Chapter 1

INTRODUCTION TO MOMENTS

1.1 Motivation

In our everyday life, each of us nearly constantly receives, processes, and analyzes a huge amount of information of various kind, significance and quality and has to make decisions based on this analysis. More than 95% of information we perceive is of optical character. 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 amount of information. An image you can take and send to your friends using a cell phone in a few seconds contains as much information as several hundreds of text pages. This is why there is an urgent need for automatic and powerful image analysis methods.

Analysis and interpretation of an image acquired by a real (i.e. non-ideal) imaging system is the key problem in many application areas such as robot vision, remote sensing, astronomy and medicine, to name a few. Since real imaging systems as well as imaging conditions are usually imperfect, the observed 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, etc. (see Figs. 1.1, 1.2, and 1.3 for illustrative examples).

Figure 1.1: Perspective distortion of the image caused by a non-perpendicular view.

In the general case, the relation between the ideal image $f(x,y)$ and the observed image $g(x,y)$ is described as $g = \mathcal{D}(f)$, where $\mathcal{D}$ is a degradation operator. Degradation operator $\mathcal{D}$ can usually be decomposed into radiometric (i.e. graylevel or color) degradation operator $\mathcal{R}$ and geometric (i.e. spatial) degradation operator $\mathcal{G}$. In real imaging systems $\mathcal{R}$ can be usually modeled by space-variant or space-invariant convolution plus noise while $\mathcal{G}$ is typically a transform of spatial coordinates (for instance perspective projection). In practice, both operators are typically either unknown or are described by a parametric model with unknown parameters. Our goal is to analyze the unknown scene $f(x,y)$, an ideal image of which is not available, by means of the sensed image $g(x,y)$ and a priori information about the degradations.

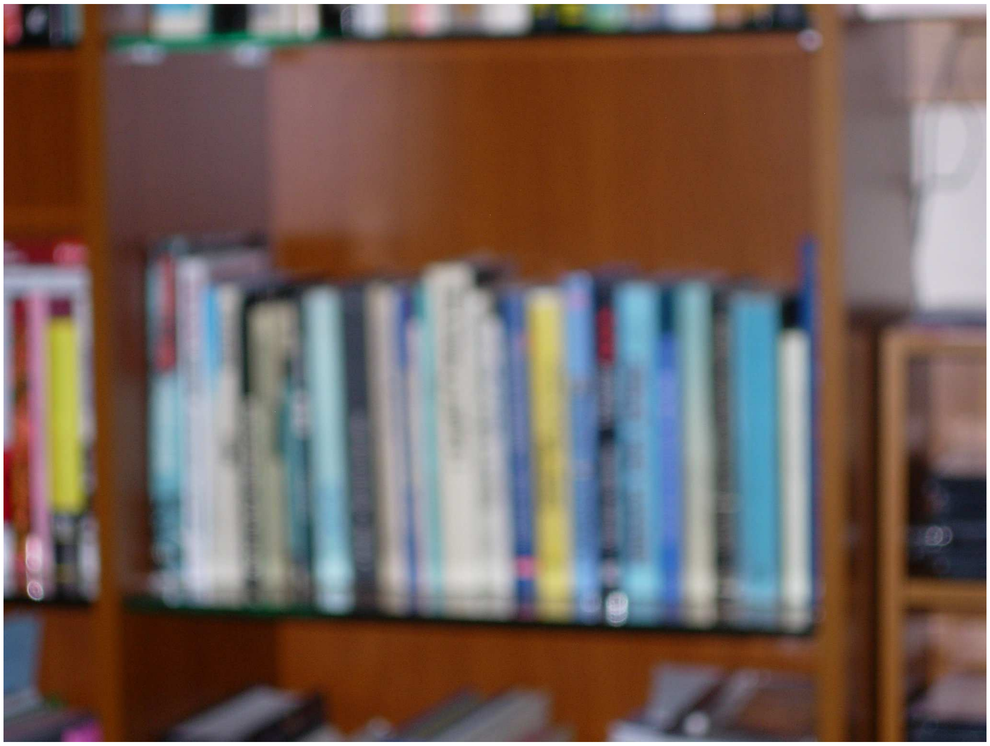

Figure 1.2: Image blurring caused by wrong focus of the camera.

