Face Annotation Using Transductive Kernel Fisher Discriminant

Jianke Zhu, Steven C.H. Hoi, and Michael R. Lyu

***Abstract*—Face annotation in images and videos enjoys many potential applications in multimedia information retrieval. Face annotation usually requires many training data labeled by hand in order to build effective classifiers. This is particularly challenging when annotating faces on large-scale collections of media data, in which huge labeling efforts would be very expensive. As a result, traditional supervised face annotation methods often suffer from insufficient training data. To attack this challenge, in this paper, we propose a novel Transductive Kernel Fisher Discriminant (TKFD) scheme for face annotation, which outperforms traditional supervised annotation methods with few training data. The main idea of our approach is to solve the Fisher's discriminant using deformed kernels incorporating the information of both labeled and unlabeled data. To evaluate the effectiveness of our method, we have conducted extensive experiments on three types of multimedia testbeds: the FRGC benchmark face dataset, the Yahoo! web image collection, and the TRECVID video data collection. The experimental results show that our TKFD algorithm is more effective than traditional supervised approaches, especially when there are very few training data.**

***Index Terms*—Face annotation, image annotation, kernel Fisher discriminant, multimedia information retrieval, supervised learning, transductive kernel Fisher discriminant, transductive learning.**

I. Introduction

IMAGE annotation enables traditional text based search engines to index and retrieve large collections of media data effectively which has received a rapid growth of research attention in recent years [1]–[6]. Although numerous research efforts have been devoted to content-based image annotation and retrieval [7], the general image annotation problem is still a very challenging research issue due to the semantic gap between low-level visual features and high-level semantic concepts [8], [9]. We are still a long way from achieving a practical solution of general image annotation for web-scale applications.

In general, image annotation can be considered a typical object detection and recognition problem, in which a variety of concept detectors can be developed and applied. Among various concept detectors, face annotation, may be one of the most important and so far the most effective components for image annotation tasks. Face annotation is a task to label the facial images, which has recently received a surge of research attention in the multimedia retrieval community due to its numerous potential applications [10]–[13]. One such application is to support the manual insertion of name labels into photo albums, which can facilitate photo management and search tasks [14]. Another significant application is the annotation of faces on web images or photos, such as web news images [11]. This would enable current text based search engines to retrieve the content of facial images effectively by text based indexing and searching ways, which can facilitate traditional content-based image retrieval [15], [7], [16]. Face annotation also has some important applications in the video domain. For example, detecting important persons in video data, such as news videos, can help content-based video retrieval tasks significantly [17], [18]. These potential applications are often very large-scale, making the face annotation tasks very challenging in practice.

Face annotation is often regarded as a supervised classification problem, in which traditional face recognition methods are directly applied to solve the problem. Traditional face recognition methods are usually based on supervised learning techniques, which typically require a large number of training faces in order to achieve satisfactory performance. In large-scale applications, it is excessively costly to manually label large amount of training data. Therefore, it is critically important to develop an effective annotation method which is able to annotate faces effectively with small numbers of training examples. Since there are usually large amounts of unlabeled data available in a given face annotation task, taking advantage of these unlabeled data would offer a worthwhile advantage. This motivates us to explore transductive learning or semi-supervised learning techniques for face annotation tasks [19], [20].

Although transductive learning and semi-supervised learning techniques have already been actively studied in machine learning communities [20], the problem of choosing an appropriate classification method for face annotation remains unsolved. The choice of classification method is of great importance for achieving satisfactory annotation performance. In traditional face recognition problems, Fisher's linear discriminant analysis [21] and its kernel variants [22] are generally regarded as the state-of-the-art methods in face recognition tasks. Considering that face annotation is closely related to face recognition, developing transductive techniques of Fisher's linear discriminant analysis is likely to be a promising solution for face annotation. To this end, we propose a novel Transductive Kernel Fisher Discriminant (TKFD) scheme, which takes advantages of both labeled and unlabeled data for face annotation tasks. The main idea of our solution is to convert

Manuscript received November 22, 2006; revised September 5, 2007. The work was supported by the Innovation and Technology under Fund ITS/084/07 and by the Research Grants Council under Earmarked Grant CUHK4150/07E. The asscoiate editor coordinating the review of this manuscript and approving it for publication was Dr. Bangalore S. Manjunath.

J. Zhu and M. R. Lyu are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China (e-mail: jkzhu@cse.cuhk.edu.hk; lyu@cse.cuhk.edu.hk).

S. C. H. Hoi is with the School of Computer Engineering of Nanyang Technological University, Singapore 639798 (e-mail: chhoi@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2007.911245

the traditional Kernel Fisher Discriminant (KFD) into a transductive technique, which still has no straightforward solution available in the literature. As we know, for Kernel Fisher Discriminant (KFD), the kernel function has an essential impact on the classification performance. Therefore, in our TKFD solution, we propose to first induce a new transductive kernel by employing kernel deformation techniques to incorporate information from unlabeled data into the original kernel, and then apply the new kernel to classification tasks based on the Kernel Fisher Discriminant. Compared with traditional KFD methods, our TKFD approach is more effective particularly when there are only a small number of labeled data. We have conducted extensive experiments to evaluate the performance of our algorithm on three kinds of testbeds, namely the Face Recognition Grand Challenge (FRGC) benchmark dataset [23], the Yahoo! news images from WWW [11], and the TRECVID 2005 video dataset [24].

The rest of this paper is organized as follows. Section II reviews the existing work on face annotation. Section III presents the main methodology of our face annotation solution. We first introduce the Kernel Fisher Discriminant, which is considered the state-of-the-art approach for traditional face recognition. We then discuss how to induce a transductive kernel using the kernel deformation principle for incorporating information from unlabeled data into an input kernel. Finally, we give the algorithm of the Transductive Kernel Fisher Discriminant for face annotation. Section IV presents the experimental evaluations of TKFD on the three kinds of testbeds. Section V discusses the limitation of our solution and some future directions. Section VI sets out our conclusion.

II. Related Work

Considerable research effort has been devoted to face annotation problems in the multimedia community recently [10]–[13], [25], [26]. Most previous studies usually assume textual information is available and there exist correspondences between visual image content and texts, such as between web images and surrounding texts [11], or video frames and closed-captions [10], [12], [13]. Consequently, face annotation has previously been regarded as a problem of finding the correlations between the texts and the image contents. Satoh *et al.* [10] proposed the first approach to associate names with faces in news videos by measuring the frequency of faces and names occurring at the same time. However, without a prior face-name association set, this method may suffer significantly from noise, especially for low-quality images.

Berg *et al.* [11], [26] collected a large number of face images from Yahoo! News channel and labeled them using some language models and clustering methods. Their approach tried to find the correspondences between faces and names in news picture-caption pairs during the clustering procedure. Encouraging results were reported on their dataset with a variety of poses, illuminations, expressions and environmental conditions. One disadvantage of their clustering approaches is that a single identity may become associated with different names in the clusters due to text noise, limiting the retrieval performance.

There is no doubt that textual information can be beneficial for face annotation tasks when it is available. However, in some situations, textual information may not always be available and may be quite noisy in real-world situations. Hence, it is important to study effective ways of exploring the visual information for face annotation tasks. To date, the research community has developed few solutions using only visual information.

In general, face annotation can be regarded as an extended face detection and recognition problem if one is considering only the visual information. Face detection and recognition has already been studied extensively in the past decade [27]. A recent survey can be found in [28].

