LOCAL ZERNIKE MOMENTS: A NEW REPRESENTATION FOR FACE RECOGNITION

Evangelos Saryyanidi\textsuperscript{a}, Volkan Da˘glı\textsuperscript{b}, Salih Cihan Tek\textsuperscript{c}, Birkan Tunç\textsuperscript{b}, Muhittin Gökmen\textsuperscript{c}

\textsuperscript{a}Department of Control Engineering, Istanbul Technical University, Istanbul. saryyanidi@itu.edu.tr
\textsuperscript{b}Informatics Institute, Istanbul Technical University, Istanbul. \{tuncbi, daglivo\}@itu.edu.tr
\textsuperscript{c}Department of Computer Engineering, Istanbul Technical University, Istanbul. \{tek, gokmen\}@itu.edu.tr

ABSTRACT

In this paper, we propose a new image representation called Local Zernike Moments (LZMs) for face recognition. In recent years, local image representations such as Gabor and Local Binary Patterns (LBP) have attracted great interest due to their success in handling difficulties of face recognition. In this study, we aim to develop an alternative representation to further improve the face recognition performance. We achieve this by utilizing Zernike Moments which have been successfully used as shape descriptors for character recognition. We modify global Zernike moments to obtain a local representation by computing the moments at every pixel of a face image by considering its local neighborhood, thus decomposing the image into a set of images, moment components, to capture the micro structure around each pixel. Our experiments on FERET face database reveal the superior performance of LZM over Gabor and LBP representations.

Index Terms— Face recognition, feature extraction, Zernike moments, regional histograms

1. INTRODUCTION

Robust 2D face recognition is one of the most active research areas in computer vision. Despite the impressive advances in recent years, it still remains as a challenging problem, mainly because of the amount of possible variations in the appearance of a face that can result from illumination conditions, viewpoint and facial expressions. Robust face recognition requires robust features with the ability to deal with such variations. Most face recognition approaches adopt feature extraction methods involving local shape descriptors like Gabor filtering-based features\cite{1, 2} and LBP\cite{3} to handle some of these variations. In this study, we aim to develop a simple yet powerful representation as an alternative to these representations by utilizing moments.

Image moments are regarded as well established shape descriptors and have been frequently used for content based retrieval and pattern recognition tasks. One of the most inspiring early studies illustrating the potential of image moments is the study of Hu\cite{4}, which shows the invariance of these variations. In this study, we aim to develop a simple yet robust feature representation for face recognition. The moments are calculated at the whole image into non-overlapping subregions and then extracting features from each of these subregions using pseudo ZMs. In this paper, we present a novel representation based on local Zernike moments (LZMs) and demonstrate the superiority of this representation for face recognition. The moments are calculated at every pixel by considering its local neighborhood. A complex moment image of equal size as the input image is obtained for each moment component resulting in a set of images, as shown in Figure 1. Then, each moment image is divided to non-overlapping subregions and phase-magnitude histograms are extracted from the complex output of each moment at each subregion. The final face representation is obtained by concatenating the extracted phase-magnitude histograms.

2. GLOBAL ZERNIKE MOMENTS

ZMs of an image are defined as the projection of the image onto a orthogonal set of polynomials called Zernike polynomials. The Zernike polynomials are defined as

\[ V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{jnm}\theta \]

where \( n \) is the order of the polynomial and \( m \) is the number of iterations such that \( |m| < n \) and \( n - |m| \) is even. The radial polynomials \( R_{nm} \) are calculated as

\[ R_{nm}(\rho) = \frac{n-|m|}{2} \sum_{s=0}^{\frac{n+|m|}{2}} \frac{(-1)^s \rho^{n-2s}}{s!(\frac{n-|m|}{2} - s)!} \]

The ZMs of a digital image \( f(i, j) \) are calculated as

\[ Z_{nm} = \frac{n+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) V^*(\rho_{ij}, \theta_{ij}) \Delta x_i \Delta y_j, \]

where \( x_i \) and \( y_j \) are the image coordinates mapped to the range \([-1, 1]\], \( \rho_{ij} = \sqrt{x_i^2 + y_j^2} \), \( \theta_{ij} = \tan^{-1} \frac{y_j}{x_i} \), and \( \Delta x_i = \Delta y_j = 2/N \sqrt{2} \). The terms \( \frac{n+1}{\pi} \), \( \Delta x_i \), and \( \Delta y_j \) are constant and they will be ignored in the remainder of this paper for clarity.

