Optik 126 (2015) 3016–3019

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Double layers sparse representation for occluded face recognition

Shuhuan Zhao, Zheng-ping Hu*

School of Information Science and Engineering, Yanshan University, Hebei Street 438, Haigang District, Qinhuangdao 066004, China

ARTICLE INFO

Article history:
Received 5 June 2014
Accepted 9 July 2015

Keywords:
Occluded face recognition
Sparse representation
Occlusion detection

ABSTRACT

To implement occluded face recognition, we proposed a new method based on sparse representation. First, it partitioned images into four blocks. For each block, unit matrix was added to extend the dictionary and then sparse representation was performed. The coefficients corresponding to the unit matrix were used to estimate the occlusion pixels. Second, the un-occluded part of each block was used to reconstruct each block and the block with the biggest sparsity was used to reconstruct the image again. Occlusion pixels were modified by the residuals. Finally the identification was performed on the un-occluded part. Experiments verify the robustness and effectiveness of our method.

1. Introduction

Image recognition, especially face recognition, has attracted a lot of researchers due to its wide application. And many methods have been proposed to solve this problem, including PCA, LDA, SVM and other related methods. Recently, sparse representation coding (SRC) is attracting more and more attention [1–3] and has gotten many satisfying results. Although these methods perform well under some controlled conditions, they fail to perform well in the situation when test data is corrupted due to occlusion. To solve this problem, Meng Yang etc. proposed to solve the problem in paper [4] by extending training samples using the difference between samples. And Paper [5] uses the image Gabor-features for SRC, which can get more compact occlusion dictionary. As a result, the computation complexity and the number of atoms are reduced. In addition, paper [6] proposes a novel low-rank matrix approximation algorithm with structural incoherence for robust face recognition. In this paper the raw training data is decomposed into a low-rank matrix and a sparse error matrix. Besides, it introduces structural incoherence between low-rank matrices which promotes the discrimination between different classes, and thus this method exhibits excellent discriminating ability. However, all these methods have a common condition that they need occluded samples in the training set. If there are no occluded samples in the training set, how to perform recognition with satisfying result? Paper [7] models the sparse coding as a sparsity-constrained robust regression problem and proposes a new scheme, namely the robust sparse coding (RSC) that seeks for the MLE (maximum likelihood estimation) solution of the sparse coding problem, which makes it much more robust to outliers than SRC. And paper [8] proposes WGSR (Modular Weighted Global Sparse Representation) for robust face recognition, in this paper image is first divided into modules and each module is processed separately by SRC, then the modular reliability is determined by the modular sparsity and residual jointly. Finally a reconstructed image from the modules weighted by their reliability is used for the robust recognition. Paper [9] proposes a new approach. It detected the presence of sunglasses/scarves first and then processed the non-occluded facial regions only. The occlusion detection problem is approached by PCA and improved support vector machines (SVM), while the non-occluded facial part is identified by block-based weighted local binary patterns (LBP). Paper [10–12] also divide image into modules but changed the strategy of division into randomized modules, multi-level modules and horizontal modules.

In this paper, we proposed a new method based on sparse representation: DLSR (Double Layer Sparse Representation). First, it partitioned images into four blocks. For each sub-dictionary, unit matrix was added to extend the dictionary and then sparse representation was performed. The coefficients corresponding to the unit matrix were used to estimate the occlusion pixels. Second, the un-occluded part of each block was used to reconstruct each block by SRC and the block with the biggest sparsity was used to reconstruct the image again. Occlusion pixels were modified by the residuals. Finally the identification was performed on the un-occluded part by SRC.

2. Sparse representation

Since our classification rule is based on SRC, for the sake of clarity we now briefly review this algorithm. SRC seeks the sparsest

* Corresponding author. Tel.:+86 13833560097.
E-mail address: hzp_ysu@163.com (Z.-p. Hu).

http://dx.doi.org/10.1016/j.ijleo.2015.07.060
0030-4026/

representation of test sample in dictionary. Suppose that there exists K subjects and each training sample can be represented as vector $v_{i,n}$, so all the samples from the ith class construct a matrix $D_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in R^{m \times n_i}$, where $m = w \times h$ means the dimension of training sample and n_i means the number of training samples from the ith class. It supposes that samples from the same class construct a linear subspace, so any test sample can be represented as linear combination of samples from the same class. For example, test sample $y \in R^m$ from class i can be represented as:

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \cdots + \alpha_{i,n} v_{i,n} \text{ where } \alpha_{i,j} \in R, j = 1, 2, \ldots, n_i \tag{1}$$