Recently, several research studies have been proposed to explore visual information for face annotation by applying face recognition techniques. These approaches are often regarded as supervised learning problems. For example, authors in [26] suggested Fisher's linear discriminant analysis for face annotation. In [12], Support Vector Machines (SVM) were employed to train and predict the probabilities of names in the transcript matching faces in the videos. However, due to the high cost of manually labeling the data, supervised learning methods usually suffer from a shortage of labeled data. Recently Yang *et al.* [13] proposed a multiple instance learning approach to alleviate the problem of limited labeled data. In this paper, we suggest addressing this issue by exploring transductive kernel learning techniques.

The key of our proposed transductive learning solution is to incorporate the information of unlabeled data into the annotation tasks. More specifically, in contrast to the linear discriminant in [26], we suggest the Kernel Fisher Discriminant technique that solves the Fisher's linear discriminant in a deformed kernel feature space. Since our algorithm includes information from both labeled and unlabeled data, it is more reliable for building effective classifiers with limited amounts of labeled data than traditional supervised learning techniques. In addition, we develop an effective face detection and alignment scheme to detect the facial regions and extract effective features for face representation from the robust Gabor wavelets features. All of these make our scheme effective in exploring the available visual information for large-scale face annotation.

III. Transductive Kernel Fisher Discriminant

A. Overview

In this section, we propose a Transductive Kernel Fisher Discriminant algorithm for face annotation. We adopt the Kernel Fisher Discriminant as the basis of our method, since it is the state-of-the-art method for traditional face recognition tasks. The main idea of our solution is to transform the supervised KFD approach into a Transductive KFD learning method via kernel transformation techniques. To induce an effective transductive kernel, we propose to employ the kernel deformation principle, which is able to effectively incorporate information from unlabeled data into a new kernel. In the subsequent parts of this paper, we first give our formulation of Kernel Fisher Discriminant and then introduce the kernel deformation principle, which has a solid theoretical basis for learning nonparametric data-dependent kernels. Based on the kernel deformation principle, we finally present our propose TKFD algorithm for solving face annotation tasks.

B. Kernel Fisher Discriminant

Fisher's linear discriminant analysis [21] and its variants [22] are generally regarded as a state-of-the-art method to deal with high-dimensional facial image data [28]. Kernel Fisher Discriminant (KFD) [29]–[31] has been suggested to solve the problem of Fisher's linear discriminant in a kernel feature space, thereby yielding a nonlinear discriminant in the input space. Comparing with other supervised learning methods such as SVMs [32], KFD enjoys the merits of outputs with natural probabilistic interpretations and better solutions for multiclass classification problems.

Let $\{\mathbf{x}_i \,|\, i = 1, \ldots, l\}$ denote the labeled data in the input space and assume the annotation task is an m-class classification problem. Let K be an $l \times l$ kernel matrix whose elements are defined as

$$[K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)]$$

where Φ is a nonlinear mapping function to form the kernel function $k(\cdot, \cdot)$ in the Reproducing Kernel Hilbert Space (RKHS).

Let $X = [\Phi(\mathbf{x}_1)\Phi(\mathbf{x}_2) \cdots \Phi(\mathbf{x}_l)]$ represent the data matrix in the feature space; then the kernel matrix K can be calculated as follows:

$$[K = X^\top X].$$

For a Kernel Fisher Discriminant problem, the total scatter matrix S_t and between-class scatter matrix S_b in the feature space are defined as follows:

$$S_t = \frac{1}{l} X X^\top \tag{1}$$

$$S_b = \frac{1}{l} X W X^\top \tag{2}$$

where the weight matrix W is an $l \times l$ positive symmetric matrix, whose elements are defined as follows:

$$W_{ij} = W_{ji} = \begin{cases} \frac{1}{|\mathcal{C}(\mathbf{x}_i)|}, & \mathcal{C}(\mathbf{x}_i) = \mathcal{C}(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\mathcal{C}(\mathbf{x}_i)$ denotes the class of data instance $\mathbf{x}_i$ and $|\mathcal{C}(\mathbf{x}_i)|$ denotes the total number of data instances in the class of $\mathbf{x}_i$.

Remark: Our definition of the weight matrix W is more general and flexible than conventional block diagonal representation. Moreover, the samples that belong to the same class are no longer required to be kept in order.

Given the above definitions, instead of maximizing the typical Fisher's discriminant criterion $J = \mathbf{tr}(S_w^{-1} S_b)$, where S_w is the within-class scatter matrix, we consider a variant [33] that can deal with small sample size problems in high dimensional input space as follows:

$$\max_{\mathcal{V}} \quad \frac{|\mathcal{V}^\top S_b \mathcal{V}|}{|\mathcal{V}^\top S_t \mathcal{V}|} \tag{4}$$

where $\mathcal{V}$ is a projection matrix.

There are several ways to solve the above optimization problem. One approach is to solve the following equivalent generalized eigen-decomposition problem:

$$\lambda S_t \mathcal{V} = S_b \mathcal{V} \tag{5}$$

and then to form the projection matrix $\mathcal{V}$ by selecting the eigenvectors with maximal eigenvalues λ. From the theory of reproducing kernels, the solution $\mathcal{V}$ lies in the span of $[\Phi(x_1)\Phi(x_2) \cdots \Phi(x_l)]$ in feature space

$$\mathcal{V} = \sum_{i=1}^{l} \mathcal{B}_i \Phi(\mathbf{x}_i) = X\mathcal{B} \tag{6}$$

where $\mathcal{B} = [\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_l]^\top$. Substituting (1), (2), (6) into (5), we obtain

$$\lambda X X^\top X \mathcal{B} = X W X^\top X \mathcal{B}.$$

Multiplying both sides by $X^\top$, we then have

$$\lambda X^\top X X^\top X \mathcal{B} = X^\top X W X^\top X \mathcal{B}.$$

Since $K = X^\top X$, we can turn (5) into the following equivalent form:

$$\lambda K K \mathcal{B} = K W K \mathcal{B}. \tag{7}$$

To ensure numerical stability of matrix inversion, we can add a regularization term in (7). Consequently, the KFD problem becomes one of solving the following equivalent eigen-decomposition problem

$$\lambda \mathcal{B} = (KK + \gamma I)^{-1} (KWK) \mathcal{B} \tag{8}$$

where γ is the regularization parameter, and I is an identity matrix.

Since the purpose of the Kernel Fisher Discriminant is to project input data into the optimal feature space, let $\mathbf{x}$ denote a data example in the input space. We can then project the high-dimensional vector $\Phi(\mathbf{x})$ into a lower dimensional space

$$\begin{aligned} \mathbf{u} = \mathcal{V} \cdot \Phi(\mathbf{x}) &= \sum_{i=1}^{l} \mathcal{B}_i (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)) \\ &= \sum_{i=1}^{l} \mathcal{B}_i k(\mathbf{x}_i, \mathbf{x}). \end{aligned}$$

Let $\mathbf{k}_\mathbf{x} \in R^l$ denote $(k(\mathbf{x}_1, \mathbf{x}) \ldots k(\mathbf{x}_l, \mathbf{x}))^\top$; then, the projected feature vector $\mathbf{u}$ can be represented by the following formula:

$$\mathbf{u} = \mathcal{B}^\top \mathbf{k}_\mathbf{x}. \tag{9}$$

C. Kernel Deformation Principle

In face annotation, conventional supervised learning methods usually require a large number of labeled data to train the model. Previous approaches attempted to solve this problem by multiple instance learning. We tackle this problem by engaging

transductive learning techniques, which can exploit the unlabeled data effectively. The kernel deformation technique [20] provides a framework for learning a data-dependent nonparametric kernel from unlabeled data. It can effectively turn a supervised learning algorithm into transductive or semi-supervised learning settings.

The main idea of the kernel deformation principle is to estimate the geometry of the underlying marginal distribution from unlabeled data, then incorporate them into the kernel deformation procedure. Thus, the resulting new kernel can take advantage of information from unlabeled data. When an input kernel is deformed according to the data distribution, the resulting kernel method may be able to achieve better performance than the original input kernel.

