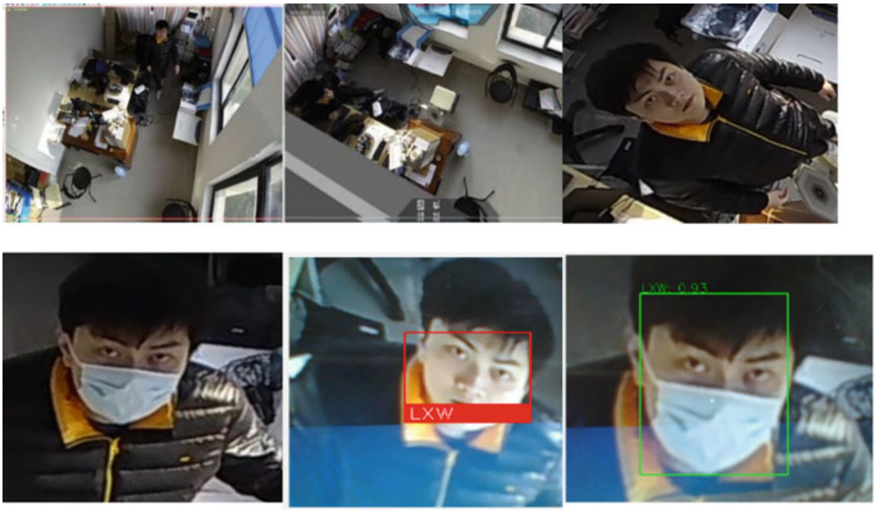


Fig. 16.4 The actually experimental results

three-dimensional model. Adding the hand-built model from SketchUp Pro to the folder where the Holographic Camera software is located to perform the 3D fusion [17, 18]. Finally, the three-dimensional model is optimized to get the following effect pictures. Meanwhile, performing 3D fusion, we must obtain images through a holographic camera as shown in Fig. 16.3. Then by fusing the image obtained by the holographic camera with the 3D model, the image after the virtual and real fusion can be obtained [19]. Based on the above three-dimensional fusion process, we can expand face recognition in the three-dimensional visualization scene. First, we obtain the face image and then achieve 3D fusion.

A homogeneous representation of camera coordinates to video image coordinates, there we use Binocular camera and we have shown the 2D-3D transformation above [20], as shown in Figs. 16.3 and 16.4.

In this way, the occluded face recognition in the 3D visualization scene is completed. From the above results, it can be seen that the establishment of a three-dimensional visualization scene has a positive effect on the recognition of occluded faces, which can assist the recognition of occluded faces. In this paper, we have collected many faces which include wearing a mask and no mask as in our face library and it helps us to further achieve 3D face recognition.

16.5 Conclusions and Perspectives

This paper mainly proposes to apply 3D technology to recognize people and use 2D scene to collect the face. And we will further statistic the accuracy of the standard face library. Through simple attempt, we solve the occluded faces problem that are difficult to recognize. During the COVID-19 pandemic, we compared contact authentication such as fingerprints, contactless face recognition authentication which has become an important tool during the May 1st Conference. The most important thing of face recognition is the biological information of the face, including facial contour, position of the nose and mouth, etc. The more feature information, the more accurate of face recognition results. However, during the epidemic, people wear masks in and out of public places, which greatly affects the accuracy of face recognition and two-dimensional occluded face recognition. After combining 3D video face recognition technology, through the recognition of face images and video, the face images in the 2D scene are obtained and we get the face library. With the help of more facial feature information, face recognition can be easier used in lots of scenes, the accuracy of face recognition can be greatly improved eventually.

Acknowledgements This research was supported by the Shanghai High-Level Base-Building Project for Industrial Technology Innovation (1021GN204005-A06).

References

1. Brunelli, R., Poggio, T.: Face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **15**(10), 1042–1052 (1993)
2. Cheng, S.C., Su, J.Y., Hsiao, K.F., et al.: Latent semantic learning with time-series cross correlation analysis for video scene detection and classification. Multimed. Tools Appl. **75**(20), 12919–12940 (2016)
3. Monteiro, J., Cardoso, J.: A cognitively-motivated framework for partial face recognition in unconstrained scenarios. Sensors **15**(1), 1903–1924 (2015)
4. Zhe, X.P., Yang, J.H., Yan, Y.X.: Research of the three-dimensional tracking and registration method based on multiobjective constraints in an AR system. Appl. Opt. **57**(32), 9625–9634 (2018)
5. Yang, Y., Xiao, F., Meng, K., et al.: Solution to render the 2D timely in the 3D visual reality. Comput. Sci. **38**(6), 279–282 (2011)
6. Wang, H., Stout, D.B., Taschereau, R., et al.: MARS: a mouse atlas registration system based on a planar x-ray projector and an optical camera. Phys. Med. Biol. **57**(19), 60–63 (2012)
7. Wang, W.: The explicit expression of axis and angle of a rotation matrix. College Math. J. **52**(1), 39–44 (2021)
8. Meng, W., Gao, Y., Lu, K., et al.: View-based discriminative probabilistic modeling for 3D object retrieval and recognition. IEEE Trans. Image Process. **22**(4), 1395–1407 (2013)
9. Mohammadzade, H., Hatzinakos, D.: Iterative closest normal point for 3D face recognition. IEEE Trans. Softw. Eng. **35**(2), 381–397 (2013)
10. Wang, Y., Pan, G., Wu, Z.: A survey of 3D face recognition. J. Comput. Aided Des. Comput. Graph. **20**(7), 819–829 (2008)
11. Shah, X.M.: Tri-view morphing. Comput. Vis. Image Underst. **96**(3), 345–366 (2004)

12. Miao, S., Zhou, Y., Wei, Z.: An efficient architecture for adaptive deblocking filter of H.264/ AVC video coding. IEEE Trans. Consum. Electron. **50**(1), 292–296 (2004)
13. Lee, J.H., Lim, K.W., Song, B.C., et al.: A fast multi-resolution block matching algorithm and its LSI architecture for low bit-rate video coding. IEEE Trans. Circuits Syst. Video Technol. **11**(12), 1289–1301 (2001)
14. Jiang, C., Ding, G., Gamal, A.E., et al.: IEEE TCCN special section editorial: machine learning and artificial intelligence for the physical layer. IEEE Trans. Cogn. Commun. Netw. **7**(1), 1–4 (2021)
15. Atitallah, A.B., Kadionik, P., Ghozzi, F., et al.: An FPGA implementation of HW/SW codesign architecture for H.263 video coding. AEU Int. J. Electron. Commun. **61**(9), 605–620 (2007)
16. Eleftheriadis, A., Jacquin, A.: Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates. Signal Process. Image Commun. **7**(3), 231–248 (1995)
17. Zhou, Z., Zhou, Y., Xiao, J.J.: Review of virtual reality enhancement technology. Sci. China: Inf. Sci. **45**(2), 157–180 (2015)
18. Cao, X., Lin, W., Xiao, J., et al.: Video synchronization and its application to object transfer. Image Vis. Comput. **28**(1), 92–100 (2010)
19. Cotte, Y., Toy, M.F., Arfire, C., et al.: Realistic 3D coherent transfer function inverse filtering of complex fields. Biomed. Opt. Express **2**(8), 2216–2230 (2011)
20. Hu, L., Li, Y., Li, T., et al.: The efficiency improved scheme for secure access control of digital video distribution. Multimed. Tools Appl. **75**(20), 1–18 (2016)

Chapter 17
Contrast Enhancement and Noise Removal from Medical Images Using a Hybrid Technique

Mansi Lather and Parvinder Singh

Abstract Medical image analysis is very important for the proper and efficient diagnosis of various disorders. Detection and diagnosis of various brain disorders are very challenging because of the complex shape and structure of the brain. Among the various medical imaging tools, Magnetic Resonance Imaging (MRI) gives the most precise image of the brain structure. But, these images suffer from low contrast and are also distorted by noise. So, these images need to be pre-processed such that accurate and precise information can be extracted from them for further analysis and detection of various brain disorders. In this paper, a hybrid approach has been proposed for pre-processing, wherein the contrast of the MRI image has been enhanced using the Minimum Mean Brightness Error Bi-Histogram Equalization (MMBEBHE) approach and denoising has been done using a combination of Wiener and bilateral filter. The results of the proposed approach have been analyzed by adding speckle and Gaussian noise considering Peak Signal to Noise Ratio (PSNR), Root Mean Square (RMS) Contrast, Structural Similarity Index Measure (SSIM), Signal to Noise Ratio (SNR), and Normalized Correlation (NC) as performance parameters.

17.1 Introduction

Every human being wants to live a healthy life. But, in our lifetime, unfortunately, we are generally infected with one or the other illness. Among the various disorders in the human body, the most dangerous and deadly disorder is the brain tumor. The brain tumor is the anomalous and unusual growth of cells present in the brain tissues. These tumors vary from each other depending on their source and position in the body. On the basis of origin, brain tumors are broadly classified into two types, one is called as primary brain tumors and the other one is called secondary or metastatic brain tumors [1]. Primary brain tumors are the tumors that are generated in the brain

M. Lather (✉) · P. Singh
Department of Computer Science and Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonepat 131039, India
e-mail: mansi.schcse@dcrustm.org

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_17

tissues while the secondary or metastatic brain tumors are the tumors that are firstly generated in other parts of the body which are affected by cancer already like the colon, skin, lungs, etc. and then these infected tumor cells move to the brain and increased there. Generally, the metastases brain tumors are caused by lung cancer, melanoma, breast cancer, and other cancers [2].

Almost 11,000 cases of brain tumors are being diagnosed every year [3]. It is important to identify and diagnose these tumors as early as possible so that the survival period of humans can increase. Disease detection and diagnosis have become easy because of recent advances in technology. Digital image processing has become an integral part of everyone's life and is playing important role in our daily life applications. Medical image processing is very useful for extracting medical information from medical images. This information can then be analyzed to extract the needed attributes which can then be used by the machine to come up with the proper diagnosis [4]. Different medical imaging tools such as X-ray, Computed Tomography (CT) scan, MRI, etc. are used to get the medical scans of the human body [5]. These scans are then utilized to detect the various disorders and malfunctioning inside the human body. Among these scans, an MRI scan gives better details of the inside structure of the brain [6].

MRI is a powerful visualization tool that allows internal anatomy images to be safely and non-invasively acquired [7]. MRI scans are produced under an MRI scanner by utilizing the magnetic, gradient, and radiofrequency coils. MRI produces four different types of MRI sequences T1, T2, Flair, and T1c [8]. MRI scan suffers from low contrast and is also affected by noise, thus hindering the accurate detection and diagnosis of the lesion region [9]. Therefore, these MRI images need to be pre-processed for accurate lesion detection and further analysis so as to come up with the correct timely diagnosis of the disorder.

Pre-processing of images improves the quality of original images making them more suitable for further analysis such as segmentation or classification of brain tumors. These pre-processed images extract the high-value features from the images and filters out the low-value features, thus improving the segmentation or classification algorithm results [10]. As part of pre-processing, we can perform denoising, contrast enhancement, skull stripping, edge preservation, image restoration, etc. In [11], as part of pre-processing, denoising is performed using median, unsharp, and Wiener filters. Skull stripping is done using the Brain Extraction Method (BET). In [12], pre-processing phase involves denoising. Noise is removed from the MRI images by using median and Wiener filters and then fusing the obtained filtered images. In [13], denoising is performed using a combination of bilateral and anisotropic diffusion filtering approaches. In [9], as part of pre-processing denoising and brain surface extraction is done. Noise is removed using an adaptive Wiener filter, and morphological operators are used to remove the non-brain tissues.

In this paper, a hybrid approach for pre-processing of MRI images has been proposed, wherein the contrast of the MRI image has been enhanced using the Minimum Mean Brightness Error Bi-Histogram Equalization (MMBEBHE) approach and denoising has been done using a combination of Wiener and bilateral filter.

The rest of the paper is organized as follows: Sect. 17.2 describes the proposed hybrid pre-processing approach; Sect. 17.3 describes the experimental results; and Sect. 17.4 concludes the paper.

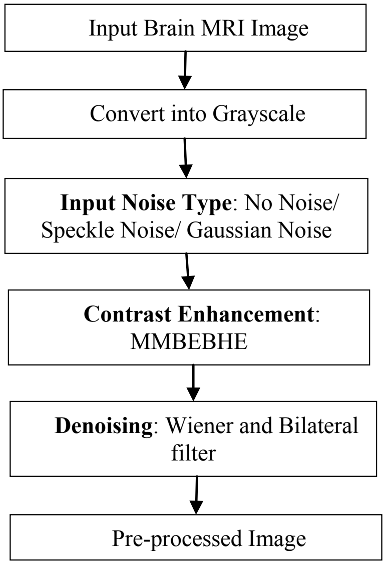
17.2 Proposed Methodology

This section describes the proposed model for pre-processing of the brain MRI images. The flowchart of the proposed approach is shown in Fig. 17.1. The input brain MRI image is first converted into grayscale; then contrast enhancement is done using the MMBEBHE approach; finally, noise is removed from the contrast-enhanced image using Wiener and bilateral filters.

17.2.1 Input MRI Image and Conversion into Grayscale

MRI image is taken as input from the BraTS 2018 dataset [14–16]. The input image is then converted into grayscale for further processing.

Fig. 17.1 Flowchart of the proposed pre-processing model


Input Brain MRI Image
Convert into Grayscale
Input Noise Type: No Noise/ Speckle Noise/ Gaussian Noise
Contrast Enhancement: MMBEBHE
Denoising: Wiener and Bilateral filter
Pre-processed Image

17.2.2 Add the Noise

In this step, we can either add no external noise or add external speckle or Gaussian noise to the original image.

17.2.3 Contrast Enhancement

MRI images usually have low contrast, making it difficult to detect the tumor area from the brain MRI images. Fundamental enhancement in contrast is needed in MRI to enhance the brightness of the image because the contrast between normal dense tissue and tumor tissue could be there, but below the threshold of human perception. In order to enhance image features against its background so as to have better visualization of image properties through an open eye, image contrast needs to be increased. In the proposed approach, contrast is enhanced using the MMBEBHE approach. MMBEBHE enhances the image contrast by modifying the histogram of an image. This approach performs separation based on the threshold level which yields minimum Absolute Mean Brightness Error (AMBE) [17].

17.2.4 Denoising

After preserving the brightness of the image and enhancing the contrast, we remove noise from the image. Noise adds distortion in medical images because noise contains irrelevant information which hampers the image authenticity and quality and misleads the actual data. Noisy medical images cannot be analyzed with accuracy and precision. Thus, it is important to remove noise from the image and at the same time preserve the contents and quality of the image. For this purpose, we denoised the image using a Wiener filter with a window size 5, and then to preserve the edge information we used a bilateral filter with a window size 5 and $\sigma_d = 3$ and $\sigma_r = 0.3$, where σ_d is the geometric spread in the domain and σ_r is the photometric spread in the range.

17.3 Experimental Results and Analysis

The proposed pre-processing approach is simulated on MATLAB 2020b, Windows 10 OS with 8 GB RAM and Intel i3 processor using BraTS 2018 dataset. Figure 17.2 shows the output of the different steps of the proposed approach when no external noise is added to the original image. Figure 17.3 shows the output of various steps when speckle noise with a variance of 0.01 is added to the original image. Figure 17.4

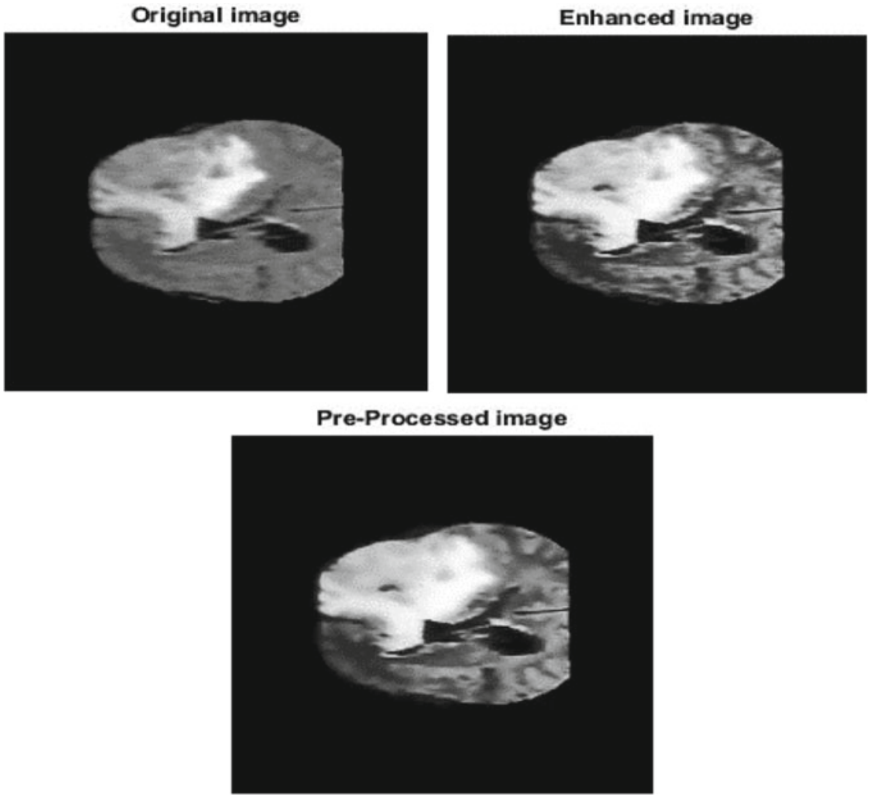


Fig. 17.2 Experimental results of the proposed pre-processing model with no input noise

shows the output results of different steps of the proposed model when Gaussian noise with variance 0.01 is added to the original image.

17.3.1 Performance Evaluation

The quality of the proposed pre-processing model is evaluated using Peak Signal to Noise Ratio (PSNR), Root Mean Square (RMS) Contrast, Signal to Noise Ratio (SNR), Structural Similarity Index Measure (SSIM), and Normalized Correlation (NC) as the performance measures.

SNR measures the ratio of signal to noise and estimates the quality of the pre-processed image compared to the original image [18]. PSNR is the measure of quality assessment between the original image and the pre-processed image. The higher the value of PSNR, the better the quality of the pre-processed image. PSNR is calculated using the formula given in Eq. 17.1, where max is the maximum value of the input

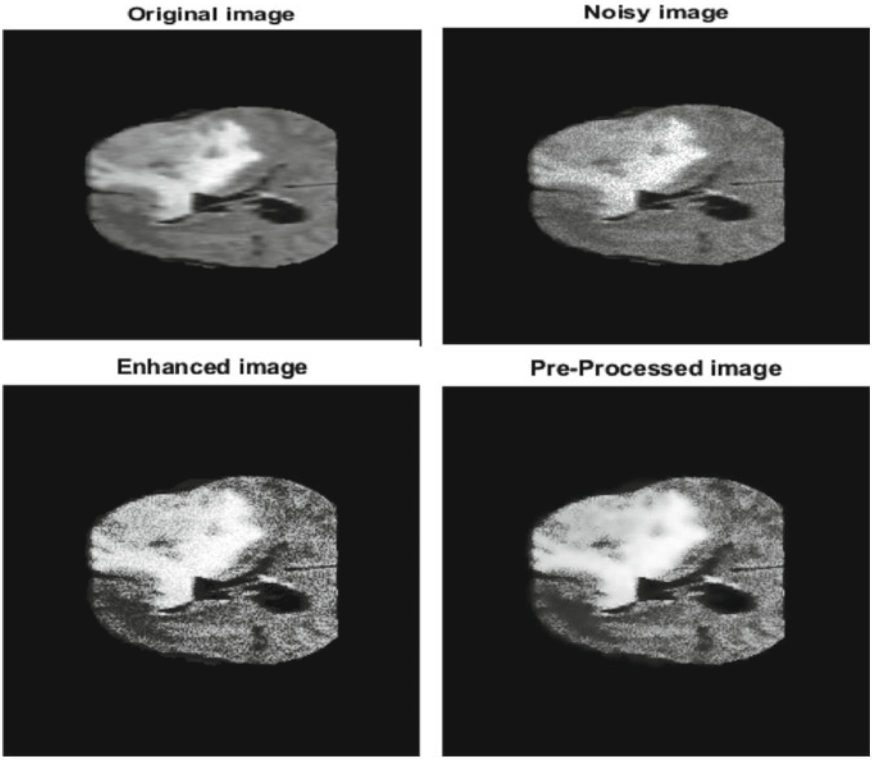


Fig. 17.3 Experimental results of the proposed pre-processing model with speckle noise

image [19].

$$\text{PSNR} = 10\log_{10}\frac{\max^2}{\text{MSE}} \tag{17.1}$$

In Eq. 17.2, MSE is the squared error between the original image and the pre-processed image and is calculated using the formula given in Eq. 17.2 [20].

$$\text{MSE} = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M \times N} \tag{17.2}$$

In the above equation, M and N are the number of rows and columns in the input image.

RMS contrast is calculated as the square root of MSE. The lower the value of RMS contrast, the better is the image quality.

SSIM is used to measure the similarity between the original image and the pre-processed image by assessing the visual impact of luminance, contrast, and structure

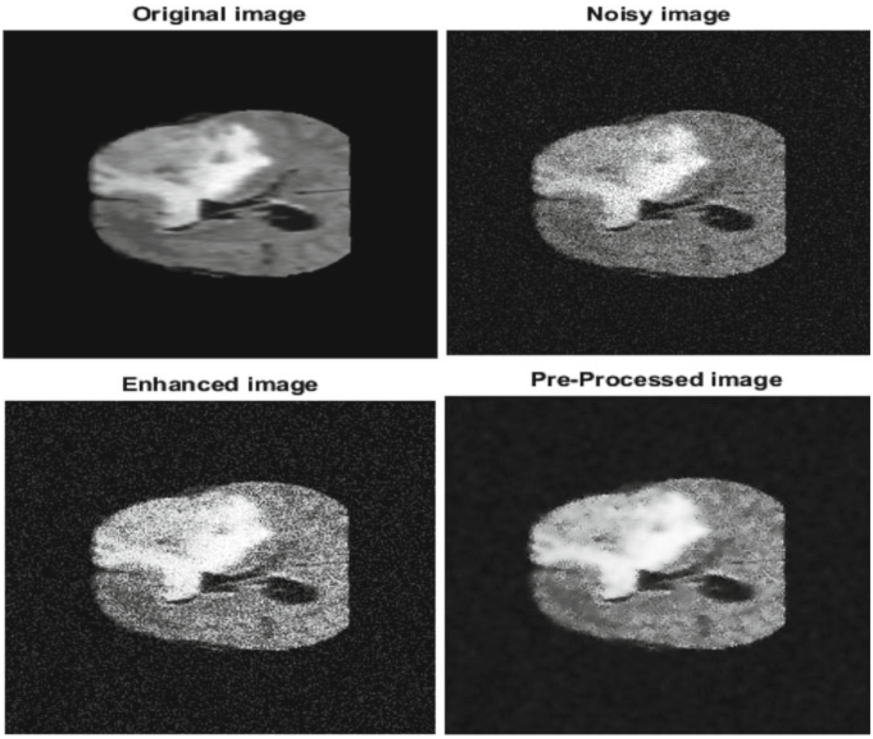


Fig. 17.4 Experimental results of the proposed pre-processing model with Gaussian noise

of the image [20]. NC is also used to measure the degree of similarity or dissimilarity between the original image and the pre-processed image [21].

The proposed pre-processing model is evaluated using the PSNR, SSIM, SNR, and NC as the performance parameters. The performance of the proposed model is analyzed by adding speckle and Gaussian noise with variance 0.01 to the original image and comparing the results of these performance parameters in the three cases. The results obtained by the proposed approach in all three cases are shown in Table 17.1.

The results of various performance parameters show that the proposed approach is very effective in pre-processing the images, thus improving their quality for further

Table 17.1 Performance parameters result of the proposed model

Noise type	SNR	PSNR	SSIM	NC
No noise	14.645	24.827	0.973	0.982
Speckle	14.504	24.743	0.951	0.980
Gaussian	12.736	22.391	0.895	0.974

Table 17.2 Comparison of the performance parameters results of the proposed model with other techniques

Techniques	Noise type	SNR	PSNR	SSIM	NC
Proposed	No Noise	13.38	23.17	0.96	0.97
(MMBEBHE + wiener + bilateral filtering)	Speckle	13.26	22.98	0.92	0.96
	Gaussian	11.16	20.31	0.87	0.95
MMBEBHE + gaussian filtering	No Noise	13.16	22.79	0.95	0.96
	Speckle	13.08	21.87	0.90	0.95
	Gaussian	10.14	20.34	0.81	0.93
MMBEBHE + median filtering	No Noise	12.87	22.50	0.95	0.96
	Speckle	13.04	21.73	0.91	0.95
	Gaussian	12.35	21.36	0.86	0.94

Table 17.3 Root mean square contrast

	MMBEBHE	BBHE	HE
RMS contrast	5.92	14.59	73.31

analysis. The proposed approach is also effective in pre-processing the images that are affected by speckle and Gaussian noise.

Table 17.2 compares the average result of the proposed approach with the other techniques with respect to all three cases of the noise type. From the table, it is clear that the proposed approach is performing better than the other two approaches in all three cases of noise type. In Table 17.3, the RMS contrast value obtained by using the MMBEBHE method in the proposed approach is compared with the other contrast enhancement techniques such as Histogram Equalization (HE) and Brightness Preserving Bi-Histogram Equalization (BBHE). The lower value of RMS contrast correlates to the better image quality. From the results of Table 17.3, it is evident that MMBEBHE results in lower value of RMS contrast. Thus, MMBEBHE provides better contrast enhancement and maximum brightness preservation as compared to HE and BBHE techniques.

17.4 Conclusion

Medical image analysis is very important for the correct and timely detection and diagnosis of various disorders in the human body. Medical image processing plays a very important role in medical image analysis. Medical imaging such as MRI scan provides precise details of the internal anatomy of human body organs. But, these images suffer from low contrast and noise effects. Thus, these medical images need to be pre-processed for further analysis. In this paper, a hybrid pre-processing approach is proposed which enhances the image contrast and preserves the brightness using the

MMBEBHE approach and then removes the noise effect from the image using Wiener and bilateral filters. The results of the proposed hybrid approach are analyzed using RMS contrast SNR, PSNR, NC, and SSIM as performance evaluation measures. The effectiveness of the proposed model is also analyzed by adding external speckle and Gaussian noise. The results of the proposed approach are compared with the other techniques for all three cases of the noise type, and it outperforms the other techniques in all three cases.

As part of a future study, we will extend this work to segment and classify the tumor region from the brain MRI images.

