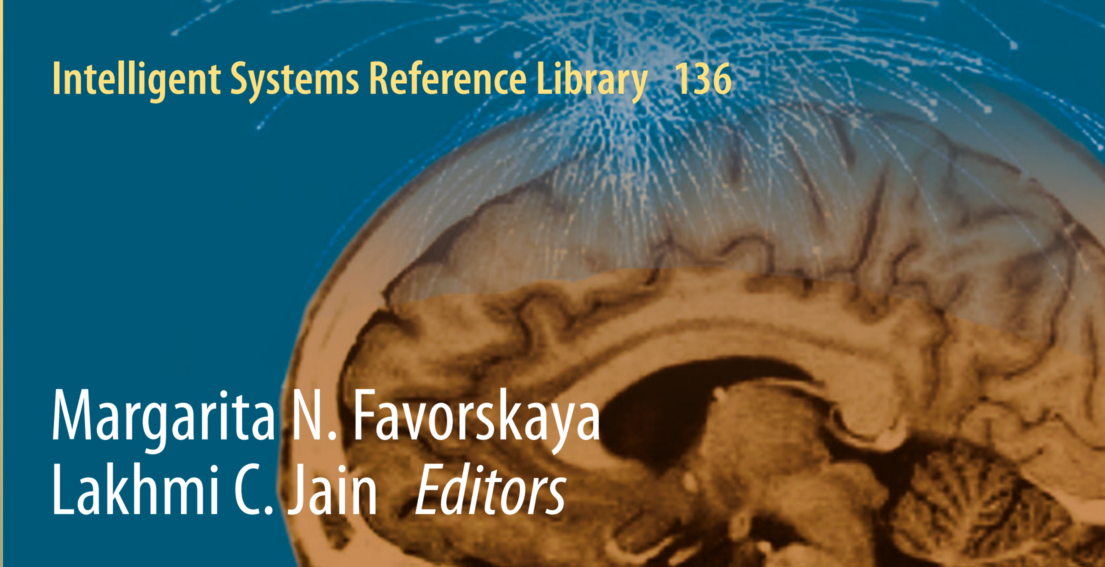

Intelligent Systems Reference Library 136

Margarita N. Favorskaya
Lakhmi C. Jain *Editors*

# Computer Vision in Control Systems-4

## Real Life Applications

Springer

# Intelligent Systems Reference Library

Volume 136

**Series editors**

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-4

## Real Life Applications

Springer

*Editors*
Margarita N. Favorskaya
Institute of Informatics and Telecommunications
Reshetnev Siberian State University of Science and Technology
Krasnoyarsk
Russia Federation

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

and

Bournemouth University
Poole
UK

and

KES International
Shoreham-by-Sea
UK

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-319-67993-8 ISBN 978-3-319-67994-5 (eBook)
https://doi.org/10.1007/978-3-319-67994-5

Library of Congress Control Number: 2017952893



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

The research book is a continuation of our previous books which are focused on the recent advances in computer vision methodologies and technical solutions using conventional and intelligent paradigms.

- Computer Vision in Control Systems-1, Mathematical Theory, ISRL Series, Volume 73, Springer-Verlag, 2015
- Computer Vision in Control Systems-2, Innovations in Practice, ISRL Series, Volume 75, Springer-Verlag, 2015
- Computer Vision in Control Systems-3, Aerial ans Satellite Image Processing, ISRL Series, Volume 135, Springer-Verlag, 2018

The research work presented in the book includes a number of real-life applications including the identification of handwritten texts, watermarking techniques, the mobile robot simultaneous localization and mapping, motion control systems of mobile robots, analysis of indoor human activity, face image quality assessment, android device controlling, medical images processing, clinical decision-making and foot progression angle detection.

The book is directed to the Ph.D. students, 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.

Krasnoyarsk, Russian Federation — Margarita N. Favorskaya
Canberra, Australia — Lakhmi C. Jain

# Contents

# About the Editors

**Dr. Margarita N. Favorskaya** is a Professor and Head of Department of Informatics and Computer Techniques at Siberian State Aerospace University, Russian Federation.

Professor Favorskaya is a member of KES organization since 2010, the IPC member and the Chair of invited sessions of international conferences. She serves as a reviewer in international journals (Neurocomputing, Knowledge Engineering and Soft Data Paradigms, Pattern Recognition Letters, Engineering Applications of Artificial Intelligence), an associate editor of Intelligent Decision Technologies Journal and Computer and Information Science Journal. She is the author or the co-author of 160 publications and 20 educational manuals in computer science. She co-edited three books for Springer recently. She supervised eight Ph.D. candidates and presently supervising five Ph.D. students.

Her main research interests are digital image and videos processing, remote sensing, pattern recognition, fractal image processing, artificial intelligence and information technologies.

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

Professor Jain founded the KES International for providing a professional community the opportunities for publications, knowledge exchange, cooperation and teaming. Involving around 5000 researchers drawn from universities and companies worldwide, KES facilitates international cooperation and generates 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
# Innovative Algorithms in Computer Vision

**Lakhmi C. Jain and Margarita N. Favorskaya**

**Abstract** This chapter contains a brief description of the methods, algorithms, and implementations applied in many fields of computer vision. The graphological analysis and identification of handwritten manuscripts are discussed using the examples of Great Russian writers. A perceptually tuned watermarking using non-subsampled shearlet transform is a contribution in the development of the watermarking techniques. The mobile robot simultaneous localization and mapping, as well as the joined processing of visual and audio information in the motion control systems of the mobile robots, are directed on the robotics' development. The ambient audiovisual monitoring based on a wide set of methods for digital processing of video sequences is another useful real life application. Processing of medical images becomes more and more complicated due to the enforced current requirements of medical practitioners.