The global form of ZMs expressed in (3) was used for face recognition in\cite{7}. The authors utilized both the phase and magnitude information. However the results reported on the FERET face database\cite{10} are not promising enough to demonstrate the real potential of ZMs.

To further investigate the potential of ZMs, we have divided the whole image into non-overlapping subregions and calculated the
ZMs of each subregion separately. Recognition rates achieved by this approach on the FaFb subset of FERET face database are given in Table 1 for various numbers of subregions. Results indicate that such a partitioning increases the performance up to a point.

<table>
<thead>
<tr>
<th># subregions</th>
<th>1²</th>
<th>3²</th>
<th>5²</th>
<th>7²</th>
<th>9²</th>
<th>11²</th>
<th>13²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>59.0</td>
<td>71.4</td>
<td>78.2</td>
<td>78.3</td>
<td>79.5</td>
<td>78.5</td>
<td>76.3</td>
</tr>
</tbody>
</table>

Table 1. Recognition rates using global ZMs for different levels of partitioning.

3. LOCAL MOMENT TRANSFORMATION

Global moment descriptors must be applied to images containing explicit shape features like characters or fingerprints to get a useful description. In this study, we define a novel transformation to stimulate the shape characteristics of raw intensity images. The proposed transformation gives a new description of the image by extracting the shape characteristics out of the texture information contained in it. Resulting moment components prove to be very useful for capturing the micropatterns of the face.

3.1. Local computation of ZMs

The key difference between our approach and the global approaches is that we calculate moments around each pixel to obtain a new image representation through a transformation referred to as LZM transformation.

We derive moment-based operators \( V_{nm}^k \), which are \( k \times k \) kernels calculated using the equation \( V_{nm}^k(i,j) = V_{nm}(\rho_{ij}, \theta_{ij}) \). These operators are similar to 2D convolution kernels used for image filtering. The LZM transformation can be defined by these kernels as follows

\[
Z_{nm}^k(i,j) = \sum_{p,q=-k+1}^{k-1} f(i-p,j-q)V_{nm}^k(p,q). \tag{4}
\]

This transformation provides a rich image representation by successfully exposing the intensity variations around each pixel. The components of the representation (i.e., complex images corresponding to different moment orders) seem to be very robust to illumination variations as shown in Figure 1. The real and imaginary parts of an exemplar \( V_{nm}^k \) filter for size \( k = 3 \), \( n = m = 1 \) are shown in Figure 2.

The number of components depends on the moment order \( n \). We impose an additional constraint on the second parameter \( m \), such as \( m \neq 0 \) since the imaginary part of the kernels \( V_{nm}^k \) becomes zero when \( m = 0 \) and this is not a desirable behavior when the outcome of (4) is used to extract the phase-magnitude histograms as explained in Section 4.2. The number of effective moment components can be calculated through the following expression:

\[
K(n) = \begin{cases} 
\frac{n(n+2)}{4} & \text{if } n \text{ is even} \\
\frac{(n+1)^2}{4} & \text{if } n \text{ is odd}
\end{cases}.
\]

3.2. Face description from moment components \( Z_{nm}^k \)

The advantage of dividing images into subregions has been expressed in many face recognition studies. In this study, we follow the same approach and divide the moment components into subregions after the proposed LZM transformation is applied. A two-step partitioning is performed. First we divide the image to \( N \times N \) equal-sized blocks beginning from the top-left of the image. Then, we divide the image to \( N - 1 \times N - 1 \) blocks of the same size as the previous ones with a grid shifted half a block size from top-left. Hence, we have \( N^2 + (N-1)^2 \) subregions for each moment component. In order to increase the robustness against illumination variations further, each subregion is z-normalized.

