Graph Modeling based Local Descriptor Selection via A Hierarchical Structure for Biometric Recognition

Xiaobo Zhang, Zhenan Sun, and Tieniu Tan
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
95, Zhongguancun East Road, Haidian District P.O.Box 2728, 100190 Beijing, P.R. China
{xbzhang, znsun, tnt}@nlpr.ia.ac.cn

Abstract

Local descriptor based image representation is widely used in biometrics and has achieved promising results. We usually extract the most distinctive local descriptors for image sparse representation due to the large feature space and the redundancy among local descriptors. In this paper, we describe the local descriptor based image representation via a graph model, in which each node is a local descriptor (we call it “atom”) and the edges denote the relationship between atoms. Based on this model, a hierarchical structure is constructed to select the most distinctive local descriptors. Two-layer structure is adopted in our work, including local selection and global selection. In the first layer, $L_1/L_q$ regularized least square regression is adopted to reduce the redundancy of local descriptors in local regions. In the second layer, AdaBoost learning is performed for local descriptor selection based on the results of the first layer. We apply this method to long-range personal identification by using binocular regions. Our method can select the distinctive local descriptors and reduce the redundancy among them, and achieve encouraging results on the collected binocular database and CASIA-Iris-Distance. Particularly, our method is about 50 times faster than the traditional AdaBoost learning based method in the experiments.

1. Introduction

Local features are powerful tools to characterize the patterns in local regions, and they have shown promising results in biometrics, e.g. Gabor representation based iris recognition [4], LBP based face recognition [2], SIFT based periocular biometrics [14]. Figure 1 shows an example of face representation based on local descriptors. However, this representation is usually high-dimension due to the large parameter space of local feature (e.g. multiple channel Gabor filters, multiple forms of LBP). Even feature representation can not tackle this problem directly due to the high computational complexity and large storage cost. Therefore, some feature selection methods are proposed to extract the most representative local descriptors to describe the image, which makes the sparse representation (See Fig. 1(d)) promising and attractive.

![Figure 1. Local descriptor based face representation.](a) A face image is from CASIA-Iris-Distance [1] and is normalized; (b) A set of local descriptors. Each local descriptors are denoted by blue dots; (c) Dense representation with a set of descriptors, denoted by green dots; (d) Sparse representation with selected local descriptors.

AdaBoost learning is one of the most widely used tools for local descriptor selection [8, 2, 18], and achieves the state-of-the-art performance, but it requires large feature pools and training samples, which makes the selection procedure time-consuming (We will detail this in Sec. 6). In this paper, we propose a graph model to describe the local descriptor representation and selection, and based on this model, a hierarchical structure is proposed to select the most distinctive local descriptors for image representation. At first, we extract the local descriptors for each image. Then, the least square regression with $L_1/L_q$ regularization is performed to remove the redundancy among local descriptors in local regions. Finally, AdaBoost learning is adopted to select the most distinctive local descriptors based on the selected ones in the first layer. We apply the proposed method to personal identification by using binocular regions at a distance.

The main contributions of this paper are three-fold:

- We model the local descriptor based image representation as a graph, and propose an efficient and effective
hierarchical structure to select the most stable and distinctive local descriptors.

- Two layers are adopted in our hierarchical structure. The first layer is to reduce the redundancy of local descriptors with local constraint, and the second layer is to select the most distinctive local descriptors globally.

- We apply the proposed method to binocular biometrics at a distance, which uses the binocular regions for long-range personal identification, and our method achieves comparable results to the state-of-the-art with much lower computational complexity.

The remainder of this paper is organized as follows. Section 2 briefly introduces some related work. Section 3 details the proposed graph modeling based image representation and local descriptor selection. Section 4 describes the proposed method. Section 5 presents the application of the proposed method on binocular biometrics. Section 6 presents experimental results and discussions. Finally, Section 7 summaries this paper.

2. Related work

AdaBoost learning is one of the most famous methods adopted for local descriptor selection. It can not only select the most distinctive feature, but also construct a cascade classifier, which is effective and efficient in real applications. Since AdaBoost [17] showed powerful ability in face detection, many variants of AdaBoost learning (Real AdaBoost, Logit AdaBoost, Gentle AdaBoost, etc.) [6] have been proposed to tackle some related topics, e.g. object detection, face recognition, iris recognition.

