Support Vector Machine with Local Summation Kernel for Robust Face Recognition

Kazuhiro Hotta
The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan
Email: hotta@ice.uec.ac.jp

Abstract

This paper presents Support Vector Machine (SVM) with local summation kernel for robust face recognition. In recent years, the effectiveness of SVM and local features is reported. However, conventional methods apply one kernel to global features. The effectiveness of local features is not used in those methods. In order to use the effectiveness of local features in SVM, one kernel is applied to local features. It is necessary to compute one kernel value from local kernels in order to use the local kernels in SVM. In this paper, the summation of local kernels is used because it is robust to occlusion. The robustness of the proposed method under partial occlusion is shown by the experiments using the occluded face images. In addition, the proposed method is compared with the global kernel based SVM. The recognition rate of the proposed method is over 80% under large occlusion, while the recognition rate of the SVM with global Gaussian kernel decreases dramatically.

1 Introduction

Face recognition has many potential applications such as security system, man-machine interface, and the search from video databases or WWW. Therefore, many researchers work actively and many face recognition methods have been proposed [1]. In these days, face recognition can be performed with high accuracy under the restricted environment [1, 2, 3]. However, the face recognition under practical environment is still difficult, because there are some obstacles such as occlusion, illumination changes, and view changes. A robust recognition method under practical environment is desired.

In recent years, the effectiveness of Support Vector Machine (SVM) [4, 5] for object recognition is reported [2, 6]. On the other hand, the effectiveness of local features is also reported [6, 7, 8]. Heisele et al. [6] compared the local component-based method with global appearance-based method and showed the effectiveness of local component. From the recent two results, it is natural to consider that the recognition method based on local features is effective. However, in the conventional recognition methods, one kernel is applied to all features extracted from one image [2, 3, 9]. Namely, the effectiveness of local features is not used in those methods. In this paper, in order to use the effectiveness of local features in SVM, one kernel is applied to local features. We propose the SVM with local kernels and show the effectiveness of the proposed method.

In order to use local kernels in SVM, it is necessary that one kernel value is computed from local kernels. The product and summation of local kernels are considered as the integration methods of local kernels which satisfy Mercer’s theorem [10, 5]. We call these as the local product kernel and local summation kernel respectively. It is considered that the local summation kernel is better than local product kernel. The reason is as follows. In local product kernel, if only one local kernel gives the low value, then the product kernel value becomes low. This represents that the local product kernel is influenced easily by noise or occlusion. On the other hand, the local summation kernel is not influenced when some local kernels give low value. This represents that local summation kernel is robust to occlusion. In practical environment, human faces are sometimes occluded by sunglasses, mask, and other objects. Therefore, the robustness to occlusion is necessary under practical environment. In this paper, local summation kernel is used to realize the robust face recognition.

To investigate the robustness to noise or occlusion, the white and black squares are added to face images randomly. The effectiveness and robustness of the proposed method is shown. In addition, the proposed method is compared with the global kernel based SVM. The recognition rate of the proposed method is over 80% under large occlusion, while the recognition rate of the SVM with global Gaussian kernel decreases dramatically.

In section 2, face recognition method based on SVM with local summation kernel is explained.
Section 3 shows the effectiveness of the proposed method. Conclusion and future works are described in section 4.

2 Face recognition method

This section explains the face recognition method based on SVM with local summation kernel. The proposed method is based on local kernels. Therefore, we use Gabor filters which can extract local appearance. The properties of Gabor filter are described in section 2.1. In section 2.2, SVM with local summation kernel is explained.

2.1 Gabor filter

In mammalian visual cortex, there are many neurons which are characterized as localized and orientation selective. It is known that Gabor filters are well fitted to the receptive field profiles of the simple cells of cat’s visual cortex.

Gabor filters are defined by

\[
\psi_k(x) = \frac{k^2}{\sigma^2} \exp \left( -\frac{k^2 x^2}{2\sigma^2} \right) \left[ \exp(i k x) - \exp(-\sigma^2/2) \right],
\]

where \( x = (x, y)^T \), \( k = k_0 \exp(i \phi) \), \( k_0 = k_{max}/f^\nu \), \( \phi = \mu \cdot \pi/4 \), and \( f = \sqrt{2} \), respectively. In the following experiments, Gabor filters of 4 different orientations are used. The size of Gabor filters is set to 9 x 9 pixels.

