PAPER

Composite Support Vector Machines with Extended Discriminative Features for Accurate Face Detection

Tae-Kyun KIM[†*a)] *and* Josef KITTLER[††], *Nonmembers*

SUMMARY This paper describes a pattern classifier for detecting frontal-view faces via learning a decision boundary. The proposed classifier consists of two major parts for improving classification accuracy: the implicit modeling of both the face and the near-face classes resulting in an extended discriminative feature set, and the subsequent composite Support Vector Machines (SVMs) for speeding up the classification. For the extended discriminative feature set, Principal Component Analysis (PCA) or Independent Component Analysis (ICA) is performed for the face and near-face classes separately. The projections and distances to the two different subspaces are complementary, which significantly enhances classification accuracy of SVM. Multiple nonlinear SVMs are trained for the local facial feature spaces considering the general multi-modal characteristic of the face space. Each component SVM has a simpler boundary than that of a single SVM for the whole face space. The most appropriate component SVM is selected by a gating mechanism based on clustering. The classification by utilizing one of the multiple SVMs guarantees good generalization performance and speeds up face detection. The proposed classifier is finally implemented to work in real-time by cascading a boosting based face detector.

key words: *pattern classification, face detection, support vector machine, independent component analysis, principal component analysis, Adaboost*

1. Introduction

In the last ten years, face detection has been widely studied for applications such as video surveillance systems and human-computer interaction, providing a large improvement in both classification accuracy (it is referred to as detection rate to a given number of false alarms in terms of detection) and speed [1]–[8]. Although face detection systems have obtained real-time speed, they still have a problem in accuracy. They often miss faces and make false alarms in complex and changing surveillance environments. Neural Network (N.N), Adaboost [6], [10] and Support Vector Machine (SVM) are representative methods to make an accurate face detector. They are all face template-based so that they do not need to localize the facial features such as eyes, nose, mouth and etc. This gives the advantage of being able to detect even low-resolution faces, which are often captured in surveillance environments.

Sung and Poggio [9] clustered the entire set of face and non-face samples into several clusters and constructed a subspace for each cluster. The distances to these subspaces were utilized as an input vector to N.N. Although the method of feature extraction was effective, N.N. generally has a high degree of freedom in its structure and this causes an overfitting to given training data. The boosting based face detectors [6], [10], [11] exploit the simple features to be computed in an integral image. The boosted classifiers using the simple features yield very rapid detection even in gray static images. They have also showed comparable accuracy to the N.N. based methods [5], [9], [12], [13], but they require exhaustive training, which takes the order of weeks and many steps of bootstrapping to collect effective training samples in [6], [10]. SVM [3], [14], [16], [19], one of the existing most accurate face detectors, is also efficient in training time and avoids overfitting by minimizing both structural and empirical risk [17], [18]. These advantages of SVM facilitate further accuracy improvement.

SVMs have generated a lot of interests in relation to face pattern classification, because it is easy to find an optimal decision boundary for given data compared with other methods. Osuna [3] successfully applied SVMs to face detection and Qi [19] enhanced their performance by using Independent Component Analysis (ICA) features. Qi obtained a larger margin of separation with fewer support vectors by using ICA compared to training SVMs directly in the image space. ICA offers the benefit of exploiting higher-order statistics of face images, whereas Principal Component Analysis (PCA), which has been still widely adopted as a representative face subspace method, assumes a Gaussian data distribution. Up to our knowledge, there is no comparative study of ICA and PCA in terms of face detection accuracy. In SVM based face detection non-linear kernels of SVMs are usually adopted for capturing nonlinear manifolds of face data and they operate by comparing an input to the full set of support vectors (SVs). As this is time consuming, some methods of speeding up non-linear SVMs are required.

In this paper, we propose an accurate face/non-face classifier which combines the implicit modeling of the face and the near-face classes and a computationally efficient SVM classifier. We propose two main ideas to enhance classification accuracy. First, density distributions of both the face and the near-face classes are estimated separately by subspace analysis. The two sets of projection coefficients obtained from the face and near-face subspaces are complementary yielding a more reasonable SVM boundary. The

Manuscript received July 30, 2004.

Manuscript revised March 17, 2005.

[†]The author is with Samsung Advanced Institute of Technology, Korea.

[††]The author is with Centre for Vision, Speech and Signal Processing, University of Surrey, U.K.

