Moment-Based Local Binary Patterns: A Novel Descriptor for Invariant Pattern Recognition Applications


Democritus University of Thrace, Department of Production Engineering and Management, 67100 Xanthi, Greece
e-mail: gpapakos@ee.duth.gr, jimk@pme.duth.gr, ekarakas@pme.duth.gr, vtourasi@pme.duth.gr

Abstract

A novel descriptor able to improve the classification capabilities of a typical pattern recognition system is proposed in this paper. The introduced descriptor is derived by incorporating two efficient region descriptors, namely image moments and Local Binary Patterns (LBP), commonly used in pattern recognition applications, in the last decades. The main idea behind this novel feature extraction methodology is the need of improved recognition capabilities, a goal achieved by the combinative uses of these descriptors. This collaboration aims to make use of the major advantages each one presents, by simultaneously complementing each other, in order to elevate their weak points. In this way, the useful properties of the moments and moment invariants regarding their robustness to the noise presence, their global information coding mechanism and their invariant behaviour under scaling, translation and rotation conditions, along with the local nature of the LBP, are combined in a single concrete methodology. As a result a novel descriptor invariant to common geometric transformations of the described object, capable to encode its local characteristics, is formed and its classification capabilities are investigated through massive experimental scenarios. The experiments have shown the superiority of the introduced
1. Introduction

A crucial part of a modern intelligent imaging system, which learns from its environment and interacts with it, is the pattern recognition procedure. In general, a pattern recognition process employs four stages: 1) image acquisition, 2) image pre-processing (denoising, filtering, etc.), 3) feature extraction and finally 4) classification. The third step is probably the most complicated and it affects the overall performance of the system. A feature extraction method (FEM) can be termed successful if the resulted features (descriptors) describe uniquely the processed object in a scene. The more successful a FEM is, the more efficient the classification is.

The discrimination power of a descriptor is measured by its ability to capture the particular characteristics of the described pattern, which distinguish it among similar or totally different objects. A difficult pattern recognition task consists of objects being quite similar and differing slightly. In these cases the descriptors need to have strong local nature in order to encode the information that distinguishes them. A considerable performance evaluation of some well-known local descriptors has been performed in [1], with very useful and constructive conclusions.

Among the most widely used local descriptors, the Local Binary Patterns (LBP) operator proposed by Ojala et al. [2], has attracted the attention of the scientists and motivated them to extent its applicability to many disciplines. The LBP operator is initially introduced as texture descriptor but it has been applied, after some
modifications, to face recognition [3,4], facial expression [5], pedestrian detection [6], etc. Although there is a specific LBP version being rotation invariant [7,8], its application in traditional pattern recognition tasks where rotated, translated and scaled objects have to be recognized, is not suggested.

Another popular FEM that is used to generate discriminative feature sets is the computation of image moments. Moments have been used successfully in many classification applications [9-12] and their ability to describe an object fully makes them a powerful tool in computer vision applications, like object recognition in robotic applications and object characterization in visual inspection based quality control systems. However, since image moments are region descriptors they provide global information of an object. Although moments of higher orders capture the object’s details, their local behaviour is quite poor.

In this work a novel FEM that constructs rotation, translation, scale invariant descriptors of local nature, is introduced. The proposed feature extraction methodology aims to utilize the local behaviour of the LBP and the invariance nature of the orthogonal image moments. The derived descriptors called Moment Based Local Binary Patterns (Mb-LBP) seem to be effective pattern features, which improve the classification ability of the traditional moment invariants and extend the performance of the original LBP in more general pattern recognition problems.

In order to investigate the behaviour of the proposed descriptors, an exhaustive experimental plan has been arranged, consisting of several pattern recognition tasks (face recognition, facial expression recognition, texture recognition, object recognition) by using several benchmark datasets from the literature.

To summarize, the contribution of this work lies on the construction of a novel LBP-like descriptor which is making use of the invariant properties the moments and
moment invariants have. This is achieved by using a novel RST (rotation, scale, translation) invariant image representation called image momentgram, which is constructed by applying a straightforward procedure. This new image representation depends on the type of the used moment family and therefore can take several forms, by incorporating different image information. This flexibility helps on finding the most appropriate moment family that better describes the problem at hand. The utility of the momentgram is to combine the main advantages of the moments and moment invariants with those of the LBP method, aiming to construct an invariant local descriptor. Based on this RST image representation the LBP operator can be applied on, in order to extract invariant feature vectors for pattern recognition applications.

The Mb-LBP descriptor shows a global behaviour that comes from the nature of the moment functions to describe the image's content in several components (“bands” - moment orders). On the other hand, the local behaviour of the Mb-LBP descriptor is based on the local information of the constructed momentgram and not that of the original image. This is the reason why the Mb-LBP descriptor is stable under rotation/scaling/translation, since these transformations are filtered by the construction of the momentgram, while the distinctive patterns' information is captured through the LBP histograms of the momentgrams.

The paper is organized as follows: Section 2 describes the most popular moment families and the corresponding moment invariants. The basic theory of LBP operator is presented in Section 3, while the proposed local descriptor along with its definitional principles is discussed in Section 4. An extensive experimental study regarding the classification performance of the proposed descriptor, in comparison with the moments, moment invariants feature vectors, the conventional LBP and other
popular descriptors, takes place in Section 5. Finally, the main conclusions are summarized and discussed in Section 6.

2. Moments and Moment Invariants

Image moments have attracted the attention of the engineering scientific community for several decades, as a powerful tool to describe the content of an image. They have been used in many research fields of the engineering life, such as pattern recognition [9,10], computer vision [11,12] and image processing [13,14] with significant results. The first introduction of image moments for classification purpose was performed by Hu [9], by introducing the concept of moment invariants for invariant pattern recognition applications. By using the geometrical, central and normalized image moments [13], Hu proposed seven measurements invariant to any translation, scaling and rotation transformation of the image be processed, called moment invariants. Since then, many attempts to develop improved moment invariants with superior classification performance have taken place lately [15,16].

Recently, the scientists have developed the orthogonal moments and moment invariants, which use as kernel functions polynomials that constitute orthogonal basis and therefore they present minimum information redundancy, meaning that different moment orders describe different image part. Such moment families are the Legendre [13], Zernike and Pseudo-Zernike [17], Fourier-Mellin [18], Tchebichef [19], Krawtchouk [20] moments. These moments can be used as image descriptors after an appropriate normalization procedure in order to achieve translation, scale and rotation invariances.

In the next sections the most representative moment families are described and their invariants are derived. The extracted moment features will be used to construct the proposed Mb-LBP descriptors for the case of each moment family.
2.1 Moments

The general computational form of a \((n+m)^{th}\) order moment of a \(NxM\) image having intensity function \(f(x, y)\), is defined as follows

\[
M_{nm} = NF \times \sum_{i=1}^{N} \sum_{j=1}^{M} Kernel_{nm}(x_i, y_j) \times f(x_i, y_j)
\]  

(1)

where \(Kernel_{nm}()\) corresponds to the moment’s kernel consisting of specific polynomials [21] of orders \(n\) and \(m\), which constitute the orthogonal basis and \(NF\) is a normalization factor (in the case of geometric moments the kernel has the form of a monomial). The type of Kernel’s polynomial gives the name to the moment family by resulting to a wide range of moment types. In the following Table 1, the main characteristics of the most representative moment families, the Geometric Moments (GMs), Zernike Moments (ZMs), Legendre Moments (LMs), Tchebichef Moments (TMs) and Krawtchouk Moments (KMs), are summarized.

