1

Motivation

In our everyday life, each of us constantly receives, processes, and analyzes a huge amount of information of various kinds, significance and quality, and has to make decisions based on this analysis. More than 95% of information we perceive is of a visual character. An image is a very powerful information medium and communication tool capable of representing complex scenes and processes in a compact and efficient way. Thanks to this, images are not only primary sources of information, but are also used for communication among people and for interaction between humans and machines. Common digital images contain enormous amounts of information. An image you can take and send to your friends using a smart phone in a few seconds contains as much information as hundreds of text pages. This is why in many application areas, such as robot vision, surveillance, medicine, remote sensing, and astronomy, there is an urgent need for automatic and powerful image analysis methods.

1.1 Image analysis by computers

Image analysis in a broad sense is a many-step process the input of which is an image while the output is a final (usually symbolic) piece of information or a decision. Typical examples are localization of human faces in the scene and recognition (against a list or a database of persons) who is who, finding and recognizing road signs in the visual field of the driver, and identification of suspect tissues in a CT or MRI image. A general image analysis flowchart is shown in Figure 1.1 and an example what the respective steps look like in the car licence plate recognition is shown in Figure 1.2.

The first three steps of image analysis—image acquisition, preprocessing, and object segmentation/detection—are comprehensively covered in classical image processing textbooks [1–5], in recent specialized monographs [6–8] and in thousands of research papers. We very briefly recall them below, but we do not deal with them further in this book. The topic of this book falls into the fourth step, the feature design. The book is devoted to one particular family of features which are based on image moments. The last step, object classification, is shortly reviewed in Chapter 2, but its deeper description is beyond the scope of this book.

In *image acquisition*, the main theoretical questions are how to choose the sampling scheme, the sampling frequency and the number of the quantization levels such that the artifacts caused by aliasing, moire, and quantization noise do not degrade the image much while keeping the

2D and 3D Image Analysis by Moments, First Edition. Jan Flusser, Tomáš Suk and Barbara Zitová.

Companion website: www.wiley.com/go/flusser/2d3dimageanalysis

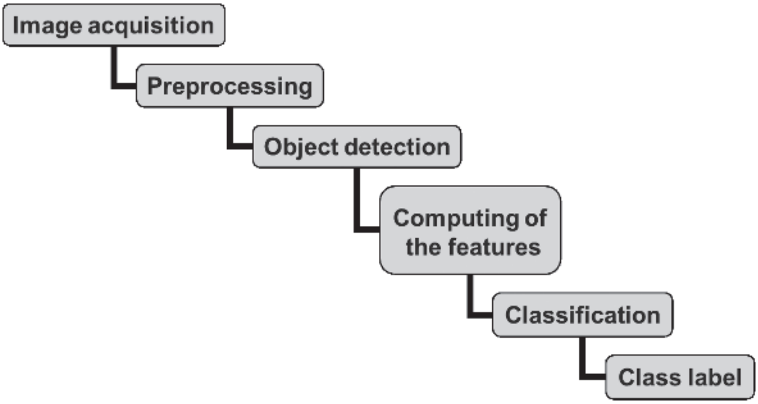

Figure 1.1 General image analysis flowchart

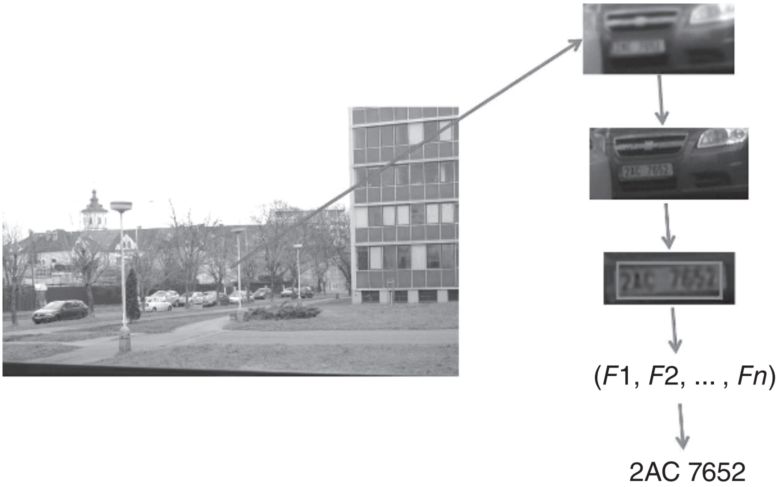

Figure 1.2 An example of the car licence plate recognition

image size reasonably low (there are of course also many technical questions about the appropriate choice of the camera and the spectral band, the objective, the memory card, the transmission line, the storage format, and so forth, which we do not discuss here).