Feature selection based on sparse model is another hot topic. It can be seemed as model fitting problem with a sparse constraint on the model parameter. The sparsity of the model parameter indicates the importance of the feature. Therefore, we can select the features by considering the sparsity of model parameter. Methods like sparse multiple kernel learning (SMKL) [15], sparse support vector machine (SSVM) [3] belong to this model. They have shown promising results in recent years.

However, one drawback of the above methods is that they only consider the discrimination of the local descriptors, whereas ignore the feature structure during selection. Recently, some researchers consider the structure information of feature, and pay more attention on structured sparsity learning based feature selection [9]. Based on different prior information, methods like group lasso [19], fused lasso [16], tree structured group lasso [10] are proposed for different kinds of applications in computer vision and bioinformatics.

3. Graph modeling of local descriptor selection

For local descriptor, there are two types of parameters: spatial parameter $s$ and descriptor parameter $f$. Spatial parameter indicates the spatial position $(x, y)$ in the image spatial space. Feature parameter indicates the position in the descriptor space, e.g. the parameters of Gabor filter, such as orientation and scale. Therefore, a local descriptor can be described as a tuple $a(s, f)$, we call this tuple "atom". Usually, the atoms are not independent to each other, e.g. the atoms close to each other in feature space is highly correlated. Therefore, an image can be represented by a tuple $T(A, R)$, in which $A$ is the set of atoms, and $R$ denotes the relationship among atoms. We describe the image with a graph $G(N, V)$ (See Fig. 2), where $N$ is the set of nodes, and $V$ is the set of edges. The nodes in this graph are atoms, and the edges indicate the relationship among the atoms.

The feature space of local descriptor is composed of the image space and the parameter space of the descriptor, therefore it is usually very large, which makes local descriptor selection high complex. In addition, the local descriptors which are close to each other in feature space may have redundancy. Therefore, based on our graph model, the goal of local descriptor selection is to reduce the redundancy of the atoms and select the most discriminative ones. Figure 2 illustrates the procedure of graph modeling based local descriptor selection. Given a graph $G$, we should remove the nodes which are not representative and the edges which are redundancy to obtain a sparse graph $G'$. A simple method is to select fixed atoms in feature space by considering some prior knowledge, for example, for face recognition, the eye region is more distinctive, so we only select the atoms in this region. A more effective way is to use machine learning methods to exploit the discriminative atoms. We focus on the second way in this paper.

The learning procedure is usually time-consuming due to the large feature space, therefore, in the following section, we will introduce a two-layer hierarchical structure for atom selection, which is effective and efficient.

![Figure 2. Procedure of graph modeling based local descriptor selection.](image)

XY denotes the image spatial space, and $F$ denotes the parameter space of local feature. $G$ is the original graph, and $G'$ is the sparse one after selection. $T$ is the original image representation based on local descriptor, and $T'$ is the one after local descriptor selection.
4. Atom selection via a hierarchical structure

4.1. Framework

Our framework of local descriptor selection is based on the above graph model. We want to select the most discriminative atoms, meanwhile we add the local constraint in feature space to reduce the redundancy of atoms. The framework consists of two layers: local selection and global selection. Figure 3 shows the proposed implementation of this framework. Given a constructed graph $G$ based on local descriptors, we first group the atoms according to their spatial relationship, and then we reduce the redundance of the atoms within each subgroup via structure sparsity learning to obtain $G'$. After that, we adopt AdaBoost learning to select the most discriminative atoms and obtain the final sparse graph $G''$.

![Figure 3. The hierarchical structure for atom selection. The red lines divide the graph into several groups according to the spatial distribution. (a) is local selection via structured sparsity learning and (b) is global selection via AdaBoost learning. Both of these two do not consider the relationship among atoms.](image)

Here, we emphasize that, because both structured sparsity learning and AdaBoost learning do not consider the relationship among atoms, therefore, there are no edges in $G'$ and $G''$ in Fig. 3. By the way, before local atom selection, we should divide the original graph into several subgraphs. For simple computation, here we use the spatial constraint to divide the whole feature space into regular subgroups.

