View Independent Face Detection Based on Combination of Local and Global Kernels

Kazuhiro HOTTA

The University of Electro-Communications,
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, JAPAN
hotta@ice.uec.ac.jp,

Abstract. In this paper, local and global kernels are combined to use the detailed and rough similarities simultaneously. In recent years, many recognition methods based on local features have been proposed. However, the combination of only local matching is not sufficient. Global viewpoint is also necessary to improve the generalization ability. In general, local feature matching measures the detailed similarity and global feature matching measures the rough similarity. Therefore, the error pattern is different in local and global features. If they are combined well, the generalization ability is improved. In the proposed method, local kernels and global kernel are combined by summation, and the combined kernel is used in SVM. The proposed method is applied to view independent face detection task. We confirm that the false positive is reduced by combining local and global kernels. The effectiveness of the proposed method is demonstrated by the comparison with only global and local kernels.

1 Introduction

View independent recognition is a big open problem in face detection and recognition task [1–3]. Since view changes induce large non-linear variations in feature space [4, 5], view independent recognition is difficult. Figure 1 shows eigen space under view changes. This space is constructed by using 5 views of some subjects. We can see non-linear variations induced by view changes. To be robust to view changes, we must cope with the non-linear variations.

Kernel-based methods can represent non-linear variations easily. Therefore, kernel-based method is used to cope with view changes [6, 7]. In particular, Support Vector Machine (SVM) is usually used in object detection task because it is a binary-classification problem. In recent years, SVM with local kernels have been proposed to use local features effectively [8–10]. It is reported that the robustness to partial occlusion is improved by using local features in SVM [11]. However, only use of local features do not achieve high generalization ability like humans. Global viewpoint is also necessary to improve accuracy further. Local kernel measures the detailed similarity and global kernel measures rough similarity. Therefore, the error patterns are different in local features and global features. If local and global kernels are combined, a good detector will be developed using both detailed and rough similarities. In this paper, we show that detection accuracy is improved by using SVM based on the combination of local and global kernels.

In general, view changes induce large appearance variations in horizontal direction. Therefore, a robust face detector under view changes can be developed if the appearance

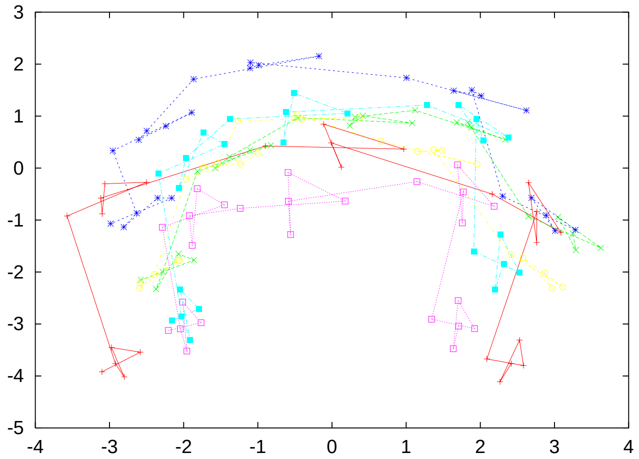


Fig. 1. Eigen space under view changes

changes in horizontal direction are represented well. To represent the appearance in horizontal direction, a face region is divided into horizontal rectangles. Local kernel is applied to each horizontal rectangle. Local kernel measures the similarity between horizontal regions like a montage picture. In this paper, these detailed similarities are combined with rough similarity of global kernel. To combine local kernels and global kernel, the summation of them is used. It is known that the summation of classifiers gives good performance [12, 13]. Of course, the summation of kernels satisfies Mercer's theorem [14, 15]. Since we use normalized polynomial kernel whose output is between 0 and 1 as a kernel function, the output of each kernel is combined fairly.

To evaluate the proposed method based on the combination of local and global kernels, many face images of various views and non-face images are gathered from some databases or WWW [16–18]. The generalization ability is evaluated by using Receiver Operating Characteristic curve. First, we compare the SVM with only global kernel and SVM with summation of local kernels. Next, the horizontal rectangle features are compared with the vertical rectangle features, and the effectiveness of horizontal rectangle features is shown. Finally, local kernels of horizontal rectangle features and global kernel are combined and used in SVM. We confirm that the combination of local and global kernels improves the generalization ability. In particular, false positive is reduced by combining the detailed and rough similarities. This result suggests that we should recognize the importance of global features again.