[*]Presently, with Machine Intelligence Laboratory, University of Cambridge, U.K.

a) E-mail: tkk22@cam.ac.uk

DOI: 10.1093/ietisy/e88–d.10.2373

distances to the two sub-spaces defined by residual errors are also utilized to augment the feature vector for SVM, which is called *extended discriminative features*. They indicate the quality of representation of the input image in the face or near-face subspaces. The use of the two residual errors helps the separation into face and non-face classes. Second, the classifier consists of component SVMs trained in local feature spaces. A face space is divided into several local spaces by *k*-means clustering and the corresponding non-face examples to each face cluster are separately collected. Then multiple nonlinear SVMs are trained for each training class. This composite architecture of SVMs, called *composite SVMs*, would be more effective in both computational cost and accuracy if face patterns were multimodally distributed. Each nonlinear component SVM has a simpler boundary than that of a single SVM for the whole face space. The classification by utilizing one of the SVMs enhances generalization performance and speeds up face detection. This work is related to the method of Sung and Poggio [9] in the aspect of clustering but the major difference is that they consider all the clusters for classification by using a fully-connected N.N. whereas in our proposal only one cluster is selected for efficient classification. Besides, there is a significant difference in the subspace method for feature extraction: The proposed method additionally exploited projection coefficients to the distance features, which are only input in their method.

The paper is organized as follows. In Sect. 2, the extended discriminative feature set is proposed by implicitly modeling the face and the near-face classes. Both representative subspace methods, PCA and ICA are considered and compared in the modeling. In Sect. 3, the composite non-linear SVMs in local feature spaces are developed for face detection. The experimental results and the real-time implementation of the proposed classifier by combining the Adaboost face detector [6], [10] are explained in Sect. 4 and the conclusion is drawn in Sect. 5.

2. Extended Discriminative Feature Set

The proposed face pattern classifier is based on learning a decision boundary using the SVM approach. The feature space is defined by ICA (or PCA) which learns statistically independent basis vectors for the near-face space as well as the face space. In high-level pattern classification, as the image data projected into this feature space is in general multimodally distributed, multiple non-linear SVMs are learned in local feature spaces. This results in a bank of SVM classifiers which are gated by a nearest mean classifier based on *k*-means clustering. The method to extract the extended discriminative feature set will be explained based on ICA. For the PCA based features, the same procedure is executed by replacing the ICA with PCA.

The exact estimation of probability densities of face and non-face classes becomes difficult due to the high dimensionality, maybe non-Gaussianity, and multimodality of the images. So here, to obtain a more reliable decision boundary, face and non-face classes are separately and implicitly modeled by unsupervised learning. Because the range of the non-face class is extremely broad, only the near-face images that lie close to the face space and are therefore easily confused with face images are considered. The near-face class consists of the non-face images which were initially collected based on their distance to an average vector of training faces and enlarged by a bootstrap technique. The extended features obtained by the projection to the near-face sub-space play a complementary role to the features of the face sub-space, providing better classification results. The projection error for each of face and near-face sub-space is simply defined by the Euclidean distance measure assuming an isotropic Gaussian data distribution. The basis vectors trained from the face class reconstruct a face image faithfully and the residual error of the face image is generally smaller than that of non-face images. The opposite is true for the near face subspace.

ICA obtained by using the extended Infomax algorithm [20] is adopted for density estimation. Let $\mathbf{X}$ be an input matrix whose rows are training face images. We seek matrix $\mathbf{U}$ such that

$$\mathbf{U} = \mathbf{W}\mathbf{X}, \tag{1}$$

where $\mathbf{W}$ is a weight matrix and $\mathbf{U}$ is an output matrix of ICA. The rows of output,$\mathbf{U}$, are also images. The update rule of the weight matrix obtained using the extended Infomax algorithm is given by

$$\Delta\mathbf{W} \propto [\mathbf{I} - \mathbf{K}\,tanh(\mathbf{U})\mathbf{U}^{\mathbf{T}} - \mathbf{U}\mathbf{U}^{\mathbf{T}}]\mathbf{W}, \tag{2}$$

where $\mathbf{K}$ is a diagonal matrix whose elements depend on whether the sources are super- or sub-Gaussian distributions. $\mathbf{K}$ is determined by the following switching criterion,

$$\mathbf{K} = diag(sign(\mathbf{E}(sech(\mathbf{U}^{\mathbf{T}}).^{2}). \times \mathbf{E}(\mathbf{U}^{\mathbf{T}}.^{2}) \\ -\mathbf{E}(tanh(\mathbf{U}^{\mathbf{T}}). \times \mathbf{U}^{\mathbf{T}}))) \tag{3}$$

where $.^2$, $.\times$ and $\mathbf{E}$ are the array power operator, array multiply operator and a row vector containing the mean value of each column respectively. In practice, ICA is performed on the main eigenvectors to solve the convergence problem that ICA exhibits for the training set in the original image space and to control the number of independent basis vectors [19], [21], [22].