Basically, the kernel deformation technique aims to deform the original RKHS $\mathcal{H}$ into a new RKHS $\tilde{\mathcal{H}}$ that can estimate the underlying marginal distribution of both labeled and unlabeled data. Working with $\tilde{\mathcal{H}}$, the new kernel $\tilde{k}$ is computed explicitly in terms of unlabeled data, and a supervised kernel method can be employed for semi-supervised inference. Given an input kernel k, the new kernel $\tilde{k}$ in $\tilde{\mathcal{H}}$ can be explicitly computed by

$$\tilde{k}(\mathbf{x},\mathbf{y}) = k(\mathbf{x},\mathbf{y}) + \boldsymbol{\kappa}_{\mathbf{y}}^{\top}\mathbf{c}(\mathbf{x})$$

where $\boldsymbol{\kappa}_{\mathbf{y}} = (k(\mathbf{x}_1,\mathbf{y})\ldots k(\mathbf{x}_n,\mathbf{y}))^{\top}$ and the coefficients $\mathbf{c}(\mathbf{x}) = (\mathbf{c}_1(\mathbf{x})\ldots\mathbf{c}_n(\mathbf{x}))^{\top}$ depend on $\mathbf{x}$. Both $\boldsymbol{\kappa}_{\mathbf{y}}$ and $\mathbf{c}$ are n-dimensional vectors, where n is the total number of data, both labeled and unlabeled. Let $G \in R^{n\times n}$ be a symmetric positive semi-definite matrix, as discussed in the next section. Now, $\mathbf{c}(\mathbf{x})$ can be computed as follows:

$$\mathbf{c}(\mathbf{x}) = -(I + G\mathcal{K})^{-1}G\boldsymbol{\kappa}_{\mathbf{x}},$$

where $\mathcal{K} \in R^{n\times n}$ is the kernel matrix with both labeled data and unlabeled data, and $\boldsymbol{\kappa}_{\mathbf{x}}$ is defined as $(k(\mathbf{x}_1,\mathbf{x})\ldots k(\mathbf{x}_n,\mathbf{x}))^{\top}$. Consequently, the explicit form of the new kernel $\tilde{k}$ can be formulated as follows:

$$\tilde{k}(\mathbf{x},\mathbf{y}) = k(\mathbf{x},\mathbf{y}) - \boldsymbol{\kappa}_{\mathbf{y}}^{\top}(I + G\mathcal{K})^{-1}G\boldsymbol{\kappa}_{\mathbf{x}}. \tag{10}$$

D. Transductive Kernel Fisher Discriminant

The idea of our Transductive Kernel Fisher Discriminant approach is to solve the Fisher's discriminant on the new RKHS, which is constructed by warping the structure in exploiting the underlying distribution of the data. To estimate the new RKHS, we consider the kernel deformation method described in Section III-C. By using the deformed kernels, we are able to transform the supervised Kernel Fisher Discriminant into transductive or semi-supervised learning forms.

It can be observed that our KFD formulation separates the kernel matrix K from the label information through the definition of the matrix W in (8). Therefore, only kernel k on labeled data is required to solve the Kernel Fisher Discriminant optimization problem, which is only a portion of the deformed kernel $\tilde{k}$. Moreover, it can be found from (10) that the deformed kernel $\tilde{k}$ works with both labeled and unlabeled data in the new RKHS. Thus, we replace K with the corresponding part in the deformed kernel $\tilde{K}$, which conveys the information of the unlabeled data. Therefore, by applying similar methodology to that used in the supervised Kernel Fisher Discriminant, we can solve the problem more effectively than supervised approaches by taking advantage of the unlabeled data. Note that (10) can be used to compute either the semi-supervised kernel or the transductive kernel. For the proposed Transductive Kernel Fisher Discriminant approach, the new deformed kernel matrix $\tilde{K} \in R^{n\times n}$ can be derived as

$$\tilde{K} = \mathcal{K} - \mathcal{K}(I + G\mathcal{K})^{-1}G\mathcal{K}. \tag{11}$$

It can be simplified through the Kailath Variant

$$\tilde{K} = (I + \mathcal{K}G)^{-1}\mathcal{K}.$$

Moreover, the above equation is equal to

$$\tilde{K} = \mathcal{K}(I + G\mathcal{K})^{-1}. \tag{12}$$

It is interesting to note that the representation in (12) is more concise and computationally more efficient than the original one in (11).

There are several choices for the symmetric positive semi-definite matrix G. As suggested in [20], the graph Laplacian method is used in this work. G is defined by L^p, where L is the Laplacian matrix of a graph and p is a degree parameter. The graph Laplacian is defined as $L = D - Q$, where

$$Q_{ij} = Q_{ji} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}}, & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$

and D is a diagonal matrix where $D_{ii} = \sum_i Q_{ij}$.

Then, we employ the deformed kernel to find the optimal projection in the new RKHS $\tilde{H}$. Let $\tilde{K}_{\mathrm{tr}} \in R^{l\times l}$ denote the matrix part of the "training-data block" in the deformed kernel matrix $\tilde{K}$; substituting it into (8)

$$\lambda\tilde{\mathcal{B}} = (\tilde{K}_{\mathrm{tr}}\tilde{K}_{\mathrm{tr}} + \gamma I)^{-1}(\tilde{K}_{\mathrm{tr}}W\tilde{K}_{\mathrm{tr}})\tilde{\mathcal{B}}.$$

Therefore, the feature vector projected from the new RKHS is derived as

$$\tilde{\mathbf{u}} = \tilde{\mathcal{B}}^{\top}\tilde{\mathbf{k}}_{\boldsymbol{x}} \tag{13}$$

where $\tilde{\mathbf{k}}_{\boldsymbol{x}} = (\tilde{k}(\mathbf{x}_1,\mathbf{x})\ldots\tilde{k}(\mathbf{x}_l,\mathbf{x}))^{\top}$. The complete TKFD algorithm is summarized in Fig. 1.

After feature vectors are extracted by this TKFD algorithm, the next step is to measure the similarity for nearest neighbor (NN) classification. The NN classifier is a nonparametric classification method, which works by finding the neighbor with the minimum distance between the query instance $\mathbf{u}$ and all labeled data instances. The query instance $\mathbf{u}$ will be classified into the class of the closest labeled instance. Since the cosine similarity $\triangle_{\cos}$ yields better results in the empirical evaluation, it is selected as the distance measure for the NN classification schemes using Kernel Fisher Discriminant, and is also used for the proposed Transductive Kernel Fisher Discriminant

$$\triangle_{\cos} = \frac{\mathbf{u}^{\top}\mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|} \tag{14}$$

Algorithm 1 Transductive Kernel Fisher Discriminant

Input

- X: input data
- k: input kernel function
- γ: regularization parameter

Output

- $\tilde{K}$: transductive kernel matrix
- $\tilde{\mathcal{B}}$: projection matrix

Procedure

1: Calculate initial kernel matrix $\mathcal{K}$: $\mathcal{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$

2: Calculate transductive kernel matrix $\tilde{K}$:

$$\tilde{K} = \mathcal{K}(I + G\mathcal{K})^{-1}$$

3: Calculate weight matrix W:

$$W_{ij} = \begin{cases} \frac{1}{|\mathcal{C}(\mathbf{x}_i)|}, & \mathcal{C}(\mathbf{x}_i) = \mathcal{C}(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases}$$

4: Find $\tilde{\mathcal{B}}$ by solving the following eigen-decomposition:

$$\lambda\tilde{\mathcal{B}} = (\tilde{K}_{tr}\tilde{K}_{tr} + \gamma I)^{-1}(\tilde{K}_{tr}W\tilde{K}_{tr})\tilde{\mathcal{B}}$$

5: Return ($\tilde{K}$, $\tilde{\mathcal{B}}$)

End

Fig. 1. Transductive kernel Fisher discriminant algorithm.

where $\mathbf{u}$ and $\mathbf{v}$ are the extracted feature vectors. Note that the methodology discussed above can be applied to solve other general multiclass classification problems. In this paper, however, we restrict its application to face annotation tasks.

