SPATIAL CO-OCCURRENCE OF LOCAL INTENSITY ORDER FOR FACE RECOGNITION

Xianbiao Qi1, Yi Lu2, Shifeng Chen2, Chun-Guang Li1 Jun Guo1

1Beijing University of Posts and Telecommunications, Beijing 100876, P.R. China
2Shenzhen Key Laboratory for Computer Vision and Pattern Recognition, Chinese Academy of Sciences
qixianbiao@gmail.com, {yi.lu, shifeng.chen}@siat.ac.cn, {lichunguang, guojun}@bupt.edu.cn.

ABSTRACT
In this paper, we propose a simple but effective spatial co-occurrence of local intensity order (CoLIO) feature for face recognition. Local intensity order (LIO) is robust to illumination variance. Spatial co-occurrence of LIO not only preserves great invariance to illumination, but also greatly enhances the discriminative power of the descriptor as CoLIO well captures the correlation between locally adjacent regions. The proposed feature has been successfully applied to two widely used face databases including AR [1] and LFW [2]. Superior performance on these two databases fully demonstrates the effectiveness of the proposed feature. Meanwhile, the extremely fast extraction speed makes the proposed feature practically useful.

Index Terms— Face Recognition, Spatial Co-occurrence, Local Intensity Order.

1. INTRODUCTION
In the past few decades, face recognition has attracted wide attention because of its great potentials in computer vision. However, in real face recognition applications, there always exist kinds of image transformations which make the face recognition problem extremely challenging. As shown in Fig. 1, real face images always contain high variance due to illumination, expression, pose, occlusion and different resolutions. Therefore, it becomes crucial to design effective features to bridge the gap between the same faces from different images.

There are two kinds of features for face recognition, the handcrafted features and the learning-based features. For the handcrafted features, local binary pattern (LBP) [3] is the most widely used for face representation. By encoding the circular binary sequence between the center point and its neighbors, LBP effectively captures the micro-patterns of the images, such as edges, spots, lines and flat regions. Meanwhile, LBP is robust to illumination as the sign of the difference between the center and its neighbors is invariant to monotonic intensity variance. Since LBP was first proposed, a lot of LBP variants have been proposed. For example, Local Gabor Phase Patterns (LGBP) [4] was proposed to capture multiple resolution and orientation information.

Many learning-based features [5] [6] [7] have been proposed in the past few years. Different from the handcrafted approaches, the learning-based methods have to learn a dictionary in advance. The dictionary is especially important for learning-based methods and always greatly affects the final recognition performance. Meanwhile, the time of feature extraction is proportional to the dictionary size. The learning-based features try to learn a representative dictionary from the training data. However, they always don’t have some satisfactory invariance to image transformations as do the handcrafted features. In contrast, the manually well-designed features usually have great invariance to image transformations. However, the discriminative power of the hand-crafted features is limited by their small amount of patterns. Thus, designing better handcrafted features could enhance the discriminative power of the descriptors and preserve great trans-
Recently, local descriptors based on local intensity order (LIO) have been proposed in [8] and have shown great performance in the experimental evaluations. Especially, the local intensity order based descriptors show great invariance to the illumination. Meanwhile, the descriptors are computationally efficient. They just need several binary comparison operations. However, the discriminative power of local intensity order patterns is limited since it just describes the relationship among the four neighbors around the center point.

Spatial co-occurrence of local features could enrich the descriptive power and boost the discriminative power of the descriptors. For instance, in [9], Qi et.al. proposed a pairwise rotation invariant co-occurrence local binary pattern (CoLBP) which considers the relationship between the two spatially adjacent LBP features. They proposed an effective strategy to encode the relative angle information between two points. The proposed CoLBP feature has shown great performance on texture and material classification tasks.

We summarize our contributions in this paper as follows:

- We propose to apply local intensity order (LIO) to face recognition. Different from the traditional LBP, LIO focuses on capturing the relationship between the neighbors, while LBP focuses on describing the relationship between the center point and its neighbors. The relative orders are stable as the intensity changes due to illumination variation are always monotonic. Therefore, LIO is robust to illumination variation.

- To enhance the discriminative power of the descriptor, we propose to capture spatial co-occurrence of local intensity order (CoLIO). The locally adjacent patterns always have strong correlations. It means that for the same person in two different images, it is much more possible to observe two patterns co-occurring in the same regions. As shown in Fig. 3, there are a lot of same co-patterns appearing in the corresponding positions of both faces. Therefore, effectively capturing such correlation will sharply enhance the discriminative power of the descriptors.