Figure 1 shows the Gabor filters of 4 different orientations and the Gabor features of a face image. The positions which give large Gabor outputs are different depending on the orientation parameter of Gabor filter. This represents the independency of Gabor features [11]. Moreover, Gabor outputs at many positions of the face image are small and only specific positions give large values. This represents the sparseness of Gabor features [12]. Independency and sparseness are the main reasons of the effectiveness for object recognition of Gabor features.

2.2 SVM with local summation kernel

First, we explain the SVM [5] briefly. SVM determines the optimal hyperplane which maximizes the margin. The margin is the distance between the hyperplane and nearest sample from hyperplane. When the training set (sample and its label) is denoted as \( S = ((x_i, y_i), \ldots, (x_L, y_L)) \), the decision function is defined by

\[
f(x) = \sum_i \alpha_i y_i x_i \cdot x, \quad \text{where } \alpha \text{ is the solutions of quadratic programming problem.}
\]

The training sample with non-zero \( \alpha \) is called support vector. This decision function assumes the linearly separable case. In the linearly non-separable case, the non-linear transform \( \Phi(x) \) is used. The training samples are mapped into high dimensional space by \( \Phi(x) \). By maximizing the margin in high dimensional space, non-linear classification can be done. If inner product \( \Phi(x) \cdot \Phi(y) \) in high dimensional space is computed by kernel \( K(x, y) \), then training and classification can be done without mapping into high dimensional space. The decision function using kernel is defined by

\[
f(x) = \sum_i \alpha_i y_i \Phi(x_i) \cdot \Phi(x) = \sum_i \alpha_i y_i K(x_i, x). \quad (2)
\]

Mercer’s theorem gives whether \( K(x, y) \) is the inner product in high dimensional space. The necessary and sufficient conditions are symmetric \( K(x, y) = K(y, x) \) and positive semi-definite of kernel matrix \( K = (K(x_i, x_j))^L_{i,j=1} \). Examples of kernel which satisfies Mercer’s theorem are Gaussian and polynomial kernel.

Next, SVM with local kernels is explained. First, we consider the type of local kernel. In the proposed method, local kernels are arranged at all positions of faces. Each local kernel plays the role of cells specialized to local features of each person’s face. In order to make the cells specialized to local features, the stimulus selectivity of Gaussian is suited. Therefore, Gaussian kernel is used as the local kernel. Local Gaussian kernel is defined by

\[
K_p(x(p), y(p)) = \exp \left( -\frac{||x(p) - y(p)||^2}{\sigma^2_p} \right), \quad (3)
\]

where \( p \) is the label of position. \( x(p) \) and \( y(p) \) represent the local features centered at position \( p \). At the simplest case, \( x(p) \) is the scalar feature of position \( p \).

In order to use local kernels in SVM, it is necessary that kernel value \( K(x, y) \) is computed from local kernels \( K_p(x(p), y(p)) \) arranged at all positions of recognition target. The product and summation of local Gaussian kernels are considered as the integration methods of local kernels which satisfy Mercer’s theorem [10, 5]. We call these two kernels as the local product kernel and local summation kernel respectively. It is considered that the local summation kernel is better than local product kernel. The reason is as follows. In local product kernel, if only one local kernel gives the low value, then the product kernel value becomes low. This represents that
the product kernel is influenced easily by noise or occlusion. Note that local product kernel corresponds to global Gaussian kernel when the variances of all local kernels are same and \( x(p) \) is the scalar feature of position \( p \). Namely, global Gaussian kernel is also influenced easily by noise or occlusion. On the other hand, local summation kernel is not influenced when some local kernels give low value. This represents that local summation kernel is robust to occlusion. Therefore, local summation kernel is used in this paper. The local summation kernel and global Gaussian kernel (local product kernel) are compared in section 3. The decision function of SVM with local summation kernel is defined by

\[
    f(x) = \sum_{i}^{L} \alpha_{i}y_{i} \frac{1}{N} \sum_{p}^{N} K_{p}(x_{i}(p), x(p)) \tag{4}
\]

where \( N \) is the number of local kernels. From equation (4), we understand that the mean of local kernels is used as the kernel value. The kernel value is normalized from 0 to 1 by dividing by the number of local kernels.

In this paper, the proposed method is applied to face recognition which is the multi-class classification problem. However, SVM is the binary classifier. In order to use SVM for multi-class problem, the pairwise (one-against-one) approach is used [9, 2]. The pairwise approach needs to train SVM of all pairs. In the case of \( K \) class classification problem, \( K(K-1)/2 \) classifiers are trained. In the test process, \( (K-1) \) times matching is required by using tournament manner.

