A Nonlinear Approach for Face Sketch Synthesis and Recognition

Qingshan Liu[1,2], Xiaoou Tang[1,3], Hongliang Jin[1,2], Hanqing Lu[2], Songde Ma[2]
[1]Department of Information Engineering, The Chinese University of Hong Kong
[2]National Laboratory of Pattern Recognition, Institute of Automation, CAS, P.R. China.
[3]Microsoft Research Asia, Beijing, P.R. China
Email: {qsliu, hljin, luhq, masd}@nlpr.ia.ac.cn, xtang@ie.cuhk.edu.hk

Abstract

Most face recognition systems focus on photo-based face recognition. In this paper, we present a face recognition system based on face sketches. The proposed system contains two elements: pseudo-sketch synthesis and sketch recognition. The pseudo-sketch generation method is based on local linear preserving of geometry between photo and sketch images, which is inspired by the idea of locally linear embedding. The nonlinear discriminate analysis is used to recognize the probe sketch from the synthesized pseudo-sketches. Experimental results on over 600 photo-sketch pairs show that the performance of the proposed method is encouraging.

1. Introduction

Face recognition has attracted much attention in recent years due to its potential applications, and a lot of algorithms and systems have been proposed [1]. However, current face recognition systems almost exclusively focus on photo-based face recognition. In this paper, we develop a face recognition system based on face sketches. This is very useful for face recognition applications where a probe face photo is not available. For example, in law enforcement application, very often only a sketch drawing based on the recollection of an eyewitness is available. We may only be able to draw a sketch based on an old photo to search for a long lost relative. To be able to search a large photo database using a drawn sketch will be very useful for the above applications. For these applications, we only focus on sketches of plain style, i.e., without exaggeration, so that the sketch can realistically describe the real subject. Figure 1 gives two samples of photo-sketch pairs.

From Figure 1, we can see that face sketches and photos are of different modalities, which bring more difficulties for face sketch recognition than normal face recognition based on photo images, for it is hard and improper to directly measure their similarities. Intuitively, we may want to recover the photo image from a sketch. However, this is an ill-pose problem, for a sketch image often only has main facial features, and many details features are ignored. An alternative approach is to generate a pseudo-sketch from a photo image [9,10,11], which is used in our scheme.

(a) (b)

Figure 1. Samples of photo-sketch pairs: (a) Photos, (b) Sketches.

Let I_p be a photo image, and I_s be a sketch image. The procedure of pseudo-sketch generation is equivalent to setting up a mapping relation P between a photo and a sketch, $I_s = P(I_p)$. Intuitively, P should be a complex nonlinear mapping. In this paper, we present a local geometry preserving based nonlinear method to approximate the mapping function P. Our method is inspired by a recently proposed manifold learning method called locally linear embedding (LLE) [2,3]. We assume that small image patches in the photo and sketch images form manifold with similar local geometry in two different image spaces, and then we can automatically generate a pseudo-sketch of a new photo from given training photo-sketch pair samples based on local nearest neighbors reconstruction.

Another element of the proposed system is sketch

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

recognition, i.e., identifying the probe sketch from the pseudo-sketches. Though we only focus on sketches of plain style, it is also inevitable to have some distortions when artists draw the sketches. In addition, the weight combination of local nearest neighbors brings some blurs to the pseudo-sketches. In order to reduce the influence of nonlinear variations due to distortions and blurs, we adopt the kernel based nonlinear discriminant analysis (KNDA) [4] for sketch recognition, which combines the nonlinear kernel trick with linear discriminant analysis (LDA). We evaluate the recognition performance on a database of 606 subjects, and compared KNDA with Bayesian subspace [5], LDA [6] and PCA [7].