IV. Experimental Results

A. Overview

In this section, we report empirical evaluations of the Transductive Kernel Fisher Discriminant algorithm with applications to face annotation tasks. To make evaluations comprehensive, we have collected three different kinds of datasets as our experimental testbeds. One is the Face Recognition Grand Challenge (FRGC) dataset [23], which was originally designed for benchmark evaluation of face recognition. The second dataset is the Yahoo! News facial images dataset, which was derived from the web [11]. The third facial image dataset is selected from the TRECVID 2005 dataset, which was originally used for benchmarking video retrieval tasks.

For performance comparison, we also implement four approaches for face annotations, i.e., Linear Discriminant Analysis (LDA), Kernel Fisher Discriminant (KFD), Support Vector Machine (SVM), and Transductive SVM. Both LDA and KFD are two typical methods for face recognition tasks. For a performance metric, average accuracy of annotation results is used for the evaluations. Precision and recall curves are also provided for comparisons. In the following text, we first show the details of our testbeds. Then we discuss our preprocessing approaches for face extraction and feature representation. Finally, we present and discuss our experimental results.

Fig. 2. FRGC image examples with controlled and uncontrolled environment. The cropped faces are placed to the right side of each original image. Each cropped image is interpolated to the size of 128 × 128.

Fig. 3. Yahoo! News face images used in our experiments. The cropped faces are placed to the right side of each original image. Each cropped image is interpolated to the size of 128 × 128.

B. Experimental Testbeds

1) FRGC Dataset: The FRGC dataset [23][1] is the state-of-the-art benchmark protocol for performance evaluation of face recognition techniques. We adopt the FRGC version 1 data set (Spring 2003) in evaluating our face annotation algorithms. This dataset contains 5660 images of either 1704 × 2272 pixels or 1200 × 1600 pixels. Since we consider the face annotation task rather than biometric identification, the standard FRGC experimental protocol is not directly applied for performance evaluation. The dataset used in our experiment consists of 1920 images, corresponding to 80 individuals selected from the original collection. Each individual has 24 controlled or uncontrolled color images. The faces are automatically detected and normalized through a face detection and extraction method, which will be detailed in Section IV-C. Fig. 2 shows geometrically normalized face images cropped from the original FRGC images, with the cropped regions resized to the size of 128 × 128. Moreover, some image processing operations are performed on these face images, such as histogram equalization, lighting correction, etc.

2) Yahoo! News Face Dataset: The Yahoo! News Face dataset was constructed by Berg *et al.* [11] from about half a million captioned news images collected from the Yahoo! News web site. It consists of large number of photographs taken in real life conditions, rather than in the controlled environments widely used in face recognition evaluation. As a result, there are a large variety of poses, illuminations, expressions and environmental conditions. After applying a face detection algorithm and processing the resulting faces, there is a total of 31 586 large well detected faces available for clustering. Each image in this set is associated with a set of names. Discarding face clusters with a small number of elements, a subset of 1248 face

[1]Accessible from http://www.frvt.org/FRGC

Fig. 4. TRECVID video dataset used in our experiments. The cropped faces are placed to right side of each original image.

clusters is obtained. In addition, there are several individuals having more than one cluster each; we merged them so that one individual corresponds to one cluster. 1940 images, corresponding to the 97 largest face clusters, are selected to form our experimental dataset, in which each individual has 20 images. As with the FRGC dataset, faces are cropped from selected images using the same face detection and extraction method in Section IV-C. Only the relevant face image is retained when there are multiple faces in one image. Fig. 3 presents examples selected Yahoo! News images and the extracted faces. All these faces are geometrically normalized.

3) TRECVID Video Data: The third dataset used in our experimental testbeds is from the TREC Video Retrieval (TRECVID) 2005 dataset [24]. The original dataset contains 277 broadcast news videos of 171 h from six channels in three languages (English, Chinese, and Arabic). The original dataset is designed for benchmarking video retrieval tasks. In our experiment, we extract the facial regions from the key frames in the original dataset. Among the detected faces, we select 31 individuals to form our face annotation dataset, which contains 867 face images in total. Fig. 4 shows examples of video frames and the extracted faces.

C. Facial Image Detection and Extraction

The major task of facial image extraction is to locate and crop the face region from the input image, then to normalize the cropped image geometrically and photometrically. In order to enable an automatic face annotation scheme, we cascade a state-of-the-art face detector [34] with Active Appearance Models (AAMs) [35], [36] to locate faces and facial features in the input images. More specifically, we employ the face detector roughly locate the facial region which is employed to initialize AAMs fitting. The image is aligned to the predefined template using the estimated centers of the eyes provided by the AAMs facial feature locator. Finally, the facial region without hair is cropped from the original image. If there are multiple faces in an image, we iteratively extract each face from those images. Fig. 5 shows two sample resulting images with two faces using the face extractor employed in this study. Note that any false detection by the face detector can be inspected by thresholding the AAMs fitting error. Figs. 2 and 3 depict some cropped sample faces using our proposed facial image extractor. The performance in terms of correct registration is greatly dependent on the image conditions. In fact, the proposed method successfully crops 99% images on the FRGC dataset. Similarly, the correct registration rate is around 80% for the Yahoo! News Face dataset, and around 85% for the TRECVID dataset.

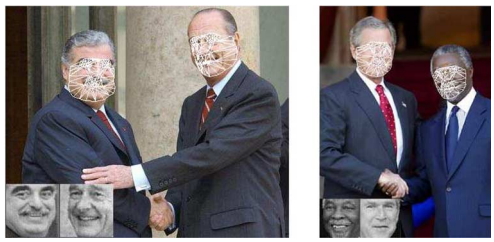


Fig. 5. AAMs fitting result on two sample images.

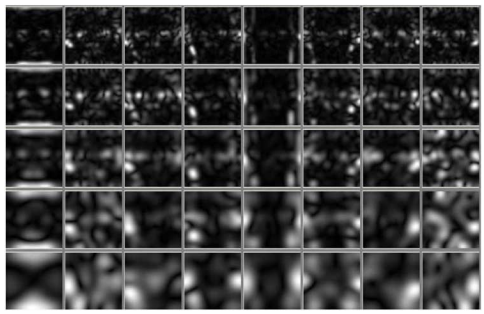


Fig. 6. Face image represented by 40 subimages of the magnitude part of the Gabor wavelet transform.

D. Feature Representation

Once facial images are extracted, the next step is to extract features and then represent them effectively in classification tasks. The feature representation techniques have been extensively studied for face detection and recognition in recent years. Many effective feature extraction methods have been proposed to address the task, such as Local Binary Pattern [37] and Gabor Wavelets Transform. Among those methods, Gabor wavelets representation of facial image has been widely accepted as a successful approach [22]. From past studies in the area of signal processing, Lades *et al.* [38] empirically surmised that good performance can be achieved by extracting Gabor wavelet features of five different scales and eight orientations. In our experiments, we employ a similar approach by applying Gabor wavelet transform on each image (scaled to 128×128) at five scales and eight orientations. Fig. 6 shows an example of 40 resulting subimages after Gabor wavelet transformation. Finally, we normalize each subimage to form a feature vector $\mathbf{x} \in \mathbf{R}^n$ with the sample scale reduced to 64, which results in a 10 240-dimensional feature vector for each facial image.

