Facial Image-based Gender Classification using Local Circular Patterns

Chen Wang, Di Huang, Yunhong Wang, Guangpeng Zhang
IRIP Lab, School of Computer Science and Engineering, Beihang Univ., Beijing, 100191, China
chen.wang.ec@gmail.com; dhuang@buaa.edu.cn; yhwang@buaa.edu.cn

Abstract

Gender is one of the most important demographic attributes of human beings, and recently automatic face-based gender classification has received increasing attentions due to its wide potential in many useful applications. To address such an issue, in this paper, we propose a novel variant of Local Binary Patterns (LBP), namely Local Circular Patterns (LCP). LCP makes use of clustering-based quantization instead of the binary coding strategy of the LBP operator, leading to an improvement in discriminative power. Meanwhile, thanks to the nature property of clustering-based quantization, LCP is more robust than LBP to noise. Experiments are carried out on the FERET database and the classification accuracy is up to 95.36%, clearly highlighting the effectiveness of the proposed method.

1 Introduction

Gender is one of the most important demographic attributes of human beings. People can easily recognize the gender of others through facial appearances; however, it is still a non-trivial task for computers. Recently, automatic face-based gender classification has received increasing attentions since it plays an important role in various applications, e.g. facial image analysis, human computer interaction, video and image retrieval etc.

During the last decade, research on face-based gender classification has grown up rapidly since it emerged. The approaches can be approximately divided into two categories: global image based and local texture based ones. The former regards the raw facial image as an entire input and uses dimensionality reduction techniques (down-sampling or subspaces) to process the image in order to finally feed the classifier. For instance, O’Tool et al. [9] applied Principal Components Analysis (PCA) to 130 grayscale facial images of size $512 \times 512$, and obtained a peak correct rate of 93.8% in a 20-dimensional subspace. Moghaddam and Yang [7] introduced a gender classification method, in which facial images were firstly down sampled into the size of $21 \times 12$, and then classified by support vector machine (SVM). An error rate of 3.38% was achieved on a dataset containing 1755 samples. The latter adopts local descriptors to represent texture changes between male and female, such as the positions of eyebrows and the shape of jawbone or jaw line [1]. Lian et al. [4] extracted Gabor features from a set of facial feature points, and used SVM for classification. Around 94% accuracy was reported on a dataset of 1991 images. As haar-like features are of low computational complexity, they were used in combination with Ada-Boost for ethnicity and gender classification [11]. Lu and Lin [5] also used haar-like features to train the classifier (AdaBoost+SVM), and tested their approach on 518 frontal facial images of size $24 \times 24$ from the FERET database, reporting an accuracy of 80%. Based on the study of these above tasks, similar to the conclusions achieved in face recognition, local texture based methods prove its superiority to global based ones, because they can better describe the differences between male and female, and are generally insensitive to pose and illumination variations.

Due to its great success in 2D face recognition, LBP as a powerful local texture descriptor has been investigated in gender classification [8], and the state-of-the-art performance was achieved. However, LBP still has limitations, and one of the most critical aspects lies in that LBP only extracts the sign of differences between neighboring pixels while ignoring the magnitude, leading to the lost in discriminative power. In this paper, we propose a method called Local Circular Patterns (LCP for simplicity), which is an improvement of traditional LBP, mainly because rather than explicitly quantizing sign or magnitude components of local texture patterns, it encodes the patterns through clustering. At the same time, due to the nature property of clustering quantization, LCP tends to be more robust than LBP to noises. Experiments are carried out on FERET, one of the most comprehensive datasets for facial analysis, and a classification accuracy of 95.36% is achieved, which clearly highlights the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section 2 introduces the proposed LCP descriptor in detail. Ex-
perimental results are presented in section 3, and section 4 concludes the paper.

## 2 LCP based Facial Feature Extraction

Since LCP is a variant of LBP, we will first recall the basic concepts of the original LBP, and then introduce the proposed LCP and LCP based facial representation.

### 2.1 LBP descriptor

The original LBP operator simply thresholds a $3 \times 3$ neighborhood by the value of the central pixel, and the signs of thresholded neighbors can form a binary number. The binary number is then transformed into a decimal number, which is treated as the label of the central pixel (Fig. 1(a)). We call this quantization scheme as binary quantization in the following. The histogram of the labels in a region is often used as a texture descriptor.

