Chapter 13
Improving the Recognition Performance of Moment Features by Selection

George A. Papakostas

Abstract This chapter deals with the selection of the most appropriate moment features used to recognize known patterns. This chapter aims to highlight the need for selection of moment features subject to their descriptive capabilities. For this purpose, some popular moment families are presented and their properties, making them suitable for pattern recognition tasks, are discussed. Two different types of feature selection algorithms, a simple Genetic Algorithm (GA) and the Relief algorithm are applied to select the moment features that better discriminate human faces and facial expressions, under several pose and illumination conditions. Appropriate experiments using four benchmark datasets have been conducted in order to investigate the theoretical assertions. An extensive experimental analysis has shown that the recognition performance of the moment features can be significantly improved by selecting them from a predefined pool, relative to a specific application.

Keywords Moment descriptors · Pattern recognition · Feature selection · Genetic algorithms · Relief algorithm

13.1 Introduction

Nowadays, many advanced intelligent systems take part into humans' daily life helping them to satisfy possible professional or entertainment needs. Thus, advanced human computer/machine interaction [3], human identity authentication [10], biometric authentication [30] and surveillance systems [28], have been developed and proposed. Such systems mainly consist of a pattern recognition procedure, which enables the system to interact with the surrounding environment.

G.A. Papakostas (✉)
Human Machines Interaction (HMI) Laboratory, Department of Computer and Informatics Engineering, Eastern Macedonia and Thrace (EMT) Institute of Technology, Ag. Loukas 65404, Kavala, Greece
e-mail: gpapak@teikav.edu.gr

U. Stańczyk and L.C. Jain (eds.), *Feature Selection for Data and Pattern Recognition*, Studies in Computational Intelligence 584,
DOI 10.1007/978-3-662-45620-0_13

The successful operation of the pattern recognition procedure is mainly based on the representation method of the real patterns in a form suitable to be manipulated and managed by the recognition module (classifier). In the case of image based systems the content of an image pattern has to be transformed in a compact numerical format (or other) by applying a feature extraction method (FEM). The role of a FEM is twofold; it performs a dimensionality reduction from the space of image pixels to a small set of numbers and it captures the discriminative characteristics of the patterns in order to distinguish them.

A popular feature extraction method for the case of image patterns is the method of moments. Image moments have proved to be efficient descriptors of the images' content, with many applications in pattern recognition [2, 22, 25, 36], computer vision [12, 23], image analysis [33, 37], image watermarking [27] etc. Among the several moment types, the orthogonal moments [4] constitute the most prominent moment features (discrimination features based on moments) due to their minimum redundancy and high reconstruction capabilities. Additionally, their inherent properties staying invariant under common geometrical transformations (rotation, scaling, translation, flipping) or incorporating such invariances through coordinates transformation, give them all the desirable advantages for any invariant pattern recognition task.

However, a common drawback is the absence of a prior knowledge regarding the number and the suitability of the used moment features being controlled by adjusting the order of the orthogonal polynomial used as kernel function. A common practice is to compute all the moment features up to a certain order and to apply the entire set of moments as discriminative features. This is an "ad hoc" practice in some sense, since the significance of each moment component in discriminating the patterns of the application is not taken into account. A possible solution to this issue is the application of an additional process that selects, from a large pool, the moment features that best perform in terms of recognition accuracy.

The aforementioned issue, of the used moments' appropriateness, constitutes the main subject of this chapter. Initially, the main properties of some representative moment features and their representation capabilities are discussed in Sect. 13.2. Section 13.3 focuses on the justification of the need for selection of the moment features that better describe the distinctive characteristics of the patterns. The selection of moment features by applying two different types of selection algorithms, a Genetic Algorithm and the Relief algorithm, is presented in Sect. 13.4. An extensive experimental study with four benchmark pattern recognition datasets and selected moment features subsets aims to justify the initial assertions in Sect. 13.5. Finally, Sect. 13.6 summarizes and discusses the main conclusions.

13.2 Image Moment Features

Geometric moments were the first type of moments introduced in image analysis and pattern recognition [9]. These moments constitute the projection of the image on the monomial $x^n y^m$ of $(n + m)$th order. However, the geometric moments suffer from

high information redundancy making them less efficient in difficult problems where more discriminative information needs to be captured, due to monomials' lack of orthogonality increasing their information redundancy.

This fact has motivated scientists to develop the orthogonal moments, which use as kernel functions orthogonal polynomials that constitute orthogonal basis. The property of orthogonality gives to the corresponding moments the feature of minimum information redundancy, meaning that different moment orders describe different image content.

Initially, the orthogonal moments defined in the continuous space were introduced [26], such as Zernike, Pseudo-Zernike, Fourier-Mellin and Legendre moments. Although these moments were widely applied in many disciplines for a long time, their performance is degraded by several approximation errors [32] generated mainly due to coordinates normalization and space granulation procedures.

Recently, enhanced orthogonal moments free from approximation errors and directly defined inside the discrete coordinate system of the image, were proposed to overcome the disadvantages of the continuous moments. The most representative moment families of discrete form are the Tchebichef [19], Krawtchouk [34] and dual Hahn [14, 37] moments.

It is worth noting that the main research directions across which most scientists work with, in the field of image moments are the following: (1) the development of new algorithms that accelerate the overall moments' computation time, (2) the improvement of the moments' accuracy by reducing the quantization and approximation errors and (3) the embodiment of invariance capabilities into the moments' computation regarding the major linear image's transformations (translation, rotation and scaling). The last direction is relative to the capabilities of the moment features to achieve high recognition rates exploiting invariant behaviour under the aforementioned three basic transformations. Herein, only the description capabilities of the moment features in terms of recognition accuracy will be studied, without paying any attention to the invariant versions of them.

The most representative orthogonal moment families of both continuous and discrete coordinate space are hereafter described and analyzed experimentally.

13.2.1 Continuous Orthogonal Moments

In the previous section it has already been mentioned that the first type of orthogonal moments for images (2-D) was defined in the continuous coordinate space of a continuous intensity function $f(x, y)$. However, in order to use those moments with digital images, which are defined in the discrete domain, an approximation was applied the so-called zeroth order approximation (ZOA). These two different definitions are as follows:

$$M_{nm} = NF_1 \iint_A Kernel_{nm}(x, y) f(x, y)\, dxdy \tag{13.1}$$

where A is the computation coordinate space, $Kernel_{nm}(\cdot)$ corresponds to the moment's kernel (product of two polynomials) consisting of specific orthogonal polynomials of order n and m, which constitute the orthogonal basis and NF_1 is a normalization factor. The type of Kernel's polynomials gives the name to the moment family and thus a wide range of different moment types can be derived.