Let $\mathbf{P_m}$ denote the matrix containing the first m eigenvectors in its columns. The PCA representation of zero-mean face images $\mathbf{X}$ is defined as $\mathbf{R_m} = \mathbf{X}\mathbf{P_m}$ and the reconstruction of $\mathbf{X}$ is obtained by $\mathbf{X_{rec}} = \mathbf{R_m}\mathbf{P_m^T}$. ICA is performed on $\mathbf{P_m}$.

$$\mathbf{W}\mathbf{P_m^T} = \mathbf{U} \quad \Rightarrow \quad \mathbf{P_m^T} = \mathbf{W}^{-1}\mathbf{U}. \tag{4}$$

Therefore the ICA reconstruction is obtained by

$$\mathbf{X_{rec}} = \mathbf{R_m}\mathbf{P_m^T} \quad \Rightarrow \quad \mathbf{X_{rec}} = (\mathbf{X}\mathbf{P_m})(\mathbf{W}^{-1}\mathbf{U}). \tag{5}$$

Finally the ICA representation of $\mathbf{X}$ is given by

$$\mathbf{B_m^{face}} = \mathbf{X}\mathbf{P_m}\mathbf{W}^{-1}. \tag{6}$$

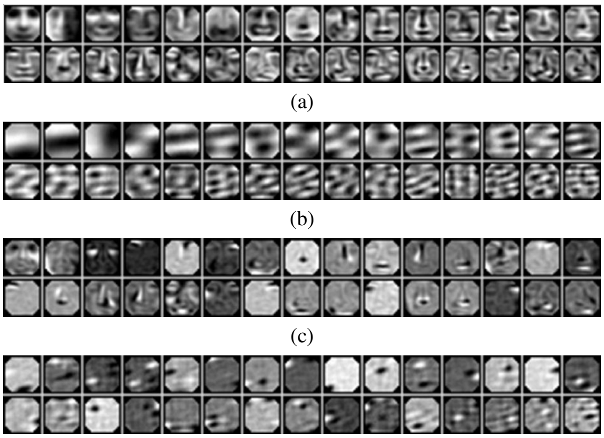


Fig. 1 Basis images: (a) First 30 PCA basis images from face class. (b) First 30 PCA basis images from near-face class. (c) Randomly selected 30 ICA basis images from face class. (d) Randomly selected 30 ICA basis images from near-face class.

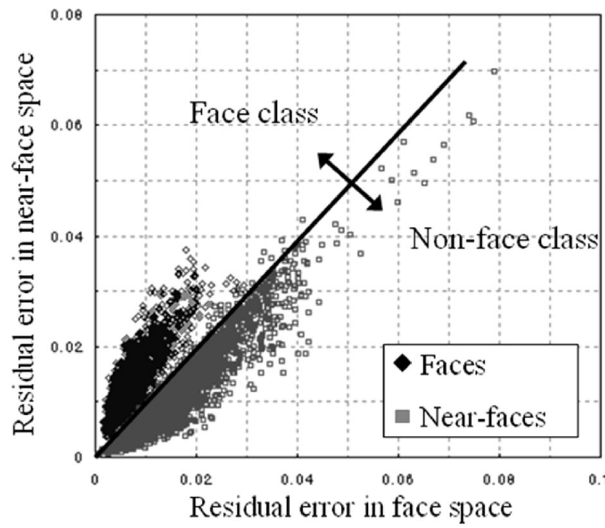


Fig. 2 Residual errors of trained face and near-face patterns.

The same procedure is repeated on the near-face images for the near-face features, $\mathbf{B}_{\mathbf{m}}^{\mathbf{near-face}}$. Figure 1 shows the learned basis images of PCA and ICA. The basis set from the near-face class consists of various filters of horizontal edges. It is different from the bases of randomly selected non-face images, which shows a set of arbitrary directional edge images. While PCA basis set appear holistic, ICA basis sets are localized.

