A ROBUST FACE DETECTOR UNDER PARTIAL OCCLUSION

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 a robust face detector under partial occlusion. In recent years, the effectiveness of Support Vector Machine (SVM) to object detection is reported. However, conventional methods apply one kernel to global features. Therefore, those methods are not robust to occlusion because global features are influenced easily by noise or occlusion. To overcome this problem, SVM with local kernels is proposed. It is used to realize a robust face detector under partial occlusion. The robustness of the proposed method under partial occlusion is shown by using the occluded face images. The proposed method can detect the faces wearing sunglasses or scarf. It is also confirmed that the proposed method is superior to the conventional SVM with global kernel.

1. INTRODUCTION

Face detection is the first essential step for automatic face recognition. Since automatic face recognition has many potential applications, face detection becomes an active research area [1, 2]. Some frontal face detection methods give high detection rate under the restricted environment [1, 2]. However, in practical environment and applications, there is some obstacles such as occlusion, illumination direction changes, and view changes. Therefore, the face detection under practical environment is difficult. The robust face detection method to some obstacles is desired. At this point, the robustness to occlusion is important, because human faces are sometimes occluded by sunglasses or mask in real world. Furthermore, the shadows on faces induced by illumination direction changes are considered as one of the occlusion.

In recent years, the effectiveness of Support Vector Machine (SVM) to object detection is reported [3, 4]. However, in the conventional approaches, one kernel is applied to global features. Therefore, those methods are not robust to occlusion, because global features are influenced easily by noise or occlusion. It is considered that local features based recognition is more robust to noise or occlusion, because local features on non-occluded region is not influenced by occlusion. If local similarities are integrated well, it is expected that the robustness to occlusion is realized. For example, Martinez realized the robust recognition under partial occlusion by integrating the local similarities [5]. From these considerations, in order to give SVM the robustness under partial occlusion, it is necessary to treat local features in SVM. In that method, the robustness to occlusion is realized by integrating the local similarities. In this paper, SVM with local kernels is proposed. It is used to realize a robust face detector under partial occlusion.

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 [6, 7]. As described above, how to integrate the local similarities (kernels) is important. It is considered that the local summation kernel is better than local product kernel. The product value of local kernels becomes low when some local kernels give low values. Therefore, the product of local kernels is not robust to noise or occlusion. On the other hand, the summation of local kernels is robust to that case, because the summation is not influenced by low values of some local kernels. Therefore, the summation of local kernels is used to integrate the local kernels. We call this the local summation kernel. In this paper, the robust face detector under partial occlusion is realized by using SVM with local summation kernel.

In recent years, a face detection method based on local SVM is proposed [8]. Although that method uses local features, SVM is applied to all features extracted from local region. Namely, the global kernel is applied to local region. Therefore, that method is different from the proposed method based on local kernels. Since that method uses all features of local region, it is not robust to occlusion.

To show the effectiveness of the propose method, the comparison with the global kernel based SVM is performed. The robustness under partial occlusion is investigated by using the face images to which a white square is added randomly. The proposed method gives high performance under partial occlusion, while the performance of the global kernel based method decreases dramatically. In addition, it is confirmed that the proposed method can detect the faces wearing sunglasses or scarf. The face with shadows is also detected correctly.

In section 2, a face detector based on SVM with local summation kernel is explained. Section 3 shows the effectiveness of the proposed method. Conclusion and future

0-7803-8554-3/04/$20.00 ©2004 IEEE.

works are described in section 4.

2. A ROBUST FACE DETECTOR TO OCCLUSION

This section explains SVM with local summation kernel for robust face detection under partial occlusion. Since the proposed method is based on local kernels, 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

The outputs of Gabor filter are regarded as sparse coding, because Gabor-like receptive fields are obtained by using the constraint which maximizes the sparseness of the response to natural images [9]. It is also reported that Gabor-like filters are obtained by using the info-max network which is able to perform independent components analysis [10]. This means that the outputs of Gabor filters are regarded as independent of each other.