The zeroth order approximation of Eq. (13.1) for a $N \times N$ image having intensity function $f(x, y)$ has the form:

$$(ZOA) : M_{nm} = NF_2 \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} Kernel_{nm}(x, y) f(x, y) \tag{13.2}$$

where NF_2 is a normalization factor and the double integral of Eq. (13.1) is substituted by a double summation, by incorporating some approximation error. The minimization of this error has been the subject of many works [31], which try to propose a discrete computation form that converges to the theoretical values. Three representative moment families of this category will be described in details in the next sections.

13.2.1.1 Zernike Moments

Zernike moments are the most widely used family of orthogonal moments due to their inherent property of being invariant to an arbitrary rotation of the image they describe. The main characteristic of this moment family is the usage of a set of complex polynomials as basis, which forms a complete orthogonal set over the interior of the unit circle $x^2 + y^2 = 1$. These polynomials have the form:

$$V_{nm}(r, \theta) = R_{nm}(r) e^{jm\theta} \tag{13.3}$$

where n is a non-negative integer and m an integer subject to the constraints $n - |m|$ even and $|m| \leq n$, $r(r = \sqrt{x^2 + y^2})$ is the length of the vector from the origin to the pixel and θ $\left(\theta = tan^{-1}(y/x)\right)$ is the angle between the vector r and x-axis in counter-clockwise direction. $R_{nm}(r)$, are the Zernike radial polynomials [35], in (r, θ) polar coordinates defined as:

$$R_{nm}(r) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} r^{n-2s} \tag{13.4}$$

Note that $R_{n(-m)}(r) = R_{nm}(r)$.

The Zernike moment of order n with repetition m for a $N \times N$ pixels size continuous image function $f(x, y)$, that vanishes outside the unit disk, has the form:

$$Z_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 V_{nm}^*(r, \theta) f(r, \theta) r dr d\theta \tag{13.5}$$

where the symbol $(*)$ corresponds to conjugate.

For a digital image, the integrals are replaced according to the zeroth order approximation Eq. (13.2) by summations to get:

$$Z_{nm} = \frac{n+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} V_{nm}^*(r_{ij}, \theta_{ij}) f(i, j). \tag{13.6}$$

The above transformation from continuous to discrete form adds some approximation errors. For this reason, several attempts [32] to decrease these approximation errors have been reported in the literature. Moreover, significant work has been done [7] in the last years towards the fast computation of the radial polynomials (Eq. 13.4) and the moments (Eq. 13.6).

13.2.1.2 Legendre Moments

The $(n+m)$th order Legendre moment [6] of an intensity function $f(x, y)$, is defined in $[-1,1]$ as:

$$L_{nm} = \frac{(2n+1)(2m+1)}{4} \int_{-1}^{1} \int_{-1}^{1} P_n(x) P_m(y) f(x, y) dx dy \tag{13.7}$$

where $P_n(x)$ is the nth order Legendre polynomial defined as:

$$P_n(x) = \sum_{k=0}^{n} \alpha_{k,n} x^k = \frac{1}{2^n n!} \left(\frac{d}{dx} \right)^n (x^2 - 1)^n \tag{13.8}$$

The above Legendre polynomials satisfy the following recursive equation:

$$\begin{aligned} P_n(x) &= \left[(2n-1) x P_{n-1}(x) - (n-1) P_{n-2}(x) \right] / n \\ P_0(x) &= 1, \quad P_1(x) = x \end{aligned} \tag{13.9}$$

The recursive formula of Eq. (13.10) permits the fast computation of the Legendre polynomials by using polynomials of lower order. In case of computing the Legendre moments of a $N \times N$ image, Eq. (13.7) takes the following discrete form:

$$L_{nm} = \frac{(2n+1)(2m+1)}{(N-1)(N-1)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} P_n(x) P_m(y) f(x,y). \tag{13.10}$$

The computation of the Legendre moments through Eq. (13.10) shows significant approximation errors as discussed lately and the resulted Legendre moments do not satisfy the properties of the theoretical ones, by affecting their ability to describe the image in process. For this reason new algorithms ensuring the accurate computation of the moments have been proposed [6].

13.2.1.3 Gaussian-Hermite Moments

Gaussian-Hermite moments are continuous moments that have been introduced in image analysis quite recently by Yang and Dai [33]. The $(n+m)$th order Gaussian-Hermite moment is defined in $(-\infty, +\infty)$ and has the form:

$$GH_{nm} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{H}_n(x;\sigma) \hat{H}_m(y;\sigma) f(x,y) dx dy \tag{13.11}$$

where

$$\hat{H}_n(x;\sigma) = \frac{1}{\sqrt{2^n n! \sigma \sqrt{\pi}}} e^{(-x^2/2\sigma^2)} H_n(x;\sigma) \tag{13.12}$$

is the weighted Hermite orthonormal polynomial of order n, derived by the ordinary Hermite polynomial $H_n(x;\sigma)$, modulated by a Gaussian function with σ variance. The ordinary Hermite polynomial of order n is defined as:

$$H_n(x) = n! \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k}{k!(n-2k)!} (2x)^{n-2k} \tag{13.13}$$

The recursive computation of the above Hermite polynomials is performed according to:

$$\begin{aligned} H_{n+1}(x) &= 2x\, H_n(x) - 2n\, H_{n-1}(x), \quad \text{for } n \geq 1 \\ H_0(x) &= 1, \quad H_1(x) = 2x \end{aligned} \tag{13.14}$$

The Gaussian-Hermite moments have proved to be of higher image representation ability [33], compared to some traditional moment families e.g. Legendre and thus their usage has been rapidly increased in many applications of the engineering life.

Due to this popularity, a faster and more accurate computation algorithm has been proposed recently [8].

13.2.2 Discrete Orthogonal Moments

The aforementioned drawback of the continuous orthogonal moments, has motivated scientists in the field of image moments to develop more accurate moment families. This goal has been achieved by the introduction of the discrete orthogonal moments being defined directly on the discrete image coordinates space. Some of the most representative discrete moment families are discussed herein.