Acknowledgements This work is part of a bilateral Indian-Bulgarian cooperation research project between Technical University of Sofia, Bulgaria and Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonepat, 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. Venu, K., Natesan, P., Sasipriyaa, N., Poorani, S.: Review on brain tumor segmentation methods using convolution neural network for MRI images. In: 2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW), pp. 291–295 https://doi.org/10.1109/I2C2SW45816.2018.8997387 (2018)
2. Cherguif, H., Riffi, J., Mahraz, M.A., Yahyaouy, A., Tairi, H.: Brain tumor segmentation based on deep learning. In: 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), pp. 1–8. IEEE, Taza, Morocco. https://doi.org/10.1109/ISACS48493.2019.9068878 (2019)
3. What is a brain tumour? The Brain Tumour Charity. https://www.thebraintumourcharity.org/brain-tumour-diagnosis-treatment/how-brain-tumours-are-diagnosed/brain-tumour-biology/what-is-a-brain-tumour/. Last accessed 19 May 2021
4. Lather, M., Singh, P.: Image processing: what, how and future. In: Singh, V., Asari, V.K., Kumar, S., Patel, R.B. (eds.) Computational Methods and Data Engineering, pp. 305–317. Springer Singapore, Singapore.https://doi.org/10.1007/978-981-15-6876-3_23 (2021)
5. Lather, M., Singh, P.: Investigating brain tumor segmentation and detection techniques. Procedia Comput. Sci. **167**, 121–130 (2020). https://doi.org/10.1016/j.procs.2020.03.189
6. Brain Tumors—Classifications, Symptoms, Diagnosis and Treatments. https://www.aans.org/. Last accessed 19 May 2021
7. Gopal, N., Karnan, M.: Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques. In: Presented at the 2010 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2010 January 29 https://doi.org/10.1109/ICCIC.2010.5705890 (2011)
8. Sravan, V., Swaraja, K., Meenakshi, K., Kora, P., Samson, M.: magnetic resonance images based brain tumor segmentation—a critical survey. In: 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), pp. 1063–1068. https://doi.org/10.1109/ICOEI48184.2020.9143045 (2020)
9. Zhang, C., Shen, X., Cheng, H., Qian, Q.: Brain tumor segmentation based on hybrid clustering and morphological operations. Int. J. Biomed. Imaging **2019**, 1–11 (2019). https://doi.org/10.1155/2019/7305832

10. Tahir, B., Iqbal, S., Usman Ghani Khan, M., Saba, T., Mehmood, Z., Anjum, A., Mahmood, T.: Feature enhancement framework for brain tumor segmentation and classification. Microsc Res Tech. **82**, 803–811. https://doi.org/10.1002/jemt.23224 (2019)
11. Chatterjee, S., Das, A.: A novel systematic approach to diagnose brain tumor using integrated type-II fuzzy logic and ANFIS (adaptive neuro-fuzzy inference system) model. Soft. Comput. **24**, 11731–11754 (2020). https://doi.org/10.1007/s00500-019-04635-7
12. Min, A., Kyu, Z.: MRI images enhancement and brain tumor segmentation. Adv. Sci. Technol. Eng. Syst. J. **3**. https://doi.org/10.25046/aj030642 (2018)
13. Anchal, A., Budhiraja, S., Goyal, B., Dogra, A., Agrawal, S.: An efficient image denoising scheme for higher noise levels using spatial domain filters. Biomed. Pharmacol. J. **11**, 625–634. https://doi.org/10.13005/bpj/1415 (2018)
14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging. **34**, 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
15. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data. **4**,(2017). https://doi.org/10.1038/sdata.2017.117
16. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al.: Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. arXiv:1811.02629 [cs, stat] (2019)
17. Chen, S.D., Ramli, A.: Minimum mean brightness error Bi-histogram equalization in contrast enhancement. Consum. Electron. IEEE Trans. **49**, 1310–1319 (2003). https://doi.org/10.1109/TCE.2003.1261234
18. Al-Najjar, Y., Chen, S.D.: Comparison of image quality assessment: PSNR, HVS, SSIM, UIQI. Int. J. Sci. Eng. Res. **3**, 1–5 (2012)
19. Compute peak signal-to-noise ratio (PSNR) between images—Simulink—MathWorks India. https://in.mathworks.com/help/vision/ref/psnr.html. Accessed 20 May 2021
20. Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. J. Comput. Commun. **7**, 8–18 (2019). https://doi.org/10.4236/jcc.2019.73002
21. Rao, Y.: Application of normalized cross correlation to image registration. Int. J. Res. Eng. Technol. **03**, 12–16 (2014). https://doi.org/10.15623/ijret.2014.0317003

Chapter 18
Application of Computer Image Recognition Technology in Ship Monitoring Direction

Renqiang Wang, Keyin Miao, Hua Deng, and Jianming Sun

Abstract With the development and utilization of the ocean, the ocean has become more and more important for human beings. Therefore, the implementation of marine ship detection plays a vital role in the national economic development, maritime traffic safety, rational use of marine resources and territorial sea safety. As a new technology, ship recognition based on optical satellite image provides a new technical means for long-distance marine dynamic monitoring system. Compared with the traditional detection system, it has many obvious advantages, such as not limited by region and time, long-distance and large-scale dynamic monitoring, short monitoring cycle, high recognition rate and so on. Therefore, it is very important to recognize and segment satellite ship image quickly, efficiently and accurately. In this paper, the u-net network is used to recognize the satellite ship image. The accuracy rate of the verification set is 94.30%, and the recall rate is 92.03%; the accuracy rate of the test set is 95.23%, and the recall rate is 94.01%. It can do a good job of ship identification and monitoring.

18.1 Introduction

Nowadays, in the coastal areas of China, port construction and water transportation are becoming increasingly busy, ship traffic flow is increasing rapidly, and water transportation conditions are more complex, which brings many hidden dangers to ship navigation and the safety of life and property at sea. The maritime traffic accidents of passenger ships occurred in Bohai Bay for several consecutive years. Strengthening the safety management of coastal waters, orderly organization of ship navigation, and coordinated command of search and rescue are the urgent problems to be solved to improve the safety management of coastal ships [1].

The integration of embedded technology, network, image processing technology and wireless transmission technology in modern science and technology makes

R. Wang (✉) · K. Miao · H. Deng · J. Sun
Navigation College, Jiangsu Maritime Institute, Nanjing 2111709, China
e-mail: wangrenqiang2009@126.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_18

the ship video remote monitoring technology have a great development. Using the existing wireless communication network (such as GPRS and CDMA) to realize the wireless transmission of video data from the ship video source to the monitoring center, the monitoring range can reach any position covered by the network, and there is no need to lay a special line at the monitoring object. With the help of China Unicom CDMA network covering the Bohai sea route, all ship video signals and navigation information are sent to the monitoring center through cdma20001x network and the data of the monitoring center can be received, so as to provide information processing guarantee and support for mobile office and on-site law enforcement, and improve the existing maritime network step by step extension [2]. The shore-based offshore vessel dynamic monitoring system is mainly composed of a vessel traffic management system (VTS), an automatic identification system (AIS), and a video surveillance system (CCTV). However, it is difficult to achieve effective long-distance coverage when the VTS has a limit ranging of 20 nautical miles; at the same time, its echo signal is easily attenuated by sea conditions. The video monitoring system (CCTV) can realize real-time video monitoring. Once a ship is found to be overloaded or in violation of the navigation rules, it can immediately notify the relevant personnel in the video monitoring area for law enforcement. In the case of bad weather conditions, the monitoring capacity will be greatly affected, and the monitoring range is limited by coastal waters, which cannot meet the needs of long-distance and large-scale monitoring [3].

In the aspect of image recognition, according to the characteristics of satellite ship image data, the RIRNet network model is designed in paper based on convolutional neural network, and carries out experiments on satellite ship data set, compares with a variety of neural network models, and analyzes and summarizes the results. In the aspect of semantic segmentation, the u-net network is used to segment the satellite ship image. On this basis, the u-net is improved by combining with the self-designed RIRNet network model.

18.2 Image Semantic Segmentation

Image semantic segmentation is a typical computer vision problem. Its purpose is to take the original image data as input and convert the target area to be recognized into a mask. It is the fact that the problems in early computer vision only found the elements such as edges (straight lines and curves) or gradients, and the pixel-level images are not to understand. Image classification and target detection are also two common computer vision problems [4]. Image classification is to identify different targets in the image; target detection is to use bounding boxes to identify the target and its position in the image; semantic segmentation is to classify each pixel to identify the target and target position in the image [5]. The application of semantic segmentation scenes involves geographic information systems, unmanned driving, medical image analysis, precision agriculture, etc. [6].

18.2.1 Semantic Segmentation Data Set

Pascal part [7]: this data set is developed from Pascal-voc 2010 data set. All categories of the original data set remain unchanged, and all images and image annotations in the original data set are retained. On this basis, the categories of the original data set are subdivided, such as bicycles are divided into wheels, seats, lights, armrests, sprockets, and other parts; cars are subdivided into front doors, back doors, windows, front wheels, rear wheels, roof, hood, mirrors, and other parts; animals are subdivided into head, ears, mouth, forelimbs, hind legs, tail, body, and other parts.

18.2.2 Evaluation Criterion

Image recognition and evaluation generally use two overall image indicators, accuracy rate and recall rate. However, the evaluation index of semantic segmentation is mostly at the pixel level, usually using the variation of pixel accuracy and intersection and union ratio. Pixel accuracy (PA) is the most intuitive indicator of semantic segmentation, which is to calculate the ratio of the number of correctly classified pixels in all categories, with the number $k + 1$, of all pixels in the image, as shown in (18.1) [8].

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{18.1}$$

where p_{ij} is the number of pixels that belong to the ith category but are classified into the jth category, and p_{ii} represents the number of positive examples that are correctly classified.

In practical application, PA cannot reflect the relationship between different categories of correctly classified pixels, and many data sets are not balanced, so the pixel accuracy cannot be well evaluated. The average pixel accuracy (MPA) is improved on the basis of PA. First calculate the pixel accuracy, and then average the total number of categories. The specific formula is shown in Fig. 18.2.

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{18.2}$$

Average intersection union ratio (MIoU): it is an improvement on the intersection union ratio (IOU), that is, to calculate the average IOU on each category. The intersection and union ratio is the ratio of the intersection and union of the predicted region and the real region. In the semantic segmentation, the ratio of the real positive example to the total number of pixels of this category is calculated, as shown in Formula 18.3.

$$MIoU = \frac{1}{k+1}\sum\nolimits_{i=0}^{k}\frac{p_{ii}}{\sum\nolimits_{j=0}^{k}p_{ij} + \sum\nolimits_{j=0}^{k}p_{ji} - p_{ii}} \tag{18.3}$$

The importance of each category in the actual data set may not be balanced. For example, the importance of each category in the unmanned driving system image is higher than that of other categories. Therefore, by weighting each category according to its importance, we can get the frequency weighted intersection union ratio (FWIoU), as shown in (18.4).

$$FWIoU = \frac{1}{\sum\nolimits_{i=0}^{k}\sum\nolimits_{j=0}^{k}p_{ij}}\sum\nolimits_{i=0}^{k}\frac{\sum\nolimits_{j=0}^{k}p_{ij}p_{ii}}{\sum\nolimits_{j=0}^{k}p_{ij} + \sum\nolimits_{j=0}^{k}p_{ji} - p_{ii}} \tag{18.4}$$

18.2.3 Ship Image Data Set

In the data set, the number of ships ranges cover 0 to 15. When the number is 0, it means the background image, and the largest number is 15. The satellite ship data set contains a total of 62,000 images, including 48,000 images without ship, accounting for 77.42%; 8730 images only contain single target, accounting for 14.08%, while the images of two or more ships account for a very small proportion. It can be seen that this data set is extremely unbalanced.

The dataset contains many different scenes, such as night, fog, snow, cloudy conditions, and floating ice on the sea. In some simple scenes, the target is relatively easy to distinguish. In fact, there are more complex scenes in the data set. The main body of the target accounts for a small proportion of the image, and there is more interference information. Even the human eye is difficult to distinguish. It is very difficult for image recognition and semantic segmentation. When extracting convolution features from smaller targets, it is easy to lose information. In the later model construction, we need to pay attention to this kind of situation. The model design is adjusted accordingly [9].

18.2.4 Recognition Method Based on Convolutional Neural Network

Convolutional neural networks generally include convolution layer, pooling layer, nonlinear activation layer, and fully connected layer. Convolutional neural network mainly realizes the invariance of displacement, scaling and deformation through local receptive field, weight sharing and down sampling. Convolution layer is the core part of convolution neural network and the key operation of feature extraction. The operation in convolution layer and convolution operation in analytical mathematics are not the same operation. It is actually cross-correlation operation. The input array is cross correlated with the convolution kernel to get the output array. The operation process of cross-correlation is to multiply the specified area of the input array and the convolution kernel from the starting position one by one in the sum, start the operation from the upper left corner of the input array, and then move the operation area from left to right and from top to bottom according to the steps, until all areas of the input array complete the operation. When the input array is filled, zero elements of corresponding size are filled around the input array according to the filling size, and then the above operation is performed.

Pooling layer is another core operation of convolutional neural network. The main function of pooling layer is to compress the features extracted by convolution, reduce the parameters of the model, and effectively reduce the size of the parameter matrix. The pooling layer is divided into maximum pooling and average pooling, that is to calculate the maximum and average values of elements in the pooling window. Consistent with the convolution operation, the pooling window starts searching from the upper left corner of the input array in a left-to-right and top-to-bottom manner. When the merge window is at any position, the maximum value of the input sub-array in the window is the element corresponding to the position in the output array. The working principle of two-dimensional average pooling is similar to two-dimensional maximum pooling, but the maximum operator is replaced by the average operator [10].

18.3 Application of Image Recognition Technology in Ship Monitoring

In order to better verify the model structure algorithm in this chapter, the results are compared with VGG-19 (3 × 3 convolution kernel, network architecture weight is 574 MB), ResNet-50 (3 × 3 convolution kernel, 50 weight layers, weight size to 102 MB), ResNet-101 (3 × 3 convolution kernel, the size of the weight is 40 MB), and Perception-v3 (3 × 3 convolution kernel, the size of the weight is 96 MB) models. Results mainly from the accuracy and recall were compared. All models adopt the same initialization method, loss function, activation function, batch size, and optimizer. The learning rate of training will be adjusted manually according to the convergence degree of different models, which will not guarantee the same learning rate of all models. For the data test results, there are the following four cases. TP: forecast positive, actual positive; FP: forecast positive, actual negative; TN: forecast negative, actual negative; FN: forecast negative, actual positive. Therefore, the formulas for calculating the True Positive Rate (TPR) and False Positive Rate (FPR) can be obtained in accordance with (18.5) and (18.6):

$$TPR = \frac{TP}{TP + FN} \tag{18.5}$$

$$FPR = \frac{FP}{FP + TN} \tag{18.6}$$

The true case rate TPR and the false positive case rate FPR are used to judge the quality of the classifier. The higher the TPR and the lower the FPR, the better the classification effect of the classifier.

18.3.1 Experimental Development Environment

The network structure designed in this chapter contains a large number of network parameters and convolution layer, which requires a large amount of calculation, and the speed of the computer with GPU will be much faster. The following will introduce the development environment in this experiment.

Processor Intel Core i7 7700hq @ 2.8GhZ;
Memory: Kingston DDR4 3200 16 g;
Graphics card: NVIDIA GeForce GTX 1080 ti;
Hard disk: Western Digital 2 TB;
Operating system: Ubuntu 16.04.3;
Deep learning framework: mxnet;
CUDA: CUDA 9.0.176;
Language: Python.

18.3.2 Data Preparation

The satellite ship data set contains 62,000 images, including 48,000 images without ship and 18,000 images with ship. Due to the imbalance of data, in order to make up for the deviation caused by the imbalance of data, 12,000 images with ship and 12,000 images without ship are selected as the training set, 3000 images with ship and 3000 images without ship are selected as the verification set. In fact, most of the images captured by satellite are non-ship images, so 8000 non-ship images and 3000 ship images are selected in the test set, which is closer to the real data.

18.3.3 Network Model Structure Classification and Identification Test of Verification Set

Recall rate indicates the probability that the target concerned is predicted correctly. This paper focuses on how many ships can be detected, so recall rate is an important index. Due to the imbalance of the satellite ship data set, the distribution of the

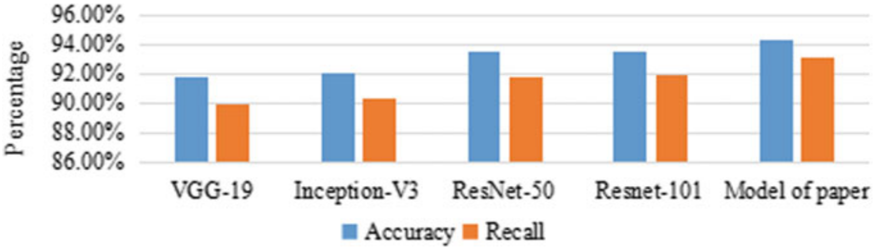


Fig. 18.1 Network models classification recognition result comparison in verification set

Table 18.1 Comparison in verification of Network models classification recognition

Models	Accuracy (%)	Recall (%)
VGG-19	91.75	89.87
Inception-V3	92.08	90.32
ResNet-50	93.44	91.71
ResNet-101	93.50	91.82
Model of paper	94.30	93.03

verification set and the test set in this chapter is different. The verification set and the training set ensure the same distribution. In order to balance the positive and negative proportion of the data, the same proportion is used. The test set is closer to the real distribution, so the accuracy and recall of the test set are more important.

According to Fig. 18.1 and Table 18.1, after testing each network model in the verification set, the test results show that the recall rate and accuracy rate of this model are the highest, 93.03 and 94.30%, respectively.

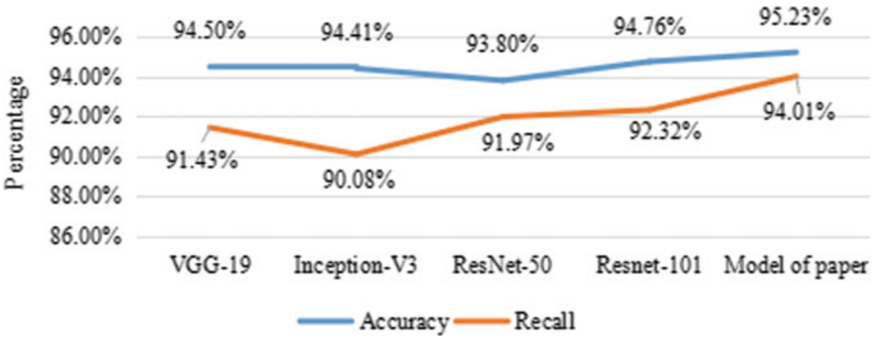


Fig. 18.2 Network models classification recognition result comparison in test set

Table 18.2 Network models classification recognition result comparison in test set

	Accuracy (%)	Recall (%)
VGG-19	94.50	91.43
Inception-V3	94.41	90.08
ResNet-50	93.80	91.97
ResNet-101	94.76	92.32
Model of paper	95.23	94.01

18.3.4 Classification and Identification Test of Each Network Model Structure in Test Set

According to Fig. 18.2 and Table 18.2, after testing each network model in the verification set, the test results show that the recall rate and accuracy rate of this model are the highest, 94.01 and 95.23%, respectively.

It can be seen from Tables 18.1 and 18.2 that the accuracy and recall rate of the identification model RIRNet in the verification set and test set are the highest. It can be seen that the accuracy of this model is similar to other models, mainly because the recall rate is 1.5–2% points higher than other models, which proves that the model has higher accuracy in the recognition of target images.

The above models, such as VGG-19, inception-v3, ResNet-54, and ResNet-101, which are compared with RIRNet, do not use transfer learning. After Imagenet pre training, the accuracy and recall rate of these models will increase by about two percentage points. Due to the limited computing resources, the model designed in this chapter is not trained by Imagenet. In practical application, the accuracy and recall rate can be further improved by pretraining.

18.4 Conclusion

This paper mainly focuses on how to carry out image recognition on satellite ship data set for ship monitoring. Based on the deep neural network framework, the existing neural network algorithms are analyzed. Through the experimental results, the effectiveness of the model is reasonably evaluated. The results of VGG-19, ResNet-34, ResNet-50, and perception-v3 models are compared. The results show that the accuracy and recall of the verification set and the test set are better than other models. The accuracy rate of verification set of RIRNet is 94.30%, and the recall rate is 92.03%; the accuracy rate of test set is 95.23%, and the recall rate is 94.01%. It can be seen that the network model designed in this chapter has achieved good results for satellite ship image recognition.

References

1. Perera, L.P.: Statistical filter based sensor and DAQ fault detection for onboard ship performance and navigation monitoring systems. IFAC PapersOnLine 49(23):323–328 (2016)
2. Bocchetti, D., Lepore, A., Palumbo, B., et al.: A statistical approach to ship fuel consumption monitoring. J. Ship Res. **59**(3), 162–171 (2015)
3. Fouda, B.M.T., Han, D., An, B., et al.: Design and implementation of software for ship monitoring system in offshore wind farms. Model. Simul. Eng. **2019**(10), 1–11 (2019)
4. Pelich, R., Chini, M., Hostache, R., et al.: Coastline detection based on Sentinel-1 time series for ship- and flood-monitoring applications. IEEE Geosci. Remote Sens. Lett. **99**, 1–5 (2020)
5. Mitsumori, W., Murakami, M.: Ship monitoring support service through the use of IoT technology offered by MES—ClassNK CMAXS e-GICSX and fleet monitor. Mar. Eng. **52**(2), 211–216 (2017)
6. Kim, D.H., Jung, B.K., Han, S.J.: A study on the monitoring method of ship hull and propeller performance by operating ship. J. Korean Soc. Mar. Environ. Saf. **26**(1), 15–21 (2020)
7. Anto, P., Hoogeland, M., Vredeveldt, A.W., et al.: E-inspection: effect of continuous hull monitoring on ship safety and crew workload. Int. Shipbuild. Progr. **66**(3), 1–24 (2019)
8. Soner, O., Akyuz, E., Celik, M.: Statistical modelling of ship operational performance monitoring problem. J. Mar. Sci. Technol. **24**(2), 543–552 (2019)
9. Jeong, J., Kim, T.H., Yang, C.S.: Construction of real-time remote ship monitoring system using Ka-band payload of COMS. Korean J. Remote Sens. **32**(3), 323–330 (2016)
10. Xiang, C., Li, B.: Research on ship intelligent manufacturing data monitoring and quality control system based on industrial Internet of Things. Int. J. Adv. Manuf. Technol. **107**(3), 983–992 (2020)

Chapter 19
Application of Remote Network Technology in Engine Room Communication of the Ship

Keyin Miao, Renqiang Wang, Jianming Sun, and Hua Deng

Abstract In recent years, the ship communication network based on computer and Internet of things (IoT) is an important part of ship modernization and intelligence. The first problem of intelligent ship is to solve the communication problem of ship engine room, so the stability of wireless communication system in ship steel environment becomes the focus of research. This paper compares and analyzes several mainstream narrowband IoT transmission modes, and selects LoRa technology as the wireless communication means. A communication test equipment based on LoRa mechanism is designed and manufactured, which is used to collect and analyze the signal strength in the engine room of a real ship. The measured results are compared with the simulation results, and then a reliable transmission path of narrowband IoT Lora wireless communication in the engine room of a ship is proposed.

19.1 Introduction

At present, more and more intelligent devices are applied to ship systems, and the explosive growth of network data in ship communication systems, and the timeliness and stability of data transmission put forward higher requirements for ship communication systems [1]. In the environment of dense equipment in ship cabins, there are higher requirements on the perforated wiring and the size of the holes. In contrast, the use of remote wireless communication between the compartments will greatly shorten the ship manufacturing period. When the cable of the wired communication system has a fault, it needs to spend a lot of time and manpower to replace and repair the cable [2]. In terms of installation cost, the deployment of ship's wired communication system accounts for a high proportion of the overall cost of the ship, the deployment of ship's automation system accounts for 1/6 of the total cost of the ship, the marine wired communication cable accounts for 1/4 of the automation system, and the cost of using remote wireless signal transmission technology is

K. Miao · R. Wang (✉) · J. Sun · H. Deng
Navigation College, Jiangsu Maritime Institute, Nanjing 2111709, China
e-mail: wangrenqiang2009@126.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_19

only 1/10 of the cost of the wired communication system. In the transformation of ship communication system, the traditional whole ship wiring ship communication system transformation is difficult and the workload is large. In contrast, the workload of wiring will be greatly reduced when the remote wireless communication system is used in the transformation of the engine room communication system.

Compared with wired communication system, remote wireless communication system has great advantages in terms of late maintenance and convenience of overhaul. Remote wireless data transmission based on IoT technology also provides a reference direction for the overall development of ship communication system. The automation level and modernization ability of ship communication have been greatly improved, which makes the ship remote wireless communication system become an important auxiliary equipment for safe navigation at sea, and also makes it possible for the remote wireless communication system to become the mainstream of ship communication system [3].

With the development of IoT technology and the wide use of sensor networks, the new communication system which integrates IoT and various technologies has been widely used. Nowadays, high-tech communication equipments are developing towards the direction of long-distance wireless and convenient. The highly information-based and fast computing of intelligent equipment and sensors provide users with shorter delay and more reliable data. With the introduction of 5G technology, it also marks that the remote wireless communication is gradually becoming the mainstream data communication system on land. The reliability of remote wireless communication is getting higher and higher, so that the remote wireless communication system can basically realize remote wireless data transmission inside the ship.

Most cabins are made of metal and steel, and the reliability of wireless signal transmission has become the main obstacle to achieving wireless coverage. In addition, the steel environment of the ship's cabin and the influence of noise during navigation and other related factors are based on the wireless signal of the IoT communication technology. It will be affected in the transmission process. Therefore, this paper studies and optimizes the reliability of computer data wireless communication in ship communication and navigation.

19.2 IoT Wireless Communication Based on LoRa Technology

19.2.1 Wireless Communication in Internet of Things

The main advantages of narrow-band IoT are low cost, low power consumption, reliable link, long transmission distance, and so on. All aspects are more advanced and scientific than other technologies. It has become a wireless communication technology with wide coverage and multi terminal access [4].

In addition to the most widely used technologies such as WiFi, Bluetooth, nb.IoT, and NFC, many wireless technologies play an indispensable role in their respective application scenarios. By comparing the transmission distance and anti-interference performance of various wireless technologies, we can see that the advantages of LoRa technology are particularly obvious, so the narrow-band Internet of things based on LoRa technology has a good prospect in the application of ship engine room monitoring system. In addition, satisfied with the different requirements of different customers for data protocols, LoRa protocol is widely used in many scenarios with its own uniqueness and wide adaptability [5]. Because of its fast network deployment, it can complete networking in a very short time. Part of the data in the engine room of the ship is confidential. LoRa Wan is a protocol specification based on LoRa technology. It adopts 128 bit symmetric encryption algorithm AES for integrity protection and data encryption, and uses LoRa technology for ship ad hoc network to ensure the safety of data transmission.

19.2.2 Introduction of LoRa Technology

LoRa is a low-power communication technology, which is based on spread spectrum technology for ultra-long distance, low-power wireless transmission. The specification of wireless Internet of things, which is jointly developed and initiated by Semtech, efficiency, IBM research and other European and American companies, has the advantages of long transmission distance, low power consumption and many networking nodes. At present, Lora mainly operates in the global free frequency band, including 433, 868, 915 MHz, etc. It does not charge communication fees and belongs to the ad hoc wireless network. LoRa communication transmits data with a bandwidth of 37.5.200 kb/s. The communication distance radius of open and non-sheltered areas such as suburbs or sea surface is 15–20 km; the communication distance radius of sheltered areas such as urban area is 5.10 km [6].

LoRa-WAN network architecture is composed of terminal node, Lora base station, web server, and data server. Six application scenarios based on this network architecture are listed. In this network architecture, a node can connect multiple terminal sensors or devices with transmission protocol in series. The base station forwards the Lora WAN protocol data between the web server, the data server and the terminal. The data is transmitted wirelessly through Lora or through TCP/IP [7].

19.2.3 Lora Terminal Working Mode

LoRa technology has three terminal working modes of class A/B/C, which can be applied to all Internet of things application scenarios.

(1) Class A (a11): the terminal of class A working mode adopts ALOHA protocol to report data on demand, which is the most basic working mode of Lora technology. After each data uplink, first open the receiving window for receiving the data sent by the server. If no data is received in the first receiving window, open the second receiving window to realize the two-way transmission of data. This kind of operation has the lowest power consumption, but it must wait for the terminal to report the data before transmitting the data. It include temperature monitoring, smoke detection, gas monitoring and other application scenarios [8].

(2) Class B (beacon): for the working mode of Class B, in addition to the random receiving window of class A, Class B will open the receiving window at a specified time. In order to achieve this function, the terminal node needs to receive time synchronization signal markers from the base station. The downlink transmission of data is completed in the fixed receiving window of the terminal, but the data transmission time becomes longer. Application scenarios include water meter, gas meter, electricity meter, etc.

