PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE @ DIRECT®

Pattern Recognition 38 (2005) 651 – 659

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Improved feature reduction in input and feature spaces

Frank Y. Shih*, Shouxian Cheng

Computer Vision Laboratory, Department of Computer Science, College of Computing Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

Received 12 May 2004; received in revised form 25 October 2004; accepted 25 October 2004

Abstract

In this paper, we present an improved feature reduction method in input and feature spaces for classification using support vector machines (SVMs). In the input space, we select a subset of input features by ranking their contributions to the decision function. In the feature space, features are ranked according to the weighted support vector in each dimension. By applying feature reduction in both input and feature spaces, we develop a fast non-linear SVM without a significant loss in performance. We have tested the proposed method on the detection of face, person, and car. Subsets of features are chosen from pixel values for face detection and from Haar wavelet features for person and car detection. The experimental results show that the proposed feature reduction method works successfully. In fact, our method performs better than the methods of using all the features and the Fisher's features in the detection of person and car. We also gain the advantage of speed.

Keywords: Feature reduction; Feature ranking; Support vector machine; Object detection

1. Introduction

Feature extraction and reduction are two primary issues in feature selection that is essential in pattern classification. Feature extraction is used to achieve high classification rates by extracting features to represent objects from raw data. Feature reduction is used to select a subset of features with preservation or improvement of classification rates. In general, it intends to speed up the classification process by keeping the most important class-relevant features.

Support vector machines (SVMs) [1,2] are founded from a mathematical point of view. While most classifiers (e.g., Bayesian, neural networks, and radial basis function (RBF)) are trained to minimize the empirical risk, SVMs are implemented to minimize the structural risk. Osuna et al. [3] applied SVMs to face detection. Heisele et al. [4] trained a 2nd-degree polynomial SVM using 10,038 faces and 36,220 non-faces, and achieved a higher detection rate than [3]. However, they need to spend several minutes to search an image for faces at different scales. Heisele et al. [5] presented two following methods to speed up face detection using SVMs: hierarchical classification and feature reduction.

Principal component analysis (PCA) features are used to reduce the dimensionality in input space. Weston et al. [6] developed a feature reduction method by minimizing the bounds on the leave-one-out error. Evgenious et al. [7] introduced a method for feature selection based on the observation that the most important features are the ones that separate the hyperplane the most.

In this paper, we present the method of feature reduction in the input and feature spaces to achieve a fast non-linear SVM without a significant loss in performance. The rest of the paper is organized as follows. Sections 2 and 3 describe the feature reduction methods in input and feature spaces, respectively, for face detection. In Section 4, we develop a fast non-linear SVM by the combination of input and feature spaces for face detection. In Section 5, we apply the

* Corresponding author. Tel.: +1 973 596 5654; fax: +1 973 596 5777.
E-mail address: shih@njit.edu (F.Y. Shih).

0031-3203/$30.00
doi:10.1016/j.patcog.2004.10.004

proposed method for person and car detection. Finally, we provide discussions and conclusions in Section 6.

2. Feature reduction in input space

2.1. Feature ranking

Principal component analysis (PCA) is widely used in image representation for dimensionality reduction. To obtain m principal components, we multiply a transformation matrix of $m \times N$ by an input pattern of $N \times 1$. The computation is costly. In this section, we propose a method of feature reduction in the input space in order to save computational time.

One way of feature reduction is using Fisher's criterion to choose a subset of features that possess a large between-class variance and a small within-class variance. For face detection, we use the within-class variance as

$$\sigma_i^2 = \frac{\sum_{j=1}^{l}(g_{j,i} - m_i)^2}{l-1}, \tag{1}$$

where l is the total number of samples, $g_{j,i}$ is the ith dimensional gray value of sample j, and m_i is the mean value of the ith dimension. We use Fisher's score for between-class measurement as

$$S_i = \left| \frac{m_{i,face} - m_{i,nonface}}{\sigma_{i,face}^2 + \sigma_{i,nonface}^2} \right|. \tag{2}$$

By selecting the features with the highest Fisher's scores, we can retain the most discriminative features between face and non-face classes.