In section 2, we explain a SVM with summation kernel of rectangle features for view independent face detection. Section 3 shows the effectiveness of the proposed method by comparison with standard SVM. Conclusions and future works are described in section 4.

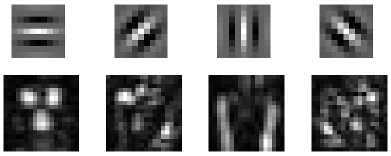


Fig. 2. Gabor filters and Gabor features

2 Proposed method

In this paper, we want to use local appearance features to utilize local kernels effectively. For this purpose, Gabor features are used. It is known that Gabor features are robust to illumination changes. In addition, they are effective to object detection and recognition [11, 19, 20]. In section 2.1, the properties of Gabor filter are explained. Section 2.2 explains SVM based on the combination of local and global kernels.

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 [21].

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 [22]. It is also reported that Gabor-like filters are obtained by independent components analysis of natural images [23].

Gabor filters are defined by

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

where $\boldsymbol{x} = (x, y)^T$, $\boldsymbol{k} = k_\nu \exp(i\phi)$, $k_\nu = k_{max}/f^\nu$, $\phi = \mu \cdot \pi/4$ and $f = \sqrt{2}$. There are some results that Gabor features of only 1 frequency level gives good detection performance [11]. Therefore, in the following experiments, Gabor filters of 4 different orientations with 1 frequency level are used in order to speed up the recognition. The size of Gabor filters is set to 9×9 pixels. Figure 2 shows the Gabor filters of 4 different orientations and the Gabor features of a frontal face image. Gabor outputs at many positions of face images are small and only specific positions give large values. This represents the sparseness of Gabor features.

2.2 SVM based on combination of local and global kernels

First, we explain the SVM [24, 14] 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 $S = ((\boldsymbol{x}_i, y_i), \ldots, (\boldsymbol{x}_L, y_L))$, the optimal hyperplane is defined by

$$f(\boldsymbol{x}) = \sum_{i \in SV} \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x} + b, \tag{2}$$

where SV is a set of support vectors, b is the threshold and α is the solutions of quadratic programming problem. The training samples with non-zero α are called support vectors.

This assumes the linearly separable case. In the linearly non-separable case, the non-linear transform $\Phi(\boldsymbol{x})$ can be used. The training samples are mapped into high dimensional space by $\Phi(\boldsymbol{x})$. By maximizing the margin in high dimensional space, non-linear classification can be done. If inner product $\Phi(\boldsymbol{x})^T\Phi(\boldsymbol{y})$ in high dimensional space is computed by kernel $K(\boldsymbol{x},\boldsymbol{y})$, then training and classification can be done without mapping into high dimensional space. The optimal hyperplane using kernel is defined by

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i\in SV} \alpha_i y_i \Phi(\boldsymbol{x}_i)^T \Phi(\boldsymbol{x}) + b, \\ &= \sum_{i\in SV} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b. \end{aligned} \tag{3}$$

Mercer's theorem gives whether $K(\boldsymbol{x},\boldsymbol{y})$ is the inner product in high dimensional space. The necessary and sufficient conditions are symmetry $K(\boldsymbol{x},\boldsymbol{y}) = K(\boldsymbol{y},\boldsymbol{x})$ and positive semi-definiteness of kernel matrix $\boldsymbol{K} = (K(\boldsymbol{x}_i,\boldsymbol{x}_j))_{i,j=1}^{L}$. If $\boldsymbol{\beta}^T\boldsymbol{K}\boldsymbol{\beta} \geq 0$ where $\beta \in \Re$ is satisfied, $\boldsymbol{K}$ is a positive semi-definite matrix. It is known that summation and production of kernels satisfies Mercer's theorem [14, 15].