(3) Class C (continuous): in Class C mode, the receiving port remains open, and it will be temporarily closed when sending data. This mode will cause unnecessary window waste, so the terminal of class C will consume more power than Class A and class B. Because the terminal is always in continuous receiving state, it can send data to the terminal at any time. It is commonly used in street lamp control and other scenes [9].

19.2.4 Radio Wave Propagation Mechanism

The equipment in the engine room of a ship is generally made of steel. The propagation of electromagnetic wave in the engine room of a ship can be summarized as reflection, transmission, and diffraction. When the electromagnetic wave is incident on the surface of the equipment or engine room, it will be reflected; when the electromagnetic wave is incident on the glass window of the central control room, it will be reflected and transmitted: when the electromagnetic wave is blocked by the edge of a sharp object or meets an object similar to the wavelength diffraction occurs [10]. The reflection coefficients of parallel and vertical electromagnetic waves are calculated by Eqs. 19.1 and 19.2:

$$\Gamma_1 = \frac{-\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}}{\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{19.1}$$

$$\Gamma_2 = \frac{\varepsilon_r + \sin\theta_i - \sqrt{\varepsilon_r - \cos^2\theta_i}}{\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{19.2}$$

where Γ_1 is the reflection coefficient of parallel electromagnetic waves, and Γ_2 is the reflection coefficient of vertical electromagnetic waves.ε_r is the reflection medium constant, and θ_i is the oblique incident angle of electromagnetic waves.

Transmission is mentioned in the content of electromagnetic wave reflection characteristics. Some electromagnetic waves are reflected at the interface of the medium, and some electromagnetic waves are transmitted through the interface of the medium. In optics, refraction is called refraction. The formula of electromagnetic wave transmission can be obtained by deriving the reflection coefficient. See Eqs. 19.3 and 19.4:

$$\Gamma_1 = \frac{2\sin\theta_i}{\sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{19.3}$$

$$\Gamma_2 = \frac{2\sin\theta_i}{\varepsilon\sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{19.4}$$

where Γ_1 is the transmission coefficient of parallel electromagnetic waves, and Γ_2 is the transmission coefficient of vertical electromagnetic waves. ε is the electromagnetic wave incident medium constant.

19.3 Communication Hardware Program Design Based on LoRa Mechanism

19.3.1 Communication Hardware Node Selection

The communication hardware selects LoRa node developed by the laboratory for energy consumption collection in plant area, park, and mine. The node has TTL and RS485 serial communication modes, supports the access of devices and sensors with Modbus RTU communication protocol, and the working temperature is minus 40°. The temperature in the cabin is generally not more than 45°. The wireless transmission module of the LoRa node terminal is the LoRa series module of Ra-01 produced by Anxin Co technology company, which is completely suitable for the cabin environment.

19.3.2 Construction of Development Environment

Keil u-Vision 4 software is used in the CPU program design of the hardware device. The software development interface is simple and friendly, and has the development environment of enhanced integrated language. According to the problems of different CPU programming encountered in user development, software developers

Table 19.1 Antenna parameters

Frequency range	Gain	Standing-wave ratio	Input impedence	Power capacity	Antenna height
433 MHz	15 dBi	≤1.5	50 Ω	100 W	205 mm

developed this kind of software with different processors that can be programmed in the same integrated development environment. The software integrates the function items needed to develop the hardware equipment used in this test, and integrates text editor, project management, program compilation and other functions. The software supports C language and assembly language for programming. This design uses C language for programming. The advantages of C language are small amount of code, simple writing, fast compiling speed, powerful language function, and good code readability.

19.4 Experimental Verification and Analysis

19.4.1 Antenna Parameter Test

It is the most direct and accurate method to test the received power of radio wave. To detect the signal strength distribution and attenuation of LoRa narrowband communication technology in the actual ship engine room, the experimental platform of LoRa mechanism communication in the actual ship engine room is built.

The equipment selected for the experimental platform includes LoRa communication node, notebook computer with uartast serial program, adjustable height tripod and antenna. The 433 MHz omnidirectional, monopole and vertically polarized copper rod antenna is used in the test. The overall height is 205 mm and the effective length is 175 mm, which is consistent with the simulation parameters (Table 19.1).

19.4.2 Data Comparison

It is found that the lowest receiving power measured in path B is the # 13 receiver, while the # 15 receiver in simulation has the lowest receiving power, which is not in the same location, but in the same area; in the measured environment, six groups of eight receivers are located near the stairs, and the receiving power fluctuates greatly, which is consistent with the trend of simulation data. #Under the Los environment, the simulation results show that the receiving power of # 25 receiver is larger than that of # 24 receiver, while the measured data shows that the receiving power of # 24 receiver is larger than that of # 25 transmitter. The data difference is caused by the different processing methods of the measured equipment and simulation software (Fig. 19.1).

In Fig. 19.2, the received power of measured data path C shows a decreasing trend, and the difference between path C.2 and simulation data is large, and the attenuation range of measured data is larger than that of simulation data. In the actual ship engine room environment and simulation environment, the # 6 receiver on path C is in the same position and is closest to the transmitter, but the receiving power of the # 6 receiver in the actual environment is much lower than that of the adjacent receiver. Comparing the received power attenuation amplitude of path B and path C in the measured environment, we can still find that the received power conversion amplitude of the receiver increases with the increase of the plane height of the receiver relative to the transmitter.

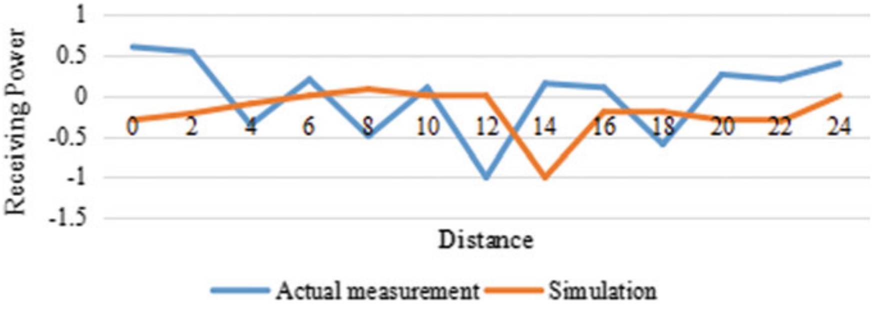


Fig. 19.1 Comparison of path B measurement and simulation

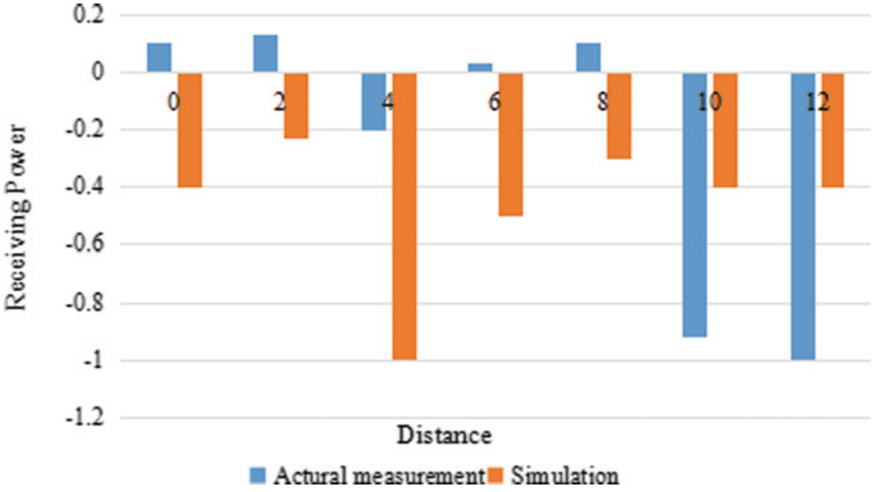


Fig. 19.2 Comparison of path C measurement and simulation

It is found that the overall trend of measured data and simulation data is consistent, but there are still differences, mainly due to the following reasons:

(1) The simulation environment uses ray tracing algorithm to get the received power. Due to the multipath effect of the radio wave in the ship engine room environment, the propagation of the transmitted wave has delay expansion. The calculation of the signal strength of the LoRa communication node is different from that of the simulation software. The signal strength of the first ray received by the LoRa module is the receiving power of the test point.

(2) Ray tracing algorithm is closely related to the surrounding environment. There are grooves, metal frames, various pipes and other obstacles in the real ship engine room, which are not reflected in the simulation environment.

(3) When the narrowband signal receiving power of ship engine room is measured, some equipment are in working state, and the environmental noise is large, so the influence of working condition and temperature is not considered in the simulation environment.

(4) The measured signal strength is seriously affected by the environment, which makes the received power result and the simulation result have larger error.

19.5 Conclusion

The development of intelligent engine room under intelligent ship needs corresponding remote wireless communication system. For the long-distance wireless communication system suitable for the environment of ship engine room, it is necessary to understand the propagation characteristics of radio waves in the engine room. In this paper, the engine room environment measurement and simulation of a real ship are studied. The signal strength acquisition program based on LoRa mechanism communication hardware equipment is designed, which is used to collect the narrowband signal strength in the engine room of a real ship. The results show that the receiver should not be placed close to the back of the obstacle, on both sides of the obstacle which is diagonal to the transmitter, and on the position with strong vibration.

References

1. Abdelmoaty, A., Dahman, G., Bousselmi, A.A., et al.: Using vertically separated MIMO in ship-to-ship communications. IEEE Access **3**, 99 (2020)
2. None. Study shows disconnect in communications at membership organizations. Nonprofit Bus. Advis. **324**, 132–133 (2016)
3. Chen, J.L., Zhang, et al.: Planetary gearbox condition monitoring of ship-based satellite communication antennas using ensemble multiwavelet analysis method. Mech. Syst. Signal PR **45**, 277–292 (2015)
4. Gray, A.W., Singer, D.J.: A hybrid agent Type-1 fuzzy logic system for set-based conceptual ship design communications and negotiations. Nav. Eng. J. **128**(1), 77–89 (2016)

5. Kim, H.-G., Joo, Y.-I., et al.: Distributed optimal path guidance system considering communication isolation and transmission power for multistory ship evacuation. J. Korean Soc. Mar. Eng. **41**(7), 677–682 (2017)
6. Perera, L.P., Mo, B.: Ship performance and navigation data compression and communication under autoencoder system architecture. J. Ocean Eng. Sci. **3**(2), 133–143 (2018)
7. None. Add membership value with a strong communications strategy. Membership Manag. Rep. **12**(3), 6–13 (2016)
8. Niwa, Y., Motogi, H., Nishizaki, C., et al.: Ship-to-ship radiocommunication trial by using wireless LAN. Int. J. E-Navig. Marit. Econ. **3**, 32–39 (2015)
9. Daniel, L., Hristov, S., Lyu, X., et al.: Design and validation of a passive Radar concept for ship detection using communication satellite signals. IEEE Trans. Aerosp. Electron. Syst. **53**(6), 3115–3134 (2019)
10. Kempton, S.B.: Transportation and communications: ship routing measures in international straits. Angew Chem. **54**(2), 560–563 (2015)

Chapter 20
Pi Test for Signal Complexity

Li Chen, Jie Cheng, and Wenshi Li

Abstract The basic functions of the golden standards of signal complexity measure are fast operation, accurate identification and no need to select parameters manually. Compared with the known spectral entropy (SE) complexity and construction creep (CC) rate of spring test for chaos, a novel candidate answer is Pi test for signal complexity. Its algorithm principle is mainly developed from the cast point proportion calculation of Pi based on Monte Carlo method. The key technologies depend on that (1) feeding the inputs of original random numbers with the periodic or chaotic signals under test, (2) replacing pie-shape with ball-shape, (3) and auxiliary application of cosine function inputs. One to three-dimensional chaotic equations are used as test cases. The results of Pi scoring for signal complexity are represented as Pi score (near zero means period; less than 2.7 signs chaos; and close to 3.14 marks random number) and Pi pie (ratio of Pi score to 3.3), rapidly agreeing with the distinguished results of Bifurcation diagram, SE complexity (the larger, the more complex) and CC rate (the two critical points are 7 and 83).

20.1 Introduction

20.1.1 Background

Research purposes: Jim Gray, winner of Turing prize in 1998, predicts that the amount of data will double every 18 months. In the future, more sensors will be deployed and more data will be generated. Data coding is facing unprecedented challenges, the first thing to deal with is efficient source coding to achieve efficient representation of information. In the era of big data, we are constantly studying it.

L. Chen · J. Cheng · W. Li (✉)
School of Electronics and Information En., Soochow University, Suzhou 215006, China
e-mail: lwshi@suda.edu.cn

W. Li
Laboratory of Modern Acoustics of MOE, Nanjing University, Nanjing 210093, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_20

Application area: For example, the human brain is the most complex structure in our human body, and the brain signal is also extremely complex. In order to better identify it, we need to find some new features of it, so as to improve the recognition rate conveniently and further analyze the brain signal, so as to make contributions to human scientific research.

For another example, we can also regard the fluctuation of the stock market as the change of signal. Is it random or periodic, or is it a chaotic state? In order to try to explore, we look for its characteristics and make effective investment!

In addition, with the rapid growth of data, more new information is waiting to be mined: Monitoring of human health information; Monitoring of machine health information… All aspects are waiting for us to carry out signal analysis, understanding and identification application, so as to carry out the feasibility of data-driven application research.

20.1.2 Theory

The most famous example of Monte Carlo method [1–4] is in Pi [5, 6] calculated by casting points within unit square surrounded pie-shape (Casting points in a circle, which is in a square). If we replace random numbers with periodic or chaotic signals, what is the new result [4]? Here the simplest criteria of signal complexity measure will be illustrated on Pi test, with Pi score and Pi pie.

To confirm the above idea, firstly three maps of chaos are selected as test cases, and then three kinds of control experiments are used with Bifurcation diagram, spectral entropy (SE) complexity and construction creep (CC) rate of spring test for chaos, at last, highlights on the detail of modified Pi scoring by Monte Carlo method are discussed.

Notice: Pi pie is a form of expression about the ratio (Pi score ratio to 3.3). Pi test is the estimate of Pi based on MC which is calculated by casting points.

20.2 Test Cases and Principles

20.2.1 Test Cases

Test case 1: Logistic map is written as formula (20.1), wherein the coefficient μ is state-dependent, which is a typical one-dimensional chaotic equation [4].

Test case 2: Lozi map is written in formula (20.2), wherein the coefficient a is state-dependent, which is a typical two-dimensional chaotic equation [7].

Test case 3: Chua's equation is as in formula (20.3), here the coefficient k is state-dependent, which is a typical two-dimensional chaotic equation [4].

$$x_{n+1} = \mu x_n(1-x_n) \tag{20.1}$$

$$x_{n+1} = 1-a|x_n| + 0.54272y_n,\ \ y_{n+1} = x_n \tag{20.2}$$

$$\mathrm{d}x/\mathrm{d}t = -2.564x + 10y + 0.5k(|x+1|-|x-1|)$$

$$\mathrm{d}y/\mathrm{d}t = x-y+z$$

$$\mathrm{d}z/\mathrm{d}t = -14.706y \tag{20.3}$$

Notice: The above three equations are optional for testing, so it is not necessary to study the coefficients of the equations or consider their practical significance. They are only used for data testing.

20.2.2 Modified Pi Scoring

The source code of the modified Pi scoring algorithm with MC is as follows [4].

```
close all; clear all;
x = x; %data under test1,2,3
y = y; %data under test1,2,3
z = cos(t); %as inputs of auxiliary data
n =length(x); %The length of data used to calculate
date = zeros(n,1);%create an array
    k = 0;%let the initial value of K be 0
    for i = 1:n    %Cycle n times
         if x(i)^2+y(i)^2<=1; %pie-shape casting points
%           if x(i)^2+y(i)^2+z(i)^2<=1; %ball-shape casting points
```

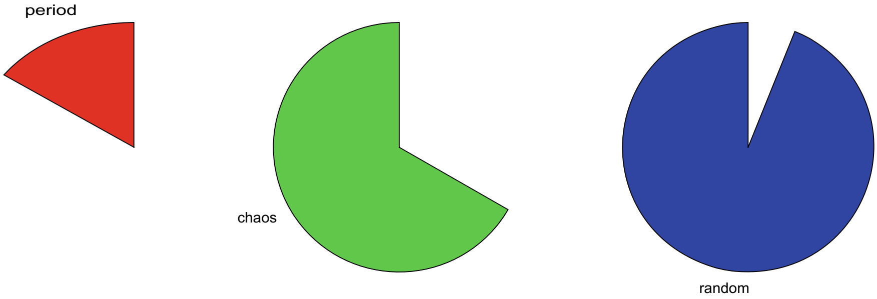


Fig. 20.1 Pi pie (ratio of Pi score) of Pi test (left for period, middle for chaos, and right random)

```
                k = k+1;%if the data point falls in a circle or a sphere with a radius of 1 centered on the origin position, the count is increased by 1

            end

              date(i,1)=4*(k/i); %Pi estimation=4*number of points in circle/ number of points in square(two-dimension)

  %             date(i,1)=6*(k/i); %Pi estimation=6*number of points in ball/ number of points in box(three-dimension)

        pii = date(n,1); %Pi score--Pi value based on the calculation

  x = [pii/3.3]; %Pi score ratio to 3.3
  cm = [1 0 0];%colormap(cm);the color of pie

  pie(x); %Pi pie
```

Key 1: Rise-dimensional of Pi calculation. From square to box, the casting points calculation of Pi expands two-input data to three dimensional (Fig. 20.1).

Key 2: Data input strategy. It is convenient for us to calculate each variable of the three-dimensional equation separately, using the cosine functions as the auxiliary input data in the box by Monte Carlo method.

Key 3: Verification of period and chaos. While the data length of random number is more than 1000 points, we can gain the convergent estimation of Pi constant. In general, we need the data lengths to be longer than 1000 during test.

20.2.3 Control Experiments

Bifurcation diagram illustrates the chaotic evolution processes from single period to multi-period to chaos, including the processes of withdrawing from chaotic states.

In brief there are basic methods of getmax instruction in MATLAB and Poincare section of relation selection of $x = y$ [8–11].

As two automatic criteria of signal complexity measures, SE complexity and CC rate are utilized individually.

SE complexity is based on the FFT instruction in MATLAB tool to control the flatness of frequency domain, which is a gold standard for identifying signal complexity. The complexity of the signal is as follows: the larger the SE value is, the more complex it is [12].

CC rate measures compressed attractor through self similar characterization, and describes the chaotic degree of the system according to the shape of the attractor, which usually takes a long time to calculate.

Wherein the corresponding rules are that SE complexity (the larger, the more complex) and CC rate (The two critical points are 7 and 84 for three classification recognition of period, chaos, and random states) [13, 14].

20.3 Results and Discussions

20.3.1 Results

In Fig. 20.2, we depict the Bifurcation diagram and SE complexity of trace states in Logistic map.

In Fig. 20.3, Pi scoring is illustrated (left in pie-shape casting and right in ball-shape casting; 4 kinds of data lengths). Wherein the test conditions are (1) initial value (0.1); (2) data length of 3000 points, abandoned first 1000 points; (3) step of 0.01; (4) control experiments of Bifurcation diagram and spectral entropy (SE) complexity.

In Fig. 20.4, we show the Bifurcation diagram and SE complexity of trace states in Lozi map.

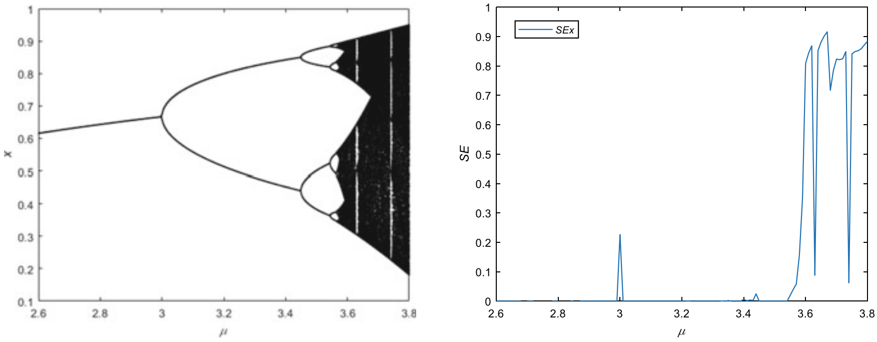


Fig. 20.2 Logistic map: Bifurcation diagram (left) and SE complexity (right)

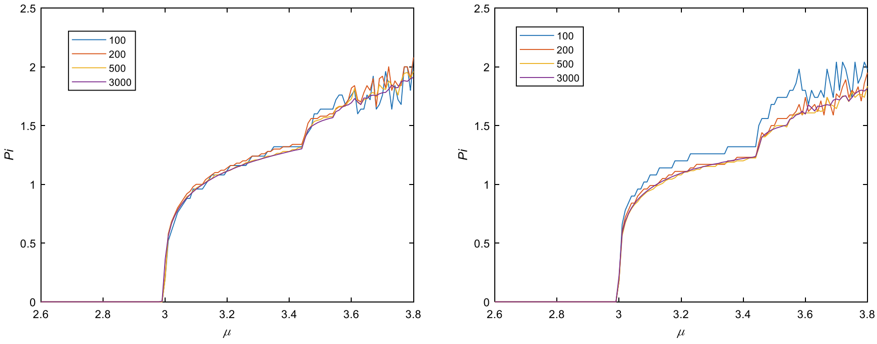


Fig. 20.3 Logistic map: Pi scoring (left in pie-shape casting and right in ball-shape casting; 4 kinds of data lengths)

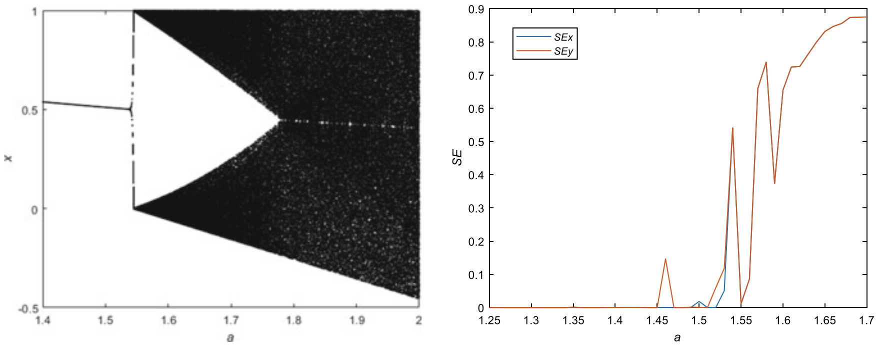


Fig. 20.4 Lozi map: bifurcation diagram (left) and SE complexity (right)

In Fig. 20.5, CC rate (left) and Pi scoring are illustrated (middle in pie-shape casting and right in ball-shape casting; 4 kinds of data lengths). Wherein the test conditions are (1) initial values (0.1, 0.1); (2) data length of 3000 points, abandoned first 1000 points; (3) step of 0.01; (4) control experiments of Bifurcation diagram and spectral entropy (SE) complexity and CC rate.

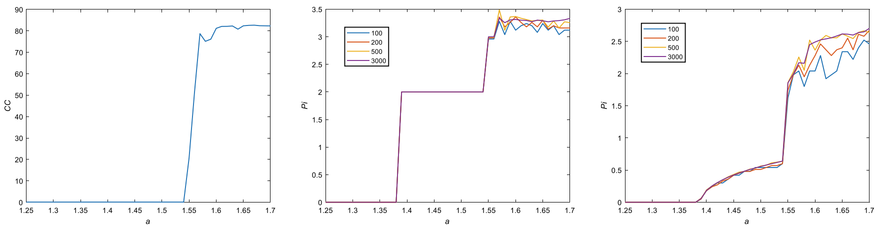


Fig. 20.5 Lozi map: CC rate (left) and Pi scoring (middle in pie-shape casting and right in ball-shape casting with 4 kinds of data lengths)

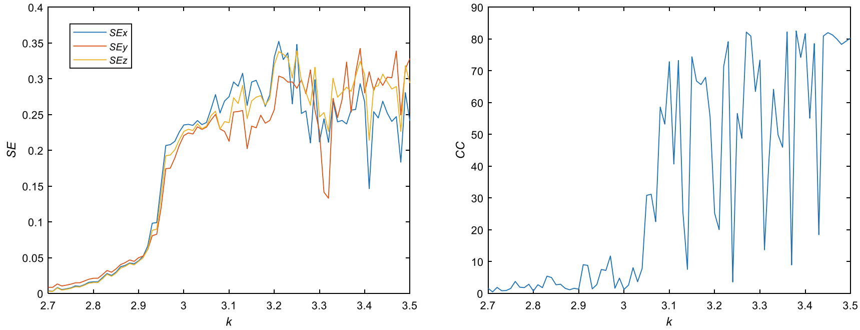


Fig. 20.6 Chua's equation: SE complexity (left) and CC rate (right)

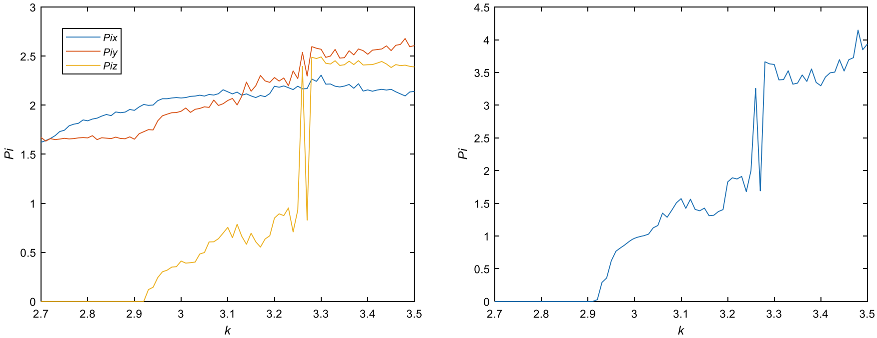


Fig. 20.7 Chua's equation: Pi scoring (left, keep ball-shape casting, set $y = z = \cos(t)$ in code) and (right, keep ball-shape casting)

From Fig. 20.6 to Fig. 20.7, we illustrated the automatic criteria of SE complexity (left) and CC rate (right, narrow-period-window probing) during scanning parameter k in Chua's equation. We set test conditions as (1) initial values (0.1, 0.1, 0.1); (2) data length of 3000 points, abandoned first 5000 points; (3) step of 0.01 in ode45 instruction.

20.3.2 Discussions

Note 1: Control tests' uniformity of Bifurcation diagram, SE complexity, and CC rate. In Fig. 20.2, Fig. 20.4, Fig. 20.5 (left), and Fig. 20.6 (k less than 3.01 marks period-states), SE complexity and CC rate signs the bifurcation points and chaotic states. Note 2: Data length. For rapid evolution of Logistic map, the data length touches 100 points (Fig. 20.3, Fig. 20.4, and Fig. 20.5). Note 3: Advantages and limitations of data inputs strategy. Comparing pie-shape with ball-shape casting, we

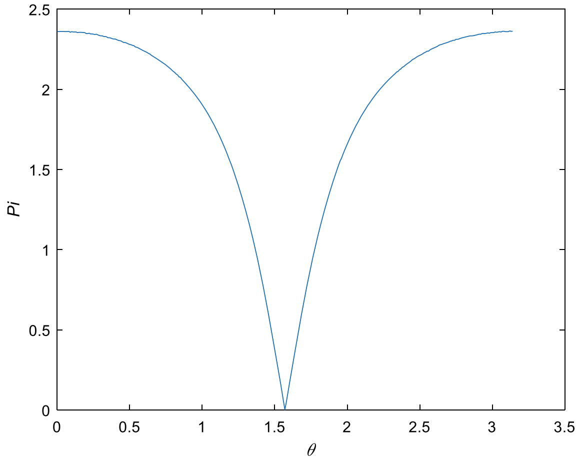


Fig. 20.8 Pi scoring of $x = \cos(t + \theta)$ recognition (keep ball-shape casting, set $y = z = \cos(t)$ in code)

suggested using the ball-shape casting in reason of lower Pi values showing (see Fig. 20.5, right versus middle). But while we check the single input of $x = \cos(t + \theta)$, there exist the blind zones of period recognition (θ: 0 ~ π/3, 2π/3 ~ π; Pi should be less than 1.7). Fortunately, there are SE complexity and CC rate as cross criteria. And the calculating time of Pi test is 0.75 times of SE complexity (Fig. 20.8).