To improve Fisher's method, we propose a 2nd-degree polynomial SVM with kernel $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^2$. The decision function for a pattern $\mathbf{x}$ is defined as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{s} \alpha_i y_i (1 + \mathbf{x}_i \cdot \mathbf{x})^2 + b \\ &= \sum_{i=1}^{s} \alpha_i y_i (1 + x_{i,1}x_1 + x_{i,2}x_2 + \cdots + x_{i,k}x_k \\ &\quad + \cdots + x_{i,N}x_N)^2 + b, \end{aligned} \tag{3}$$

where s is the total number of support vectors, $\mathbf{x}_i$ is the ith support vector, and $x_{i,k}$ and x_k are respectively the kth dimension for the support vector $\mathbf{x}_i$ and the pattern $\mathbf{x}$. The component in the kth dimension (where $k = 1, 2, \ldots, N$) is

$$\begin{aligned} f(\mathbf{x}, k) &= \sum_{i=1}^{s} \alpha_i y_i [2x_k x_{i,k}(1 + x_{i,1}x_1 + \cdots + x_{i,k-1}x_{k-1} \\ &\quad + x_{i,k+1}x_{k+1} + \cdots + x_{i,N}x_N) + x_k^2 x_{i,k}^2] \\ &= \sum_{i=1}^{s} \alpha_i y_i [2x_k x_{i,k}(1 + x_{i,1}x_1 \\ &\quad + \cdots + x_{i,N}x_N) - x_k^2 x_{i,k}^2]. \end{aligned} \tag{4}$$

We use the largest m contributions to the decision function out of the original N features. The contribution can be obtained by

$$F(k) = \int_V f(\mathbf{x}, k)\, \mathrm{d}P(\mathbf{x}), \tag{5}$$

where V denotes the input space and $P(\mathbf{x})$ denotes the probability distribution function. Since $P(\mathbf{x})$ is unknown, we approximate $F(k)$ using a summation over the support vectors as

$$\begin{aligned} F(k) = \sum_{i=1}^{s} \Bigg| &\sum_{j=1}^{s} \alpha_j y_j [2x_{i,k}x_{j,k}(1 + x_{j,1}x_{i,1} \\ &+ \cdots + x_{j,N}x_{i,N}) - x_{i,k}^2 x_{j,k}^2] \Bigg|. \end{aligned} \tag{6}$$

2.2. Experimental results

We adopt a face image database from the Center for Biological and Computational Learning at Massachusetts Institute of Technology (MIT), which contains 2429 face training samples, 472 face testing samples, and 23,573 non-face testing samples. We randomly collected 15,228 non-face training samples from the images that do not contain faces. The size of all these samples is 19×19. A 2nd-degree polynomial SVM with kernel $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^2$ is used in our experiments.

In order to remove background pixels, a mask is applied to extract only the face. Prior to classification, we perform image normalization and histogram equalization. The image normalization is used to normalize the gray-level distribution by the Gaussian function with zero mean and one variance. The histogram equalization uses a transformation function equal to the cumulative distribution to produce an image whose gray levels have a uniform density. Fig. 1 shows (a) a face image, (b) the mask, and (c) and (d) the images after normalization and histogram equalization, respectively.

Fig. 2 shows the receiver operating characteristic (ROC) curves of using two following methods to obtain PCA features: one uses both face and non-face training samples,

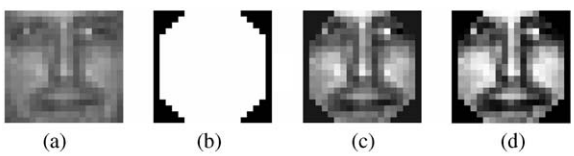

Fig. 1. (a) Original face image, (b) the mask, (c) normalized image, and (d) histogram equalized image.

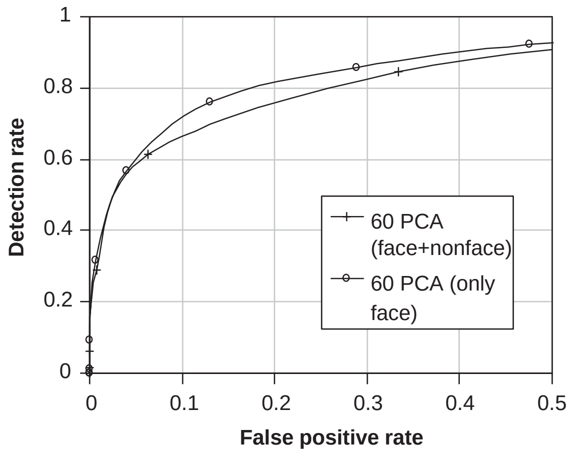

Fig. 2. ROC curves of using two following methods to obtain PCA features: one uses both face and non-face training samples, and the other uses only face training samples.

and the other uses only face training samples. The ROC curve is defined as shifting the SVM hyperplane by changing the threshold value b. We perform face classification on the testing set and calculate the false positive and the detection rates. The horizontal axis shows the false positive rate over 23,573 non-face testing samples. The vertical axis shows the detection rate over 472 face testing samples. We observe that using only positive training samples performs better than using both positive and negative training samples. We also test on the 3600 faces extracted from FERET database and obtain the same result. The reason is that only training face samples are used in the calculation of the transformation matrix, and later for testing the input samples are projected on the face space, so that better classification results can be achieved in separating face and non-face classes.