2. Related Work

Recently, an effective face sketch recognition system is developed based on an assumption that the mapping P between a photo and a sketch can be approximated as a linear process [9,10]. According to PCA reconstruction, a new face photo and sketch can be represented by a linear combination of training photo-sketch pair samples T_p and T_s respectively, $I_p = T_p C_p$ and $I_s = T_s C_s$, where each column of T_p and T_s represents a training photo sample and the corresponding sketch sample respectively, and the linear combination coefficient C_p and C_s are column vectors obtained by PCA eigentransform. Based on the linear assumption, a sketch should have a similar linear reconstruction to its corresponding photo image, $C_s \approx C_p$, so the pseudo-sketch of a photo can be synthesized with $T_s C_p$. This technique is called eigentransform based pseudo-sketch synthesis. Three distances based on $C_p - C_s$ were defined for sketch recognition. To further improve the performance, in [11], shape and texture information in a photo are separated, and such an eigentransform is conducted on shape and texture respectively. Finally, a Bayesian subspace classifier is employed to recognize the probe sketch from the pseudo-sketches.

However, it is not difficult to notice that approximating the sketch drawing process of an artist by a linear process is not accurate. This process may be better estimated as a nonlinear process. The proposed method in this paper differs from [9,10,11] in that two nonlinear techniques are presented for pseudo-sketch synthesis and sketch recognition, i.e., the local geometry preserving based pseudo-sketch synthesis and the KNDA based sketch recognition.

3. The Pseudo-Sketch Synthesis

Because sketch and photo images are in different modalities, it is difficult to directly measure their similarities. In the proposed system, we first generate the pseudo-sketch of a photo image, and then match the probe sketch with the pseudo-sketches. Inspired by the idea of LLE, we present a local geometry preserving based method to learn the mapping relation between photos and sketches from a training set.

LLE is a promising manifold learning method, and is widely used for nonlinear dimension reduction of high-dimensional data and image analysis [2,3,12]. The basic idea of LLE is to compute neighbor-preserving mapping between a high-dimensional original data space and a low-dimensional feature space, based on simple geometric intuition that each data and its neighbors lie on or close to a locally patch of the manifold. According to this idea, we present a method based on the idea of local geometry preserving for pseudo-sketch synthesis, with the help of training photo and sketch image pairs, T_p and T_s.

Due to the complexity of face structure, we adopt a patch-based strategy as [12,13]. We divide the photo and sketch images into N small overlapping image patches in the same way (the photo and sketch images are geometrically normalized by fixing the locations of eyes in our experiments). We denote photo and sketch image patches as I_p^t and I_s^t, $t = 1,2,...,N$. In this paper, we only consider sketch images of plain style only, so we can assume that corresponding photo and sketch image patches form manifolds with similar local geometry in two different image spaces. Thus, similar to LLE, for each photo image patch I_p^t, we first fit it with its K nearest neighbors from training samples T_p^t, and calculate the reconstruction weights, and then its corresponding sketch patch I_s^t can be estimated from training sketch samples T_s^t by preserving the local geometry. The sketch synthesis algorithm is summarized as follows:

For $t = 1,2,...N$

(1) For a photo patch I_p^t, find its K nearest neighbors $\hat{I}_{pk}^t \in T_p^t$, $k = 1,2,...,K$.

(2). Compute the reconstruction weights of the neighbors, $w_{pk}^t, k = 1,2,...,K$, that minimize the error of reconstructing I_p^t.

(3). Based on local geometry preserving, estimate its sketch patch I_s^t using the corresponding sketch patches $\hat{I}_{sk}^t \in T_s^t$ of the K nearest neighbors $\hat{I}_{pk}^t$ and the reconstruction weights w_{pk}^t, $k = 1,2,...,K$.

END.

In step (1), the Euclidean distance is used to find the nearest neighbors, and the local reconstruction based on the K nearest neighbors in step (2) is achieved by minimizing,

$$\varepsilon^t(w) = \left\| I_p^t - \sum_{k=1}^{K} w_{pk}^t \hat{I}_{pk}^t \right\|^2, \qquad (1)$$

subject to $\sum_{k=1}^{K} w_{pk}^t = 1$. This is a constrained least squares problem. By defining a $K \times K$ matrix Q,

$$Q(i,j) = (I_p^t - \hat{I}_{pi}^t)^T (I_p^t - \hat{I}_{pj}^t), \qquad (2)$$

and $R = Q^{-1}$, this constrained least squares problem has the following close-form solution [3]:

$$w_{pk}^t = \frac{\sum_{m=1}^{K} R(k,m)}{\sum_{i=1}^{K} \sum_{j=1}^{K} R(i,j)}, \qquad (3)$$

where $k = 1,2,...,K$.

