LEARNING A DECISION BOUNDARY FOR FACE DETECTION

Tae-Kyun Kim, Donggeon Kong and Sang-Ryong Kim

Human Computer Interaction Lab., Samsung AIT, Yongin, Korea

ABSTRACT

This paper describes a pattern classification approach for detecting frontal-view faces via learning a decision boundary. The classification can be achieved either by explicit estimation of density functions of two classes, face and non-face or by direct learning of a classification function (decision boundary). The latter is a more effective approach, when the number of training available examples is small, compared to the dimensionality of image space. The proposed method consists of a implicit modeling of both face and near-face classes using Independent Component Analysis (ICA), and a subsequent classification stage based on the decision boundary estimation using Support Vector Machine (SVM). Multiple nonlinear SVMs are trained for local subspaces, considering the general non-Gaussian and multi-modal characteristic of face space. This parallelization of SVMs reduces computational cost of on-line classification, since the locally trained SVM has small number of support vectors compared to the SVM trained on entire data space. We showed that the proposed algorithm is superior to the simple combination of ICA and SVM, both in accuracy and computational burden.

1. INTRODUCTION

Recently face detection has been widely studied for interesting applications such as surveillance system and human-computer interaction. Among various approaches to face detection, especially learning a decision boundary directly from both face and non-face samples has been principally adopted and has achieved many successful results. This is a simpler approach in that it does not require a difficult problem of estimation of two densities. Vapnik [2] and Gong [3] indicated that estimating density functions of faces and non-faces is under-constrained when training data are sparse compared to the dimensionality of the image space.

Mainly two approaches, multi-layer perceptrons (MLPs) and support vector machines (SVMs), are used to find a decision boundary by learning from examples. Many studies [4,5] have successfully applied MLPs to face detection problem. In [4] Rowley et al. designed a MLP-based face detector using a local receptive field-like connectivity. 12 vectors of normalised Mahalanobis distances in local Principal Component Analysis (PCA) sub-spaces were used as inputs to MLPs in [5]. However, MLPs are rather domain-specific and thus they result in poor generalization. It is also difficult to learn iteratively due to high complexity of learning. Recently SVM, which is trained based on minimisation of both structural and empirical risk, have a great attention for face detection. The SVM-based methods [1,6,7] yielded comparable or better performance than the MLP-based approaches.

Osuna [6] successfully applied the image space to SVMs for face detection. Qi et al. [1] enhanced performance of SVM-based face detector by using ICA features instead of raw image as input of SVM classifier. Qi obtained a larger margin of separation and fewer support vectors by using ICA. It has a few advantages over other unsupervised density estimation techniques by exploiting higher-order statistics and has been also adopted in face recognition [8] resulting in better performance than PCA which has been widely used for feature extraction. Study of Romdhani [7] was focused on the speed up of non-linear SVM. It is time consuming that the non-linear SVM operates by comparing an input to a full set of support vectors when considering the huge amount of observation windows to be searched.

In this paper, we propose a novel face detector based on combination of ICA feature extraction and SVM classifier. We contribute to face detection in two aspects. First, our method utilizes information on both face and near-face classes for the feature extraction, as well as learning a classification function. The extracted features obtained from two ICA spaces helped the SVM to learn a more reasonable boundary. Second, our classifier consists of parallel SVMs, trained in local sub-spaces. Face space is divided by k-means clustering algorithm and corresponding non-face examples to each face cluster are separately collected. Then multiple SVMs are trained for each training class. The parallelization of SVMs may be more effective in both accuracy and computational cost when the face space is multi-modal distributed.

2. IMPLICIT MODELING OF FACE AND NEAR-FACE CLASSES

As mentioned before, exact estimation of pdfs of face and non-face classes becomes difficult due to the high dimensionality, 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. It is shown in Section 4 that features obtained from the near-face class play the complemantary role of features from the face-class providing better classification results. ICA using an extended infomax algorithm [9] is adopted for unsupervised density estimations. ICA is performed on main eigenvectors of PCA to solve the problem that ICA is difficult to converge for the given entire data set and to control the number of independent basis vectors [1,8]. Residual error space in each ICA space is simply considered using Euclidean distance measure assuming an isotropic Gaussian data distribution. The independent 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 alternative case is also same. Near-face patterns were initially collected by matching an average face pattern and enlarged by a bootstrap technique [5]. Detailed procedure is shown below.