We test on different preprocessing methods: image normalization, histogram equalization, and without preprocessing, and obtain that using normalization or equalization can produce better results than without preprocessing. Therefore, in the following experiments, we will use image normalization as the pre-processing method and use positive training samples to calculate PCA values.

By using the normalized 2429 face and 15,228 non-face training samples and taking all the 283 gray values as input to train the 2nd-degree SVM, we obtain 252 and 514 support vectors for face and non-face classes, respectively. By using these support vectors in Eq. (6), we obtain $F(k)$, where $k = 1, 2, \ldots, 283$. Fig. 3 shows the ROC curves for different features in input space. We compare our ranking method of using 100 features with the methods of using all the 283 gray values, 100 PCA features, Fisher's scores, and the 100 features selected by Evgenious et al. [7]. We observe that

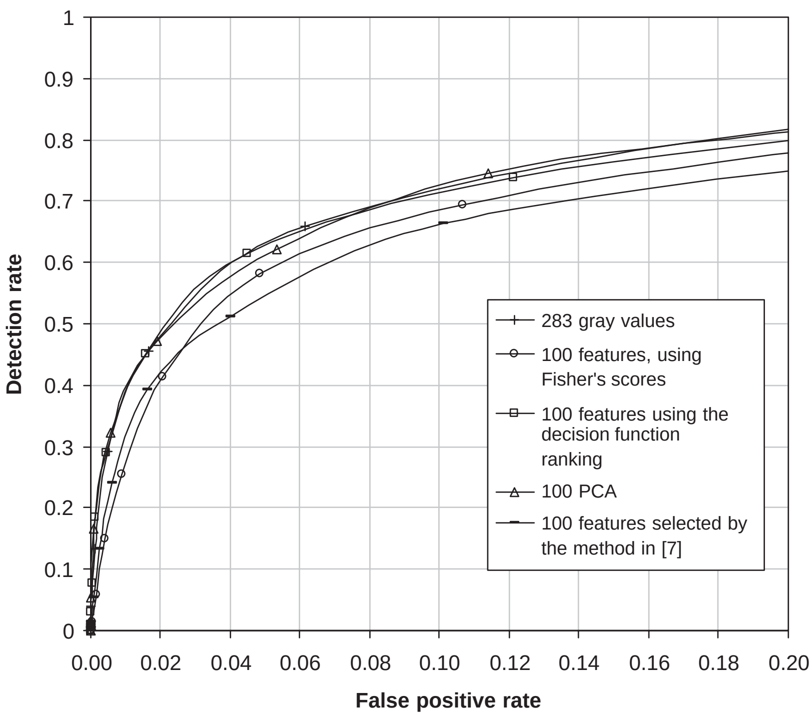

Fig. 3. ROC curves for comparisons of using different features in input space.

using our method performs similarly as using all the 283 features and using 100 PCA values; however, it is better than using the 100 features by Fisher's scores and by [7].

3. Feature reduction in feature space

3.1. Feature ranking

In feature space, the decision function *f(x)* of SVMs is defined as

$$f(\mathbf{x}) = \sum_{i=1}^{s} \alpha_i y_i (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)) + b = \mathbf{w} \cdot \Phi(\mathbf{x}) + b, \tag{7}$$

where $\mathbf{w}$ is the support vector. For a 2nd-degree polynomial SVM with the input space of dimension N and kernel $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^2$, the feature space is given by

$$\Phi(\mathbf{x}) = \Big(\sqrt{2}x_1, \ldots, \sqrt{2}x_N, x_1^2, \ldots, x_N^2, \sqrt{2}x_1x_2, \ldots, \sqrt{2}x_{N-1}x_N\Big) \tag{8}$$

of dimension $P = N(N+3)/2$.

