Chapter 1
Development of Mathematical Theory in Computer Vision

Margarita N. Favorskaya and Lakhmi C. Jain

Abstract This chapter presents a brief description of chapters devoted to the theoretical development of computer vision. Original investigations in mathematical morphology, estimations of structural changes, the hierarchical adaptive Karhunen-Loeve and projective transforms, among others, provide the great contribution in mathematical foundations of computer vision. Each theoretical chapter involves practical implementations, which demonstrate the merit of proposed methods in practice.

Keywords Computer vision · Image processing · Videos processing

1.1 Introduction

In the past decades, computer vision techniques have progressed significantly and are widely used in many implementations of control systems. Great advances have been made in image filtering, segmentation, pattern recognition, and events understanding. However, the excellent mathematical models and methods cannot be directly applied in many practical situations. The majority of efforts focus on designing the efficient and real-time methods to analyze images and video data on various levels of processing. The contemporary solutions based on advanced mathematical achievements emphasize on more information and visual monitoring

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

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 and L.C. Jain (eds.), *Computer Vision in Control Systems-1*,
Intelligent Systems Reference Library 73, DOI 10.1007/978-3-319-10653-3_1

in natural and human environment. The goal of current investigations is designing such observation models, which are close to realistic visualization and interpretation of events in our world.

1.2 Chapters Included in the Book

The main purpose of this research book is to present a sample of research results on recent advances in computer vision. This book includes eleven chapters on the "Mathematical Theory" aspect of the computer vision.

Chapter 2 introduces the morphological framework as a very wide theoretical platform for creation of mid-level image analysis tools for specialized computer vision applications. It utilizes the structural image modeling and decides some image filtering, segmentation, and comparison problems. Mathematical Morphology (MM) by Serra [1] and Matheron [2] is still the most well-known version of MM until these days. Another morphological approach proposed by Pyt'ev is based on geometrical and algebraic reasoning. In the framework of Pyt'ev morphology, images are considered as piecewise-constant 2D functions [3]. The tessellation of image frame by a set of non-intersected connected regions with constant intensities determines the "shape" of the image. The main idea of this approach is the projection of one image onto the shape of other image. The detection of morphological changing is performed by comparison of image and its projection to the reference image. Such morphological tools are invariant relative to image intensity transforms and stable relative to noise. The idempotent operators such as morphological filters or projectors are introduced using a concept of figure filling by structured elements. In other version, morphological filters are based on merging of grayscale image connected regions ("flat" zones). The continuous binary morphology is based on computational geometry and provides very fast tool for computation of continuous figure skeletons using approximation of 2D binary image by region border polygons and calculation of Voronoi diagram for segments these polygons [4]. A skeletal representation of the figure is formed as its skeleton and the radial function determined in skeleton points. The projective morphology is a generalized framework based on Serra's MM and Pyt'ev's Morphological Analysis. It combines the ideas of both morphological approaches and allows construction of some new morphological systems and operators based on different image decompositions and transforms and/or criterions (energy functions). Criterion-based projective morphological filters are implemented using numeric optimization techniques (linear programming, dynamic programming, graph cutting, and so on) [5]. The morphological spectrum as a multi-scale morphological shape analysis tool based on "granulometry" also contains in this chapter.

Chapter 3 discusses the criteria of Mean Structural Similarity Index Measure (MSSIM) and the developed Mean Nonparametric Structural Similarity Index Measure (MNSSIM), as well as the spectral algorithm for detecting structural changes in a frame, which have been used to good effect in video codec analysis [6].

These criteria provide the estimations for structural (texture) variations of images. The growing popularity of these criteria is proved by their quite appropriate compliance with the human vision system [7]. The detection of variations in the image segment structure is based on spectral and correlation analysis of space-time fields. At present, the quasi-optimum heuristic algorithms applying variations of field correlation features, non-invariance of spectrum in various bases (in relation to a segment movement and change of their texture features) exist. The different estimation methods and algorithms for images presented by numbered blocks as well as the criteria and metrics being a basis to detect these differences are investigated. In this chapter, the reader can find practical examples using of pixel and spectral algorithms in image analysis.