By the term *scene analysis* we usually understand a complex process consisting of three basic stages. First, the image is preprocessed, segmented and objects of potential interest are detected. Secondly, the extracted objects are "recognized", which means they are mathematically described and classified as elements of a certain class from the set of pre-defined object classes. Finally, spatial relations among the objects can be analyzed. The first stage contains traditional image processing methods and is exhaustively covered in standard textbooks [1], [2], [3]. The classification stage is independent of the original data and is carried out in the space of descriptors. This part is comprehensively reviewed in the famous Duda-Hart-Stork book [4]. For the last stage we again refer to [3].

Figure 1.3: Image distortion caused by a non-linear deformation of the scene.

1.2 What are invariants?

Recognition of objects and patterns that are deformed in various ways has been a goal of much recent research. There are basically three major approaches to this problem – brute force, image normalization, and invariant features. In the brute force approach we search the parametric space of all possible image degradations. That means the training set of each class should contain not only all class representatives but also all their rotated, scaled, blurred, and deformed versions. Clearly, this approach would lead to extreme time complexity and is practically inapplicable. In the normalization approach, the objects are transformed into a certain standard position before they enter the classifier. This is very efficient in the classification stage but the object normalization itself usually requires solving difficult inverse problems that are often ill-conditioned or even ill-posed. For instance, in case of image blurring, "normalization" means in fact blind deconvolution [5] and in case of spatial image deformation, "normalization" requires to perform registration of the image to some reference frame [6].

The approach using invariant features appears to be the most promising and has been used extensively. Its basic idea is to describe the objects by a set of measurable quantities called *invariants* that are insensitive to particular deformations and that provide enough discrimination power to distinguish among objects belonging to different classes. From mathematical point of view, invariant I is a functional defined on the space of all admissible image functions which does not change its value under degradation operator $\mathcal{D}$, i.e. which satisfies the condition $I(f) = I(\mathcal{D}(f))$ for any image function f. This property is called *invariance*. In practice, in order to accommodate the influence of imperfect segmentation, intra-class variability and noise, we usually formulate this requirement as a weaker constraint: $I(f)$ should not be significantly different from $I(\mathcal{D}(f))$. Another desirable property of I, as important as invariance, is *discriminability*. For objects belonging to different classes, I must have significantly different values. Clearly, these two requirements are antagonistic – the broader the invariance, the less discrimination power and vice versa. Choosing a proper trade-off between invariance and discrimination power is a very important task in feature-based object recognition (see Fig. 1.4 for an example of desired situation).

Usually one invariant does not provide enough discrimination power and several invariants $I_1, \ldots, I_n$ must be used simultaneously. Then we speak about an *invariant vector*. In this way, each object is represented by a point in an n-dimensional metric space called *feature space* or *invariant space*.

1.2.1 Categories of the invariants

The existing invariant features used for describing 2D objects can be categorized from various points of view. Most straightforward is the categorization according to the type of invariance. We recognize translation, rotation, scaling, affine, projective, and elastic geometric invariants. Radiometric invariants exist with respect to linear contrast stretching, non-linear intensity transforms, and to convolution.

Categorization according to the mathematical tools used may be as follows.