20.4 Conclusions

The simplest Pi test for signal complexity had been proposed compared with Bifurcation diagram, SE complexity, and CC rate. By the understanding of modified Pi computing of Monte Carlo method, Pi test uses Pi scores to verify period (in value range of 0–1.6), chaos (1.7–2.9) and random states (3.0–3.3) in test case studies, and their Pi pie features the ratio of Pi score to number 3.3. Although the calculating time of Pi test is just 0.75 times of SE complexity consumption, there are blind zones of the cosine recognition in its 2/3 phase change (Keynote: The multi-criteria strategy is to apply SE and CC). Further identification validity tests should pay attention to both more test cases researching and future health feature modeling of micro-nano electronics.

References

1. Hu, X., Chen, X., Parks, G.T., Yao, W.: Review of improved Monte Carlo methods in uncertainty-based design optimization for aerospace vehicles. Prog. Aerosp. Sci. **86**, 20–27 (2016)
2. Charles, J.M., Robert, J.G.: A review on Monte Carlo simulation methods as they apply to

mutation and selection as formulated in Wright-Fisher models of evolutionary genetics. Math. Biosci. **211**(2), 205–225 (2008)

3. Milligan, G.W.: A review of Monte Carlo tests of cluster analysis. Multivar. Behav. Res. **16**(3), 379–407 (1981)
4. Li, W.S.: Case Study of Micro-nano Electronics Modeling (in Chinese), 2nd edn. Soochow University Press, Suzhou (2019)
5. Montero, R.S.: State of the art of compactness and circularity measures. Int. Math. Forum **4**(27), 1305–1335 (2009)
6. Herrera-Navarro, A.M., Hernández, H.J., Peregrina-Barreto, H., Manríquez-Guerrero, F., Terol-Villalobos, I.R.: A new measure of circularity based on distribution of the radius. Computación y Sistemas **17**(4), 515–526 (2013)
7. Sprott, J.C.: Maximally complex simple attractors. Chaos 17, 033124–1–6 (2007)
8. Fischi, J., Nilchiani, R., Wade, J.: Dynamic complexity measures for use in complexity-based system design. IEEE Syst. J. **11**(4), 2018–2027 (2017)
9. Ishii, R., Canuet, L., Aoki, Y., Hata, M., Iwase, M., Ikeda, S., Nishida, K., Iked, M.: Healthy and pathological brain aging: from the perspective of oscillations, functional connectivity, and signal complexity. Neuropsychobiology **75**(4), 151–161 (2017)
10. Tang, L., Lv H.L., Yang, F.M., Yu, L.A.: Complexity testing techniques for time series data: a comprehensive literature review. Chaos Solitons Fractals **81**, 117–135 (2015)
11. Wade, J., Heydari, B.: Complexity: definition and reduction techniques: some simple thoughts on complex systems. Complex Syst. Des. Manag. **1234**(18), 213–226 (2014)
12. Su, K.H., He, S.B., He, Y., Yin, L.Z.: Complexity analysis of chaotic pseudo-random sequences based on spectral entropy algorithm. Acta Phys. Sin. **62**(1), 27–34 (2013)
13. Wu, S.L., Li, Y.T., Li, W.S., Li, L.: Chaos criteria design based on three-threshold sign function. Chin. J. Electron. **28**(2), 364–369 (2019)
14. Cai, J.W., Li, Y.T., Li, W.S., Li, L.: Two entropy-based criteria design for signal complexity measures. Chin. J. Electron. **28**(6), 1139–1143 (2019)

Chapter 21
A Method of Respiratory Monitoring Based on Knowledge Graph

Yiying Zhang, Delong Wang, Baoxian Zhou, and Yiyang Liu

Abstract Sleep is an indispensable part of people's daily life, so it is particularly important to monitor their sleep state. Based on this, this paper proposes a respiratory monitoring method based on a knowledge graph to monitor the sleep state. Firstly, the data collected by the sensor terminal and related disease information are identified by named entity recognition and relationship extraction, and the knowledge graph of respiratory monitoring data is constructed, and then it is visualized through the graph database Neo4j. Finally, on the basis of the existing knowledge graph, the relationship model between respiratory monitoring index and disease is established to realize the perceptual analysis of human health state, so as to achieve the purpose of disease prediction.

21.1 Introduction

In recent years, with the continuous improvement of medical standards and quality of life, people have increased requirements for sleep quality. Therefore, monitoring the sleep state to predict diseases to achieve the purpose of improving sleep has become a common means [1]. However, there are many kinds of sleep monitoring indicators and monitoring methods, so the sleep monitoring method is designed to accurately predict which disease the monitored person will get while monitoring people's sleep. It has become an important focus to improve people's sleep quality.

Although the existing respiratory monitoring methods have become more and more perfect, there are few methods involving knowledge graph, and the accuracy of disease prediction still needs to be raised. In this paper, based on the respiratory monitoring data collected from the respiratory sensor terminal, named entity recognition and relationship extraction are carried out, and the knowledge graph of respiratory monitoring data is constructed, on the basis of which the correlation model between

Y. Zhang · D. Wang (✉) · B. Zhou · Y. Liu
College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China
e-mail: 17862328885@163.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_21

respiratory monitoring index and disease is established. to achieve the prediction of diseases, so as to improve people's sleep quality.

21.2 Research Status

The technology of monitoring sleep status has received widespread attention in the domestic industry. In earlier studies, there were technical methods for the detection of sleep apnea using the Back Propagation (BP) neural network. The prediction accuracy of the intelligent detection system of sleep respiratory syndrome developed by this method can reach about 88.5%. In recent years, polysomnography (Polysomnography, PSG) is commonly used to monitor sleep apnea syndrome [2]. This method monitors patients through sensors connected to various parts of the body.

With the increasing demand for sleep monitoring technology, portable sleep monitoring system came into being. The portable sleep monitoring system is divided into Electroencephalogram (EEG) sleep monitoring system and non-EEG sleep monitoring system. Among them, the non-EEG sleep monitoring system is represented by the intelligent bracelet, and the working principle of the intelligent bracelet is that a three-axis accelerometer is built into the intelligent bracelet to judge the sleep status according to the range of activity during sleep [3]. The smart bracelet has the advantages of small size, easy to wear, low power consumption, low price, and so on. However, the smart bracelet will record long bed rest as sleep time, while insomnia patients have a long bed rest state. Therefore, there will be some errors in the evaluation of sleep [4]. Therefore, designing a method to monitor people's sleep in order to prevent the occurrence of sleep diseases is of great significance to improve people's sleep quality.

21.3 Construction of Respiratory Monitoring Model

The respiratory monitoring model based on knowledge graph is mainly designed from four aspects, including named entity recognition, entity relationship extraction, graph database storage, and knowledge update [5, 6]. Because most of the data received by the respiratory sensor terminal is text data, which belongs to unstructured data, the data basis of knowledge graph construction is aimed at unstructured data [7], and the architecture design is shown in the following Fig. 21.1.

In this paper, the named entity recognition of the respiratory monitoring index is realized by using Bidirectional Long Short-term Memory and Conditional Random Field (BiLSTM-CRF) knowledge extraction model. Firstly, the word embedding model is used to convert the text data collected by the sensor terminal into word vectors, then BiLSTM neural network is used to extract text features of word vectors [8], and finally, CRF is used to label the extracted corpus to achieve accurate recognition of named entities [9].

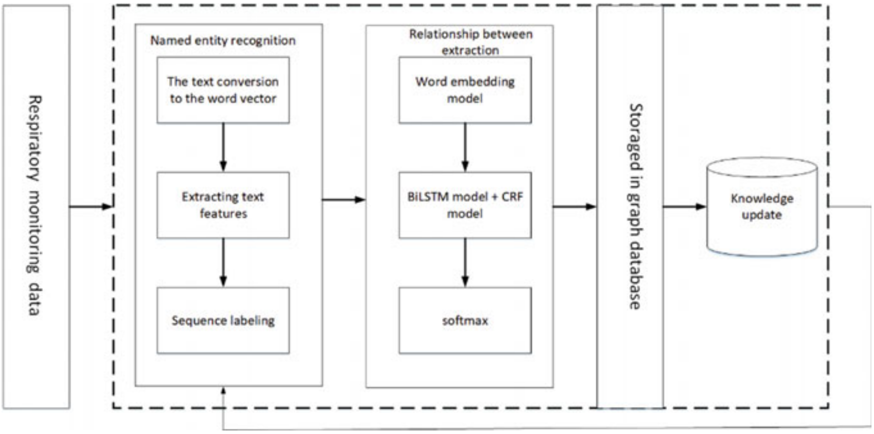


Fig. 21.1 Architecture design of respiratory monitoring model

In this paper, on the basis of named entity recognition, the text features extracted by BiLSTM and the entity tags predicted by CRF are used as the input of softmax function layer, and the problem of relationship classification is transformed into the problem of maximum probability, so as to realize the accurate extraction of the relationship between respiratory symptoms and disease. Finally, the graph database Neo4j is used to store entities and relationships, and dynamically update the knowledge graph, so that it can guide the related work of disease prediction.

21.4 Respiratory Monitoring Method

Based on the respiratory monitoring model based on knowledge graph, a respiratory monitoring method is designed to realize the perceptual analysis of human health, so as to achieve the purpose of disease prediction. The overall implementation scheme is shown in the following Fig. 21.2.

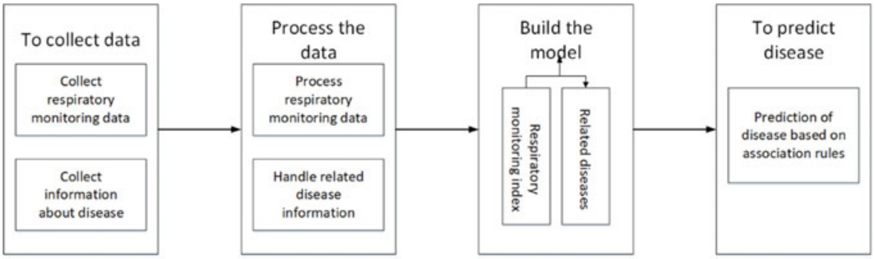

To collect data
Collect respiratory monitoring data
Collect information about disease
Process the data
Process respiratory monitoring data
Handle related disease information
Build the model
Respiratory monitoring index
Related diseases
To predict disease
Prediction of disease based on association rules

Fig. 21.2 Block diagram of the overall implementation scheme

21.4.1 Collect and Process Data

Collect respiratory monitoring data from the knowledge graph that has been constructed, and collect relevant disease information from the Internet. The respiratory monitoring data used in this paper are all collected by the respiratory sensor terminal, from which four respiratory monitoring indicators: respiratory intensity, snoring, heartbeat frequency, and pulse beating times are selected as the research object. There are many diseases related to sleep breathing. In this paper, six common diseases are selected to study: hypertension, angina pectoris, myocardial infarction, arrhythmia, ischemic cardiac arrest, and stroke.

Because the Apriori algorithm is not suitable for continuous variables, the respiratory monitoring data and related disease information are discretized in order to analyze and solve the problem.

21.4.2 Design of Respiratory Monitoring Method

Once the data are collected and processed, the disease can be predicted based on the data, as shown in the following Fig. 21.3.

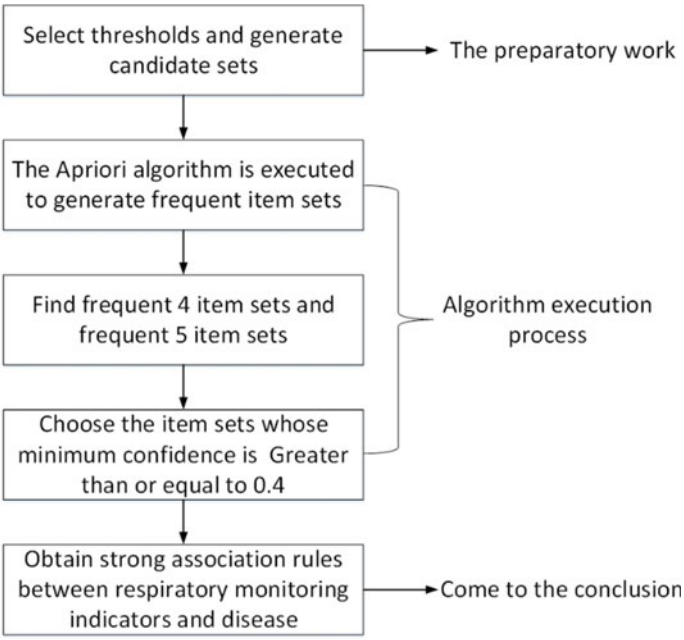

Select thresholds and generate candidate sets
The preparatory work
The Apriori algorithm is executed to generate frequent item sets
Find frequent 4 item sets and frequent 5 item sets
Algorithm execution process
Choose the item sets whose minimum confidence is Greater than or equal to 0.4
Obtain strong association rules between respiratory monitoring indicators and disease
Come to the conclusion

Fig. 21.3 Disease prediction process

Step 1: Select the threshold. Using the Apriori algorithm to construct the correlation model between respiratory monitoring index and disease, the first and most important step is the selection of minimum support and minimum confidence, which plays a significant role in the accuracy of the experimental results. After performing a large number of experiments and consulting a certain number of literature, it is decided to set the minimum support to 0.04 and the minimum confidence to 0.4. Because the purpose of this paper is to study the relationship between respiratory monitoring index and disease, the respiratory monitoring index is set as the antecedent of the rule, and the disease is set as the posterior of the rule, so as to predict the possible diseases in different respiratory states [10].

Step 2: We have previously processed the required data, so it directly scans the entire data set, obtains each item, and generates the candidate set C1. The items in C1 that are less than the preset minimum support of 0.04 are removed to generate a frequent item set L1. Because L1 contains only one item, it cannot be associated, that is, disease prediction cannot be carried out, so L1 is connected and pruned to generate candidate 2 item set C2.

Step 3: Executes Apriori algorithm on the candidate set C2, and obtains frequent 2 item set L2, frequent 3 item set L3, frequent 4 item set L4, and frequent 5 item set L5 of respiratory monitoring index and disease. The specific process is shown in the following Fig. 21.4.

(1) Each item in the candidate set C2 is counted, and then the items that do not meet the requirements are eliminated according to the minimum support degree, so as to obtain the frequency 2 item set L2 of respiratory monitoring index and disease.

(2) Perform pruning and join operations on L2 to generate the set C3 of candidate 3 item set, then scan all data sets and count each item in C3. Similarly, the items that do not meet the requirements are removed from C3 according to the minimum support degree, and frequent 3 item set L3 of respiratory monitoring index and disease are obtained.

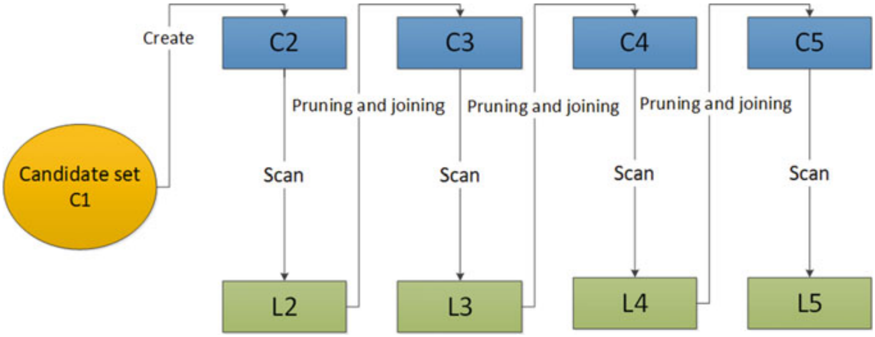


Fig. 21.4 Generate frequent item sets

(3) The frequent 4 item set L4 of respiratory monitoring index and disease was obtained by the same method.
(4) The frequent 5 item set L5 of respiratory monitoring index and disease was obtained.

Step 4: Finds frequent 4 item set and frequent 5 item set from all generated item sets. Because the frequent 2 item set involves only one respiratory monitoring index, and the frequent 3 item set only involves two respiratory monitoring indexes, there will be a big deviation in the accuracy of disease prediction results. Therefore, in order to make the prediction results more accurate, frequent 4 item set and frequent 5 item set are selected as research objects to predict possible diseases.

Step 5: From the frequent 4 item set and frequent 5 item set, we can find the item set whose minimum confidence is greater than or equal to our preset minimum confidence level 0.4, so that we can get the strong association rules between respiratory monitoring indexes and diseases, which can be used to predict diseases.

21.5 Simulation Experiment Design

In this paper, four commonly used respiratory monitoring indicators, namely, respiratory rate, snoring, heart rate, and pulse beat times, were selected as the research objects. The experiment compared the accuracy of Apriori algorithm, Multivariate State Estimation Technique (MSET) algorithm, and Neural Network in disease prediction under the above respiratory monitoring indexes. The experimental results are shown in the following Figs. 21.5, 21.6, 21.7, and 21.8.

As can be seen from the experimental results, compared with the other two algorithms, the method in this paper has higher accuracy in disease prediction, up to 97%.

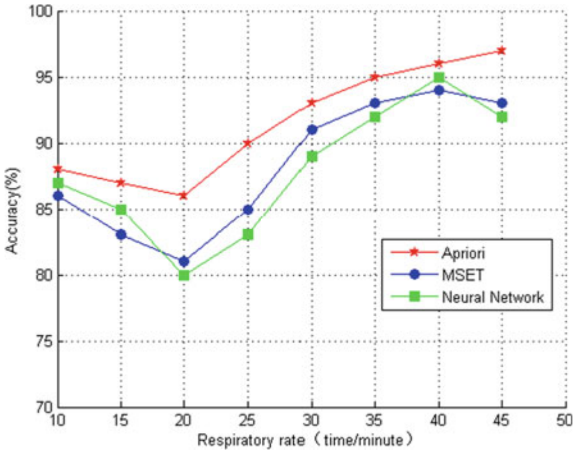


Fig. 21.5 Respiratory disease prediction accuracy at different respiratory rate

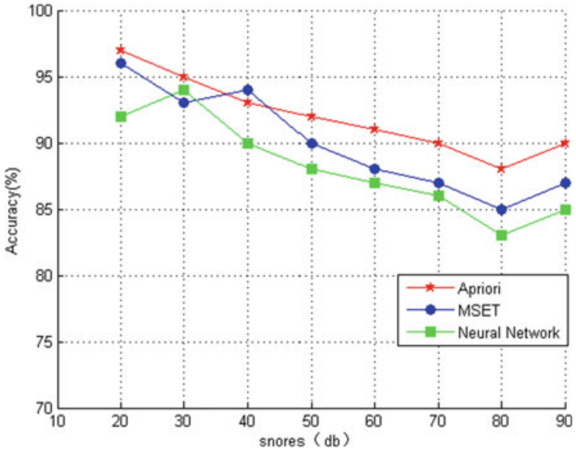


Fig. 21.6 Respiratory disease prediction accuracy at different respiratory snores

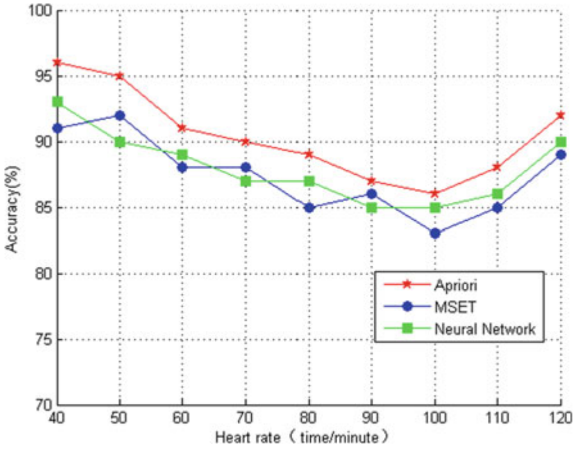


Fig. 21.7 Respiratory disease prediction accuracy at different respiratory heart rate

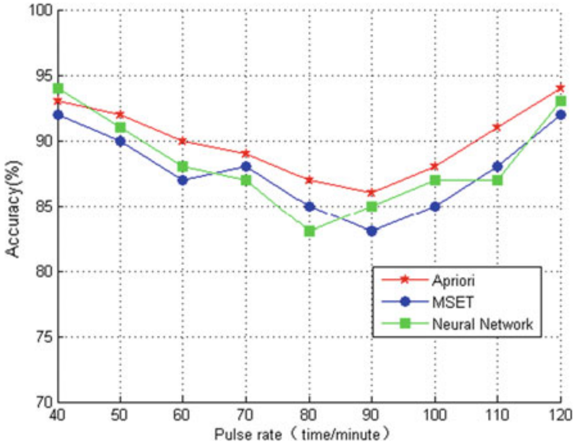


Fig. 21.8 Respiratory disease prediction accuracy at different respiratory pulse rate

21.6 Summary

This paper proposes a respiratory monitoring method based on knowledge graph. In this paper, we build the respiratory monitoring data knowledge graph and establish the relationship model between respiratory monitoring indicators and disease, digging out the strong association rules of respiratory monitoring indicators and of disease, achieving the perception of human health status analysis, so as to achieve the goal of predicting disease. In addition, this paper also designs experiments to compare with MSET algorithm and Neural Network in disease prediction. The results show that this method has high accuracy.

Acknowledgements This research was supported by the Tianjin Science and Technology Project (No. 20YDTPJC01410).

References

1. Xu, X.T.: Monitoring of human physiological signals during sleep. Med. Inf. **17**(6), 326–329 (2004)
2. Feng, Y.J.: Discussion on polysomnography and nursing intervention. Syst. Med. **2**(26), 151–153 (2017)
3. Zhao, F.C.: Research on the design and algorithm of a home sleep monitoring system. Chongqing Univ. 11–12 (2017)
4. Li, H.Y.: Application evaluation of smart bracelet. World J. Sleep Med. **1**(6), 341–344 (2014)
5. Liu, Y.: Construction of knowledge graph of medical encyclopedia. Chin. J. Med. Libr. Inf. **27**(6), 28–34 (2018)
6. Huang, M.X.: Research on entity recognition and knowledge graph construction based on electronic medical record. Res. Appl. Comput. **36**(12), 221–225 (2019)
7. Zan, H.Y.: Construction of Chinese medical knowledge graph based on multi-source text. J. Zhengzhou Univ. **52**(2), 48–54 (2020)
8. Wu, H.: Chinese named entity recognition based on transfer learning and BilstM-CRF. J. Small Microcomput. Syst. **40**(6), 1142–1147 (2019)
9. Li, G.: Entity identification of Chinese electronic medical records based on BilSTM-CRF network and dictionary resources. Mod. Intell. **40**(4), 3–12+58 (2020)
10. Li, M.L.: Study on the correlation model between air quality and chronic disease. Yangtze Univ. 59–60 (2018)

Part IV
New Approaches in Communications and Computer Technologies

Chapter 22
Modern Communication Technology-Based Optimization of Ship Remote Control Data Management Platform

Jiabao Du, Renqiang Wang, Yongqian Yang, Keyin Miao, and Hua Deng

Abstract By using Microsoft ASP.NET Web development technology, the system is designed based on B/S structure for ship remote control data management platform. The background database table of the system is designed and tested, and the test results show that the system can well meet the detection requirements. If the communication distance is less than 20 km, the early warning success rate in various test environments is higher than 95%, and the signal data transmission rate is about 0.17 s, which is less than 0.3 s, which meets the requirements.

22.1 Introduction

The ship itself is a relatively independent system, which can rely on its own reserves of energy and materials to operate independently on the water and can get in touch with the shore through communication equipment. However, it is dangerous for ships to operate independently in inland rivers or even on the sea, and the ship itself often carries relevant operators. Once the ship breaks down or accidents occur, the social impact is particularly great. Therefore, people put forward the concept of ship monitoring and management, and envisage to improve the safety of ship operation by using comprehensive technical means and optimized management mode [1]. With the warning of all kinds of ship safety accidents at home and abroad, as well as the idea of "people-oriented" put forward in our country, the industry pays more and more attention to the working conditions and environment of ship workers, so the concept and technology of ship monitoring management will always develop with the development of ships.

The automobile ferry is one of the important means of transportation across the Yangtze River. With the increasingly close economic and cultural exchanges between cities along the Yangtze River and the popularity of automobiles in China, the demand for automobile ferry is also rising [2]. However, the ferry itself has the characteristics

J. Du · R. Wang (✉) · Y. Yang · K. Miao · H. Deng
Navigation College, Jiangsu Maritime Institute, Nanjing 211170, China
e-mail: wangrenqiang2009@126.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_22

of bad working environment and aging equipment. In recent years, a series of ferry safety accidents have not only brought bad effects to the ferry company, but also brought huge losses to the society. At the same time, affected by the diversion factors of the Yangtze River Bridge, the ferry is also under greater competitive pressure. Therefore, only through advanced technology to improve the safety of ferry ships, while reducing the cost of enterprise operation and management, can we improve the market competitiveness of the Yangtze River Ferry, and achieve the goal of pursuing benefits in safety [3].

Based on the above background, a remote data management system is designed based on B/S architecture. At the same time, it makes full use of the equipment on the ferry ships, which improves the automation level of ferry ships and reduces the development cost.

22.2 B/S Structure and ADO.NET Technology

The early control system is limited by technology, usually the data to be processed and the data processing program are concentrated on the same computer, this machine is the central host of the system, the rest of the terminals in the system have no data processing ability, can only passively accept the information from the central host. The remote control between the central host and the terminal is realized by a special network, special transmission control protocol, RS-232C, GPIB, and other related transmission control protocols and corresponding hardware [4]. This closed architecture is called centralized control.

In this architecture, the central host of the system bears the pressure of data processing of the whole system. Once the central host fails, the whole system will not run normally. And the excessive concentration of functions makes it difficult for the system to expand, and at the same time, it is impossible to share information between systems. Compared with the centralized control system, the distributed system has the characteristics of resource sharing, convenient maintenance, and higher reliability, and uses the general network protocol to replace the special network protocol. At present, there are two main architectures of a distributed system: C/S and B/S [5].

22.2.1 B/S Structure

B/S structure is a special C/S structure based on HTTP protocol [6]. Among them, the presentation layer is the browser, which is responsible for providing the user visual interface and is presented in the form of browser web pages. The middle layer is the web server, which stores the system logic and connects the presentation layer and the data layer, and the data layer is the database [7]. The B/S structure separates the system logic function and the interface display function of the client in the original C/S structure, and puts the system logic function into the web server as the middle layer, which solves the problem of uneven system load in the two-tier structure. Compared with C/S structure, the advantages of B/S structure are the following [8]:

(1) Compared with C/S structure of fat client/thin server mode, B/S structure is thin client/fat server mode. The client is only responsible for the display of information, and the maintenance and upgrade of the system only need to be done on the system server. It greatly reduces the workload of system management.

(2) The data access logic layer is used to isolate the front-end web page presentation layer and the back-end data source. In this way, the presentation layer is only responsible for the presentation of data and visual elements, the logic layer encapsulates the logic code of system operation with classes, and the underlying data sources are separated by the logic layer. At the same time, confidential data such as data architecture can be further protected.

(3) Using the mature cross platform architecture such as Java and. Net, one development can be used on multiple platforms. It effectively overcomes the disadvantage of poor generality of C/S structure [9].

(4) Using browser as client, the interface is friendly and easy to operate. Users can enter the system on any computer with a browser installed, without special training.

