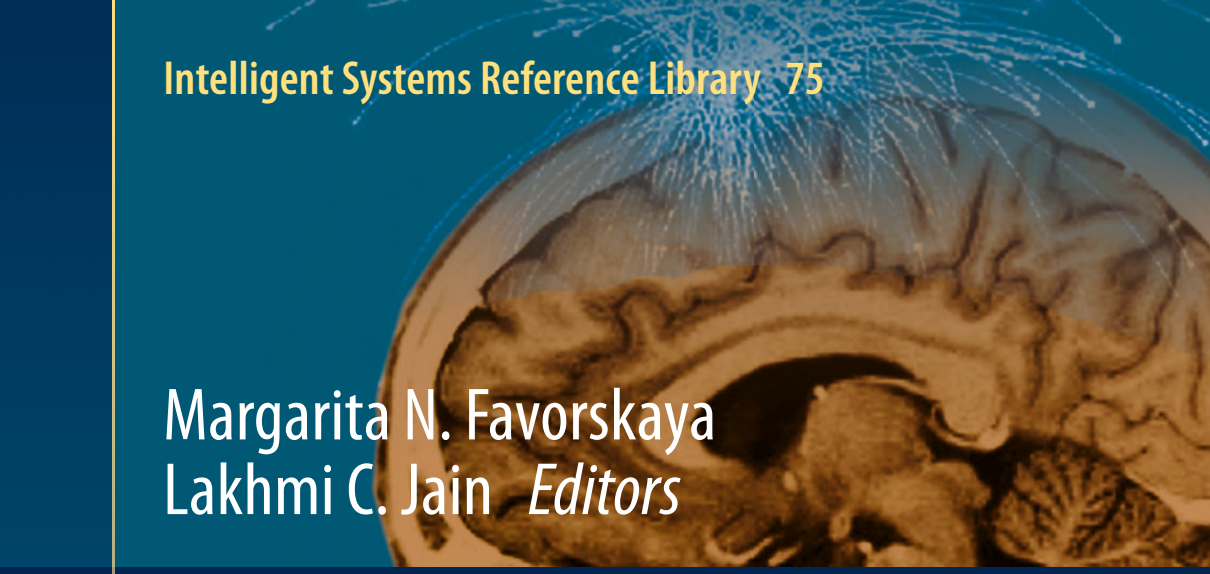
Intelligent Systems Reference Library 75



Margarita N. Favorskaya
Lakhmi C. Jain *Editors*

Computer Vision in Control Systems-2

Innovations in Practice

Springer

Intelligent Systems Reference Library

Volume 75

About this Series

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included.

More information about this series at http://www.springer.com/series/8578

Margarita N. Favorskaya · Lakhmi C. Jain
Editors

Computer Vision in Control Systems-2

Innovations in Practice

Springer

Editors
Margarita N. Favorskaya
Department of Informatics and Computer Techniques
Siberian State Aerospace University
Krasnoyarsk
Russia

Lakhmi C. Jain
Faculty of Education, Science, Technology and Mathematics
University of Canberra
Canberra
Australia

ISSN 1868-4394
ISBN 978-3-319-11429-3
DOI 10.1007/978-3-319-11430-9

ISSN 1868-4408 (electronic)
ISBN 978-3-319-11430-9 (eBook)

Library of Congress Control Number: 2014951912

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

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

Foreword

In keeping with the overview of the mathematics underlying computer vision given in Volume 1, this volume introduces a rich landscape that spans a broad spectrum of computer vision applications. This landscape includes a very original view of human action recognition, a rich selection of novel methods in robot navigation systems, promising applications of face recognition methods, visual panorama reconstructions using images captured by mobile robot cameras, motion estimation methods based on salient feature points, intelligent robot motion control-based robotic visual perception (closely related to human perception), a novel approach to automatic surveillance based on the description of complex scenes and inter-object relationships, innovations in avionics, in-situ position estimation in navigating autonomous underwater vehicles, a thorough coverage of digital image filtration methods, as well as a good overview of image segmentation systems for monitoring such things as the Earth's surface, disease diagnosis, and technical object safety. As a result, this volume provides an in-depth view of computer vision applications in harmony with the computer vision theory in Volume 1. Appropriately, this volume begins with a detailed survey of the contributions by chapter authors: *Practical Matters in Computer Vision* (L.C. Jain, M. Favorskaya). This initial chapter is followed by a host of interesting practical views of computer vision.

The central motifs in this volume reflect a sensitive and remarkable understanding of computer vision in practice. These motifs are threefold.

1. Image Structures

Human Action Recognition (S. Al-Ali, M. Milanova, H. Al-Rizzo, V.L. Fox),
Efficient Denoising Algorithms for Intelligent Recognition Systems (A. Priorov, K. Tumanov, V. Kolokhov),
Image Segmentation Based on Two-Dimensional Markov Chains (E. Medvedeva, E. Kurbatova).

2. Image and Video Measurement

Real Time Audience Analysis System (V. Khryashchev, L. Shmaglit, A. Shemyakov),
Panorama Construction from Multi-view Cameras in Outdoor Scenes (L.C. Jain, M. Favorskaya, D. Novikov),
Real-Time Method of Contextual Image Description and Its Application in Robot Navigation and Intelligent Control (K.I. Kiy),
Perception of Audio Visual Information for Mobile Robot Motion Control Systems (S. Pleshkova, A. Bekiarski, S.S. Dehkharghani, K. Peeva).