**Keywords** Graphological analysis · Digital watermarking · Simultaneous localization and mapping · Visual and audio decision making · Indoor human activity · Face image quality assessment · Eye detection and tracking Medical image processing · Clinical decision support system · Gait monitoring

L.C. Jain (✉)
Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, ACT 2601, Australia
e-mail: jainlakhmi@gmail.com

L.C. Jain
Bournemouth University, Poole, UK

M.N. Favorskaya
Institute of Informatics and Telecommunications, Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy ave., Krasnoyarsk 660037, Russian Federation
e-mail: favorskaya@sibsau.ru



M.N. Favorskaya and L.C. Jain (eds.), *Computer Vision in Control Systems-4*,
Intelligent Systems Reference Library 136,
https://doi.org/10.1007/978-3-319-67994-5_1

## 1.1 Introduction

The core of any control vision system is the algorithms for raw data processing in such manner that it will be possible to obtain the non-evidence dependences in big data volumes, essential improvement of visual data, or reliable recognition of the objects of interest. Each task requires the special approaches in creation of innovative algorithms, as well as the great experimental work before its implementation in real devices. Sometimes, visual data are not enough for decision making. In these cases, audio data can be attracted successfully. The spectrum of real life applications is very wide. Therefore, this book is an attempt to contribute in some spheres of human activity including culture, robotics, human interactions, and medicine.

## 1.2 Chapters Included in the Book

Chapter 2 includes the issues of graphological analysis and identification of handwritten texts on the examples of the author's calligraphy courtesy provided by the Manuscript Department of the Institute of Russian Literature (Pushkin's House) of the Russian Academy of Sciences [1]. The authors discuss the main challenges of this problem due to great variety of historical texts, when even the samples of writing of some letters can greatly vary depending on the speed of writing and the writing tool used (goose quill, pen point, or pencil). Since a textual analyst follows the own manual algorithm for handwritten text identification, it is possible to design the automated information storage and retrieval system as a useful software tool in order to reduce such long process. Additionally to the existing methods and algorithms of bitmap image transformation, two methods were developed. Vectorization of bitmap images of handwritten texts builds the orientation field of image contours using the following steps: the pre-filtering of the bitmap image, construction of the orientation field for this bitmap image, filtering and extrapolation of the direction field, and searching for and tracking of image contours using the created orientation field. The core idea of this algorithm is to trace the contours in the original image that provides information about the direction of a line or a stroke near a given point. Method of vector dynamic parameterization permits to obtain more information about letter images. This method try to restore the dynamic information about the movement of a pen, when writing a letter, including the indication of the start and end points, movement direction, and the number of cyclic letter outlines. Thus, the vector dynamic representation provides the ability to visualize each letter in the form of a 3D image with $x$, $y$ and $t$ coordinates, where $t$ is a time. Also the reader can find the detailed description of architecture of the designed software tool based on the client-server technique [2]. The system interacts with the database (around 400 Mb memory), which is pre-filled with a collection of different versions of the author's calligraphy of the classics of Russian literature A.S. Pushkin and A.S. Griboyedov (images of individual letters and their ligaments/ligatures).

Chapter 3 investigates the issues of perceptually tuned watermarking using non-subsampled shearlet transform. The multiple classification and criteria of the watermarking techniques permitted to chose very promising transformation called as digital shearlet transform and its modifications for embedding/extraction of gray-scale or color watermarks into the host images [3]. The tampering attacks categorized as the global and local ones distort a watermark; more several types of attacks can be applied for the same watermark. This causes a necessity to develop the robust watermarking algorithms, also considering the contents of the host image and watermark [4]. Such algorithms ought to find a decision in a contradictory desire to embed the maximum data volume (an image payload) and a visibility. The multiplicative watermarking techniques support a balance of the signal magnitude (fidelity) and the quality of the watermarked image (robustness) at an acceptable level. The multi-scale decomposition analysis is provided by many wavelet-like methods, including ridgelet, curvelet, brushlet, wedgelet, beamlet, contourlet, bandelet, directionlet, and shearlet. The use of the digital shearlet transform produces a variety of sub-bands for inserting the secret data due to the shearlets' multi-resolution property. These sub-bands are correlated in the digital wavelet transform but not correlated in the digital shearlet transform that makes a watermarking more secure procedure. At the same time, the non-subsampled shearlet transform is a fully shift-invariant, multi-scale, and multi-directional expansion of the shearlet transform. The proposed algorithm of a watermark embedding and extraction using a watermarking scrambling via Arnold's transform and paradigms of a human visual system demonstrates good experimental results on 44 color and monochrome images from the database Miscellaneous of University of Southern California. The scaling, disproportionate scaling, rotation, translation, and cropping attacks were simulated, and the impact of the attacks on a watermark was estimated by bit error rate and the peak signal to noise ration metrics.

Chapter 4 conducts the issues of the Simultaneous Localization And Mapping (SLAM) algorithm improvement for the task of the mobile robot simultaneous localization and mapping, which consists of two parts: the evaluation of the trajectory of the robot movement and evaluation the locations of landmarks, which depend on the coordinates of the robot at the time of each measurement. [5]. The authors consider the FastSLAM modification using the nonlinear models through the first order Taylor series expansion at the mean of the landmarks' observations. The probabilistic properties of the SLAM algorithm were considered. The proposed algorithm computes a more accurate mean and uncertainty of the landmarks moving nonlinearly. The method of calibration of Kinect-like cameras, depth map restoration using a modified interpolation technique, and filtering the noise in the RGB images are the main preprocessing methods. Additionally, the improved resampling algorithm for the particle filtering through the adaptive thresholding based on the data of the effective particle number evolution was developed. The proposed algorithm runs in real time and shows good accuracy and robustness in comparison with other modern SLAM systems. The average errors in calculating the displacement of the camera between successive frames using the adaptive threshold values is about 24% less than in the case of the strict thresholding.

