HYBRID INDEPENDENT COMPONENT ANALYSIS AND SUPPORT VECTOR MACHINE LEARNING SCHEME FOR FACE DETECTION

Yuan Qi, David Doermann, and Daniel DeMenthon

The Laboratory for Language and Media Processing,
University of Maryland, College Park, MD 20742
E-mail: yuanqi@media.mit.edu, {doermann, daniel}@cfar.umd.edu

ABSTRACT

In this paper we propose a new hybrid unsupervised / supervised learning scheme that integrates Independent Component Analysis (ICA) with the Support Vector Machine (SVM) approach and apply this new learning scheme to the face detection problem. In low-level feature extraction, ICA produces independent image bases that emphasize edge information in the image data. In high-level classification, SVM classifies the ICA features as a face or non-faces. Our experimental results show that by using ICA features we obtained a larger margin of separation and fewer support vectors than by training SVM directly on the image data. This indicates better generalization performance, which was verified in our experiments.

1. INTRODUCTION

By exploiting higher-order statistics, ICA can not only separate mixed signals, but also effectively extract low-level features in signals. The independent ICA basis emphasizes the edge information in images [2], and unlike wavelets, ICA basis is adaptive to the training data. ICA feature extraction is also biologically plausible. As shown in [1, 9], an important characteristic of sensory processing in the brain is redundancy reduction. ICA algorithms have been successfully used in face recognition [14, 7].

SVM is a powerful supervised learning algorithm, which is rooted in statistical learning theory [15]. By minimizing the sum of the empirical risk and the complexity of the hypothesis space, SVM gives good generalization performance on many pattern recognition problems [3].

In this paper, we propose a hybrid unsupervised / supervised learning scheme that integrates ICA with SVM, and apply it to the face detection problem. Our experimental results show that by using ICA features instead of original image data, a larger margin of separation and fewer support vectors are obtained in SVM training. This indicates better generalization performance, which is demonstrated in our experiments.

2. HYBRID ICA/SVM LEARNING SCHEME

In this section we describe the hybrid ICA/SVM learning scheme, beginning with ICA feature extraction.

2.1. ICA Feature Extraction

Let us assume a linear mixture model

$$\mathbf{x} = A\mathbf{s}$$

where components of s are independent sources and unknown, A is unknown, and x is observed. ICA tries to estimate the matrix $\mathbf{W}$ in the reconstruction model

$$\mathbf{y} = W\mathbf{x},$$

Applying the natural gradient algorithm [12] to minimize the Kullback-Leibler divergence between the source signal vector s and its estimate $\mathbf{y}$ leads to the following learning rules:

$$\mathbf{Q} = \mathbf{I} - \mathbf{g}(\mathbf{y}(n))\mathbf{y}^T(n), \tag{1}$$

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \eta(n)\mathbf{Q}\mathbf{W}^T(n), \tag{2}$$

where $\mathbf{I}$ is the identity matrix, $\eta(n)$ is the learning rate, and $\mathbf{g}(\mathbf{y}) = (g(y_1), \cdots, g(y_n))^T$ is a nonlinear function. When the mixed signals have a super-Gaussian distribution, we can simply let $g(z) = 2\tanh(z)$.

Applying the rules above, we find the statistically independent basis images and then represent the image by the coefficients of the projection on those basis images. To control the number of independent basis images, we use the main eigenvectors of the training images instead of the original images to train ICA basis, following [14]. Since we assume the images are linear combinations of independent sources in the ICA mixture model, we do not lose information by replacing the original images with their eigenvectors that are linear combinations of the original images. Specifically, we have the following procedure.

First by viewing each image as an observation of a random vector, we compute the matrix P_m, whose columns are the eigenvectors that correspond to the maximal m eigenvalues. We then take P_m^T as the mixture $\mathbf{x}$ and apply the ICA algorithm as follows:

$$\begin{aligned} WP_m^T &= \mathbf{y} \\ \Rightarrow \quad P_m^T &= W^{-1}\mathbf{y} \end{aligned} \tag{3}$$

where each row of $\mathbf{y}$ represents an independent image basis.

From P_m, a minimum squared error approximation of $\mathbf{x}$ is obtained by

$$\mathbf{x_{rec}} = \mathbf{x}P_m P_m^T. \tag{4}$$

Substituting equation (3) into equation (4), we get

$$\mathbf{x_{rec}} = \mathbf{x}P_m W^{-1}\mathbf{y}. \tag{5}$$

The rows of $\mathbf{x}P_m W^{-1}$ are the coefficients for the linear combination of independent basis images in $\mathbf{y}$. Thus for the representation of a test image, which is a row vector $I_{1\times N}$, the ICA representation is

$$\mathbf{c} = IP_m W^{-1} \tag{6}$$

where $P_m W^{-1}$ is obtained during the ICA training procedure.