Next, we consider the type of kernel function. Gaussian kernel gives good performance when the parameter is set well. However, the optimal parameter selection is not easy task. It is reported that normalized polynomial kernel gives the comparable performance with Gaussian kernel using optimal parameter [25]. In addition, parameter dependency of normalized polynomial kernel is much lower than that of Gaussian kernel. Of course, normalized kernel satisfies Mercer's theorem [26]. Therefore, we use normalized polynomial kernel as the kernel function. Normalized polynomial kernel is defined as

$$K(\boldsymbol{x},\boldsymbol{y}) = \frac{(1+\boldsymbol{x}^T\boldsymbol{y})^d}{\sqrt{(1+\boldsymbol{x}^T\boldsymbol{x})^d\,(1+\boldsymbol{y}^T\boldsymbol{y})^d}}. \tag{4}$$

By normalizing the output of standard polynomial kernel, the kernel value is between 0 and 1 like Gaussian kernel. In the following experiments, the optimal value of d is determined by using the error rate to validation set.

In this paper, normalized polynomial kernel is applied to horizontal rectangle features and global features. Local kernel applied to horizontal rectangle features is defined as

$$K_{lp}(\boldsymbol{x},\boldsymbol{y}) = K_l(A_p^T\boldsymbol{x}, A_p^T\boldsymbol{y}) = \frac{(1+\boldsymbol{x}^T A_p A_p^T \boldsymbol{y})^d}{\sqrt{(1+\boldsymbol{x}^T A_p A_p^T \boldsymbol{x})^d\,(1+\boldsymbol{y}^T A_p A_p^T \boldsymbol{y})^d}}, \tag{5}$$

where A_p is the diagonal matrix and p is the p-th horizontal rectangle features. We assign 1 to only diagonals that correspond to the elements used as local features in local kernel. The other elements in A_p are 0. For example, $A_p(1,1)$ is 1 when the first element of a feature vector is used in local kernel. On the other hand, global kernel applied to global features is denoted as $K_g(\boldsymbol{x},\boldsymbol{y})$. This is the same as equation (4) when $\boldsymbol{x}$ and $\boldsymbol{y}$ are global features.

The local and global kernels are combined and used as a kernel in SVM. In this paper, the summation is used to combine the kernels. It is reported that the summation kernel is more robust to partial occlusion than the production kernel [11]. In [11], only local kernels with same size are combined. In this paper, local and global kernels are combined to use the detailed and rough similarities simultaneously. We demonstrate that the combination of local and global kernel improves the generalization ability. The combined kernel is defined as

$$K_{comb}(\boldsymbol{x},\boldsymbol{y}) = K_g(\boldsymbol{x},\boldsymbol{y}) + \sum_p K_{lp}(\boldsymbol{x},\boldsymbol{y}). \tag{6}$$

The proof that this kernel satisfies Mercer's theorem is easy [14]. When K_g and K_l satisfy Mercer's theorem, $\boldsymbol{\beta}^T \boldsymbol{K}_g \boldsymbol{\beta} \geq 0$ and $\boldsymbol{\beta}^T \boldsymbol{K}_l \boldsymbol{\beta} \geq 0$ where $\boldsymbol{K}_g$ is the kernel matrix of K_g, $\boldsymbol{K}_l$ is the kernel matrix of K_l and $\boldsymbol{\beta} \in \Re$. Therefore, the summation of them satisfies Mercer's theorem as follows.

$$\boldsymbol{\beta}^T(\boldsymbol{K}_g + \boldsymbol{K}_l)\boldsymbol{\beta} = \boldsymbol{\beta}^T \boldsymbol{K}_g \boldsymbol{\beta} + \boldsymbol{\beta}^T \boldsymbol{K}_l \boldsymbol{\beta} \geq 0. \tag{7}$$

In the following experiments, the combination of local and global kernels is compared with only global kernel and the summation of only local kernels. The effectiveness of combination of local and global kernels is demonstrated in section 3.