Once a finite number of moments up to a specific order \(n_{max}\) is computed, the original image can be reconstructed by applying a simple formula, inverse to Eq.(1), defined as:

\[
\hat{f}(x, y) = \sum_{n=0}^{n_{max}} \sum_{m=0}^{m_{max}} Kernel_{nm}(x, y) \times M_{nm}
\]  

(2)

In theory, if one computes all image moments and uses them in Eq.(2), the reconstructed image is identical to the original one with minimum reconstruction error.

Table 1. Main characteristics of orthogonal moment families.

<table>
<thead>
<tr>
<th>Moment Family</th>
<th>Kernel Form</th>
<th>Properties</th>
<th>Factor (NF) Normalization</th>
<th>Type</th>
<th>Coordinate System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric Moments (GMs) [13]</td>
<td>$Kernel_{nm}(r, \theta) = P_n(x) \times P_m(y)$</td>
<td>$P_n(x) = x^n, P_m(x) = x^m$</td>
<td>1</td>
<td>continuous/discrete</td>
<td>[-1,1]/image dimensions</td>
</tr>
<tr>
<td>Zernike Moments (ZMs) [17]</td>
<td>$Kernel_{nm}(r, \theta) = R_{nm}(r, \theta) \times e^{-jnm\theta}$</td>
<td>$R_{nm}(r, \theta) = \sum_{s=0}^{n</td>
<td>m</td>
<td>} (-1)^s \cdot \frac{(n-s)!}{s!(n-m</td>
<td>s</td>
</tr>
<tr>
<td>Legendre Moments (LMs) [13]</td>
<td>$Kernel_{nm}(x, y) = L_n(x) \times L_m(y)$</td>
<td>$L_n(x) = \frac{1}{2^n n!} \frac{d^n(x^2-1)^n}{dx^n}, L_m(x) = \frac{1}{2^m m!} \frac{d^m(x^2-1)^m}{dx^m}$</td>
<td>$(2n+1)(2m+1)$</td>
<td>continuous</td>
<td>[-1,1]</td>
</tr>
<tr>
<td>Tchebichef Moments (TMs) [19]</td>
<td>$Kernel_{nm}(x, y) = t_n(x) \times t_m(y)$</td>
<td>$t_n(x) = \sum_{k=0}^{n} (-1)^{n-k} \binom{N-1-k}{n-k} \frac{n+k}{n} \binom{x}{k}$ and $t_m(y) = \sum_{k=0}^{m} (-1)^{m-k} \binom{M-1-k}{m-k} \frac{m+k}{m} \binom{y}{k}$</td>
<td>$F(n,N)\times F(m,M)$ where $F(n,N) = \binom{N+n}{2n}$ and $F(m,M) = \binom{M+m}{2m}$</td>
<td>discrete</td>
<td>image dimensions</td>
</tr>
<tr>
<td>Krawtchouk Moments (KMs) [20]</td>
<td>$Kernel_{nm}(x, y) = K_n(x; p_1, N) \times K_m(y; p_2, M)$</td>
<td>$K_n(x; p_1, N) = \sum_{k=0}^{N} a_{k,n,p_1} x^k, K_m(x; p_2, M) = \sum_{k=0}^{M} a_{k,m,p_2} x^k$</td>
<td>1</td>
<td>discrete</td>
<td>image dimensions</td>
</tr>
</tbody>
</table>
2.2 Moment Invariants

Apart from the ability of the moments to describe the content of an image in a statistical fashion and to reconstruct it (orthogonal moments) perfectly according to Eq.(2), they can also be used to distinguish a set of patterns belonging to different categories (classes). This property makes them suitable for many artificial intelligence applications such as biometrics, visual inspection or surveillance, quality control, robotic vision and guidance, biomedical diagnosis, mechanical fault diagnosis, etc. However, in order to use the moments to classify visual objects, they have to ensure high recognition rates for all possible object’s orientations.

Mainly, there are two methodologies used to ensure the invariance under common geometric transformations (rotation, scaling and translation) either by image coordinates normalization and description through the geometric moment invariants [13,22] or by developing new computation formulas which incorporate these useful properties inherently [22]. The former strategy is applied next for deriving the moment invariants of each moment family, since it can be applied in each moment family in the same way.

Initially, by applying coordinates normalization [23], the GMs of Eq.(1) are transformed to invariant quantities called Geometric Moment Invariants (GMIs). The GMIs exhibit translation invariance by moving the origin to the centre of mass of the image \((\bar{x}, \bar{y})\), rotation invariance by estimating and altering the rotation angle and scaling invariance by normalizing each moment with the zeroth order geometrical moment as depicted in the first row of Table 2. However, more detailed information regarding the derivation of moment invariants can be found in [13,21,23].

Finally, the corresponding invariants of each moment family are derived by expressing them in terms of GMIs, as presented in Table 2.

Table 2. Main characteristics of orthogonal moment invariants.

<table>
<thead>
<tr>
<th>Moment Family</th>
<th>Invariants</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric Moment Invariants (GMIs)</td>
<td>$GMI_{nm} = GM_{00}^{-1} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \left[ (x-x) \cos \theta + (y-y) \sin \theta \right]^{n} \left[ (y-y) \cos \theta - (x-x) \sin \theta \right]^{m} f(x,y)$</td>
<td>$\gamma = \frac{n+m+1}{2}, \ \bar{x} = \frac{GM_{00}}{GM_{00}}, \ \bar{y} = \frac{GM_{00}}{GM_{00}}, \ \theta = \frac{1}{2} \tan^{-1} \left( \frac{2\mu_{1}}{\mu_{2} - \mu_{0}} \right), \ \mu_{nm} = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} (x-x)^{n} (y-y)^{m} f(x,y)$</td>
</tr>
<tr>
<td>Zernike Moment Invariants (ZMIs)</td>
<td>$ZMI_{nm} = \frac{n+1}{\pi} \sum_{k=m}^{n} B_{nmk} \sum_{i=0}^{m} \sum_{j=0}^{m} w^{i} i^{m} GMI_{k-2i-j,2i+j}$</td>
<td>$w = \left{ \begin{array}{ll} -i, &amp; m &gt; 0 \ i, &amp; m \leq 0 \end{array} \right., \ i = \sqrt{-1}, \ s = \frac{1}{2} (k-m), B_{nmk} = \frac{(-1)^{\frac{k-m}{2}} (n+k)!}{(n-k)!(k+m)! (k-m)!}$</td>
</tr>
<tr>
<td>Legendre Moments (LMIs)</td>
<td>$LMI_{nm} = \frac{(2n+1)(2m+1)}{N \times M} \sum_{i=0}^{n} \sum_{j=0}^{m} a_{n,i} a_{m,j} GMI_{ij}$</td>
<td>$a_{00} = 1, a_{11} = 1, a_{p+1,0} = \frac{(-p)a_{p-1,0}}{p+1}, \ p \geq 1$</td>
</tr>
<tr>
<td>Tchebichef Moment Invariants (TMIs)</td>
<td>$TMI_{nm} = \frac{1}{\beta(n,M)\beta(m,M)} \sum_{k=0}^{n} \sum_{l=0}^{m} T_{n,k} T_{m,l} \sum_{i=0}^{k} \sum_{j=0}^{l} s(i,j) GMI_{ij}$</td>
<td>$T_{n,k} = \frac{(n+k)!}{(n-k)! \times k!}, \ s(0,0) = 1, \ s(k,0) = s(0,i) = 0, \ k, i \geq 1$</td>
</tr>
<tr>
<td>Krawtchouk Moment Invariants (KMs)</td>
<td>$KMI_{nm} = \left[ \rho(n) \rho(m) \right]^{2} \sum_{i=0}^{n} \sum_{j=0}^{m} d_{n,p,i} d_{m,p,j} V_{ij}$</td>
<td>$V_{nm} = \sum_{p=0}^{n} \sum_{q=0}^{m} \binom{n}{p} \binom{m}{q} \left( \frac{N \times M}{2} \right)^{p+q-1} \left( \frac{N - n - p}{2} \right) \left( \frac{M - m - q}{2} \right) GMI_{pq}$</td>
</tr>
</tbody>
</table>

\[ d_{i,0,p} = 1 \]
\[ d_{k,n,p} = (-1)^{k} \sum_{i=0}^{n} Y_{i,k+1,p} s(i+1,n), \ n \geq 1, \ Y_{i,n,p} = \frac{(-k)^{n}}{(N-n) \cdot p^{n} \cdot n!} \]
3. Local Binary Patterns (LBP)