Suppose that we train a 2nd-degree SVM using face and non-face samples to obtain s support vectors. The support vector in the feature space can be represented as

$$\mathbf{w} = \sum_{i=1}^{s} \alpha_i y_i \Phi(\mathbf{x}_i) = (w_1,\ w_2, \ldots, w_P). \tag{9}$$

One way to select a subset of features is to rank $|w_k|$, for $k = 1, 2, \ldots, P$. In this section, we propose an improved method of using $\left|w_k \int_V |x_k^*| \mathrm{d}p(x_k^*)\right|$, where x_k^* denotes the kth dimension of $\mathbf{x}$ in the feature space V. Since the distribution function $\mathrm{d}p(x_k^*)$ is unknown, we use the ranking function $R(k)$ as

$$R(k) = \left| w_k \sum_{i=1}^{s} |x_{i,k}^*| \right|, \tag{10}$$

where $x_{i,k}^*$ denotes the kth dimension of $\mathbf{x}_i$. The decision function of q features is calculated as

$$f(\mathbf{x}, q) = \mathbf{w}(q) \cdot \Phi(\mathbf{x}, q) + b, \tag{11}$$

where $\mathbf{w}(q)$ is the selected q features in $\mathbf{w}$, and $\Phi(\mathbf{x}, q)$ is the corresponding q features in $\mathbf{x}$.

For a pattern $\mathbf{x}$, we calculate the difference of two decision values of using all the features and using the subset of q features as

$$\Delta f_q(\mathbf{x}) = |f(\mathbf{x}) - f(\mathbf{x}, q)|. \tag{12}$$

We sum up the differences over all the support vectors as

$$\Delta F_q = \sum_{i=1}^{s} \Delta f_q(\mathbf{x}_i). \tag{13}$$

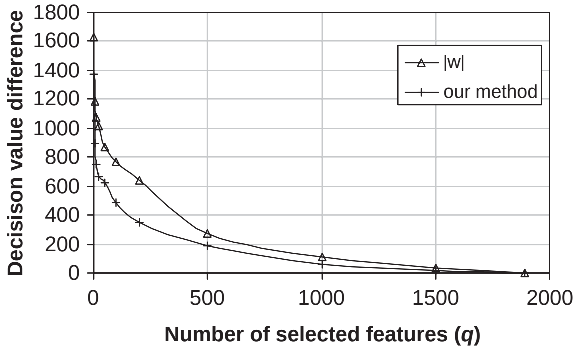

Fig. 4. Decision value differences relative to the subset of q features in the feature space selected by two ranking methods.

3.2. Experimental results

In the experiments, we train a 2nd-degree polynomial SVM using 60 PCA values in the input space. The training samples are the same as in Section 2.2. We obtain 289 and 412 support vectors for face and non-face classes, respectively. The 1890 features in the feature space can be calculated by Eq. (8). The support vector in the feature space, $\mathbf{w} = (w_1, w_2, \ldots w_{1890})$, can be calculated by Eq. (9). The value of ΔF_q for different q can be obtained by Eq. (13). Fig. 4 shows the variance of ΔF_q by varying the number of feature subsets using two following ranking methods: one by $|w_k|$ and the other by our proposed method. We observe that our method leads to a faster decrease in difference.

Given a pattern $\mathbf{x}$, the decision value of using the selected q features can be calculated by Eq. (11). When $q = 300$, 500, and 1000, we illustrate the results in Fig. 5. We observe that using the selected 500 or 1000 features can achieve almost the same performance as using all the 1890 features. However, using 300 features is insufficient to achieve a good performance.

4. Combination of input and feature spaces

In this section, we present the method of feature reduction by combining the input and feature spaces. First, we choose m features from the N input space as described in Section 2, and train the 2nd-degree polynomial SVM. Next, we select q features from $P = m(m+3)/2$ features in the feature space to calculate the decision value. We use two following methods to compute the decision values: one by Eq. (3) in input space and the other by Eq. (7) in feature space. In Eq. (3), the number of multiplications required to calculate the decision function is $(N+1)s$. Note that s is the total number of support vectors. In Eqs. (7) and (8), the total number of multiplications required is $(N+3)N$. If $(N+1)s > (N+3)N$, it is more efficient to implement the 2nd-degree polynomial SVM in the feature space; otherwise,

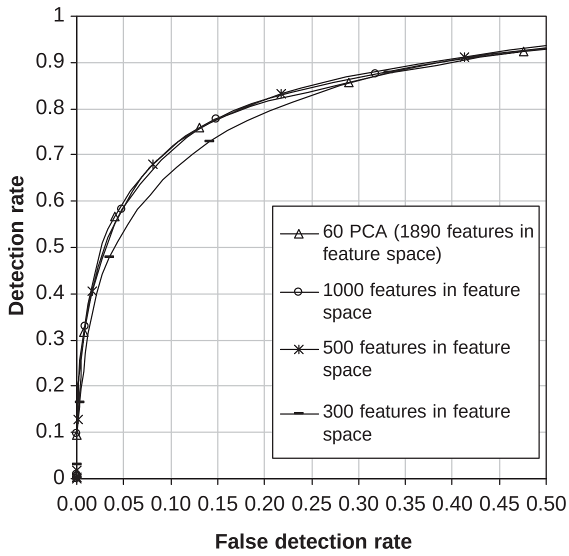

Fig. 5. ROC curves for different numbers of features in the feature space.