The basic LBP was later extended to uniform pattern [8]. A uniform pattern contains at most two transitions between 0 and 1, each of which occupies a single label, while the non-uniform ones are all accumulated into another single label. It was stated that these uniform patterns are fundamental patterns providing a vast majority of all $3 \times 3$ patterns in the observed textures [8].

### 2.2 Clustering-based quantization

As depicted in Fig. 2, for one pixel $t$ with its 8 neighboring ones $\{t_1, t_2, \ldots, t_8\}$ located on the circle with a radius value of 1, the local circular pattern ($LCP_{8,1}$) $p$ is defined as $p = (t_1 - t, t_2 - t, \ldots, t_8 - t)^T$. Given $N$ training local circular patterns $p_i, i = 1, 2, \ldots, N$, K-means clustering is performed to find a partition $C = \{c_1, c_2, \ldots, c_k\}$ by minimizing the following function,

$$ J(C) = \sum_{i=1}^{k} \sum_{p_j \in c_i} D(p_j, \mu_i) $$

where $D$ represents the distance function, and $\mu_i$ is the center of $c_i$. Then a new local circular pattern $p'$ can be quantized into the nearest cluster center.

$$ l(p') = \arg\min_i D(p', \mu_i) $$

K-means is a greedy algorithm that can only converge to a local minimum, but recent study has shown that K-means can converge to the global optimum with a large probability when clusters are well separated [6].

### 2.3 LCP based facial representation

Like LBP, for facial representation, the facial image is divided into some rectangular areas where LCP based histograms are extracted, and these histograms are further concatenated into a single one to encode both local and global information (Fig. 4).
2.4 Multi-scale extension

The LBP operator was later extended with different sizes of local neighborhoods to deal with various scales. The local neighborhood is defined as a set of sampling points evenly spaced on a circle that is centered at the pixel to be labeled. The sampling points that do not fall exactly on the pixels are expressed by using the bilinear interpolation technique; thus allowing any radius value and any number of points. Fig.1(b) shows different LBP neighborhoods. The notation \( (P, R) \) denotes the neighborhood of \( P \) sampling points on a circle with radius \( R \). Using the same protocol, LCP can also handle different sampling points and scales so that we can incorporate more distinctive information to represent texture variations between male and female.

3 Evaluation

3.1 Experimental settings

FERET [10] is a standard dataset used for face analysis system evaluation. Our experiments are carried out on a subset of the FERET database, which is made up of 3585 frontal facial images from 1203 subjects. Training and test sets are randomly divided with no subjects in common, and these two sets are thus independent. Male and female images are evenly separated in the two sets: the training set contains 1774 images from 601 subjects (246 women and 355 men), whilst the test set consists of 1811 images from 602 subjects (246 women and 356 men). The positions of two pupils of each facial image provided by the dataset are utilized to crop faces. Then an average mask is used to eliminate non-face regions and segment face out. Finally images are normalized to the size of 120 × 120 pixels. Fig.5 shows a few examples of normalized facial images.

45 facial images from training set are randomly chosen for K-means clustering to produce cluster centers. The number of clusters is set to 59 for 8 neighboring points and 243 for 16 neighboring points, so as to compare to LBP with the same dimensionality. Then both LBP and LCP features are extracted for different scales: (8, 1), (16, 2) and (16, 3). Each code image is divided into small rectangular regions of size 12 × 12 pixels.

SVM with a linear kernel is used for the classification. The implementation is achieved with source code of the LIBSVM developed by Chang and Lin [2].

3.2 Parameter selection

To set a proper distance for clustering, we randomly selected 20 samples from the FERET database and extracted the local circular patterns as plotted in Fig. 6. The results obtained with \( L^2 \) distance and \( L^1 \) distance are different, the results of \( L^2 \) distance are more evenly spaced than those of \( L^1 \) distance, while \( L^1 \) distance results focus on the concentrations of the data and pay less attention to the outer of the distribution. So the cluster centers obtained by \( L^1 \) distance should be more representative.

On the other hand, we compare computational cost for each distance metric. Experiments are carried out on a PC with Intel Core i3 CPU using matlab. With 314960 training local circular patterns, the computational time taken by K-means clustering with \( L^2 \) distance repeating 10 times is 12233 seconds, while that for \( L^1 \) distance is only 600 seconds. Therefore \( L^1 \) distance is computationally more efficient than \( L^2 \) distance as expected. Thus we use \( L^1 \) distance in our following experiments.