Local Binary Patterns (LBP) operator has been introduced by Ojala et al. [2], for texture analysis purposes. LBP is a powerful illumination invariant local descriptor, which a binary code that describes the local texture pattern is constructed by thresholding a neighbourhood of a centred pixel. If the gray value of each neighbourhood’s pixel is greater than the value of the centred one, this pixel takes the value 1 in a thresholding fashion; otherwise it takes the 0 value. In this way a thresholded image is constructed by applying specific weights in each pixel position so a gray-scale value in the range [0,255], which describes this small image region, is derived. By applying this operator to all available image regions of the same size, a type of histogram describing the intensity distribution of the entire image, is formed. In the following Fig.1, a typical $3 \times 3$ neighbourhood corresponding to a small image region is presented along with the image resulted after thresholding in respect to the centred pixel and the weights of each pixel used in order to compute the LBP decimal value of this region.

\[
LBP = \sum_{i=0}^{7} b_i \times 2^i
\]

where $b_i = \{0,1\}$ corresponds to the binary value of the binary image derived by applying thresholding.

**Figure 1.** Computation of the LBP for a small gray-scale image portion.

The transformation of the binary code to a decimal value is achieved by assigning specific weights to the pixels belonging to the neighbourhood in process, according to their position and by applying the following formula:
A more general form of the LBP operator for a neighbourhood with \( P \) pixels \( \{g_p\} \), \( p=0,..P-1 \) in a distance \( R \) from the centred pixel \( g_c \) [8], is the following:

\[
LBP_{p,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \times 2^p
\]  
\[(4)\]

with

\[
s(g_p - g_c) = \begin{cases} 1, & g_p - g_c \geq 0 \\ 0, & g_p - g_c < 0 \end{cases}
\]  
\[(5)\]

being the thresholding procedure.

The histogram formed by applying the basic LBP operator of Eq.(3) in each \( 3\times3 \) region of the image, constitutes a signature of the image’s content, which encloses the intensity dependencies of the pixels belonging in the region. The resulted histogram is insensitive to illumination and gray-scale variations, invariant under image translation, computed easily and thus can be used as local descriptor in order to distinguish several image patterns [2].

Although the basic LBP operator gave a novel point of view in texture analysis, there was an additional need to enhance its operation by making it invariant to rotation transformation of the image in process. As a result, a new version of the LBP operator is proposed by Ojala et al. [7,8], extending its application to classification problems where rotated objects are encountered.

The rotation invariant local binary pattern operator is defined as [8]:

\[
LBP_{p,R}^{\text{inv}} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) \text{ if } U(LBP_{p,R}) \leq 2 \\ P + 1 \text{ otherwise}, \end{cases}
\]  
\[(6)\]

where

\[
U(LBP_{p,R}) = \left| s(g_{p-1} - g_c) - s(g_0 - g_c) \right| + \sum_{p=1}^{P-1} \left| s(g_p - g_c) - s(g_{p-1} - g_c) \right|
\]  
\[(7)\]

The above rotation invariant LBP operator constructs \( P+1 \) uniform binary patterns and can be implemented by creating a look-up table that converts the basic LBP codes into
their LBP\textsuperscript{riu2} (superscript \textsuperscript{riu2} reflects the use of rotation invariant “uniform” patterns) components [8].

In this work, the basic (Eq. (3)) and rotation invariant (Eq. (6)) local binary patterns operators are used to construct local descriptors, by applying them not on the original pattern images but on a novel image representation that ensures global and local information encoding and scaling invariance, simultaneously.

4. Moment-Based Local Binary Patterns (Mb-LBP)

The application of the LBP operators, defined in the previous sections, on the original image pattern, has a significant limitation regarding the invariance properties of the resulted descriptors. While translation and rotation invariances are achieved through the construction procedure of the local binary patterns, less attention has been paid to the scaling invariance. This deficiency of the previously defined LBP operators makes them inappropriate to more general pattern recognition applications, where scaled objects have to be further recognized.

Moreover, the local behaviour of the traditional LBP operators comprises a drawback in applications where the objects under recognition are totally different and more global information is needed in order to distinct them. An ideal descriptor has to combine local and global information in order to ensure high recognition rates for similar as well as for totally different objects.

The proposed novel local descriptor handles the above issues, by providing a local behaviour through the application of the LBP operator and encoding the global information through the representation of the original image by a moment and/or moment invariant description called momentgram, as analysed in the next section.
4.1 Momentgram

The main idea behind the concept of the momentgram is the construction of a new image representation in another working space different from the spatial image domain, which incorporates the moments’ distribution of the image.

In this way, several momentgrams can be derived for an image regarding the moments’ type (total, sliced), family and whether they are invariant under certain geometric transformations.

4.1.1 Total Momentgram (TMgram)

The main characteristics of the Total Momentgram (TMgram) are summarized in the following definition:

**Definition 1:** A Total Momentgram (TMgram) is an image that shows how the moments of an image are distributed along to their order \( n \) and repetition \( m \).

The mathematical representation of the (TMgram) is:

\[
\text{Momentgram}_{\text{Total}}^{\text{Family}} (n, m) = |M_{nm}|
\]  

(8)

where *Family* is the moment family and \( n, m \) are the moment’s order and repetition, respectively. From the above formula is obvious that the TMgram corresponds to a grayscale image size of \( n \times m \) pixels, after appropriate normalization. This means that the \((i,j)\) pixel of the momentgram has as intensity value the normalized magnitude of the corresponding moment of \( i \) order and \( j \) repetition. If the moment invariants of the corresponding moment family, defined in section 2, are used instead of normal moments, the Invariant Total Momentgram (ITMgram) is constructed, which is invariant under translation, rotation and scaling of the image.

4.1.2 Sliced Momentgram (SMgram)
Recently, the authors have proposed a new methodology for the acceleration of the moments’ computation [25,26] and 2-D DCT transform [27], which also provides a mechanism to form novel moment invariants [28] that improve the pattern recognition rate. This methodology is based on the ISR (Intensity Slice Representation) [25-28] representation according to which a gray-scale image can be considered as the resultant of non-overlapped image slices, whose pixels have specific intensities. Based on this representation, we can decompose the original image into several slices, from which we can then reconstruct it, by applying fundamental mathematical operations.

Based on the ISR representation, the intensity function \( f(x,y) \) of a gray-scale image can be defined as an expansion of the intensity functions of the slices as:

\[
  f(x,y) = \sum_{i=1}^{s} f_i(x,y)
\]

(9)

where \( s \) is the number of slices (equal to the number of different intensity values) and \( f_i(x,y) \) is the intensity function of the \( i^{th} \) slice. In the case of a binary image \( s \) is 1 and thus \( f(x,y)=f_1(x,y) \).

In the general case of gray-scale images, each of the extracted slices can be considered as a two-level image and thus the IBR algorithm [25] can be applied directly, in order to decompose each slice into a number of non-overlapped binary blocks.