In step (3), the sketch patch I_s^t can be estimated based on w_{pk}^t and the corresponding K sketch patches $\hat{I}_{sk}^t$ in T_s^t.

$$I_s^t = \sum_{k=1}^{K} w_{pk}^t \hat{I}_{sk}^t. \qquad (4)$$

In order to enforce local compatibility and smoothness between adjacent synthetic sketch patches, we use an average for overlapped regions in the final reconstructed results.

From above description, we can see that there are three parameters: the number of neighbors K, the patch size, and the degree of overlapping between adjacent patches. We find that blur appears when the number of neighbors K is too large. Figure 2 shows the comparison of $K = 5,25,50$ with a patch size 13×13. In our experiments, we set $K = 5$. The idea of overlapping is to keep smooth transition between two adjacent patches. We keep 2/3 region overlapping. For the patch size, we find that some details are disappeared when the size is too large, and noise appears with too small patches. In addition, the size of patches should be related to the size of the photo and sketch images. In our experiments, both the face photo and sketch images have a size of 160×100, and we set the patch size as 13×13. Figure 3 gives a comparison of pseudo-sketches with different patch sizes.

(a) (b) (c)

Figure 2. Comparison of pseudo-sketches with different neighbor sizes: (a) $K = 5$, (b) $K = 25$, (c) $K = 50$.

(a) (b) (c)

Figure 3. Comparison of pseudo-sketches with different patch sizes: (a) 7×7, (b) 13×13, (c) 19×19.

4. Sketch Recognition

Face sketch recognition is to measure similarities between the probe sketch and the pseudo-sketches from photos. In [11], Bayesian subspace classifier is used to improve the sketch recognition performance, in which the distribution of the difference between a real sketch drawn by an artist and its corresponding pseudo-sketch is assumed to be a Gaussian distribution [5]. In this paper, we adopt the KNDA based nonlinear discriminative classifier for sketch recognition, in order to better describe nonlinear variations due to distortions and blurs in the real sketches drawn by artists and pseudo-sketches.

KNDA is a nonlinear version of LDA, and it is widely used in pattern recognition [14,15,16]. The idea of KNDA is first to map the input data X into an implicit feature space F with the nonlinear kernel trick, $\phi: x \in R^D \to \phi(x) \in F$, and then LDA is performed in F to extract nonlinear discriminating features of the original data. In implementation, the implicit feature vector ϕ

does not need to be explicitly calculated; instead it is embodied by computing the inner product of two vectors in F using a kernel function, $k(x_1, x_2) = (\phi(x_1) \cdot \phi(x_2))$ [4].

Assuming the training set has N images and C classes. X_c represents the sample set of the c-th class with N_c samples. Define the within class scatter $S_w^\phi = \sum_{i=1}^{C} \sum_{\phi(x_j)\in X_i} (\phi(x_j) - \mu_i)(\phi(x_j) - \mu_i)^T$ and the between class scatter $S_b^\phi = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T$ in F, where μ_i and μ are the mean of the i-th class samples and the mean of all the samples in F respectively, and $\phi(x_j) \in X_i$ means x_j belongs to the i-th class. Performing LDA in F is equivalent to maximizing the following objective function:

$$J(W^\phi) = \arg\max_{W^\phi} \frac{|(W^\phi)^T S_b^\phi W^\phi|}{|(W^\phi)^T S_w^\phi W^\phi|}. \qquad (5)$$

Because W^ϕ is a linear transformation in F, any solution $w^\phi \in W^\phi$ can be represented by $w^\phi = \sum_{i=1}^{N} \alpha_i \phi(x_i)$ due to linear transform property. Define the column vector $K_i = (k(x_i, x_1), k(x_i, x_2), \ldots k(x_i, x_N))^T$. Then equation (4) can be rewritten as:

$$J(\alpha) = \arg\max_{\alpha} \frac{|\alpha^T G_b \alpha|}{|\alpha^T G_w \alpha|}, \qquad (6)$$

where $G_b = \sum_{i=1}^{C} N_i(m_i - \overline{m})(m_i - \overline{m})^T$, $m_i = \frac{1}{N_i}\sum_{j=1}^{N_i} K_j$ with $K_j \in X_i$ which represents x_j belonging to the i-th class, $\overline{m}$ is the mean of all the m_i , and $G_w = \sum_{i=1}^{C} \sum_{K_j \in X_i} (K_j - m_i)(K_j - m_i)^T$. Thus, the problem of KNDA is converted into finding the leading eigenvectors of $G_w^{-1} G_b$.