3.3 Experimental Results

Table 1 shows the performance of different local texture features and we can see that LBP and LCP achieve better results than Gabor features. Table 2 demonstrates the accuracies of LBP and LCP with different scales, and it can be seen that the proposed LCP outperforms LBP at all parameter settings, highlighting the improvement in discriminative power to distinguish males from females. Table 3 displays the multi-scale results of LCP and LBP combining the ones of individual scale ((8, 1), (16, 2) and (16, 3)) by using different score level fusion.
strategies. In Table 3, we can see that on the one hand, for both LCP and LBP, multi-scale strategy is an effective way to further improve the classification accuracy. On the other hand, the geometric average based fusion strategy reaches the best performance, and the final accuracy is up to 95.36%.

It should be noted that since the tasks on facial image based gender classification were either evaluated on a private dataset or a subset of a public one, it is difficult to compare the figures fairly. Our experiments are direct evidences to show that the LCP based facial representation is more effective than LBP based ones. In addition, to compare with the other work, we use a more difficult experimental protocol, i.e. a larger test set (1811 images instead of 518 [5]) and a smaller ratio between training and testing samples (0.98 instead of 5.78 [7]).

### Table 1. Performance comparison for different local texture based methods.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Dimensionality</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabor</td>
<td>10240</td>
<td>91.99%</td>
</tr>
<tr>
<td>LBP$_{8,1}$</td>
<td>5900</td>
<td>93.26%</td>
</tr>
<tr>
<td>LCP$_{8,1}$</td>
<td>5900</td>
<td>94.64%</td>
</tr>
</tbody>
</table>

### Table 2. Comparison between LBP & LCP.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Dimensionality</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP$_{8,1}$</td>
<td>5900</td>
<td>93.26%</td>
</tr>
<tr>
<td>LCP$_{8,1}$</td>
<td>5900</td>
<td>94.64%</td>
</tr>
<tr>
<td>LBP$_{16,2}$</td>
<td>8748</td>
<td>91.66%</td>
</tr>
<tr>
<td>LCP$_{16,2}$</td>
<td>8748</td>
<td>93.48%</td>
</tr>
<tr>
<td>LBP$_{16,3}$</td>
<td>8748</td>
<td>91.77%</td>
</tr>
<tr>
<td>LCP$_{16,3}$</td>
<td>8748</td>
<td>93.48%</td>
</tr>
</tbody>
</table>

### Table 3. Comparison between Multi-scale LBP & LCP by using different score level fusion schemes.

<table>
<thead>
<tr>
<th>Fusion mode</th>
<th>Feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic average</td>
<td>LBP</td>
<td>93.65%</td>
</tr>
<tr>
<td></td>
<td>LCP</td>
<td>95.31%</td>
</tr>
<tr>
<td>Geometric average</td>
<td>LBP</td>
<td>93.71%</td>
</tr>
<tr>
<td></td>
<td>LCP</td>
<td>95.36%</td>
</tr>
<tr>
<td>Min rule</td>
<td>LBP</td>
<td>91.44%</td>
</tr>
<tr>
<td></td>
<td>LCP</td>
<td>93.65%</td>
</tr>
<tr>
<td>Max rule</td>
<td>LBP</td>
<td>92.27%</td>
</tr>
<tr>
<td></td>
<td>LCP</td>
<td>93.65%</td>
</tr>
</tbody>
</table>

### 4 Conclusion

In this paper, we have presented a novel variant of traditional LBP, called clustering-based quantization of LCP. Rather than explicitly quantizing the sign or magnitude components of local patterns, it quantizes the local patterns through clustering. As a result, LCP possesses a more powerful discriminative ability than LBP. Moreover, it tends to be more robust to noises. Experiments were carried out on the FERET database, and the classification accuracy was up to 95.36%, clearly highlighting the effectiveness of the proposed method.

In our future work, we will study the importance of different facial regions by machine learning techniques to investigate the possible increase in performance.

### Acknowledgments

This work is in part jointly funded by the French research agency, Agence Nationale de Recherche (ANR) and Natural Science Foundation of China (NSFC), in the 3D Face Analyzer project (grant ANR 2010 INTB 0301 01; grant NSFC 61061130560); the 3D Face Interpreter project supported by the LIA 2MCSI lab between the group of Ecoles Centrales and Beihang University; the Fundamental Research Funds for the Central Universities.

### References