By using the ISR representation scheme, the computation of the \((n+m)^{th}\) order moment (Eq.(1)) of a gray-scale image \( f(x,y) \), can be performed according to the equation

\[
  M_{nm} = f_1M_{nm}^1 + f_2M_{nm}^2 + ... + f_sM_{nm}^s
\]

(10)

where \( f_i \) and \( M_{nm}^i, \ i=1,2,...,s \) are the intensity functions of the slices and the corresponding \((n+m)^{th}\) order moments of the \( i^{th} \) binary slice, respectively.

The corresponding moment of a binary slice \( M_{nm}^i \) is the moment computed by considering a block representation of the image [25], as follows
\[ M_{nm}^j = \sum_{j=0}^{k-1} M_{nm}(b_j) = NF \times \sum_{j=0}^{k-1} \sum_{x=x_{1,b_j}}^{x_{2,b_j}} \sum_{y=y_{1,b_j}}^{y_{2,b_j}} \text{Poly}_n(x) \times \text{Poly}_m(y) \]

\[ = NF \times \sum_{j=0}^{k-1} \left( \sum_{x=x_{1,b_j}}^{x_{2,b_j}} \text{Poly}_n(x) \right) \left( \sum_{y=y_{1,b_j}}^{y_{2,b_j}} \text{Poly}_m(y) \right) \]  \( (11) \)

where \( \text{Poly}_n(\cdot) \), \( \text{Poly}_m(\cdot) \) are the \( n^{th} \) and \( m^{th} \) orders of the same polynomial of the used moment family (see Table 1), \( x_{1,b_j}, x_{2,b_j} \) and \( y_{1,b_j}, y_{2,b_j} \) are the coordinates of the block \( b_j \), with respect to the horizontal and vertical axes, respectively. The reader can refer to [25-28] for more detailed information regarding the computation of the slice moments.

In this context, the Sliced Momentgram (SMgram) is determined according to the following definition:

**Definition 2:** A Sliced Momentgram is an image that shows the contribution of each slice’s moment in constructing the total moment of the original image.

\[ \text{Momentgram}_{\text{Sliced Family}}^{\text{index, slice}}(M_{nm}) = \frac{M_{\text{slice}}}{M_{nm}} \]

\( (12) \)

where \( \text{index} \) is the computation sequence of each moment \( (\text{index} = I(n=0,m=0), 2(n=0,m=1), \ldots) \), \( \text{slice} \) is one of the 255 possible intensity slices of the image and \( n,m \) the moment’s order and repetition, respectively. By applying an appropriate normalization a SMgram can be considered as a gray-scale image size of \( (\text{index}) \times (\text{slice}) \) pixels.

As in the case of the ITMgram, if the moments of each slice are replaced by the corresponding slice moment invariants, the Invariant Sliced Momentgram (ISMgram) is derived, which is invariant under translation, rotation and scaling of the image.

### 4.1.3 An Example

In order to understand the form of a momentgram, a sample image from the JAFFE [29] dataset, which contains images of seven facial expressions, is selected and its possible
momentgrams are constructed. The original image is depicted in Fig.2, while the resulted momentgrams for the case of the prescribed moment families are illustrated in Fig.3.

**Figure 2.** An image sample of the JAFFE [31] dataset.

Figure 3, illustrates the extracted momentgrams (by using a moments set from (n,m)=(0,0) up to (30,30)), after applying a typical normalization procedure in order to map the moment and moment invariants values into the gray-scale intensities range [0,255], of the above sample image for the case of each moment family. The first letter in front of the momentgrams’ title corresponds to the initial letter of the used moment family (G–Geometric, Z–Zernike, L–Legendre, T–Tchebichef and K–Krawtchouk).

![Momentgrams](image.png)

**Figure 3.** Momentgrams of the JAFFE sample image using (a) Geometric (1st row), (b) Zernike (2nd row), (c) Legendre (3rd row), (d) Tchebichef (4th row) and (e) Krawtchouk (5th row) moment families.
A careful study of the above momentgrams can lead to the conclusion that this new representation transforms the original image to specific patterns, which depend on the nature of each moment family and the moments’ order and repetition.

The formed patterns that correspond to the ITMgrams and ISMgrams are invariant under the basic geometric image transformations, since they are constructed by using the moment invariants of a specific moment family. Moreover, the discrimination ability of each momentgram is ensured by the discrimination power of the moments, and their classification performance will be defined through appropriate experiments in the next section.

### 4.2 Proposed Descriptor

The novel descriptor, which is proposed in this manuscript, is formed by applying the basic LBP operator on the introduced momentgrams instead of the original image commonly performed. The representation of an image by its momentgram exhibits some very useful properties against the original image, which enhance the discrimination ability of the LBP operator.

The main advantage of the momentgram constructed by the moment invariants is its property to remain the same under translation, rotation and scaling of the image’s content and therefore the LBP histogram computed on this momentgram is more suitable for invariant pattern recognition than the conventional LBP one.

Moreover, the momentgram captures global information about the content of the original image, which in collaboration with the local behaviour of the LBP operator, as a result combined information is provided. All in all, the enhanced LBP histogram in this case, encloses global as well as local information, regarding the object under description.

Apart from the two main properties of the new descriptor, the representation of an image by its momentgram gives more flexibility to the description capabilities of the LBP
operator, through the selection of several moment families and moment types, relatively to each application needs.

The following Fig. 4 illustrates the computational flowchart of the proposed descriptor, which constructs an enhanced LBP histogram able to distinguish patterns concerning a specific pattern recognition task.

**Figure 4.** Computational flowchart of the proposed descriptor.

The proposed operator that produces the novel descriptor which is an enhanced LBP histogram, is called *Moment-based LBP (Mb-LBP)* and based on the several free parameters that control its nature, the following nomenclature is used to describe it.

$$Mb - LBP_{\text{family}}^{\text{type}(\text{order})}$$

$$Mb - LBP_{\text{family}}^{(i)(\text{type})(\text{order})}$$

where *family* is the initial letter of the moment family used to construct the momentgram (*G: Geometric, Z: Zernike, L: Legendre, T: Tchebichef and K: Krawtchouk*), *type* is the moments’ type (*t: total* or *s: slice*) and *order* is the order (*n*) up to which the moments are computed (the repetition parameter (*m*) takes values according to *p* and the additional constraints that hold for each moment family). It is noted that only the second operator is invariant under translation, rotation and scaling, a property which is indicated by the index *i*.

In the next section, an extensive experimental study of the discriminative capabilities the proposed operator exhibits for different type of pattern recognition problems, is taking place.
5. Experimental Study

In order to investigate the classification performance and behaviour of the proposed descriptor, a set of appropriate experiments has been arranged. For the experimental purposes, specific software has been developed in C++ language that implements the momentgrams construction. Moreover, the MATLAB implementations for the LBP [30] and the built-in k-NN classifier [31] are used, while all experiments are executed in a Pentium 3.3GHz PC with 2GB RAM.

5.1 Datasets

For the sake of the experiments, representative datasets defining different pattern recognition problems are selected from the literature. Among these, there is a face recognition dataset (ORL [32]), a facial expression recognition dataset (JAFFE [29]), a texture recognition dataset (Ponce [33]) and three object recognition datasets which are subsets of the COIL-20 [34] dataset generated by applying additional geometric transformations and by adding Gaussian noise. The main characteristics of the datasets are summarized in the following Table 3.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Num. Classes</th>
<th>Instances / Class</th>
<th>Total Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 ORL [32]</td>
<td>face recognition</td>
<td>40</td>
<td>10</td>
<td>400</td>
</tr>
<tr>
<td>D2 JAFFE [29]</td>
<td>facial expression recognition</td>
<td>7</td>
<td>30,29,32,31,30,31,30</td>
<td>213</td>
</tr>
<tr>
<td>D3 Ponce [33]</td>
<td>texture classification</td>
<td>25</td>
<td>40</td>
<td>1000</td>
</tr>
<tr>
<td>D4 COIL-R [34]</td>
<td>object recognition</td>
<td>10</td>
<td>12</td>
<td>120</td>
</tr>
<tr>
<td>D5 COIL-S [34]</td>
<td>object recognition</td>
<td>10</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>D6 COIL-RN [34]</td>
<td>object recognition</td>
<td>10</td>
<td>12</td>
<td>120</td>
</tr>
</tbody>
</table>

In the case of the face and facial expression recognition datasets, an extra pre-processing task has to be performed on the images, in order to isolate the image’s part, which includes the main face’s information, by discarding the useless content. Moreover,
the Haar Cascade Classifier [35] is used in the case of JAFFE dataset in order to detect the faces and to extract blocks size of 160x140 pixels as image regions of interest, while the ORL dataset remains unchanged since the useful information is centred inside the original image.