Practically, KNDA produces a nonlinear subspace. We project the probe sketch and pseudo-sketch into this subspace, and measure their similarity based on distance in the subspace, $d = \left\| y^r - y^p \right\|$, where y^r and y^p are the projections of the probe sketch and pseudo-sketch, $y = \alpha^T (k(x_1, x), k(x_2, x), \ldots, k(x_n, x))^T$.

Notice, each class has two samples: a real sketch drawn by an artist and a pseudo-sketch in our study.

5. Experiments

We conduct the experiments on a database of 606 persons [11]. Each person has a frontal face photo image and a sketch image with plain style drawn by an artist. We divide the database into two sets. One is a training set containing 306 persons, and the other is a testing set that is composed of the rest 300 persons. The training set has two functions. One is for generating the pseudo-sketches of photo images in the testing set, and another is for learning classifiers. Of Course, as for the training set, we also need pseudo-sketches in learning classifiers. In order to keep the training set and testing set separated, we use the leave-one-out strategy to generate the pseudo-sketches in the training set.

5.1. Pseudo-Sketch Synthesis Performance

Figure 4 (c) shows some results of pseudo-sketch synthesis based on local geometry preserving, and comparison with the sketches drawn by artists and obtained by the eigentransform method in [9,10]. We can see that the proposed method is better than the eigentransform method. Basically, the pseudo-sketch reconstructed by the proposed method can well approximate the real sketches. Some blurs exist because we use the weighted sum of K nearest neighbors and overlapping between two adjacent patches for smoothness. The proposed method uses the idea of local geometry preserving to approximate the complex nonlinear mapping between photo and sketch patches, while the eigentranform method simply regards the mapping as a linear process. Though in [11], the synthesis effect of the eigentransform method can be improved by separating shape and texture, it is based on geometrical feature points, so some facial information cannot well be reconstructed.

5.2 Sketch Recognition Performance

Like most face recognition approaches, we remove the hair and background with a mask before doing recognition, for they are unreliable factors. The polynomial kernel function is selected for KNDA, $k(x_1, x_2) = (a(x_1 \cdot x_2) + b)^d$, and the parameters are set as $a = 10^{-3}$, $b = 1$, and $d = 3$.

Besides the Bayesian subspace method, we also compare KNDA with LDA [6] and PCA [7]. As for

Bayesian subspace, we respectively reserve 90% and 95% energy in the PCA principal subspace, which means the ratio of energy, $\sum_{i=1}^{M}\lambda_i / \sum_{i=1}^{N}\lambda_i$, being equivalent to 0.9 and 0.95 respectively, where $\sum_{i=1}^{M}\lambda_i$ represents the sum of eigenvalues in the principal subspace, and $\sum_{i=1}^{N}\lambda_i$ is the sum of all the eigenvalues. For simplify, we denote them as Bayes90 and Bayes95 in the following. Similar to KNDA, the distance based similarity measure is used for LDA and PCA. Because there are not enough training samples to