E. Experimental Settings and Implementation Details

In our experiments, three datasets are used for performance evaluations. Table I summarizes the details of these testbeds. In

TABLE I
FACE IMAGE DATASETS USED IN THE EXPERIMENTS

Dataset	# total images	# classes	# images per class
FRGC	1920	80	24
Yahoo! News	1940	97	20
TRECVID 2005	867	31	11∼111

the experimental evaluations, each dataset is partitioned into a labeled set and an unlabeled set. For each transductive learning setting, the training set comprises $l + u$ data examples for each class, where l is the number of labeled data and u is the number of unlabeled data for the class. For each supervised learning setting, the training set only considers l labeled examples for each class.

The LDA algorithm is used as the baseline method for evaluating the performance of the proposed face annotation approach. The implemented baseline method [2] is similar to the Fisherfaces method [21], which applies LDA after PCA dimensionality reduction. We also implement SVM and Transductive SVM (TSVM) for comparison. As mentioned in [39], finding an exact optimal solution for TSVM is NP-hard; a great deal of research effort has been devoted to the approximation algorithm. We consider the LapSVM [20] as the reference TSVM method, which has demonstrated better performance than the other popular approaches, such as the TSVM in $\mathrm{SVM}^{\mathrm{light}}$ [40] and the Low Density Separation (LDS) [41].

It is worth noting that LDA is a feature extraction method rather than a classifier itself. It is often followed by some simple classifiers, such as k-NN, to solve the pattern classification problems. Other sophisticated classification techniques can also be engaged as the classifiers on the extracted features. Similarly, the extracted features by KFD and TKFD could also be used by other kernel-based classifiers.

We set up the following experimental protocol for all tests. The number of labeled examples of each class, l, is gradually increased from 1 to 7, and the rest examples are considered as the unlabeled data. A variation of the tenfold cross validation approach is performed in the experiments. For each evaluation round, the labeled data are randomly selected. We use the same kernel and regularization parameters for both KFD and TKFD. The linear kernel is used for all the experiments. The regularization parameter $\lambda = 0.001$ is fixed for all experiments to enable an objective comparison and reduce the complexity in choosing model parameters. For SVM and TSVM, the linear kernel is also used in the experiment, and the regularization parameter C is set to 100. For TSVM and TKFD, the Laplacian graph is constructed based on the Euclidean nearest neighborhood. For selecting the eigenspaces of KFD and TKFD, we choose the eigenvectors corresponding to 98% of the total variations.

In our experiments, all the compared methods were implemented in Matlab and evaluated on a PC with a 3.0 GHz single processor and 2 GB memory.

[2]A regularization term is added into the LDA optimization $(S_b + \gamma I)^{-1} S_w$ in order to ensure numerical stability, where $\gamma = 0.001$. In addition, Euclidean distance is employed as the similarity measurement.

F. Experiment-I: Evaluation on FRGC Face Dataset

Table II presents the experimental results of different settings. From the experimental results, we first observe that all the kernel methods, SVM, TSVM, KFD and TKFD, outperform the baseline LDA method significantly on different features. For example, when the number of labeled example is equal to 5, the LDA method only achieved overall 43.9% accuracy on intensity and 71.3% on Gabor features, while four other kernel methods are able to achieve significantly better performance. These results show that kernel techniques are generally much powerful than the linear ones in face annotation tasks. Second, we can observe that the Gabor features are more efficient than the intensity features in all cases. Further, comparing the two kernel methods, KFD and TKFD, we found that the proposed TKFD method performs better than the supervised KFD method given the same number of labeled data in most cases. The improvements are particularly significant when there are smaller numbers of labeled examples. More impressively, for the cases with one and two labeled examples, the TKFD method is able to outperform the supervised KFD method with 96% and 15% improvements, respectively. Finally, the proposed TKFD approach outperforms TSVM in most cases except for the single example case. This is because there is no intra-class information for KFD and TKFD methods when there is only one sample in a class. Therefore, the KFD and TKFD may not work effectively in this case.

In order to look into the details of the empirical comparison, we also plot the precision-recall curve of the annotation results in Fig. 7. These experimental results show that the TKFD approach consistently outperforms the supervised KFD and SVM methods. In contrast to the other state-of-the-art semi-supervised method, TKFD is comparable to TSVM in the performance of retrieval precision, and better than TSVM in the performance of retrieval recall. This verifies that our proposed TKFD algorithm is effective to improving traditional supervised KFD methods over the challenge of insufficient training samples.

G. Experiment-II: Evaluation on Yahoo! News Image Dataset

Using the Yahoo! News image dataset, we conduct evaluations similar to the FRGC approach. Table III shows the experimental results of overall annotation accuracy. From the results, we found that the annotation task on Web images is more challenging than the FRGC faces. Specifically, given the setting of 5 labeled examples per class, the TKFD method achieved only 53.9% average accuracy on the Yahoo! News image dataset of 96 classes, while it achieved 82.4% average accuracy on the FRGC dataset of 80 classes. This is because the images in the FRGC dataset are usually taken in some controlled environment, while the images collected in the Yahoo! News image dataset have more variants of different lighting conditions and orientations. Looking into the performance comparison, we also found the two kernel methods are considerably better than the LDA method in most cases, and the TKFD method outperforms the supervised KFD in most cases. For the cases with one and two labeled examples per class, the TKFD method is able to respectively outperform the KFD method by 77.5% and 10.0%. To examine the retrieval performance of precision and recall, we also plot the corresponding curves in Fig. 8, in which the proposed

TABLE II
AVERAGE ACCURACY OF ANNOTATION PERFORMANCE ON THE FRGC DATASET (%)

Label	Intensity					Gabor				
Size	LDA	SVM	TSVM	KFD	TKFD	LDA	SVM	TSVM	KFD	TKFD
1	13.3 ± 4.8	12.5 ± 1.3	21.3 ± 1.8	12.2 ± 1.4	**22.5** ± 1.5	16.5 ± 8.7	33.2 ± 1.5	**43.4 ± 2.0**	18.4 ± 1.8	36.1 ± 1.5
2	20.3 ± 6.6	36.0 ± 1.7	39.4 ± 2.0	36.5 ± 1.1	**40.4** ± 1.3	34.0 ± 11.3	49.7 ± 1.5	60.4 ± 1.7	53.1 ± 2.2	**60.9 ± 1.3**
3	34.1 ± 2.8	46.7 ± 2.8	49.2 ± 2.5	47.3 ± 1.4	**49.3** ± 1.7	53.9 ± 6.1	62.2 ± 1.9	71.1 ± 2.0	69.0 ± 1.5	**72.4 ± 1.1**
4	36.8 ± 3.2	53.2 ± 2.2	**55.3 ± 2.0**	52.9 ± 2.1	55.2 ± 1.8	62.8 ± 7.2	71.6 ± 1.0	77.9 ± 0.9	77.7 ± 1.4	**79.6 ± 1.2**
5	43.9 ± 4.0	57.2 ± 1.9	59.0 ± 1.9	59.0 ± 1.9	**59.4** ± 1.6	71.3 ± 3.1	75.7 ± 1.7	80.8 ± 1.4	81.7 ± 1.2	**82.4 ± 1.3**
6	45.6 ± 4.9	60.7 ± 2.0	62.6 ± 1.8	62.3 ± 1.1	**63.8** ± 1.0	73.4 ± 3.4	80.0 ± 1.2	83.1 ± 1.1	84.3 ± 1.0	**84.5 ± 0.9**
7	50.5 ± 3.3	63.0 ± 1.5	64.3 ± 1.8	64.4 ± 1.9	**65.1** ± 1.8	78.3 ± 0.9	83.0 ± 1.1	85.6 ± 1.4	87.0 ± 1.0	**87.3 ± 1.0**