The three object recognition datasets are constructed by choosing the first 10 patterns of the well-known COIL-20 [34] dataset. These patterns have been rotated 11 times by 30 degrees each time, by forming the COIL-R dataset (12 instances/class: 11 rotated, 1 original) and scaled 4 times from 50% to 150% by a step of 25% in order to construct the COIL-S dataset (5 instances/class: 4 scaled, 1 original). Finally, the COIL-RN is constructed by adding Gaussian noise with \((m,\sigma^2)=(0,0.04)\) to the COIL-R dataset.

In order to better understand the differences between the above datasets, four samples of each dataset are illustrated in the following Fig.5.
The selection of the datasets has been performed by taking into account the different conditions under which each problem is considered. For example, the face recognition dataset includes cases with highly changed illumination conditions, while in the case of the facial expressions recognition dataset there are local regions which carry high discriminative information. Moreover, owing to the fact that the traditional LBP operator was firstly applied on texture patterns, where local discriminative information need to be captured, the Ponce [33] texture dataset is selected. Eventually, the object recognition datasets are constructed in the context of the modelling of common geometric transformations (translation, rotation, scaling), usually occurring in generic pattern recognition applications.

By studying the information coding and discriminative abilities of the proposed descriptor on these various pattern recognition problems, useful conclusions about the improvement of the conventional LBP operator are drawn.

5.2 Classification

The next step, after the feature vectors establishment, is to perform the classification of the objects to the appropriate classes, in other words the recognition of each object. This last processing is performed by specific modules called classifiers and their performance is highly depended on their structure and the discrimination power of the features being used.

For the sake of the experiments, a simple classifier (the k-NN classifier [36]) is selected as the classification module in all pattern recognition problems under investigation.
However, in order to analyse the behaviour of each descriptor in conjunction with the classifier’s operation, several distance measures are selected to build different classifier’s configurations. These measures evaluate the distance of each object in process from the patterns representing problem’s classes and were used in order to avoid any configuration bias that might lead to wrong conclusions regarding the appropriateness of the Mb-LBP descriptor. Since each metric measures the distance or the similarity between two patterns from a different point of view, this variety of metrics is used to investigate the discrimination capabilities of the introduced descriptor. The object is decided to belong to the specific class having the least distance from its pattern.

Five well-known metrics from the literature, the Euclidean [37], Manhattan [38], Chebyshev [39] distances, Chi Square Statistics ($\chi^2$) [3], Pearson Correlation Coefficient [40], are selected and presented in the following,

\[
\begin{align*}
\text{Euclidean Distance} & \quad d_1(p, s) = \sqrt{\sum_{i=1}^{n} (p_i - s_i)^2} \\
\text{Manhattan Distance} & \quad d_2(p, s) = \sum_{i=1}^{n} |p_i - s_i| \\
\text{Chebyshev Distance} & \quad d_3(p, s) = \max_{i} \{|p_i - s_i|\} \\
\text{Chi Square Statistics} & \quad d_4(p, s) = \sum_{i=1}^{n} \frac{(p_i - s_i)^2}{p_i + s_i} \\
\text{Pearson Correlation Coefficient} & \quad d_5(p, s) = \frac{\sum_{i=1}^{n} (p_i - \bar{p})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n} (p_i - \bar{p})^2} \sqrt{\sum_{i=1}^{n} (s_i - \bar{s})^2}}
\end{align*}
\]

The above formulas measure the distance between two vectors $p = [p_1, p_2, p_3, \ldots, p_n]$ and $s = [s_1, s_2, s_3, \ldots, s_n]$, which are defined in the $\mathbb{R}^n$ space and $\bar{p}, \bar{s}$ are the mean values of $p$ and $s$ vectors, respectively.
It has to be noted that the above measures tend to 0 for the case of two equal vectors, except $d_5$ which gives 1, since it counts the similarity of the two vectors. In the simulations following this section, these vectors correspond to the histograms computed by applying the proposed Mb-LBP and the conventional LBP operators.

5.3 Simulations

The classification procedure using the k-NN classifier is performed with $k=1$ for all datasets and the patterns’ classes are considered to be the first instance of each class (D3,D4,D5,D6) or the centre of each class (D1,D2) computed by applying the same distance metric with that used in the classifier in each case. Moreover, in all the experiments, each dataset has been randomly divided into two subsets, the training (70%) and the testing (30%) sets.

The Feature Vectors (Table 4, FV - this nomenclature is used to describe the quantities extracted by applying a descriptor and are used as inputs to the classifiers, for the rest of this section), which will be compared in terms of their classification performance, are constructed by applying: the basic and three rotation invariant (LBP$_{8,1}$, LBP$_{16,2}$, LBP$_{24,3}$) LBP operators, the traditional Moments and Moment Invariants, the total and sliced Mb-LBP operators invariant or not to common geometric transformations.

Moreover, the moment based feature vectors will be studied for all different moment families of section 2 and for several moment orders.

Table 4. Feature vectors under comparison.

<table>
<thead>
<tr>
<th>Feature Vector ID</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV1</td>
<td>LBP</td>
<td>Basic LBP operator</td>
</tr>
<tr>
<td>FV2</td>
<td>LBP$_{8,1}$</td>
<td>Rotation invariant LBP operator consisting of 8 pixels in 1 pixel distance from the center neighbourhood</td>
</tr>
<tr>
<td>FV3</td>
<td>LBP$_{16,2}$</td>
<td>Rotation invariant LBP operator</td>
</tr>
</tbody>
</table>
The simulations are conducted along three different directions: 1) initially the usefulness of each distance measure is investigated, 2) the classification capabilities of all possible momentgrams are examined in comparison with the conventional LBP operators and 3) the same comparison is repeated only between the best momentgram and the LBP operators, for the case of face and facial expressions datasets, by applying an additional image division into square blocks strategy. Moreover, a separate comparative study of the proposed Mb-LBP descriptor with well-known descriptors from the literature, complete the overall experimental study.

### 5.3.1 Direction I (Effect of Distance Metric)
Before proceeding with the direct comparison of the several feature vectors, it is useful to investigate the effect of each distance metric to the classifier’s performance, in each dataset.

The following Table 5, depicts the distances’ scores (times of each distance that give the maximum classification rate) in a bold face, in conjunction with the minimum, maximum and mean classification rate (in %), for all the feature vectors under investigation in the case of each dataset. It is worth mentioning that in the cases where the distance having the maximum rate is more than one, the score is assigned to all these distances simultaneously.