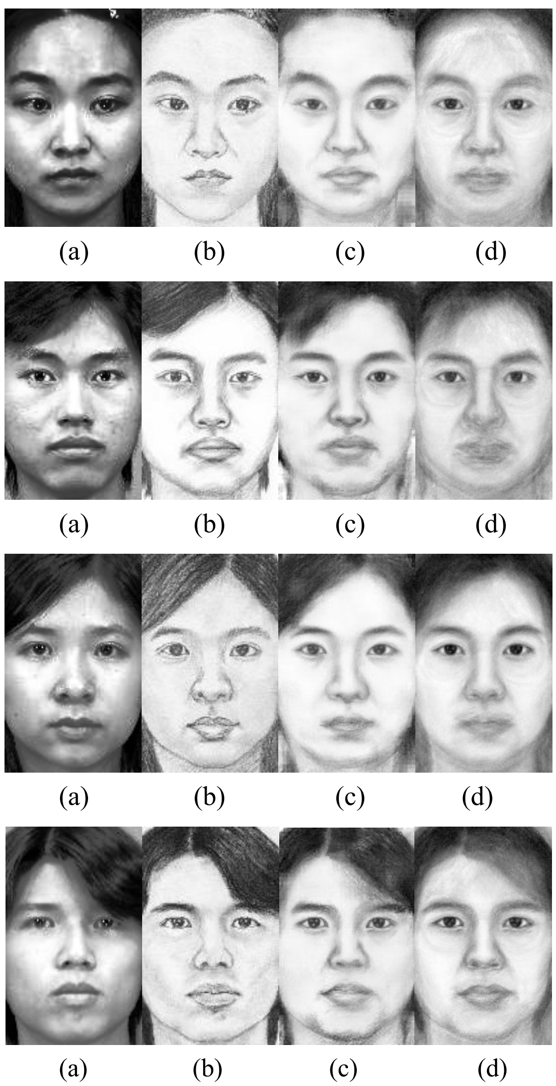

Figure 4. Some pseudo-sketch synthesis results: (a) photo images, (b) sketches drawn by artists, (c) pseudo-sketches with the proposed method, (d) pseudo-sketches with the eigentransform method.

guarantee the within class scatter in LDA and K_W in KNDA non-singularity, we adopt the trick of enhanced LDA for LDA [6], and we use $K'_W = K_W + \mu I$ to replace K_W for KNDA, where μ is a small constant (10^{-4}) and I is the identity matrix [14].

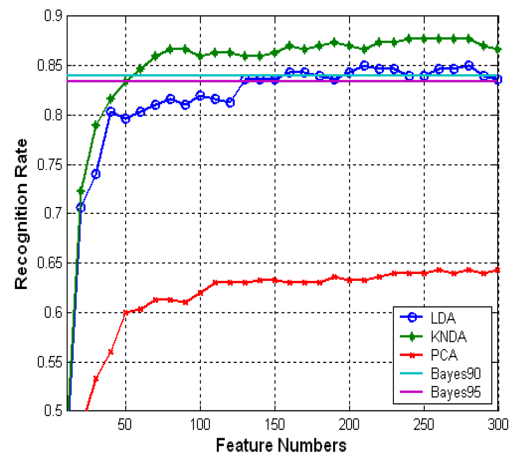

Figure 5. Comparison of recognition rate.

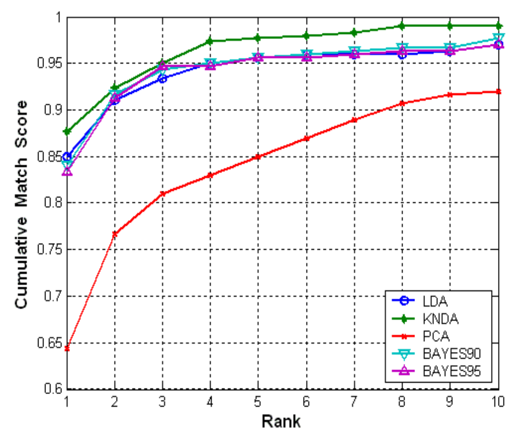

Figure 6. Comparison of cumulative match scores.

Figure 5 reports the recognition accuracy of KNDA, Bayesian, LDA and PCA. We can see that the performance of KNDA is better than the other three methods. KNDA has the highest recognition rate 87.67%, while the best recognition rates of LDA and PCA are 85% and 64.33% respectively. Bayes90 gets the recognition rate of 84 % and Bayes95 is 83.33%. In Figure 6, we use the cumulative score to evaluate the performance of the algorithms. Actually, Figure 6 reports the results of rank 1. It can be found that KNDA also has better performance than Bayesian subspace, LDA and PCA. KNDA can reach 99 % at rank 8, while the other three are still below 98 % up to rank 10. PCA has the worst performance, for it is optimal for reconstruction but not for discriminating one class from the others. Though LDA is based on the goal of separating one class from the others, it is a linear method in nature. The performance of Bayesian subspace is similar to that of LDA, since it also assumes the distribution of the intra-personal difference as a Guassian. In practice, there exist complex nonlinear variations due to man-made distortions and blurs in the real sketches and pseudo-sketches.