2. SPATIAL CO-OCCURRENCE OF LOCAL INTENSITY ORDER

2.1. Local Intensity Order

An effective representation of faces is extremely important for face recognition due to the existing kinds of image transformations in the real world. The illumination, expression and pose variations will greatly change the local intensity values. However, fortunately, there always exist some invariant properties in the local regions. Thus, when designing effective features, we should pay more attention to capturing such invariant properties of the local regions.

The underlying hypothesis of LBP is that the sign of the difference between two locally adjacent pixels is not affected by illumination variation. LBP focuses on describing the relationship between the center and its neighbors by encoding a circular binary sequence. But, LBP ignores the relationship between the neighbors. Similar with LBP, LIO assumes that the local intensity order is robust to illumination variation. Instead of focusing on capturing the relationship between the center points and their neighbors, LIO focuses on capturing the relationship among the neighbors. Here, the LIO feature defined on four points could be defined as follows:

\[ f(A) = \text{Order}(A_1, A_2, A_3, A_4). \]  \hfill (1)

Where \( A_1, A_2, A_3 \) and \( A_4 \) are the four neighbors of the point \( A \) as shown on the left side of Fig. 2, and \( \text{Order}() \) is the order function.

The LIO feature defined on four points has 24 \((4! = 24)\) patterns in total. As shown on the left side of Fig. 2, the descending sorted order of \( (A_1 = 62, A_2 = 65, A_3 = 58, A_4 = 65) \)
and they belong to the same person. On the right side of Fig. 2, we show some same co-patterns. Therefore, capturing the spatial correlation using co-occurrence is necessary.

For the point $A$, we could ascertain the point $B$ according to the predefined template. The template could be defined as the point $B$ is 2 pixels away from the point $A$ along the horizontal axis as shown on the right side of Fig. 2. Thus, according to the template, we could calculate the co-patterns for each point except for the points on the boundary. The overall CoLIO patterns contain 576 patterns. Because of the high correlation in local regions, some LIO patterns don’t co-occur. Thus, the final block-based CoLIO histogram is sparse.

### 2.3. Face Representation Based on CoLIO

In this subsection, we present the face representation based on CoLIO in detail. The detailed framework has been shown in Fig. 4. The whole process consists of the following five steps.

**Step 1:** For the input image, the LIO feature is extracted for each point except for the ones on the boundary. The computational cost for LIO is extremely small as LIO doesn’t need any interpolations like LBP, and it just needs six binary comparisons for each point.

**Step 2:** When the LIO patterns for each point are obtained, we could calculate the CoLIO patterns for each point according to the predefined template. The CoLIO feature has 576 patterns. For a lot of learning-based features [5][6], they always use the codebook size 512 or 1024.

**Step 3:** Following the traditional face representation methods [11][5][6], we divide the face into blocks. Since the face images are aligned, the block-wise face representation could better reflect the structure information of the faces. Only the faces which have better correspondences in local structures will have higher similarity scores.

**Step 4:** The histogram statistics of CoLIO patterns are calculated in each block. The histograms of each block are concatenated to form the final face representation.

![Fig. 4. The framework of face representation based on CoLIO.](image)
3. EXPERIMENTS

3.1. Datasets and Implemental Details

**AR Database:** The AR database contains 126 individuals and over 4000 frontal images. For each individual, 26 pictures are taken in two different sessions with 13 images each session. Follow [12] [13], we use the images of 120 individuals (65 men and 55 women). The facial portion of each image is cropped out and then resized to 100*80. Fig. 5 shows 13 samples from the first session of one individual. For this database, we apply the same face representation setup as [12]. The original gray images are used. The image is divided into 5*3 windows with each block 20*27.

**Labeled Faces in the Wild (LFW) Database:** The LFW database [2] was recently introduced as a benchmark dataset for face verification in the unconstrained environments. It contains 13233 images of celebrities. The LFW database consists of ten subsets with 300 intra-personal/extra-personal pairs in each subset. At each test, one subset is used for test and the rest nine subsets are used for training. The final average recognition performance is reported. The LFW gray images automatically aligned by Wolf et al. [14] are cropped to 114*80 around the center. DoG filters with \( \sigma_1 = 2.0 \) and \( \sigma_2 = 4.0 \) are used to obtain the DoG images. In Fig. 6, we show some cropped DoG image samples. Removing several boundary pixels, the image is divided into 10*6 blocks.