Chapter 4 investigates a novel approach to process a single image or sequences of frames through the Hierarchical Adaptive Karhunen-Loeve (KL)-based Transform (HA-KLT). This approach is suitable for image block coding and for inter-frame processing of correlated frames in groups [8]. The basic aim of a new transform is to achieve a decorrelation of the image blocks, respectively of all frames in the processed group. This is realized by a multiple applying of the HA-KLT. After each level of the hierarchical transform, all sub-blocks (respectively groups) are rearranged so, that the components with highest correlation, which are obtained in the preceding level, would be placed in a new sub-blocks of the current level. The kernel of the multi-level transform is the Adaptive KL Transform (AKLT). The AKLT with a transform matrix of size 2×2 and 3×3 is used for the processing of the image sub-blocks and the pixels with same position in the sub-groups of frames respectively. The algebraic method for the calculation of the elements of the AKLT transform matrix of size 2×2 and 3×3 is presented in this chapter. The 2D and 3D HA-KLT algorithms for the blocks of a single image and for inter-frame processing of sequences (groups) of frames are also developed [9]. The computational complexity of these algorithms is compared with the "classic" KLT. On the one hand, the proposed approach ensures a higher accuracy of color segmentation in all cases, when a distribution of color vectors is not Gaussian. This is achieved by using a polynomial kernel for the color space expansion, after which the HA-KLT is applied to the expanded color vectors. In result, a decorrelation of the transformed vectors and an information concentration in their first components are achieved. On the other hand, this permits to reduce a number of components of the transformed vectors, retaining the first two only. In a new 2D space, the color vectors clusterization in respect to RGB space is enhanced, and they can be classified with high accuracy by using the support vector machine algorithm or other similar methods [10]. The HA-KLT method is a basis for the creation of novel efficient algorithms for a fusion of 3D images in face recognition task, an objects tracing in videos, a compression with movement compensation and without visual quality loss of TV and multi-view visual information, medical and multispectral images, etc [11].

Chapter 5 provides the design of object-invariant cores, which correspond to all types of spatially compact object images (previously segmented from a background), under the affine and projective transformations caused by an image projection through the spherical (or almost spherical) lenses being the traditional parts of photo- and video-cameras [12]. The object-invariant core is synthesized by means of truncating the high-frequency harmonics in a spatial image spectrum. These rejected high-frequency harmonics present the object peculiarities, while the rest (extremely low-frequency) harmonics contain the information about spatial image transformations. It is shown that such object-independent core is mathematically described by elliptic paraboloid (quadratic parabola in 1D image projection). All parameters of affine geometric transformation (except a rotation and a mirror-like reflection) are measured analytically from this object invariant core. The parameters of rotation and mirror-like reflection are calculated from the cyclic narrow-band harmonic cores of image projection on the angular coordinate in a polar system. While the 6-parametric affine transformation is entirely linear, the full projective transformation contains additionally a nonlinear part described by two additional parameters. Due to this nonlinearity, the specific parameters of projective transformation cannot be measured analytically. A novel iterative optimization procedure is proposed to measure all parameters of projective transformation [13]. It is proposed to measure the missing parameters of projective transformation by a displacement of object-invariant core under the test transformations. The convergence of iterative measurement procedure is rigorously proven. At the end of the chapter, the examples of practical applications for automatic measurement of all projective transformation parameters are presented.