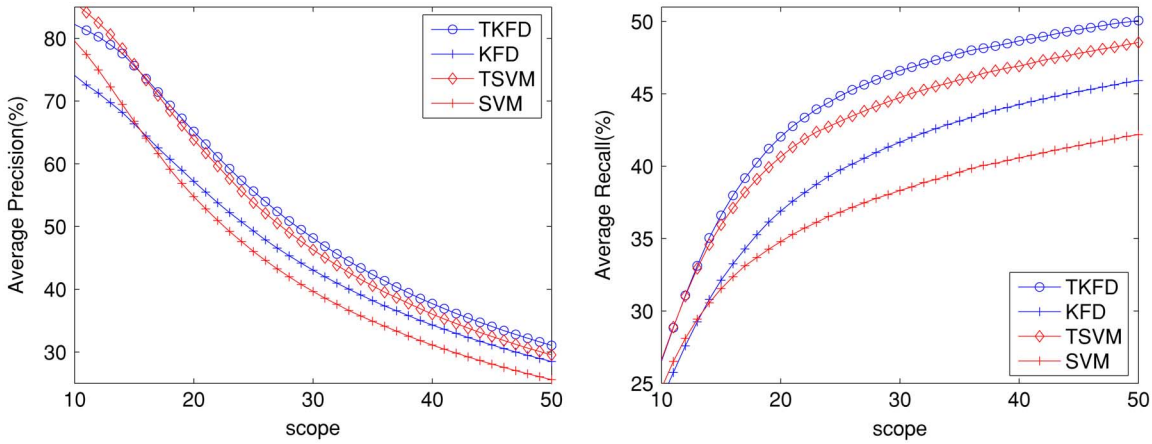


Fig. 7. Precision-Recall curves of annotation results with two labeled examples per person on the FRGC dataset. (a) Average Precision. (b) Average Recall.

TABLE III
AVERAGE ACCURACY OF ANNOTATION PERFORMANCE ON THE YAHOO! NEWS FACE DATASET (%)

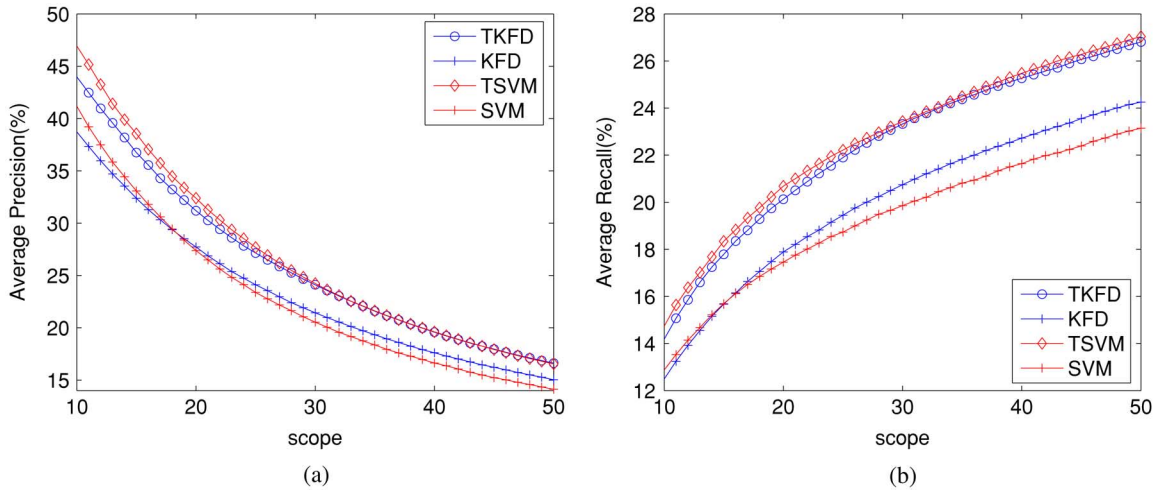
Label	Intensity					Gabor				
Size	LDA	SVM	TSVM	KFD	TKFD	LDA	SVM	TSVM	KFD	TKFD
1	8.3 ± 1.1	7.2 ± 0.9	9.6 ± 1.1	7.2 ± 1.1	**10.3 ± 1.2**	11.5 ± 0.9	18.1 ± 1.6	**23.1 ± 1.9**	10.7 ± 1.0	19.0 ± 1.2
2	13.0 ± 1.5	17.6 ± 0.6	19.2 ± 0.6	17.8 ± 1.2	**19.4 ± 1.0**	24.4 ± 0.7	28.8 ± 1.2	**34.0 ± 1.1**	29.9 ± 1.2	32.9 ± 1.6
3	16.1 ± 1.2	22.4 ± 1.4	**24.2 ± 1.6**	21.6 ± 1.2	23.1 ± 1.2	32.0 ± 2.0	35.3 ± 1.3	41.2 ± 1.1	41.1 ± 1.4	**42.4 ± 1.3**
4	18.0 ± 2.3	24.8 ± 1.1	26.5 ± 1.4	25.2 ± 1.3	**26.7 ± 1.3**	36.8 ± 1.9	40.4 ± 1.2	47.0 ± 1.4	47.5 ± 1.1	**48.2 ± 1.4**
5	19.8 ± 1.7	26.8 ± 1.2	28.1 ± 1.1	27.2 ± 1.4	**28.2 ± 1.1**	41.5 ± 1.6	44.8 ± 1.1	51.1 ± 1.8	53.8 ± 1.3	**53.9 ± 1.2**
6	21.2 ± 1.7	30.1 ± 1.2	30.7 ± 1.4	29.7 ± 1.5	**31.1 ± 1.3**	45.2 ± 1.3	48.2 ± 1.0	54.9 ± 0.9	57.2 ± 1.5	**57.3 ± 1.2**
7	22.8 ± 2.0	30.4 ± 0.7	31.5 ± 0.9	31.4 ± 1.5	**31.6 ± 1.3**	48.4 ± 1.0	50.9 ± 0.9	57.9 ± 1.4	60.5 ± 0.9	**61.6 ± 1.2**

TKFD is significantly better than the supervised methods, SVM and KFD, and is slightly better than the other semi-supervised method, i.e., TSVMs.

H. Experiment-III: Evaluation on Trecvid Video Dataset

The final experimental evaluation is on the TRECVID 2005 video dataset. Similar evaluations are conducted. Table IV and Fig. 9 show the experimental results. From the empirical results, we can see that the overall annotation performance is rather promising in this dataset. Specifically, for the case of five labeled examples per class, the TKFD method achieves an average accuracy of 83.0%, which is better than the results achieved on the other two datasets. One reason is because the number of classes used in the TRECVID dataset is smaller than with the other two datasets. Thus, the annotation task becomes relatively easy. In comparison with other annotation methods, we found similar results to those observed on the previous datasets. Precision-recall curves of annotation results with two labeled examples each person on the TRECVID 2005 dataset

Based on the promising empirical results on the three datasets, we can conclude that our proposed TKFD algorithm