3. Image-Based Signal Analysis

Adaptive Surveillance Algorithms Based on the Situation Analysis (N. Kim, N. Bodunkov),
Enhanced, Synthetic and Combined Vision Technologies for Civil Aviation (O. Vygolov, S. Zheltov),
Navigation of Autonomous Underwater Vehicles Using Acoustic and Visual Data Processing (I. Burdinsky, A. Myagotin).

A comprehensive view of applications in pattern recognition and image analysis are given in this volume. These applications are ably presented by the volume contributors.

I strongly recommend this volume and its companion volume as a concise and very original introduction to image analysis and its applications.

June 2014

James F. Peters
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, MB, Canada

and

Faculty of Arts and Sciences
Department of Mathematics
Adıyaman University
Adıyaman, Turkey

Preface

The research book is a continuation of our previous book, which is focused on the recent advances in computer vision methodologies and technical solutions using conventional and intelligent paradigms. The contemporary solutions based on advanced mathematical achievements emphasize more information and visual monitoring in natural and human environment. The real challenge of designing such observation models is to make them close to realistic visualization and interpretation of events in our world.

This book presents some of the research results from some of the most respectable researchers in the field of computer vision including some innovative applications in practice. The contributions include the recent methodologies for human action recognition, real-time audience analysis system, panorama construction from multiview cameras in outdoor scenes, real-time applications in robot navigation and intelligent control, adaptive surveillance algorithms, vision technologies for civil aviation, navigation of autonomous underwater vehicles, denoising algorithms for intelligent recognition systems, and image segmentation based on 2D Markov chains.

The book is directed to professors, researchers, and software developers working in the areas of digital video processing and computer vision technologies.

We wish to express our gratitude to the authors and reviewers for their contribution. The assistance provided by Springer-Verlag is acknowledged.

Russia — Margarita N. Favorskaya
Australia — Lakhmi C. Jain

Contents

About the Editors

Margarita N. Favorskaya received her engineering diploma from Rybinsk State Aviation Technological University, Russia, in 1980 and was awarded a Ph.D. by S.-Petersburg State University of Aerospace Instrumentation, S.-Petersburg, in 1985. Since 1986 she worked as an Associate Professor of Siberian State Aerospace University, Krasnoyarsk. Margarita Favorskaya defended her doctorial dissertation in Siberian Federal University in 2011. Since 2011 she is a Professor and a Head of Department of Informatics and Computer Techniques at Siberian State Aerospace University.

Her main research interests are digital image and videos processing, pattern recognition, fractal image processing, artificial intelligence, information technologies, and remote sensing. She is the author or the co-author of nearly 130 scientific publications and 20 educational manuals in these fields. Margarita Favorskaya is a member of KES organization, IPC member of International Conferences, and Co-Chair of Invited Sessions. She serves as a Reviewer, a Guest Editor, and an Associate Editor in International Journals.

Lakhmi C. Jain is with the Faculty of Education, Science, Technology, and Mathematics at the University of Canberra, Australia and University of South Australia, Australia. He is a Fellow of the Institution of Engineers Australia.

Dr. Jain founded the KES International for providing a professional community the opportunities for publications, knowledge exchange, cooperation, and teaming. Involving around 5,000 researchers drawn from universities and companies world-wide, KES facilitates international cooperation and generate synergy in teaching and research. KES regularly provides networking opportunities for professional community through one of the largest conferences of its kind in the area of KES. www.kesinternational.org.

His interests focus on the artificial intelligence paradigms and their applications in complex systems, security, e-education, e-healthcare, unmanned air vehicles, and intelligent agents.

Chapter 1
Practical Matters in Computer Vision

Lakhmi C. Jain and Margarita N. Favorskaya

Abstract A brief description of researches close to implementation in technical systems is represented in this chapter. Human action recognition and audience analysis systems as well as smart software tool for panorama construction help for a well-being of a human. The application of novel methods in robot navigation systems and the perception of audio visual information for mobile robots are the issues of other innovative investigations. The adaptive comprehensive surveillance algorithms for situation analysis, the enhanced, synthetic, and combined vision technologies for civil aviation, and the navigation techniques reflect the recent achievements in machine vision for robotics and autonomous vehicles. Also the efficient denoising algorithms and the image segmentation based on 2D Markov chains are useful in intelligent recognition systems.

Keywords Video surveillance · Face recognition · Panorama construction · Robot navigation · Avionics · Autonomous vehicle

1.1 Introduction

Video surveillance has a wide variety of applications in outdoor and indoor environment. Even a single video sequence provides the redundant data while information from multi-cameras combining the data from other sensors (acoustic signals, deep of scene data, odometer data, tactile information, etc.) is transformed in large scale data. On the other hand, in practice only the recognition of some objects or events as well as

L.C. Jain (✉)
Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, ACT 2601, Australia
e-mail: lakhmi.jain@unisa.edu.au

M.N. Favorskaya
Institute of Informatics and Telecommunications, Siberian State Aerospace University, 31 Krasnoyarsky Rabochy, Krasnoyarsk 660014, Russian Federation
e-mail: favorskaya@sibsau.ru

M.N. Favorskaya and L.C. Jain (eds.), *Computer Vision in Control Systems-2*, Intelligent Systems Reference Library 75, DOI 10.1007/978-3-319-11430-9_1

a scene classification is required. Therefore, a development of intelligent systems is the urgent goal for scientific community in computer vision scope. The requirements of real-time implementation with the desired degree of accuracy are the main contradictory criteria for most vision-based systems. Often the analysis of rich experimental results permits to find a way for the reasonable simplification of high-cost algorithms in order to receive the innovative practical decisions.