Chapter 6 presents a way of energy analysis for image and video sequence processing as a preliminary processing in vision systems [14]. Usually the object movements are determined by the analysis of an Inter-Frame Difference (IFD) in video signals. It is the simplest universal method. However, it doesn't exhaust opportunities for intelligent processing, especially in extremely low luminance. The IFD of energy spectrums and phase-energy spectrums are considered as an alternative analysis. The phase-energy spectrum is a product of partial derivatives in spatial phase-frequency spectrum over their spatial frequencies. It provides the detailed information about motion in finite frames [15]. The modeling of the IFD of frequency responses shows the necessity of analysis for pixels located near the moving boundaries. A processing of such pixels intensities increases a probability of movement's detection. Also distortions of moving object's shape, movement's characteristics, and a quantity of moving objects are possible to define based on the analysis of the IFD types. The phase-energy spectrums are used for edges analysis, if any movement is detected in a scene. The analysis of the energy spectrums is applied to design the effective 2D filters. The changes of the energetic indexes in static images determine the efficiency function on a whole set of impulse responses of the filter. The function of efficiency has a positively certain quadratic form with the coefficients of energy spectrum decomposition into the 2D Fourier series over the cosines. The analysis of stationary points by using this function of efficiency allows to synthesize the optimum and the quasi-optimum 2D filters. The proposed way of energy analysis provides some novel possibilities, for example, the detection of objects with extremely small contrast image.

Chapter 7 studies on how visual motion can be estimated at the lowest overall uncertainty of measurement across the entire range of useful sensor sizes (in artificial systems) [16] or the entire range of receptive fields (in biological systems). In other words, the following is an attempt to develop an economic normative theory of motion-sensitive systems. Such norms are derived for efficient design of systems, and then the norms are compared with facts of biological vision. This approach from the first principles of measurement and parsimony helps to understand the forces that shape the characteristics of biological vision. These characteristics include the spatiotemporal contrast sensitivity function, the adaptive transformations of this function caused by stimulus change, and also some characteristics of the higher-level perceptual processes such as perceptual organization [17]. In the following, the minimax strategy is implemented by assuming the maximal (worst-case) uncertainty of measurement on the sensors that span the entire range of the useful spatial and temporal scales. This strategy is used in two ways. First, the consequences of Gabor's uncertainty relation are investigated by assuming that the uncertainty of measurement is as high as possible. Second, the outcomes of measurement on different sensors are anticipated by adding their component uncertainties.

Chapter 8 presents the segmentation of natural images as a challenging task in image processing. Many methods have been proposed in the literature regarding algorithms for segmentation of such images [18]. Many of algorithms are complex in nature and inefficient in practice with unaltered images. In order to efficiently use the algorithms it is beneficial to pre-process the natural images. However, natural images often involve subjects and background that are not easily quantified with crisp pre-processing parameters. A partial solution to the problem of segmenting complex images is to use features that discriminate in the active contour algorithms [19]. These feature descriptions range from curvature to the orientation of level sets and usually result in better segmentation. An unfortunate side effect of using feature discriminates is that the complexity of the algorithm greatly increases resulting in even higher computational cost and difficulty in implementing the method. The goal is to develop a morphological level set active contour segmentation method that can robustly and efficiently segment multiphase textural images of high complexity [20]. To do this the usage of region statistics inside and outside the contour, membership functions from fuzzy logic methodology, and a Gaussian kernel function are required. In this chapter, a number of existing methods for shape feature extraction and representation are presented. At the end, application examples for using object shape representation in application for object recognition and human activity recognition are show.

Chapter 9 is devoted to digital video stabilization oriented on removal of intentional motions from video sequences caused by camera vibrations under strong wind in static scenes, by motion of robots unstable platforms in dynamic scenes, or jitters during a human hand-held shooting [21]. The analysis of dynamic scenes is required in advanced intelligent methods and directly depends from a problem statement. Several sequential stages connect with the choice of anchor frame, local and global motion estimations, and the jitters compensation algorithm. The choice of anchor frame into static scenes may be random with duration 1 s or 24 frames. In the case of dynamic scene, the additional problem of scenes' separation should be solved for receiving a 'good' anchor frame. Most existing methods and algorithms do not work in real time. For investigation purposes, a non-real time approach is developed, however practical applications need in fast and reliable solutions. Several strategies are used for Local Motion Vectors (LMVs) building based on the keypoint detectors and block-matching algorithm [22]. The application of fuzzy logic operators improves the separation results between the unwanted motion and the real motion of rigid objects. For dynamic scenes, the kurtosis estimations are calculated and tracking curves are built in the case of small vibrations, and frame interpolation is applied, if vibrations have large values. The fuzzy model based on triangular, trapezoidal, and *S*-shape memberships partitions the LMVs concerning them to an unwanted camera motion and objects motion into a scene. The output of fuzzy logic model indicates a final reliability of matching quality by using the Takagi-Sugeno-Kang model. Such zero-order fuzzy model generates the quality index (a value in the range [0, 1]). The quality of the points matching is classified into four categories: excellent, good, medium, and bad. Therefore, fuzzy logic is used for improvement of local and global motion estimations and determines the novelty of approach. The similar procedure is applied for estimation of Global Motion Vectors (GMVs). The corrective algorithm compensates the unwanted motion into frames. Thereby, the scene is aligned. For restoration of current frame, pixels are shifted on a value of Accumulated Motion Vector (AMV) of unwanted motion. However, the sizes of stabilized frames became less relatively the original video sequence and the restoration of "missing" frame edges is required.