We use histogram representation to combine the information gathered from different moment components. This representation has been shown to be very successful for face recognition [2, 3]. Figure 3 illustrates the whole process.

4. EXPERIMENTAL RESULTS

We have carried out extensive experiments to assess the performance of our approach. The complex output of ZMs is represented using phase-magnitude histograms. The performance of the method is substantially increased by applying the LZM transformation twice.

4.1. Experimental Setup

In our method, there are several parameters that need to be determined, such as the size of ZM-based operator \( k \), the moment order \( n \), the grid size \( N \times N \) and the number of bins in the histogram \( b \). After extensive experiments, we found that good performance is obtained when setting \( k = 7, n = 4, N = 10, b = 24 \). When the moment order, \( n \), is set to 4 the number of resulting moments, \( K \), is 6, corresponding to moment components \( Z_{11}^k, Z_{22}^k, Z_{33}^k, Z_{42}^k, Z_{34}^k \). The length of the final feature vector is \( (N^2 + (N-1)^2) \times b \times K = 26064 \).

All tests are performed on the FERET face database [10] using standard FERET evaluation protocol: a gallery set of 1196 images of 1196 subjects and four probe sets: FaFb (1195 images), FaFc
Fig. 3. An illustration of the proposed method. Only magnitude images are shown. During histogram construction both magnitude and phase values are considered.

(194 images), Dup1 (722 images) and Dup2 (234 images). The input images are normalized to have zero mean and unit variance after they are cropped, and their sizes are fixed to $130 \times 150$. The final feature vectors are compared using $L_1$ distance.

4.2. Local phase-magnitude histograms (H-LZM)

Several types of histograms can be employed to utilize the output of LZM transformation, including magnitude histograms (MHs) and phase-magnitude histograms (PMHs). Our experiments have led us to conclude that PMHs perform substantially better than MHs. The PMH of a moment component is extracted as follows: The angle interval of $[0, 2\pi]$ is divided into $b$ bins. Then, the magnitude value $|Z_{nm}(i,j)|$ at each pixel location $(i,j)$ is added to the bin corresponding to phase value $(\mathcal{Z}_{nm}(i,j))$ of the same pixel location.

The PMHs of moment components are extracted at each subregion as illustrated in Figure 3. Each local PMH is normalized to have unit norm, and all normalized local PMHs are concatenated to get the final feature vector. Most face recognition methods that involve partitioning use a weight map similar to the one in [3] to assign a different weight for each subregion. In this study we adopt a similar weighting procedure. We use the weight map proposed in [3] by simply interpolating it to fit to our grid size. Although it is possible to determine the optimum weights for each subregion of each moment component, this was left for future work.

Furthermore, we perform a different, pixel-wise weighting inside each subregion when computing the PMHs. We use a gaussian weighting kernel of same size as a subregion with $\sigma = 8$. We multiply the magnitude of each pixel with the corresponding kernel weight before adding it to the relevant histogram bin.

The performance of the proposed method is given in Table 2 with the abbreviation H-LZM. Although the results are acceptable, they do not show the full potential of the LZMs. The reason of the relatively low performance is discussed in the next section, where a solution to increase the performance is also described.

4.3. Cascaded LZM transformation (H-LZM$^2$)

In general, image moments perform better on images containing explicit shape information like fingerprints or character images. As indicated before, the proposed LZM transformation gives the description of shape in a micro scale around each pixel. In the previous experiments, we have tried to obtain statistics from this new description by calculating PMHs. Since the resulting description exhibits shape information, it is possible to obtain more useful statistics by using another feature extraction method before computing the histograms.