Chapter 5 implies the development of fast parallel algorithms for joined processing of visual and audio information in the motion control systems of the mobile robots [6]. The main goal is the path planning of a mobile robot in real time implementation using three types of information obtained from the video camera, microphones, and laser range finders [7]. Such algorithms ought to have the appropriate precision and effectiveness in scenes with the moving objects, people, and obstacles. Fast parallel algorithms are implemented using the multiprocessor hardware structure (the well known NVIDIA Graphics Processing Unit (GPU) processor) and software platform (Compute Unified Device Architecture (CUDA) platform). Many mathematical operations widely used in the images, audio, and other signal processing, such as the matrix multiplication, Fast Fourier transform, correlation and convolution, can be executed faster in a comparison to the traditional Central Processing Unit (CPU). During this research, three algorithms were developed. The Algorithm 1 is based on the SLAM method and the mobile robot audio visual perception and attention [8]. The SLAM method collects and keeps the information for all previous executive locations for a decision making about the following locations of a mobile robot. The Extended Kalman Filter SLAM (EKF SLAM) was modified for parallel execution. Algorithm 2 and Algorithm 3 are the developed fast parallel algorithm based on the mobile robot audio perception and attention for a decision making in motion control system managed by a speaking person [9]. They consider the current audio features defining an angle of sound source arrival or location of sound source. The Steered Response Power (SRP) with PHAse Transform (PHAT) called as SRP-PHAT was applied for this purpose. The Algorithm 1, Algorithm 2, and Algorithm 3 demonstrate clearly the advantages and real time execution of using NIVIDA GPU and CUDA platform instead of CPU in the developed fast algorithm for a decision making in motion control system of a mobile robot based on audio visual information.

Chapter 6 promotes the indoor ambient audiovisual monitoring in the intelligent meeting room with a goal to determine the time events, when the states of the participants' activities are changing [10]. The chapter contains a description of some intelligent meeting rooms equipped with audio and video recording equipment [11]. However, the main discussion deals with the methods and algorithms of image and speech analysis, such as the recognition and identification of faces, detection and tracking of participants, identification of participants' positions, voice activity detection, localization and separation of sound sources, identification of speakers, and recognition of speech and non-speech acoustic events [12]. Also the combined methods are possible, for example, the estimation of the position and orientation of the speaker's head and multimodal recognition of emotions. Such audiovisual monitoring permits to control the events (a new participant enters the meeting room, the presentation begins, and the audience is given the floor) taking place in the meeting room. The proposed image processing includes the removal of low-quality frames, illumination normalization, elimination of image blur, cleaning of digital noise, and automatic segmentation and recognition of the participants' faces. In experimental section, one can find a digital comparison of three basic techniques for face recognition, such as the principal components analysis, linear discriminant analysis, and local binary pattern analysis. The authors developed a method of audiovisual recording of participant activity based on the multifunction system of video monitoring and multichannel system of audio localization that detects the position of the sound source evaluating the phase difference of signals recorded by pairs of microphones from four arrays [13].

Chapter 7 contributes in the audience analysis systems [14] and gender recognition systems [15] using the assessment of face image quality. This is a continuation of investigations of these authors during last two decades. Several standards, such as ISO/IEC 19794-5 and ICAO 9303, contain a description of parameters that provide a decision of the image suitability in the automatic face recognition systems. All parameters are grouped into two classes the textural features (sharpness, contrast and light intensity, compression ratio, and other distortions) and face features (symmetry, position, rotation, eyes visibility, and the presence of glare or shadows on the face). For such algorithms, an assessment of face image quality is a cornerstone because a success of face verification/identification depends strongly from the selected lucky frames. Three types of the quality metrics were considered, such as the texture-based, symmetry-based, and universal metrics. The authors developed the no-reference image quality assessment algorithms based on the analysis of texture and the landmark points' symmetry in the facial images [16]. Note the original methodology of the experiments' statement. Two sets were formed called as Top1 with the single best image of a person and Top3 considering three high quality images chosen by the experts and objective metrics. The proposed algorithms were tested using standard LIVE and TID2013 image database. The performance results show that the proposed algorithms are highly competitive for audience analysis systems and, moreover, have very low computational complexity making it well suited for real-time applications. Additionally, novel gender classification algorithm was suggested based on non-linear SVM classifier. It includes several steps, among which are the scale-invariant feature transform, histogram of oriented gradients, Gabor filters, and pre-selection of blocks with the corresponding positive experimental results.