22.2.2 ADO.NET Technology

ADO.NET is a data source access technology under the .Net environment provided by Microsoft. It supports Microsoft SQL Server data sources and access to open data sources through OLE DB and XML. When the user needs to query, modify or delete the contents in the database, the program will ADO.NET to connect these data sources and perform the corresponding operation.

22.3 System Design

The monitoring object of the Yangtze River Ferry is the ferry itself, and the main body of management is the supervision department on shore. Therefore, a complete remote monitoring and management system should include three subsystems: ship terminal system, shore terminal system, and wireless communication system between ship and shore [10].

22.3.1 Ship Terminal System

The ship terminal system mainly collects ship data and preprocesses the collected data. Therefore, the ship terminal system includes two sub modules: data acquisition module and data preprocessing module.

The data includes real-time data and input data is collected in the data acquisition module.

(1) Host monitoring data

Main engine speed, main engine lubricating oil inlet pressure, main engine cooling water outlet temperature, main engine lubricating oil temperature, main engine exhaust temperature, main engine cooling water inlet pressure.

(2) Generator monitoring data

Oil inlet pressure of generator prime mover, oil temperature of generator prime mover, cooling water temperature of generator prime mover, cooling water inlet pressure of generator prime mover, generator voltage, generator current, generator frequency, generator power, and generator speed.

(3) Rudder and propeller monitoring data

The real-time switch data of steam ferry include: pressure of rudder propeller, oil temperature of rudder propeller, and hydraulic temperature of rudder propeller.

Generator overload state, power loss state of main power supply of springboard, oil level of springboard mailbox, blockage of high pressure filter of springboard, clutch oil pressure, clutch oil temperature, clutch oil temperature, etc.

The main equipment of data acquisition submodule includes various sensors used for data acquisition, programmable logic controller (PLC) used for bottom equipment control, industrial control computer of data processing device and ups which can ensure the industrial control computer to continue to work under power failure.

22.3.2 Ship Shore Communication Subsystem

The communication system of ferry vessel is limited by its own conditions and working environment, which is different from shore wireless communication and has the following characteristics:

(1) There are many monitoring points on the ferry, and the amount of data collected is large, so the communication system is required to have greater data transmission capacity;
(2) The electromagnetic interference of various electrical equipment on the ferry is serious, and the communication system should have the good anti-interference ability;

(3) The ship terminal communication equipment should have high IP (dust and waterproof) level;
(4) The communication system should have a certain stability to ensure that the data transmission will not be interrupted in the process of moving the ferry.

In order to meet the needs of video data collection, China Mobile's 4G network card was selected as the system's network communication equipment, according to the characteristics of the ship's communication system.

22.3.3 Shore Terminal System

The main functions of the shore terminal system are remote data storage, ferry information monitoring, and management. The remote data storage function includes two aspects:

(1) Current data update. The system receives real-time data from the wireless equipment and updates the records in the database automatically. It is equivalent to the backup management of information on steam ferry.
(2) Data management. While updating the current data, the data is also added to the historical database.

There are two aspects in the information monitoring and management of the ferry company

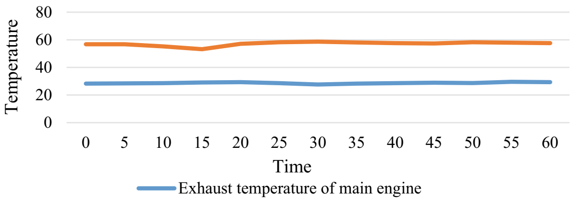
(1) Real-time monitoring of ferry equipment information. By monitoring the information of the ferry equipment, we can understand the working condition of the ferry.
(2) Safety status management of steam ferry. By consulting the safety certificate and life-saving equipment on the ferry. It can find out the hidden danger in time.

22.4 System Function Test and Analysis

22.4.1 Real-Time Monitoring Function Test Temperature

As shown in Fig. 22.1, the modified figure shows the equipment temperature real-time monitoring function, which can stably detect various temperature data and provide guarantee for temperature control. According to the data in the figure, the function can operate normally. Other functions, such as personnel registration function and data query function, will not be described again.

Fig. 22.1 Real-time monitoring interface



22.4.2 *Early Warning Function Test*

The network transmission capacity determines the real-time operation speed and stability of the system, and the early warning success rate is used to measure the reliability of the platform.

Accordingly, the distance between the terminal node and the coordination device is set to 20 km, 50 km, and 100 km, respectively (Fig. 22.2).

As shown in Fig. 22.1, when the communication distance is less than 20 km, the early warning success rate in all environments is higher than 95%; when the communication distance is 50 and 100 km, the early warning success rate in all environments is higher than 90%.

As shown in Table 22.1, the success rate of sensor data transmission remains above 90% in both indoor and outdoor environments. Even in the case of 20 km from the coast, the transmission success rate can be as high as 97%.

In future actual engineering projects, the quality and efficiency of sensors can be improved to ensure the success rate of transmission and corresponding reliability.

Fig. 22.2 Communication success rate with different environments

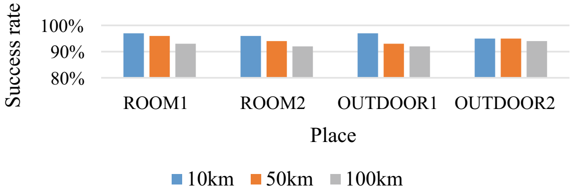


Table 22.1 Functional testing

Function	Functional description	Testing
Data storage	Backup and storage of ship personnel, equipment, and other data	Pass
Data deletion	Deletion of data on ship personnel, equipment, etc	Pass
Data registration	Registration of data on ship personnel, equipment, etc	Pass
Navigation	Tracking of ships' navigation	Pass

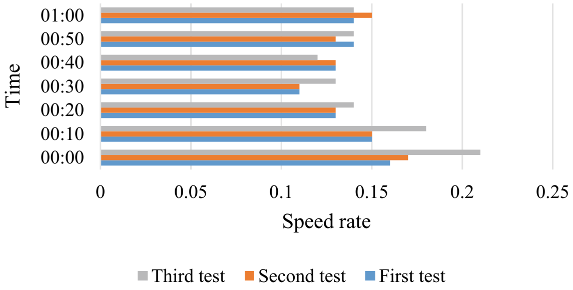


Fig. 22.3 Platform data transmission rate test (20 km)

Table 22.2 Communication success rate in different environments

	100 m (%)	200 m (%)	300 m (%)
ROOM1	97	96	93
ROOM2	96	94	92
OUTDOOR1	97	93	92
OUTDOOR 2	95	95	94

Table 22.3 Platform data transmission rate test (20 km)

	First test	Second test	Third test
0:00	0.16	0.17	0.21
0:10	0.15	0.15	0.18
0:20	0.13	0.13	0.14
0:30	0.11	0.11	0.13
0:40	0.13	0.13	0.12
0:50	0.14	0.13	0.14
1:00	0.14	0.15	0.14

22.4.3 Transmission Rate Test

As shown in Fig. 22.3, the test was conducted on the basis of a distance of 20 km from the shore. In terms of rate, the transmission rate tends to be stable with the increase of time. Through three tests, it is found that the transmission rate is basically stable at about 0.17 s, meeting the basic requirements of less than 0.3 s. The specific data are shown in Table 22.2 (Table 22.3).

22.5 Conclusion

Based on the actual background of the research, this paper fully studies the safety factors of the ferry ship and the safety management dilemma faced by the ferry company. The design goal, design principle, and feasibility of developing a safety management system for Qidu company are analyzed. The physical architecture and web architecture of the system are designed, and the functional modules of the system are determined according to the needs of users. At first, the background database table of the system is designed and tested, and the test results show that the system can well meet the detection requirements.

References

1. Liu, W., Liu, S., Zhao, J., et al.: A remote sensing data management system for sea area usage management in China. Ocean Coast. Manag. **152**(2), 163–174 (2018)
2. André, O. de Araujo, Filho, M.R., Pires, F.C.M.: A ship construction cost data management system. Mar. Syst. Ocean Technol. **12**(3), 93–103 (2017)
3. Design of grassland multi-source remote sensing data management system. Geomat. Sci. Technol. **07**(4), 186–194 (2019)
4. Wyant, M.C., Bretherton, C.S., Wood, R., et al.: Global and regional modeling of clouds and aerosols in the marine boundary layer during VOCALS: the VOCA intercomparison. Atmos. Chem. Phys. **15**(1), 153–172 (2015)
5. Benjamin, B., Paul, A., Böhner, J., et al.: Mapping local climate zones for a worldwide database of the form and function of cities. Int. J. Geo-Inf. **4**(1), 199–219 (2015)
6. Ozdenizci, B., Ok, K., Coskun, V.: A tokenization-based communication architecture for HCE-enabled NFC services. Mob. Inf. Syst. **5**, 1–20 (2016)
7. Chai, W.T., Xiao, B.J., Yuan, Q.P., et al.: The design of remote discharge scenario management system on EAST. Fusion Eng. Des. **112**(11), 1049–1054 (2016)
8. Dhamaniya, A., Sonu, M., Krishnanunni, M., et al.: Development of web based road accident data management system in GIS environment: a case study. J. Ind. Soc. Remote Sens. **44**(5), 1–8 (2016)
9. Wiersma, R., Grelewicz, Z., Belcher, A., et al.: SU-E-T-11: a cloud based CT and LINAC QA data management system. Med. Phys. **42**(6), 3326–3333 (2015)
10. He, D., Kumar, N., Zeadally, S., et al.: Certificateless provable data possession scheme for cloud-based smart grid data management systems. IEEE Trans. Ind. Inf. **99** (2017)

Chapter 23
A Cognitive Radio Adaptive Communication Platform

Li Miao

Abstract Using ALTERA's DE3 FPGA development board, a cognitive radio voice signal adaptive communication platform was designed and developed. FPGA mainly completes channel information feedback, adaptive communication, and digital signal processing functions; voice chip mainly completes voice signal acquisition and simulation Digital/digital-analog conversion function. The experimental results show that the cognitive radio communication platform can adaptively adjust parameters to adapt to the channel state, verify the effectiveness of the adaptive coding and modulation algorithm, and have good channel feedback and adaptive communication performance.

23.1 Introduction

As the human society enters the information age, under the constant promotion of new technologies and demands, communication technology is developing rapidly and has put forward higher requirements for the reliability of communication. The adaptive modulation and coding (AMC) technology in cognitive radio provides a new idea for communication anti-interference. Adding an adaptive coding and modulation module to the digital voice communication system can intuitively verify the anti-interference performance of the system through the transmission quality of the voice signal.

In the literature [1], the author briefly studied the adaptive connection scheme using constant power and adaptive modulation and coding (AMC) scheme to ensure the bit error rate (BER) requirements, the research background is the wireless fading interference propagation channel. Literature [2] introduced the cooperative AMC technology with SINR (Signal to Interference plus Noise Ratio) as the threshold to realize the cognitive radio network architecture.

L. Miao (✉)
Information Science Academy of CETC, Beijing, China
e-mail: limiao_cetc@163.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_23

This article designs and implements a cognitive adaptive communication platform based on FPGA and TLV320AIC23 voice chip, and uses ALTERA's DE3 development board and Stratix3 EP2SE260F1152C2 chip to verify the system's adaptive communication function.

23.2 Adaptive Communication Scheme and the Selection of the Hardware Set

The scheme of the cognitive radio adaptive communication platform based on voice signals is shown in Fig. 23.1. It is mainly composed of the following parts: voice collection part (microphone), dedicated voice chip, adaptive coding and modulation communication module, channel (including interference and noise), voice output part (speaker). Among them, the data buffer and adaptive coding and modulation are completed by FPGA. At the transmitting end, the microphone collects the analog voice signal, performs A/D conversion of the analog voice signal through the voice chip, and sends the signal to the channel for transmission after the signal is encoded and modulated through the FPGA. The receiving end goes through the reverse process and receives the sound of the sending end through the speaker.

Among them, the microphone collects the voice signal as the source, processed by the voice chip, buffered, coded and modulated, and sent to the channel. The receiver completes the real-time signal-to-noise ratio (SNR) estimation, and selects the coding and modulation method to be sent at the next moment according to the threshold list, and feeds it back to the sender. At the same time, the receiving end performs demodulation and decoding, restores the voice signal through the voice processing chip, and sends it out by the speaker.

It can be seen from Fig. 23.1 that the main functions of the adaptive communication platform are realized by FPGA, the core of which is the adaptive coding and modulation strategy and the real-time estimation algorithm of SNR. The system

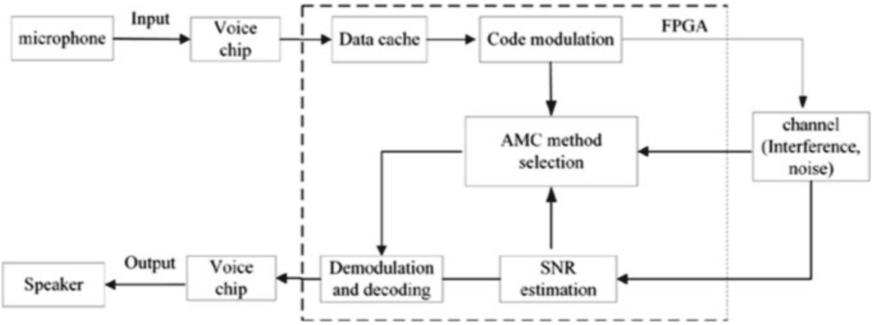


Fig. 23.1 Cognitive radio adaptive communication platform composition block diagram

communication frequency is 50 MHz, and the voice signal acquisition rate is 128 kbps. The symbol rate is 256 kbps. The clock of the modem module is 16 MHz.

The Altera DE3 development board is embedded with the Stratix III series FPGA chip EP2SE260F1152C2, which contains 254.4 K logic units, 14,688 K memory, 768 18 × 18-bit multiplier modules, and a 50 MHz crystal oscillator external to the FPGA chip [3]. The DE3 development board and FPGA chip resources meet the index requirements of this system.

TLV320AIC23 (hereinafter referred to as AIC23) is a high-performance stereo audio Codec chip launched by TI. The analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC) components of AIC23 are highly integrated into the chip and can support sampling rates from 8 to 96 kHz. The digital transmission word length can be 16 bit, 20 bit, 24 bit and 32 bit, suitable for the cognitive radio adaptive communication platform.

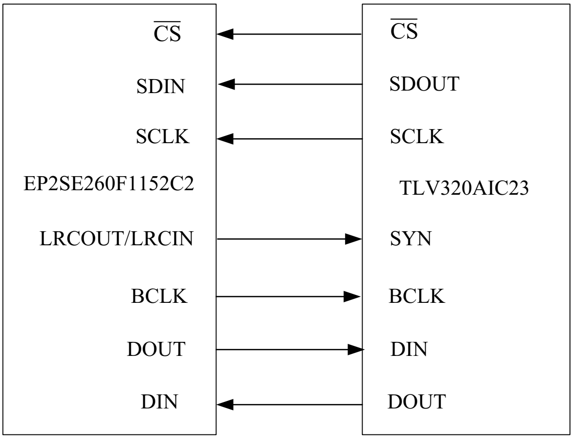
23.2.1 Voice Chip Configuration

There are two interfaces between FPGA and AIC23: digital audio control interface and data communication interface [4], which complete the setting of the working parameters of the voice chip and the transmission of audio data, respectively. The block diagram of the connection between FPGA and AIC23 is shown in Fig. 23.2. The following describes the configuration of these two interfaces, respectively.

The AIC23 control interface has two working modes: I2C and SPI. When MODE is low, it works in I2C mode; when MODE is high, it works in SPI mode [3].

In this article, the AIC23 control interface selects the SPI mode. The internal registers of the AIC23 can be configured through the FPGA control interface to change its working state, and the Verilog language programming is used to realize the control of the voice chip. FPGA outputs to AIC23: chip selection signal, clock

Fig. 23.2 TLV320AIC23 and FPGA interface



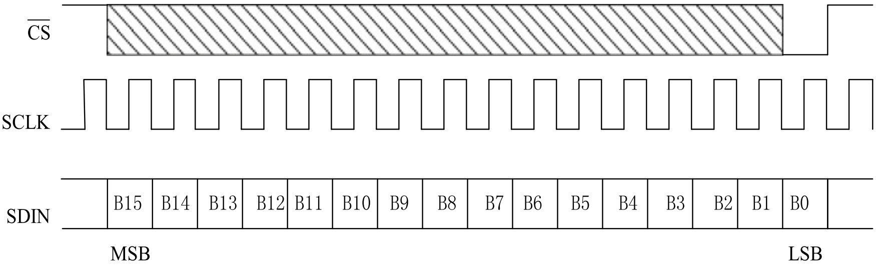


Fig. 23.3 Timing diagram of SPI mode

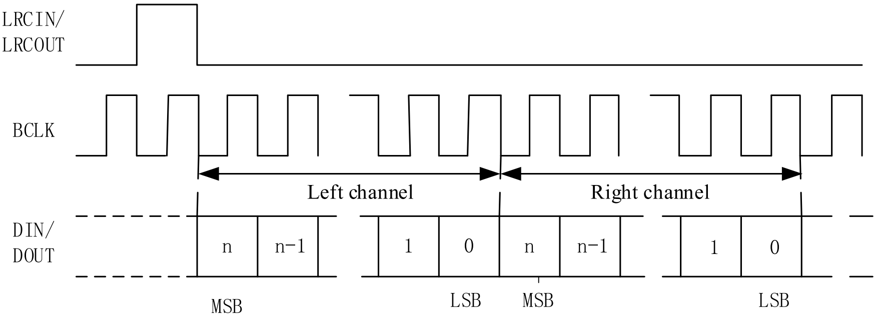


Fig. 23.4 Timing diagram of DSP mode

signal SCLK, control register input data SDIN [4]. The SPI mode sequence of AIC23 [4] is shown in Fig. 23.3.

AIC23 supports 4 audio data formats: left-justified mode, right-justified mode, I2S mode, and DSP mode. We choose DSP mode in this system scheme. The digital audio interface includes output clock BCLK, data input and output DIN and DOUT, and synchronization signals LRCIN and LRCOUT [4].

AIC23 begins to transmit data on the falling edge of the sync signal, and DIN and DOUT begin to change at the same time on the falling edge of the clock signal BCLK. The timing diagram of the DSP mode is shown in Fig. 23.4.

23.2.2 Adaptive Modulation and Coding Function Module

Through the introduction of LDPC channel coding technology, the adaptive modulation technology is combined with the channel coding technology, and the system reliability and anti-noise performance are improved by using redundant coding and channel coding with strong error correction capabilities [5]. The adaptive modulation and coding technology can adaptively select the appropriate modulation and coding scheme (MCS) according to the channel conditions, and maximize the use of spectrum resources under the condition of limited power [6].

The anti-noise and interference capabilities of different modulation methods can be expressed by their error performance in the additive white Gaussian noise (AWGN) channel [7]. The coherent demodulation error performance function of different modulation methods in AWGN channel can be expressed as

$$\begin{aligned}
P_{BPSK}(\gamma) &= Q(\sqrt{2\gamma}) \\
P_{QPSK}(\gamma) &= Q(\sqrt{\gamma}) \\
P_{16QAM}(\gamma) &= \frac{1}{4}\left[Q(\sqrt{\gamma/5}) + Q(3\sqrt{\gamma/5})\right] + \frac{1}{2}Q(\sqrt{\gamma/5}) \\
P_{64QAM}(\gamma) &= \frac{1}{12}\left[Q(\sqrt{\gamma/21}) + Q(3\sqrt{\gamma/21}) + Q(5\sqrt{\gamma/21}) + Q(7\sqrt{\gamma/21})\right] \\
&\quad + \frac{1}{6}\left[Q(\sqrt{\gamma/21}) + Q(3\sqrt{\gamma/21})\right] + \frac{1}{12}\left[Q(5\sqrt{\gamma/21}) + Q(7\sqrt{\gamma/21})\right] \\
&\quad + \frac{1}{3}Q(\sqrt{\gamma/21}) + \frac{1}{4}Q(3\sqrt{\gamma/21}) - \frac{1}{4}Q(5\sqrt{\gamma/21}) - \frac{1}{6}Q(7\sqrt{\gamma/21}) \\
&\quad + \frac{1}{6}Q(9\sqrt{\gamma/21}) + \frac{1}{12}Q(11\sqrt{\gamma/21}) - \frac{1}{12}Q(13\sqrt{\gamma/21}) \qquad (23.1)
\end{aligned}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-t^2/2}dt = \frac{1}{2}\left[1 - erf\left(x\Big/\sqrt{2}\right)\right] \qquad (23.2)$$

In the above formula, γ represents the signal-to-noise ratio per bit, t represents communication time.

This cognitive radio adaptive communication platform uses LDPC coding with a code rate of 1/2, and four modulation methods of BPSK\QPSK\16QAM\64QAM, which form a total of 4 modulation and coding schemes (MCS), as shown in Table 23.1. For the voice signal transmitted in this scheme, based on the BER 10^{-3} as the threshold, the mapping relationship between MCS and SNR interval is established according to the SNR-BER curve obtained by MATLAB simulation. By real-time estimation of the SNR at the receiving end, the corresponding modulation and coding combination scheme is selected according to the interval of the estimated SNR, and the MCS scheme is fed back to the transmitting end as the modulation and coding method used for the next communication.

Table 23.1 Modulation and coding scheme (MCS)

MCS	LDPC code rate	LDPC code length	Modulation
MCS1	1/2	1152	BPSK
MCS2	1/2	1152	QPSK
MCS3	1/2	2304	16QAM
MCS4	1/2	2304	64QAM

23.2.3 Experimental Results of the Cognitive Radio Adaptive Communication Platform

The Tektronix AFG3235 arbitrary function generator is used to generate Gaussian white noise, and the Agilent ESG4438C signal generator is used to generate interference signals. The two are loaded into the channel through a combiner and filter to simulate the actual channel with interference and noise.

The working process of the whole system is as follows: the voice is collected by the microphone as the input signal, sampled by the TLV320AIC23 chip, the data is input to the FPGA chip for buffering, encoding, modulation, and sending to the channel. The SNR estimation, signal demodulation, and decoding are completed at the receiving end, output to the TLV320AIC23 digital voice chip to restore the voice, and output the voice through the power amplifier, and according to the interval of the estimated value of the SNR at this time, adaptively select the next modulation and coding scheme, and complete the test process of the voice signal.

23.3 Conclusions

This paper designs and develops a cognitive radio adaptive communication platform based on FPGA and dedicated voice chip TLV320AIC23, and verifies the platform on the DE3 development board with Stratix3 EP2SE260F1152C2 chip as the core.

This paper presents an adaptive coding and modulation system model, implements FPGA hardware implementation of the AMC system, gives the implementation scheme of baseband coding and modulation, and the implementation process of key modules and hardware simulation results to verify the adaptive coding and the effectiveness of the modulation algorithm.

Experimental test results show that when there is interference and noise in the channel, the voice signal transmission quality will be greatly affected. After the cognitive radio adaptive modulation and coding module, the system can adaptively adjust the communication parameters according to the interval of the SNR. It ensures that the voice signal is complete and clear, with high resolution, and the system's BER requirements are guaranteed.

References

1. Webb, W.T., Steele, R.: Bandwidth-efficient QAM schemes for Rayleigh fading channels. Part I **138**(3), 169–175 (1991)
2. Steele, R., Webb, W.T.: Viable rate QAM for data transmissions over mobile radio channels. Keynote paper, Wireless '91, Calgary, Alberta (1991)
3. Altera DE3 Prototyping System User Manual
4. Texas Instrument Inc. TLV320AIC23B Stereo Audio CODEC 8-to 96-kHz, With Integrated Headphone Amplifier. 2004. 2
5. Haykin, S., Thomson, D.J., Reed, J.H.: Spectrum sensing for cognitive radio. In: Proceedings of the IEEE, pp. 849–877 (2009)
6. Vartiainen, J., Lehtomaki, J., Saarnisaari, H., et al.: Estimation of signal detection threshold by CME algorithms. In: Vehicular Technology Conference, pp. 1654–1658 (2004)
7. Khalaf, Z., Nafkha, A., Palicot, J., et al.: Hybrid spectrum sensing architecture for cognitive radio equipment. In: Telecommunications (AICT), 2010 Sixth Advanced International Conference. IEEE Explore, pp. 46–51 (2010)

Chapter 24
Billing System and 5G Network Slicing Service

Zhenfeng Gao and Guodong Shan

Abstract By referring to the international mainstream operational support models such as eTOM (Enhanced Telecom Operations Map) and NGOSS (Next Generation Operations System and Software), combined with the iterative development of large domestic projects and the experience of the new system construction of emerging telecom carriers, the enterprise service model and data model are transformed into the enterprise system architecture for the telecom full-service. Through the summary of the requirements of the system in 5G application scenarios, the transformation scheme and data structure design are proposed under the 4G deployment network. In the development of the new system, problems are solved such as the 5G service-launch for users, 4G VoLTE (Voice over Long-Term Evolution) and NSA (Non-Standalone) users service-launch in the SA (Standalone) area, the support of 5G slice scene, and the automatic service-launch of SA by default.

24.1 Introduction

In the age of 5G (the 5th Generation Mobile Communication Technology) communication, immense changes have taken place in the network architecture of the core network and the bearer network. The 5G communication core network adopts the SBA (Service-based Architecture) micro-service architecture. Compared with the communication network in the traditional mode. A large number of NF (Network Functional Entities) are set in the 5G network architecture. And all the NF (Network Functional Entities) is fully capable of independent operation and automated management. NF (Network Functional Entities) of all its configuration can be fully combined with the actual needs of the business to achieve mutual communication. The significant evolution of the network system has greatly increased the difficulty of data collection and checking processes of the business operation support system. In order

Z. Gao (✉) · G. Shan
State Grid Information & Telecommunication Co., Ltd., Beijing Branch, Beijing 100052, China
e-mail: gzfmyh@126.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_24

to elaborate control of different businesses, a large number of new software interfaces are needed. Compared with the 4G (the 4th generation mobile communication technology) technique, the 5G network usually carries out the application of many advanced technologies such as network function resource pooling, edge computing, network slicing, etc. Therefore, it has made a big leap in the scalability of the overall network architecture and the possibility of service bearing. The 5G network will mainly apply three application scenarios, and the actual distribution control in different scenarios is also quite different. The user support system for all nodes in the entire network needs to timely collect data, stepwise check and uniformly processing, underlying support for various business scenarios.

Distribution in the supporting system of each module component is the key to build the business support system. Therefore, the business support system construction must sufficiently protect the whole system, and architecture need to be attributed to rationality, stability, reliability, and continuity. The essential for promotion of 5G network service is upgrading network service with low-cost based on the current business system. Meanwhile, the further mining potential of the business operation support system have become pivotal to the telecom carriers' impetus for the innovation development. Though further developing the subsystems of the business system, the construction cost and time cycle can be significantly reduced.

24.2 Billing System

In the process of construction of billing model, the functional areas of the system need to be well planned from the perspective of the top-level design, within clearly defined the connection between functional entity and the body of data. Taking the model as the core element by using the body of data as the benchmark carrying out the top-down processing system design. Meantime, the content decomposition of different types of the body data implements further refinement of data with the ending of the whole model construction.

24.2.1 The Design of Billing

In order to provide accurate supporting billing system or application for the whole service of telecom, the data of CDR (Call Detail Record) will be applied into the billing processes flow as information acquisition, pricing, accounting, and billing for users in service and have management functions such as customer's expense check, expense warning, and arrears settlement.

24.2.1.1 Pricing Approval

Pricing approval is calculated based on pricing plans, usage of repetitive costs, non-recurring expenses, and various discount rates. The proper pricing rate must refer to the users' basic personal information while calculating.