Let $\mathbf{X}$ be an input matrix whose rows are training images.

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

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 using an extended infomax algorithm [9] is described by

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

where $\mathbf{K}$ is the diagonal matrix whose elements differ to sources with either super- or sub-Gaussian distributions. $\mathbf{K}$ is determined by a following switching criterion [9].

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

where .^, .× and $\mathbf{E}$ are array power operator, array multiply operator and a row vector containing the mean value of each column.

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

$$\mathbf{W}\mathbf{P}_m^{\mathrm{T}} = \mathbf{U} \ \Rightarrow\ \mathbf{P}_m^{\mathrm{T}} = \mathbf{W}^{-1}\mathbf{U}. \tag{4}$$

Therefore the ICA reconstruction is obtained by

$$\mathbf{X}_{rec} = \mathbf{R}_m\mathbf{P}_m^{\mathrm{T}} \ \Rightarrow\ \mathbf{X}_{rec} = (\mathbf{X}\mathbf{P}_m)(\mathbf{W}^{-1}\mathbf{U}). \tag{5}$$

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

$$\mathbf{B}_m = \mathbf{X}\mathbf{P}_m\mathbf{W}^{-1}\ . \tag{6}$$

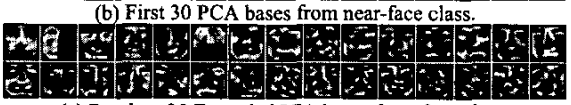
A residual error from the ICA space is utilised as a feature. Euclidean distance measure describes the error under the assumption of isotropic Gaussian distribution.

$$\varepsilon = \left|\mathbf{X} - \mathbf{X}_{rec}\right|^2 = \left|\mathbf{X} - \mathbf{X}\mathbf{P}_m\mathbf{W}^{-1}\mathbf{U}\right|^2 = \left|\mathbf{X}(\mathbf{I} - \mathbf{P}_m\mathbf{W}^{-1}\mathbf{U})\right|^2. \tag{7}$$

Figure 1 shows the learned basis images of PCA and extended ICA. As shown in Figure 1 (b), 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 edges. Both PCA and extended ICA basis sets appear holistic. In some face-related studies, locality of bases and representations is considered to be better for pattern classifications. We got local ICA bases by considering only super-Gaussian sources in learning. However, local and holistic ICA bases were similar in performance. Figure 2 shows the trained local ICA basis images from the face class.

(a) First 30 PCA bases from face class.

(b) First 30 PCA bases from near-face class.



(c) Random 30 Extended ICA bases from face class.

(d) Random 30 Extended ICA bases from near-face class.

Figure 1. Learned basis images.

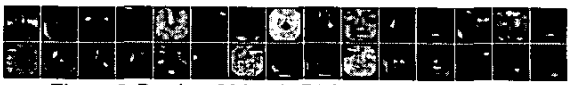


Figure 2. Random 30 local ICA bases from face class.

Two features of the residual errors in ICA face and near-face spaces are shown to be very effective in discrimination between the trained face and near-face patterns in Figure 3. These features may play a dominant role in learning a decision boundary. The other features obtained from two ICA representations of the face and near-face classes may enhance the quality of the decision boundary.

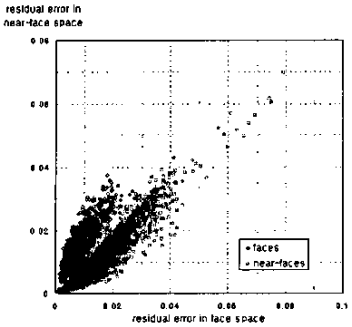


Figure 3. Residual errors of trained faces and near-faces.