Chapter 8 contains a description of real time eye blink detection method for Android device controlling and its implementation using standard libraries like OpenCV conversion procedures. The Human Computer Interaction (HCI) or Mobile Computer Interaction (MCI) may be realized in different ways. The authors chose the issue of the eye blinking detection that has a significant role in the human mobile interaction, computer interaction, driving safety, and health care. The heuristic propositions concluded from the theory and practice of the HCI were used during design. The eye tracking is the fastest non-invasive method of measuring user attention [17]. The best definition of eye tracking can be the estimation of user gaze's direction, which is very difficult problem especially in the real world. The proposed method consists of five main parts including the mobile camera processing, face detection, eye detection, eye tracking, and blink detection. At the initial step, a video captured from the front camera of any Android device is converted to grayscale frame using OpenCV, and then stored in special mobile folder to be used later for face detection step. For face detection, Haar classifiers are used, while the eyes detection is based on the trained AdaBoost and Haar feature algorithms. The corneal reflection and pupil center of the eyes have been used as the most important eye features in the proposed method in order to track the movement of the eyes. During eye blink detection step, each black color pixel in the grayscale frame represented by 1, while each gray or white color pixel will be represented by 0. The median blur filtering is applied in this step. The mobile activity is controlled in case the eye blinks by several ways, such as the sending a text message, turning on the alarm system, opening a web browser or making a phone call. The experiments with different distances between a human face and phone and various lighting conditions (indoor/outdoor) were conducted. The overall and detection accuracy had reached 98%. For each frame, the average execution time was 12.30 ms that provides a real time execution.

Chapter 9 contains the rich experimental study of the medical image processing and morphological analysis in urology and plastic surgery (hernioplasty). At present, it is necessary not only to identify the localization of the calculus but also to determine the density and configuration of the calculus and evaluate the functional state of the urinary tract above and below the obstruction. For this purpose, the methodology based on a novel method for color coding of contour representation obtained by the digital shearlet transform was developed. Medical images contain the noise of different nature. In order to improve medical images that can have a high resolution the optimized in implementation algorithms of the most frequently used filters, such as the mean filter, Gaussian filter, median filter, and 2D Cleaner filter, were developed. The study of properties of filters' functioning permitted to use all three type of parallelism, viz. the data, algorithmic, and functional parallelism. A comparison of the optimized and ordinary implementations of noise reduction filters shows great speed improvement of the optimized implementations (around 3–20 times). The highest increase in the processing speed was achieved for the median filter. Thus, for the small kernel $3 \times 3$ the acceleration was about 8 times and for the large kernel $11 \times 11$ is around 70 times. For the parallel implementation, the OpenMP standard was used. The chapter includes the pseudo-code of some program procedures. For contour representation, the simple conventional methods, such as Roberts, Prewitt, Sobel, and more complex ones (LoG and Canny methods) were tested. However, the main attention was paid to use digital shearlet transform in order to obtain the best results in contour extraction [18]. This approach was enforced by the novel color coding algorithm. It is based on the color selection and density distribution of the isolines in an image corresponds to the known technique of building elastic maps using the spatial data. As a result, the image accuracy of estimates in urology and plastic surgery (hernioplasty) was increased up 10–25% in averaged. In urology, the proposed color coding method increased accuracy, especially in complex cases of multiple stones in the kidney. In hernioplasty, the color coding allows to conduct more efficient analysis of the tissue regeneration by controlling the variability of texture with improved accuracy.

Chapter 10 investigates a study of medical image processing and analysis in clinical decision support systems regarding the cervix oncological changes diagnostics [19, 20]. The novelty of the proposed approach is based on the combined two known approaches that analyze the images obtained in white light illumination and fluorescent images separately. Chapter provides the detailed description of that field and rationality of three classes to define the diagnosis, such as Norm, Chronic Nonspecific Inflammation (CNI), and Cervical Intraepithelial Neoplasia in various types of oncological changes (CIN I, CIN II, CIN III) as a single class. First, the ordinary procedures to improve a quality of the images (noise reduction, contrast enhancement, etc.) are executed. Second, the special medical imaging procedures (the matching medical images taken under different lighting conditions, automatic segmentation of regions of interest, and removal of highlights in the images) are developed. It is worth noting that many algorithms were developed, for example shift compensation algorithm for the multispectral images and their analysis in a combination with the images obtained in white light based on a phase correlation in Fourier domain, segmentation of region of interests using Gauss filtering and morphological operations of dilatation and erosion. Practical recommendations for different image processing are included in this chapter. For classification task, some famous strategies were tested in real images provided by South Korea clinics. These investigations show that the best specificity, sensitivity, and accuracy are obtained using the random decision forest strategy. This result is connected with the high degree of image variability obtained from different patients with differences not only due to pathology but also due to differences in the age, menopause, and other features of woman physical condition. The final results in the sensitivity and specificity exceed the results of inexperienced physicians that demonstrates a possibility of practical application of the pathology maps for colposcopist examination.

Chapter 11 involves some results in Foot Progression Angle (FPA) detection, which is an important measurement in clinical gait analysis [21]. The proposed Visual Feature Matching (VFM) model is a solution for long-term real time FPA monitoring in home or community like environments for the patients with movement disorders or abnormal gait. The FPA is calculated as the angle between the foot vector, line joining the heel point center and the second metatarsal head [22]. Thus, the efforts of the authors were directed to the accurate estimation of foot orientation. Image calibration and rectification are the tools used to eliminate two types of distortion, such as the lens distortion and perspective distortion caused by the optical lens and position of the camera relative to the subject. The classic approach for foot feature extraction and matching includes the algorithm for feature extraction, for example, using the SURF descriptor and matching algorithm, in this case, the statistically robust M-estimator sample consensus algorithm. However, the authors proposed an alternative approach, when a pair of paper strips, instead of shoes, is used in the investigations because of different color and textile of participant's shoes. Some estimators for the FPA measurements were obtained.

The authors proposed the original equipment, when the camera is mounted downward on the torso that has a few advantages.