A residual error for the sub-space is exploited as an additional feature for SVM. The Euclidean distance measure describes the error under the assumption of isotropic Gaussian distribution by

$$\begin{aligned} \varepsilon &= (\mathbf{x} - \mathbf{x_{rec}})(\mathbf{x} - \mathbf{x_{rec}})^{\mathbf{T}} \\ &= (\mathbf{x}(\mathbf{I} - \mathbf{P_m}\mathbf{W}^{-1}\mathbf{U}))(\mathbf{x}(\mathbf{I} - \mathbf{P_m}\mathbf{W}^{-1}\mathbf{U}))^{\mathbf{T}}. \end{aligned} \tag{7}$$

The two features defined by the residual errors for the face and the near-face sub-spaces are shown to be very effective in discriminating between the trained face and near-face patterns in Fig. 2. While only one error can not discriminate two classes well, the relationship of the error of the face class to that of the near-face class, which can be considered as an angle of a line in the figure, dichotomizes the training classes correctly. These features might play dominant roles in learning a decision boundary in conjunction with the other features obtained by the projections to sub-spaces of the face and near-face classes. The extended feature set enhances the separability of the two classes. Finally the proposed description for a given image is presented by

$$\mathbf{y} = \{\mathbf{b}_{\mathbf{m}}^{\mathbf{face}}, \mathbf{b}_{\mathbf{m}}^{\mathbf{near-face}}, \varepsilon^{\mathbf{face}}, \varepsilon^{\mathbf{near-face}}\}, \tag{8}$$

where $\mathbf{b}$ is a row vector of the representation matrix $\mathbf{B}$ in (6). The experimental section gives the comparison of PCA and ICA for the extended discriminative feature set for face detection.

3. Composite SVMs in Local Feature Spaces

SVMs implicitly map the data (in our case, the data is the extended feature vector described in (8)) into a dot product space via a nonlinear mapping function. Then the SVMs learn a hyperplane that separates the data by a large margin. A test pattern, $\mathbf{x_{test}}$, is classified as a face or non face by using the trained SVMs.

$$f(\mathbf{x_{test}}) = sign\left(\sum_{i=1}^{l} c_i \lambda_i \mathbf{K}(\mathbf{x_{test}}, \mathbf{x}_i) + b\right), \tag{9}$$

where c_i is the class label for the i-th training feature vector $\mathbf{x}_i$, λ_i and b are constants which are determined by learning, $\mathbf{K}$ is a kernel function and l is the number of support vectors (SVs).

In classification using non-linear SVMs, run-time complexity is proportional to the number of SVs. It is time-consuming to apply a full SVM classifier to every pixel of an image for face localization. The composite SVMs simplify the decision boundaries and reduce the number of SVs, thus speeding up the algorithm. Suppose that the overall distribution of the training data is multi-modal. In this case, a more complicated decision boundary is required as shown in the left image of Fig. 3. It has a large number of SVs resulting in high computation costs. Moreover it is difficult to learn such a complex boundary in a high dimensional space yielding poor performance for a new data set. If the SVMs are trained in each local sub-space, a decision boundary becomes much simpler as shown in the upper and lower right images of Fig. 3. The classification into face and near-face classes might be multi-modally distributed and thus the composite scheme of SVMs trained in local sub-spaces improves the speed of processing and generalization performance.

For the division of a space into local subspaces, k-means clustering algorithm was utilized. After the cluster-

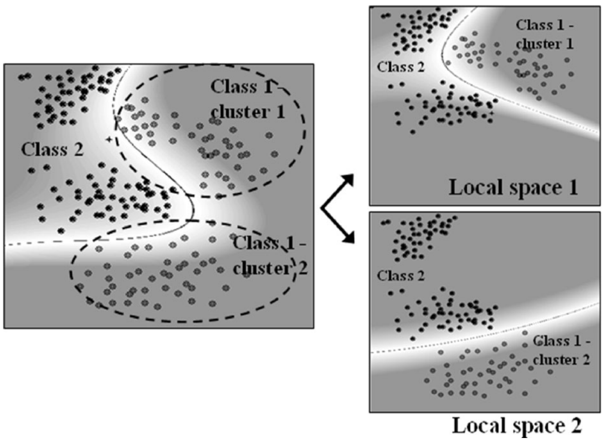


Fig. 3 Learning a decision boundary in local sub-space.