**Table 5. Classification performance of the distance measures in each dataset.**

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Distances</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>20</td>
<td>12.73</td>
<td>98.79</td>
<td>5</td>
<td>12.12</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12.12</td>
<td>89.09</td>
<td>30</td>
<td>14.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>46.73</td>
<td>36.70</td>
<td>48.30</td>
<td>43</td>
</tr>
<tr>
<td>D2</td>
<td>17</td>
<td>15.49</td>
<td>72.77</td>
<td>26</td>
<td>16.90</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>33.02</td>
<td>77.93</td>
<td>57.75</td>
<td>32</td>
</tr>
<tr>
<td>D3</td>
<td>25</td>
<td>11.30</td>
<td>96.80</td>
<td>25</td>
<td>11.20</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>39.66</td>
<td>51.30</td>
<td>28.77</td>
<td>22</td>
</tr>
<tr>
<td>D4</td>
<td>50</td>
<td>12.50</td>
<td>100.00</td>
<td>35</td>
<td>11.67</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>61.54</td>
<td>100.00</td>
<td>55.63</td>
<td>34</td>
</tr>
<tr>
<td>D5</td>
<td>55</td>
<td>12.00</td>
<td>100.00</td>
<td>82</td>
<td>12.00</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>55.45</td>
<td>100.00</td>
<td>51.89</td>
<td>87</td>
</tr>
<tr>
<td>D6</td>
<td>29</td>
<td>15.83</td>
<td>97.50</td>
<td>75</td>
<td>16.67</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>39.47</td>
<td>99.17</td>
<td>35.93</td>
<td>26</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>Scores</strong></td>
<td><strong>196</strong></td>
<td><strong>339</strong></td>
<td><strong>157</strong></td>
<td><strong>527</strong></td>
<td><strong>244</strong></td>
</tr>
</tbody>
</table>

The classification rates along with the scores of each distance depicted in the above Table 5, correspond to a set of 184 (4(LBP FVs) + 6(families) × 5(Mb-LBP FVs) × 6(orders: from 5 to 30 with increment 5)) different experiments, arranged for each dataset and each distance.

A careful study of the above results can lead to some interesting conclusions, regarding the suitability of the distances, in classifying patterns of several datasets using feature
vectors of diverse discrimination capabilities. Considering the scores each distance achieves, it is concluded that distance $d_4$ (Chi Square Statistics) performs better (527 times) than the other ones, in the most datasets, while in the cases where it is less effective, its performance is very close to the outperformed distances. The superiority of $d_4$ holds also regarding the mean classification rate, which is higher than the rest distances, a fact which reinforces the credibility of the distance.

Moreover, distance $d_2$ (Manhattan Distance) seems to work well for the case of D4-D6 object recognition datasets, where basic geometric transformations (translation, rotation and scaling) of the objects have been incorporated. This result shows that the invariant properties of the feature vectors, regarding these geometric distortions, are preserved while in the case of the other distances, these properties are corrupted.

Before making a decision on the appropriateness of the five distances, it is useful to study the same performance in relation to the type of feature vectors used to classify the instances of each dataset. The following Table 6, summarizes the times (scores) a feature vector shows the best classification rate along with the minimum, maximum and mean rates for each distance case.

<table>
<thead>
<tr>
<th>Feature Vectors</th>
<th>Distances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d_1$</td>
</tr>
<tr>
<td>$FV1$</td>
<td>30.00</td>
</tr>
<tr>
<td></td>
<td>79.00</td>
</tr>
<tr>
<td></td>
<td>48.35</td>
</tr>
<tr>
<td>$FV2$</td>
<td>18.78</td>
</tr>
<tr>
<td></td>
<td>60.00</td>
</tr>
<tr>
<td></td>
<td>34.48</td>
</tr>
<tr>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>46.66</td>
</tr>
<tr>
<td>$FV4$</td>
<td>26.76</td>
</tr>
<tr>
<td></td>
<td>80.00</td>
</tr>
<tr>
<td></td>
<td>41.79</td>
</tr>
<tr>
<td>$FV5$</td>
<td>17.50</td>
</tr>
<tr>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>52.39</td>
</tr>
<tr>
<td>$FV6$</td>
<td>12.25</td>
</tr>
<tr>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>55.28</td>
</tr>
</tbody>
</table>
The results of the above table are consistent with those of Table 5 and emphasize the superiority of the \( d_4 \) generally. The maximum classification rate is achieved when this distance is used in the k-NN classifier, for the set of all different feature vectors, as compared with the other distances.

Moreover, Table 6 shows an early result regarding the performance of the proposed descriptors (FV7- FV10) which present many scores, when they are used in conjunction with the \( d_4 \) distance, while their mean classification rates over all the datasets are the higher among those of the other features. Since, the set of the proposed descriptors (FV7-FV10) consists of the features extracted by all moment families for various moment orders, it is concluded that almost in all cases there is a feature vector of a specific moment family and order that performs better than the conventional descriptors (LBP, moments and moment invariants), which is constructed by applying the proposed operator of Fig. 4.

The previous study answers to the question of what is the most appropriate distance for the k-NN classifier, when the histogram based feature vectors are used to classify the instances of the ten different datasets considered. The answer is that the \( d_4 \) (Chi Square Statistics) distance gives in general the highest classification rates, a conclusion which is consistent with the selection of the Ahonen et al. [3], where the same distance is used for the case of the histogram derived by the LBP operator.
As far as the classification capabilities of the proposed descriptor are concerned, significant outcomes have been drawn through the above study, showing that it is superior to the conventional LBP, moments and moment invariants, for the datasets it was applied on.
5.3.2 Direction II (Classification Capabilities of Mb-LBP Descriptor)

Considering the $d_4$ (Chi Square Statistics) as the distance incorporated in the k-NN classifier, it is constructive for the overall performance analysis of the proposed descriptor, to investigate the behaviour of the resulted feature vectors in relation to the moment family which is used to construct the corresponding momentgrams.

The following Table 7, summarizes the times (scores) a feature vector of each moment family shows the best classification rate along with the minimum, maximum and mean rates in each case. Since the LBP based features vectors (FV1-FV4) are independent of the moment family, they are not included in Table 6 and only the original moments (FV5), moment invariants (FV6) and the Mb-LBP descriptors (FV7-FV10) are considered.

<table>
<thead>
<tr>
<th>Feature Vectors</th>
<th>Moment Family</th>
<th>Geometric</th>
<th>Legendre</th>
<th>Krawtchouk</th>
<th>Tchebichef</th>
<th>Zernike</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV5</td>
<td>0 6 3 9 19</td>
<td>25.00</td>
<td>15.83</td>
<td>20.83</td>
<td>22.00</td>
<td>25.50</td>
</tr>
<tr>
<td>FV6</td>
<td>4 6 2 13 7</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>FV7</td>
<td>1 4 1 10 3</td>
<td>58.50</td>
<td>36.00</td>
<td>97.75</td>
<td>96.00</td>
<td>86.66</td>
</tr>
<tr>
<td>FV8</td>
<td>5 1 0 0 1</td>
<td>61.66</td>
<td>80.00</td>
<td>83.33</td>
<td>75.00</td>
<td>94.16</td>
</tr>
<tr>
<td>FV9</td>
<td>1 6 6 11 0</td>
<td>100.00</td>
<td>98.00</td>
<td>100.00</td>
<td>100.00</td>
<td>77.50</td>
</tr>
<tr>
<td>FV10</td>
<td>6 6 0 4</td>
<td>91.66</td>
<td>80.00</td>
<td>78.33</td>
<td>93.33</td>
<td>96.66</td>
</tr>
<tr>
<td><strong>Total Scores</strong></td>
<td><strong>17 29 12 43 34</strong></td>
<td><strong>6.00</strong></td>
<td><strong>20.65</strong></td>
<td><strong>21.12</strong></td>
<td><strong>23.94</strong></td>
<td><strong>22.06</strong></td>
</tr>
</tbody>
</table>

From the above Table 7, the supremacy of the feature vectors produced by using the Tchebichef moments family is clear, when compared to the other families. This
outperformance holds regarding not only the reported scores but also the absolute classification rates, which are higher.