3 Experiments

This section shows the experimental results. First, image database is described in section 3.1. Next, section 3.2 demonstrates the effectiveness of the proposed method by the comparison with global kernel based SVM.

3.1 Image database

This paper uses the ORL face database [13] which consists of 40 individuals \( \times \) 10 images. In the following experiments, this database is divided into 3 sets. The first set consists of 40 individuals \( \times \) 5 images. This set is used for training the pairwise SVM. The second set consists of 40 individuals \( \times \) 2 images. This set is used for selecting the optimal parameters of SVM. The third set consists of 40 individuals \( \times \) 3 images. This set is used for evaluating performance. The number of face images becomes small by dividing the database. To avoid this, the number of face images is increased by shifting the original face images 1 pixel vertically and horizontally. In addition, the mirroring of face image is also used. The number of face images is increased 10 times by these processings. The shifting and mirroring of face images are performed only to the first and second sets.

To investigate the robustness to partial occlusion, a square of \( M \times M \) pixels is added to face images in the third data set randomly. In this experiment, the white and black squares are used. The \( M \) is changed from 0 to 50. \( M = 0 \) means non-occlusion. Examples of occluded face images are shown in Figure 2. From left to right, \( M \) is changed from 10 to 50 at the interval of 10. The recognition rate to the face images occluded by the white and black squares is computed as the mean recognition rate of both cases. The size of these face images is \( 112 \times 92 \) pixels. Gabor features are extracted at the interval of one pixel. As a result, 8,736 (= 52 (height) \( \times \) 42 (width) \( \times \) 4 (orientations)) dimensional Gabor features are obtained from one image.

3.2 Performance evaluation

First, the local summation kernel and global Gaussian kernel are compared. Since the proposed method uses Gaussian kernel as the local kernel, this is equivalent to the comparison of local kernel and global kernel. In this experiment, the scalar feature of position \( p \) is used as local feature and the variances of all local kernels are same. Therefore, the local product kernel corresponds to global Gaussian kernel. The result is shown in Figure 3 (a). The horizontal axis represents \( M \) (the size of squares). The vertical axis represents the true recognition rate. From Figure 3 (a), we understand that global Gaussian kernel (local product kernel) is influenced easily by occlusion. In practical environment, human faces are sometimes occluded by sunglass, mask, and other objects. Therefore, the face recognition method using global Gaussian kernel does not work well under practical environment. On the other hand, the proposed local summation kernel is robust to occlusion. The recognition rate is over 80% under large occlusion. This result shows that the proposed method is appropriate for the real-world applications. Furthermore, the performance of the local summation kernel is better than that of the global Gaussian kernel when there is not any occlusion (\( M=0 \)). This shows the effectiveness of the local kernels.

Second, the comparison with global polynomial
kernel is performed. The global polynomial kernel has the summation of the local features \( K(x, y) = \left( 1 + \sum_{p} x(p) \cdot y(p) \right)^d \). Therefore, it is expected that the global polynomial kernel is also robust to occlusion. In this experiment, \( d \) is set to 2 which gives the highest performance to the second face set for parameter setting. The result is shown in Figure 3 (b). From Figure 3 (b), the global polynomial kernel is robust to occlusion. However, its performance is worse than that of the proposed method. The difference of recognition rate is over 20% under large occlusion. This result shows that \( x(p) \cdot y(p) \) in polynomial kernel is not sufficient as the model of the local features. The proposed method uses the Gaussian for modeling the local features. In other words, the local cells tuned to the specific position of each person’s face are made by local Gaussian kernel. By this modeling, high accuracy is obtained.

From these experiments, the effectiveness and robustness of the proposed method are demonstrated.

4 Conclusions

We propose the SVM with local summation kernel for robust face recognition. The robustness to occlusion is obtained by using the summation of local kernels. The effectiveness of the local kernel is shown by the comparison with global kernel based SVM. The recognition rate of the proposed method is over 80% under large occlusion, while the recognition rate of the SVM with global kernel decreases dramatically.

In this paper, a scalar local feature is used in local kernel. It is expected that the performance is improved further by changing the size of local kernel. However, the optimal size of local kernels depends on the position and recognition target. For the robust and accurate recognition, it is necessary to determine these parameters automatically. To solve this problem, we may be able to use non-negative matrix factorization [14] which decomposes the faces into local parts. This is one of the future works.

References