**Classifier:** For the AR dataset, the “nearest neighbor” classifier is applied, and \( \chi^2 \) distance measurement is used. For the LFW dataset, for the direct evaluation on the original feature space, we also use \( \chi^2 \) distance measurement and simple threshold to classify. The threshold is choosen as it best classifies the other nine training splits. Thus, we will also evaluate the performance of CoLIO using whiten PCA (WPCA) on the LFW dataset.

**Computational Cost:** The extraction time for the proposed CoLIO feature is extremely short. The extraction of CoLIO doesn’t need any interpolation operators like LBP and just needs six binary comparisons. It is also much faster than the learning-based features which are very time consuming in codebook assignment. On a desktop computer with dual-core 2.8G CPU, our matlab implementation of CoLIO takes about 3.7 milliseconds to process an image of 114*80. LBP1 takes about 5.9 milliseconds to process an image.

3.2. Comparison of Different Co-occurrence Configurations

Here, we evaluate the performance of different co-occurrence configurations on the AR and LFW. We denote the five features as follows: (LIO): the single point LIO; (CoLIO_X1): CoLIO along the horizontal axis with 1-pixel distance between the point \( A \) and \( B \); (CoLIO_X2): CoLIO along the horizontal axis with 2-pixels distance; (CoLIO_Y1): CoLIO along the vertical axis with 1-pixel distance; (CoLIO_Y2): CoLIO along the vertical axis with 2-pixels distance.

On the AR database, instead of using subsets like [12] [13], we conduct experiments on all 120 persons with 26 faces each individual. Here, we design eight groups of experiments. For all eight setups, the rest images outside the training samples will be used for test. The eight experimental configurations are “C1: The first image in session 1 used for training and the rest for testing”, “C2: The first four images in session 1 used for training”, “C3: The first seven images in session 1 used for training”, “C4: The first four images in session 1 used for training”, “C5”, “C6”, “C7”, “C8” use the corresponding training samples from session 2 as “C1”, “C2”, “C3”, “C4”.

For the LFW database, we just use the original features (No PCA) to evaluate 5 different configurations.

---

1http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab
The experimental results of these five features on AR database are shown in Tab. 1, and the results on the LFW database are shown in Fig. 7. From Tab. 1, we could obtain the following observations. Firstly, these four types of co-occurrence features sharply outperform the single point LIO feature. Specifically, for the configuration “C1”, CoLIO_Y2 sharply outperforms the single LIO feature by about 14%. The huge improvement proves that our proposed co-occurrence scheme could greatly boost the discriminative power of the features. Secondly, we could find that the scheme of point B being 2 pixels away from point A always achieves better performance than the scheme of just 1-pixel distance. Finally, we can see that on the AR database, the co-occurrence scheme along the horizontal axis achieves similar performance with the vertical co-occurrence scheme.

![Fig. 7. The ROC curve for five different configurations on the LFW database without using whiten PCA.](image)

### 3.3. Comparison with State-of-the-art Methods

**Experiments on the AR Database:** Here, we compare CoLIO with some related state-of-the-art methods. We divide the experiments on the AR database into two parts. First, we totally follow the experimental setup of [12] [13] and use the first 7 images of session 1 to conduct the experiments. It contains 840 in total. In the second part, to further evaluate the proposed feature, we conduct the evaluation on larger data (120*26=3120 images) and compare CoLIO with LBP.

The 7 images from session 1 contain (a) neutral expression, (b) smile, (c) anger, (d) scream, (e) left light on, (f) right light on and (g) both light on. For the experiment 1, (a, b, c, d) are used for training and (e, f, g) are used for testing. For the experiment 2, the training set and the test set are exchanged. The experimental results are shown in Tab. 2

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy(Experiment 1)</th>
<th>Accuracy(Experiment 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIO</td>
<td>69.0</td>
<td>68.89</td>
</tr>
<tr>
<td>DICA</td>
<td>99.72</td>
<td>99.75</td>
</tr>
<tr>
<td>LBP</td>
<td>98.33</td>
<td>98.75</td>
</tr>
<tr>
<td>HMLBP</td>
<td>98.89</td>
<td>96.67</td>
</tr>
<tr>
<td>CoLIO</td>
<td>99.72</td>
<td>99.58</td>
</tr>
</tbody>
</table>

It should be noted that, here, we use LBP_{8,2} and the true performance of LBP_{8,2} is higher than the implementation of [12]. In [12], they first divide the images into blocks and then compute the LBP features for each block. In this way, a lot of boundary pixels are abandoned. For example, for a block of 8*8, LBP_{8,2} just calculates the 4*4 region around the center and discard most boundary pixels. According to our framework (Fig. 4), the blocking is done after the LBP is calculated for the whole image.