Since real imaging systems as well as imaging conditions are usually imperfect, the acquired image represents only a degraded version of the original scene. Various kinds of degradations (geometric as well as graylevel/color) are introduced into the image during the acquisition process by such factors as imaging geometry, lens aberration, wrong focus, motion of the scene, systematic and random sensor errors, noise, etc. (see Figure 1.3 for the general scheme and an illustrative example). Removal or at least suppression of these degradations is a subject of *image preprocessing*. Historically, image preprocessing was one of the first topics systematically studied in digital image processing (already in the very first monograph [9] there was a chapter

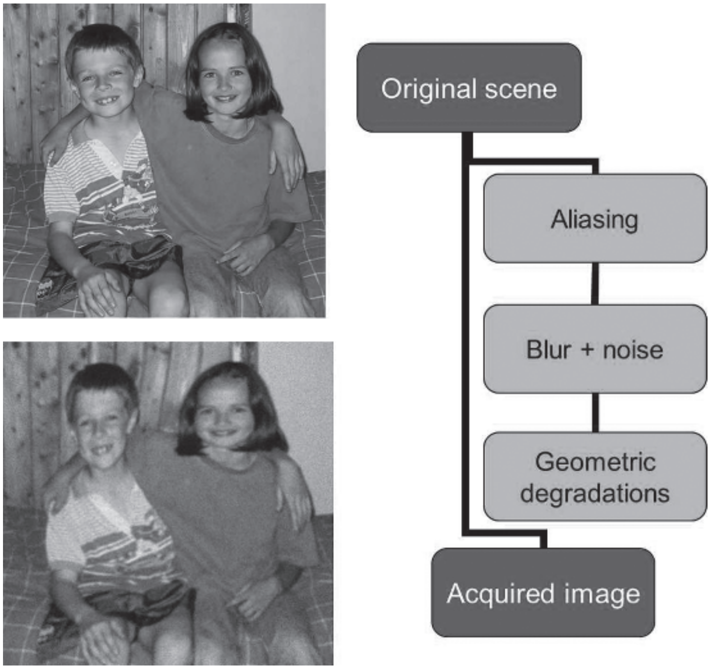

Figure 1.3 Image acquisition process with degradations

devoted to this topic) because even simple preprocessing methods were able to enhance the visual quality of the images and were feasible on old computers. The first two steps, image acquisition and preprocessing, are in the literature often categorized into *low-level processing*. The characteristic feature of the low-level methods is that both their input and output are digital images. On the contrary, in *high-level processing*, the input is a digital image (often an output of some preprocessing) while the output is a symbolic (i.e. high-level) information, such as the coordinates of the detected objects, the list of boundary pixels, etc.

Object detection is a typical example of high-level processing. The goal is to localize the objects of interest in the image and separate (segment) them from the other objects and from the background[1]. Hundreds of segmentation methods have been described in the literature. Some of them are universal, but most of them were designed for specific families of objects such as characters, logos, cars, faces, human silhouettes, roads, etc. A good basic survey of object detection and segmentation methods can be found in [5] and in the references therein.

Feature definition and computing is probably the most challenging part of image analysis. The features should provide an accurate and unambiguous quantitative description of the

[1] Formally, the terms such as "object", "boundary", "interior", "background", and others are subject of the discrete topology we are working in. Various discrete topologies differ from one another by the notion of neighboring pixels and hence by the notion of connectivity. However, here we use the above terms intuitively without specifying the particular topology.

objects[2]. The feature values are elements of the *feature space* which should be for the sake of efficient computation of low dimensionality. The design of the features is highly dependent on the type of objects, on the conditions under which the images have been acquired, on the type and the quality of preprocessing, and on the application area. There is no unique "optimal" solution.