Motivated by this property, we modify the method explained so far by replacing the input intensity images in Figure 3 with LZM transformed images. In other words, we apply the LZM transformation twice: First to stimulate the local shape characteristics; second, to describe the local shape statistics of the transformed images. The moment components obtained by the cascaded LZM transformation are shown in Figure 4, while the block diagram of the cascaded transform is shown in Figure 5.
directly leads to better results and therefore report the results only for these inputs in Table 2, where H-LZM²-I, H-LZM²-R, and H-LZM²-IR indicate results with imaginary, real, and combined inputs, respectively. As expected, preprocessing the input image to extract the shape characteristics increases the performance substantially. The moment orders used in this approach are \( n_1 = 4 \) for the first stage, and \( n_2 = 4 \) for the second stage. Kernel sizes are \( k_1 = 5 \) and \( k_2 = 7 \), respectively. The length of the final feature vector is \( (N^2 + (N - 1)^2) \times b \times K_1 \times K_2 \times 2 = 312768 \) when combining real and imaginary parts.

### Table 2. Recognition rates with different settings.

<table>
<thead>
<tr>
<th>Method</th>
<th>FaFb</th>
<th>FaFc</th>
<th>Dup I</th>
<th>Dup II</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-LZM</td>
<td>95.0</td>
<td>87.1</td>
<td>73.8</td>
<td>70.9</td>
</tr>
<tr>
<td>H-LZM (Weighted)</td>
<td>97.5</td>
<td>95.4</td>
<td>78.5</td>
<td>76.1</td>
</tr>
<tr>
<td>H-LZM²-I</td>
<td>96.2</td>
<td>97.9</td>
<td>79.6</td>
<td>76.9</td>
</tr>
<tr>
<td>H-LZM²-I (Weighted)</td>
<td>98.7</td>
<td>99.5</td>
<td>83.9</td>
<td>82.5</td>
</tr>
<tr>
<td>H-LZM²-R</td>
<td>96.2</td>
<td>96.9</td>
<td>77.4</td>
<td>73.5</td>
</tr>
<tr>
<td>H-LZM²-R (Weighted)</td>
<td>98.7</td>
<td>99.0</td>
<td>83.2</td>
<td>81.2</td>
</tr>
<tr>
<td>H-LZM²-IR</td>
<td>96.3</td>
<td>97.9</td>
<td>79.9</td>
<td>76.5</td>
</tr>
<tr>
<td>H-LZM²-IR (Weighted)</td>
<td>98.7</td>
<td>99.5</td>
<td>84.8</td>
<td>82.5</td>
</tr>
</tbody>
</table>

### Table 3. Recognition rates compared to other methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>FaFb</th>
<th>FaFc</th>
<th>Dup I</th>
<th>Dup II</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP [3]</td>
<td>93.0</td>
<td>81.0</td>
<td>61.0</td>
<td>50.0</td>
</tr>
<tr>
<td>LBP (Weighted) [3]</td>
<td>97.0</td>
<td>79.0</td>
<td>66.0</td>
<td>64.0</td>
</tr>
<tr>
<td>HGPP [2]</td>
<td>97.6</td>
<td>98.9</td>
<td>77.7</td>
<td>76.1</td>
</tr>
<tr>
<td>HGPP (Weighted) [2]</td>
<td>97.5</td>
<td>99.5</td>
<td>79.5</td>
<td>77.8</td>
</tr>
<tr>
<td>H-LZM²-IR</td>
<td>96.3</td>
<td>97.9</td>
<td>79.9</td>
<td>76.5</td>
</tr>
<tr>
<td>H-LZM²-IR (Weighted)</td>
<td>98.7</td>
<td>99.5</td>
<td>84.8</td>
<td>82.5</td>
</tr>
</tbody>
</table>

### 6. REFERENCES