3. PARALLEL NON-LINEAR SVMS IN LOCAL SUB-SPACES

In classifications by using the non-linear SVMs, run-time complexity is proportional to the number of support vectors (SVs). It is time-consuming to apply full SVM classifier to every pixel of image for face localization. Parallelization of SVMs reduces number of SVs and simplifies decision boundaries speeding up the algorithm. SVMs implicitly map the data (in our case, the data is feature vector described in Section 2) into a dot product space via a nonlinear mapping function. Then the SVMs learn a hyperplane which separates the data by a large margin. A test pattern, $\mathbf{x}_{test}$, is classified as a face or not by using the trained SVMs.

$$f(\mathbf{x}_{test}) = sign\left(\sum_{i=1}^{l} y_i \lambda_i \mathbf{K}(\mathbf{x}_{test}, \mathbf{x}_i) + b\right), \quad (8)$$

where y_i and $\mathbf{x}_i$ are a class label and a training feature vector respectively, λ_i and b are constants which are decided by learning, $\mathbf{K}$ is a polynomial kernel with a degree 2 and l is the number of SVs.

Suppose that overall distribution of training data is non-Gaussian and multi-modal. In this case, a complicated decision boundary is required as shown in the left image of Figure 4. It has a large number of SVs resulting in expensive computation cost and it is actually difficult to learn such a complex boundary in a high dimensional space. 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 Figure 4. Classification in a local sub-space has the advantages of using smaller number of SVs with a larger margin and possibly using a linear SVM.

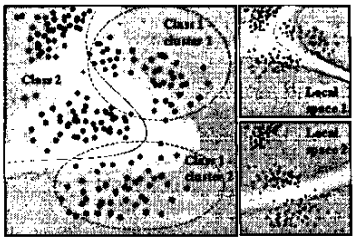


Figure 4. Learning a decision boundary in local sub-space.

Face and near-face classes may be multi-modal distributed and thus the parallel scheme of SVMs trained in local sub-spaces may improve the accuracy and speed of the classification. For the division of image space into local subspaces, K-means clustering algorithm is utilized. After the clustering, covariance matrices of clusters are computed. The near-face samples used for the training are separated by the nearest neighboring algorithm with Mahalanobis distances to the face centroids, as

$$M(\mathbf{x};\mathbf{u}_i,\Sigma_i) = \frac{1}{2}(d \cdot ln2\pi + ln|\Sigma_i| + (\mathbf{x}-\mathbf{u}_i)^T \Sigma_i^{-1}(\mathbf{x}-\mathbf{u}_i)), \quad (9)$$

where d is vector space dimensionality, $\mathbf{u}_i$ and Σ_i are the centroid and covariance matrix of a face cluster respectively. Both face and near-face classes are divided into ten subspaces. 10 SVMs are trained on the each set of face and near-face samples. For the classification of a new pattern, the nearest SVM is only applied to the pattern. Figure 5 shows 10 centroid images of the face class. They include intrinsic changes for identity and expression, and extrinsic changes for illumination and viewing geometry of faces.

Figure 5. Face centroid images

4. EXPERIMENTAL RESULTS AND DISCUSSION

The training set consists of various face and near-face images. The images were normalized to 20×20 pixels, histogram equalized, and reduced to 300 dimensional vectors by oval masking. The training set of face images included artificially rotated, shifted, resized faces. The near-face images were initially selected by matching an average template of faces and then enlarged by a bootstrap technique. To detect faces in images, the proposed classifier is performed on observation windows at every possible position and scale. Two sets of grey images were used for the performance evaluation. Set A contained 400

high-quality images with one face per image from the Olivetti image database. Set B contained 36 images of mixed quality, with a total of 172 faces from the Rowley [5] test set. Set A involved 1684800 pattern windows, while Set B involved 6178110.

4.1. Comparison of Feature Extraction Methods

We selected the first 50 PCA eigenvectors that span the face or near-face class with most energy. 50 elements of the ICA representation and a residual error of the face class were selected as features. 51 features were similarly selected from the near-face class. Totally 102 features represent the feature vector. It is shown in Figure 5 and Table 1 that this expansion of ICA features from the near-face class and residual error space largely increases the detection performance. The SVM was trained on 2,000 faces, 3,000 near-faces and 10,000 random non-faces by using a polynomial kernel with a degree 2. Our result is comparable to Rowley's, Osuna's [6] and Romdhani's [7], although the performance evaluation was performed on a slightly different training and test set.