Moreover, the 0 scores shown some feature vectors for specific moment family e.g. FV5 for Geometric family, mean that zero times these feature vectors, when the specific moment family is used, show the best classification rate among the other FV-family combinations. The presence of these 0’s is relative to the used distance metric, while for other distances these 0’s might not hold and 0’s in other cases might be appeared.

Apart from the main conclusion that is drawn from the above table, concerning the superiority of the Tchebichef moments family for all feature vectors, an additional significant outcome is also deduced. For the set of all experiments, there is always a winner feature vector (of some moment family and order), which is formed by applying the proposed methodology, a fact that enforces the usefulness of the introduced descriptor since it encloses discriminative information.

For space saving purposes, it is not possible to illustrate the performance of each moment family’s feature vectors for all datasets and only the case of the Tchebichef moment family is considered and depicted in the following Fig.6. In this figure the classification rates of the traditional LBP and the Tchebichef moments, moment invariants and Mb-LBP descriptors, for several moments’ order and each dataset, are illustrated.
By studying the above feature vectors’ performance, some very interesting conclusions are drawn, regarding their classification capabilities in each pattern recognition dataset.

The original Tchebichef moments (TMs) present a remarkable performance for the case of the D1(ORL), D2(JAFFE) and D3(Ponce) datasets, where the geometric invariances do not play a significant role, since in these cases the transformations are of
different nature (illumination, pose variations, etc.). For the rest datasets the Tchebichef moment invariants (TMIs) are superior to the TMs, since they are invariant under translation, rotation and scaling of the images’ content. Based on these observations it is concluded that the TMs perform better than the TMIs, for non-geometric distortions of the images.

The $Mb-LBP_t^{(order)}$, which is constructed by applying the proposed operator, classifies correctly more patterns than the TMs (Table 1) for low moment orders for the case of D1 dataset (Fig.6a), while it outperforms them totally for the case of the D2(JAFFE) (Fig.6b) dataset (87.79% instead of 71%) by using less moments. This happens due to the additional capability of the $Mb-LBP_t^{(order)}$ to enclose local information about the pattern it describes. Similar conclusions are derived through the study of the performance for the rest datasets, where the proposed feature vector presents higher or equal classification rates in the most cases. Generally, the application of the proposed operator on the momentgram image representation seems to improve the classification abilities of the moments by a significant factor.

While the superiority of the Mb-LBP feature vectors over the corresponding moment and moment invariants descriptors is somewhat expected, since the former vectors combine global and local discriminative information, their supremacy against the traditional LBP histograms is worth noticing. For each dataset a Mb-LBP feature vector is better than any LBP descriptor and in almost all datasets, this superiority is followed by a significant higher classification rates.

The outperformance of the Mb-LBP feature vectors is spread along two distinct directions. Firstly, in problems where the geometric invariance properties are not demanded, such as D1-D3, the proposed descriptors are more efficient since they enclose global as well as local information about the patterns’ classes, while the LBP descriptors describe local information regarding a small image portion. The combination of global
and local information gives additional discrimination capabilities to the constructed descriptors and makes them capable to distinguish similar or totally different patterns.

On the contrary, in more general pattern recognition problems where there is a need to recognize translated, rotated and scaled objects, the LBP descriptors perform poor and only the rotation invariant LBP is robust. Especially, in D5(COIL-S) dataset which includes scaled objects the LBP descriptors fail to classify correctly many patterns. In these cases, the invariant Mb-LBP feature vectors correctly classify the patterns with very high classification rates, since they are constructed by using the family’s moment invariants which are invariant under translation, rotation and scaling of the dataset’s patterns.

In order to further analyse the classification behaviour of the proposed descriptors in comparison to the LBP ones, the ROC curves [36,41] of all classes of the D2(JAFFE) dataset are extracted and depicted in Fig.7.

A ROC curve [41] is a plot with the false positive rate (incorrectly labelled as positive) on the X axis and the true positive rate (correctly labelled as positive) on the Y axis, by encapsulating all information contained in the confusion matrix. The desirable form of a ROC curve is a point to (0,1) meaning that the classifier classifies all positive and negative cases correctly (0 (none) false positive rate and 1 (all) true positive rate). Based on this definition, the feature vector of whom the ROC curve is as close as possible to the (0,1) point, shows better classification capabilities.

The 1st column of the Fig.7 presents the ROC curves of the LBP and the best of the Mb-LBP descriptors, while the 2nd column shows the ROC curves of all Mb-LBP descriptors for the same classes.
In the above ROC curves the Mb-LBP descriptors of 25th order are selected, since in this order the classification rate is close to the mean rate for all moment orders from 5 to 30 with increment 5. The right column of Fig. 7 shows the performance of the Mb-LBP descriptors, where it is obvious that the $Mb-LBP^{25}_T$ one classifies almost all the classes with minimum false positive and maximum true positive rates.

Figure 7. ROC curves of each class of the D2(JAFFE) dataset.
By comparing the classification behaviour of the LBP descriptors with the $M_b-LBP_{25}$, it is evident that the latter behaves better in all classes except the class 7, where the basic LBP descriptor is marginal better.

Concluding, from the previous analysis it is claimed that the proposed Mb-LBP descriptors can be used as effective surrogates of the traditional LBP ones and the newly proposed operator can significantly improve the classification capabilities of the LBP extracted histograms.

### 5.3.3 Direction III (Face and Facial Recognition Cases)

For the case of face recognition Ahonen et al. [3] proposed an alternative policy regarding the application of the LBP operator. They extend the histogram of the basic LBP operator into a *spatially enhanced histogram*, which is constructed by dividing the image into several rectangular blocks where the basic LBP operator is applied separately. The resulted histograms are then combined by concatenating them in order to form the spatially enhanced histogram. The form of the JAFFE sample image of Fig.2 after applying a block division strategy is depicted in the following Fig.8.

![Figure 8](image)

*Figure 8.* The JAFFE sample image of Fig.2 divided in rectangular blocks.

Based on this methodology, the classification experiments of the previous section are repeated only for the case of D1(ORL) and D2(JAFFE) datasets, by dividing the images into $5 \times 4$ blocks of different sizes for each dataset (D1: $22 \times 23$ and D2: $32 \times 35$, pixels) and the performance of all feature vectors under comparison is summarized in the following Table 8. The order of the moment based descriptors is selected to be 6, in order to construct feature vectors with the same length (980 features).

Table 8. Classification performance (rates in % ) of the spatially enhanced histogram based descriptors.

<table>
<thead>
<tr>
<th>Feature Vectors</th>
<th>D1</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YALE Dataset</td>
<td>JAFFE Dataset</td>
</tr>
<tr>
<td>Division</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>LBP</td>
<td>83.03</td>
<td><strong>100.00</strong></td>
</tr>
<tr>
<td>$LBP_{8,1}$</td>
<td>49.69</td>
<td>91.51</td>
</tr>
<tr>
<td>$LBP_{16,2}$</td>
<td>56.96</td>
<td>96.36</td>
</tr>
<tr>
<td>$LBP_{24,3}$</td>
<td>59.39</td>
<td>96.96</td>
</tr>
<tr>
<td>TMs</td>
<td>95.15</td>
<td>99.39</td>
</tr>
<tr>
<td>TMIs</td>
<td>27.27</td>
<td>63.00</td>
</tr>
<tr>
<td>$Mb – LBP_{T}^{t6}$</td>
<td><strong>96.69</strong></td>
<td><strong>100.00</strong></td>
</tr>
<tr>
<td>$Mb – LBP_{F}^{t6}$</td>
<td>76.36</td>
<td>86.00</td>
</tr>
<tr>
<td>$Mb – LBP_{T}^{t6}$</td>
<td>32.12</td>
<td>75.75</td>
</tr>
<tr>
<td>$Mb – LBP_{F}^{t6}$</td>
<td>52.12</td>
<td>84.24</td>
</tr>
</tbody>
</table>

From the results of the above table, it is clear that the division of the images into several rectangular blocks and the computation of the spatially enhanced histogram based descriptors, improves significantly the recognition capabilities of the LBP feature vectors, while for the case of Mb-LBP descriptors this improvement is lower. However, the supremacy of the $Mb – LBP_{T}^{t6}$ still holds and its efficiency is boosted, by giving the best rates (bold faced rates) in all cases.