4.2. Atom selection in local regions

We want to reduce the redundancy of atoms and obtain a sparse structure in each subgraph. Down-sampling is the common method in implementation, however, it is a hard selection strategy with strong prior knowledge. Therefore, it can not eliminate the redundancy of atoms as well as keep the most representative ones. Here, we adopt $L_1/L_2$ regularization based least square regression for local selection. The atoms within each subgroup via structure sparsity learning to obtain $G'$. After that, we adopt AdaBoost learning to select the most discriminative atoms and obtain the final sparse graph $G''$.

Feature selection is one of the most effective and efficient methods to feature selection. The general idea of AdaBoost is described as follows [8]:

Given $n$ samples $\{x_i, y_i\}_{i=1}^n (x \in \mathbb{R}^d), y \in \{+1, -1\}$ with associated weight $\{w(x_i)\}_{i=1}^n$, and $\Phi = \{\phi_m(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}\}_{m=1}^M$ is the feature pool, $P_+^w(\phi_m(x)), P_-^w(\phi_m(x))$ are the positive and negative distribution of $\phi_m(x)$, the goal is to automatically learn a small set of the most discriminative features $\{\phi(t)\}_{i=1}^T$ from the feature pool, and construct an ensemble classifier:

$$H(x) = \text{sign} \left( \sum_{t=1}^T h_t(\phi_t(x)) \right)$$

Here, we take the intra-class matching scores as positive samples and the inter-class matching scores as negative samples. And all the atoms in the feature space construct the feature pool. We adopt Real AdaBoost for learning. In Real AdaBoost, the weak learner $\phi(t)$ is defined as:

$$\phi(t) = \arg \min_{\phi \in \Phi} \sum_{j=1}^n \sqrt{P_+^w(j)P_-^w(j)}$$

and the component classifier $h_t(\phi_t)$ is defined as:

$$h_t(\phi_t) = \frac{1}{2} \ln \left( \frac{P_+^w(j)}{P_-^w(j)} \right)$$

5. Long-range personal identification by using binocular regions

5.1. Binocular biometrics

Human ocular region contains rich and discriminative information which has been used for personal identification in the past few years. Iris recognition is one promising kinds of ocular biometrics. Traditional iris imaging systems require users’ high cooperation, which limits the application extension of iris recognition [4]. Although some non-cooperative systems [13, 5] have emerged in last few years,
the performance may degrade due to the low quality iris images captured under uncontrolled environments. Recently, Park et al. [14] reported that the periocular regions can be used for personal identification. In addition, in face recognition, it has been reported that the eye regions are the most distinctive and stable regions in face [7]. What’s more, in some specific applications, only partial face regions, particularly eye regions can be captured (e.g., Arab women have to wear mask on their faces, and all regions are occluded except for eye regions.). In our work, we use the binocular regions for long-range personal identification.

5.2. Local descriptor extraction and selection

Here, we adopt the multiple channel Gabor filters to characterize the binocular regions due to its psychophysical relevance. Gabor filters can effectively capture the local structure in terms of spatial frequency (scale), spatial localization, and orientation. The kernel of 2-D Gabor filter can be formulated as follows:

$$
\psi_{\mu,\nu}(z) = \frac{1}{\sigma^2} e^{-\frac{1}{2} \| \mathbf{k}_{\mu,\nu} \|^2} e^{-\frac{1}{2} \| \mathbf{k}_{\mu,\nu} \|^2} e^{i \phi_{\mu}} e^{i \nu z}$$

where \( \mu \) and \( \nu \) define the orientation and scale of the Gabor kernels, \( z = (x, y) \), \( \| \cdot \| \) denotes the norm operator, and the wave vector \( \mathbf{k}_{\mu,\nu} \) is defined as follows:

$$
\mathbf{k}_{\mu,\nu} = k_{\nu} e^{i \phi_{\mu}}
$$

where \( k_{\nu} = k_{\text{max}} / f \) and \( \phi_{\mu} = \mu / \pi, k_{\text{max}} \) is the maximum frequency, and \( f \) is the spacing factor between kernels in the frequency domain.

After Gabor filtering, the response of each Gabor filter on each pixel is taken as a local descriptor, and the feature vector of each image is composed of all the local descriptors. We perform the proposed method to select the most discriminative local descriptors for image representation as described in Sec. 4.