Classification/recognition of the object is the last stage of the image analysis pipeline. It is entirely performed in the feature space. Each unknown object, now being represented by a point in the feature space, is classified as an element of a certain class. The classes can be specified in advance by their representative samples, which create a *training set*; in such a case we speak about *supervised classification*. Alternatively, if no training set is available, the classes are formed from the unknown objects based on their distribution in the feature space. This case is called *unsupervised classification* or *clustering*, and in visual object recognition this approach is rare. Unlike the feature design, the classification algorithms are application independent—they consider neither the nature of the original data nor the physical meaning of the features. Classification methods are comprehensively reviewed in the famous Duda-Hart-Stork book [10] and in the most recent monograph [11]. The use of the classification methods is not restricted to image analysis. We can find numerous applications in artificial intelligence, decision making, social sciences, statistical data analysis, and in many other areas beyond the scope of this book.

1.2 Humans, computers, and object recognition

Why is image analysis so easy and natural for human beings and so difficult for machines? There are several reasons for this performance difference. Incorporating lifetime experience, applying information fusion, the ability of contextual perception, and perceptual robustness are the key elements. Current research directions are often motivated by these factors, but up to now the results have been achieved only in laboratory conditions and have not reached the quality of human beings yet. Probably the most serious difference is that humans incorporate their lifetime knowledge into the recognition process and do so in a very efficient way. If you see an object you have already met (even a long time ago), you are able to "retrieve" this information from your brain almost immediately (it appears that the "retrieval time" does not depend on the number of the objects stored in the brain) and use it as a hint for the classification. Your brain is also able to supply the missing information if only a part of the object is visible. This ability helps us a lot when recognizing partially occluded objects. Computers could in principle do the same, but it would require massive learning on large-scale databases, which would lead not only to a time-consuming learning stage but also to relatively slow search/retrieval of the learned objects. To overcome that, several projects have been started recently which are trying to substitute human life experience by shared knowledge available at search engines (see [12] for example) but this research is still at the beginning stages.

Our ability to combine supplementary information of other kinds in parallel with vision makes human perception efficient, too. We also have other senses (hearing, smell, taste, and

[2] There exists an alternative way of object description by its structure. The respective recognition methods are then called structural or syntactic ones. They use the apparatus of formal language theory. Each object class is given by a language, and each object is a word. An object is assigned to the class if its word belongs to the class language. This approach is useful if the structural information is more appropriate than a quantitative description.

touch), and we are very good in combining them with the visual input and making a decision by fusing all these input channels. This helps us a lot, especially if the visual information is insufficient such as in the darkness. Modern-day robots are often equipped with such sensors, even working in modalities we are not able to perceive. However the usage of such multi-modal robots is limited by their cost and, what is more important, by the non-triviality of efficient fusion of the acquired information.

Contextual perception used in object recognition gives humans yet another advantage comparing to computers. Computer algorithms either do not use the context at all because they work with segmented isolated objects or classify the objects in the neighborhood, and the contextual relations are used as additional features or prior probability estimators (for instance, if we recognize a car in front of the unknown object and another car behind it, the prior probability that the object in question is also a car is high). Such investigation of the context is, however, time expensive. Humans perceive, recognize, and evaluate the context very quickly which gives them the ability of "recognizing" even such objects that are not visible at all at the moment. Imagine you see a man with a leash in his hand on the picture. You can, based on context (a man in the a park) and your prior knowledge, quickly deduce that the animal on the other end of the leash, that cannot be seen in the photo, is a dog (which actually may not be true in all cases). This kind of thinking is very complex and thus difficult to implement on a computer.

Yet another advantage of human vision over computer methods is its high robustness to the change of the orientation, scale, and pose of the object. Most people can easily recognize a place or a person on a photograph that is upside down or mirrored. The algorithms can do that as well, but it requires the use of special features, which are insensitive to such modifications. Standard features may change significantly under spatial image transformations, which may lead to misclassifications.

From this short introduction, we can see that automatic object recognition is a challenging and complex task, important for numerous application areas, which requires sophisticated algorithms in all its stages. This book contributes to the solution of this problem by systematic research of one particular type of features, which are known as *image moments*.

1.3 Outline of the book

This book deals systematically with moments and moment invariants of 2D and 3D images and with their use in object description, recognition, and in other applications.

Chapter 2 introduces basic terms and concepts of object recognition such as feature spaces and related norms, equivalences, and space partitions. Invariant based approaches are introduced together with normalization methods. Eight basic categories of invariants are reviewed together with illustrative examples of their usage. We also recall main supervised and unsupervised classifiers as well as related topics of classifier fusion and reduction of the feature space dimensionality.

