Sparse residue for occluded face image reconstruction and classification

Jinghua Wang\textsuperscript{a}, Yong Xu\textsuperscript{b}, Jane You\textsuperscript{a}
\textsuperscript{a) Department of Computing, the Hong Kong Polytechnic University, Hong Kong}
jinghuawng@gmail.com
\textsuperscript{b) Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China}

Abstract

Occlusion problem is one of remaining challenges in face recognition. This work expresses an occluded image as the summation of a non-occluded image and a sparse occlusion. By solving a \( l_1 \) norm minimization problem, we isolate the sparse occlusion from the face image, and simultaneously reconstruct the image. The reconstructed image is same to the original one in most pixels. To classify an occluded image with unknown identity, we first linearly express it using the images of every person, and then make decision based on the residues of the expressions. This paper also presents the relationship between the proposed method and the popular methods. The experiments validate the feasibility of the proposed method.

1. Introduction

The methods that can handle the occlusion problem roughly fall into three groups: completion-based method, partition-based method, and holistic method. The completion-based method [1-2] detects the occluded part and rectifies them before classification. These methods need an additional procedure of occlusion detection and completion. The partition-based method partitions the images into several blocks and computes the local similarity [3-4]. The optimal partition of the face images is still unknown. More popular are the holistic methods [5-7] which directly manipulate the occluded images.

Linear regression-based classification (LRC) [5] is a class specific method. It represents the occluded image as a linear combination of face images from every person. The image is classified based on the representation error in terms of \( l_2 \) -norm. Sparse representation-based classification (SRC) [6] and collaborative representation based classification (CRC) [7] represent the occluded image as a linear combination of all the face images. The image is classified into the class that contributed most in the representation.

In this paper, we propose one more method for occluded face recognition reconstruction and recognition by minimizing the class-specific representation error in terms of \( l_1 \) -norm. Our method not only separates the sparsely occlusion from the image for classification, but also well recovers the image. In this class-specific method, the class that yields the sparsest residue recovers the occluded image. This method takes the recovery ability to assess how one class is similar to an occluded image. With no assumption on the shape of the occlusion, the proposed method can work well for kinds of occlusions. As detailed in 3.3, the proposed method is related to the popular methods, including SRC [6], CRC [7], LRC[5], and NS [8].

Section 2 presents the basic idea. Section 3 proposes our method. Section 4 performs experiments. Section 5 concludes this paper.

2. Basic idea

Figure 1 intra-class and inter-class distance

In figure 1, the intra-class distance between \( a_1 \) and \( a_2 \) is small and the inter-class distance between \( a_1 \) (or \( a_2 \)) and \( b \) is large. However, The intra-class distance between \( a_1 \) and \( a_3 \) is larger than the inter-class distance between \( a_1 \) and \( b \), due to the occlusion. Euclidean distance is no longer effective to assess the similarity between images.

Every pixel contributes to the similarity between two images. The occluded pixels in an image introduce distrustful similarity. We should make decision only based on the trustful similarity introduced by the non-occluded pixels.
One can identify the similarity between $a_1$ and $a_3$ (or $a_4$) in figure 1 and consider them as images of the same person, though their distance is very large. This is because 1) a large portion of these two images are very similar to each other; and 2) we are sure that this similarity is trustworthy. When we group $a_1$ and $a_3$ into the same class, we are making decision based on the trustworthy similarity. The appearance of distrustful dissimilarity deteriorates the Euclidean distance as a good measurement of similarity. Discarding the distrustful similarity is helpful for classification.

Figure 2: the relationship between occluded and non-occluded images

Equation (1) states the relationship between an occluded face image and a non-occluded face image (as shown in figure 2):

$$f_i = f + (r_i - r_j)$$  \hspace{1cm} (1)

where $f$ and $f_i$ are respectively non-occluded and occluded face images, $r_i$ and $r_j$ are the occlusion and the occluded part of $f$. Here, we assume the non-occluded pixels of $f$ and $f_i$ are the same. The appearance of the sunglasses significantly exaggerates the Euclidean distance between $f$ and $f_i$. We can reduce their distance by reconstruct the occluded part.

Compared with the whole face images ($f$ and $f_i$), the occlusion (i.e. $r_i$) consists of many fewer pixels. We consider the occlusion and the face image are of the same size, and most pixels of the occlusion are zeros. Then, the occlusion is a sparse image. Based on this observation, we proposed an occluded image reconstruction and recognition method in section 3.

3. Occluded face image reconstruction and recognition

3.1 Reconstruction

Assume $x'_i, x'_j, \ldots$ are $n$ non-occluded training images of the $i$th person, and $y$ is another non-occluded image of the same person. Denote an occluded version of $y$ as $y'$.