Since we don't know which class the test sample belongs to, it identifies a dictionary as $D = [D_1, D_2, \ldots, D_k] = \left[v_{1,1}, v_{1,2}, \ldots, v_{k,n}\right] \in R^{m \times N}$, where m denotes the dimension of training sample, k denotes the number of objects, and $N = \sum_{i=1}^{k} n_i$ refers to the total number of training samples. The test sample y can be represented as the linear combination of all training samples:

$$y = Dx_0 \in R^m \tag{2}$$

where $x_0 = \left[0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n}, 0, \ldots, 0\right]^T \in R^N$ denotes the vector of coefficients. The non-zero atoms in x_0 correspond to the training sample from the same class as the test sample. In image recognition, Eq. (2) is usually underdetermined, that is $m < N$, so there will be many solutions. Since we know that x is sparse, we can restrain the equation by min-l_0 norm:

$$(l_0): \hat{x}_0 = arg \min \|x\|_0 \; s.t. \; Dx = y \tag{3}$$

Because Eq. (3) is a NP-hard problem, according to sparse representation and compressive sensing, l_0-norm can be replaced by l_1 norm only if x_0 is sparse enough, so there is

$$(l_1): \hat{x}_1 = arg \min \|x\|_1 \; s.t. \; Dx = y \tag{4}$$

Since there is noise in image, the linear combination of training samples cannot reconstruct test sample accurately, so it permits the existence of error and defines the limit of error as ε. Eq. (4) can be converted into the following form:

$$(l_1): \hat{x}_1 = arg \min \|x\|_1 \; s.t. \; \left\|Dx - y\right\|_2 \le \varepsilon \tag{5}$$

Finally, the residuals of each class are computed and the test sample is classified into the class with lowest residual:

$$\text{identify}(y) = \text{argmin}_i r_i(y) \tag{6}$$

where $r_i(y) = \left\|y - D\delta_i(x)\right\|_2$, $i = 1, 2, \ldots, k$, $\delta_i(x) = \left[\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n}\right]^T$ denotes the vector of coefficient of ith class.

3. Double layer sparse representation for occluded image recognition

SRC is widely used for its robustness and effectiveness. But when the image is occluded, the performance of global SRC will decline. To solve this problem, a block-based method DLSR (Double Layer Sparse Representation) was proposed. For the training and test images as Fig. 1(a) shows, we partitioned them into 4 blocks as Fig. 1(b) shows. So the dictionary D consisted by the training set can be converted into 4 sub-dictionaries as $D = [D_1, D_2, D_3, D_4]$. So did the test image y and it can be written as $y = [y_1, y_2, y_3, y_4]$. To detect the occlusion pixels, unit matrix was added to each sub-dictionary. Then the extended dictionary can be rewritten as $DI = [DI_1, DI_2, DI_3, DI_4]$. Then sparse representation was performed on each block

$$[xD_i, xI_i] = \arg\min \left\|x_i\right\|_2, s.t. \left\|DI_i x_i - y_i\right\|_2 \le \varepsilon \text{ 其中 } 1 \le i \le 4 \tag{7}$$

in which $x_i = [xD_i, xI_i]$, xD_i and xI_i mean the representation coefficients of test image block y_i on sub-dictionary D_i and unit dictionary I, respectively. According to the theory of representation xI_i denoted the coefficient of each pixel on unit matrix. Because the occlusion pixels cannot be reconstructed well by training samples, so the corresponding coefficient in xI_i should be large. Then the occlusion pixels can be estimated. In this paper, coefficients corresponding to the unit matrix were connected as $xI = [xI_1, xI_2, xI_3, xI_4]$. Portion of the pixels corresponding the maximum values were selected as occlusion pixels. The roughly estimated occlusion was showed in Fig. 1(c). But this estimation was not detailed enough. On this base, this paper reconstructed each block by un-occluded part. Fig. 1(d) shows the reconstruction image. From the image, we can see that blocks with occlusion cannot be reconstructed well. So we selected the block with the highest sparsity to reconstruct the global image as Fig. 1(e) shows. The sparsity can be calculated by Eq. (8). The image was much clearer than Fig. 1(d) and the residual image was showed in Fig. 1(f). Occlusion part can be modified by the residual and the final occlusion pixels were as Fig. 1(g) shows. Obviously, this occlusion estimation was more accurate. The final identification was performed on the un-occluded part of the image by SRC.

$$S_k = \frac{\sqrt{C}}{\sqrt{C} - 1} \left(\frac{\left\|x_k\right\|_2}{\left\|x_k\right\|_1} - \frac{1}{\sqrt{C}} \right) \tag{8}$$

4. Experiments

To verify the effectiveness of the proposed, we do experiments on AR and Yale B database. SRC and SPP are selected as benchmark. The parameter λ in each method is set as 0.1.

4.1. AR database

The AR database consists of over 4000 frontal-face images from 126 subjects, among which 26 pictures were taken in two separate sessions for each subject. In this experiment, 119 subjects are selected. And for each subject we choose 7 images without occlusion for training and 3 images with scarf and 3 images with sunglasses for testing. All the chosen images have a little difference in expression, hair type, and occlusion etc. Fig. 2 shows some images of AR database, in which images on the first line are training samples and images on the second line are test images. Before the experiment, all the images are down-sampled to reduce dimensions.

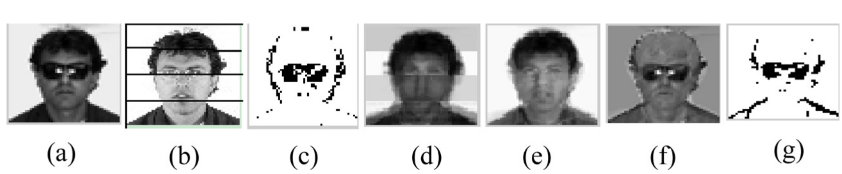

Fig. 1. Intermediate result images of our method.