3 Experiments

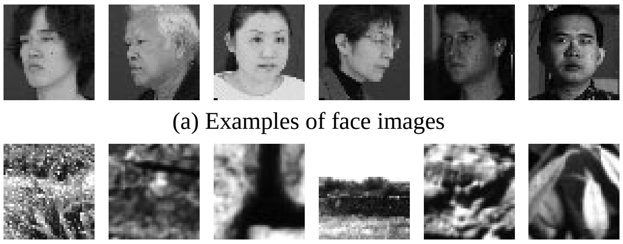
In section 3.1, Image database used in this paper is described. Section 3.2 shows the evaluation result on face detection. First, the results of only local kernels and global kernel are shown. Next, the effectiveness of horizontal rectangle features under view changes is shown by the comparison with vertical rectangle features. Finally, we demonstrate that the combination of local and global kernels improves generalization ability.

3.1 Image database

In this paper, HOIP face database[1] and PICS database [16] are used as face images. The face images with little shadows obtained from PIE face database [17] are also used. The face regions of these images are cropped by using the positions of eyes, nose and mouth. Examples of face images are shown in Figure 3 (a). The face images captured under various environment are included. The size of these images is 44×44 pixels. Gabor features are extracted at an interval of 1 pixel from 44×44 pixel's images. Since the size of Gabor filters is 9×9 pixels, the peripheral 4 pixels of each side of a image can not extract Gabor features. As a result, $1,296$ (= 18 (height) $\times$ 18 (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 $2,010$ face images of various views. 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 the true 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. 3. Face and non-face images

On the other hand, the non-face images are obtained by PICS database [16], pbic database [18] and WWW. The $17,750$ images with 44×44 pixels are cropped randomly from PICS and WWW images. Examples of non-face images are shown in Figure 3 (b). These images are divided into 2 sets. The first set which includes $8,875$ images is used for training the SVM. The second set (remaining images) is used for selecting the parameters of SVM. The 100 pbic images are used for evaluating false positive rate.

3.2 Evaluation

First, we explain how to evaluate the performance. Face detection has two measures for evaluation; false positive rate (FPR) and true positive rate (TPR). False positive is that non-face sample is misclassified as the face class. True positive is that face sample is classified correctly. To evaluate two measures simultaneously, Receiver Operating Characteristic (ROC) curve is used [27]. When the threshold b in equation (3) is changed, FPR and TPR are changed. Therefore, the performance of a classifier becomes a curve on FPR-TPR space. 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 44×44 pixels are obtained from 100 pbic images. All regions are used to compute false positive rate. On the other hand, true positive rate is computed by using the third face set described in previous section. In the following experiments, the threshold is changed until TPR achieves 99%.

In this paper, local kernels are applied to horizontal rectangle features with non-overlapped manner. Figure 4 shows how to apply local kernels. The horizontal rectangles on Gabor features show local kernels. Since Gabor features extracted from one sample is 18 (height) $\times$ 18 (width) $\times$ 4 (orientations) dimensional features, local kernels are applied to $3 \times 18 \times 4$, $6 \times 18 \times 4$ and $9 \times 18 \times 4$ dimensional horizontal rectangle regions. When local kernels are applied to $3 \times 18 \times 4$ dimensional region, 6 local kernels are summed because local kernels are arranged with non-overlapped manner.

First, we evaluate SVM with only global kernel and SVM with summation of only local kernels. Figure 5 shows the ROC curves of them. The horizontal axis shows FPR. The vertical axis shows TPR. High TPR and low FPR means good performance. Therefore, upper left curve is the best. Namely, SVM with local kernels of $6 \times 18 \times 4$ dimensions is the best. The horizontal rectangle regions of $6 \times 18 \times 4$ dimensions cover eyes, nose and mouth well. Therefore, this kernel gives the superior performance than others.