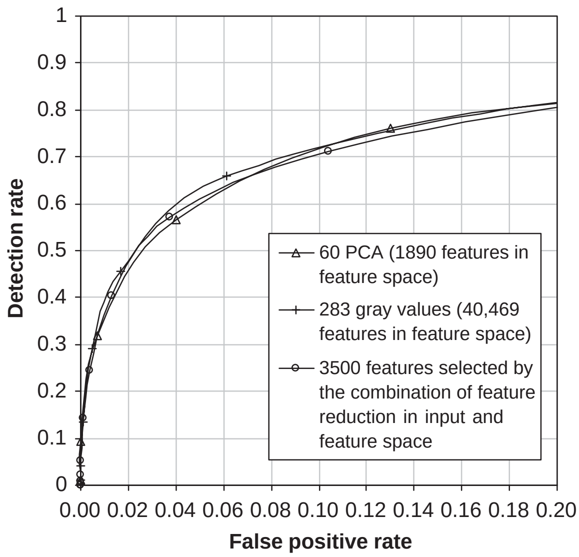

Fig. 6. ROC curves of using the proposed feature reduction method, 60 PCA, and all the 283 gray values.

in the input space. This is evidenced by our experiments because the number of support vectors is more than 700, that is much larger than N. Note that $N = 283$, 60, or 100 indicates all the gray-value features, 60 PCA values, or 100 features, respectively.

We train the SVM using the selected 100 features as described in Section 2.2 to obtain 244 and 469 support vectors for face and non-face classes, respectively. Fig. 6 shows the

Table 1
Comparisons of the number of features and the number of multiplications

Methods	Number of features in input space	Number of features in feature space	Number of multiplications
All gray values	283	40,469	80,938
PCA	60	1890	20,760
Our method	100	3500	8650

comparisons of our combinational method using 3500 features in feature space, 60 PCA values, and all the 283 gray values. We observe that using our combinational method can obtain competitive results as using 60 PCA values or all 283 gray values. Apparently, our method gains the advantage of speed.

Table 1 lists the number of features used in the input and feature spaces and the number of multiplications required in calculating the decision values for comparing our method with the methods of using all the gray values and using PCA features. From Table 1, we observe that using PCA for feature reduction in the input and feature spaces, a speed-up factor of 4.08 can be achieved. However, using our method, a speed-up factor of 9.36 can be achieved. Note that once the features in the feature space are determined, we do not need to project the input space on the whole feature space, but on the selected feature space. This can further reduce the computation; i.e., only 7000 multiplications are required instead of 8650.

5. Performance on person and car detection

In this section, we extend our experiments to person and car detection to further justify the validity of the proposed feature reduction method. Instead of selecting features from gray values in face detection, we use Haar wavelet features in person and car detection since the positive samples of person and car contain complex background, while the face samples contain only the face part.

5.1. Person detection

We adopt 924 person images of size 128×64 used in Ref. [7]. Some examples are shown in Fig. 7. We use 700 images for training and the remaining 224 images for testing. We also randomly extract 6000 non-person training samples and 3000 non-person testing images.

We use Haar wavelet representation in person detection since it can produce better results than using pixels or PCA [7]. The Haar feature is defined as the difference between two sums of pixel values in white rectangular and gray rectangular areas. We calculate vertical, horizontal and diagonal Haar features at scale of 32×32 and 16×16, respectively.

Fig. 7. Some examples of person images.

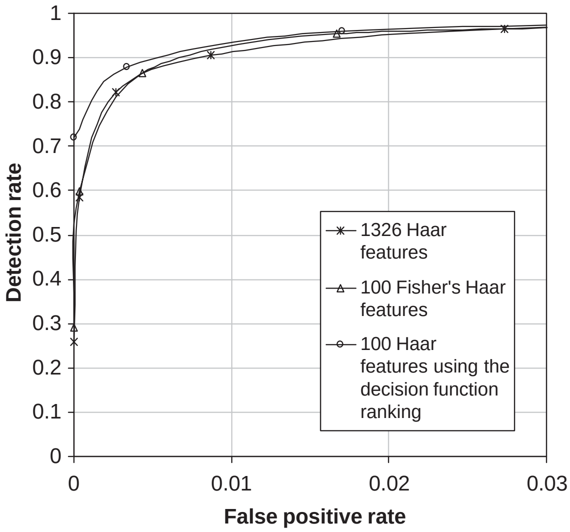

Fig. 8. ROC curves for person detection using the proposed feature reduction method in input space, Fisher's method and all the 1326 Haar features.