Price approval refers to the CDR (Call Detail Record) bill information of the usage events such as voice, short message, traffic, value-added services, etc., which is queried according to the content of the contract and the accumulated cost. Through the validity review and standardization of the CDR (Call Detail Record) bill, determine the corresponding rate standard, and calculate the total cost of the combined bill according to the rate standard. Finally, the approved bill will be sent to the account processing module. For the abnormal data found in each step, the data can be traced back to the normal use event, so that the price processing of the bill can be carried out again. At the same time, the calculation results of the phone bill should be synchronized with the CRM (Customer Relationship Management) system. The approval price shall include the following contents:

(1) Maintenance Rate: to ensure the integrity of the rate pricing system, and provide the rate modification function.
(2) Real-time price approval: track the real-time CDR (Call Detail Record) of users, and price each calling record according to the contract content.
(3) Multiple price approval: different pricing strategies applied for the same batch of calling detail records in multiple price approval.
(4) Pre-authorization of deduction: with the real-time rating function, it provides billing service authorization for the account level balance (prepaid users) or credit level (postpaid users) of the specified account.
(5) Discount rate of pricing: The function to discount for one calling event based on special attributes in the pricing process.
(6) Repeat price approval: Price approval is repeated once the records are refused.

24.2.1.2 Billing and Bill Generation

Preferential management: confirm, combine, calculate, and distribute all the preferential terms; This includes cross-product offers, prioritization for multiple offer applications, loyalty programs across billing cycles, etc.

(1) Pricing and Fee Calculation: Support error recovery and rebilling; It must be able to support the processing of service quality and partner data, including hierarchical revenue structure, income distribution, compensation, etc. [1].
(2) Charging on demand: According to the actual requirements of the front end, we can combine the costs of specific customer accounts and deal with one-time discounts.
(3) Refunding: The process of recalculating the bill data involved according to the conditions of refunding. It is a means of correcting erroneous data for various reasons.

(4) Promotion Management: Define and manage all typical discount programs and additional attributes, such as rental fees, minimum spending, service activation and discontinuation fees, and fixed periods. A product portfolio that can achieve the target customer group of a promotion program. Promotions are sometimes used for fee bundles for a limited period of time (separate from discount plans).

(5) Tax Management: To manage the tax levied by the government on services provided by operators. Support different levels (region, country) for multiple types of taxes (e.g., telecom, sales, value-added tax).

(6) Bills generation: can generate bills in standard format according to all charges.

(7) Distribution: bill will be able to print bill delivery format data sent to the bill for printing or file transfer, will display the billing format data written to the file to provide online, according to the feature extraction and distribute formatted data to a media/equipment, such as the invoice printer, electronic display terminals, and charging management interface.

(8) Post-billing data distribution: Distribution of billed data to ERP (Enterprise Resource Planning), commission settlement or other external applications via online or batch interfaces.

(9) Messages and Insert Messages: Provides the ability to insert marketing or promotional messages and graphics into a bill, providing a dynamic way to insert information according to the needs of a customer group, and to insert information according to a particular customer type or region. Insertions can also be triggered by events, such as late payment or new account opening.

(10) Billing Backup: Backup the customer's historical billing data.

(11) Reports and Reports: Includes report generation, browsing, and printing of a complete set of off-the-books reports. These reports are mainly related to the summary of financial data, customer service data, payment data, etc., in addition to some useful system reports for system administrators, such as: server status, responsible for the control report of the batch process of the billing function.

(12) Hierarchy Billing: Calculates total expenses and subtotals at the same level below for corporate customers, applies company-wide discounts, generates cost center reports and aggregates billing, and is able to break expenses down to different accounts at different levels based on business rules set up.

(13) Refunds: Calculate the total amount refundable during the billing process.

24.2.2 Overview Flow of Billing and Accounting

Description of the overall process of billing accounting is shown in Fig. 24.1 (Bill DB (Bill DataBase), CRM (Customer Relationship Management), CDR (Call Detail Record)):

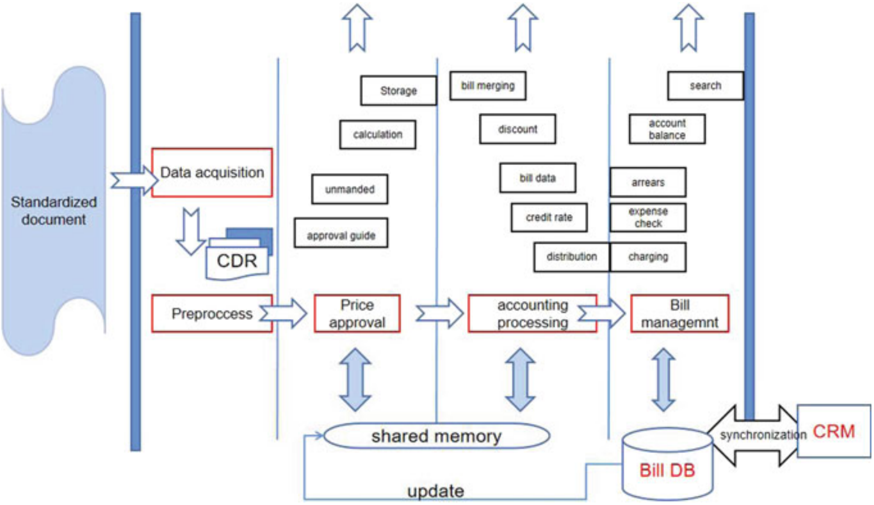


Fig. 24.1 Overview flow chart of billing and accounting

(1) The data acquisition module of the billing accounting system is responsible for collecting billing phone bills, and synchronizes the data to the billing preprocessing module while backing up the users' bill data.

(2) After the original data is collected, it is formatted and recognized through the preprocessing module, and classified according to the consistency of the data. Finally, a record format file is formed according to the unified format and relevant standards.

(3) The approval module of the billing accounting system calculates the cost of the user's bill according to the package configuration and forms a detailed list of the user.

(4) The accounting processing module of the billing accounting system completes the process from approving bills to generating comprehensive bills for users, including functions such as closing accounts, real-time billing, real-time credit control, monthly billing, and so on.

(5) The charging module of the billing accounting system realizes the functions of collecting user fees, displaying bills, and printing bills.

24.3 Design of 5G Slicing

Once the slices of the 5G network need unified operation and management, different telecom service management functions will be the system foundations. Coordinating the core network, access network, and all kinds of basic networks and systems, users need to provide the exclusive and custom network logic. The application of network slicing belongs to a brand new business model, both for operators and users. The

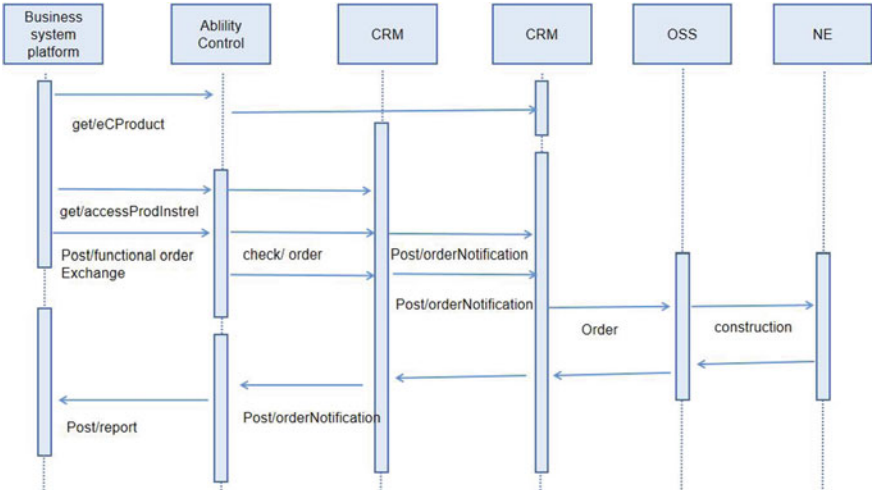


Fig. 24.2 Flow chart of 5G slicing

flow of 5G slicing is shown in Fig. 24.2 (OSS (Operation Support Systems), NE (Net Element)).

The later operation and maintenance process of network construction will still involve a large number of network slicing modules and a high degree of complexity of template arrangement strategy, collaborated between different networks (4G and 5G) with compatibility of heterogeneous devices.

In the 5G era, user demands in various segmented industries also present personalized characteristics. Taking full advantage of network slicing technology can bring unprecedented challenges to the construction of a network and operation support network for the entire technology center, considering network resources allocation.

(1) Unique section: the section identification ID is generated by CRM (Customer Relationship Management) in accordance with international standards. Unique section identification identifies a section, and relevant systems need to do handling.
(2) Slice management: CRM (Customer Relationship Management) system uniformly manages enterprise slices, forms the CRM (Customer Relationship Management) point acceptance, and generates the slice identification ID, which is sent to each network side through the network operating system for launching. CRM (Customer Relationship Management) or business platform can accept slices by calling the acceptance ability; after successful acceptance by CRM (Customer Relationship Management) system, slice instance data will be issued;
(3) Member management: the CRM (Customer Relationship Management) system uniformly manages enterprise-level slice members, and the business platform calls the CRM (Customer Relationship Management) system's acceptance

capacity for membership inclusion or withdrawal; ensure that members are incorporated into a slice consistent with the enterprise.

(4) Members can be included only after the sections are accepted successfully. Similarly, before the slicing service is disassembled, the full number of members need to exit the slicing.

24.4 5G Network Service

For 5G service, end-to-end, custom-made on-demand, and isolation are its most important characteristics in the practical application process. So, the suspect mainly refers to the application process of network deception that does not need to be equipped with the core network. In addition, it also needs to involve the access network transmission network to set up some management network; The so-called on-demand customization mainly refers to the development of slices that can be fully combined with the failure of functions, business, quality of service, and connection relations, and at the same time, it can be managed according to the requirements of the whole life cycle of slices. As for its isolation, it mainly includes security isolation and resource isolation.

At present, the application process of starting point mainly includes the main application scenarios of EMBB (Enhanced Mobile Broadband), MMTC (Massive Large Connection) and URLLC (Low Latency and High Reliability). At the beginning, the 5G network mainly provides EMBB (Enhanced Mobile Broadband) scenes and related service types, among which augmented reality and high-definition video port are the most typical application types. This part of the application time is to put forward a high demand for bandwidth [2].

For URLLC (Low Latency and High Reliability), its most typical application scenarios mainly include industrial control, UAV (Unmanned Aerial Vehicle) control, and even intelligent control. This type of scenario is usually only for sensitive services, so it must be required that its application scenario can demonstrate high reliability.

Under the premise of no change of card or number, when 4G or 5G NSA (Non-Standalone) users change 5G SA (Standalone) terminals, they will automatically launch 5G SA service by default [3, 4]. The UDM (Unified Data Management) interface of the IT (Information Technology) system needs to be extended [5]. The IT system can receive the request of 5G SA (Standalone) automatic launching sent by the UDM (Unified Data Management) and trigger the SA automatic silent opening process. The user service launching process is shown in Fig. 24.3 (gNB (generation Node B), 5G C (the 5G core network), eNB (eNodeB), EPC (Evolved Packet Core)).

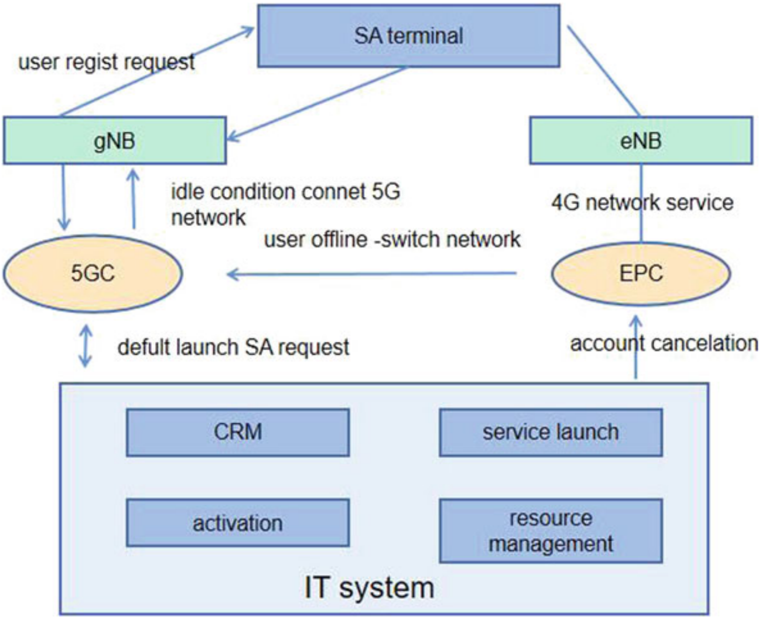


Fig. 24.3 Schematic diagram of the user service process startup process

24.5 Conclusion

From the perspective of the whole life cycle of software, in addition to the substantial innovation of communication technology, the update or adjustment of the system is mainly due to the change of requirements. Through the overall architecture design, function planning, and business logic analysis of the business support system, this paper avoids the unclear requirements and inconsistent data in the construction process of the system, At the same time, it also avoids launching various island systems to support new requirements, which not only prolongs the life cycle of software functions, but also forms a standard capability service unit. In the same software life cycle, it effectively improves the input–output ratio of software and reduces unnecessary waste in the development process. Through the overall consideration of the whole business function, the unified consideration of data structure, and the overall planning of logical function, this paper ensures that each functional module has unit output capability.

In this paper, based on a real system design and implementation, using the operation support system module of existing support for the business logic, and combines the 5G business characteristics and puts forward new requirements to support system, proposed to improve the functional architecture, data structure and the function design. The loose coupling of module functions between systems is realized by means of micro-service architecture. It not only meets the normal development of the existing telecom services, but also meets the requirements for the launching and billing of the service support system of 5G networks. Moreover, it improves the overall development speed, operation and maintenance efficiency, and system availability by greatly reducing the system development cost. It can provide upgrading experience for enterprises that have already built BOSS (Business & Operation Support) system and construction ideas for enterprises that have not built BOSS (Business & Operation Support) system. On this basis, it is expected to provide information support for the construction of telecom enterprises' full-service operation support system.

References

1. Jiajia, W.: Research on Integration Scheme of Billing Accounting System of Beijing Unicom, Beijing University of Posts and Telecommunications (2011)
2. Chufeng, W., Yuan, D.: 5G experimental scale network strategy and network construction scheme. In: Telecommunication Science, pp. 141–149 (2019)
3. Lanfang, W., Wentao, D., Yongqian, L., Yizhou, H.: Mobility management scheme under the background of NSA networking co-construction and sharing. In: Design Technology of Posts and Telecommunications, pp. 68–74 (2020)
4. Jun, L.: Evaluation and analysis of 5G NSA and SA networking scheme. In: China Telecom Express, pp. 8–12 (2020)
5. Yi, L., GuangHai, L., GuoPing, X., QingLiang, L.: Research on interoperation strategy between NSA and SA mode in 5G network. In: Posts and Telecommunications Design Technology, pp. 55–60 (2020)

Chapter 25
Location Models for Public Healthcare Facilities in India

Manoj Panwar and Kavita Rathi

Abstract Healthcare facility locations are decided in response to demography, geography, existing health care and its situation, socio-economic, cultural and political conditions, and disease pattern. Public healthcare facility locations in India are made to take advantage of political fronts. The application of scientific location-allocation models helps in achieving both the real objectives of the public healthcare system as well as political benefits. The paper presents the factors deciding the location of a healthcare facility, application of facility location models in general, and on the healthcare system and basic model from discrete facility models for direct use with minimum input required for decision-making for healthcare facility locations. The result of the application of the heuristic model shows efficiency in comparison to the existing practice of decision-making. The papers conclude with the benefits of using the location-allocation models in the healthcare system.

25.1 Background

Facility planning consists of two components: facility location and facility design. The static component, i.e., location of the facility has been continuously challenged by various planning theories and location-allocation models from the inception of facility location problems because of continues changes in the dynamic nature of demand and supply. The temporal and spatial dynamism of social/public facility demand for optimum location has to be satisfied for present and future, keeping the envelope and supporting infrastructure facility as static. Health facilities are the main pillar of the social infrastructure. Healthcare facilities are developed in response to demography, geography, existing health care and its situation, socio-economic, cultural and political conditions, and disease pattern. The basic facility of health is complicated as it is difficult to cover the entire population facilities as it is not uniformly distributed. Lower access and inferior quality of service adversely

M. Panwar (✉) · K. Rathi
Deenbandhu Chhotu Ram University of Science and Technology, Murthal 131039, India
e-mail: manojpanwar.arch@dcrustm.org

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_25

Table 25.1 Population norms for provision of the primary healthcare system in India

Level of healthcare facility	Population norms	
	Plain area	Hilly/Tribal/Difficult area
Sub-Centre	5000	3000
Primary Health Centre	30,000	20,000
Community Health Centre	120,000	80,000

Source Ministry of Health and Family Welfare, Government of India [5]

affect health outcomes [1]. Physical accessibility of public healthcare facilities influences the use of service, especially in rural areas [2]. The variation in the quality of service provision by the same level facilities impacts the proximity criteria of using a health facility [3]. The provision of basic health infrastructure improves the health conditions of the population in lower economic strata in comparison to specialized services. Reliability and good access to healthcare facilities encourage people to pay for health services [4].

The healthcare systems in India do not bring about a systemized system for spatial planning. The location problems in India have been studied for a single level of facilities only with population count as the only factor for decision-making (Table 25.1).

25.1.1 Classifications of Healthcare Facilities in India

The healthcare facilities in India exist as an amalgamation of multilevel hierarchical systems. Based on specialization and level of care, health care has been classified as Primary, Secondary, and Tertiary care. Based on function and the coverage of the population, healthcare systems have been classified as Health Posts, Health Centre, District Hospital, Regional Hospital, and Central Hospital (the nomenclature changes from state to state). Based on the ownerships, healthcare systems have been classified as public healthcare facilities, private voluntary healthcare facilities, private commercial healthcare facilities, and public–private partnership modules. The objective of a healthcare facility is more dependent on ownership classification and plays a vital role in the location-allocation of the healthcare system for a region. With the Hierarchical systems as the basis of classification, the healthcare systems can be classified as Coherent hierarchical system, Nested hierarchical system, Non-nested hierarchical system, Referral hierarchical system, and Non-referral system. The healthcare system in India is extremely complex, having all types of hierarchical systems available in the form of private healthcare providers and public healthcare systems.

25.1.2 Factors Affecting the Location of a Healthcare Facility

Planning commission of India identified “Universal access and access to an adequate level without an excessive burden; fair distribution of financial costs for access and fair distribution of burden in rationing care and capacity and a constant search for improvement; training providers for competence empathy and accountability, the pursuit of quality care and cost-effective use of the results of relevant research; and special attention to vulnerable groups such as children, women, disabled, aged, etc.” as four criteria in the healthcare system, which makes healthcare system ideal [5].

The need/demand of the population helps in defining the number of healthcare facilities. Existing facilities in the region (public/private) help in need identification of more facilities. Trends in Region 3D’s Disease, Deaths, and Disasters, help in determining the kind of facilities required for that region, i.e., the speciality of the facility. The socio-economic level of the population served, and geographical setting (urban or rural) define the scale of economies. Approaches/transportation infrastructure decides the accessibility of facility and temporal distances. Availability of other resources like water and disposal services helps in the identification of other infrastructure facility requirements. Formal affiliations with other hospitals and referral patterns decide the hierarchal system. Availability of well-qualified staff impacts the service provision at all facilities of equal level. Functional and cultural needs of the community and climate help in the identification of disease patterns. All these factors affect the location of a healthcare facility and decide the type of facility.

25.1.3 Application of Location-Allocation Models

A large number of location-allocation models have been used for optimizing the benefits and resources in various disciplines, including healthcare facility planning. The location-allocation models use multi-criteria decision-making techniques, which include Linear Programming, Dynamic Programming, Analytical Hierarchical Processing, Game Theory, Branch and Bound, and Multi-objective Multi-criteria [6, 7]. Many researchers aggregated several objectives into one objective function. The healthcare facility location problem is similar and has been attempted by using multi-criteria decision-making. Heuristics and Meta-heuristics are other techniques developed for finding solutions to georeferenced problems [7]. The demographic, economic, and geographic variations led to the development of heuristics, which require geography-specific details on various parameters [8, 9]. Despite all the studies on location models involving a set of constraints and different objective functions, all the models developed all over the research world are having the same objective of achieving the best location by using the three basic terms (accessibility, availability, and adaptability). The way of defining an objective and the way to achieve these three set goals with the set of constraints keep on changing trying to get closer to the real-world scenarios modelling. The accessibility model considers access to care to be affordable with the right healthcare resources in the right place at the right time. Availability models even accounted for facilities, beds, doctors, healthcare staff, vehicles, and supporting facilities. Such models have been used for fixing ambulance locations. Adaptability models recognize the future conditions to find the solutions to a range of future scenarios [6, 10, 11]. Other models have also been developed, but most are based on these basic models or are heuristic models like the Mayope algorithm [12] and Neighbouring search algorithm [13] but related to a particular geographic area.

All these models remain complex to be understood by the common public, and the facility planning community instead of leading decision-making for actual objectives is driven by political motives [8, 9]. In all the facility location models, the objective function becomes the function of ownership; the nature of the hierarchical system of facilities; transportation infrastructure and cost to visit facility. The objectives of location-allocation of private sector facilities are a function of profit or market share; however, the goals and objectives of public facility location are complex due to social objectives inclusion [6, 13]. Şahin and Süral [14] precisely wrote, "Main objectives of the facility location models are mainly optimizing the expected system performance, minimizing the worst-case performance and/or minimizing the maximum regret". The purpose of provision of social services is defied, especially in rural areas, where private profit-making healthcare providers avoid setting up any facility due to the scale of economies [2].

Facility location-allocation models are classified in Network Location Models, Discrete Location Models, Stochastic Demand/Congestion Models, Continuous Location Models, and Competitive Location Models [6, 13, 14]. Network Location Models and Discrete Location Models are important in location-allocation problems involving the geographic location of healthcare facilities. Network Location Models consider the network, with the possible location of facilities including nodes, as well as edges of the network as the probable facility location, however, the set of potential facility locations is discrete (often consisting of nodes of the underlying networks) in Discrete Location Models.

In the Indian context, the absence of any formal analysis by using the scientific processes and generation of alternatives due to the complexity of the available models, location-allocation decisions related to public facilities (School, Healthcare building locations) are generally taken political leaders or locally by government officers projected on public requests without pragmatic considerations. The decision made by using the scientific location models will improve healthcare service provision by optimizing the investment costs, operating costs, user satisfaction, and travel cost to users and environmental pollution.

25.2 Method

Discrete facility location models [13, 15] assume aggregation of demand to discrete points and discrete probable candidate locations for siting healthcare facilities. The objective of a set covering model is to minimize the cost of the facilities by covering all demand nodes. Following data input is required for formulating a set covering location model.

i The set of aggregated demand points;
j Discrete location sites which are competent for facility location;
f_j fixed cost of locating a facility at any site (in case of multi-facility, the cost of building facility of the same level).

Besides, we need the following decision variable:

$$X_j = \begin{cases} 1 \; if \;\; we \;\; locate \;\; a \;\; facility \;\; at \;\; site \;\; j \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad if \;\; not \end{cases} \tag{25.1}$$

Two major problems that occur with the set covering model are that the cost of covering all demands is often prohibitive increasing the number of facilities required to cover all demands, and the model fails to distinguish the weights of demand from different demand points. Church and Davis [16] addressed these concerns in maximal covering problem by adding two additional inputs in the form of weights of demand points and restricting the number of facilities. Maximal covering location problem maximizes the number of covered demands. The model considers the demand rather than considering the demand as a node by locating exactly the proposed number of facilities with standard integrality constraints. Both set covering model and maximal covering location problem treat facilities and regional nodes as binary, however, the coverage is more than the binary, and optimizing the average distance (physical, temporal, and psychological) any user travel to avail a healthcare facility is more apt objective. Optimization of average/total demand travel distance is the objective function in P-median problems.

25.2.1 P-centre Model

The P-centre model [17] addresses the optimization of the number of facilities and average distance coverage through the integration of multi-objectives into heuristics as an iterative process. The model considers the optimization of coverage distance by finding the optimal number of facilities. Following additional data, the input is required to formulate P centre model.

$$d_{ij} = distance \;\; from \;\; demand \;\; node \;\; i \;\; to \;\; candidate \;\; site \;\; location \;\; j;$$

$$Y_{ij} = \begin{cases} 1\ if\ the\ demand\ node\ i\ to\ candidate\ site\ location\ j \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad if\ not \end{cases} \tag{25.2}$$

The P-median problem is a multi-dimensional task problem and considers the demand as constant in time and minimizes the demand weighted total distance with the objective function presented in Eq. 25.1 and its constraints from Eqs. (25.3) to (25.8).

$$Minimize \sum_{j \in J} \sum_{i \in I} h_i d_{ij} Y_{ij} \tag{25.3}$$

where h_i is Demand at node i,

Subject to

assigning the demand nodes to at least one facility

$$\sum_{j \in J} Y_{ij} = 1, \forall i \in I \tag{25.4}$$

assigning demand nodes to open facility

$$Y_{ij} - X_j \leq 0; \forall i \in I; \forall j \in J \tag{25.5}$$

iterating P facilities and locating them at discrete points

$$\sum_{j \in J} X_j = P \tag{25.6}$$

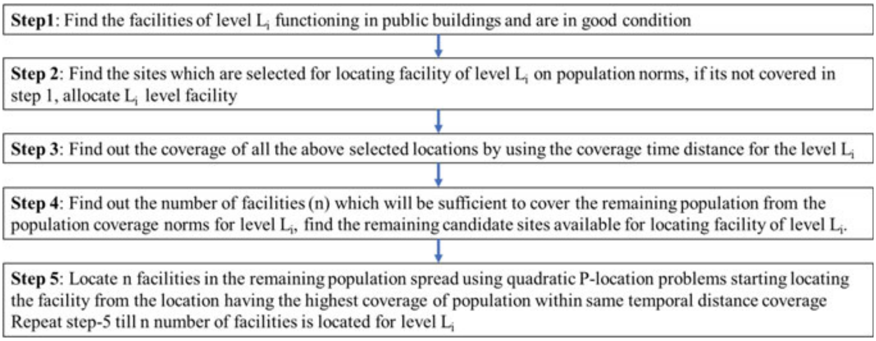
with standard integrality constraints

$$X_j \in \{0, 1\}; \forall j \in J \tag{25.7}$$

$$Y_{ij} \in \{0, 1\}; \forall i \in I; \forall j \in J \tag{25.8}$$

25.2.2 *Algorithm of Proposed Heuristic Model*

Find the facilities of level L_i which are good in condition and are functioning in public buildings. Find the sites which are selected for locating facility of level L_i on population norms presented in Table 25.1. Find out the coverage of all the above-selected locations by using the coverage time distance for the level L_i. Find out the



Step1: Find the facilities of level L_i functioning in public buildings and are in good condition

Step 2: Find the sites which are selected for locating facility of level L_i on population norms, if its not covered in step 1, allocate L_i level facility

Step 3: Find out the coverage of all the above selected locations by using the coverage time distance for the level L_i

Step 4: Find out the number of facilities (n) which will be sufficient to cover the remaining population from the population coverage norms for level L_i, find the remaining candidate sites available for locating facility of level L_i.

Step 5: Locate n facilities in the remaining population spread using quadratic P-location problems starting locating the facility from the location having the highest coverage of population within same temporal distance coverage Repeat step-5 till n number of facilities is located for level L_i

Fig. 25.1 Algorithm of proposed heuristic model for locating healthcare facility on discrete locations in a region

number of facilities (n) which will be sufficient to cover the remaining population from the population coverage norms for level L_i; find the remaining candidate sites available for locating facility of level L_i. Locate n facilities in the remaining population spread using quadratic P-location problems starting locating the facility from the location having the highest coverage of population within the same temporal distance coverage and repeat till n number of facilities are located for level L_i. Find out the nearest facility allocation for all demand points to ensure demand point assignment to the nearest facility. Find out the population to be served by each facility after the closest assignment, and find out the difference in the needed resources and available resources (like the number of more beds required at the new assigned facilities and at existing facilities), and supply the required resources at the location of the Facility. The algorithm of the proposed heuristic model for locating healthcare facilities in discrete locations in a region is presented in Fig. 25.1.