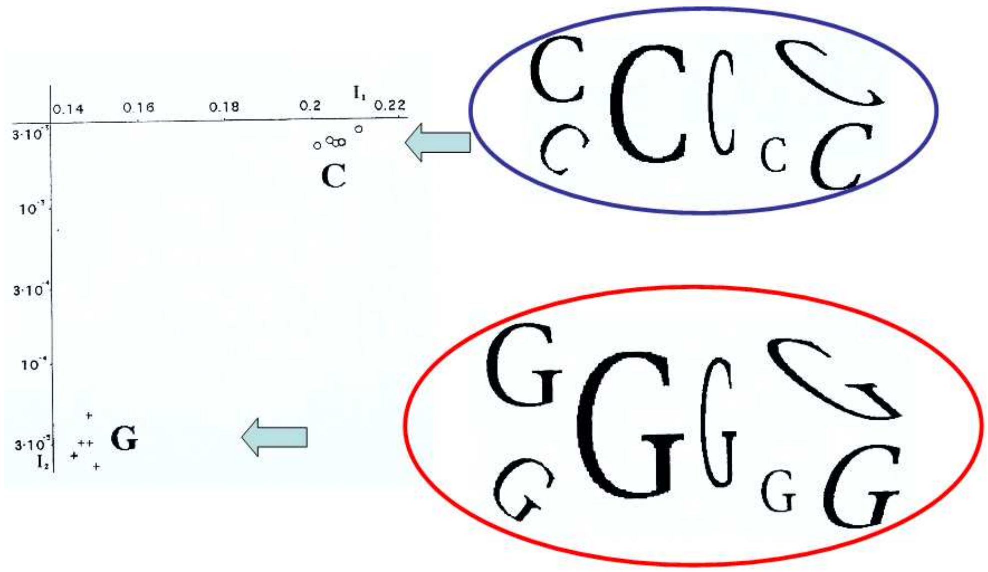

Figure 1.4: Two-dimensional feature space with two classes, almost an ideal example. Each class forms a compact cluster (the features are invariant) and the clusters are well separated (the features are discriminative).

- *Simple shape descriptors* – compactness, convexity, elongation, etc. [3].
- *Transform coefficient features* are calculated from a certain transform of the image – Fourier descriptors [7], [8], Hadamard descriptors, Radon transform coefficients, and wavelet-based features [9], [10].
- *Point set invariants* use positions of dominant points [11], [12], [13], [14].
- *Differential invariants* employ derivatives of the object boundary [15], [16], [17], [18], [19].
- *Moment invariants* are special functions of image moments.

Another viewpoint reflects what part of the object is needed to calculate the invariant.

- *Global* invariants are calculated from the whole image (including background if no segmentation was performed). Most of them include projections of the image onto certain basis functions and are calculated by integration. Comparing to local invariants, global invariants are much more robust with respect to noise, inaccurate boundary detection and other similar factors. On the other hand, their serious drawback is the fact, that a local change of the image influences values of all invariants and is not "localized" in a few components only. This is why global invariants cannot be used when the studied object is partially occluded by another object and/or when a part of it is out of the visual field. Moment invariants fall into this category.

- *Local* invariants are, on the contrary, calculated from a certain neighborhood of dominant points only. Differential invariants are typical representatives of this category. The object boundary is detected first and then the invariants are calculated for each boundary point as functions of the boundary derivatives. Thanks to this, the invariants at any given point depend only on the shape of the boundary in its immediate vicinity. If the rest of the object undergoes any change, the local invariants are not affected. This property makes them a seemingly perfect tool for recognition of partially occluded objects but due to their extreme vulnerability to discretization errors, segmentation inaccuracies, and noise it is difficult to actually implement and use them in practice.

- *Semi-Local* invariants attempt to keep positive properties of the two above groups and to avoid the negative ones. They divide the object into stable parts (most often this division is based on inflection points or vertices of the boundary) and describe each part by some kind of global invariants. The whole object is then characterized by a string of vectors of invariants and recognition under occlusion is performed by maximum substring matching. This modern and practically applicable approach was used in various modifications in [20], [21], [22], [23], [24], [25], [26].

In this book, we focus on object description and recognition by means of moments and moment invariants. The history of moment invariants began many years before the appearance of the first computers, in the

19th century under the framework of the group theory and of the theory of algebraic invariants. The theory of algebraic invariants was thoroughly studied by famous German mathematicians P.A. Gordan and D. Hilbert [27] and was further developed in the 20th century in [28] and [29], among others.

Moment invariants were first introduced to the pattern recognition and image processing community in 1962 [30], when Hu employed the results of the theory of algebraic invariants and derived his seven famous invariants to rotation of 2-D objects. Since that time, hundreds of papers have been devoted to various improvements, extensions and generalizations of moment invariants and also to their use in many areas of application. Moment invariants have become one of the most important and most frequently used shape descriptors. Even though they suffer from certain intrinsic limitations (the worst of which is their globalness, which prevents direct utilization for occluded object recognition), they frequently serve as "first-choice descriptors" and as a reference method for evaluating the performance of other shape descriptors. Despite a tremendous effort and huge number of published papers, many open problems remain to be resolved.

1.3 What are moments?

Moments are scalar quantities used for hundreds of years to characterize a function and to capture its significant features. They have been widely used in statistics for description of the shape of a probability density function and in classic rigid-body mechanics to measure the mass distribution of a body. From the mathematical point of view, moments are "projections" of a function onto a polynomial basis (similarly, Fourier transform is a projection onto a basis of harmonic functions). For the sake of clarity, we introduce some basic terms and propositions, which we will use throughout the book.