1.2 Chapters Included in the Book

Two volumes "Mathematical Theory" and "Innovations in Practice" are included in the presented book. Ten original chapters are devoted to development the human surveillance monitoring systems, algorithms and software tools for robots' navigation in various environments, and hardware for novel avionics solutions based on enhanced vision technologies.

Chapter 2 presents a fast growing field of research—a human action recognition extracted from videos as an important scope with numerous applications in the area of computer vision [1]. Applications for human action recognition include video surveillance, video indexing, and human-computer interface. In the case, where a video is segmented to contain a single implementation of human activity, the objective of the system is to classify video data into its respective activity category [2]. In this chapter, a number of existing methods for human activity recognition are discussed. First, the object extraction is done in each frame by performing background subtraction. Second, the frame normalization is applied by using a horizontal frame alignment and a resizing. Third, an Aligned Motion Image (AMI) is computed. Fourth, the video normalization is applied in each video allowing the cropping boundary box to go around the AMI, and subsequently unifying the size for all videos. Finally, a structure similarity measurement is applied to find the similarity between the different AMIs [3]. At the end, an application example of a new algorithm for human action recognition is presented. These results are very close to the recognition of a human observer. The result is about 96.774 % of correct recognitions by using all frames or "complete sequence" in each video. The result is about 98.925 % by using the first 30 frames or "sub sequence" from each video. The obtained experimental results demonstrate a high level of accuracy and efficiency of the proposed methodology.

Chapter 3 is devoted to automatic video data analysis of face and gender recognition. The promising practical applications of face recognition algorithms can be used in modern biometric control system, visitors calculation systems, throughput control on the entrance of office buildings, airports, automatic systems of accident prevention, intelligent human-computer interfaces, and some others [4]. The gender recognition can be applied to collect and estimate the demographic indicators [5]. Besides that it can be an important pre-processing step of person identification, the gender recognition allows twice to reduce the number of candidates for analysis, and thus twice to accelerate the identification process. A human age estimation is

another problem in the field of computer vision, which is connected with face analysis [6]. Among its possible applications one should note an electronic customer relationship management, a security control, a surveillance monitoring, and biometrics. In order to organize a completely automatic system, the classification algorithms are utilized in the combination with a face detection algorithm, which selects candidates for further analysis [7]. The quality of face detection step is critical to the final result of the whole system, as inaccuracies at face position determination can lead to wrong decisions at the stage of recognition. To solve the task of face detection, the AdaBoost classifier is utilized. The detected fragments are preprocessed to align their luminance characteristics and transform them into a uniform scale. On the next stage, the detected and preprocessed image fragments are passed to the input of gender recognition classifier, which makes a decision on their belonging to one of two classes. The same fragments are also analyzed by the age estimation algorithm, which divides them into several age groups. To estimate the period of a person's stay in the range of camera's visibility, a face tracking algorithm is proposed. This chapter describes the main algorithmic techniques utilized in different stages of the proposed video data analysis system, which provides collection and processing of information about the audience in real time. The level of gender and age classification accuracy are estimated in real-life situations. The algorithms proposed in this chapter incorporate the universal machine learning techniques, and thus can be applied to solve other object classification tasks.

Chapter 4 investigates the issues of panorama construction by using images received from several cameras, which are maintained on mobile robots, or by handheld shooting [8]. The main researches presented in this chapter connect with geometrical alignment of selected frames, color enhancement in shadow and bright areas, and seamless frames stitching. The accuracy of each algorithm is achieved by a high computational cost and non-real time realization. Therefore, several ways have proposed to increase the time execution with a suitable accuracy reduction [9]. A stitching of frames is the main core in the panorama construction, which includes the procedures for detection and matching of the similar regions. In this research, the feature points approach is applied, particularly in speeding robust feature detector. The estimation of feature point's correspondences is executed by using famous RANdom SAmple Consensus (RANSAC) algorithm. The luminance enhancement of selected frames is the necessary processing stage. The classical retinex algorithm normalizes dark areas and provides a result image with large contract values. The Enhanced Multi-Scale Retinex (EMSR) algorithm equalizes adaptively the spectral ranges of dark and bright areas for a single frame [10]. A special function stretches the spectral ranges with low and high intensity values due to a reduction of the middle spectral range. The improvement of visibility into the stitching areas is often required in final panoramic images [11]. The point-based rendering attracts a high interest in geometric modeling as an alternative to triangle meshes. The following improvement of blending is connected with the multi-scale or the multi-orientation sub-bands with a number of bands, not more than 3–4.

The main idea is to blend the sun-bands of image with various blending degree. The rich experimental materials are represented in this chapter.