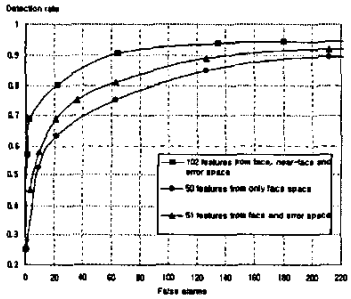


Figure 5. ROC curves on Set B

Table 1. Test results of feature extraction methods

	Set A		Set B	
	Detect Rate	False Alarms	Detect Rate	False Alarms
Typical*	97.2%	3	84.3%	127
Proposed**	98.5%	4	90.1%	62

* : 50 ICA features from face class. ** : 100 ICA features from face and near-face class and 2 residual errors.

4.2. Test Results of the Parallel SVMs

In Table 2, weighted averaged values of margin and number of SVs were calculated for the parallelized schemes by

$$value = \frac{1}{10}\sum_{k=1}^{10} \mathbf{P}(\mathbf{x}_{test} \mid \mathbf{C}_k) \cdot F(k), \qquad (10)$$

where $\mathbf{P}(\mathbf{x}_{test} \mid \mathbf{C}_k)$ is the likelihood of local sub-space $\mathbf{C}_k$ and $F(k)$ is the margin or number of SVs of each SVM.

Table 2. Learning and test results of SVMs on Set B

	Margin	# of SVs	Detect Rate	False Alarms
One SVM	0.2158	439	93.5%	120
Parallel Non-linear SVMs	0.5945	119.67	93.5%	145
Parallel linear SVMs	0.2213	157	93.5%	1989

It is noted that the proposed parallelization of SVMs provides a larger margin and a smaller number of SVs making the algorithm to be 3.7 times faster and have a similar face detection performance compared to one SVM trained from the entire data set. However, linear SVMs were not enough to learn reliable decision boundaries of 10 face and near-face sub-spaces.

5. CONCLUSION

We have presented a novel method of learning a decision boundary with a hybrid ICA and SVM approach. Face, near-face and residual error spaces are utilized together for the feature extraction using ICA. This largely increased the detection accuracy. Considering the speed of the algorithm, multiple SVMs were trained in local subspaces divided by K-means clustering. This parallelization of SVMs makes the algorithm approximately 4 times faster. The problems of deciding an optimal number of local sub-spaces and reducing the dimension of feature vectors still remain a further work.

6. REFERENCES

[1] Y.Qi, D.Doermann, and D.DeMenthon, "Hybrid ICA and SVM learning scheme for face detection", *ICASSP*, Utah, May 2001.

[2] V.Vapnik, *The nature of statistical learning theory*, Springer-Verlag, New York, 1995.

[3] S.Gong, S.J.McKenna and A.Psarrou, *Dynamic Vision : From Images to Face Recognition*, Imperial College Press, 2000.

[4] H.A. Rowley and T. Kanade, "Neural network-based face detection", *IEEE Trans. on PAMI*, vol. 20, no. 1, Jan. 1998.

[5] Kah-Kay Sung and Tomaso Poggio, "Example-based learning for view-based human face detection", *IEEE Trans. on PAMI*, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[6] Edgar Osuna, Robert Freund and Federico Girosi, "Training support vector machines: an application to face detection", *CVPR*, San Juan, June 1997.

[7] S. Romdhani, P. Torr, B. Scholkopf and A. Blake, "Computationally efficient face detection", *ICCV*, 2001.

[8] M.S. Bartlett and T.J. Sejnowski, "Independent component representations for face recognition", *SPIE Conf. On Human Vision and Electronic Imaging III*, San Jose, Jan. 1998.

[9] 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 Computation*, vol. 11(2), pp. 417-441, 1999.