Gabor filters are defined by

$$\psi_{\mathbf{k}}(\mathbf{x}) = \frac{\mathbf{k^2}}{\sigma^2}\exp\left(\frac{-\mathbf{k^2x^2}}{2\sigma^2}\right)\left[\exp(i\mathbf{kx}) - \exp\left(-\sigma^2/2\right)\right], \quad (1)$$

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

2.2. SVM with local summation kernel

First, we explain the SVM [7] briefly. SVM determines the optimal hyperplane which maximizes the margin. The margin is the distance between hyperplane and nearest sample from it. When the training set (sample and its label) is denoted as $S = ((\mathbf{x_i}, \mathbf{y_i}), \ldots, (\mathbf{x_L}, \mathbf{y_L}))$, the decision function is defined by $f(\mathbf{x}) = \sum_{\mathbf{i}}^{\mathbf{L}} \alpha_\mathbf{i}\mathbf{y_i}\mathbf{x_i^T}\mathbf{x}$. α is the solutions of quadratic programming problem. The training sample with non-zero α is called support vector. This decision function assumes the linearly separable case. In the linearly non-separable case, we can use the non-linear transform $\Phi(\mathbf{x})$. The training data is mapped into high dimensional space by $\Phi(\mathbf{x})$. By maximizing the margin in high dimensional space, non-linear classification can be done. If inner product $\Phi(\mathbf{x})^{\mathbf{T}}\mathbf{\Phi}(\mathbf{y})$ in high dimensional space is computed by kernel $K(\mathbf{x}, \mathbf{y})$, then training and classification can be done without mapping into high dimensional space. The decision function using kernel is defined by

$$f(\mathbf{x}) = \sum_{\mathbf{i}}^{\mathbf{L}} \alpha_\mathbf{i}\mathbf{y_i}\mathbf{\Phi}(\mathbf{x_i})^{\mathbf{T}}\mathbf{\Phi}(\mathbf{x}) = \sum_{\mathbf{i}}^{\mathbf{L}} \alpha_\mathbf{i}\mathbf{y_i}\mathbf{K}(\mathbf{x_i}, \mathbf{x}). \quad (2)$$

Mercer's theorem gives whether $K(\mathbf{x}, \mathbf{y})$ is the inner product in high dimensional space. The necessary and sufficient conditions are symmetry $K(\mathbf{x}, \mathbf{y}) = \mathbf{K}(\mathbf{y}, \mathbf{x})$ and positive semi-definiteness of kernel matrix $\mathbf{K} = (\mathbf{K}(\mathbf{x_i}, \mathbf{x_j}))_{\mathbf{i,j=1}}^{\mathbf{L}}$. Examples of kernel which satisfy 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

$$K_p(\mathbf{x(p)}, \mathbf{y(p)}) = \exp\left(-\frac{||\mathbf{x(p)} - \mathbf{y(p)}||^\mathbf{2}}{\sigma_p^2}\right), \quad (3)$$

where p is the label of position. $\mathbf{x(p)}$ and $\mathbf{y(p)}$ represent the feature vector of local region centered at position p. At the simplest case, $x(p)$ and $y(p)$ are the scaler feature of position p.

In order to use local kernels in SVM, it is necessary that kernel value $K(\mathbf{x}, \mathbf{y})$ is computed from local kernels $K_p(\mathbf{x(p)}, \mathbf{y(p)})$ arranged at all positions of recognition target. The product and summation of local Gaussian kernels are considered as the integration method of local kernels. The product and summation of local Gaussian kernel satisfy Mercer's theorem [6, 7]. 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 some local kernels give low values, 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 scaler 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

$$f(\mathbf{x}) = \sum_{\mathbf{i}}^{\mathbf{L}} \alpha_\mathbf{i}\mathbf{y_i}\,\frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{p}}^{\mathbf{N}} \mathbf{K_p}(\mathbf{x_i(p)}, \mathbf{x(p)}), \quad (4)$$

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.

3. EXPERIMENTS

This section shows the effectiveness and robustness of the proposed method. First, image database is described in sec-

tion 3.1. Next, the comparison results and face detection results are shown in section 3.2.

3.1. Image Database

In this paper, HOIP face database[1] and CMU test face database [11] are used as face images. The face regions of these images are cropped by using the positions of eyes, nose, and mouth. The number of cropped face images is 933 (HOIP:300 and CMU:633). Since the number of HOIP face images is small, their mirroring images are also used. Examples of face images are shown in Figure 1 (a). The face images captured under various environment are included. The size of these images is 38×38 pixels. Gabor features are extracted at an interval of 1 pixel from 38×38 pixel's images. As a result, $900 (= 15$ (height) $\times 15$ (width) $\times 4$ (orientations)) dimensional Gabor features are obtained from one image.