Chapter 5 studies the motion estimation methods based on salient feature points such as Harris corner points, scale invariant feature points, etc. in robot navigation systems. Such artifacts as a fast motion with changing directions, overlapping of moving objects, and a complex background require a novel technique in real-time navigation and control tasks [12]. This technique ought to be able to cope with detection and tracking of boundary curves of the main objects in the image and the objects themselves. In fast motion (e.g. in sport games), a human vision provides the navigation of a sportsman based on a small number of features such as critical objects perceived as contrast (frequently, colored) blobs [13]. The idea to provide such description into a real-time computer vision is in the backbone of the proposed method. It is well known that the lack of real-time techniques of stable image segmentation limits the capabilities of mobile robots to understand scenes and solve the localization (categorization) problem. The chapter describes a version of the required technique and its application in indoor (outdoor) robot navigation. The proposed method makes it possible to select objects in complex scenes and provide their recognition based on the obtained generalized geometric descriptions in the language of collections of intervals [14]. The main restriction is connected with overall low saturation of colors in the image. Several applications of a novel real-time method of contextual image description are presented in the chapter. In the first application, a robot is navigated using artificial color visual landmarks (e.g. ones composed of colored rectangles). Visual landmark may be put on a wall, be in hands of a human, or be mounted on another robot (moving landmarks). In the second application, a robot finds doors (opened or closed), windows, walls, etc., while running in indoor environment.

Chapter 6 introduces a precise mobile robot motion control with clear orientation in the area of robot perception and observation [15]. In this chapter, the mobile robot audio and visual systems are outlined. They use data from corresponding audio (microphone array) and video (mono, stereo, or infrared cameras) sensors, accompanied with laser range finder sensor. The audio and video information captured from the sensors is used in the perception audio visual model. The proposed model performs a joint processing of audio visual information and determines a current mobile robot position (current space coordinates) in the area of robot perception and observation. The captured from audio visual sensors information is estimated with suitable algorithms developed for a speech and image quality estimation [16]. The preprocessing methods for increasing a quality and minimizing the errors of mobile robot position are developed. The current space coordinates, determined from a laser range finder, are used as supplementary information of mobile robot position. Also space coordinates are applied for error calculation and comparison with the results from audio visual mobile robot motion control. In the development of the mobile robot perception audio visual model, some methods are used [17]. The RANSAC method estimates parameters of a mathematical model from a set of observed audio visual coordinate data. The method of direction of arrival determines a localization of sound source direction

with microphone array of speaker sending voice commands to the mobile robot. The method for speech recognition classifies the voice commands sending from the speaker to the robot. The current mobile robot position is calculated from a joint usage of perceived audio visual information. It is used in appropriate algorithms for mobile robot navigation, motion control, and objects tracking: a map-based or map less methods, path planning and obstacle avoidance, simultaneous localization and mapping, data fusion, etc. The error, accuracy, and precision of the proposed mobile robot motion control with perception of audio visual information are analyzed and estimated from the results of the numerous experimental tests, which are presented at the end of this chapter. The experiments are carried out mainly with simulations of the algorithms listed above. The parallel computing methods in implementation of the developed algorithms permit to reach a real time robot navigation and motion control using perceived audio visual information from the mobile robot audio visual sensors.

Chapter 7 provides some automatic surveillance solutions under an uncertainty and a variability of characteristics for an observed scene such as inaccuracy definition of objects coordinates and the mutual location of objects, illumination distortions and partial or complete objects overlapping, shadows and noises. The implementation of adaptive comprehensive algorithm for image processing and analysis is one of directions improving the efficiency of surveillance under uncertain conditions [18]. The algorithm provides a possibility to observe the object actions under complex and changing surveillance conditions. Also such approach is useful for situation analysis in less informative scenes. The novelty of this approach is based on the original descriptions of objects into complex scenes and inter-objects relationships. Such descriptions involve the elements of language contingency management that makes them more robust to various destabilizing factors. Algorithms of situation analysis provide a decision-making based on the choice of valid extracted features in accordance with the target task [19]. The entropy estimations (initial, current, and final) permit to decrease an uncertainty of complex scenes with large numbers of moving objects. The situation analysis reduces the initial and/or current entropy estimations during the procedure of object detection [20]. The software tool for a situation analysis is realized by using specialized database, knowledge base, and model descriptions. Database contains the descriptions of observed objects and inter-object relationships (spatial, temporal, causal, etc.). Knowledge base includes production rules describing the causal relations between objects and situations. It provides the final decision-making for control system. The model descriptions are built by considering the target tasks and environment models based on apriori and current information. The designed adaptive algorithms may be used in many applications for mobile robots and vehicles of various types including unmanned aerial vehicles.

Chapter 8 presents the innovations in avionics solutions aimed to enhance a flight visibility and a situational awareness of a flight crew. Such solutions are based on Enhanced Vision System (EVS), Synthetic Vision System (SVS), and Combined Vision System (CVS) [21]. These systems provide a supplemental view of external cabin space for a flight crew using technical vision, computer

graphics, and augmented reality [22]. The chapter addresses the main aspects of the EVS/SVS/CVS technologies development and includes the main topics such as the EVS/SVS/CVS typical applications, the overview of well-known commercial and experimental systems, the cores of advanced EVS/SVS/CVS technologies, among others. The EVS generates an enhanced image of external cabin space in a real time. The main EVS functions connect with image enhancement, image fusion from visible and infrared ranges, interconnections of multispectral visual information, creation of superresolution image by using a set of low resolution frames, video stabilization, automatic binding and visual combination of enhanced image with a flight symbology, automatic runway and obstacles detection in the landing zone and taxiing, etc. The digital terrain modeling based on a photogrammetric processing of 2D/3D data and a synthesized image creation based on navigation, terrain, and obstacles layers fusion are the main tasks of the SVS [23]. The representation of topological map as vector graphical patterns of flight corresponding to real environment is produced by the on-board computer equipment. It is needed to create 3D view of external cabin space with enough scene depth to sense a relative distance to real objects by a flight crew. The automatic recognition and the flight symbols representation of potentially dangerous events are also provided by the SVS. The CVS binds and integrates the sensory and geospatial information to implement a human-machine interaction based on virtual and augmented realities. Such enhanced images are shown on the primary flight display. The chapter contains some examples of the EVS, the SVS, and the CVS images, which have been obtained during the research and development program initiated by Russian State Research Institute of Aviation Systems "GosNIIAS".