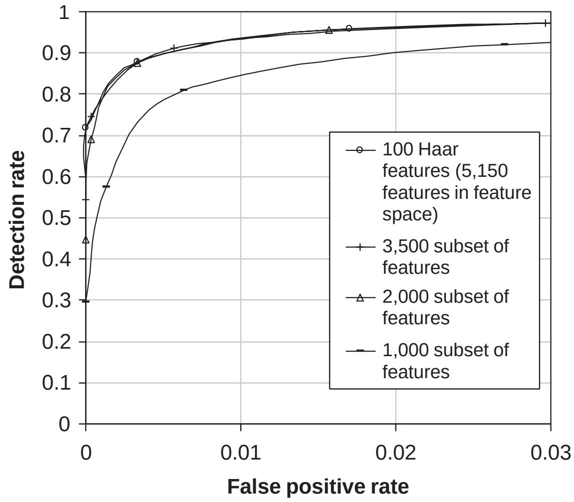

Fig. 9. ROC curves for person detection using the proposed feature reduction method in feature space.

Totally, we obtain 1326 Haar features for each person image. We normalize the Haar features in between 0 and 1.

Fig. 8 shows the results of using the three methods: all the 1326 Haar features, the subset of 100 features selected using our method, and the 100 features using Fisher's score. The x-axis indicates the false positive rate over 3000 non-person testing samples. The y-axis indicates the detection rate over 224 person images. We observe that when false positive rate > 0.004, using 100 Fisher's features can obtain slightly better results than using all the 1326 features, while using 100 features selected by the proposed method can always perform better than using all the 1326 features.

Fig. 9 shows that using the subset of 1000 features in feature space, the results are obviously worse than using all the 5150 features. Using the subset of 2000 features, at very low false detection rates (i.e., less than 0.002), the performance is worse than using all the 5150 features, while using the subset of 3500 features, we can obtain almost the same results as using all the 5150 features.

For person detection, after training a 2nd-degree SVM using all the 1326 Haar features, we obtain 486 support vectors, among them 136 for person and 350 for non-person classes. It is more efficient to implement the 2nd-degree SVM in input space since $(N+1)s < (N+3)N$, where $N = 1326$ and $s = 486$. Therefore, it takes about 644,922 multiplications to classify a pattern. If we use the proposed feature reduction method in input space to select 100 Haar features and in feature space to select 3500 features, it takes about 8650 multiplications. That is about 74.5 times faster.

5.2. Car detection

We adopt 516 car images of size 128×128 used in [8]. Some examples are shown in Fig. 10. For each car image, we also obtain its mirror image. Totally, we have 1032 car images. We use 700 car images as positive training samples and the remaining 332 images as testing samples. We also randomly extract 6000 non-car training samples and 3000 non-car testing images.

Fig. 10. Some examples of car images.

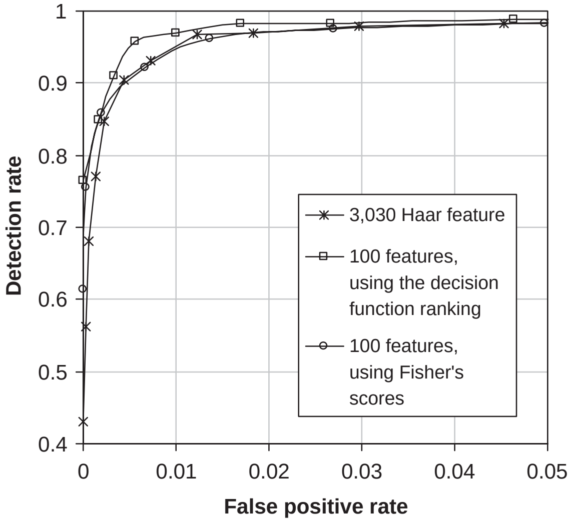

Fig. 11. ROC curves for car detection using the proposed feature reduction method in input space, Fisher's method, and all the 3030 Haar features.