Based on our implementation, multi-scale with LBP_{8,2}, LBP_{8,3} and LBP_{8,4} doesn’t further improve the performance of the single LBP_{8,2} on the AR database.

From Tab. 2, we could find that the proposed feature achieves the best performance on both experiments. First, the proposed feature significantly outperforms the PCA and ICA based methods. Meanwhile, it also outperforms the DICA method. DICA [13] is a training based method in feature extraction, so it requires more training samples to learn representative features. Small training samples will greatly affect the final recognition performance. So the performance decreases a lot from experiment 1 to experiment 2 when the training samples decrease from 4 to 3. However, the proposed CoLIO does not need any training stage and is not sensitive to the number of the training samples.

For experiment 1, 480 samples are used for training and 360 samples are used for testing. For the LBP feature, the accuracy of 98.33% means that LBP misclassifies 6 images of 360. However, CoLIO just misclassifies 1 image. It equals...
that CoLIO decreases the error rate by 83% compared with LBP. For experiment 2, LBP misclassifies 6 images and CoLIO just misclassifies 2 images. It means that CoLIO decreases the error rate by 67% compared with LBP.

Although the proposed feature has shown great improvement in two aforementioned experiments, the size of the used data is small. In order to fully evaluate CoLIO, we apply CoLIO on the all 120 persons with 26 samples for each individual and compare CoLIO with LBP under eight experimental setups as Sec. 3.2. The results is shown in Tab. 3.

### Table 3. Recognition Accuracy( % ) of different methods on the larger AR database with 3120 images.

<table>
<thead>
<tr>
<th>Methods</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>74.4</td>
<td>74.4</td>
<td>76.4</td>
<td>76.9</td>
<td>75.9</td>
<td>76.2</td>
<td>76.8</td>
<td>76.6</td>
</tr>
<tr>
<td>CoLIO</td>
<td>83.0</td>
<td>82.6</td>
<td>84.2</td>
<td>80.8</td>
<td>82.1</td>
<td>83.0</td>
<td>83.5</td>
<td>79.7</td>
</tr>
</tbody>
</table>

According to Tab. 3, CoLIO sharply outperforms LBP feature on all eight experiments. Specifically, using the configuration “C1”, CoLIO achieves 83.0% which outperforms the performance of LBP(74.4%) by about 8.6%.

### Experiments on the LFW Database: In this subsection, we conduct experiments on the LFW dataset and compare the proposed feature with some newly published state-of-the-art methods [11] [16] [17] [7]. Here, we first evaluate the performance on the original feature space. Then we use whiten PCA to remove the redundancy and obtain a compact descriptor. The experimetal results are shown in Tab. 4.

### Table 4. Face verification performance on the LFW.

<table>
<thead>
<tr>
<th>Methods(No WPCA)</th>
<th>Performance</th>
<th>Methods(WPCA)</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1-like [18]</td>
<td>64.21</td>
<td>V1-like/MKL [19]</td>
<td>79.35</td>
</tr>
<tr>
<td>POEM-HS(Flip) [17]</td>
<td>75.22</td>
<td>CoLIO (Ours)</td>
<td>81.55</td>
</tr>
<tr>
<td>LARK [16]</td>
<td>72.22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LHS [7]</td>
<td>73.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CoLIO (Ours)</td>
<td>74.45</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CoLIO(Flip) (Ours)</td>
<td>75.72</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

From Tab. 4, it is clear that the CoLIO feature outperforms some newly proposed state-of-the-art features, including LARK, POEM and LHS. Specifically, CoLIO achieves 74.45% which outperforms V1-like by about 6.3%, LARK by 2.2% and POEM by 0.8%. The LFW database contains significant pose variation. In the paper [17], Vu et al. indicated that flipping images could improve the verification performance. Using the introduced flipping method, CoLIO gets 75.72% which further improves original CoLIO by 1.3%.

Combined with whiten PCA, the performance of CoLIO improves from 74.45% to 81.55%. CoLIO(WPCA) outperforms other WPCA based features including LE [5] with 81.22% and POEM [17] with 81.13%.

### 4. CONCLUSION

A simple and effective feature is proposed for face recognition in this paper. The proposed feature is based on capturing spatial co-occurrence of local intensity order. Spatial co-occurrence could well describe the correlation between locally adjacent features. Spatial co-occurrence of two features could provide much more information than their individual occurrences. The proposed feature is successfully applied to two face datasets including AR and LFW face databases. Superior performance is achieved in both databases.

### 5. REFERENCES