Chapter 9 discusses the accurate in-situ position estimation of navigation system for an Autonomous Underwater Vehicle (AUV). Among the different navigation principles, acoustic and vision-based ones became the most popular [24]. The acoustic navigation uses the Time-Of-Flight (TOF) measurements of ultrasound waves propagating from a stationary buoy or a set of buoys with known location to a hydrophone, which is maintained on AUV. The vision-based navigation is based on the analysis of snapshots series receiving from an on-board optical camera. The operating range of the acoustic systems is typically from 1 m up to 10 km, while the working range of the optical navigation is reduced to several cm. A vehicle mission is to take a certain orientation in the space relatively to an underwater target. An operation model includes two steps called as a long-distant and a near-distant guidance. The long-distant guidance is based on the TOF measurements of acoustic signals, which can be one-way synchronous or one-way asynchronous signals from an acoustic buoy or two-way asynchronous guidance, when the AUV transmits a pilot signal, which is replicated by the buoy and is registered back on the vehicle side. The one-way asynchronous guidance was selected. Such approach does not require any synchronization and provides a possibility to navigate multiple AUVs. In general, the long-distant guidance can be viewed as a problem of time delay minimization between the transducer (acoustic buoy) and the receiver (AUV's hydrophone) [25]. This problem is divided into two sub-tasks including the accurate TOF measurements and the estimations translation into engine control signals.

Addressing the first task, a cross-correlation technique based on pseudo-noise sequences is applied. For the second task, a Proportional-Integral-Derivative (PID) controller is used. The further positioning is continued by an image analysis for series of digital snapshots. The near-distant guidance aims to improve the position of the AUV and justify its course in accordance to an underwater target [26]. An image processing algorithm computes a lateral transition, rotation, and scaling between a snapshot and a stored target sample. An input snapshot is convolved with a Gaussian kernel in order to reduce an image noise and remove the non-significant details. The normalized gradient image is transformed to binary image by a pre-defined threshold. A target center is calculated as a center of mass. Then the unknown angular difference and scale factor are estimated in a log-polar system. The total computational complexity of the developed algorithm is $O(MN \log(\max(M, N)))$, where $M * N$ is the number of pixels in the digital image. The PID controller is used to manipulate the position and orientation of the AUV during a near-distant guidance mode. In order to evaluate an accuracy and robustness of the developed model, a 3D simulator was implemented, where the series of numerical experiments with different underwater targets was carried out. The experiments have shown a high accuracy of the proposed approach, which may be successfully used in real AUV navigation system.

Chapter 10 examines various digital image filtration algorithms [27], which are useful in great variety of video devices – cameras, mobile phones, scanners. These algorithms eliminate the distortions and other blurring effects and improve "raw" images for further specific applications. Among the most known filtration models are an Additive White Gaussian Noise (AWGN) model and a mixed noise model. As the AWGN model is suitable for description of effect of multiple noise sources caused by digital devices. A mixed noise model describes better a Complementary Metal-Oxide Semiconductor (CMOS) matrix noise. Therefore, these models were taken as the primary for investigations. For denoising purposes, algorithms based on Principal Component Analysis (PCA) and non-local processing were used [28]. The idea of the PCA is in a change of image representation in order to reduce the data dimensionality with minimizing of the mean square error. The core functions are a block representation of image, a blocks' transform into a set of the PCA coefficients, a coefficients processing, and an image reconstruction. The non-local processing is connected with the evaluation of a certain image's pixel using an average of all weighted values of pixels in a neighborhood. The calculation of pixel weight is based on a degree of similarity between neighborhoods in a noisy and denoisy images. However, most of the modern filtration algorithms implemented for grayscale or color images cannot be directly applied for "raw" (color filter array) images having the dependencies between color components. In addition, each of them possesses its own pros and cons in terms of image reconstruction quality. The main problems of the quality of reconstructed images, which researchers try to evaluate, are a Gibbs effect, which becomes highly noticeable in images containing objects with a high brightness contrast in outer edges, and an edge blurring in images. Both of these effects highly degrade an image perception and could not be suited for high demands [29]. The chapter provides the analysis of such effects and

their compensation in a filtration stage. The listed features were formulated in order to implement a series of denoising algorithms, each of those was specifically designed to achieve a high image quality. They can be applied in multimedia transmission systems, digital video broadcasting, pattern recognition, object tracing, and other practical applications. The recommendations of further development and limitations of use are situated at the end of chapter.