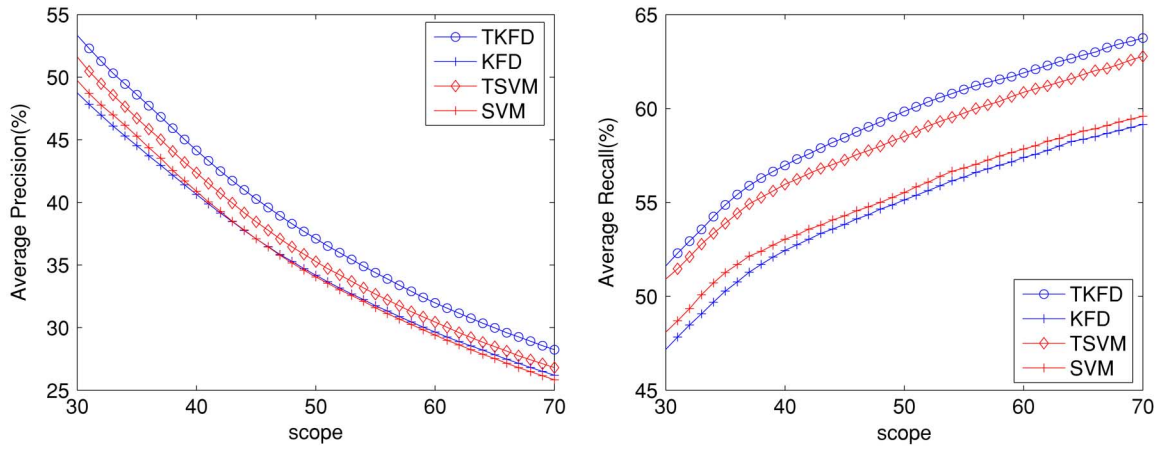


(a) (b)

Fig. 8. Precision-recall curves of annotation results with 2 labeled examples each person on the Yahoo! News photo dataset. (a) Average Precision. (b) Average Recall.

TABLE IV
AVERAGE ACCURACY OF ANNOTATION PERFORMANCE ON THE TRECVID 2005 FACE DATASET (%)

Label	Intensity					Gabor				
Size	LDA	SVM	TSVM	KFD	TKFD	LDA	SVM	TSVM	KFD	TKFD
1	35.8 ± 3.4	41.3 ± 2.3	**44.2 ± 2.7**	34.3 ± 2.9	40.0 ± 3.4	46.3 ± 1.3	51.3 ± 2.9	**55.0 ± 4.4**	45.9 ± 3.7	54.1 ± 4.0
2	48.8 ± 3.7	51.2 ± 2.7	54.0 ± 3.2	49.8 ± 4.1	**55.4 ± 2.6**	60.7 ± 1.9	59.1 ± 1.1	65.6 ± 3.8	61.4 ± 2.2	**68.2 ± 2.5**
3	56.8 ± 2.8	57.3 ± 2.3	60.0 ± 2.4	56.7 ± 2.7	**60.1 ± 2.7**	68.7 ± 3.3	68.6 ± 3.4	**74.8 ± 3.5**	69.6 ± 2.5	74.4 ± 2.2
4	64.5 ± 2.3	61.6 ± 4.2	63.6 ± 3.9	64.0 ± 0.8	**68.3 ± 0.8**	74.0 ± 3.0	73.7 ± 2.1	76.7 ± 2.3	74.6 ± 2.4	**79.2 ± 2.0**
5	65.0 ± 1.6	64.6 ± 2.3	66.9 ± 2.2	66.5 ± 3.0	**69.7 ± 3.3**	79.4 ± 1.7	77.6 ± 2.0	80.4 ± 1.4	79.5 ± 2.3	**83.0 ± 1.8**
6	68.6 ± 2.3	67.7 ± 2.0	69.7 ± 2.4	67.9 ± 2.7	**71.4 ± 2.1**	80.1 ± 2.2	81.1 ± 2.2	82.3 ± 1.7	81.4 ± 2.4	**84.6 ± 2.0**
7	70.7 ± 1.8	70.0 ± 3.1	72.4 ± 3.3	70.5 ± 2.3	**73.1 ± 2.2**	83.7 ± 1.7	83.5 ± 1.8	84.3 ± 1.7	84.3 ± 1.4	**87.1 ± 0.9**



(a) (b)

Fig. 9. Precision-recall curves of annotation result with two labeled examples each person on the TRECVID 2005 dataset.

is more effective than the traditional supervised KFD and SVM methods for face annotation when dealing with a small number of labeled examples, which is a critical advantage for large-scale face annotation applications.

V. DISCUSSIONS AND FUTURE WORK

We have proposed a comprehensive scheme for face annotation by a novel Transductive Kernel Fisher Analysis algorithm. Although the promising experimental results validated the effectiveness of our methodology, we should address limitations and future directions to improve our current approach. First of all, we focused our attention only on exploring the visual information for the face annotation task. In future work, we can combine other annotation approaches studied in the textual domain [10], [25] for improving the annotation performance if the textual information is available. Second, we employed

the kernel deformation principle for learning transductive kernels in the TKFD algorithm. In future work, we can extend the TKFD algorithm to other kernel learning techniques [42]–[44]. For example, we may consider the kernel alignment techniques for combining multiple input kernels instead of using only a single input kernel as in the current solution [42]. We may also study spectral kernel learning techniques to achieve better transductive kernels for annotation tasks [43]. Finally, to minimize the human effort of labeling training data, we can study active learning techniques to provide users the most informative examples for labeling during the annotation tasks [45]–[48].

VI. Conclusion

In this paper, we proposed a novel transductive learning algorithm for face annotation. In contrast to traditional approaches using supervised learning methods, we proposed the Transductive Kernel Fisher Discriminant (TKFD) algorithm, which employs the kernel deformation techniques to exploit both labeled and unlabeled data effectively for annotation tasks. The TKFD algorithm is more effective than traditional supervised annotation methods with a small set of training data, since it can take advantage of information from unlabeled data. To apply the TKFD to face annotation tasks effectively, we developed a comprehensive face annotation scheme using state-of-the-art face detection and feature extraction techniques. We conducted extensive evaluations on three kinds of testbeds. The promising experimental results showed that our method is more effective than conventional approaches, especially for dealing with the cases having only a limited amount of labeled data, which is critical for large-scale face annotation tasks.

References

[1] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Proc. Eur. Conf. Comput. Vision*, 2002, pp. 97–112.

[2] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," in *Proc. 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, 2003, pp. 119–126.

[3] D. Blei and M. I. Jordan, "Modeling annotated data," in *Proc. 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, 2003, pp. 127–134.

[4] V. Lavrenko, R. Manmatha, and J. Jeon, "A model for learning the semantics of pictures," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS'03)*, 2003.

[5] F. Kang and R. Jin, "Symmetric statistical translation models for automatic image annotation," in *Proc. 2005 SIAM Conf. Data Min. (SDM 2005)*, Newport Beach, CA, 2005.

[6] K.-S. Goh and E. Y. Chang, "One, two class svms for multiclass image annotation," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 10, 2005.

[7] R. Brunelli and O. Mich, "Image retrieval by examples," *IEEE Trans. Multimedia*, vol. 2, no. 3, pp. 164–171, Sep. 2000.

[8] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[9] Y. Lu, H. Zhang, L. Wenyin, and C. Hu, "Joint semantics and feature based image retrieval using relevance feedback," *IEEE Trans. Multimedia*, vol. 5, no. 3, pp. 339–347, Sep. 2003.

[10] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and detecting faces in news videos," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 22–35, Jan. 1999.

[11] T. L. Berg, A. C. Berg, J. Edwards, and D. Forsyth, "Who's in the picture," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005.

[12] J. Yang and A. G. Hauptmann, "Naming every individual in news video monologues," in *Proc. 12th Annu. ACM Int. Conf. Multimedia*, New York, 2004, pp. 580–587.

[13] J. Yang, R. Yan, and A. G. Hauptmann, "Multiple instance learning for labeling faces in broadcasting news video," in *Proc. 13th Annu. ACM Int. Conf. Multimedia*, New York, 2005, pp. 31–40.

[14] L. Zhang, L. Chen, M. Li, and H. Zhang, "Automated annotation of human faces in family albums," in *Proc. ACM Multimedia Conf.*, 2003.