13.2.2.1 Tchebichef Moments

This moment family is the first proposed in the literature by Mukundan et al. [19]. The $(n+m)$th order Tchebichef moment of a $N \times N$ image having intensity function $f(x, y)$ is defined as:

$$T_{nm} = \frac{1}{\tilde{\rho}(n, N)\tilde{\rho}(m, N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \tilde{t}_n(x)\tilde{t}_m(y) f(x, y) \tag{13.15}$$

where $\tilde{t}_n(x)$ is the nth order normalized Tchebichef polynomial, introduced in order to ensure numerical stability and moments' limited dynamical range, defined as follows:

$$\tilde{t}_n(x) = \frac{t_n(x)}{\beta(n, N)} \tag{13.16}$$

where the ordinary Tchebichef polynomial of n order has the form:

$$\begin{aligned} t_n(x) &= (1-N)_n \, {}_3F_2(-n, -x, 1+n; 1, 1-N; 1) \\ &= \sum_{k=0}^{n} (-1)^{n-k} \binom{N-1-k}{n-k} \binom{n+k}{n} \binom{x}{k}. \end{aligned} \tag{13.17}$$

In the above formulas, ${}_3F_2$, is the generalized hypergeometric function, $n, x = 0, 1, 2, \ldots, N-1, N$ is the image size and $\beta(n, N)$ is a suitable constant independent of x that serves as a scaling factor, such as N^n. Moreover, $\tilde{\rho}(n, N)$ is the normalized norm of the Tchebichef polynomials defined as:

$$\tilde{\rho}(n, N) = \frac{\rho(n, N)}{\beta(n, N)^2} \tag{13.18}$$

with

$$\rho(n, N) = (2n)! \binom{N+n}{2n+1}, \quad n = 0, 1, \ldots, N-1. \tag{13.19}$$

The computation speed of Tchebichef moments can be accelerated by using the following recursive formula:

$$\begin{aligned} n\tilde{t}_n(x) &= (2n-1)\,\tilde{t}_{n-1}(x) - (n-1)\left(1 - \tfrac{(n-1)^2}{N^2}\right)\tilde{t}_{n-2}(x) \\ \tilde{t}_0(x) &= 1, \quad \tilde{t}_1(x) = \tfrac{2x+1-N}{N} \end{aligned} \tag{13.20}$$

Tchebichef moments have proved to be superior to Zernike and Legendre moments in describing objects, while their robustness in the presence of high noise levels makes them appropriate to real-time pattern classification and computer vision applications.

13.2.2.2 Krawtchouk Moments

The Krawtchouk orthogonal moments are the second type of discrete moments introduced in image analysis by Yap et al. [34]. The $(n+m)$th order orthogonal discrete Krawtchouk moment of a $N \times N$ image having intensity function $f(x, y)$ is defined as:

$$K_{nm} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} \overline{K}_n(x; p_1, N-1)\,\overline{K}_m(y; p_2, N-1) f(x, y) \tag{13.21}$$

where

$$\overline{K}_n(x; p, N) = K_n(x; p, N)\sqrt{\frac{w(x; p, N)}{\rho(n; p, N)}} \tag{13.22}$$

is the weighted Krawtchouk polynomial of n order, used to reduce the numerical fluctuations presented in the ordinary Krawtchouk polynomials, defined as:

$$K_n(x; p, N) = {}_2F_1\left(-n, -x; -N; \frac{1}{p}\right) = \sum_{k=0}^{N} \alpha_{k,n,p} x^k. \tag{13.23}$$

In Eq. (13.22) $\rho(n; p, N)$ is the norm of the Krawtchouk polynomials having the form:

$$\rho(n; p, N) = (-1)^n \left(\frac{1-p}{p}\right)^n \frac{n!}{(-N)_n}, \; n = 1, \ldots, N \tag{13.24}$$

and $w(x; p, N)$ is the weight function of the Krawtchouk moments,

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x} \tag{13.25}$$

In Eq. (13.24) the symbol $(\cdot)_n$ is the Pochhammer symbol, which for the general case is defined as $(\alpha)_k = \alpha(\alpha+1)\ldots(\alpha+k+1)$.

In practice, the computation of the weighted Krawtchouk polynomials is not performed through Eq. (13.22), since this is a very time consuming procedure; instead, a recursive algorithm [34] is applied:

$$p\,(N-n)\,\overline{K}_{n+1}\,(x;p,N) = A\,(Np-2np+n-x)\,\overline{K}_n\,(x;p,N) - Bn\,(1-p)\,\overline{K}_{n-1}\,(x;p,N) \tag{13.26}$$

where

$$A = \sqrt{\frac{p\,(N-n)}{(1-p)\,(n+1)}}, \quad B = \sqrt{\frac{p^2\,(N-n)\,(N-n+1)}{(1-p)^2\,(n+1)\,n}} \tag{13.27}$$

and

$$\overline{K}_0\,(x;p,N) = \sqrt{\frac{w\,(x;p,N)}{\rho\,(0;p,N)}}, \quad \overline{K}_1\,(x;p,N) = \left(1-\frac{x}{pN}\right)\sqrt{\frac{w\,(x;p,N)}{\rho\,(1;p,N)}} \tag{13.28}$$

The Krawtchouk moments proved to be effective local descriptors, since they can describe the local features of an image, unlike the other moment families, which capture only the global features of the objects they describe. This locality property is controlled by appropriate adjustment of the p_1,p_2 parameters of Eq. (13.21).

13.2.2.3 Dual Hahn Moments

The $(n+m)$th order orthogonal dual Hahn moment [37] of a $N \times N$ image having intensity function $f\,(x,y)$ is defined as:

$$W_{nm} = \sum_{x=a}^{b-1}\sum_{y=a}^{b-1} \hat{W}_n^{(c)}\,(x,a,b)\,\hat{W}_m^{(c)}\,(y,a,b) f\,(x,y)\,, \quad n,m = 0,1,\ldots,N-1 \tag{13.29}$$

where $-\frac{1}{2} < a < b,\ |c| < 1+a,\ b = a+N$ and

$$\hat{W}_n^{(c)}(s,a,b) = W_n^{(c)}(s,a,b)\sqrt{\frac{\rho(s)}{d_n^2}\left[\Delta x\left(s-\frac{1}{2}\right)\right]} \tag{13.30}$$

is the nth order weighted dual Hahn polynomial used to reduce the numerical instabilities caused by the ordinary dual Hahn polynomials, defined as:

$$W_n^{(c)}(s,a,b) = \frac{(a-b+1)_n\,(a+c+1)_n}{n!}$$
$$_3F_2\left(-n, a-s, a+s+1; a-b+1, a+c+1; 1\right) \tag{13.31}$$

for $n = 0, 1, \ldots, N-1$, $s = a, a+1, \ldots, b-1$, where $_3F_2$ is the generalized hypergeometric function given by:

$$_3F_2\left(a_1, a_2, a_3; b_1, b_2; z\right) = \sum_{k=0}^{\infty} \frac{(a_1)_k\,(a_2)_k\,(a_3)_k}{(b_1)_k\,(b_2)_k} \frac{z^k}{k!} \tag{13.32}$$

In the above formulas $\rho(s)$ is the weighting function defined in terms of the gamma function $\Gamma(\cdot)$ as:

$$\rho(s) = \frac{\Gamma(a+s+1)\,\Gamma(c+s+1)}{\Gamma(s-a+1)\,\Gamma(b-s)\,\Gamma(b+s+1)\,\Gamma(s-c+1)} \tag{13.33}$$

and

$$d_n^2 = \frac{\Gamma(a+c+n+1)}{n!\,(b-a-n-1)!\Gamma(b-c-n)},\ n = 0, 1, \ldots, N-1. \tag{13.34}$$

It is obvious from the above equations that the computation of dual Hahn polynomials is a time consuming task, so efficient recursive algorithms need to be used [37].