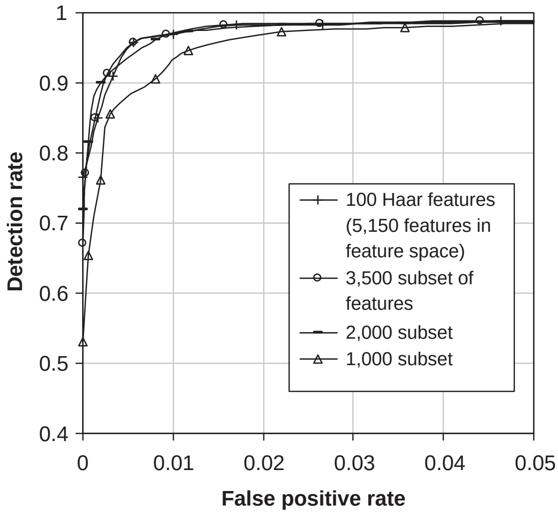

Fig. 12. ROC curves for car detection using the proposed feature reduction method in feature space.

We also use Haar wavelet features for car detection and obtain 3030 Haar features for each car image. We normalize the Haar features in between 0 and 1. Fig. 11 shows the results of using the three methods: all the 3030 Haar features, the subset of 100 features selected by our method, the 100 features by Fisher's score. We observe that using the subset of 100 Fisher's features can produce a little better result than using all the 3030 features. Using the subset of 100 features from the proposed method in input space can produce better results than both the Fisher's method and all the 3030 features.

Fig. 12 shows that using the subset of 1000 features in feature space, the result is obviously worse than using all the 5150 features, while using the subsets of 2000 and 3500 features, we can obtain similar results as using all the 5150 features.

For car detection, after training a 2nd-degree SVM using all the 3030 Haar features, we obtain 260 support vectors, among them 88 for car and 172 for non-car classes. It is more efficient to implement the 2nd-degree SVM in input space since $(N+1)s < (N+3)N$, where $N = 3030$ and $s = 260$. Therefore, it takes about 788,060 multiplications to classify a pattern. If we use the proposed feature reduction method in input space to select 100 Haar features and in feature space to select 3500 features, it takes about 8650 multiplications. That is about 90 times faster.

6. Discussions and conclusions

In this paper we intend to select a subset of features by ranking their contributions to the decision function of SVMs. The features that have large contributions indicate the importance of relevance in classification. This leads to a fast 2nd-degree SVM without a significant loss in performance. We have tested our method on the detection of face, person, and car. The experimental results show that the proposed feature reduction method works successfully. In fact, our method performs better in the detection of person and car. It is because they contain complex background and our method

Fig. 13. The 100 selected pixels are shown as white pixels.

only selects the majority of target features rather than the background features. Our method also gains the advantage of speed. For face detection, the subset is chosen directly from 283 gray values instead of the PCA features that require matrix computation. For person detection, we select 100 Haar features instead of all the 1032 Haar features. For car detection, we select 100 Haar features instead of all the 3030 Haar features. Furthermore, in the feature-space method, we choose a subset of 3500 features instead of all the 5150 features.

For our feature reduction method in input space, the number of features required to obtain good results is problem dependent. We have experimented to select 100 gray values from all the 283 gray values for face detection, and 100 Haar features from 1326 and 3030 Haar features for person and car detection, respectively. For face detection problem, the selected features are shown in Fig. 13 as white pixels. We can see that most of these features are located at eyes, nose, cheek, and the corner of mouth. These pixels are the most relevant ones that distinguish face from non-face images. For car and person detection, the features near the boundary capture the difference between the object and background, while the features on the object capture the object property. It is the reason why we obtain better results by using a subset of features than using all the Haar features. Six examples of the selected Haar features are shown in Fig. 14. For a domain problem, if all the features in input space are necessary to obtain good results, then the proposed feature reduction method in input space cannot be applied. On the other hand, if the dimensionality of input space is very large and we know that it is possible to obtain good results using only a subset of features, then we cannot apply this feature reduction method directly. In this case, we need to use other feature reduction methods to reduce the dimensionality first.

For our feature reduction method in feature space, we have experimented on selecting 1000 features from 1890 features corresponding to 60 PCA values in input space and on selecting 3500 features from 5150 features corresponding to 100 gray values in input space for face detection. For person and car detection, we have experimented on selecting 3500 features from 5150 features corresponding to 100 Haar features in input space. Experimental results show that our method works successfully. However, there is one limitation to apply our feature reduction method in feature space. The feature reduction method in feature space can only be applied if it is more efficient to implement the 2nd-degree SVM in feature space than in input space; i.e., the condition $(N+1)s > (N+3)N$ must be satisfied, where N is the dimensionality of input space and s is the total number of support vectors.

Acknowledgements

The authors would like to express our appreciation to anonymous reviewers for providing valuable comments to improve this paper.