Chapter 11 provides the methods of image segmentation in systems for monitoring of the Earth surface, disease diagnosis, and safety of various large technical objects [30, 31]. A complex inhomogeneous background in image and sometimes a low signal-to-noise ratio do not allow to apply the simple solutions. It is proposed to use the mathematical theory of conditional Markov processes, the representation of g-bit grayscale images as a set of g-binary planes, and the entropy approach for calculation the state elements probabilities [32]. This approach has allowed to develop the novel efficient methods of contour and texture segmentation for objects of interest in images. For each element, certain information content is calculated to detect the contours of objects. By comparison the obtained values with a predetermined threshold, one can decide, the given point belongs for a contour or not. The developed method of contour segmentation using the calculated value of information content detects the objects of interest with a high accuracy and requires significantly less computational resources than conventional methods (Canny, Laplacian of Gaussian, Roberts, Prewitt, and Sobel). To detect the object contours in image transmitted over a noisy radio channel, it is necessary the efficient filtering of images. For image restoration distorted by the white Gaussian noise, it is proposed to use 2D nonlinear filtering algorithm. Having more accurate estimates of the image's elements states, and taking into account the transitions probability between elements, it is possible to outline the edges of objects of interest. The proposed method is effective under the signal-to-noise ratio for an input of receiver device up to 9 dB. Also the method of texture segmentation has been proposed for the determination of extensive areas with similar statistical characteristics of satellite images such as areas of the forest, the areas of urban developments, etc. The novel method is based on the calculation of average transition probabilities in the image's elements using a slicing window. Experimental results confirm that the developed algorithms for objects and texture areas detection in images are effective in terms of quality and processing speed.

1.3 Conclusion

All included chapters contain the innovative decisions in computer vision for control and surveillance systems. To receive the real-time implementations is the main goal for close to practice tasks. The efforts direct on the development of robust, exact, and fast methods and algorithms. Many of represented works have an experimental software implementation as a result of previous many-years researches. The chapters include the recent methodologies for human action recognition,

real-time audience analysis system, panorama construction from multi-view cameras in outdoor scenes, real-time applications in robot navigation and intelligent control, adaptive surveillance algorithms, enhanced, synthetic and combined vision technologies for civil aviation, navigation of autonomous underwater vehicles, efficient denoising algorithms for intelligent recognition systems, image segmentation based on 2D Markov chains. Each chapter explains in detail algorithmic and software/hardware implementations in the chosen area of researches.

References

1. Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. ACM Comput. Surv. 43(3): 16:1–16:43
2. Dalal N, Triggs B, Schmid C (2006) Human detection using oriented histograms of flow and appearance. In: European conference on computer vision (ECCV 2006), pp 428–441
3. Amraji N, Mu L, Milanova M (2011) Shape–based human actions recognition in videos. In: 14th international conference on human-computer interaction: design and development approaches, vol. 1, pp 539–546
4. Li SZ, Jain AK (2005) Handbook of face recognition. Springer, Berlin
5. Makinen E, Raisamo R (2008) An experimental comparison of gender classification methods. Pattern Recogn Lett 29(10):1544–1556
6. Fu Y, Huang TS (2010) Age synthesis and estimation via faces: a survey. IEEE Trans Pattern Anal Mach Intell 32(11):1955–1976
7. Khryashchev V, Ganin A, Golubev M, Shmaglit L (2013) Audience analysis system on the basis of face detection, tracking and classification techniques. In: International multi-conference of engineers and computer scientists (IMECS 2013) 1:446–450
8. Haenselmann T, Busse M, Kopf S, King T, Effelsberg W (2009) Multi perspective panoramic imaging. Image Vis Comput 27(4):391–401
9. Kwon OS, Ha YH (2010) Panoramic video using Scale Invariant Feature Transform with embedded color-Invariant values. IEEE Trans Consum Electron 56(2):792–798
10. Favorskaya M, Pakhirka A (2012) A way for color image enhancement under complex luminance conditions. In: Watanabe T, Watada J, Takahashi N, Howlett RJ, Jain LC (eds) Intelligent interactive multimedia: systems and services. Springer, Berlin
11. Zhao G, Lin L, Tang Y (2013) A new optimal seam finding method based on tensor analysis for automatic panorama construction. Pattern Recogn Lett 34(3):308–314
12. Bonin-Font F, Ortiz A, Oliver G (2008) Visual navigation for mobile robots: a survey. J Intell Robot Syst 53(1):263–296
13. Kiy KI, Dickmanns ED (2004) A color vision system for analysis of road scenes. In: IEEE intelligent vehicle'04 symposium, pp 54–59
14. Kiy KI (2010) A new real-time method for description and segmentation of color images. Pattern Recogn Image Anal Adv Math Theory Appl 20(2): 169–176
15. Jarvis R (2008) Intelligent robotics: past, present and future. Int J Comput Sci Appl Technomathematics Res Found 5(3):23–35
16. Bekiarski Al, Pleshkova Sn (2009) Microphone array beamforming for mobile robot. In: 8th WSEAS international conference on circuits, systems, electronics, control and signal processing (CSECS'2009), pp 146–149
17. Dehkharghani SSh, Bekiarski Al, Pleshkova Sn (2012) Application of probabilistic methods in mobile robots audio visual motion control combined with laser range finder distance measurements. In: Biolek D, Volkov K, Ng KM (eds) Advances in circuits, systems, automation and mechanics. WSEAS Press, Greece