25.2.3 *Model Assumptions*

The space in which facilities have to be located is inclusive; the area solution is discrete; all parameters are deterministic; demand points and candidate facility sites are assumed to be points; the facilities of level L will offer service of the same quality; and accessibility is considered as a link, linear or non-linear.

25.3 Results and Discussion

The author developed a heuristic approach and tested it by using healthcare facility data for Sonipat district of Haryana, India, with three different scenarios, namely existing facility coverage in 2011, best case, i.e., 100% coverage of the population, and facility location using existing healthcare facility and future proposal to optimize coverage. The physical distance of 3 km, 8 km, 12 km, and 15 km has been used for analysing the ideal facility locations for subcentres, primary health centres, community health centres, and general hospitals, respectively, for covering 100% demand points with the optimal number of facilities by using hierarchical analytical process utilizing the identified parameters. The temporal distances of 5 min, 10 min, 15 min, and 25 min by vehicle using different speeds for different road types considering existing road infrastructural facilities with the help of GIS software for reaching to sub-centres, primary health centres, community health centres, and general hospitals, respectively. The results by using the heuristic model showing the efficiency of ideal case, existing healthcare facilities in 2011, and existing and proposed healthcare facilities at a different level of the hierarchical healthcare system for the District of Sonipat, Haryana, India, are presented in Table 25.2, and georeferenced results are presented in Fig. 25.2.

It can be observed from the results of the model that the ideal case for covering 100% demand points by the number of subcentres, primary health centres, community health centres, and general hospitals facilities would have remained 138, 21, 7, and 4, respectively. However, the existing facilities planned without using any mathematical modelling are 140, 29, 7, and 3 with a coverage of 95.92%, 88.97%, 82.80%, and 76.62%, respectively. The coverage by the existing facilities in 2011 is too low despite having a higher number of facilities on the same levels of hierarchy in the healthcare system. The proposed heuristic model by considering the existing facilities added very few health facilities at a different level, i.e., 05, 02, 02, and 01 and the total facilities at subcentres, primary health centres, community health centres, and general hospitals reached 145, 31, 9, and 4, respectively, to make the temporal coverage to 99.97%, 96.27%, 94.89%, and 95.75% at the respective levels of healthcare hierarchy. The average temporal distance coverage also showed a drastic reduction from 2.13 to 1.73 min for sub-centres, from 7.1 to 5.88 min for primary health centres, from 12.27 to 10.98 min for community health centres, and from 27.96 to 15.72 min for general hospitals. The temporal distance coverage for most distant 10 and 20% demand points also show a drastic reduction. Therefore, the importance of models in healthcare location planning is well established and validated by using the hierarchical facility data for Sonipat District of the State of Haryana, India.

25.4 Conclusion

Health facilities tend to be too dispersed and too distant for many consumers in rural areas as compared to urban, and most of the existing models focus only on the coverage factor of the population considering cost factor associated with the population and so miss out the coverage factor by distance or accessibility time. The distance-based standards are missing for the planning of healthcare systems except for the trauma centres on highways. User, demand points, facilities, space and metrics

Table 25.2 Results by using heuristic model showing the efficiency of ideal case, existing healthcare facilities in 2011, and existing and proposed healthcare facilities at a different level of the hierarchical healthcare system for the District of Sonipat, Haryana, India

Sr. No	Description		Level of facility			
			Sub- Centre	Primary Health Centre	Community Health Centre	General Hospital
1	Ideal case	Number of facilities of level i (Demand covered in %)	138 (100)	21 (100)	7 (100)	4 (100)
		Average physical coverage distance for all demand points (Kms)	1.55	6.4	10.4	12.6
2	Healthcare facilities in 2011	Number of facilities of level i (Demand covered in %)	140 (95.92)	29 (88.97)	7 (82.80)	3 (76.62)
		Average time coverage distance for all demand points	2.13	7.1	12.27	27.96
		Most distant 10% locations avg	6.32	13.68	23.77	68.62
		Most distant 20% locations avg	5.57	13.45	21.12	59.64
3	Existing and proposed healthcare facilities by using the heuristic model	Number of facilities of level i (Demand covered in %)	145 (99.97)	31 (96.27)	9 (94.89)	4 (95.75)
		Average time coverage distance for all demand points	1.73	5.88	10.98	15.72
		Most distant 10% locations avg	4.97	11.38	20	28.32
		Most distant 20% locations avg	4.49	10.37	18.08	25.85

(continued)

Table 25.2 (continued)

Sr. No	Description		Level of facility			
			Sub- Centre	Primary Health Centre	Community Health Centre	General Hospital
		Max distance	8	15	27	33

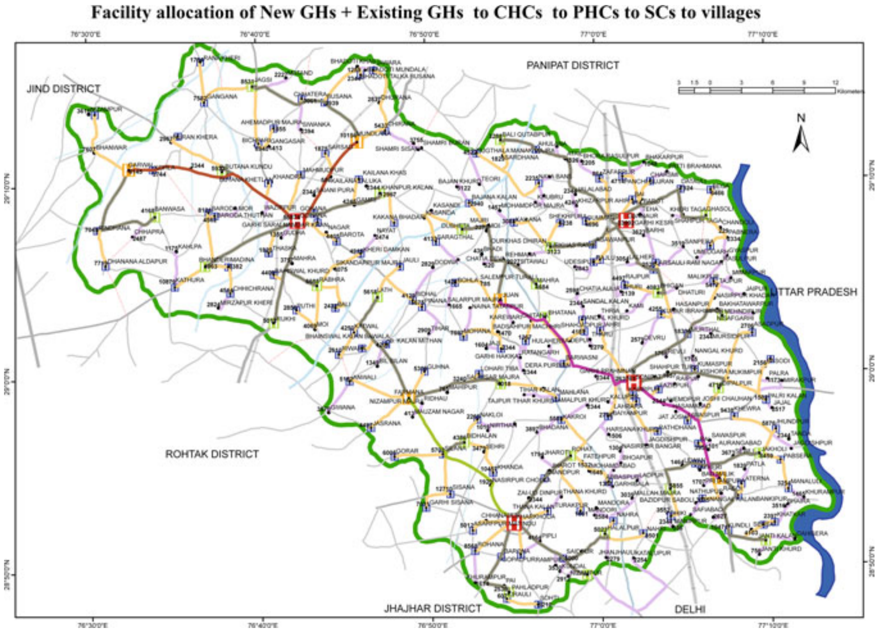


Fig. 25.2 Georeferenced results proposed heuristic model showing existing healthcare facilities in 2011, and proposed healthcare facilities at a different level of the hierarchical healthcare system for the District of Sonipat, Haryana, India

are the four essential components for describing location problems. The objective function of the location-allocation models can be formulated by the identification of basic societal needs apart from optimizing the cost. The application of basic scientific facility location-allocation models can help make political announcements for pseudo-optimal if not optimal benefits of the social facility while fulfilling the political agenda. The increase in coverage of health facilities with optimal numbers will have an impact on financial, infrastructural, social, and environmental fronts too due to a reduction in travel distance and resource requirement.

References

1. Khan, M.M., Ali, D., Ferdousy Z., Al-Mamun, A.: A cost-minimization approach to planning the geographical distribution of health facilities. Health Policy Plan. **16**(3), 264–272 (2001). https://doi.org/10.1093/heapol/16.3.264
2. Noorali, R., Luby, S., Rahbar, M.H.: Does use of a government service depend on distance from the health facility? Health Policy Plan. **14**(2), 191–197 (1999). https://doi.org/10.1093/heapol/14.2.191
3. Akin, J.S., Hutchinson, P.: Health-care facility choice and the phenomenon of bypassing. Health Policy Plan. **14**(2), 135–151 (1999). https://doi.org/10.1093/heapol/14.2.135
4. Alderman, H., Lavy, V.C., Alderman, H., Lavy, V.: Household responses to public health services: cost and quality tradeoffs. World Bank Res. Obs. **11**(1), 3–22 (1996)
5. G. MoHUA, CPWD, Compendium of Norms for Designing of Hospitals & Medical Institutions. Cent. Public Work. Dep. **1**, 1–236 (2019)
6. Drezner, Z., Hamacher, H.: Facility Location: Applications and Theory. Springer, Berlin (2002)
7. Villegas, J.G., Palacios, F., Medaglia, A.L.: Solution methods for the bi-objective (cost-coverage) unconstrained facility location problem with an illustrative example. Ann. Oper. Res. **147**(1), 109–141 (2006). https://doi.org/10.1007/s10479-006-0061-4
8. Panwar, M., Rathi, K.: Sustainable hierarchical facility location planning with reference to public healthcare system in Sonipat, India. In: Advancements in Sustainable Practices and Innovations in Renewable Energy, pp. 29–34 (2013)
9. Panwar, M., Rathi, K.: Social sustainability: contextual facility location planning model for multi-facility hierarchical healthcare system in India. Int. J. Appl. Eng. Res. **9**(3), Special Issue (2014)
10. Caccetta, L., Dzator, M.: Heuristic methods for locating emergency facilities. MODSIM05—Int. Congr. Model. Simul. Adv. Appl. Manag. Decis. Making, Proc., 1744–1750 (2005)
11. Teixeira, J.C., Antunes, A.P.: A hierarchical location model for public facility planning. Eur. J. Oper. Res. **185**(1), 92–104 (2008). https://doi.org/10.1016/j.ejor.2006.12.027
12. Coast, G., Dzator, M.: An efficient modified Greedy algorithm for the P-median problem. In: International Congress on Modelling and Simulation, pp. 1855–1861 (2015). https://doi.org/10.36334/modsim.2015.j11.dzator
13. Farahani, R.Z., Hekmatfar, M.: Facility Location. Physica-Verlag Heidelberg (2009)
14. Şahin, G., Süral, H.: A review of hierarchical facility location models. Comput. Oper. Res. **34**(8), 2310–2331 (2007). https://doi.org/10.1016/j.cor.2005.09.005
15. Current, J., Daskin, M., Schilling, D.: Discrete Network Location Models (2002)
16. Church, R.L., Davis, R.R.: The fixed charge maximal covering location problem. Pap. Reg. Sci. 71(3), 199–215 (1992). https://doi.org/10.1007/BF01434264
17. Suzuki, A., Drezner, Z.: The P-center location problem in an area. Locat. Sci. **4**(1–2), 69–82 (1996). https://doi.org/10.1016/S0966-8349(96)00012-5

Chapter 26
Natural Language Query for Power Grid Information Model

Bing Wu, Jinhao Cao, Yuanbin Song, Junyi Chu, Fulin Li, and Sipeng Li

Abstract A building information model often provides the functional and physical data of an electrical facility for the downstream construction, operation and maintenance of the built power grid infrastructure. Therefore, the rapid and convenient query of the required information from the design model is crucial for all the project participants. However, the query of the design data from a BIM model is frequently burdensome and tedious. Moreover, the Grid Information Modeling (GIM) schema, developed by the China State Grid for describing electrical equipment with more engineering details, exaggerates the difficulty of querying the design model. This study applies the Natural Language Interface to Database (NLIDB) approach for querying data from the Neo4j graph database that fuses both IFC data for architectural or structural design and GIM data for electrical equipment. Meanwhile, this study also develops a tool to automatically convert the natural language questions into Cypher queries. In addition, a knowledge graph is also developed for linking the semantic elements extracted from the natural language questions with the IFC semantics stored in the Neo4j database.

26.1 Background

Design information is a crucial resource for the construction and operation management of electrical infrastructure, but it has been sealed by a few proprietary CAD formats for a long time until the wide application of the Building Information

B. Wu
Economic and Technological Research Institute, State Grid Zhejiang Electric Power Co. Ltd, Hangzhou 310001, China

J. Cao · Y. Song (✉) · J. Chu · F. Li
Department of Transportation Engineering, School of Naval Architecture, Shanghai Jiao Tong University, Ocean & Civil Engineering, Shanghai 200240, China
e-mail: ybsong@sjtu.edu.cn

S. Li
Lishui Power Supply Company, State Grid Zhejiang Electric Power Co. Ltd, Lishui 323000, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_26

Modeling (BIM) approach. A building information model is the digital representation of the physical and functional characteristics of an engineering project, and such a model is often represented by the Industry Foundation Class (IFC) format, an open representation schema for sharing design information among trades. Besides the 3D geometric data, an IFC file also contains data of the compositional, physical, and functional attributes of an electrical facility. Moreover, the China State Grid issued the domestic BIM standards, often called Grid Information Modeling (GIM) [1], to simplify the description of electrical devices or equipment, which exaggerates the difficulty of the query of BIM data.

The research of Natural Language Interface to Database (NLIDB) provides a new means to acquire data from BIM design models [2]. NLIDB tools can automatically generate database queries by translating natural language sentences into a structured format. These tools may play an increasingly important role as designers and engineers seek to obtain information from design model databases without the assistance of computer experts with specific domain expertise or knowledge of formal query languages [3]. Natural language query can greatly reduce the time and cost for designers and engineers. Compared with English text, Chinese sentences are more difficult for processing since there is no gap between Chinese words or characters. Therefore, it is often required to segment a Chinese sentence into a sequence of words with the prevailing tools like Jieba [4], LTP2, and CoreNLP [5].

Although Recurrent Neural Network (RNN) is more suitable for evaluating a sequence of data than Convolutional Neural Network (CNN) [6], the traditional RNN approach may not capture the semantic relationship between two words with long distance in a sentence, nor can it solve the network training problem of gradient disappearance and the gradient explosion. On the other hand, Long Short-Term Memory (LSTM) network can make up for the shortcomings of the RNN model with the usage of gate units [7]. Furthermore, Bi-LSTM applies both forward and backward processing for capturing the richer context for a word [8]. In addition, many studies implied that pretrained NLP models can be successfully used as the base for specific applications with supplementary corpus. In general, there are two typical approaches for utilizing pretrained NLP models, i.e. feature-based (for example, ELMo) and fine-tuning (for example, BERT [9]).

Meanwhile, the Memory Augmented Policy Optimization (MAPO) model and its improved version MAPOX were developed to convert natural language into formal query language [10]. Meanwhile, Dong and Lapata utilized a supervised learning strategy to resolve the NL2SQL problem, proposing two alternatives, Seq2Seq and Seq2Tree [11]. Li et al. used a decomposition strategy for joint extraction of multiple relations and entities from design codes [12]. Nevertheless, the approach of converting the extracted semantic relations from engineers' questions into BIM database query scripts should be further studied. Therefore, this study explores the approach of using natural language to query information models of power grid projects.

26.2 Framework of Natural Language Query

Figure 26.1 illustrates the framework of a natural language query on a grid information model. The left part of the diagram shows the procedure that a zipped GIM [13] file that is converted into a graph database. In detail, the non-ifc files, in the format of China Grid Information Modeling standard, are first converted into .ifc files, using the method elaborated in the succeeding section. And then all .ifc files are imported into a Neo4j graph database using an IFC file parser and a sequence of Cypher codes programmed in C#. Moreover, a knowledge graph is specially developed for bridging the semantic elements extracted from the natural language questions with the IFC terms defined in the GIM database.

At the same time, the right part of Fig. 26.1 presents the process flow for translating a natural language question into a Cypher query script. The semantic information is extracted from a natural language question or query and then expressed in the format of triplets. Subsequently, these triplets are automatically converted into Cypher scripts using the templates addressed in the succeeding section. Finally, the Cypher script is executed on a knowledge enhanced graph database.

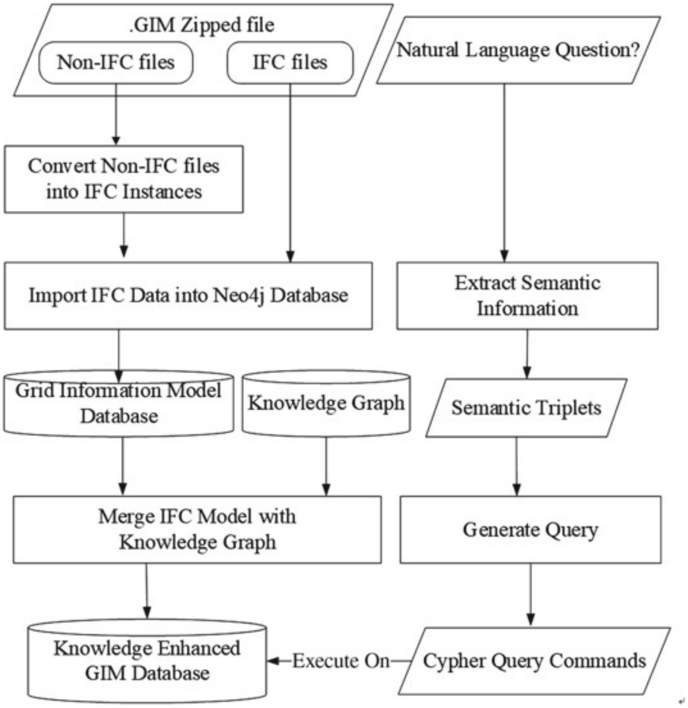

.GIM Zipped file
Non-IFC files
IFC files
Natural Language Question?
Convert Non-IFC files into IFC Instances
Import IFC Data into Neo4j Database
Extract Semantic Information
Grid Information Model Database
Knowledge Graph
Semantic Triplets
Merge IFC Model with Knowledge Graph
Generate Query
Knowledge Enhanced GIM Database
Execute On
Cypher Query Commands

Fig. 26.1 Framework of natural language query on grid information model

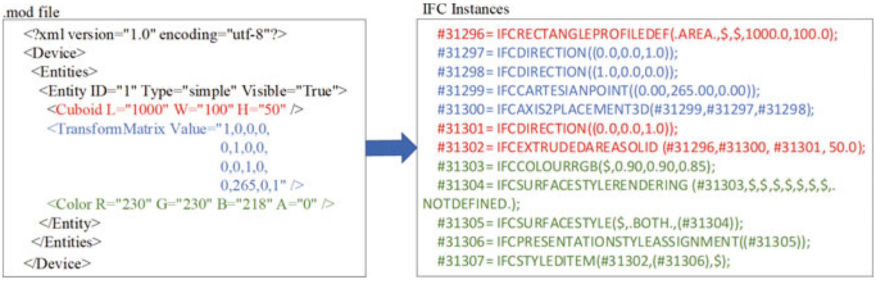
26.3 Knowledge Enhanced GIM Database

26.3.1 Conversion from GIM to IFC

A GIM file is actually a zipped file containing 4 folder files: CBM, DEV, PHM and MOD [13]. In each folder, a file uses the GUID as its unique name, and its content is encoded in UTF-8. Specifically, a .mod file depicts the parametric shape with its transformation matrix for local placement and its RGB color. A .phm file describes the structure of multiple.mod files for a more complex component, and one.phm file can further reference other .phm files. The files in the DEV folder define the individual device/equipment, or a system of devices. At the same time, the files in the CBM folder describe the organization of subsystems of a gird engineering project, which also comprises .ifc files for architectural, structural, pipe, and ventilation systems. In addition, a .fam file is used for depicting the material, physical, and functional attributes of either parts or devices or systems.

Since the IFC schema has much richer semantic constructors than GIM, the non-ifc BIM data files for representing electrical devices are first converted into IFC instances. Figure 26.2 illustrates a typical case of converting a .mod file into the corresponding IFC components.

The parametric representation of a cuboid in the .mod file is converted to a number of IFC instances. The rectangle profile of the cuboid is represented by the IfcRectangleProfileDef instance (#31,296), and its extrusion direction by the IfcDirection instance (#31,301), and furthermore the cuboid shape is described by the IfcExtrudedAreaSolid instance (#31,302). Then, the 4*4 transformation matrix for local placement of the cuboid is also converted into the IfcAxis2Placement3D instance (#31,300), which is further described by another three IFC instances, original point (#31,299), X direction (#31,298), and Z direction (#31,297). In addition, the color of the shape is depicted by the IfcColourRGB instance (#31,303).



.mod file

```
<?xml version="1.0" encoding="utf-8"?>
<Device>
 <Entities>
  <Entity ID="1" Type="simple" Visible="True">
   <Cuboid L="1000" W="100" H="50" />
   <TransformMatrix Value="1,0,0,0,
                           0,1,0,0,
                           0,0,1,0,
                           0,265,0,1" />
   <Color R="230" G="230" B="218" A="0" />
  </Entity>
 </Entities>
</Device>
```

IFC Instances

```
  #31296= IFCRECTANGLEPROFILEDEF(.AREA.,$,$,1000.0,100.0);
  #31297= IFCDIRECTION((0.0,0.0,1.0));
  #31298= IFCDIRECTION((1.0,0.0,0.0));
  #31299= IFCCARTESIANPOINT((0.00,265.00,0.00));
  #31300= IFCAXIS2PLACEMENT3D(#31299,#31297,#31298);
  #31301= IFCDIRECTION((0.0,0.0,1.0));
  #31302= IFCEXTRUDEDAREASOLID (#31296,#31300, #31301, 50.0);
  #31303= IFCCOLOURRGB($,0.90,0.90,0.85);
  #31304= IFCSURFACESTYLERENDERING (#31303,$,$,$,$,$,$,$,.
NOTDEFINED.);
  #31305= IFCSURFACESTYLE($,.BOTH.,(#31304));
  #31306= IFCPRESENTATIONSTYLEASSIGNMENT((#31305));
  #31307= IFCSTYLEDITEM(#31302,(#31306),$);
```

Fig. 26.2 Conversion of MOD file into IFC instances

26.3.2 Knowledge Graph for Enhancing GIM Graph Database

In order to correlate the semantic elements extracted from natural language queries with the class names defined in the IFC schema, a knowledge graph is defined. The knowledge graph in Fig. 26.3 illustrates the key concepts: *Element, Attribute, System, Space, Material, Comparison, and MathFunction*. Each core concept can be further described with its subcategory concepts or hyponyms.

In detail, the category of *Element* has 10 subcategories, like *Building Element, Civil Element, Distribution Element, Feature Element, Furnishing Element, Geographic Element,* etc. Moreover, the *Building Element* can be further subcategorized into *Architectural Element* and *Structural Element*. Subsequently, the category of *Structural Element* can have 6 subcategories of element classes, i.e. *Foundation, Pile, Structural Beam, Structural Column, Structural Slab, and Structural Wall.*

In the knowledge graph, each hypernym associates with its hyponyms by the Subcategory relationship. Meanwhile, there exists HasProperty relationship between Element and Material and Attribute. At the same time, the concept of Space has Contain relationship with Element. These semantic relationships between core concepts are also coded as Neo4j relationships.

The key concepts in the knowledge graph further associate with one or more IFC classes. For example, *Structural Beam connects with* IfcBeam via the ReferenceIfcClass relationship. Then, using the Neo4j Cypher script can create a connection between the IFC classes in the knowledge graph and the IFC instances stored in the GIM database. For example, the IfcBeam class in the knowledge graph can be

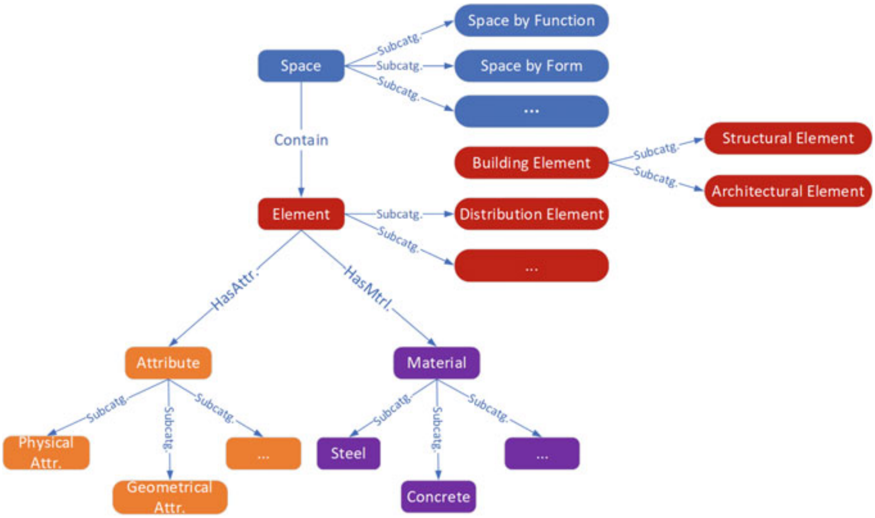


Fig. 26.3 Part of Knowledge Graph for Enhancing GIM database

used for linking the IFC instance labeled by IfcBeam in the GIM database. Meanwhile, the Hot Rolled H Steel material concept is connects with IfcPropertySet class, which further associates with the IFC instances labeled with IfcPropertySet in the GIM database. In this way, the knowledge graph acts as the bridge between the semantic entities (in the natural language questions) and the IFC instances (in the GIM database).

26.4 Automatic Generation of Cypher Script

The joint extraction model developed by Li et al. [12] is used to extract semantic information from natural language questions. Figure 26.4 illustrates the information extraction pipeline that composes four components, i.e. character embedding, shared semantic encoder, subject extractor, and object and predicate extractor. Through the pipeline, a query sentence written in natural language can be automatically converted into a number of triplets.

The character embedding, the first module in the aforementioned pipeline, is utilized for transforming each character in the Chinese question into a real vector, herein called a character vector. Subsequently, the shared semantic encoder is applied to learn the context features of each character, literally called task-shared features. Specifically, a Bi-LSTM model is used to encode the association between a character and its surrounding characters (on both left-hand and right-hand sides). Subsequently, the subject extractor module uses the task-shared features to identify all candidate-named entities that have the opportunity to act as subjects. And then the associated object entities and predicate relations, for each subject entity identified, are simultaneously identified by the object and predicate extractor using task-shared features. Finally, the semantic information in the natural language question can be automatically extracted into a set of triplets.

Figure 26.5 presents the conversion of the Chinese question “钢结构件的总重量是多少?” to semantic triplets. In English, the Chinese question means to query

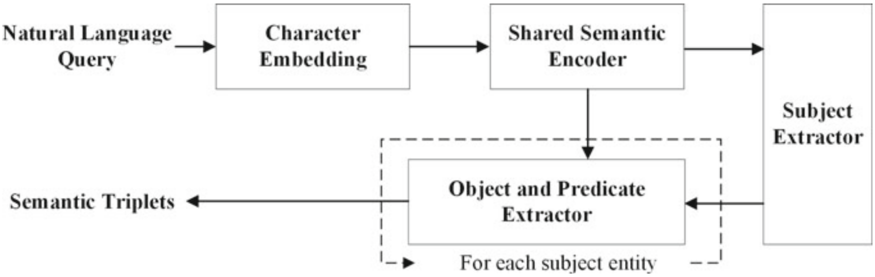
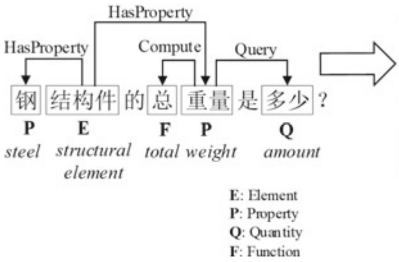


Fig. 26.4 Framework of the joint extraction model



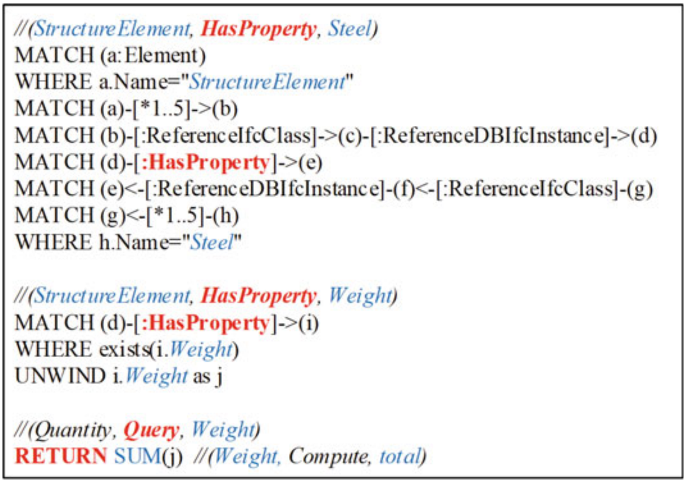
Subject entity	Predicate relation	Object entity
"结构件"#1(1,3)	HasProperty	"钢"#1(0,0)
"结构件"#1(1,3)	HasProperty	"重量"#1(6,7)
"重量"#1(6,7)	Query	"多少"#1(9,10)
"重量"#1(6,7)	Compute	"总"#1(5,5)

Format: ["named entity" #index (start position, end position)]

Fig. 26.5 Conversion of natural language question into semantic triplets

the total weight of the steel structural elements. Using the pipeline in Fig. 26.4, the natural language is converted into 4 semantic triplets.