Definition 1: By an *image function* (or *image*) we understand any piece-wise continuous real function $f(x, y)$ of two variables defined on a compact support $D \subset \mathbb{R} \times \mathbb{R}$ and having a finite nonzero integral.

Definition 2[1]**:** *General moment* $M_{pq}^{(f)}$ of an image $f(x, y)$, where p, q are non-negative integers and $r = p + q$ is called the *order* of the moment, is defined as

$$M_{pq}^{(f)} = \int\int_D p_{pq}(x, y) f(x, y) dx dy, \tag{1.1}$$

where $p_{00}(x, y), p_{10}(x, y), \ldots, p_{kj}(x, y), \ldots$ are polynomial basis functions defined on D. (We omit the superscript $^{(f)}$ if there is no danger of confusion.)

Depending on the polynomial basis used, we recognize various systems of moments.

1.3.1 Geometric and complex moments

The most common choice is a standard power basis $p_{kj}(x, y) = x^k y^j$ that leads to *geometric moments*

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x, y) \mathrm{d}x\mathrm{d}y. \tag{1.2}$$

Geometric moments of low orders have an intuitive meaning – m_{00} is a "mass" of the image (for binary images, m_{00} is an area of the object), m_{10}/m_{00} and m_{01}/m_{00} define the *center of gravity* or *centroid* of the image. Second-order moments m_{20} and m_{02} describe the "distribution of mass" of the image with respect to the coordinate axes. In mechanics they are called the *moments of inertia*. Another popular mechanical quantity, the *radius of gyration* with respect to an axis, can be also expressed in terms of moments as $\sqrt{m_{20}/m_{00}}$ and $\sqrt{m_{02}/m_{00}}$, respectively.

If the image is considered a probability density function (i.e. its values are normalized such that $m_{00} = 1$), then m_{10} and m_{01} are the mean values. In case of zero means, m_{20} and m_{02} are *variances* of horizontal and vertical projections and m_{11} is a *covariance* between them. In this way, the second-order moments define the orientation of the image. As will be seen later, second-order geometric moments can be used to find the

[1] In some papers one can find extended versions of Definition 2 that include various scalar factors and/or weighting functions in the integrand. We introduce such extensions in Chapter 6.

normalized position of an image. In statistics, two higher-order moment characteristics have been commonly used – the *skewness* and the *kurtosis*. The skewness of the horizontal projection is defined as $m_{30}/\sqrt{m_{20}^3}$ and that of vertical projection as $m_{03}/\sqrt{m_{02}^3}$. The skewness measures the deviation of the respective projection from symmetry. If the projection is symmetric with respect to the mean (i.e. to the origin in this case), then the corresponding skewness equals zero. The kurtosis measures the "peakedness" of the probability density function and is again defined separately for each projection – the horizontal kurtosis as m_{40}/m_{20}^2 and the vertical kurtosis as m_{04}/m_{02}^2.

Characterization of the image by means of geometric moments is complete in the following sense. For any image function, geometric moments of all orders do exist and are finite. The image function can be exactly reconstructed from the set of its moments (this assertion is known as *the uniqueness theorem*).

Another popular choice of the polynomial basis $p_{kj}(x,y) = (x+iy)^k(x-iy)^j$, where i is the imaginary unit, leads to *complex moments*

$$c_{pq} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x+iy)^p(x-iy)^q f(x,y)dxdy. \tag{1.3}$$

Geometric moments and complex moments carry the same amount of information. Each complex moment can be expressed in terms of geometric moments of the same order as

$$c_{pq} = \sum_{k=0}^{p}\sum_{j=0}^{q}\binom{p}{k}\binom{q}{j}(-1)^{q-j}\cdot i^{p+q-k-j}\cdot m_{k+j,p+q-k-j} \tag{1.4}$$

and vice versa[2]

$$m_{pq} = \frac{1}{2^{p+q}i^q}\sum_{k=0}^{p}\sum_{j=0}^{q}\binom{p}{k}\binom{q}{j}(-1)^{q-j}\cdot c_{k+j,p+q-k-j}. \tag{1.5}$$

Complex moments are introduced because they behave favorably under image rotation. This property can be advantageously employed when constructing invariants with respect to rotation, as will be shown in the following chapter.

[2] While the proof of (1.4) is straightforward, the proof of (1.5) requires first to express x and y as $x = ((x+iy)+(x-iy))/2$ and $y = ((x+iy)-(x-iy))/2i$.