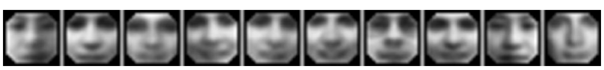


Fig. 4 Centroid images of face clusters.

ing of the training face images, the centroids of the k face clusters $\mathbf{u}^i, i = 1, \ldots, k$ are stored. The near-face samples utilized for the training are assigned to the seperated face subspaces according to Euclidean distances to the face centroids as $E(\mathbf{x}, \mathbf{u}^i) = |\mathbf{x} - \mathbf{u}^i|^2$. The near-face samples were assigned to the nearest face clusters. In this way both the face and the near-face training patterns are divided into k subsets. k nonlinear SVMs are then trained on each set of the face and near-face samples. As shown in Fig. 4, the centroid images of the face class include intrinsic changes for identity and expression, and extrinsic changes for illumination and viewing geometry of faces.

For the classification of a new pattern $\mathbf{x_{test}}$, the nearest SVM to the test pattern is selected and applied. Suppose that k SVMs are trained for each sub-set of face and near-face samples as,

$$[\{c_i^1, \lambda_i^1, \mathbf{x}_i^1, \mathbf{K}^1, b^1, \mathbf{u}^1 | i = 1, \ldots, l_1\}, \ldots, \\ \{c_i^k, \lambda_i^k, \mathbf{x}_i^k, \mathbf{K}^k, b^k, \mathbf{u}^k | i = 1, \ldots, l_k\}] \tag{10}$$

The nearest SVM to $\mathbf{x_{test}}$ is found by

$$i_s = \min_{arg\ i} E(\mathbf{x_{test}}, \mathbf{u}^i) \tag{11}$$

and the decision is made based on

$$f(\mathbf{x_{test}}) = sign\left(\sum_{j=1}^{l_{i_s}} c_j^{i_s} \lambda_j^{i_s} \mathbf{K}^{i_s}(\mathbf{x_{test}}, \mathbf{x}_j^{i_s}) + b_j^{i_s}\right). \tag{12}$$

4. Experimental Results

4.1 Experimental Setup

To show a viability of the proposed algorithm, the various face/non-face pattern classifiers, which have different features or different SVM architectures, were trained under the same condition and evaluated on two sets of gray images. Set A contained 400 high-quality images with one face per image from the Olivetti laboratory image database [23]. Set B contained the randomly selected 36 images of mixed quality of 172 faces from the Rowley [12] test set. Set A involved 1684800 pattern windows, while Set B involved 6178110. For comparison of classifiers in terms of detection performance, the classifiers are applied to image windows at all possible positions and scales and faces are detected with a common simple refinement process [12], [13].

In the training, various face, near-face and random non-face images were collected and normalized to 20×20 pixel resolution. The training face images include artificially rotated, shifted and resized faces. The near-face patterns were identified among non-face patterns as those exhibiting a small vector distance to the average vector of training faces and enlarged by a bootstrap technique. 2,000 faces, 3,000 near-face patterns and 10,000 random non-face patterns were finally collected. It is noted that the training set as well as the test set is common for methodological comparison of the proposed method and the other conventional approaches. Although the training and test set are not as large as those in [4], [6], [12], [13], they are enough to show the comparative benefits of the different methods.

4.2 Evaluation of the Composite SVMs with the Extended Discriminative Features

4.2.1 Comparison of Feature Extraction Methods

SVMs with different feature sets are compared in terms of face detection accuracy. For the extended discriminative feature set, we selected 50 significant eigenvectors to represent the each of the face and near-face training set. This number was chosen to enough represent the training image space in the sense of reconstruction errors. Thus, ICA yielded 50 independent components. Consequently the extended discriminative feature set consists of 102 features: 50 projection coefficients in the face subspace, 50 projection coefficients in the near-face subspace and 2 residual errors. A single non-linear SVM was commonly trained to discriminate between the face training images and the near/non-face training images. The SVM has a polynomial kernel of degree 2. The trade-off constant for the face class was typically set to 2 and the constant for both the near-face and the random non-face class was set to 1.

Table 1 shows the detection results of the SVM classifiers combined with different feature sets: 50 projection features in the PCA/ICA face subspace, which have been applied to SVM in the previous face detection method [19], and the extended PCA/ICA feature sets. The proposed extended features significantly increase the detection performance for both PCA and ICA representation. In Fig. 5, the performances of SVMs with the ICA feature sets such as just 50 face features, 100 face/near-face features, and 51 face/residual features, show that the near-face features and the residual features independently contributed to the ac-

Table 1 Detection results of the combined SVM with the different feature sets.