[15] S. C. Hoi and M. R. Lyu, "A novel log-based relevance feedback technique in content-based image retrieval," in *Proc. ACM Int. Conf. Multimedia (MM2004)*, New York, Oct. 10–16, 2004.

[16] S. C. Hoi, M. R. Lyu, and R. Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 4, pp. 509–524, Apr. 2006.

[17] P. Over, W. Kraaij, and A. F. Smeaton, "Trecvid 2005 an overview," in *Proc. TRECVID Workshop*, 2005.

[18] N. Vasconcelos and A. Lippman, "A multiresolution manifold distance for invariant image similarity," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 127–142, Feb. 2005.

[19] T. Joachims, "Transductive inference for text classification using support vector machines," in *Int. Conf. Mach. Learning (ICML)*, CA, USA, 1999, pp. 200–209.

[20] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proc. 22nd Int. Conf. Mach. Learning (ICML'05)*, New York, 2005, pp. 824–831.

[21] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[22] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[23] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, *Overview Face Recognition Grand Challenge*, vol. 1, pp. 947–954, 2005.

[24] TREC Video Retrieval Evaluation [Online]. Available: http://www-nlpir.nist.gov/projects/trecvid/ TRECVID

[25] R. Houghton, "Named faces: Putting names to faces," *IEEE Intell. Syst.*, vol. 14, no. 5, pp. 45–50, Sep. 1999.

[26] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y. W. Teh, E. G. Learned-Miller, and D. A. Forsyth, "Names and faces in the news," *CVPR (2)*, pp. 848–854, 2004.

[27] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1090–1104, 2000.

[28] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[29] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller, "Fisher discriminant analysis with kernels," in *Proc. IEEE NN For Signal Process. Workshop*, 1999, pp. 41–48.

[30] S. Mika, G. Rätsch, and K.-R. Müller, "A mathematical programming approach to the kernel Fisher algorithm," in *Advances in Neural Information Processing Systems 13*. Cambridge, MA: MIT Press, 2001, pp. 591–597.

[31] Q. Liu, H. Lu, and S. Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Trans.Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004.

[32] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, pp. 61–74.

[33] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1990.

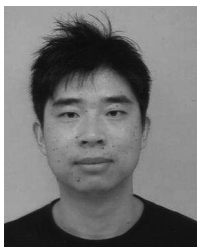
[34] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May. 2004.

[35] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.

[36] J. Zhu, S. C. Hoi, and M. R. Lyu, "Real-time non-rigid shape recovery via active appearance models for augmented reality," in *Proc. 9th Eur. Conf. Comput. Vision (ECCV2006)*, Graz, Austria, May 7–13, 2006, pp. 86–197.

[37] T. Ahonen, A. Hadid, and M. Pietikainen, *Face Recognition Local Binary Patterns*, vol. 1, pp. 469–481, 2004.

[38] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300–311, Mar. 1993.
[39] X. Zhu, Semi-Supervised learning literature survey Semi-supervised learning literature survey, Tech. Rep., 2005 [Online]. Available: http://www.cs.wisc.edu/jerryzhu/pub/ssl_survey.pdf
[40] T. Joachims, "Transductive inference for text classification using support vector machines," in *ICML'99: Proc. 16th Int. Conf. Mach. Learning*, San Francisco, CA, 1999, pp. 200–209.
[41] O. Chapelle and A. Zien, *Semi-Supervised Classification Low Density Separation*, pp. 57–64, 2005 [Online]. Available: http://www.kyb.tuebingen.mpg.de/bs/people/chapelle/lds/
[42] G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. Jordan, "Learning the kernel matrix with semi-definite programming," *J. MLR*, vol. 5, pp. 27–72, 2004.
[43] S. C. Hoi, M. R. Lyu, and E. Y. Chang, "Learning the unified kernel machines for classification," in *Proc. KDD2006*, 2006, pp. 187–196.
[44] T. Zhang and R. K. Ando, "Analysis of spectral kernel design based semi-supervised learning," in *Proc. NIPS*, 2005.
[45] S. C. Hoi, R. Jin, and M. R. Lyu, "Large-scale text categorization by batch mode active learning," in *Proc. WWW2006*, 2006, pp. 633–642.
[46] R. Jin, J. Y. Chai, and L. Si, "Effective automatic image annotation via a coherent language model and active learning," in *Proc. 12th Annu. ACM Int. Conf. Multimedia*, New York, 2004, pp. 892–899.
[47] S. C. H. Hoi and M. R. Lyu, "A semi-supervised active learning framework for image retrieval," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognition (CVPR'05)*, 2005, pp. 302–309.
[48] S. C. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma, "Learning distance metrics with contextual constraints for image retrieval," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR2006)*, New York, Jun. 17–22, 2006, pp. 2072–2078.



Jianke Zhu received the B.S. degree in mechatronics and computer engineering from the Beijing University of Chemical Technology, Beijing, China, in 2001. He received the M.S. degree in electrical and electronics engineering from University of Macau in 2005. He is currently pursuing the Ph.D. degree in the Computer Science and Engineering Department, Chinese University of Hong Kong.

His research interests are in pattern recognition, computer vision, and statistical machine learning.

Steven C. H. Hoi received the B.S. degree in computer science from Tsinghua University, Beijing, China, and the M.S. and and Ph.D. degrees in computer science and engineering from the Chinese University of Hong Kong, Hong Kong, China.

He is currently an Assistant Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests are in multimedia information retrieval, statistical machine learning, and data mining.

Michael R. Lyu received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1981; the M.S. degree in computer engineering from University of California, Santa Barbara, in 1985; and the Ph.D. degree in computer science from the University of California, Los Angeles, in 1988.

He is currently a Professor in the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong, China. He is also Director of the Video over Internet and Wireless (VIEW) Technologies Laboratory. He was with the Jet Propulsion Laboratory as a Technical Staff Member from 1988 to 1990. From 1990 to 1992, he was with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, as an Assistant Professor. From 1992 to 1995, he was a Member of Technical Staff in the applied research area of Bell Communications Research (Bellcore), Morristown, NJ. From 1995 to 1997, he was a Research Member of Technical Staff at Bell Laboratories, Murray Hill, NJ. His research interests include software reliability engineering, distributed systems, fault-tolerant computing, mobile networks, Web technologies, multimedia information processing, and E-commerce systems. He has published over 270 refereed journal and conference papers in these areas. He has participated in more than 30 industrial projects and helped to develop many commercial systems and software tools. He was the editor of two book volumes: *Software Fault Tolerance* (New York: Wiley, 1995) and *The Handbook of Software Reliability Engineering* (New York: IEEE and New McGraw-Hill, 1996).

Dr. Lyu received Best Paper Awards at ISSRE'98 and ISSRE'2003. Dr. Lyu initiated the First International Symposium on Software Reliability Engineering (ISSRE) in 1990. He was the Program Chair for ISSRE'96 and General Chair for ISSRE'2001. He was also PRDC'99 Program Co-Chair, WWW10 Program Co-Chair, SRDS'2005 Program Co-Chair, PRDC'2005 General Co-Chair, and ICEBE'2007 Program Co-Chair, and served in program committees for many other conferences including HASE, ICECCS, ISIT, FTCS, DSN, ICDSN, EUROMICRO, APSEC, PRDC, PSAM, ICCCN, ISESE, and WI. He has been frequently invited as a keynote or tutorial speaker to conferences and workshops in the U.S., Europe, and Asia. He has been on the Editorial Board of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON RELIABILITY, the *Journal of Information Science and Engineering*, and *Software Testing, Verification & Reliability Journal*. Dr. Lyu is an AAAS Fellow for his contributions to software reliability engineering and software fault tolerance.