## 1.3 Conclusions

In this book, reader will find many original and innovative algorithms from many fields of computer vision. All chapters involve great experimental material with the corresponding explanation of results that makes the proposed methods more valuable. Culture, robotics, community interactions, and medicine are not the closed list of real life application, where computer vision helps to improve a quantity of a human life. A variety of the presented innovative algorithms is wide: from the graphological analysis of handwritten texts, perceptually tuned image watermarking, mobile robots surveillance, and ambient audiovisual monitoring of human meetings to the medical image processing considering the current achievements in computer vision.

## References

1. Volkov, D.M., Mironovsky, L.A., Reshetnikova, N.N.: Automated system of graphological research of manuscripts. In: XIII International Scientific and Engineering Workshop “Modern Technologies in Problems of Control, Automation and Information processing”, pp. 482–483 (in Russian) (2004)
2. Artemov, I.V., Volkov, D.M., Kondakova, I.A., Nikitina, A.A., Reshetnikova, N.N., Soloviev, N.V.: Information Retrieval System of graphological analysis and identification of handwritten texts. The certificate of registration of the industry development. FSSI “State Coordinating Center for Information Technologies, an industry fund of algorithms and programs.” Moscow, pp. 1–10 (in Russian) (2006)
3. Favorskaya, M.N., Savchina, E.I.: Content preserving watermarking for medical images using shearlet transform and SVD. Int Arch Photogramm Remote Sens Spatial Inf Sci, XLII-2/W4, 101–108 (2017)
4. Favorskaya, M., Oreshkina, E.: Digital gray-scale watermarking based on biometrics. In: Damiani E, Howlett RJ, Jain LC, Gallo L, De Pietro G (eds.) Intelligent Interactive Multimedia Systems and Services, SIST, vol. 40, pp. 203–214 Springer International Publishing, Switzerland (2015)
5. Prozorov, A., Priorov, A.: Three-dimensional reconstruction of a scene with the use of monocular vision. Meas. Tech. **57**(10), 1137–1143 (2015)
6. Pleshkova, Sn., Bekiarski, Al.: Audio visual attention models in the mobile robots navigation. In: Kountchev, R., Nakamatsu, K. (eds.) New Approaches in Intelligent Image Analysis, ISRL, vol. 108 pp. 253–294 Springer International Publishing, Switzerland (2016)
7. Dehkharghani, S.S., Bekiarski, A., Pleshkova, S.: Application of probabilistic methods in mobile robots audio visual motion control combined with laser range finder distance measurements. In: 11th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing (CSECS’2012), pp. 91–98 (2012)

8. Al, Bekiarski: Visual mobile robots perception for motion control. In: Kountchev, R., Nakamatsu, K. (eds.) Advances in Reasoning-Based Image Processing Intelligent Systems, ISRL, vol. 29, pp. 173–209. Springer, Berlin (2012)
9. Venkov, P., Bekiarski, Al., Dehkharghani, S.S., Pleshkova, Sn.: Search and tracking of targets with mobile robot by using audio-visual information. In: International Conference on Automation and Informatics (CAI'2010), Sofia, pp. 463–469 (2010)
10. Ronzhin Al.L., Budkov V.Yu.: Determination and recording of active speaker in meeting room. In: 14th 17th International Conference Speech and Computer (SPECOM'2011), Moscow, Kazan, Russia, pp. 361–366 (2011)
11. Yusupov, R.M., AnL, Ronzhin, Prischepa, M., AlL, Ronzhin: Models and hardware software solutions for automatic control of intelligent hall. Automat. Remote Control **72**(7), 1389–1397 (2011)
12. Ronzhin AlL.: Audiovisual recording system for e-learning applications. In: International Conference on Computer Graphics Theory and Applications (GRAPP'2012), pp. 515–518 (2012)
13. Ronzhin, A., Budkov, V., Karpov, A.: Multichannel system of audio-visual support of remote mobile participant at e-meeting. In: Balandin, S., Dunaytsev, R., Koucheryavy, Y. (eds.) Smart Spaces and Next Generation Wired/Wireless Networking, LNCS, vol. 6294, pp. 62–71 (2010)
14. Khryashchev, V., Ganin, A., Golubev, M., Shmaglit, L.: Audience analysis system on the basis of face detection, tracking and classification techniques. In: International MultiConference of Engineers and Computer Scientists Hong Kong, LNECS, pp. 446–450 (2013)
15. Khryashchev, V., Priorov, A., Shmaglit, L., Golubev, M.: Gender Recognition via Face Area Analysis. World Congress on Engineering and Computer Science, Berkeley, USA, pp. 645–649 (2012)
16. Nenakhov, I., Khryashchev, V., Priorov, A.: No-reference image quality assessment based on local binary patterns. In: 14th IEEE East-West Design & Test Symposium, pp. 529–532 (2016)
17. Anwar, S., Milanova, M., Bigazzi, A., Bocchi, L., Guazzini, A.: Real time intention recognition. In: 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON'2016). (2016). doi:10.1109/IECON.2016.7794016
18. Cadena, L., Espinosa, N., Cadena, F., Kirillova, S., Barkova, D., Zotin, A.: Processing medical images by new several mathematics shearlet transform. In: International MultiConference of Engineers and Computer Scientists (IMECS'2016), vol. I, pp. 369–371 (2016)
19. Muhuri, S., Bhattacharjee, M. (2014). Automated identification and analysis of cervical cancer. In: 3rd World Conference on Applied Science, Engineering and Technology, pp. 516–520
20. Liang, M., Zheng, G., Huang, X., Milledge, G., Tokuta, A.: Identification of abnormal cervical regions from colposcopy image sequences. In: 21st International Conference on Computer Graphics, Visualization and Computer Vision (WSCG'2013), pp. 130–136 (2013)
21. Hinman, R.S., Hunt, M.A., Simic, M., Bennell, K.L.: Exercise, gait retraining, footwear and insoles for knee osteoarthritis. Curr. Phys. Med. Rehabil. Rep. **1**, 21–28 (2013)
22. Simic, M., Wrigley, T., Hinman, R.S., Hunt, M., Bennell, K.: Altering foot progression angle in people with medial knee osteoarthritis: the effects of varying toe-in and toe-out angles are mediated by pain and malalignment. Osteoarth. Cartil. **21**(9), 1272–1280 (2013)