	Set A Detect rate/False alarms	**Set B** Detect rate/False alarms
PCA-SVM	N/A	85.4%/136
ICA-SVM	97.2%/3	84.3%/127
Extended PCA-SVM	N/A	90.1%/59
Extended ICA-SVM	98.5%/4	90.1%/62

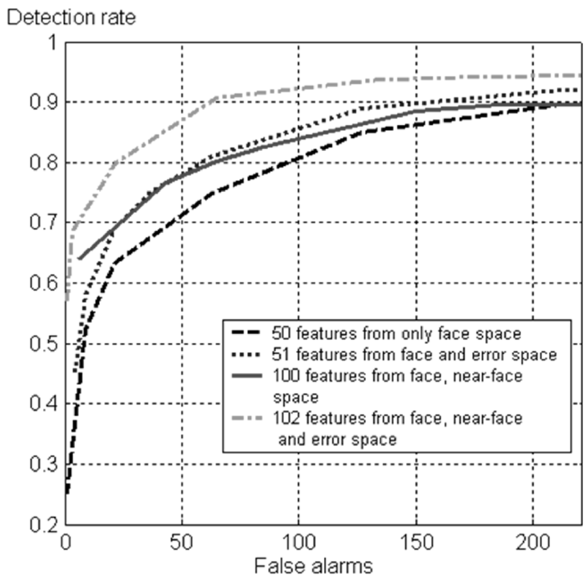


Fig. 5 ROC curves on set B.

Table 2 Training and test results of the different SVM architectures on set B.

	Training Margin/# of SVs	**Testing** Detect rate(in %)/False alarms
Single Nonlinear SVM	0.215/439	93.5%/120
Composite Nonlinear SVMs	0.594/119.6	93.5%/135
Composite Linear SVMs	0.221/157	93.5%/1989

curacy improvement. While the extended feature set gave a significant improvement over the conventional projection coefficient features in face subspace, ICA yielded similar performance with PCA both in the extended and conventional feature schemes.

4.2.2 Comparison of SVM Architectures

Table 2 shows the learning and test results of the different SVM structures with the extended discriminative feature set on Set B. For the proposed composite scheme, a weighted average of margins and the numbers of SVs were calculated by

$$h = \frac{1}{10} \sum_{k=1}^{10} \frac{m_k}{m} \cdot F(k) \tag{13}$$

with the weights determined by the relative cluster sizes, i.e. m_k is the number of training patters assigned to the k-th cluster and m is the cardinality of the training set. $F(k)$ is the margin or number of SVs of the k-th SVM. In this experiment, the number of clusters was arbitrarily set to 10. It is noted that the proposed composite scheme of SVMs provides a larger margin and a smaller number of SVs making the algorithm to be approximately 4 times faster and have a similar face detection performance compared to the single SVM trained from the entire data set. The composite structure is also expected to work well for a new data set with the benefit of the larger margin. Linear SVMs were not powerful enough to learn reliable decision boundaries for the 10 face and near-face sub-spaces. Intuitively one might expect that the local SVMs would also enhance the detection performance. However this hypothesis proved to be untrue on this test set due to the fact that local decision boundaries lose some part of the global boundary and this can have a detrimental effect on test patterns. The detection accuracy can be further enhanced by using overlapping training sets for each local SVM and selecting the optimal number of local classes on the given data, which would be an interesting extension to this work.

4.3 On Real-Time Implementation

The proposed face/non-face classifier is combined with a pre-processing stage for real-time detection performance. The role of the pre-stage is to largely reduce the search range of the proposed classifier which has a relatively heavy computation but elaborate decision. The face detectors based on Adaboost algorithm [6], [10] can be a good preprocessor in this study. The speed of the Adaboost detector is about 66 ms on 320×240 images for a Pentium IV 1 GHz PC, which enables real-time performance of the proposed detector.

To briefly explain the combination of the two face detectors, the Adaboost face detector with a threshold adjustment is applied to eliminate a huge number of candidate windows for the proposed detector. The windows that pass this test are then submitted to the proposed classifier to perform the final classification of the content into face and non-face categories. The Adaboost detector consists of multiple cascade layer classifiers itself and each layer classifier has a certain threshold to decide whether the pattern coming from the previous layers is a face or not. The threshold of the last layer classifier of the Adaboost detector was adjusted to control the number of candidate windows about the proposed classifier. The Adaboost detector was trained on the same data set which was exploited to learn the proposed classifier. The training set was depicted in Sect. 4.1. Figure 6 shows the detection performance curves of the Adaboost detector and the combined algorithm. The detection rate of the pure Adaboost was similar to the SVM classifier with just 50 face features of PCA or ICA. The combined algorithm achieved a higher detection rate owing to higher accuracy of the proposed classifier. The speed of the combined al-