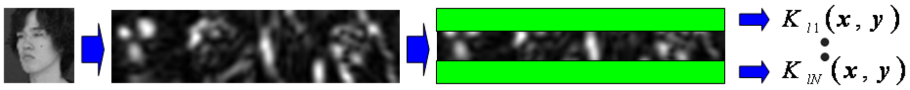

$K_{l1}(\boldsymbol{x}, \boldsymbol{y})$
$K_{lN}(\boldsymbol{x}, \boldsymbol{y})$

Fig. 4. How to apply local kernels

To show the effectiveness of horizontal rectangle features under view changes, it is compared with vertical rectangle features. In the previous experiment, horizontal features of $6 \times 18 \times 4$ dimensions gives the best result. Therefore, local kernels are applied to the vertical rectangle features of $18 \times 6 \times 4$ dimensional regions with non-overlapped manner. How to evaluate is the same as the previous experiment. Figure 6 shows the comparison result. We understand that horizontal rectangle features gives much better accuracy under view changes than vertical rectangle features. This is because the view changes include large appearance variations in horizontal direction. Therefore, a robust detector under view changes can be developed if horizontal appearance variations are represented well.

Finally, we combine global kernel and local kernels of horizontal rectangle features to improve the generalization ability further. Local kernel measures the detailed similarity and global kernel measures the rough similarity. Therefore, there is the case that local features based method mis-classifies the samples which are classified easily by global features. By combining local and global kernels, both properties of local and global features are used simultaneously. In this experiment, we combine global kernel and local kernel of $6 \times 18 \times 4$ dimensions. Figure 7 shows the ROC curve. To compare the performance, the ROC curves of only global kernel and the summation of only local kernels of $6 \times 18 \times 4$ dimensions are also shown. Figure 7 demonstrates that the combination of local and global kernels improves the generalization ability. In particular, the number of false positive is reduced by combining local and global kernels. This result means that false positive samples which are similar locally are classified correctly by using rough similarity. Namely, the properties of local and global kernels are combined well in SVM.

4 Conclusion

In this paper, local and global kernels are combined to use the detailed and rough similarities simultaneously. The recognition methods based on local features have been proposed in recent years. However, the use of only local similarities is not sufficient. Global viewpoint is also necessary to improve the generalization ability because error pattern is not the same in local and global features. In the experiments, the false positive is reduced by combining local and global kernels, the generalization ability is improved. This result suggests that we should recognize the importance of global features again.

In this paper, the proposed method is applied to view independent face detection task. However, the proposed idea can be applied to other recognition tasks. It may be useful in context-based recognition [28, 29]. Global context information can be introduced as a global kernel directly. The combination of local features and global context may improve the recognition accuracy. The proposed method will be applied to general object recognition task [30, 31], and the effectiveness of global context will be evaluated.