# Chapter 2
# Graphological Analysis and Identification of Handwritten Texts

**Leonid A. Mironovsky, Alexander V. Nikitin, Nina N. Reshetnikova and Nikolay V. Soloviev**

**Abstract** The problem of recognition of handwriting text is still far from its final solution. The existing systems of recognition of handwritten texts are usually developed for some special applications. The difficulties are caused by recognition of the conjoint writing because a variability of handwritings is the highest and often it is necessary to solve the problem of delimitation of the separate letters. In this chapter, along with to the known methods of the handwritten fragments' analysis, it is offered to use the developed methods of vectorization of raster images and vector dynamic parameterization. Also, a description of the automated information storage and retrieval system for the graphological analysis and identification of unintelligible fragments of handwritten texts is given. The system contains a database of handwriting samples with variants of the author's calligraphy from the Manuscript Department of the Institute of Russian Literature (Pushkin's House) of the Russian Academy of Sciences.

**Keywords** Graphological analysis · Handwritten text · Text segmentation
Comparison of words · Symbols and ligatures · Dynamic parameterization
Drafts autographs · Automated information retrieval system

L.A. Mironovsky (✉) · A.V. Nikitin · N.N. Reshetnikova · N.V. Soloviev
St. Petersburg State University of Aerospace Instrumentation,
67 Bol. Morskaya St, Saint Petersburg 190000, Russian Federation
e-mail: miron@aanet.ru

A.V. Nikitin
e-mail: nike51@mail.ru

N.N. Reshetnikova
e-mail: reni_07@list.ru

N.V. Soloviev
e-mail: famsol@yandex.ru

 
M.N. Favorskaya and L.C. Jain (eds.), *Computer Vision in Control Systems-4*,
Intelligent Systems Reference Library 136,
https://doi.org/10.1007/978-3-319-67994-5_2

## 2.1 Introduction

The first attempts to solve the problem of handwritten or printed text recognition in bitmap images were made more than fifty years ago, almost immediately after appearance of the devices that could upload images into computer memory [1, 2]. In this context, the text recognition means an automated process for obtaining of ASCII codes of symbols (letters, numbers, and punctuation marks). Selection and partial recognition of image segments in the photos, diagrams, plots, mathematical formulas, and tables are also possible.

Today there are a lot of software products that successfully recognize the printed characters [3, 4]. For bitmap images with high resolution (300–600 dpi), contrast, and sharpness, the number of recognition mistakes does not exceed 0.5% [5]. Also, there are a lot of software products for processing of images to be recognized in order to increase their contrast and sharpness with removal possible noise and defects [6]. In this case, the successful recognition can be implemented easily thanks to the fact that a bitmap image of printed text could be easily segmented into images of separate characters and the images of similar characters are nearly identical.

The process of recognition of printed texts is commonly referred to as Optical Character Recognition (OCR) [7] even though the OCR means recognition of any text, printed or handwritten [8]. It is notable that apart from the OCR software, which was designed for image processing using the desktop image scanners, some devices were designed specifically for recognition of the printed texts [9]. The problem of recognition of the mixed printed and handwritten texts is also solved quite successfully [4]. The main difference of such text from printed ones lies in the fact that images of similar characters can significantly differ from each other. To solve this problem, some text recognition software tools include the learning algorithms. In this case, a software tool attempts to learn the specific handwriting style from an image, which significantly increases the chance for successful recognition [10].

The most difficult task is the recognition of joined-up handwriting. This process is known as Hand Writing Recognition (HWR) [11]. The main differences of handwritten texts from the printed ones are described below. First, the lines of handwritten texts are often not parallel, which is especially common for texts written on an unlined sheet. There also can be partial overlapping of the adjacent lines and words/letters in these lines. Second, even though the words in handwritten texts are separated with spaces, some words are not written jointly, which means that some letters or even letter fragments can be separated with spaces. Finally, even one person’s handwriting depends on many factors. For example, a readability of a word in the official document can differ a lot from the same word written in a personal note book. Moreover, a way of the letter writing greatly depends on the writing of adjacent letters. The first two of the mentioned above reasons complicate the segmentation of images with handwritten texts into separate words, while the third one complicates the recognition of letters.

The chapter is organized as follows. The related works are reviewed in Sect. 2.2. Section 2.3 provides the analysis and identification of handwritten texts including a problem formulation. A formation of alphabets based on the handwriting samples is considered in Sect. 2.4. The methods of bitmap binary image based on the graphological analysis and identification are represented in Sect. 2.5, while the methods of the graphological analysis and identification based on vectorization of bitmap images are developed in Sect. 2.6. Section 2.7 includes a description of architecture of the designed software tool. Section 2.8 concluded the chapter.