13.2.3 Image Reconstruction by the Method of Moments

A significant and useful property of the orthogonal moments is their ability to reconstruct the image they describe. The reconstruction of a $N \times N$ image by using moment orders up to n_{max} and m_{max} is described by the following inverse formula of Eq. (13.2):

$$\hat{f}(x,y) = \sum_{n=0}^{n_{max}} \sum_{m=0}^{m_{max}} \overline{Kernel_{nm}}(x,y)\,M_{nm} \tag{13.35}$$

where M_{nm} is the $(n+m)$th order moment and $\overline{Kernel_{nm}}(x,y)$ the kernel function of the used polynomial family, which is not always the same as Eq. (13.2), e.g. for the case of the Zernike moments it is equal to the conjugate of $Kernel_{nm}(x,y)$.

It is worth to note that the reconstruction ability of each moment family is an indication of its information compactness, which is highly connected with the amount of image's information enclosed by the moment coefficients. An extensive analysis of the reconstruction performance of the orthogonal moments has been reported in [21].

13.2.4 Moments Interpretation

The conventional characterization of the orthogonal image moments defines them as the projections of an image to the orthogonal basis of the used polynomials. However, this definition encloses more mathematical than engineering or computer science oriented knowledge. According to a different perspective from an engineering and computer science point of view, the orthogonal moments represent the similarity between the image and a number of image patterns formed by the kernel function of each moment family.

In order to better understand the latter proposition, the image patterns derived from the kernels of the pre-analyzed moment families need to be calculated. These image patterns are called *basis images* and they are computed by applying the following formula:

$$\Phi_{nm} = \left[\mathbf{Poly}_n\right]^T \mathbf{Poly}_m \tag{13.36}$$

where $\mathbf{Poly}_n$ and $\mathbf{Poly}_m$ are vectors with the values of the nth and mth order polynomial for each image pixel. The computed basis images for the case of an 8×8 square image having constant intensity equal to 1, for the cases of the continuous orthogonal moments (up to order 7) are depicted in Figs. 13.1, 13.2 and 13.3.

By examining the basis images of Figs. 13.1, 13.2 and 13.3 it is noticeable that the patterns provided by each moment type are totally different. The Zernike moments (Fig. 13.1) generate circular patterns due to the used radial polynomials Eq. (13.2), while some images do not exist due to the constraints hold between the n and m parameters.

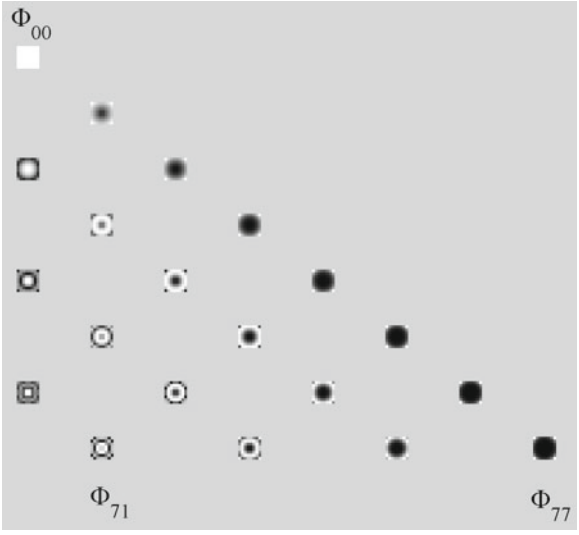

Fig. 13.1 Basis images of the continuous orthogonal Zernike moments

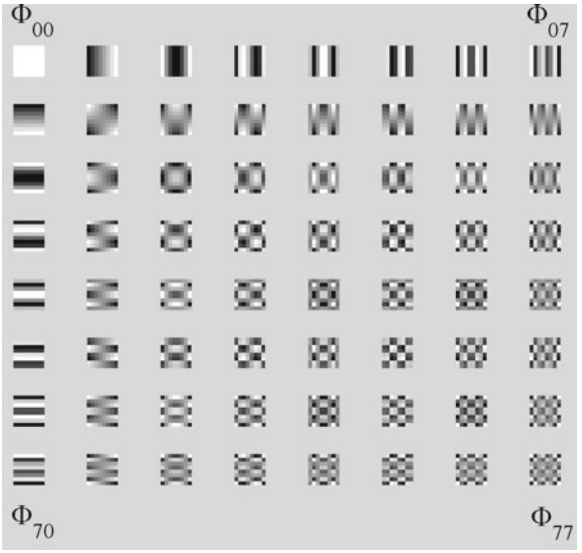

Fig. 13.2 Basis images of the continuous orthogonal Legendre moments

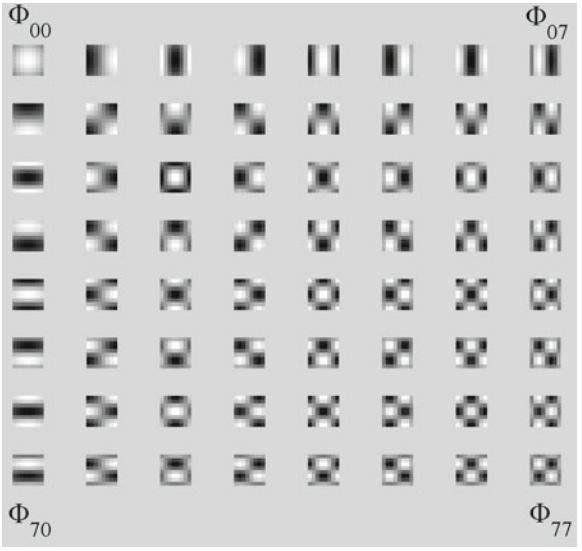

Fig. 13.3 Basis images of the continuous orthogonal Gaussian-Hermite ($\sigma = 1$) moments