1.3.2 Orthogonal moments

If the polynomial basis $\{p_{kj}(x,y)\}$ is orthogonal, i.e. if its elements satisfy the condition of orthogonality

$$\iint_{\Omega} p_{pq}(x,y) \cdot p_{mn}(x,y) dxdy = 0 \tag{1.6}$$

or weighted orthogonality

$$\iint_{\Omega} w(x,y) \cdot p_{pq}(x,y) \cdot p_{mn}(x,y) dxdy = 0 \tag{1.7}$$

for any indexes $p \neq m$ or $q \neq n$, we speak about *orthogonal (OG) moments*. Ω is the area of orthogonality.

In theory, all polynomial bases of the same degree are equivalent because they generate the same space of functions. Any moment with respect to a certain basis can be expressed in terms of moments with respect to any other basis. From this point of view, OG moments of any type are equivalent to geometric moments.

However, a significant difference appears when considering stability and computational issues in a discrete domain. Standard powers are nearly dependent both for small and large values of the exponent and increase rapidly in range as the order increases. This leads to correlated geometric moments and to the need for high computational precision. Using lower precision results in unreliable computation of geometric moments. OG moments can capture the image features in an improved, non-redundant way. They also have the advantage of requiring lower computing precision because we can evaluate them using recurrent relations, without expressing them in terms of standard powers.

Unlike geometric moments, OG moments are coordinates of f in the polynomial basis in the common sense used in linear algebra. Thanks to this, the image reconstruction from OG moments can be performed easily as

$$f(x,y) = \sum_{k,j} M_{kj} \cdot p_{kj}(x,y).$$

Moreover, this reconstruction is "optimal" because it minimizes the mean-square error when using only a finite set of moments. On the other hand, image reconstruction from geometric moments cannot be performed directly in the spatial domain. It is carried out in the Fourier domain using the fact that geometric moments form Taylor coefficients of the Fourier transform $F(u, v)$

$$F(u,v) = \sum_p \sum_q \frac{(-2\pi i)^{p+q}}{p!q!} m_{pq} u^p v^q.$$

(To prove this, expand the kernel of the Fourier transform $e^{-2\pi i(ux+vy)}$ into a power series.) Reconstruction of $f(x, y)$ is then achieved via inverse Fourier transform.

We will discus various OG moments and their properties in detail in Chapter 6. Their usage for stable implementation of implicit invariants will be shown in Chapter 4 and practical applications will be demonstrated in Chapter 8.

1.4 The outline of the book

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

Chapters 2 – 5 are devoted to four classes of moment invariants. In Chapter 2, we introduce moment invariants with respect to the simplest spatial transforms – translation, rotation and scaling. We recall the classical Hu invariants first and then 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.

Chapter 3 deals with moment invariants to the affine transform of spatial coordinates. We present three main approaches showing how to derive them – the graph method, the method of normalized moments, and the solution of the Cayley-Aronhold equation. Relationships between invariants from different methods are mentioned and the dependency of generated invariants is studied. We describe a technique used for elimination of reducible and dependent invariants. Finally, numerical experiments illustrating the performance of the affine moment invariants are carried out and a brief generalization to color images and to 3D images is proposed.

In Chapter 4, we introduce a novel concept of so-called implicit invariants to elastic deformations. Implicit invariants measure the similarity between two images factorized by admissible image deformations. For many types of image deformations traditional invariants do not exist but implicit invariants can be used as features for object recognition. In the chapter, we present implicit moment invariants with respect to the polynomial transform of spatial coordinates and demonstrate their performance in artificial as well as real experiments.

Chapter 5 deals with a completely different kind of moment invariants, with invariants to convolution/blurring. We derive invariants with respect to image blur regardless of the convolution kernel provided it has a certain degree 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.

Chapter 6 presents a survey of various types of orthogonal moments. They are divided into two groups, the first being moments orthogonal on a rectangle and the second orthogonal on a unit disk. We review Legendre, Chebyshev, Gegenbauer, Jacobi, Laguerre, 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 7, we focus on computational issues. Since the computing complexity of all moment invariants is determined by the computing complexity of 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. These ”elementary shapes” can be pixel rows or their segments, square and rectangular blocks, among others. A moment of the object is then calculated as a sum of moments of all regions. The other group is based on Green's theorem, which evaluates the double integral over the object by means of single integration along the object boundary.