In [9,10,11], extensive experiments show that the eigentransform method outperforms the conventional eigenface based method [7] and EGM based method [8]. In [11], in order to further improve its performance, shape and texture information of the photos are separated, and the Bayesian subspace classifier is performed. We call it the modified eigentransform method. Using a testing set containing 300 photo-sketch pairs, the first match for the both conventional methods is no more than 30%, and the tenth rank is no more than 60% [11]. The modified eigentransform method gets the recognition rate of 81.33% at rank 1 and the best recognition of 97% within rank 10, while the corresponding performances of the proposed method obtains 87.67% and 99%.

6. Conclusions.

We proposed a new face recognition system based on face sketch image probing. It has potential applications in law enforcement and other photo search application. The proposed system contains two elements: pseudo-sketch synthesis and face sketch recognition. Based on local geometry preserving, pseudo-sketches of photos can be automatically synthesized with the help of training photo-sketch pair samples. The KNDA based nonlinear discriminating classifier is adopted to match the probe sketch with the pseudo-sketches. Experimental results show the effectiveness of the proposed method.

Acknowledgements

This work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong SAR and a Hong Kong RGC/NSFC joint project N_CUHK 409/03 (NSFC No. 60318003), and partially supported by NSFC Grant No. 60405005.

Reference

[1]. W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, Face Recognition: A Literature Survey, *CS-Tech Report-4167*, University of Maryland, 2000.

[2]. S. T. Roweis and L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290(5500), pp.2323-2326, 2000.

[3]. D. Ridder and R. Duin, Locally Linear Embedding For Classification, *Tech-report, PH-2002-01*, Imaging Science & Technology Department, Delft University of Technology, Netherlands.

[4]. S. Mika, G. Ratsch, and J. Weston. Fisher Discriminant Analysis with Kernels. *Neural Networks for Signal Processing IX*, pp 41-48, 1999.

[5]. B. Moghaddam, T. Jebara, and A. Pentland, Bayesian Face Recognition, *Pattern Recognition*, Vol. 33, pp. 1771-1782, 2000.

[6]. C. J. Liu and H. Wechsler, Enhanced Fisher Linear Discriminant Models For Face Recognition, *Proc. of. Int. Conf. Pattern Recognition,* 1998.

[7]. M. Turk and A. Pentland, Eigenfaces For Recognition, *J.of Cognitive Neuroscience*, Vol. 3, No. 2, pp. 71-86, 1991.

[8]. L. Wiskott, J. M. Fellous, N. Kruger, and C. Malsburg, Face Recognition by Elastic Bunch Graph Matching, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 775-779, July, 1997.

[9]. X. Tang and X. Wang, Face Photo Recognition Using Sketch, *Proc. of. Int. Conf. Image Processing,* 2002.

[10]. X. Tang and X. Wang, Face Sketch Recognition, *IEEE Trans. On Circuit System and Video Technology,* Vol.14, pp.50-57, 2004.

[11]. X. Tang and X. Wang, Face Sketch Synthesis and Recognition, Proc. of. *Int. Conf. Computer Vision,* 2003.

[12]. H. Chang, D. Yeung, and Y. Xiong, Super- Resolution Through Neigbhor Embedding, *Proc. of. Int. Conf. Computer Vision and Pattern Recognition*, 2004.

[13]. W. T. Freeman, T. R. Jones, and E. C. Pasztor, Example- based Super-Resolution, *Tech-Report TR-2001-030*, Mitsubishi Electric Research Laboratories, 2001.

[14]. Q. S. Liu, R. Huang, H. Q. Lu, S. D. Ma, Face Recognition Using Kernel Based Fisher Discriminant Analysis, *Proc. of. Int. Conf. Automatic Face and Gesture Recognition*, 2002.

[15]. M. H. Yang, Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods, *Proc. of. Int. Conf. Automatic Face and Gesture Recognition*, pp. 215-220, 2002.

[16]. Y. Wu, T. S. Huang, and K. Toyama, Self- Supervised Learning for Object based on Kernel Discriminant-EM Algorithm, *Proc. of Int. Conf. Computer Vision*, 2001.