Concerning the basis images of the rest continuous moments it can be noted that the formed patterns include more details as the moment order increases, while for low order these patterns are coarse.This observation deals with what is commonly

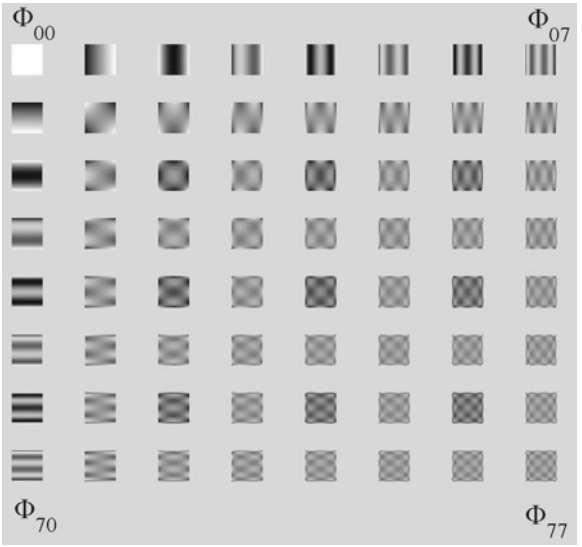

Fig. 13.4 Basis images of the discrete orthogonal Tchebichef moments

stated regarding the description capabilities of different moment orders. The corresponding basis images for the case of the discrete orthogonal moments are illustrated in Figs. 13.4, 13.5 and 13.6.

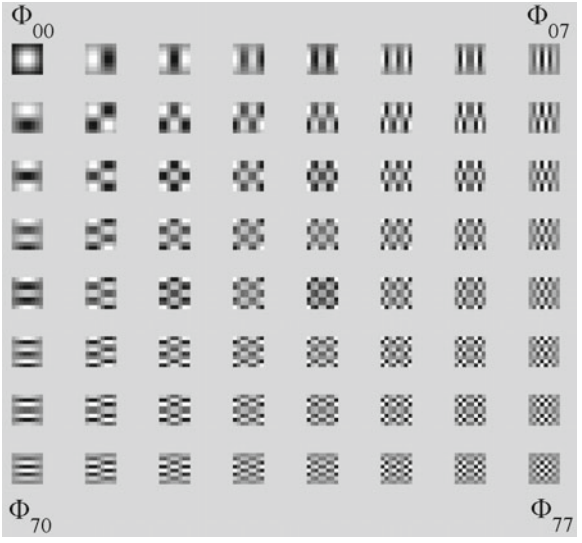

Fig. 13.5 Basis images of the discrete orthogonal Krawtchouk ($p_1 = p_2 = 0.5$) moments

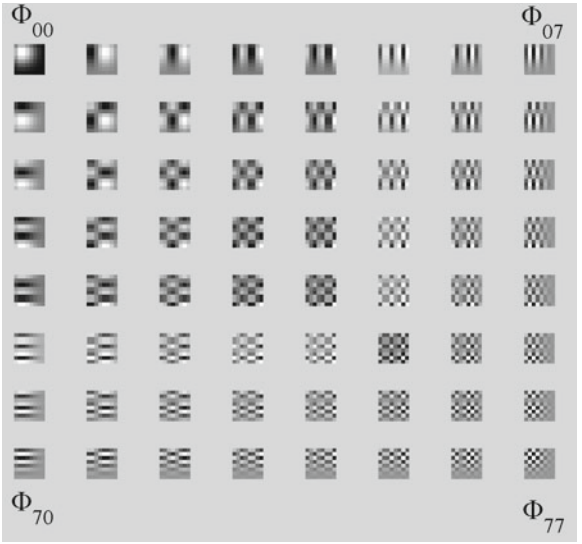

Fig. 13.6 Basis images of the discrete orthogonal dual Hahn ($a = c = 0, b = 8$) moments

The local behaviour of the Krawtchouk and dual Hahn moments is obvious from the above basis images of the discrete orthogonal moments. This constitutes the most advantageous property of these moments, since they localize the window of interest in a specific image portion. The previous analysis considering the basis images of the orthogonal moments highlights the different representation capabilities of each moment type for different moment orders. According to this study each moment carries different image's content and therefore the selection of the moments that better discriminate some patterns seems to be a reasonable and inevitable process.

13.3 Is There a Need for Moments' Selection?

Apart from the previous analysis of moments' representation capabilities, the execution of a certain experiment regarding moments' description ability, would be constructive to further highlight the need for selection of the most appropriate moment sets.

For this purpose the well known Lena benchmark image with 64 $\times$ 64 pixels size in grayscale format, is reconstructed with various sets of Tchebichef moments having different orders. The reconstruction results for moment sets consisting of 10 *different* orders are depicted in Fig. 13.7. It has to be noted that in each reconstructed image the intensities are normalized into the [0, 255] range for illustration purposes.

From the reconstructed images of Fig. 13.7, it can be deduced that as the moment order increases more detailed information of the image's content emerges.

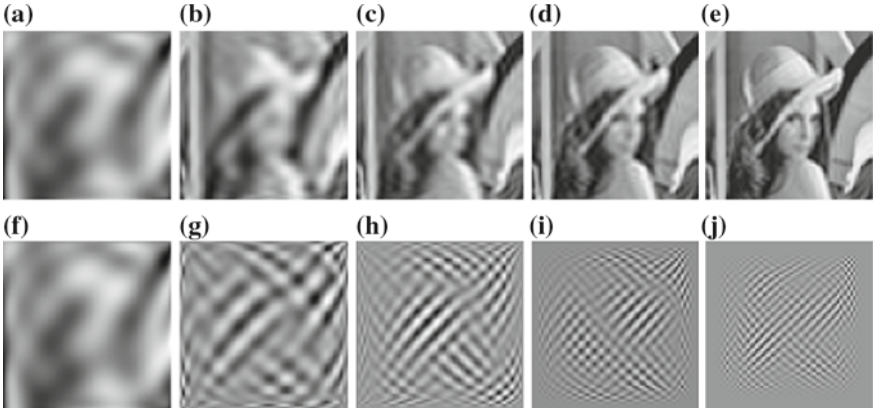

Fig. 13.7 Lena image reconstruction using several sets of Tchebichef moments of various orders: **a** 0–9, **b** 0–19, **c** 0–29, **d** 0–39, **e** 0–49, **f** 0–9, **g** 10–19, **h** 20–29, **i** 30–39 and **j** 40–49

For example, the orders' range 0–9 (1st row of Fig. 13.7) is able to reconstruct a quite coarse image's content, while by adding the next 10 orders (0–19) some detailed information is incorporated. This observation is in agreement with the image's content described by each moment set (2nd row of Fig. 13.7), where it is obvious that the higher order moments sets model the high frequency pixels variations.