Chapter 10 examines the problems of transforming information and studying data connected with processing and transmitting images. The strip-method for storage and noise-immune transmission of images is studied [23]. Before transmission, the matrix transformations of an original image are executed, during which the image fragments are mixed and superimposed on each other. The transformed image is transmitted over a communication channel, where it is distorted with a pulse noise, the latter being for example a possible reason for a complete loss of separate image fragments. In the process of receiving a signal at the receiving end, an inverse transformation is performed. At the end of this transformation, the reconstruction of the image takes place. If it is possible to provide a uniform distribution of the pulse noise over the whole area the image occupies (without any changes of its energy), then a noticeable decrease of noise amplitude will take place and an acceptable quality of all fragments of the image reconstructed. In this chapter, many tasks are considered such as versions of the two-sided strip-transformation of images, choice of optimal transformation matrices, investigation of root images of the strip-transformation, and illustration of capabilities of the method suggested using particular examples. In order to get the maximum decrease of the pulse noise amplitude, it is necessary to achieve a uniform distribution of the noise over the image by applying the inverse transformation at the receiving end of the communication channel. This will allow information about distorted or "lost" fragments to be reconstructed. Now a problem of determining the type of the transformation matrices A and B arises. The solution of this problem will provide the possibility to minimize the noise amplitude in the reconstructed image. A well-known solution of this problem is related to the cases of n, which can be divided by four, i.e. the so-called normalized Hadamard matrices [24]. The less-known solution for even n, not divisible by four, consists in so-called C-matrices (Conference-matrices). Such matrices have a zero diagonal and their remaining elements are equal to ± 1.

Chapter 11 provides a discussion about the generalized criterion of efficiency for telecommunication systems [25]. Besides the partial criteria, there exists also a need in developing generalized ones allowing to compare various telecommunication systems and to choose the most efficient ones among them. To this end, the generalized criteria should consider and incorporate the partial ones, establish certain relationships between them, and hence possess the highest possible objectivity. Such criteria should be rather simple, easily computable, and provide the way to compare the telecommunication systems within a definite numerical scale, that is, they should be normalized [26]. The chapter develops a generalized criterion to estimate the efficiency of telecommunication systems that can be applied to economics information systems, too. The criterion combines evaluation of such special properties as the information quantity, noise immunity, the data transmission speed, and the transmission cost. In contrast to other criteria, the proposed one is non-dimensional and normalized, thus estimating a telecommunication system by means of real number between 0 and 1. The design of the developed criterion based upon the concept of conditional entropy is rather simple. It allows one to calculate the system's characteristic value with sufficient accuracy for practice, thus comparing various telecommunication systems to transfer the economic information. The generalized criterion is composed as a product of some partial criteria, which permits one to estimate the telecommunication systems not only as a whole, but also with respect to their partial characteristics, such as their productivity, reliability, and transmission cost.