18. Tulum K, Durak U, Yder SK (2009) Situation aware UAV mission route planning. In: IEEE aerospace conference, pp 1–12
19. Osipov GS, Smirnov IV, Tikhomirov IA (2012) Formal methods of situational analysis: experience from their use. Autom Doc Math Linguist 46(5):183–194
20. Leishman RC, McLain TW, Beard RW (2014) Relative navigation approach for vision-based aerial GPS-denied navigation. J Intell Rob Syst 74(1–2):97–111
21. Bailey RE (2012) Awareness and detection of traffic and obstacles using synthetic and enhanced vision systems. NASA technical memorandum, 2012-217324 NASA, pp 54–60
22. Kumar SV, Kashyap SK, Kumar NS (2014) Detection of runway and obstacles using electro-optical and infrared sensors before landing. Defense Sci J 64(1):67–76
23. Vizilter Yu, Zheltov SY (2012) Geometrical correlation and matching of 2D image shapes. ISPRS Ann Photogrammetry Remote Sens Spat Inf Sci 1–3:191–196
24. Sangekar M, Thornton B, Ura T (2012) Wide area seafloor observation using an autonomous landing vehicle with adaptive resolution capability. Oceans 2012:1–9
25. Burdinsky IN (2012) Guidance algorithm for an autonomous unmanned underwater vehicle to a given target. Optoelectron Instrum Data Process 48(1):69–74
26. Bezruchko F, Burdinky I, Myagotin A (2011) Global extremum searching algorithm for the AUV guidance toward an acoustic buoy. IEEE OCEANS'2011, pp 1–7
27. Buades A, Coll B, Morel JM (2005) A review of image denoising algorithms, with a new one. Multiscale Model Simul 4:490–530
28. Katkovnik V, Foi A, Egiazarian K, Astola J (2010) From local kernel to nonlocal multiple-model image denoising. Int J Comput Vision 86(8):1–32
29. Priorov A, Tumanov K, Volokhov V, Sergeev E, Mochalov I (2013) Applications of image filtration based on principal component analysis and nonlocal image processing. IAENG Int J Comput Sci 40(2):62–80
30. Martin D, Fowlkes C, Malik J (2004) Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans Pattern Anal Mach Intell 26(5):530–549
31. Wang H, Dong Y (2008) an improved image segmentation algorithm based on otsu method. In: International symposium on photoelectronic detection and imaging: related technologies and applications, SPIE 6625, pp 1–8
32. Petrov EP, Trubin IS, Medvedeva EV, Smolskiy SM (2013) Mathematical models of video-sequences of digital half-tone images. In: Atayero AA, Sheluhin OI (eds) Integrated models for information communication system and networks: design and development. IGI Global, Hershey

Chapter 2
Human Action Recognition: Contour-Based and Silhouette-Based Approaches

Salim Al-Ali, Mariofanna Milanova, Hussain Al-Rizzo and Victoria Lynn Fox

Abstract Human action recognition in videos is a desired field in computer vision applications since it can be applied in human computer interaction, surveillance monitors, robot vision, etc. Two approaches of features are investigated in this chapter. First approach is a contour-based type. Four features are investigated in this approach such as Cartesian Coordinate Features (CCF), Fourier Descriptors Features (FDF), Centroid-Distance Features (CDF), and Chord-Length Features (CLF). The second approach is a silhouette-based type. Three features are investigated in this approach such as Histogram of Oriented Gradients (HOG), Histogram of Oriented Optical Flow (HOOF), and Structural Similarity Index Measure (SSIM) features. All these features are simple to compute, efficient to classify, and fast to calculate. Therefore, these features demonstrate a promising field for human action recognition. Moreover, the classification is achieved using two classifiers: K-Nearest-Neighbor (KNN) and Support Vector Machine (SVM). The experimental results demonstrated that these features have a promising potential and useful for the human action recognition in videos.

S. Al-Ali (✉) · M. Milanova
Department of Computer Science, University of Arkansas at Little Rock,
2801 S. University Avenue, Little Rock, AR 72204, USA
e-mail: sgsaeed@ualr.edu

M. Milanova
e-mail: mgmilanova@ualr.edu

H. Al-Rizzo
Department of System Engineering, University of Arkansas at Little Rock,
2801 S. University Avenue, Little Rock, AR 72204, USA
e-mail: hmalrizzo@ualr.edu

V.L. Fox
Department of Applied Science, University of Arkansas at Little Rock,
2801 S. University Avenue, Little Rock, AR 72204, USA
e-mail: vlfox@ualr.edu

M.N. Favorskaya and L.C. Jain (eds.), *Computer Vision in Control Systems-2*,
Intelligent Systems Reference Library 75, DOI 10.1007/978-3-319-11430-9_2

Keywords Human action recognition · Contour-based features · Silhouette-based features · K-Nearest-Neigbor · Supprt Vector Machine · Image and video understanding · Machine learning · Data mining

2.1 Introduction

Currently, computer application fields are playing significant role in multiple aspects of our lives. One important field is a computer vision, which has received a lot of attention during the past three decades due its wide applications. The human action recognition is an important goal of research on computer vision and image processing. Identifying, annotating, recognizing, and clustering human actions in videos have captured more and more attention because of its useful applications that support many different applications such as human–computer interaction, robot vision machine, human surveillance monitoring system, multimedia indexing and retrieval, entertainment environments, and healthcare systems [1, 2, 3].