Considering the above analysis, the moments of low orders are not so useful in discriminating patterns which differ slightly, since their differences are described in the high moment orders. For example, if a second image of the above Lena benchmark is constructed with Lena having her eyes closed, the two image patterns could not be discriminated by the low order moments but high orders are needed.

Therefore, it is evident that the appropriate set of moments, better discriminating some specified patterns, depends on the application and thus a selection procedure considering patterns' modalities is inevitable.

13.4 Moment Features Selection

By examining the recent literature in the field of image moments, one can reach the conclusion that little work has been done towards the moments' selection [13, 20].

The main selection method applied to all the aforementioned works is the Genetic Algorithm (GA), proved as an efficient wrapper selection technique [24] taking into account the classification model applied to recognize the patterns. Genetic Algorithms are a great example of evolutionary computation mimicking the evolutionary process that characterizes the evolution of living organisms [5]. However, the main disadvantage of the GA-based selection is the high computation time need to converge the algorithm to a suitable solution.

The aforementioned drawback of the GA-based selection makes the filter selection methods [24] an attractive alternative approach. These methods do not use the mining model (they are independent of the classification model), instead the internal data properties/characteristics (dependency, correlation etc.) are taken into consideration.

For the sake of the experiments presented in the next section, the GA-based (wrapper) and the Relief [15] algorithm (filter) selection methods will be applied for the selection of the proper moment subsets that better discriminate the patterns of some benchmark pattern recognition datasets. These two algorithms are briefly discussed in the next sections.

13.4.1 GA-Based Selection

The main operational element of a Genetic Algorithm is the chromosome. The chromosome corresponds to a candidate solution to the problem at hand, consisting of the set of variables appropriately coded. For the case of the GA-based moment selection method, the chromosome consists of the indices (Fig. 13.8) of a predefined number of moments. The indices correspond to the moment id belonging to the moment feature vector, which is constructed by arranging the computing moments according to the zigzag scanning operation [11].

Initially, a pool of 100 moment features is constructed by computing all the moments up to a specific order. Considering that a number of n moments are required to be selected, the kth chromosome structure of the GA is depicted in Fig. 13.8.

Furthermore, the objective function being minimized by the GA is equal to the recognition error (Wrong Recognized Patterns/Total Patterns) derived when the selected moment sets are fed to the classifier model.

13.4.2 Relief Algorithm

Relief algorithm [15] is a popular feature selection method due to its simplicity. It is based on the computation of the relevance between pairs of feature vectors. The relevance is measured by applying the L-dimensional Euclidean distance, where L is the dimension of the feature vectors being compared. This algorithm selects those features which are relevant subject to a defined threshold in linear time.

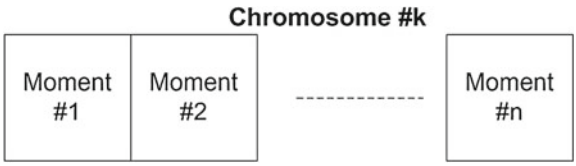

Fig. 13.8 Chromosome structure

13.5 Experimental Study

A set of experiments were conducted in order to study the impact of the selected moment subsets on the overall recognition performance in several pattern recognition problems. For this purpose, appropriate software was developed in the MATLAB 2012b environment, while all experiments were executed in an Intel i5 3.3 GHz PC with 8 GB RAM. Moreover, four well known benchmark datasets were used to evaluate the initial assertion of the moments' selection significance, towards the improvement of the recognition accuracy.

13.5.1 Benchmark Datasets

Four benchmark datasets are used in order to investigate the selection performance of the Relief and Genetic Algorithms in selecting moment subsets of various sizes. The considered datasets are the Yale face recognition dataset [1], a subset [22] of the Terravic thermal infrared face recognition [18], the JAFFE [17] and RADBOUD [16] facial expressions datasets. Three sample images (different classes) from each dataset are depicted in Fig. 13.9, while the properties of each dataset are summarized in Table 13.1.

13.5.2 Datasets Pre-processing

It is worth noting that before computing the moment features, the images need to be pre-processed in order to remove irrelevant image information (background, hair, ears, etc.) and to isolate the image's part, which includes the main face information. For this purpose, the Viola-Jones face detector [29] is applied being followed by an ellipse masking [11], for the case of Yale, JAFFE and Radboud datasets. The aforementioned face detector fails to detect faces in thermal infrared images and

Table 13.1 Benchmark datasets properties

Dataset	Type	Number of classes	Samples/class	Total samples
Yale	Face recognition	15	11	165
Terravic	Thermal infrared face recognition	10	70	700
JAFFE	Facial expression recognition	7	30, 29, 32, 31, 30, 31, 30	213
Radboud	–	8	67	536

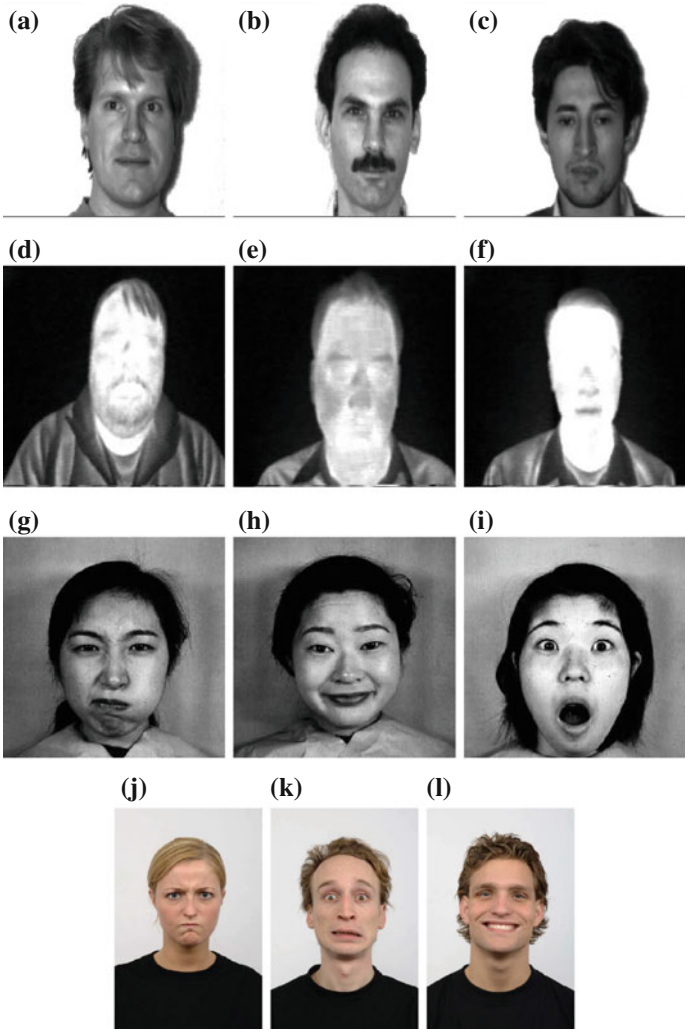

Fig. 13.9 Samples of the four benchmarks: **a–c** Yale, **d–f** Terravic, **g–i** JAFFE and **j–l** RADBOUD

thus the procedure proposed in [22] is applied for the Terravic dataset. The outcome of this processing stage is a cropped image of 128 $\times$ 128 pixels size, which includes the most relative information to the face image content.