Figure 1 shows a set of learned independent basis images obtained from training images. These basis images look like localized oriented filters which emphasize image edge information.

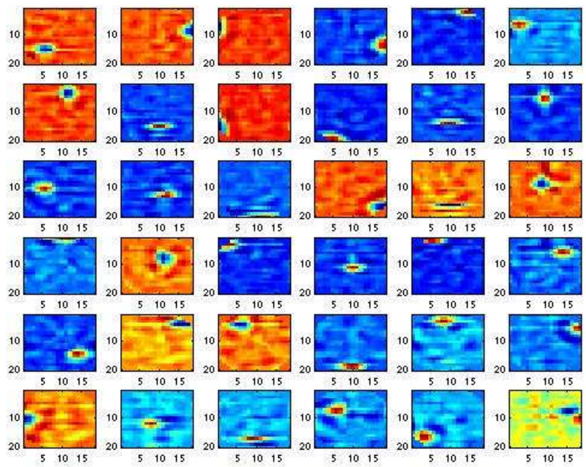

Fig. 1. Independent Basis Images

2.2. SVM Classification of ICA Features

After obtaining ICA features, we build the SVM training set $\{\mathbf{c}_i, d_i\}_{i=1}^l$ where $d_i = \{1, -1\}$ is the class type of ICA feature $\mathbf{c}_i$. l is the size of the training set.

Generally, SVM has the following form:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{l} y_i \lambda_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*\right) \tag{7}$$

where the parameters are obtained by maximizing the objective function

$$Q(\Lambda) = \sum_{i=1}^{l} \lambda_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{8}$$

subject to the constraints:

$$\begin{aligned} &\sum_{i=1}^{l} \lambda_i y_i = 0 \\ 0 \le \lambda_i \le C \qquad &\text{for } i = 1, 2, \ldots, l. \end{aligned} \tag{9}$$

By solving the above quadratic programming problem, SVM tries to maximize the margin between data points in the two classes and minimize the training errors simultaneously. According to a statistical learning theory developed by Vapnik [15], the following risk bound holds $\forall \alpha \in \Lambda$

$$R(\alpha) \le R_{emp}(\alpha) + \sqrt{\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}} \tag{10}$$

with probability $1-\eta$ where $0 \le \eta \le 1$, Λ is the parameter space and h is the Vapnik-Chervonenkis (VC) dimension of $f(\cdot, \alpha)$). The VC dimension h is a key concept described in [15], which measures the complexity of the hypothesis space.

Note that for SVM and given input vectors, the larger the separation margin between positive and negative examples, the smaller value of the upper bound of the VC dimension h [15]. Formally, the margin of separation is defined as the distance between the two classes, i.e.,

$$\text{Margin} = \frac{2}{||\mathbf{w}||} \tag{11}$$

An alternative bound on the actual risk of SVM is:

$$E[P(error)] \le \frac{E[\text{Number of support vectors}]}{\text{Number of training samples}} \tag{12}$$

It is reported that this bound is tighter than the risk bound, (i.e., Equation (10)), though it is not as predictive as the risk bound in test errors [3].

3. THE HYBRID ICA/SVM BASED FACE DETECTION SYSTEM

The face detection problem has been addressed by many researchers using different approaches, such as statistical modeling, template matching, and color and motion information [13, 11, 10, 8]. The main difference of our detection system from others is the use of a hybrid ICA/SVM learning scheme. The training part of our system consists of the following steps:

1. In a training set, 20×20 face and non-face patterns are labeled as 1 and -1 respectively. These face patterns include faces with different facial expressions and under different views; (Figures 2 and 3).

2. The image blocks are histogram equalized and pixels close to the block boundary are removed to reduce background noise and compensate for illumination difference [11, 10].

3. An ICA algorithm is applied to learn the independent image basis and generate the training ICA features.

4. Using the ICA features, the SVM is trained for classification. Since it is difficult to find a good representative set of non-face patterns, a bootstrapping technique is used to add mis-classified non-face patterns into the training set, and then the SVM is re-trained to get a better decision plane.