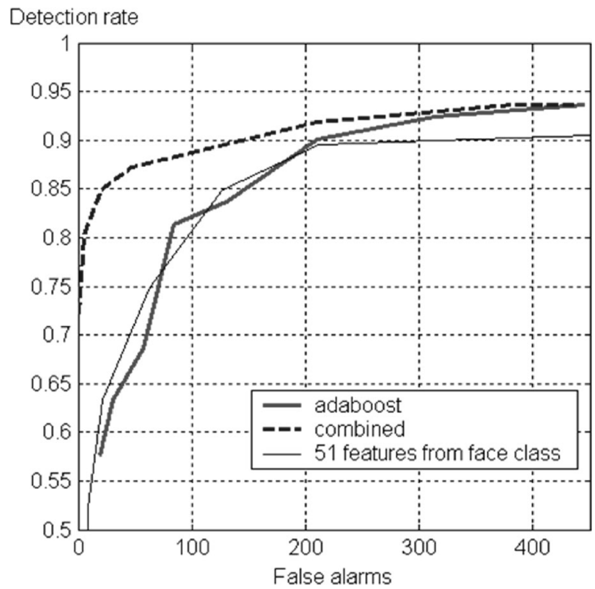


Fig. 6 ROC curves of the combined algorithm of the proposed classifier and the Adaboost detector for the set B.

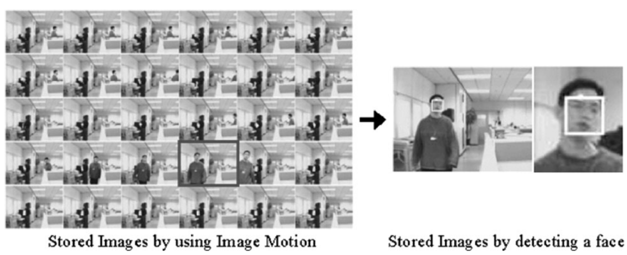


Fig. 7 Efficient image storing by detecting faces in the video surveillance system.

gorithm is similar to that of the Adaboost detector because the Adaboost as a preprocessor throws out most non-faces and the proposed SVM based classifier is much lightened by using the composite scheme.

5. Conclusion

We have presented an efficient method of face detection for surveillance applications. The developed face detection technology can help the conventional video recording systems to relieve their excessive storing by selecting key frames which involve a face as shown in Fig. 7. The face, near-face and error spaces are jointly considered for the extraction of the extended feature set using ICA or PCA. This extended feature vector significantly increased the detection accuracy. To speed up the algorithm, multiple SVMs were trained in local subspaces defined by k-means clustering. This composite SVM structure makes the elaborative classifier approximately 4 times faster. The problem of choosing the optimal number of local sub-spaces still remains open issues for future work. We also have implemented the classifier for real-time detection of faces. The suggestion involves using the new algorithm in a chain, with the Adaboost face detector performing the initial filtering of candidate face images. The proposed face/non-face classifier, which is slower but more accurate, performs the final classification of the content of the candidate windows into face and non-face categories. In this way the speed of the cascade face detector is not compromised.

Acknowledgments

This research was supported by the national project of HCI Lab in Samsung Advanced Institute of Technology, Korea. The authors would like to thank Changkyu Choi, J. Matas and Vojta Franc for their valuable discussion and comments.

References

[1] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated person tracking using stereo, color and pattern detection," Proc. Computer Vision and Pattern Recognition, pp.601–609, Santa Barbara, 1998.

[2] R.-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," Proc. International Conf. Image Processing, vol.1, pp.1046–1049, Greece, 2001.

[3] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," Proc. Computer Vision and Pattern Recognition, pp.130–136, San Juan, 1997.

[4] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," Proc. Computer Vision and Pattern Recognition, vol.1, pp.746–751, 2000.

[5] H. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," Proc. Computer Vision and Pattern Recognition, p.963, Santa Barbara, 1998.

[6] P. Viola and M. Jones, "Robust real-time face detection," Proc. International Conf. Computer Vision, vol.2, p.747, Vancouver, 2001.

[7] T.-K. Kim, D. Kong, and S.-R. Kim, "Learning a decision boundary for face detection," International Conference on Image Processing, pp.920–923, Rochester, NY, Sept. 2002.