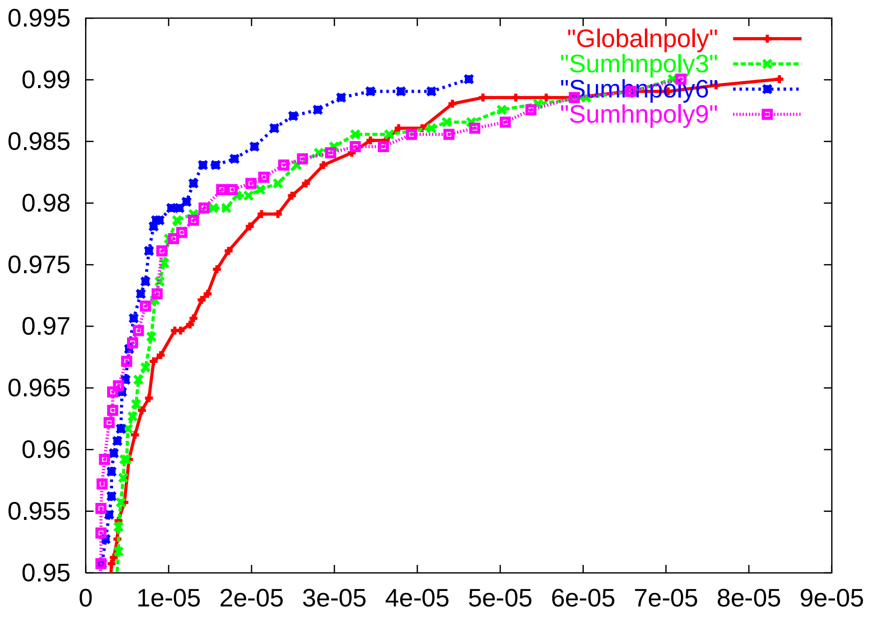


Fig. 5. Comparison results using ROC curve

References

1. E.Hjelmas and B.K.Low, "Face detection: A survey," *Computer Vision and Image Understanding* **83**(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* **24**(1), pp. 34–58, 2002.
3. W.Zhao, R.Chellappa, P.J.Phillips, and A.Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys* **35**(4), pp. 399–458, 2003.
4. P.J.Phillips, P.Grother, R.J.Micheals, D.H.Blackburn, E.Tabassi, and J.M.Bone, "Frvt2002: Evaluation report," tech. rep., NISTIR6965, http://frvt.org, 2003.
5. H.Murase and S.K.Nayar, "Visual learning and recognition of 3d objects from appearance," *International Journal of Computer Vision* **14**(1), pp. 5–24, 1995.
6. 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*, pp. 300–305, 2000.
7. K.Hotta, "View independent video-based face recognition using posterior probability in kernel fisher discriminant space," in *Proc. 3rd International Conference on Advances in Pattern Recognition (Lectures Notes in Computer Science, Vol.3687)*, pp. 103–111, 2005.
8. K.Hotta, "Support vector machine with local summation kernel for robust face recognition," in *Proc. 17th International Conference on Pattern Recognition*, pp. 482–485, 2004.
9. C.Wallraven and B.Caputo, "Recognition with local features: the kernel recipe," in *Proc. 9th IEEE International Conference on Computer Vision*, pp. 257–264, 2003.
10. S.Boughorbel, J.-P.Tarel, and F.Fleuret, "Non-mecer kernels for svm object recognition," in *Proc. British Machine Vision Conference*, 2004.