Fig. 2. Face Patterns with Different Facial Expressions Used in Training

Fig. 3. Face Patterns Under Slightly Different Views Used in Training

To speed up the detection procedure we first apply a skin color filter [5], which in essence is a thresholding method in image hue and saturation space, to locate face candidate areas. In our experiments, we found skin color is not robust to image or video coding distortion, so it is only used as an preprocessing step to eliminate non-skin areas in an image. And we exploit some morphological operations to remove small blobs and improve the binary face candidate mask. Also, we re-scale the image since we do not know the size of faces in the images. We extract ICA features from a sliding window over the scaled candidate areas after the similar preprocessing as in the training part, and perform SVM classification. Finally after a simple post-processing to remove some spurious detection, the final detection result is produced.

4. EXPERIMENTAL RESULTS

In Table 1 we compared the hybrid ICA/SVM training results with direct SVM training results without using ICA feature extraction when choosing $C = 230$ in Equation (9). From this table we see that the hybrid ICA/SVM training achieves much larger margin and fewer support vectors than the direct SVM training while $R_{emp} = 0$ for both approaches. This suggests hybrid ICA/SVM has better generalization capacity according to Equation (10) and (12).

Detection System	Margin	Number of Support Vectors
Hybrid ICA/SVM Detection	306.08	910
Direct SVM Detection	7.29	916

Table 1. Comparison of Training Results. Note: both approaches achieves zero training errors

We then tested our hybrid ICA/SVM detection system on 820 face images from the LAMP face database [6] and from the Essex facial image database [4], as well as on 100884 non-face image blocks which we obtained from the LAMP face database and the web. In the LAMP face database, the face images were recorded from broadcast television. In the Essex facial image database, face examples have expression changes and position changes. The test results are summarized in Table 2. From this table, we see that the hybrid learning scheme effectively reduced the test errors. In Figure 4, we give a detection example on a color image.

Detection System	Number of Miss Detections	Number of False Detections
Hybrid ICA/SVM Detection	39	54
Direct SVM Detection	41	252

Table 2. Test Results

5. ACKNOWLEDGMENTS

The partial support of the Department of Defense under contract MDA 9049-6c-1250 is gratefully acknowledged.

6. CONCLUSION

In this paper, we have presented a new hybrid supervised / unsupervised learning scheme that integrates ICA and SVM to address the face detection problem. From Figure (1), we see that the ICA bases emphasize edge information in the

Fig. 4. Face Detection Example: Final Result

image data. Our experimental results demonstrated that by using ICA features instead of the original image data, SVM effectively enlarged the separation margin and reduced the number of support vectors in training and finally produced fewer test errors. Finally, we would like to point out that as a general learning scheme, the hybrid ICA/SVM scheme can be applied to other pattern recognition problems than face detection.

7. REFERENCES

[1] H.B. Barlow. What is the computational goal of the neocortex ? In *Large-scale Neuronal Theories of the Brain*. MIT Press, Cambridge, MA, 1994.

[2] A. J. Bell and T. J. Sejnowski. Edges are the independent components of natural scenes. *NIPS 9*, page 381, Cambridge, MA, 1997. MIT Press.

[3] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2), 1998.

[4] The Essex facial image database, http://cswww.essex.ac.uk/mv/projects.html.

[5] M. M. Fleck, D. A. Forsyth, and C. Bregler. Finding naked people. In *Proc. 4'th European Conf on Computer Vision*, 1996.

[6] The ViPER video database, the Language and Media Processing Lab, University of Maryland at College Park, http://lamp.cfar.umd.edu

[7] C. Liu and H. Wechsler. Comparative assessment of independent component analysis (ICA) for face recognition. In *Proc. of Int'l Conf. on Audio- and Video-based Biometric Person Authentication*, Washington D. C., March 22-24 1999.

[8] B. Moghaddam and A.P. Pentland. Probabilistic visual learning for object recognition. *IEEE Trans. on PAMI* , 19(7):696–710, July 1997.

[9] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? In *Vision Research*, 37:33113325, 1997.

[10] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection. In *Proc. CVPR'97*, 1997.

[11] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. on PAMI*, 20(1):23–38, January 1998.

[12] A. Cichocki S. Amari and H. H. Yang. A new learning algorithm for blind signal separation. In *NIPS 8*, pp. 752–763, Cambridge, MA, 1996. MIT Press.

[13] K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Trans. on PAMI*, 20(1):39–51, January 1998.

[14] M.S. Bartlett, H.M. Lades, and T.J. Sejnowski. Independent component representations for face recognition. In *SPIE Conf. on Human Vision and Electronic Imaging III*, vol. 3299, pp. 528–539, 1998.

[15] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.