The semantic triplets are then converted into Cypher query scripts via the three predefined mapping templates as shown in Fig. 26.6. The predict "HasProperty" in the first triplet is translated into Neo4j relationship [:HasProperty] and a set of MATCH commands is simultaneously inferenced to generate a linkage paths with the assistance of the knowledge graph. In detail, the subject "Structure Element" can be searched from the core concept Element in the knowledge graph, while the object entity "Steel" becomes the property constraint to locate the IFC entities (see the WHERE condition). Meanwhile, the second triple is to locate all IFC elements with the property "Weight", and all the found weight values are organized into a set, named j defined with UNWIND. Finally, the third triplet indicates the goal of the information searching, and the fourth triplet defines the mathematic function SUM of the found weight values, i.e. the total weight.



```
//(StructureElement, HasProperty, Steel)
MATCH (a:Element)
WHERE a.Name="StructureElement"
MATCH (a)-[*1..5]->(b)
MATCH (b)-[:ReferenceIfcClass]->(c)-[:ReferenceDBIfcInstance]->(d)
MATCH (d)-[:HasProperty]->(e)
MATCH (e)<-[:ReferenceDBIfcInstance]-(f)<-[:ReferenceIfcClass]-(g)
MATCH (g)<-[*1..5]-(h)
WHERE h.Name="Steel"

//(StructureElement, HasProperty, Weight)
MATCH (d)-[:HasProperty]->(i)
WHERE exists(i.Weight)
UNWIND i.Weight as j

//(Quantity, Query, Weight)
RETURN SUM(j) //(Weight, Compute, total)
```

Fig. 26.6 Conversion templates of semantic triplets into Cypher query

26.5 Conclusions

The designers and engineers frequently feel it burdensome and tedious to search information from the prevailing BIM design model of a grid engineering project with multiple query sentences. In this regard, a framework of an automatic conversion from natural language questions into Neo4J Cypher scripts has been developed in this study. Those non-IFC files contained in a .gim file are first converted into .ifc files, and then all the .ifc files are imported into the Neo4j graph database to achieve faster information retrieval. And then, the joint extraction model is used for extracting semantic information from a natural language question into a set of triplets that can be further converted into Cypher scripts by the mapping templates. In addition, a knowledge graph is also developed to connect the semantic entities in the question with the IFC classes. Consequently, the BIM design model can be more effectively and conveniently queried by natural language questions.

Since this research is still in its initial stage, the joint extraction model will be further trained with more labeled questions, and more mapping templates for generating Cypher scripts will also be developed.

References

1. Jung, N., Lee, G.: Automated classification of building information modeling (BIM) case studies by BIM use based on natural language processing (NLP) and unsupervised learning. Advanced Engineering Informatics (41), 100917, (2019).
2. Agrawal, A., Kakde, O.: Object-Relational Database Based Category Data Model for Natural Language Interface to Database. International Journal of Artificial Intelligence and Applications **1**(2), 35–41 (2011)
3. Sun C.: A natural language interface for querying graph databases. Massachusetts Institute of Technology, Thesis, (2018).
4. Zhang X., Wu P., Cai J., Wang K.: A Contrastive Study of Chinese Text Segmentation Tools in Marketing Notification Texts. Journal of Physics, Conference Series 1302(2), (2010).
5. Manning, C., Surdeanu, M., Bauer, J.: The Stanford CoreNLP Natural Language Processing Toolkit. In: 52nd Annual Meeting of the Association-for-Computational-Linguistics (ACL), Baltimore, MD. 55–60, (2014).
6. Liu, X., Hou, S., Qin, Z., Liu, S., Zhang J.: Relation extraction for coal mine safety information using recurrent neural networks with bidirectional minimal gated unit. EURASIP Journal on Wireless Communications and Networking, 55, (2021).
7. Wang, J., Zhang, L., Chen, Y., Yi, Z.: A New Delay Connection for Long Short-Term Memory Networks. Int. J. Neural Syst. **28**(6), 1750061 (2017)
8. Adnen, M., Mounir, Z.: BLSTM-API: Bi-LSTM Recurrent Neural Network-Based Approach for Arabic Paraphrase Identification. Arab. J. Sci. Eng. **46**, 4163–4174 (2021)
9. Sur, C.: RBN: Enhancement in language attribute prediction using global representation of natural language transfer learning technology like Google BERT. SN Applied Sciences, 22(2), (2020).
10. Liang, C., Norouzi, M., Berant, J.: Memory Augmented Policy Optimization for Program Synthesis and Semantic Parsing. 32nd Conference on Neural Information Processing Systems (NIPS). (2018). Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, Canada December, 10015–10027, (2018).

11. Dong L., Lapata M.: Language to Logical Form with Neural Attention. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 33–43, (2016).
12. Li F., Song Y., Shan Y.: Joint Extraction of Multiple Relations and Entities from Building Code Clauses. Applied Sciences 10(20), (2020).
13. China Electricity Council: Interactive specification for the three-dimensional design model of power transmission and transformation project (2020).

Chapter 27
Time Power Law Mapping of Signal Complexity Measure

Zeyang Mao and Wenshi Li

Abstract The complexity measure is to describe the cognitive cost of a system, such as a one-dimensional time-domain signal in periodic, chaotic, or random state. Its research paradigms depend on the understandings of symbolic dynamics, entropy expansion, FFTs' combination, and the complex network. The new trends of computing aesthetics show both the group expansion with formula and without artificial selection of parameters. Following the power law of time in gold rate, we give a practical algorithm with a new feature plot and its complex number. The contrast criteria are spectral entropy complexity and approximate entropy. Three chaotic equations are used as test examples. The results reveal new complex patterns and novel complex numerical ranges, obtaining basically consistent recognition effects compared with upper control criteria. The calculation complexities of all running algorithms are reported. This work illustrates new progress in our brainchildren lab on automatic measuring of signal complexity.

27.1 Introduction

Complex behaviors are embedded in complex science (typical cases as heterogeneity, nonlinearity, and chaos) [1–4]. The complexity measure is to describe quantitatively the cognitive computing consumption or predictive cost of any output data of the system under test. It was called the search for complexity calculus by American mathematician Strogatz. The simple form of 0–1 test [5] for chaos in stochastic process principle, the algorithm complexity of *pq* plot is O(n). The *pq* plot can clearly appreciate the qualitative difference between regular and chaotic dynamics; periodic or quasi-periodic changes in time series $x(n)$ would end up with a circular trajectory in the *pq* plot. The O(n) is estimated by the number of operation units of the algorithm

Z. Mao · W. Li (✉)
School of Electronics and Information En, Soochow University, Suzhou 215006, China
e-mail: lwshi@suda.edu.cn

W. Li
Laboratory of Modern Acoustics of MOE, Nanjing University, Nanjing 210093, China

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3_27

and inspires us to establish a new group extension [5–7]. Geometry idea such as manifolds reminds us of a new pattern in the above space [8, 9]. So far, the geometric statistical invariants of the characteristic shape can be measured quantitatively. We propose a practical method with feature plot and complex numbers without manual parameter selection, performing three-category recognition of periodic, chaotic, or pseudorandom signal and expecting geometric and numerical expressions. Compared with spectral entropy complexity and approximate entropy, we combine the time power law principle with the gold rate rule. The analytical blocks of the new mapping are divided into the definitions of feature plot and its complex number.

27.2 Measuring Principles and Test Cases

27.2.1 The Main Ideas of SE and ApEn

Spectral entropy (SE) complexity investigates the flatness of the spectrum in the frequency domain. First, the algorithm idea is to remove the mean component, then calculate the power spectrum probability, and finally get the normalized Shannon entropy value which is characterized by no manual selection parameter. The more the *SE* value is, the higher is the complexity of data under test. Here, the algorithm time complexity is O($n$log(n)) [10].

Approximate entropy (ApEn) analyzes quantitatively the complexity of time-domain signals and calculates the probability of generating new patterns in the signal (large values correspond to higher complexity). The core of the $ApEn(\tau,m,r)$ algorithm is to first obtain the logarithm mean value of the numbers under the r threshold of the two-dimensional vector distance, and then change it into one dimension. Then to repeat the foregoing, the difference limit of the two logarithmic mean values is defined as the approximate entropy. Generally, the threshold r ranges from 0.1 to 2.5 times the standard deviation, and the time complexity of this algorithm is O($n^{1.5}$) [11].

27.2.2 The Principles of Power Law and Gold Rate

In the power law principles, exponential growths may be a manifestation of the self-reinforcing law of natural evolution, involving specific factors such as matter, energy, life, and the complexities existing in them, with famous examples of Price's law and Moore's law [12]. And the gold rate rule is the most known among various algorithms, because of the strict proportionality, artistry, and harmony. In scientific experiments, the 0.618 method is used commonly in optimal schemes. In the above view, this work tends to change the time factor c in the 0–1 test for chaos, observing its impacts on our new time power law models carefully. After scanning parameter *c*, we choose two times gold rate as the fixed time power law [6, 7].

27.2.3 The Principles of Feature Plot and Complex Number

The complexity measure is for one-dimensional time-domain data $x(n)$ of length N. Next, we propose a measure coined complex number (CN). *CN* signature rule: less than 11 indicates periodic state, greater than or equal to 11 features chaotic state, and greater than 300 marks random number.

First, we define the discrete expressions of the two characteristic dimensions in fixed time power law mapping with j^c and $\cos(j^c)$; new *pq* plot is written as

$$\begin{aligned} p(n) &= \sum_{j=1}^{n} x(j) j^c n = 1, 2, \ldots, N \\ q(n) &= \sum_{j=1}^{n} x(j) \cos(j^c), n = 1, 2, \ldots, \mathrm{N}. \end{aligned} \tag{27.1}$$

wherein the chaotic sensitivity factor *c* is selected actually to be twice the gold rate ($c = 1.236$).

Then define the characteristic length (CL) value *CL*:

$$CL = (p_{\max} - p_{\min})^{dt}/_{N}. \tag{27.2}$$

wherein the projection length of the feature dimension *p*, the calculation step length *dt*, and the data length N are used. Define the statistical measure *CN* (coined complex number) as

$$CN = 1000/CL. \tag{27.3}$$

27.2.4 Test Cases

Test case 1: Logistic map is used as a typical case in [6, 7] to verify respective theories; the formula is as follows:

$$x(n+1) = kx(n)(1 - x(n)). \tag{27.4}$$

wherein the coefficient k = 2.5 ~ 4. The initial value is [0.01].

Test case 2: Lorenz equation is

$$\begin{aligned} \mathrm{d}x/\mathrm{d}t &= \alpha(y - x) \\ \mathrm{d}y/\mathrm{d}t &= -xz(24 - 4k)x + kz \\ \mathrm{d}z/\mathrm{d}t &= xy - \beta z \end{aligned} \tag{27.5}$$

wherein the parameters $\alpha = 10$, $\beta = 8/3$, and $k = -2 \sim 8$.

Test case 3: Chua's equation is

$$\begin{aligned} \mathrm{d}x/\mathrm{d}t &= -2.564x + 10y + 0.5k(|x+1| - |x-1|) \\ \mathrm{d}y/\mathrm{d}t &= x - y + z \\ \mathrm{d}z/\mathrm{d}t &= -14.706y \end{aligned} \tag{27.6}$$

wherein the parameter $k = 2.6 \sim 3.5$.

Calculation conditions: Initial values are [0.01, 0.01, 0.01]. Iterative step length is 0.01. For sampling time series $x(n)$, we select data points between 1000 and 5000, and the data length, N = 4000. Approximate entropy parameter selection: time delay $\tau = 1$, embedding dimension $m = 2$, and threshold $r = 0.2\sigma$) [11].

27.3 Results and Discussions

27.3.1 Verification Based on 3 Chaotic Equations

In Fig. 27.1, the new *pq* plot compares and expresses four signal states, involving $k = 3.0$ in periodic state and $k = 3.7$ in chaotic state of the logistic map, the normal distribution, and T distribution of pseudo-random numbers. Finger-shape spots are featured in Fig. 27.1. The differences in projection lengths appear on the *p*-axis, and the differences in projection heights lie on the *q*-axis. Thus, the geometric distribution patterns from long to short and from low to high are easy to identify four known signal states by naked eyes.

To compare Fig. 27.2(left), (m), and (right), *CN* values distinguish the bifurcation points of the Logistic map like the bifurcation diagram and change larger after marching into chaos at $k = 3.6$. In detail, while complex number *CN* is small (increasing from close to 8 to 11), the initial period and its bifurcation are quantified in the zone of $k = [2.5, 3.55]$, till $k = [3.56, 4]$, and *CN* becomes larger (more than 11); chaotic mapping is captured by *CN* values. There are consistent diagnosing results in the control test of *SE* and *ApEn* values. Here, the nonlinear dynamics in entering chaos through bifurcation are all detected by three kinds of complexity measures of *CN*, *SE*, and *ApEn*.

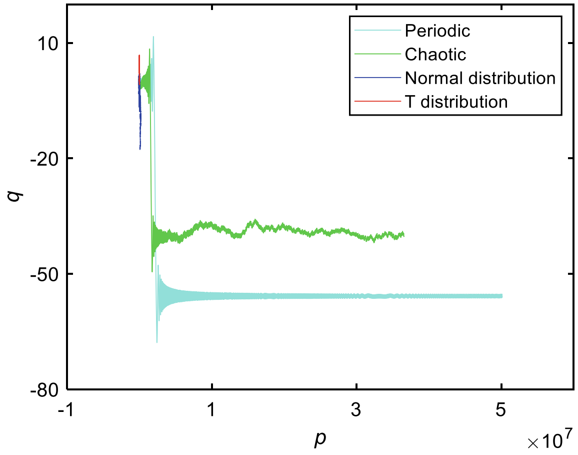


Fig. 27.1 The *pq* plot of the time power law mapping of the logistic map (with solid *c* = 1.236)

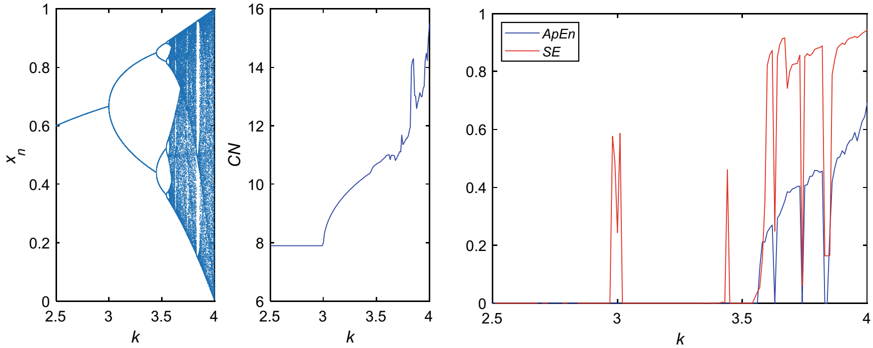


Fig. 27.2 Logistic map measuring: (left) Bifurcation diagram; (middle) *CN* values; (right) *SE* and *ApEn* values

Furthermore in Fig. 27.3(left) and (right), when the known parameter *k* lies within $-1.59 = < k < = 7.75$, the Lorenz system is locked up in a chaos state. Here, ththree features (periodic, chaotic or pseudo-random signal can be clearly distinguished)for signal complexity measuring mach well.

Finally, we examine Chua's Eq. (27.6) and observe the changing rule of the *CN* values during the evolution from a single period to double scroll (see Fig. 27.4(left) and (right)). While *k* sticks at 2.7, 2.95, 3.02, 3.11, or 3.45, the phase diagrams of Chua's equation show single-period, double-period, three-period, single-scroll, and double-scroll states, respectively.

The low values of *CN* in Fig. 27.4a slowly rise (representing a periodic state) and at $k = 3.11$, $CN = 11$ indicates that Chua's equation had entered into chaos (single scroll). The *CN* value suddenly increases at $k = 3.28$ which means that the complexity of the system begins to increase greatly. In Fig. 27.4(right), the higher values of *SE* and *ApEn* describe the chaotic performances well, while the low-value

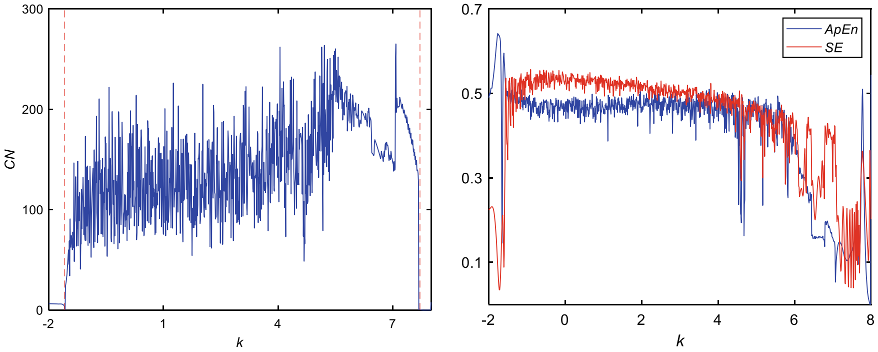


Fig. 27.3 Lorenz equation measuring: (left) *CN* values; (right) *SE* and *ApEn* values

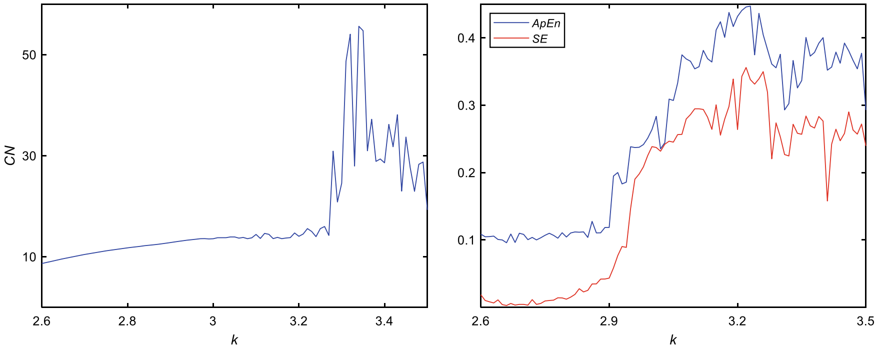


Fig. 27.4 Chua's equation measuring: **a** *CN* values, **b** *SE* and *ApEn* values

fast-rising characteristics intuitively distinguish the evolution of the periodic state into a new chaotic state.

27.3.2 *Best Factor* C *Scanning Based on Chua's Equation*

Four values of parameter c are selected as 1.1, 1.2, 1.236, and 1.3. The responding new pq plots are shown from Fig. 27.5a to (right-down).

It is obvious that the naked-eye recognition effects of Fig. 27.5(left-up) and (right-up) are not well; the geometric patterns of 6 categories are seriously aliased. Looking at Fig. 27.5(left-down) and (right-down), we can see the geometric discrimination performances in the two projection axes. We chose actually the key value of $c = 1.236$ in Fig. 27.5(l–d). Using the p-axis projection distance, we construct a simple feature length as the basis for the modeling and expression of a new metric *CN*.

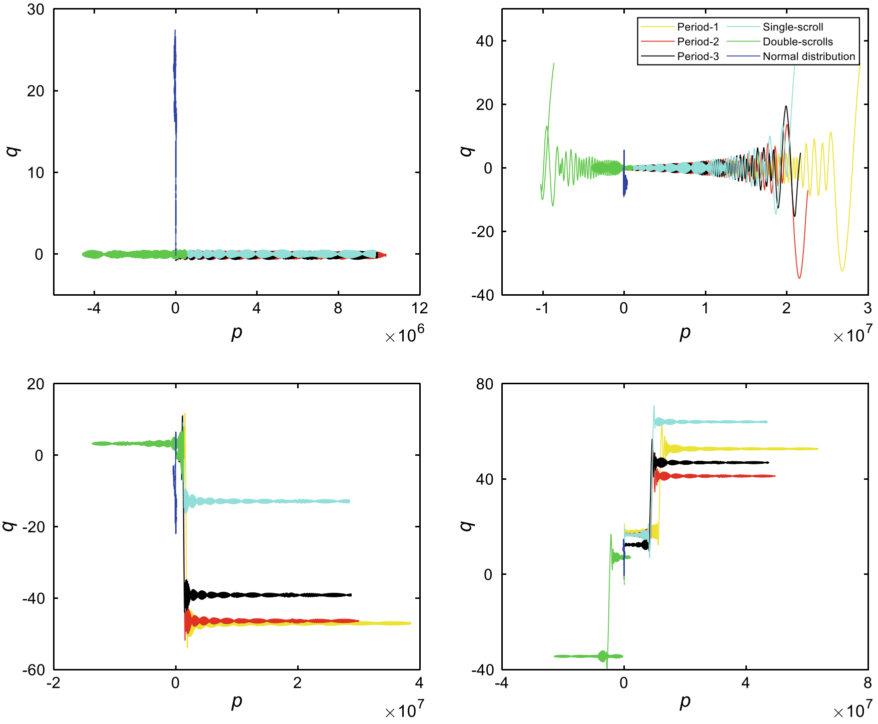


Fig. 27.5 Parameter c scanning in pq plots of Chua's equation: (left-up) $c = 1.1$; (right-up) $c = 1.2$; (left-down) $c = 1.236$; (right-down) $c = 1.3$

27.3.3 Identification of Pseudo-random Numbers

We analyze the ability of distinguishing pseudo-random numbers based on our *CN* metric. Fifteen kinds of distributions of pseudo-random numbers (data length of N = 4000) are selected to calculate their average *CN* values 10 times, and the *SE* values and *ApEn* values are compared.

Table 27.1 shows (1) The *SE* algorithm has achieved all correct recognition rates in distinguishing pseudo-random numbers. Although the *SE* values are stable (from 0.94 to 0.95), they have lost their different distributions of 15 pseudo-random numbers; (2) The discrimination results of *ApEn* values are between 0.3 and 1.7; (3) *CN* values are between 15.8 and 1444.7, which can better identify the different complexity performance distances. It is considered cautiously that *CN* can only recognize the latter five pseudo-random numbers shown in Table 27.1.

Finally, the calculation times of the three complexity measuring algorithms are compared. The simulation uses MATLAB R2018a software (CPU: Ryzen 7 4800U, 1.9 GHz) to count the calculation times of the above 3 test cases, to keep the system running environment consistent as far as possible, and to repeat every experiment 10 times and take the average results list in Table 27.2.

Table 27.1 *SE*, *ApEn*, and *CN* values of 15 pseudo-random numbers

Random number	*SE*	*ApEn*	*CN*
Uniform distribution	0.95	0.35	15.8
Rayleigh distribution	0.95	0.32	26.5
Non-central F distribution	0.94	0.36	31.1
Beta distribution	0.94	0.47	35.0
Poisson's distribution	0.94	1.27	50.3
Geometric distribution	0.95	1.70	62.5
Hypergeometric distribution	0.94	0.71	73.0
Exponential distribution	0.94	0.73	73.8
Weibull distribution	0.94	1.26	116.7
Lognormal distribution	0.95	1.21	186.7
Binomial distribution	0.94	1.29	418.6
Chi square distribution	0.94	1.39	423.9
F distribution	0.95	1.8	504.6
Standard normal distribution	0.94	0.3	1210.9
T distribution	0.95	0.4	1444.7

Table 27.2 Running time means of *SE*, *ApEn*, and *CN*

Test cases	*SE*/s	*ApEn*/s	*CN*/s
Logistic map	0.1	55.7	2.6
Lorenz equation	106.9	390.2	250.3
Chua's equation	4.8	37.5	96.2

In Table 27.2, the approximation of the calculation times shows that the calculation time complexity of the *CN* algorithm is $O(n^{1.3})$, which is between the *SE* algorithm and the *ApEn* algorithm.

27.4 Conclusions

The geometric characteristics of the fixed-time power law mapping in the new *pq* plot are finger shapes, with long, medium, and short fingers corresponding to period, chaos, and random, and it is relatively easy to perform three classification recognition depends on the finger length.

The new complex number (*CN*) combines three factors of the projection length of time power law expansion feature dimension, the discrete time step, and the data length. Generally, the *CN* value 11 is used as the threshold for judging periodic and chaotic states, and the value 300 is the threshold for judging chaos and Gaussian randomness. We can apply the *CN* algorithm in many areas such as recognition of EEG signals or speech signals.

Comparing the *CN* algorithm with Spectral Entropy complexity and Approximate Entropy, we had proposed a practical criterion of signal complexity measure, especially digging out five states of the fifteen kinds of distributions of pseudo-random numbers.

Of course, the CN algorithm's stability performance for data length changes and the comparison with newer complexity features need more chaotic test cases and deeper application scenarios for further verification.

References

1. Li, T.Y., Yorke, J.A.: Period three implies chaos. Am. Math. Mon. **82**(10), 985–992 (1975)
2. Crutchfield, J.P.: Between order and chaos. Nat. Phys. **8**(1), 17–24 (2012)
3. Leung, H.: Chaotic Signal Processing. Higher Education Press, Beijing (2014)
4. Li, W.S.: Case Study of Micro-nano Electronics Modeling (in Chinese), 2nd edn. Soochow University Press, Suzhou (2019)
5. Bernardini, D., Litak, G.: An overview of 0–1 test for chaos. J. Braz. Soc. Mech. Sci. Eng. **38**, 1433–1450 (2016)
6. Wu, S.L., Li, Y.T., Li, W.S., Li, L.: Chaos criteria design based on three-threshold sign function. Chin. J. Electron. **28**(2), 364–369 (2019)
7. Cai, J.W., Li, Y.T., Li, W.S., Li, L.: Two entropy-based criteria design for signal complexity measures. Chin. J. Electron. **28**(6), 1139–1143 (2019)
8. Tang, L., Lv, H.L., Yang, F.M., Yu, L.A.: Complexity testing techniques for time series data: a comprehensive literature review. Chaos, Solit. Fract. **81**, 117–135 (2015)
9. Wade, J., Heydari, B.: Complexity: definition and reduction techniques: some simple thoughts on complex systems. Complex Syst. Design Manag. **1234**(18), 213–226 (2014)
10. Su, K.H., He, S.B., He, Y., Yin, L.Z.: Complexity analysis of chaotic pseudo-random sequences based on spectral entropy algorithm. Acta Physica Sinica **62**(1), 27–34 (2013)
11. Pincus, S.M.: Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. **88**(6), 2297–2301 (1991)
12. Cramer, F.: Chaos and Order: The Complex Structure of Living System. Deutsche Verlags-Anstalt, München (1988)

Author Index

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Singapore Pte Ltd. 2022

R. Kountchev et al. (eds.), *New Approaches for Multidimensional Signal Processing*,
Smart Innovation, Systems and Technologies 270,
https://doi.org/10.1007/978-981-16-8558-3