In the following experiments, these face images are divided into 3 sets. Each set includes 100 images of HOIP database and 211 images of CMU database. The first set is used for training the SVM. The second set is used for selecting the parameters of SVM. The optimal parameters are determined by using the error rate to the second set. The third set is used for evaluating performance. The number of face images become small by dividing the database. Therefore, the number of face images is increased by shifting the original face images 1 pixel vertically and horizontally. The number of face images is increased 5 times by this processing. The shifting of face images is performed only to the first and second face sets.

On the other hand, the non-face images are obtained by PICS database [12], pbic database [13], and WWW. The $17,750$ images with 38×38 pixels are cropped randomly from PICS and WWW images. Examples of non-face images are shown in Figure 1 (b). These images are divided into 2 sets. The first set which includes $13,350$ images is used for training the SVM. The second set (remaining images) is used for selecting the parameters. To improve the performance, non-face images are gathered by bootstrap [14]. SVM with different kernels may have the different similarity measure. Therefore, the bootstrap of each classifier is performed to the same images independently. If the same non-face region is selected by the classifiers with different kernels, then one of them remains and the others are eliminated. The $10,619$ non-face images are gathered by bootstrap. These images are also used for training the SVM. The 100 pbic images are used for evaluating performance.

Face detection has two measures for evaluation; false positive rate and true positive rate. False positive is that non-face is misclassified as the face class. True positive is that face is classified correctly. To evaluate two measures simultaneously, Receiver Operating Characteristic (ROC) curve is used [8]. In this paper, 100 images obtained from the pbic database are used to evaluate false positive rate. The trained face detector is applied to 100 images at an interval of 1 pixel. The $7,670,478$ non-face regions of 38×38 pixels are obtained from 100 pbic images. All regions are used to compute false positive rate.

[1]The facial data in this paper are used by permission of Softpia Japan, Research and Development Division, HOIP Laboratory. It is strictly prohibited to copy, use, or distribute the facial data without permission.

(a) Examples of face images

(b) Examples of non-face images

Fig. 1. Face and non-face images

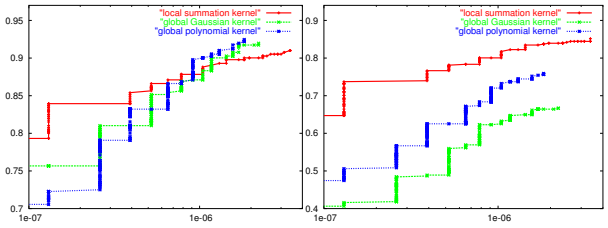
Fig. 2. Examples of occluded face images

On the other hand, true positive rate is computed by using the third face set. In order to investigate the robustness to occlusion, a white square of $M \times M$ pixels is added to face image randomly. We evaluate a case in which M is 0, 5, and 10. $M = 0$ means non-occlusion. Examples of occluded face images are shown in Figure 2.

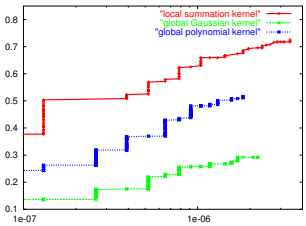
3.2. Performance Evaluation

The proposed method is compared with global Gaussian and global polynomial kernel. Since the proposed method uses Gaussian kernel as local kernels, the comparison with global Gaussian kernel is equivalent to the comparison of the local kernel and global kernel. On the other hand, global polynomial kernel also has the summation of local features $K(\mathbf{x}, \mathbf{y}) = \left(\mathbf{1} + \sum_{\mathbf{p}}^{\mathbf{N}} \mathbf{x}(\mathbf{p}) \cdot \mathbf{y}(\mathbf{p})\right)^{\mathbf{d}}$. Therefore, it is expected that global polynomial kernel is also robust to occlusion. In this experiment, d is set to 2 which gives the highest performance to the data set for parameter setting. The comparison result (ROC curve) is shown in Figure 3. The horizontal axis represents false positive rate on logarithmic scale. The vertical axis represents true positive rate. High true positive rate and low false positive rate means good performance. Therefore, upper left curve is the best. Figure 3 (a) shows the results of non-occlusion case. In the non-occlusion case, the global kernel based methods give better performance. However, at lower false positive rate, the proposed method gives better performance. This shows that faces are classified with high similarity.