## 2.2 Related Works

The interest to the problem of recognition of the joined-up handwritten texts is unabated, which is evident from the great amount of publications on this topic [12–33]. There are two main approaches that support two different ways to obtain the handwritten texts called as “on-line” recognition (when the characters are written with a pen in a special screen, for example a tablet PC screen) and “off-line” recognition (when a document with a handwritten text already exists). Generally, the on-line recognition is easier task because the problem of text segmentation into words does not appear due to the on-line mode of capturing a lot of clues that can be successfully used to learn the users’ handwritings. The existing on-line hardware and software recognition products are effectively demonstrate this proposition [34].

The off-line text recognition is a more complicated task. As it is listed in many reviews [13, 35–37], the classic methods for selection of the handwriting features, creation of a database with samples, and features identification do not work successfully in some cases [38]. It is known [39] that the artificial neural networks are able to identify the generalized features of recognizable images. In this regard, the application of artificial neural networks to the off-line recognition of handwritten texts seems to be a good idea [40–52]. The reasons for mistakes lie in the facts that the same characters can be written in different ways by different people and it is very hard to identify the separate characters in the handwritten texts.

While recognizing a handwritten text, people consider its context, which means that they take into account the information they obtained from the parts of the same text, which they read before. It is known [53] that a human being can easily read a word if it is a part of a discourse (normal text) but the same person will have problems while indentifying the same word in a text, consisting of random words. It is even more difficult for a human being to identify a separate letter even though that same letter in a text can be read easily. Presumably, this is the recognition systems’ ability to understand the meaning of handwritten texts that will allow for a breakthrough in the science but the existing intelligent systems are still far from solving this task.

Sometimes a person, who reads a handwritten text, faces a problem of recognizing a separate word that can be one of several possible similar-looking words, each of which does not violate the rules of grammar and retains the meaning of the text. For example, this problem can be faced by a literary scholar, who works with a handwritten archive of a writer [54]. Such person visually compares an unidentified word fragment with similar-looking fragments of words that are already recognized in order to find the right one. The system described below in this chapter was designed to speed up this process and increase the correctness of recognition results.

It is obvious that the problem of handwritten text recognition is far from being completely solved. The existing handwritten text recognition systems are usually developed for some special cases or applications based on the specific features of the texts to be recognized. These specific features are the quality of handwritten texts and size of vocabulary. The main quality groups of handwritten texts are texts written with the block letters and those written with the cursive handwriting. The cursive handwriting, in its turn, falls into the separate letters handwriting and joined-up handwriting.

The most difficult problem is the recognition of joined-up handwriting because in this case a variability of handwriting samples for different letters is very high and the task of identifying the borders of each letter becomes very complicated. The volume of vocabulary, i.e. a number of different words in the text, is also very important. Vocabularies of authors' handwriting styles can be of great value, when literary texts are recognized. Such vocabularies are not comprised only of a large number of words used in a literary text but can also include sets of the author's calligraphy variants (ways of writing of different letters and links between them—ligatures) in order to define the correctness of this or that word. The use of computers to recognize the illegible handwritten texts can significantly speed up the process of finding the letters, ligatures, and words and increases an objectivity of comparison of the reference fragments of handwritten texts (taking into account the variability of handwriting samples) with ones under recognition.

## 2.3 Analysis and Identification of Handwritten Texts: Problem Formulation

Reading the handwritten drafts is often a very difficult task. Let us take draft manuscripts written by A.S. Pushkin and A.S. Griboedov as an example. These manuscripts are kept at the Manuscript Department of the Institute of Russian Literature (Pushkin's House) of the Russian Academy of Sciences. Just looking at the reproduction of Pushkin's draft manuscripts in the Big Academic Collection of his works (16 volumes printed in 1937–1949), one can see a huge number of notes and <illegible text> signs, that bespeak of the uncertainty of editors' interpretation of the author's handwriting [54].

There are longstanding academic disputes about some authors' handwritten drafts. Moreover, the samples of writing of some letters can greatly vary depending on the speed of writing and the writing tool used (goose quill, pen point, or pencil). A textual analyst has to create a vocabulary of various samples of the author's calligraphy (ways of writing separate letters and links between them—ligatures) in order to read correctly this or that word. Figure 2.1 shows examples of A.S. Pushkin's handwritten drafts of the "Poltava" poem (courtesy provided by the Manuscript Department of the Institute of Russian Literature of the Russian Academy of Sciences).

Usually, while the analyzed fragment of handwritten text cannot be interpreted unambiguously, a textual analyst performs the following steps:

Step 1. Selection a possible combination of letters in accordance with the grammar rules, semantic content of the text, and known individual specifics of author's writing.
Step 2. Searching the author's fair copies to pick text fragments that matches the selected combination of letters.
Step 3. Finding a fragment that matches the selected example from the analyzed handwritten text.
Step 4. Visually comparison the sample with the recognized/analyzed fragment of handwritten text.
Step 5. If the fragments do not match, repeats Steps 2–4 or maybe even Steps 1–4.

The use of advanced computer means and technologies can significantly speed up the process of finding samples (Steps 1 and 2), facilitate in objective integral assessment of a degree of closeness of compared fragments, and reduce the subjective judgment due to a human factor.