6. Experiments

The goal of the proposed hierarchical structure based method (HS) is to select distinctive local descriptors with low computational complexity, meanwhile achieve high recognition accuracy. We compare it to the traditional AdaBoost learning based method (Ada), which is the widely used and the state-of-the-art local descriptor selection method. In this section, we analyze the computational complexity and recognition accuracy, as well as the spatial distribution of selected local descriptors in two layers in our method.

6.1. Experimental data

6.1.1 Binocular database

We use our long-range multi-modal biometric recognition system[5] to obtain high resolution binocular images at a distance under near infrared (NIR) illumination. This imaging system mainly consists of two wide-range web cameras, a narrow-range high resolution NIR camera, a pan-tilt-zoom (PTZ) unit and a NIR light source. Based on this system, we collect a dataset including 13480 binocular images from 674 subjects. Each image contains full binocular region with resolution 2352 × 1728. These images are captured at 3 meters away, and each subject has 20 images.

Figure 4(b) shows one example in this database.

6.1.2 CASIA-Iris-Distance

Another database we adopted in our experiments is the CASIA-Iris-Distance database [1], which has been released by Institute of Automation, Chinese Academy of Sciences. Images from CASIA-Iris-Distance database were captured by the system [5] designed by CASIA. This database consists of 2,567 NIR high resolution (2352×1728) images from 142 subjects, and each subject has more than 10 images. Each image contains two eyes and a partial frontal face. Figure 4(b) shows one example in this database.

Figure 4. Images used in our experiments. (a) One image in our binocular database; (b) One image in CASIA-Iris-Distance; (c) the normalized binocular image of (b).

6.2. Experiment settings

Before feature extraction, we should preprocess the original binocular images. Firstly, We adopt Haar feature and AdaBoost learning for eye detection. Secondly, we simply adopt the well-known integral-differential operator proposed by Daugman [4] to obtain the center of iris. Then, we estimate the rotation angle via the two iris centers. Finally, we obtain the normalized binocular image based on the scale factor and rotation angle similar to face normalization. The center of two iris is the center of the normalized image, and the connection line of the two irises is in horizontal direction.

Considering feature extraction, we use 5 scales and 8 orientations multiple channel Gabor filters to construct the local descriptors, namely \( \nu \in \{0, \ldots, 4\} \) and \( \mu \in \{0, \ldots, 7\} \). The other parameters of Gabor filters are set as follows: \( \sigma = 2\pi \), \( k_{\text{max}} = \pi / 2 \) and \( f = \sqrt{2} \). The size of normalized binocular image is 48 × 128 pixels, therefore the total number of local descriptors extracted for one image is 245,760. We divide the normalized binocular image into 8 × 8 non-overlapped windows for structured sparsity learning. For the structure sparsity learning, we set the regularization factor \( \lambda = 0.5 \),
and the norm factor \( q = 2 \). We divide the collected binocular database into 2 non-overlapped groups. The first group is training set, which consists of 6000 images from 300 subjects in binocular database. And the others images are used for testing. The CASIA-Iris-Distance database is also used for testing in our experiments.

6.3. Experiment results

6.3.1 Recognition accuracy

To show the effectiveness of the proposed method, we evaluate the performance of the two methods (Ada, HS) in terms of discriminative index \( (d') \), operator receiver characteristic (ROC) curve, and equal error rate (EER). The discriminative index \( d' \) is calculated as follows:

\[
d' = \frac{|m_I - m_E|}{\sqrt{(\sigma_I^2 + \sigma_E^2)/2}}
\]

where \( m_I \) and \( m_E \) denote the average values of intra-class matching scores and inter-class matching scores, and \( \sigma_I \) and \( \sigma_E \) are the corresponding standard deviations.

<table>
<thead>
<tr>
<th>Method</th>
<th>Binocular</th>
<th>CASIA-Iris-Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ada</td>
<td>4.0887</td>
<td>0.0350</td>
</tr>
<tr>
<td>HS</td>
<td>3.9489</td>
<td>0.0326</td>
</tr>
</tbody>
</table>

EER, EER

Table 1. EERs and \( d' \)s on the binocular database and the CASIA-

Figure 5. ROC curves on (a) the binocular database and (b) the CASIA-Iris-Distance database.