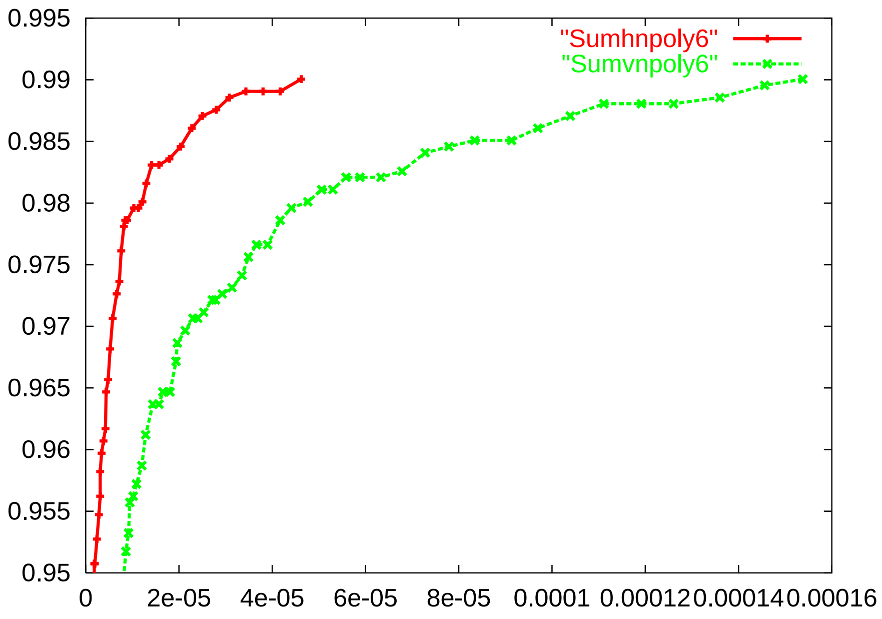


Fig. 6. Comparison horizontal and vertical rectangle features

11. K.Hotta, "A robust face detector under partial occlusion," in *Proc. IEEE International Conference on Image Processing*, pp. 597–600, 2004.
12. J.Kittler, M.Hatef, R.P.W.Duin, and J.Matas, "On combining classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(3), pp. 226–239, 1998.
13. P.Viola and M.J.Jones, "Robust real-time face detection," *International Journal of Computer Vision* **57**(2), pp. 137–154, 2004.
14. N.Cristianini and J.Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, 2000.
15. D.Haussler, "Convolution kernels on discrete structures," tech. rep., UCSC-CRL-99-10, 1999.
16. *The Psychological Image Collection at Stiring University*. http://pics.psych.stir.ac.uk/.
17. T.Sim, S.Baker, and M.Bsat, "The cmu pose, illumination, and expression (pie) database," in *Proc. fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 53–58, 2002.
18. *Pedstrian and Bicycle Information Center Image Library*. http://www.pedebikeimages.org/ Dan Burden.
19. M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R. P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computer* **42**(3), pp. 300–311, 1993.
20. T.Serre, L.Wolf, and T.Poggio, "Object recognition with features inspired by visual cortex," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 994–1000, 2005.
21. J.P.Jones and L.A.Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in the cat striate cortex," *J. Neurophysiology* **58**, pp. 1233–1258, 1987.
22. B.A.Olshausen and D.J.Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**(13), pp. 607–609, 1996.

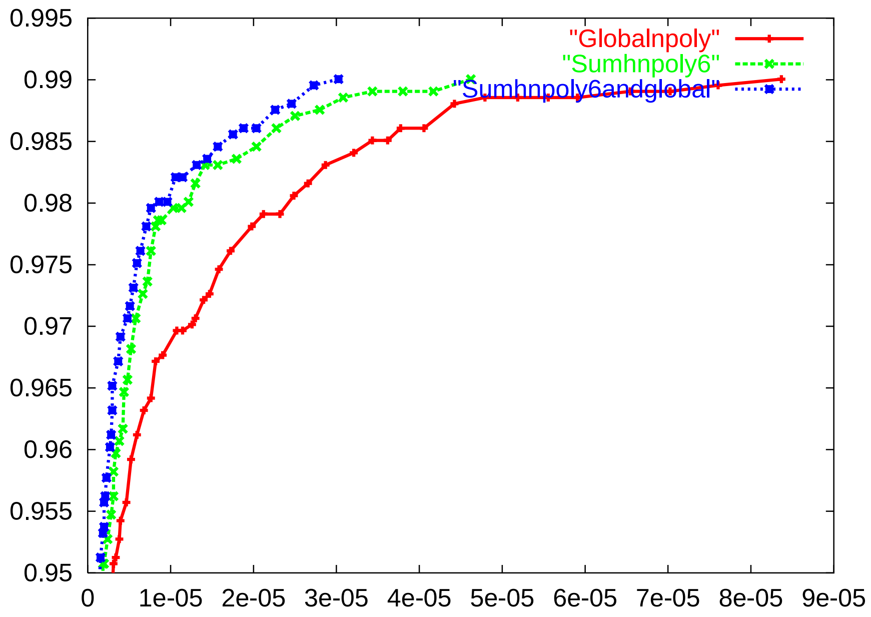


Fig. 7. Combination of local and global kernels

23. A.J.Bell and T.J.Sejnowski, “Edes are the ’independent components’ of natural scenes,” *Vision Research* **37**(23), pp. 3327–3338, 1997.
24. V.N.Vapnik, *Statistical Learning Theory*, John Wiley & Sons, 1998.
25. R.Debnath and H.Takahashi, “Kernel selection for the support vector machine,” *IEICE Trans. Info. & Syst.* **E87-D**(12), pp. 2903–2904, 2004.
26. J.Shawe-Taylor and N.Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
27. 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*, pp. 657–662, 2001.
28. A.Torralba, K.P.Murphy, W.T.Freeman, and M.A.Rubin, “Context-based vision system for place and object recogtnition,” in *Proc. International Conference on Computer Vision*, pp. 273–280, 2003.
29. A. E.B.Sudderth, W.T.Freeman, and A.S.Willsky, “Learning hierarchical models of scenes, object, and parts,” in *Proc. International Conference on Computer Vision*, pp. 1331–1338, 2005.
30. L.Fei-Fei, R.Fergus, and P.Perona, “Learning generative visual models from few training examples: an incremenal bayesian approach tested on 1010 object categories,” in *Proc. IEEE CVPR Workshop of Generative Model Based Vision*, pp. 994–1000, 2004.
31. *Caltech 101 database*. http:/www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html.