Equation (1) expresses an occluded image as an addition of a non-occluded image and an occlusion. For $y'$, the non-occluded image $y$ is not necessarily a training image. The methods [5, 9] states $y$ can be linearly represented by the training images as follows

$$y = \sum_{i=1}^{n} \alpha_i x'_i$$  \hspace{1cm} (2)

We can add a sparse residue $r$ to both sides of (2), and obtain the following equation

$$y + r = \sum_{i=1}^{n} \alpha_i x'_i + r$$  \hspace{1cm} (3)

The equation (3) shows the relationship between an occluded image and the training images. The key problem is how to determine an optimal solution $\alpha$ among the solution space.

The following model yields the sparsest residue

$$\min_{\alpha} \| y - \sum_{i=1}^{n} \alpha_i x'_i \|$$  \hspace{1cm} (4)

This model can be interpreted as follows: seek a linear combination of $x'_i, x'_j, \ldots$ who has most pixels same to $y$. The NP hard minimization problem (4) is conditional replaceable by a $l_1$ norm minimization problem [10], as follows

$$\min_{\alpha} \| y - \sum_{i=1}^{n} \alpha_i x'_i \|$$  \hspace{1cm} (5)

Study [11] shows that the proper coefficient is the unique solution of the above minimization problem (5), provided that the number of corrupt pixels is not too large. In this paper, we solve the minimization problem in (5) using 11-magic [12].

We can calculate the non-occluded image $y$ using $y = \sum_{i=1}^{n} \alpha_i x'_i$, and the occlusion using $r = y - \sum_{i=1}^{n} \alpha_i x'_i$.

3.2 recognition

Assume $x'_j (1 \leq i \leq c, 1 \leq j \leq n_i)$ is the $j$th image of the $i$th person, where $n_i$ is the number of the images belonging to the $i$th person and $c$ is the number of persons. The image $y$ is an occluded face image to be classified.

To classify an occluded image $y$, we can perform the following procedures:

**Step 1.** For each $1 \leq i \leq c$, calculate the coefficient $\alpha$ by solving a minimization problem (5), and calculate the reconstructed image and the corresponding residue;

**Step 2.** Compute the block distance between $y$ and $y_j (1 \leq i \leq c)$, i.e. the $l_1$ norm of $r_i$, $\| r_i \|_1 = \| y - y_j \|_1$;

**Step 3.** Classify $y$ into the $k$th class, if $\| r_i \|_1 = \min_{1 \leq i \leq c} \| r_i \|_1$.

Figure 3: Reconstruction results and residues
If the test image \( y \) is the image of the \( i \) th class, the reconstructed image form the \( i \) th class is the intrapersonal reconstruction and the rest are interpersonal reconstruction. Figure 3(a) is an occluded image. Figure 3(b) shows an intrapersonal reconstruction result and (b-g) show five typical interpersonal reconstruction results. The second line shows the reconstruction residues.

The intrapersonal reconstruction result is not only quite similar to the original image, but also the most clear. While most details in (b) are clear, those in the interpersonal reconstruction results (c-g) are blurring, particularly in the eye part and the mouth part.

The intrapersonal residue (b1) is sparser than the rest (c1-g1). In fig 3 b1, most of the nonzero values scatter in the area of occlusion. In the interpersonal residues, however, the nonzero values scatter not only in the occluded area but also in the rest areas. This can explain why the interpersonal reconstruction results are blurring.

### 3.3 Relation to the other methods

The same to the proposed method, LRC [5] is a class-specific model and constructs \( c \) linear expressions. Regarding to all the training images, each linear expression of the test sample is a sparse one (with the coefficients corresponding face images of the rest persons are zeros). Different from the SRC [6] that gains the unsupervised sparsity, the sparsity in the proposed method and LRC is a kind of supervised sparsity [13]. For occluded images, the proposed method yields a reconstructed image that has most pixels equal to the original one. The LRC cannot well reconstruct the image.

Though both the SRC and the proposed method can handle the occlusion problem, they have distinct difference 1) SRC expresses the occlusion using the face images that associate different persons from the test image; 2) the proposed method takes the occlusion as the residue of the linear expression.

To summary, all of LRC, SRC, and the proposed method seek a sparse representation for the test sample, and make decision based on the representations. Both SRC and CRC emphasize the globally representation by minimizing the \( l_2 \)-norm residue, i.e. the test sample is well represented around all pixels. Differently, the proposed method allows the representation has large local deviations.