The human action recognition in videos is a computer method for recognizing and identifying, what kind of action is happening in videos. In order to design and implement this program, there are many challenges such as foreground object, background scene, and camera setting. The foreground object, which is the human in this case, has many variations such as size, colour, shape, static or moving object, etc. The background scene, which is a whole image in a frame except the foreground object, has many variations such as lighting, occlusion, cluttered, static or moving background scene (based on camera setting). The camera setting is an important factor in the human action recognition because it has its own recording variations such as static or moving in all (left, right, up, or down) directions, zooming (in or out), speeds of recording (slow or high), recording types (2D or 3D), colors types in recording videos (black/white, colored, or grayscale color), etc. Moreover, the same action is performed in different ways by the same person, for example, the speed of walking is different although that the walking action is for the same person. Another challenging problem, which is more difficult and very realistic, is that the same action performed by different people. Although of these challenges, the human action recognition in videos is desired and required for many computer vision applications.

In recognizing human actions in videos, many researches have been reported in this field as shown in the survey paper [4], however, there still need to improve and develop new effective approaches. The presented chapter is a new investigation of two main features approaches (contour-based and silhouette-based) for human action recognition in videos.

In this chapter, main structure of human action recognition is defined. The structure mainly consists of three stages: human object tracking, feature extraction, and action classification. Some examples regarding literature researches of these three stages and the recent related works are given in Sect. 2.2. Subsequently,

background subtraction [5, 6] is explained in detail as an example for stage of human object tracking in videos in Sect. 2.3. Next, two approaches: contour-based and silhouette based for feature extraction from the tracked human object are described in Sects. 2.4 and 2.5, respectively. The final stage in the human action recognition is action classification stage that used to classify and identify the action happening in human action testing video. Two classifiers: KNN [7, 8, 9], and SVM [10, 11, 12, 13] are used as examples for action classification stage. More details about the classifiers and their based techniques are explained in Sect. 2.6. Two modes (training and testing) of the presented algorithm for human action recognition are described in Sect. 2.7. Experimental results are discussed in Sect. 2.8. Finally, Sect. 2.9 conducted the conclusion.

2.2 Human Action Recognition in Videos

The main goal of human action recognition in videos is to identify the unknown actions happening in these videos. This goal is achieved by analyzing the frames of these videos to form and build a series of discriminant features that can be classified efficiently in term of accuracy, speed, and simplicity. The main structure of the human action recognition consists of three main stages: human object tracking, feature extraction, and action classification. The first stage has to answer the question of how to detect or segment and track the human object in each frame of the video sequences. The second stage has to answer the question of how to extract, represent, and then build feature vector from the tracked human object that result from the first stage. The third stage has to answer the question of how to classify extracted features from the second stage by applying an effective classification algorithm. Sometimes, this stage supported by data mining process to reduce dimensionality of the extracted features. In the next sections, answers and details about these three stages will be provided. The main structure of a human action recognition system is depicted in Fig. 2.1.

Section 2.2.1 addresses the human object tracking. The issues of feature extraction and action classification are discussed in Sects. 2.2.2 and 2.2.3, respectively. A literature review about human action recognition in Weizmann dataset is represented Sect. 2.2.4.

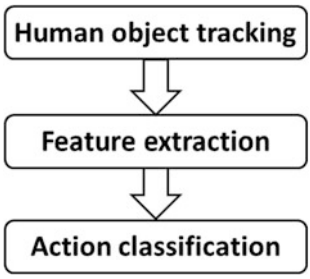


Fig. 2.1 Main structure of the human action recognition

2.2.1 Human Object Tracking

The human object tracking is a process of tracking a human object moving over sequence (time) of digital images (frames) in videos [14]. Generally, this process consists of two components: frame processing (local) and video processing (global). The first is achieved by the human object detection or segmentation. The human detection is the process of locating a human object in a frame of video. The segmentation of human object is the process of partitioning a frame into multiple segments (areas or sets). One of these separated segments represents the human object. Both detection and segmentation are mainly related to one frame (digital image) in videos and, therefore, are called frame processing. The second component is achieved by applying frame processing over all-frames (video) or sub-frames (sub video), thus, it is called video processing.

During the past three decades, many researchers solved problem of tracking objects in still image, in a frame, or videos. These solutions are achieved by several ways: point detection, image segmentation, and background modeling. First, the point detection is used for tracking based on some interesting points such as corners, or intersection points such as Harris detector [15], Scale-Invariant Feature Transform (SIFT) [16], affine invariant interest point detector [17], kernel-based object tracking [18], and Kanade-Lucas-Tomasi (KLT) detector [19]. Second, the image segmentation is a process to partition a digital image (frame) into multiple segments (sets of separated areas), used to track an object such as mean-shift [20], graph-cut [21], and active-contours [22]. Third, the background modeling is also another process used in the tracking. The goal is to obtain and build a model for the background scene. Then, the object extraction is achieved by subtracting each frame from this model, such as running Gaussian average [23], temporal median filter [24, 25], Mixture Of Gaussian (MOG) [26, 27], eigenbackground [28], and dynamic texture background [5]. More details and examples for human object tracking using a background subtraction in Weizmann human action dataset [2] are explained in Sect. 2.3.

2.2.2 Feature Extraction

The feature extraction is a process of extracting a set of features to represent some useful measurements or characteristics of a frame or video. These features are computed carefully from a frame, sub-frames, or all-frames in video efficiently in order to capture most important meaningful details. The goal of feature extraction is to provide a classifier by good feature in terms of accuracy and speed. This goal can be achieved by two ways: minimizing feature details as much as possible and at the same time maximizing features discrimination in order to increase accuracy and speed of classification in the next stage. There are several ways to enhance the feature extraction [29]. First, extracting spatial information is more related to frame