13.5.3 Simulations

A large scale experimental study has taken place in order to extract useful conclusions towards the improvement of the moment features' recognition performance through the application of a selection mechanism. The Genetic Algorithm settings are: population size 50, maximum generations 30, crossover with probability 0.8 and 2 points, mutation probability 0.01 and Stochastic Universal Approximation (SUS) selection method. The k-NN classifier ($k = 1$) is selected as the prediction model in the case of the GA-based selection. Moreover, a 10-fold cross-validation technique is applied in all datasets, while the moments are selected from a pool of the first 100 (up to 9 order for all moments, except Zernike computed up to 18 order) computed moments for each moment family.

The corresponding mean recognition rates for each dataset are summarized in Tables 13.2, 13.3, 13.4 and 13.5. The best moment family along with the best selection method is presented in these results.

By examining Tables 13.2, 13.3, 13.4 and 13.5 it is deduced that in almost all cases the selected moment features show better or equal, in the worst case, performance than the non selected (NoSel.) moments. This outcome enforces the initial assertion regarding the needs for moment features selection. Moreover, among the two examined selection methods, the GA-based one seems to be more efficient for small sized moment subsets (up to 25–30), while the Relief algorithm is superior for larger subsets (greater than 30–40). This observation can be justified by the fact that for large moment subsets (greater than 30) the optimization problem, which needs to be solved by the GA, is quite difficult. One solution to this limitation is to use more advanced versions of the algorithm, where adaptive crossover and/or mutation operators could guide the algorithm to more optimum solutions.

Table 13.2 Recognition performance of moment features subsets for the Yale dataset

Yale dataset			
Number of moments	Moment type	Recognition rate (%)	Selection method
5	Krawtchouk	76.00	GA
10	Krawtchouk	86.66	GA
15	Zernike	88.00	GA
20	Zernike	87.33	GA
25	Dual Hahn	84.66	GA
30	Krawtchouk	72.66	Relief
40	Krawtchouk	75.33	Relief
50	Krawtchouk	74.66	Relief
60	Krawtchouk	76.66	Relief
70	Dual Hahn	76.00	Relief

Table 13.3 Recognition performance of moment features subsets for the Terravic dataset

Terravic dataset			
Number of moments	Moment type	Recognition rate (%)	Selection method
5	All	100.00	NoSel./GA
10	All	100.00	NoSel./GA
15	All	100.00	NoSel./GA
20	All	100.00	NoSel./GA
25	All	100.00	NoSel./GA
30	All	100.00	NoSel./Relief
40	All	100.00	NoSel./Relief
50	All	100.00	NoSel./Relief
60	All	100.00	NoSel./Relief
70	All	100.00	NoSel./Relief

Table 13.4 Recognition performance of moment features subsets for the JAFFE dataset

JAFFE Dataset			
Number of moments	Moment type	Recognition rate (%)	Selection method
5	Legendre	71.66	GA
10	Krawtchouk	79.90	GA
15	Legendre	78.90	GA
20	Krawtchouk	77.47	GA
25	Krawtchouk	69.33	GA
30	Krawtchouk	53.92	GA
40	Dual Hahn	47.35	Relief
50	Zernike	46.88	Relief
60	Gaussian-Hermite	46.80	Relief
70	Legendre	46.88	NoSel.

From the above results, it can be observed that the increase of the number of moments used to discriminate the patterns does not always improve the recognition accuracy. In almost all the cases a subset of 10–25 moment features is able to achieve the highest recognition rate.

In order to draw a conclusion regarding the optimal settings, ensuring the best solution to each dataset, the most effective configuration in each case is summarized in Table 13.6.

The results of Table 13.6 show again the outperformance of the GA-based selection method over the Relief one, while its recognition accuracy is in agreement with the state of the art methods [11, 22, 25]. As far as the performance of the moment families is concerned, it is obvious that Zernike moments are the most efficient family, while a moments' subset of size lower than 25 is adequate to ensure an acceptable recognition

Table 13.5 Recognition performance of moment features subsets for the Radboud dataset

Radboud dataset			
Number of moments	Moment type	Recognition rate (%)	Selection method
5	Zernike	48.21	GA
10	Zernike	55.75	GA
15	Zernike	61.96	GA
20	Zernike	61.38	GA
25	Zernike	62.14	GA
30	Zernike	54.10	GA
40	Zernike	53.30	Relief
50	Zernike	53.66	Relief
60	Zernike	53.30	Relief
70	Zernike	51.47	No Sel./Relief

Table 13.6 Best configuration for each dataset

Dataset	Number of moments	Moment type	Recognition rate (%)	Selection method
Yale	15	Zernike	88.00	GA
Terravic	5	All	100.00	No Sel./GA
JAFFE	10	Krawtchouk	79.90	GA
Radboud	25	Zernike	62.14	GA

accuracy. However, Krawtchouk moments show a significant performance leading to the conclusion that the locality property plays an important role to capture the local characteristics of the patterns. More work has to be done in this direction of describing the local information by the method of moments.

13.6 Conclusions

A detailed discussion of the main properties of the most representative image orthogonal moment families was presented in the previous sections. Through an in depth analysis of the representation capabilities of the orthogonal moments, the need for selection of moment features for improved recognition accuracy is highlighted. Finally, an extensive experimental study on well known benchmark datasets has resulted in useful conclusions regarding the initial assertion of moment's selection and the description capability of each moment family. The GA-based selection method has shown superior performance to the Relief algorithm, mainly for low number of moments, while for high number of features the latter algorithm seems to be the suitable choice.

Moreover, the Zernike moments seem to be the most discriminant moment family compared to other moments, since they recognized the patterns of the three datasets with the highest rate. Also, one additional outcome of the experiments was the outperformance of the Krawtchouk moments to the JAFFE dataset. This result set the basis of a future study regarding the selection of moments belonging to different moment families, in order to take advantage of the properties each family presents.