1.3 Conclusion

The chapter has provided a briefly description of ten chapters with original mathematical investigations in computer vision techniques applied in advanced control systems. All included chapters involve the recent achievements in mathematical morphological theory, advanced criteria for structural similarity and the efficiency for telecommunication systems, the analysis of energy spectrums, complicated image transforms such as hierarchical adaptive Karhunen-Loeve transform and projective transform, optimal measurement of visual motion based on perception theory, intelligent methods for digital video stabilization, approaches for transmitting images based on Hadamard matrices Each chapter of the book explores experimental results, illustrating its use and applicability.

References

1. Serra J (1988) Image analysis and mathematical morphology. Theoretical advances. Academic Press, London
2. Matheron G (1975) Random sets and integral geometry. Wiley, New York
3. Pyt'ev Y (1998) Methods for morphological analysis of color images. Pattern Recognit Image Anal 8(4):517–531
4. Mestetskiy L (2010) Skeleton representation based on compound Bezier curves. In: 5th international conference on computer vision theory and applications (VISAPP'2010), vol 1. INSTICC Press, pp 44–51
5. Vizilter YV (2009) Design of data segmentation and data compression operators based on projective morphological decompositions. J Comput Syst Sci Int 48(3):415–429
6. Sheikh HR, Bovik FC (2011) Image information and visual quality. IEEE Trans Image Process 20(1):88–98
7. Pratt WK (2001) Digital image processing. PIKS inside, 3rd edn. Wiley, New York
8. Dony R (2001) Karhunen-Loeve transform. In: Rao K, Yip P (eds) The transform and data compression handbook. CRC Press, Boca Raton, London, New York, Washington, DC
9. Kountchev R, Kountcheva R (2013) Decorrelation of multispectral images, based on hierarchical adaptive PCA. Int J WSEAS Trans Sign Proc 3(9):120–137
10. Ivanov P, Kountchev R (2013) Hierarchical principal component analysis-based transformation of multispectral images. Int J Reasoning-Based Intell Syst (IJRIS) 5(4):260–273
11. Rao K, Kim D, Hwang J (2010) Fast Fourier transforms: algorithms and applications. Springer, Dordrecht, Heidelberg, London, New York
12. Shapiro LG, Stockman GC (2001) Computer vision. Prentice Hall, Upper Saddle River, New Jersey
13. Lutsiv V (2009) Method of iteratively compensating projective image distortions. J Opt Technol 76(7):417–422
14. Jähne B (2005) Digital image processing, 6th edn. Springer, Berlin
15. Xiao Z, Hou Z (2004) Phase based feature detector consistent with human visual system characteristics. Pattern Recogn Lett 25(10):1115–1121
16. Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. W. H. Freeman, San Francisco
17. Gepshtein S, Tyukin I, Kubovy M (2011) A failure of the proximity principle in the perception of motion. Humana Mente 17:21–34
18. Ilea DE, Whelan PF (2011) Image segmentation based on the integration of colour-texture descriptors—a review. Pattern Recogn 44(10):2479–2501
19. Marquez-Neila P, Baumela L, Alvarez L (2014) A morphological approach to curvature-based evolution of curves and surfaces. IEEE Trans Pattern Anal Mach Intell 36(1):2–17
20. Fox V, Milanova M, Al-Ali S (2013) A morphological multiphase active contour for vascular segmentation. Int J Bioinform Biosci 3(3):1–12
21. Battiato S, Lukac R (2008) Video stabilization techniques. Encyclopedia of multimedia. Springer, New York, pp 941–945
22. Rawat P, Singhai J (2011) Review of motion estimation and video stabilization techniques for hand held mobile video. Int J Signal Image Process 2(2):159–168
23. Mironowsky LA, Slaev VA (2011) Strip-method for image and signal transformation. De Gruyter, Berlin
24. Hadamard J (1893) Resolution d'une Question Relative aux Determinants. Bull Sci Math ser 2 17(1):240–246
25. Yüksel S, Başar T (2013) Stochastic networked control systems. Springer, New York, Heidelberg
26. Govindan M, Tang CM (2010) Information system evaluation: an ongoing measure. Int J Business Inform Syst 6(3):336–353