Chapters 3 – 6 are devoted to four classes of moment invariants. In Chapter 3, we introduce 2D moment invariants with respect to the simplest spatial transformations—translation, rotation, and scaling. We recall the classical Hu invariants first, and then we present a general method for constructing invariants of arbitrary orders by means of complex moments.

We prove the existence of a relatively small basis of invariants that is complete and independent. We also show an alternative approach—constructing invariants via normalization. We discuss the difficulties which the recognition of symmetric objects poses and present moment invariants suitable for such cases.

In Chapter 4 we introduce 3D moment invariants with respect to translation, rotation, and scaling. We present the derivation of the invariants by means of three approaches—the tensor method, the expansion into spherical harmonics, and the object normalization. Similarly as in 2D, the symmetry issues are also discussed there.

Chapter 5 deals with moment invariants to the affine transformation of spatial coordinates. We present four main approaches showing how to derive them in 2D—the graph method, the method of normalized moments, the transvectants, and the solution of the Cayley-Aronhold equation. Relationships between the invariants produced by different methods are mentioned, and the dependency among the invariants is studied. We describe a technique used for elimination of reducible and dependent invariants. Numerical experiments illustrating the performance of the affine moment invariants are carried out, and a generalization to color images, vector fields, and 3D images is proposed.

Chapter 6 deals with a completely different kind of moment invariants, with invariants to image blurring. We introduce the theory of projection operators, which allows us to derive invariants with respect to image blur regardless of the particular convolution kernel, provided it has a certain type of symmetry. We also derive so-called combined invariants, which are invariant to composite geometric and blur degradations. Knowing these features, we can recognize objects in the degraded scene without any restoration/deblurring.

Chapter 7 presents a survey of various types of orthogonal moments. The moments orthogonal on a rectangle/cube as well as the moments orthogonal on a unit disk/sphere are described. We review Legendre, Chebyshev, Gegenbauer, Jacobi, Laguerre, Gaussian-Hermite, Krawtchouk, dual Hahn, Racah, Zernike, pseudo-Zernike, and Fourier-Mellin polynomials and moments. The use of the moments orthogonal on a disk in the capacity of rotation invariants is discussed. The second part of the chapter is devoted to image reconstruction from its moments. We explain why orthogonal moments are more suitable for reconstruction than geometric ones, and a comparison of reconstructing power of different orthogonal moments is presented.

In Chapter 8, we focus on computational issues. Since the computing complexity of all moment invariants is determined by the computing complexity of the moments, efficient algorithms for moment calculations are of prime importance. There are basically two major groups of methods. The first one consists of methods that attempt to decompose the object into non-overlapping regions of a simple shape, the moments of which can be computed very fast. The moment of the object is then calculated as a sum of the moments of all regions. The other group is based on Green's theorem, which evaluates the integral over the object by means of a less-dimensional integration over the object boundary. We present efficient algorithms for binary and graylevel objects and for geometric as well as for selected orthogonal moments.

Chapter 9 is devoted to various applications of moments and moment invariants in image analysis. We demonstrate their use in image registration, object recognition, medical imaging, content-based image retrieval, focus/defocus measurement, forensic applications, robot navigation, digital watermarking, and others.

Chapter 10 contains a summary and a concluding discussion.

References

[1] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. Academic Press, 2nd ed., 1982.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice Hall, 1977.

[3] W. K. Pratt, *Digital Image Processing*. Wiley Interscience, 4th ed., 2007.

[4] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 3rd ed., 2007.

[5] M. Šonka, V. Hlaváč, and R. Boyle, *Image Processing, Analysis and Machine Vision*. Thomson, 3rd ed., 2007.

[6] P. Campisi and K. Egiazarian, *Blind Image Deconvolution: Theory and Applications*. CRC Press, 2007.

[7] P. Milanfar, *Super-Resolution Imaging*. CRC Press, 2011.

[8] A. N. Rajagopalan and R. Chellappa, *Motion Deblurring: Algorithms and Systems*. Cambridge University Press, 2014.

[9] A. Rosenfeld, *Picture Processing by Computer*. Academic Press, 1969.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley Interscience, 2nd ed., 2001.

[11] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 4th ed., 2009.

[12] X. Chen, A. Shrivastava, and A. Gupta, "NEIL: Extracting visual knowledge from web data," in *The 14th International Conference on Computer Vision ICCV'13*, pp. 1409–1416, 2013.