The automatic analysis and identification of illegible fragments of handwritten text requires to solve a number of problems [55–60]:

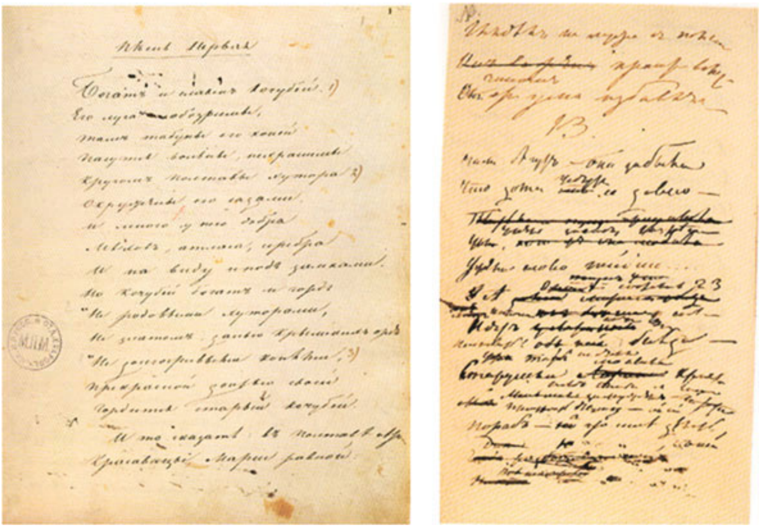

**Fig. 2.1** Examples of A.S. Pushkin's handwriting

- Identification of the type of representation of the script fragment using the bitmap binary image with the suppressed brightness distortions and vector representation of an image.
- Analysis and selection of acceptable geometric transformations of an image in order to achieve the best possible matching.
- Development of methods for assessment of a degree of closeness of images during identification.
- Development of architecture for a computer system to store and retrieve information needed for the handwriting analysis and identification of illegible fragments of handwritten text.
- Creation of a database of alphabets based on samples of the author's handwriting taking into account the variability of author's calligraphy and difference of samples of the same letters found in the different author's texts.
- Creation of a bank of efficient methods for the recognition, handwriting analysis, and identification of handwritten texts.
- Design of a system of queries to work with the bank of methods and database of alphabets based on the author's handwriting samples and ensures access to them using prospective technologies of augmented reality through the intuitive interface.

Consider some of the mentioned above tasks in details. We will illustrate studies and experiments with examples of fair copies and drafts of handwritten scripts written by A.S. Pushkin, the classic of Russian literature, whose works always have a big number of non-text fragments and noises.

## 2.4 Formation of Alphabets Based on Handwriting Samples

Consider a formation of alphabets using samples of the author's handwritings. This is one of the most important and time-consuming tasks that requires direct involvement of specialists in textual analysis. The main procedures that need to be implemented to form an author's handwriting alphabet are the following:

- Input of handwritten text. Samples of handwritten text are represented on a computer screen as black-and-white or color bitmap images.
- Pre-processing of images. Elimination of image defects and noises on the background, and selection of text zones for analysis.
- Text segmentation that includes a breaking the text into the separate lines, lines into words, and words into letters and ligatures. Segmentation should be carried out more than once in cases of controversies in the interpretation of analysis results, since the intervals between letters in handwritten texts are often wider than those between words, whereas words can be connected with each other.

- Saving the letters, ligatures, and words as the author's handwriting samples in a database.

Selection of handwritten text fragments can be made using either widely known computer software (bitmap graphics editors like Adobe Photoshop or GIMP as an open source software) or specialized applications with a range of functions for textual analysis (horizontal line adjustment, selection of different combinations of segments, etc.) with the intuitive interface.

To address the task of alphabet formation, a technique for pre-processing of images of handwritten text fragments with elimination of image defects and background noises was developed and tested. The most commonly used tools for input of static visual information are the digital cameras and scanners (hand and flatbed).

Consider a still image on a flat media—rectangle sheet of paper (Figs. 2.1, 2.2, and 2.5). Our input device is a flatbed scanner that allows to obtain both color and grayscale images of various resolutions. The comparative analysis of the color, grayscale, and binary images of handwritten text showed that the color images do not have noticeable advantages over grayscale ones in terms of use for text recognition [57, 58]. Based on the conducted experiments, it was found that the most suitable resolution for analysis of images of handwritten texts is 200–300 pixels per inch (dpi). Further increase of resolution does not lead to noticeable increase in an image quality, while its decrease impairs image representation on computer screens and sometimes leads to the loss of small text elements.

Preliminary image processing, as a rule, includes the suppression of noises and image binarization. The latter operation is needed for segmentation, i.e. selection of

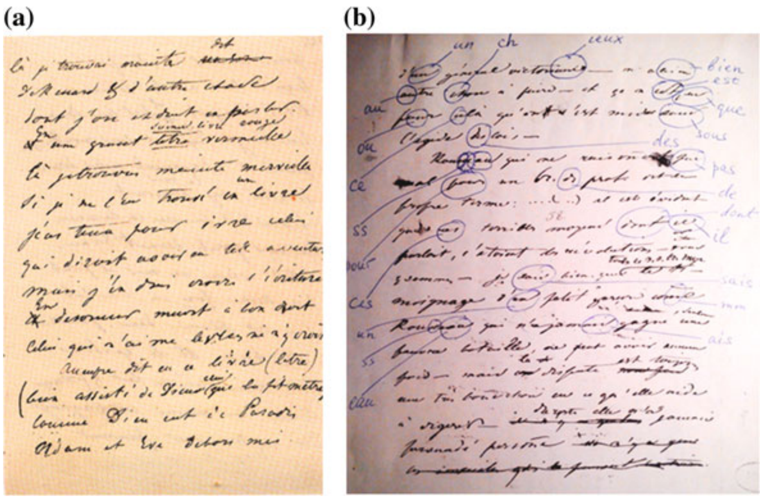

**Fig. 2.2** Parts of "Roman de Renard", a French literary classic of the 12–13th centuries: **a** original image, **b** marked image