[8] T.-K. Kim, S.-U. Lee, J.-H. Lee, S.-C. Kee, and S.-R. Kim, "Integrated approach of multiple face detection for video surveillance," Proc. International Conf. Pattern Recognition, p.20394, Quebec, 2002.

[9] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detectio," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.1, pp.39–51, 1998.

[10] P. Viola and M.Jones,"Robust real-time object detection," Proc. Second International Workshop on Statistical and Computational Theories of Vision, Vancouver, 2001.

[11] R. Lienhart and J. Maydt,"An extended set of Haar-like features for rapid object detection," Proc. International Conf. Image Processing, pp.900–903, Rochester, NY, 2002.

[12] H.A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.1, pp.23–38, 1998.

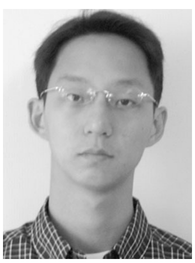
[13] H.A. Rowley, Neural Network-Based Face Detection, Ph.D. Thesis, Computer Science Department, Carnegie Mellon University, CMU-CS-99-117, May 1999.

[14] S. Romdhani, P. Torr, B. Scholkopf, and A. Blake, "Computationally efficient face detection," Proc. International Conf. Computer Vision, vol.2, pp.695–700, Vancouver, 2001.

[15] J. Terrillon, M.N. Shirazi, M. Sadek, H. Fukamachi, and S. Akamatsu, "Invariant face detection with support vector machines," Proc. International Conf. on Pattern Recognition, vol. 4, pp.210–217, Barcelona, 2000.

[16] B. Heisele, T. Serre, S. Mukherjee, and T. Poggio, "Feature reduc-

tion and hierarchy of classifiers for fast object detection in video images," Proc. Computer Vision and Pattern Recognition, vol.2, pp.18–24, 2001.
[17] V. Vapnik, The nature of statistical learning theory, Springer-Verlag, New York, 1995.
[18] S. Gong, S.J. McKenna, and A. Psarrou, Dynamic Vision: From Images to Face Recognition, Imperial College Press, 2000.
[19] Y. Qi, D. Doermann, and D. DeMenthon,"Hybrid independent component analysis and support vector machine learning scheme for face detection," Proc. ICASSP, vol.3, pp.1481–1484, Utah, 2001.
[20] T.W. Lee, M. Girolami, and T.J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," Neural Comput., vol.11, no.2, pp.417–441, 1999.
[21] M.S. Bartlett and T.J. Sejnowski, "Independent component representations for face recognition," Prof. SPIE Conf. on Human Vision and Electronic Imaging III, vol.3299, pp.528–539, San Jose, 1998.
[22] T.-K. Kim, H. Kim, W. Hwang, S.-C. Kee, and J. Kittler, "Independent component analysis in a facial local residue space," Proc. Computer Vision and Pattern Recognition, p.579, Madison, 2003.
[23] F. Samaria and A. Harter,"Parameterization of a stochastic model for human face identification," Proc. 2nd IEEE Workshop Application Computer Vision, Sarasota, FL, 1994.



Tae-Kyun Kim is a PhD student of Machine Intelligence Laboratory at the University of Cambridge. He received the BSc and a MSc degrees from the Dept. of EECS at the Korea Advanced Institute of Science and Technology (KAIST) in 1998 and 2000, respectively. He worked as a research staff member at Samsung Advanced Institute of Technology, Korea during 2000 - 2004. His research interests include computer vision, statistical pattern classification and machine learning. He reviews for the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and was the Korea delegate of MPEG-7. The joint proposal of face descriptor of Samsung and NEC, for which he developed the main algorithms, was accepted as the international standard of ISO/IEC JTC1/SC29/WG11.

Josef Kittler is Professor of Machine Intelligence, and Director of the Centre for Vision, Speech and Signal Processing at the University of Surrey. He has worked on various theoretical aspects of Pattern Recognition and Image Analysis, and on many applications including personal identity authentication, automatic inspection, target detection, detection of microcalcifications in digital mammograms, video coding and retrieval, remote sensing, robot vision, speech recognition, and document processing. He has co-authored a book with the title "Pattern Recognition: A statistical approach" published by Prentice-Hall and published more than 500 papers. He is a member of the Editorial Boards of Image and Vision Computing, Pattern Recognition Letters, Pattern Recognition and Artificial Intelligence, Pattern Analysis and Applications, and Machine Vision and Applications.