We present efficient algorithms for binary and graylevel objects and for geometric as well as selected orthogonal moments.

Chapter 8 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, and digital watermarking.

Bibliography

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

[2] W. K. Pratt, *Digital Image Processing*. New York: Wiley Interscience, 4th ed., 2007.

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

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

[5] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.

[6] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[7] C. C. Lin and R. Chellapa, "Classification of partial 2-D shapes using Fourier descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 686–690, 1987.

[8] K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger, "Application of affine-invariant Fourier descriptors to recognition of 3-D objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 640–647, 1990.

[9] Q. M. Tieng and W. W. Boles, "An application of wavelet-based affine-invariant representation," *Pattern Recognition Letters*, vol. 16, no. 12, pp. 1287–1296, 1995.

[10] M. Khalil and M. Bayeoumi, "A dyadic wavelet affine invariant function for 2D shape recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1152–1163, 2001.

[11] J. L. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*. Cambridge, Massachusetts: MIT Press, 1992.

[12] T. Suk and J. Flusser, "Vertex-based features for recognition of projectively deformed polygons," *Pattern Recognition*, vol. 29, no. 3, pp. 361–367, 1996.

[13] R. Lenz and P. Meer, "Point configuration invariants under simultaneous projective and permutation transformations," *Pattern Recognition*, vol. 27, no. 11, pp. 1523–1532, 1994.

[14] N. S. V. Rao, W. Wu, and C. W. Glover, "Algorithms for recognizing planar polygonal configurations using perspective images," *IEEE Trans. Robotics and Automation*, vol. 8, no. 4, pp. 480–486, 1992.

[15] E. Wilczynski, *Projective Differential Geometry of Curves and Ruled Surfaces*. Leipzig: B. G. Teubner, 1906.

[16] I. Weiss, "Projective invariants of shapes," in *Proc. Computer Vision and Pattern Recognition CVPR '88*, pp. 1125–1134, IEEE Computer Society, 1988.

[17] C. A. Rothwell, A. Zisserman, D. A. Forsyth, and J. L. Mundy, "Canonical frames for planar object recognition," in *Proc. 2nd European Conf. Computer Vision ECCV'92*, pp. 757–772, Springer, 1992.

[18] I. Weiss, "Differential invariants without derivatives," in *Proc. 11th Int'l Conf. Pattern Recognition ICPR'92*, pp. 394–398, 1992.

[19] F. Mokhtarian and S. Abbasi, "Shape similarity retrieval under affine transforms," *Pattern Recognition*, vol. 35, no. 1, pp. 31–41, 2002.

[20] W. S. Ibrahim Ali and F. S. Cohen, "Registering coronal histological 2-D sections of a rat brain with coronal sections of a 3-D brain atlas using geometric curve invariants and B-spline representanion," *IEEE Trans. Medical Imaging*, vol. 17, no. 6, pp. 957–966, 1998.

[21] Z. Yang and F. Cohen, "Image registration and object recognition using affine invariants and convex hulls," *IEEE Trans. Image Processing*, vol. 8, no. 7, pp. 934–946, 1999.

[22] J. Flusser, "Affine invariants of convex polygons," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 1117–1118, 2002.

[23] C. A. Rothwell, A. Zisserman, D. A. Forsyth, and J. L. Mundy, "Fast recognition using algebraic invariants," in *Geometric Invariance in Computer Vision* (J. L. Mundy and A. Zisserman, eds.), pp. 398–407, MIT Press, 1992.

[24] Y. Lamdan, J. Schwartz, and H. Wolfson, "Object recognition by affine invariant matching," in *Proc. Computer Vision and Pattern Recognition CVPR'88*, pp. 335–344, IEEE Computer Society, 1988.

[25] F. Krolupper and J. Flusser, "Polygonal shape description for recognition of partially occluded objects," *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1002–1011, 2007.

[26] O. Horáček, J. Kamenický, and J. Flusser, "Recognition of partially occluded and deformed binary objects," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 360–369, 2008.

[27] D. Hilbert, *Theory of Algebraic Invariants*. Cambridge: Cambridge University Press, 1993.

[28] G. B. Gurevich, *Foundations of the Theory of Algebraic Invariants*. Groningen, The Netherlands: Nordhoff, 1964.

[29] I. Schur, *Vorlesungen über Invariantentheorie*. Berlin: Springer, 1968. (in German).

[30] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.