Conclusively, an initial claim was set and proved both theoretically and experimentally, by establishing the selection of moment features as a mandatory processing step of any modern pattern recognition system.

References

1. Belhumeur, P.N., Kriegman, D.J.: The Yale face database. http://cvc.yale.edu/projects/yalefaces/yalefaces.html (1997)
2. Chen, B.J., Shu, H.Z., Zhang, H., Chen, G., Toumoulin, C., Dillenseger, J.L., Luo, L.M.: Quaternion Zernike moments and their invariants for color image analysis and object recognition. Signal Process. **92**(2), 308–318 (2012)
3. Cipolla, R., Pentland, A.: Computer vision for human-machine interaction. Cambridge University Press, Cambridge (1998)
4. Flusser, J., Zitova, B., Suk, T.: Moments and Moment Invariants in Pattern Recognition. Wiley, Chichester (2009)
5. Holland, J.H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. U Michigan Press, Ann Arbor (1975)
6. Hosny, K.M.: Exact Legendre moment computation for gray level images. Pattern Recognit. **40**(12), 3597–3605 (2007)
7. Hosny, K.M.: Fast computation of accurate Zernike moments. J. Real-Time Image Process. **3**(1–2), 97–107 (2008)
8. Hosny, K.M.: Fast computation of accurate Gaussian-Hermite moments for image processing applications. Digit. Signal Process. **22**(3), 476–485 (2012)
9. Hu, M.K.: Visual pattern recognition by moment invariants. IRE Trans. Inf. Theory **8**(2), 179–187 (1962)
10. Jain, A.K., Hong, L., Pankanti, S., Bolle, R.: An identity-authentication system using fingerprints. Proc. IEEE **85**(9), 1365–1388 (1997)
11. Kaburlasos, V.G., Papadakis, S.E., Papakostas, G.A.: Lattice computing extension of the FAM neural classifier for human facial expression recognition. IEEE Trans. Neural Netw. Learn. Syst. **24**(10), 1526–1538 (2013)
12. Kadir, A., Nugroho, L.E., Santosa, P.I.: Experiments of Zernike moments for leaf identification. J. Theor. Appl. Inf. Technol. **41**(1), 82–93 (2012)
13. Kanan, H.R., Faez, K.: GA-based optimal selection of PZMI features for face recognition. Appl. Math. Comput. **205**(2), 706–715 (2008)
14. Karakasis, E.G., Papakostas, G.A., Koulouriotis, D.E., Tourassis, V.D.: Generalized dual Hahn moment invariants. Pattern Recognit. **46**(7), 1998–2014 (2013)
15. Kira, K., Rendell, L.A.: A practical approach to feature selection. In: Proceedings of the 9th International Workshop on Machine Learning, pp. 249–256. Morgan Kaufmann Publishers Inc. (1992)
16. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., van Knippenberg, A.: Presentation and validation of the Radboud faces database. Cogn. Emot. **24**(8), 1377–1388 (2010)

17. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205. IEEE (1998)
18. Miezianko, R.: Terravic research infrared database. In: IEEE OTCBVS WS Series Bench. IEEE (2005)
19. Mukundan, R., Ong, S.H., Lee, P.A.: Image analysis by Tchebichef moments. IEEE Trans. Image Process. **10**(9), 1357–1364 (2001)
20. Papakostas, G.A., Boutalis, Y.S., Mertzios, B.G.: Evolutionary selection of Zernike moment sets in image processing. In: Proceedings of the 10th International Workshop on Systems, Signals and Image Processing (IWSSIP'03), pp. 10–11 (2003)
21. Papakostas, G.A., Karakasis, E.G., Koulouriotis, D.E.: Orthogonal image moment invariants: highly discriminative features for pattern recognition applications. In: Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies, pp. 34–52. IGI Global (2012)
22. Papakostas, G.A., Kaburlasos, V.G., Pachidis, T.: Thermal infrared face recognition based on lattice computing (LC) techniques. In: 2013 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 1–6. IEEE (2013)
23. Papakostas, G.A., Boutalis, Y.S., Karras, D.A., Mertzios, B.G.: A new class of Zernike moments for computer vision applications. Inf. Sci. **177**(13), 2802–2819 (2007)
24. Papakostas, G.A., Koulouriotis, D.E., Polydoros, A.S., Tourassis, V.D.: Evolutionary feature subset selection for pattern recognition applications. In: Evolutionary Algorithms, pp. 443–458. InTech (2011)
25. Papakostas, G.A., Koulouriotis, D.E., Karakasis, E.G., Tourassis, V.D.: Moment-based local binary patterns: a novel descriptor for invariant pattern recognition applications. Neurocomputing **99**(1), 358–371 (2013)
26. Teague, M.R.: Image analysis via the general theory of moments. JOSA **70**(8), 920–930 (1980)
27. Tsougenis, E.D., Papakostas, G.A., Koulouriotis, D.E., Tourassis, V.D.: Performance evaluation of moment-based watermarking methods: a review. J. Syst. Softw. **85**(8), 1864–1884 (2012)
28. Velasin, S.A., Remagnino, P.: Intelligent Distributed Video Surveillance Systems, vol. 5. IET, London (2006)
29. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
30. Wayman, J., Jain, A., Maltoni, D., Maio, D.: An introduction to biometric authentication systems. Biometric Systems, pp. 1–20. Springer, London (2005)
31. Wee, C.Y., Paramesran, R.: On the computational aspects of Zernike moments. Image Vis. Comput. **25**(6), 967–980 (2007)
32. Xin, Y., Pawlak, M., Liao, S.: Accurate computation of Zernike moments in polar coordinates. IEEE Trans. Image Process. **16**(2), 581–587 (2007)
33. Yang, B., Dai, M.: Image analysis by Gaussian-Hermite moments. Signal Process. **91**(10), 2290–2303 (2011)
34. Yap, P.T., Paramesran, R., Ong, S.H.: Image analysis by Krawtchouk moments. IEEE Trans. Image Process. **12**(11), 1367–1377 (2003)
35. Zernike, V.F.: Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. Physica **1**(7), 689–704 (1934)
36. Zhang, F., Liu, S.Q., Wang, D.B., Guan, W.: Aircraft recognition in infrared image using wavelet moment invariants. Image Vis. Comput. **27**(4), 313–318 (2009)
37. Zhu, H., Shu, H., Zhou, J., Luo, L., Coatrieux, J.L.: Image analysis by discrete orthogonal dual Hahn moments. Pattern Recognit. Lett. **28**(13), 1688–1704 (2007)