The proposed method can also be viewed as a latent local method. Different from [3-4] that divide the face images into several portions explicitly, the proposed method implicitly divides the occluded image into two parts: trustful face image and distrustful face image. The occlusion is the residue of the intrapersonal expression. With no assumption on the shape of the occlusion, the proposed method can work well for kinds of occlusions.

The proposed method can be regarded as a kind of NS classifier. Compared with the traditional NS [8], the proposed method uses a new metric (calculated using (5)) to measure the distance between a test sample and a class.

### 4. Experiments

We conduct experiments on two popular databases: AR [14] and extended YaleB [15]. We crop and normalize all of the face images. The sizes of the AR and YaleB images are respectively \( 50 \times 40 \) and \( 32 \times 32 \).

We compare our method with five popular methods: NN, NS [8], LRC [5], SRC [6], CRC [7].

#### 4.1 Face recognition without occlusion

A subset of AR database [14] consists of 1, 399 images of 100 persons (about 14 images for each person). The seven images from the first session of each person are used for training, and the other seven from the second session are used for testing. We use the eigenface as the face feature. We set the parameter \( \lambda \) to be 1e-2 in CRC. The schemes of SRC and LRC are design respectively based on [6] and [5]. Table 1 lists the classification accuracy of different methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>NN</th>
<th>NS</th>
<th>LRC</th>
<th>SRC</th>
<th>CRC</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>63.4</td>
<td>68.1</td>
<td>86.6</td>
<td>89.7</td>
<td>80.4</td>
<td>90.1</td>
</tr>
</tbody>
</table>

Table 1 classification accuracy (%) of different methods on AR images without occlusion

#### 4.2 Face recognition with disguise

This section investigates the robustness of different methods against occlusion on a subset of the AR database. This subset consists of 2, 600 images of 100 persons.
persons from two sessions. In each session, there are 7 images without occlusion, 3 images occluded by sunglasses, 3 by scarf. In our experiments, we take the 14 non-occluded images of each person as the training samples, and the occluded images as the testing samples. In this experiment, we use the local binary pattern (LBP) feature [16] to represent the face image. We do not partition the images in this experiment. Table 3 lists the classification accuracies of different methods.

Table 3 classification accuracy (%) of different methods on AR images with occlusion

<table>
<thead>
<tr>
<th>Methods</th>
<th>SVM</th>
<th>LRC</th>
<th>SRC</th>
<th>CRC</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunglasses</td>
<td>63.5</td>
<td>93.6</td>
<td>91.0</td>
<td>68.5</td>
<td>95.8</td>
</tr>
<tr>
<td>Scarf</td>
<td>24.8</td>
<td>53.4</td>
<td>59.5</td>
<td>80.0</td>
<td>79.8</td>
</tr>
</tbody>
</table>

![Figure 4 samples of the synthetic occlusions](image)

We conduct another experiment to simulate the random noise. The training set also consists of the 14 non-occluded images. We first linear combine the training images, then pollute the training images with random noise and take them as the testing images. We choose 15 lines (40% of the 50 lines) and replace each pixels of them with random numbers in the interval (0,255). To imitate the scarf, we randomly pollute the lines 1-15. To imitate the sunglasses, we randomly pollute the lines 24-39. Figure 4 shows 12 samples. Table 4 lists the classification accuracy of different methods.

Table 4 classification accuracy (%) of different methods on AR images with synthetic occlusions

<table>
<thead>
<tr>
<th>Methods</th>
<th>SVM</th>
<th>LRC</th>
<th>SRC</th>
<th>CRC</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scarf</td>
<td>74.6</td>
<td>93.0</td>
<td>95.5</td>
<td>92.3</td>
<td>100</td>
</tr>
<tr>
<td>Sunglasses</td>
<td>68.5</td>
<td>89.1</td>
<td>93.4</td>
<td>90.6</td>
<td>100</td>
</tr>
</tbody>
</table>

As can be seen from table 4, we know the proposed method can achieve the highest classification accuracy. This is because, for each test sample, some certain linear combination of the training set is totally the same to it except in the occlusion part. In our experiments, the intrapersonal reconstruction is very close to the non-occluded version of the test sample.

5. Conclusion

This paper considers an occluded image as a summation of a non-occluded image and a sparse occlusion. The appearance of the occlusion degrades the performance of traditional subspace methods. We propose a method for occluded image reconstruction and classification. The proposed method linearly expresses the occluded image by the training images of the different persons, and classifies the image based on the residue. Before validating the proposed method in experiments, we analyze the relationship between the proposed method and the popular methods.

Reference