Figure 3 (b) shows the performance when a white square of 5×5 pixels is added. In the occluded case, the performance of the global kernel based SVM decreases dramatically. This result shows that global kernel is influenced eas-



(a) Non-occlusion (b) occlusion (5×5 pixels)



(c) occlusion (10×10 pixels)

Fig. 3. Comparison result

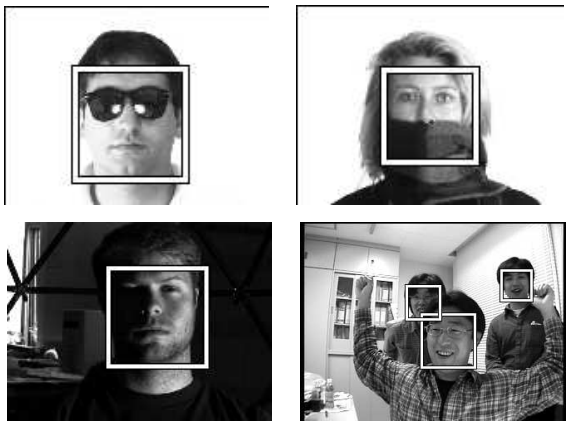


Fig. 4. Examples of face detection results

ily by occlusion. On the other hand, the performance of the proposed method gives high performance. This means that local summation kernel is robust to occlusion. Similarly, in the case of the occlusion by a white square of 10×10 pixels, the proposed method gives the best performance. The performance of global polynomial kernel is better than that of global Gaussian kernel, because polynomial kernel has the summation of local features. However, the modeling of local features is not sufficient. Therefore, its performance is worse than the proposed method.

Finally, the face detection results of the proposed method are shown in Figure 4. We understand that the faces wearing sunglasses or scarf are detected correctly. These images are obtained from AR face database [15]. The shadow on a face is considered as one of the occlusion. Therefore, the face with shadow is also detected correctly. This image is obtained from Yale face database B [16]. In addition, some faces with occlusion are detected correctly.

4. CONCLUSIONS

We present a robust face detector under partial occlusion. The conventional face detection methods based on SVM use global kernel. Those methods are not robust to occlusion because global features (kernel) are influenced easily by occlusion. In order to be robust under partial occlusion, local summation kernel is introduced in SVM.

The robustness to occlusion is confirmed by experiments using the occluded face images. The proposed method gives high performance under partial occlusion, while the performance of the global kernel based SVM decreases dramatically. In addition, the effectiveness and robustness of the proposed method are shown by some face detection results.

5. REFERENCES

[1] E.Hjelmas and B.K.Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, no. 2, pp. 236–274, 2001.

[2] M.-H.Yang, D.Kriegman, and N.Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[3] E.Osuna, R.Freund, and F.Girosi, "Training support vector machines: an application to face detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 130–136.

[4] Y.Li, S.Gong, and H.Liddell, "Support vector regression and classification based multi-view face detection," in *Proc. fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 300–305.

[5] A.M.Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[6] D.Haussler, "Convolution kernels on discrete structures," Tech. Rep., UCSC-CRL-99-10, 1999.

[7] N.Cristianini and J.Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.

[8] B.Heisele, T.Serre, M.Pontil, and T.Poggio, "Component-based face detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 657–662.

[9] B.A.Olshausen and D.J.Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 13, pp. 607–609, 1996.

[10] A.J.Bell and T.J.Sejnowski, "Edes are the 'independent components' of natural scenes," *Vision Research*, vol. 37, no. 23, pp. 3327–3338, 1997.

[11] H.A.Rowley, S.Baluja, and T.Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[12] *The Psychological Image Collection at Stiring University*, http://pics.psych.stir.ac.uk/.

[13] *Pedstrian and Bicycle Information Center Image Library*, http://www.pedebikeimages.org/ Dan Burden.

[14] K.Sung and T.Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[15] A.M.Martinez and R.Benavente, *The AR face database*, CVC Technical Report 24, 1998.

[16] A.S.Georghiades, P.N.Belhumeur, and D.J.Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.