From Tab. 6.3.1 and Fig. 5 we can see, the proposed method (HS) achieves comparable results with the traditional AdaBoost learning (Ada) on CASIA-Iris-Distance and the binocular database, which shows the effectiveness of our method. However, the performances of both algorithms on CASIA-Iris-Distance are worse than those on the binocular database. Maybe it is because that all the training images are from the binocular database, the selected local descriptors favor the binocular database much more. Another reason may be that, the CASIA-Iris-Distance database is used for long-range iris recognition, and to capture clear iris texture, the aperture of the camera is set a little large, which results in that part of the binocular region is overexposed, therefore the discrimination of skin patterns of binocular regions are weakened (See Fig. 4(b)).

6.3.2 Computational complexity

One notable advantage of the proposed method is its low computational complexity. As we all know that, in binary-class AdaBoost learning, we need large number of negative samples and positive samples. In local descriptor selection, the negative samples are the matching scores of the inter-class pairs, and the positive samples are the matching scores of the intra-class pairs. Therefore, the computational cost of AdaBoost learning is decided by the size of feature pool and the matching algorithm. The time cost of this procedure can be calculated as \( F \times C \times M_1 + L_t \), where \( C \) is the number of comparisons, \( F \) is the size of feature pool, \( M_1 \) is the time cost of single match procedure, and \( L_t \) is the time cost of AdaBoost Learning.

In hierarchical structure, the time cost is from two layers: structured sparsity learning and AdaBoost learning. The time cost of first layer depends on the number \( G \) of groups and the time cost \( S_t \) of \( L_1/L_0 \) optimization\(^1\), and the computation of time cost of the second layer is similar to the Ada.

From Tab. 6.3.3, we can see that the size of feature pool affects the time cost significantly. Both the total time of matching and the time cost of AdaBoost learning of the proposed method decrease significantly. Particularly, our method is about 50 times faster than the traditional AdaBoost based method.

6.3.3 Distribution of selected local descriptors

![Figure 6. An example of selected atoms in a local region after structured sparsity learning. (a) the normalized binocular image. The red box indicates one local region (8 × 8) in the normalized binocular image; (b) the distribution of selected atoms (denoted by red dots) in feature space. XY denotes the image spatial space, and Z axis denotes the index of Gabor filter.](image)

The distribution of selected atoms indicates the most discriminative local descriptors in spatial space. We analyze the distribution in two stages in the hierarchical structure: atom selection in local regions and atom selection in global space.

The distribution of selected atoms in the first stage reflects the importance of these atoms in local regions. Fig-

\(^1\)Liu and Ye [12] proposed an efficient method for solving \( L_1/L_0 \) norm regularization. In our experiments, we adopt their algorithms and tool box [11]
ure 6 shows one example of selected atoms in a 8 × 8 window in the normalized binocular image. We can see that, the selected atoms distribute sparsely in feature space, which confirms that the structured sparsity learning can reduce the correlation among atoms closed to each other in feature space. Figure 7 shows the selection result of AdaBoost learning in the second layer. From Fig. 7(a), we can see that the selected atoms distribute sparsely in feature space. From Fig. 7(b), we can see that there are few selected atoms in iris, pupil and sclera, which suggests that the patterns of these regions are not stable and distinctive in the low resolution normalized binocular images (48 × 128).

Figure 7. Illustration of selected atoms after AdaBoost learning. (a) the distribution of selected atoms in feature space. XY denotes the image spatial space and Z axis denotes the index of Gabor filter; (b) the distribution of selected atoms in image spatial space.

7. Conclusions

In this paper, we have described a graph model to represent an image with local descriptors, in which each node is a local descriptor and the edges represent the relationship among these local descriptors. Based on this model, we adopt a two-layer hierarchical structure for local descriptor selection. In the first layer, structured sparsity learning is adopted to select the key local descriptors to reduce the redundancy in local regions. In the second layer, AdaBoost learning is performed to select the most discriminative local descriptors for high recognition performance. Extensive experiments show the effectiveness and efficiency of the proposed method for long-range personal identification by using binocular regions. Our method can be also used for many other related applications based on local descriptor representation.

Acknowledgements

This work is funded by National Natural Science Foundation of China (Grant No. 60736018, 61075024) and International S&T Cooperation Program of China (Grant NO. 2010DFB14110).

References