This experimental study confirms the usefulness and the potentiality of the proposed descriptor in recognizing patterns belonging to complicated problems, such as face and facial expressions recognition. The usage of the momentgram image representation provides alternative and more discriminative information, where the application of the LBP operator constructs highly efficient feature vectors.

5.3.4 Comparison with other descriptors

From the previous analysis it is obvious that the proposed Mb-LBP descriptor improves the discrimination capabilities of the conventional LBP one, in almost all the
datasets. In order to further investigate the performance of the Mb-LBP histogram-based features an additional comparative study, with pattern descriptors known from the literature needs to be conducted. For this purpose, the HOG, HOG-LBP [42] and LBP-HF [43] descriptors are selected due to their popularity. Since the HOG-LBP descriptor uses the \( LBP_{8,1}^{u2} \) (uniform - u2) operator which is not rotation invariant, the HOG-LBP is adopted by using the rotation invariant (riu2) \( LBP_{8,1}, LBP_{16,2}, LBP_{24,3}[8] \) operators in order to construct rotation invariant HOG-LBP descriptors. Furthermore, the LBP-HF descriptor [43] ensures rotation invariance by applying the Fourier transform on the extracted uniform (u2) LBP histogram of the image. At this point, it is worth mentioning that the application of those feature extraction procedures construct feature vectors of different lengths (different number of bins) than the proposed descriptor as depicted in Table 9.

The classification results of the HOG, HOG-LBP and LBP-HF descriptors along with those of the best Mb-LBP one (\( Mb - LBP_{T}^{r6} \)) for the case of the face and facial recognition datasets, described in the previous section, where the procedure of the image division in 5×4 blocks is conducted, are summarized in the following Table 9.

### Table 9. Descriptors’ comparison results (classification rates in %) for the case of facial and face recognition datasets.

<table>
<thead>
<tr>
<th>Feature Vectors</th>
<th>Number of bins /cell</th>
<th>D1 YALE Dataset</th>
<th>D2 JAFFE Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>256</td>
<td>100.00</td>
<td>96.24</td>
</tr>
<tr>
<td>HOG</td>
<td>9</td>
<td>99.39</td>
<td>65.73</td>
</tr>
<tr>
<td>HOG – LBP(_{8,1}^{u2}) [45]</td>
<td>59</td>
<td>100.00</td>
<td>78.40</td>
</tr>
<tr>
<td>HOG – LBP(_{8,1})</td>
<td>19</td>
<td>99.39</td>
<td>71.36</td>
</tr>
<tr>
<td>HOG – LBP(_{16,2})</td>
<td>27</td>
<td>99.39</td>
<td>70.89</td>
</tr>
<tr>
<td>HOG – LBP(_{24,3})</td>
<td>35</td>
<td>99.39</td>
<td>73.24</td>
</tr>
<tr>
<td>LBP – HF [46]</td>
<td>59</td>
<td>99.75</td>
<td>77.93</td>
</tr>
<tr>
<td>Mb – LBP(_{T}^{r6})</td>
<td>256</td>
<td>100.00</td>
<td>99.53</td>
</tr>
</tbody>
</table>

38
The results of the above Table 9 clearly show that the combination of the HOG and LBP descriptors in a concatenation fashion marginal improves their individual classification performance with the $HOG - LBP_{8,1}^{u^2}$ [42] being the most efficient of them, the LBP-HF descriptor shows lower classification rate, while the proposed Mb-LBP descriptor outperforms them. From the above results it seems that the mechanisms that ensures rotation invariance in the construction of the -HOG-LBP$_{8,1}$, HOG-LBP$_{16,2}$, HOG-LBP$_{24,3}$ and LBP-HF descriptors, the rotation invariant operators LBP$_{8,1}$, LBP$_{16,2}$, LBP$_{24,3}$ and the Fourier transform respectively, decrease the discrimination capabilities of these descriptors, something which does not hold for the case of the proposed $Mb - LBP_{T}^{r^6}$ descriptor.

By comparing the most efficient descriptors of the previous experiment ($HOG - LBP_{8,1}^{u^2}$, $HOG - LBP_{24,3}$, $LBP - HF$, $Mb - LBP_{T}^{r^6}$ ) on the datasets D4-D6 that incorporate geometric transformations, the following results of Table 10 are derived:

**Table 10.** Descriptors’ comparison results (classification rates in %) for the case of D4-D6 datasets.

<table>
<thead>
<tr>
<th>Feature Vectors</th>
<th>COIL-R (D4)</th>
<th>COIL-S (D5)</th>
<th>COIL-RN (D6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$HOG - LBP_{8,1}^{u^2}$ [45]</td>
<td>89.16</td>
<td>36.00</td>
<td>27.50</td>
</tr>
<tr>
<td>$HOG - LBP_{24,3}$</td>
<td>32.50</td>
<td>98.00</td>
<td>33.33</td>
</tr>
<tr>
<td>$LBP - HF$ [46]</td>
<td>90.00</td>
<td>32.00</td>
<td>30.00</td>
</tr>
<tr>
<td>$Mb - LBP_{T}^{r^6}$</td>
<td><strong>100.00</strong></td>
<td><strong>100.00</strong></td>
<td><strong>95.00</strong></td>
</tr>
</tbody>
</table>

One can conclude to the same by examining the classification rates of Table 10. The performance of the proposed $Mb - LBP_{T}^{r(order)}$ descriptor is significantly better in all the datasets especially in those where the scaling transformation is applied, where all the competitive descriptors are not invariant.
Concluding, it can be stated that the proposed Mb-LBP descriptor can be used as a surrogate not only of the conventional LBP, but also of other descriptors from the literature such as HOG, HOG-LBP and LBP-HF.

6. Conclusion

A new methodology for the generation of high discriminative feature vectors was proposed in the previous sections. The introduced descriptor consists of two main processing modules, the momentgram construction and the application of the LBP operator on it. This two-step technique is making use of the main advantages of the image expansion on its moments or moment invariants and the LBP local descriptor. As a result an enhanced LBP histogram, which is invariant under common geometric transformations (translation, rotation and scaling) of the described object, enclosing global as well as local information of the described object, is derived.

Moreover, the proposed descriptor gives further flexibility to the designer of a pattern recognition system, since by selecting the appropriate moment family along with the moment type in the momentgram construction, improved classification accuracy can be achieved relative to a specific application.

The experimental study shows the improvement to the discrimination ability of the LBP descriptor when applied on a momentgram instead of the original image, by a significant factor. Besides, it enables the application of the LBP operator on more pattern recognition problems where translation, rotation and scaling invariances are necessary.

The proposed Mb-LBP descriptor is highly competitive, as a comparison study with other descriptors from the literature has highlighted its outperformance in all the datasets under examination.

Future research has to be scheduled in the direction of improving the classification performance of the Mb-LBP descriptors derived by the SMgrams, while the application of
the proposed descriptor to several problems in the field of artificial intelligence (biometrics, biomedical image processing, robotic vision, etc.) could establish it as an efficient pattern descriptor.

References