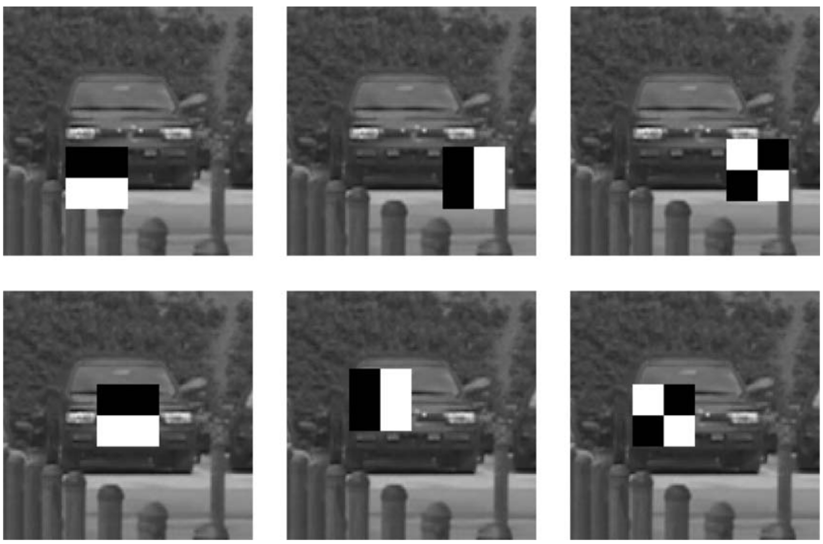

Fig. 14. Six examples of selected Haar features.

References

[1] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learning 20 (1995) 273–297.

[2] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

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

[4] B. Heisele, T. Poggio, M. Pontil, Face detection in still gray images, A. I. Memo 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA, 2000.

[5] B. Heisele, T. Serre, S. Prentice, T. Poggio, Hierarchical classification and feature reduction for fast face detection with support vector machines, Pattern Recognition 36 (9) (2003) 2007–2017.

[6] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, V. Vapnik, Feature selection for support vector machines, Advances in Neural Information Processing System, vol. 13, MIT Press, Cambridge, MA, 2000, pp. 668–674.

[7] T. Evgenious, M. Pontil, C. Papageorgiou, T. Poggio, Image representations and feature selection for multimedia database search, IEEE Trans. Knowledge and Data Engineering 15 (4) (2003) 911–920.

[8] C. Papageorgiou, T. Poggio, A trainable system for object detection, Int. J. Comput. Vision 38 (2000) 15–33.

About the Author—FRANK Y. SHIH received the B.S. degree from National Cheng-Kung University, Taiwan, in 1980, the M.S. degree from the State University of New York at Stony Brook, in 1984, and the Ph.D. degree from Purdue University, West Lafayette, Indiana, in 1987, all in Electrical Engineering. He is presently a professor jointly appointed in the Department of Computer Science (CS), the Department of Electrical and Computer Engineering (ECE), and the Department of Biomedical Engineering at New Jersey Institute of Technology, Newark, NJ. He currently serves as the Associate Chair of the CS Department and the Director of Computer Vision Laboratory. Dr. Shih is currently on the Editorial Board of the International Journal of Pattern Recognition, the International Journal of Pattern Recognition and Artificial Intelligence, the International Journal of Internet Protocol Technology, and the Journal of Internet Technology. He previously served as an Associate Editor of the International Journal of Systems Integration and the International Journal of Information Sciences. He has contributed as a steering member, committee member, and session chair for numerous technical conferences and workshops. He was the recipient of the Research Initiation Award from the National Science Foundation in 1991. He won the Honorable Mention Award from the International Pattern Recognition Society for Outstanding Paper and also won the Best Paper Award in the International Symposium on Multimedia Information Processing. He has received several awards for distinguished research at New Jersey Institute of Technology. He has served several times in the Proposal Review Panel of the National Science Foundation on Computer Vision and Machine Intelligence. Dr. Shih holds the IEEE senior membership. He has published over 130 technical papers in well-known prestigious journals and conferences. His current research interests include image processing, computer vision, computer graphics, artificial intelligence, bioinformatics, information security, robotics, fuzzy logic, and neural networks.

About the Author—SHOUXIAN CHENG received the B.S. degree in Material Science from Beijing University of Chemical Technology, Beijing, China, in 1995, and the M.S. degree in Computer Science from New Jersey Institute of Technology in 2001. Mr. Cheng is currently a Ph.D. candidate in Computer Science, New Jersey Institute of Technology. His current research interests include image processing, pattern recognition and machine learning.