Learning Face Recognition Classifiers by Incorporating Human Perceptions

Chu-Song Chen, Shih-Liang Yeh, and Chang-Ming Tsai

TR-IIS-04-006
June 18, 2004

Institute of Information Science
Academia Sinica

Learning Face Recognition Classifiers by Incorporating Human Perceptions

Chu-Song Chen, Shih-Liang Yeh, and Chang-Ming Tsai

Institute of Information Science, Academia Sinica, Nankang 115, Taipei, Taiwan.
song@iis.sinica.edu.tw, slyeh@iis.sinica.edu.tw, tomyt@iis.sinica.edu.tw

Abstract

We propose a novel approach that employs human perception to improve the learning for face-recognition. By considering that the training examples are finite samples on the face manifold, we introduce the concept of appearance morphing curve that contains significant information for identifying facial images of different persons. By employing this concept, an effective method is introduced to increase automatic face recognition performances.

1. Introduction

Many of the existing automatic face recognition (AFR) researches focus on the use of statistics-based learning methods to train a classifier for human face identification [2][4][6]. In the main loop of the statistics-based learning process for AFR, collection of training examples is inevitable but was somewhat overlooked. While the training examples were collected, many existing learning algorithms can be chosen to apply to them, and several classifiers are thus trained. Then, a classifier with the best performance was selected as the face identifier. In the learning process, training examples are essentially fixed once they have been collected (although they have ever possibly been separated into training and validation data sets). In such a typical scenario of classifier syntheses, no further information is available to refine the obtained face identifier's performance as well as its generalization ability. Even if we have found that the generalization performance is not satisfied, we have no way to refine it except new training examples are added.

By carefully doing the above statistics-based learning process, the performance of the face identifier generated can be considerably increased but, up to most recently, is still worse than human's identification ability. To human beings, we can accurately distinguish person identities from their faces. Although the exact algorithm or function run in human brain is still not very clear to our best knowledge, we can treat it as a black box and record its input-output relationship. Based on this observation, a central issue being addressed in this paper is that, how human perception abilities can be incorporated for supplementing an AFR task?

Note that in AFR, once a face identifier is synthesized, the identifier shall not refer any human opinion for identity verifications, otherwise the recognition process is not truly automatic. However, in AFR it is allowable to exploit human perception before a face identifier has been synthesized. By doing so, a better face identifier is expected to being constructed.

In this paper, we introduce a systematic procedure to incorporate human perception in the learning process of AFR. More specifically, we will first examine the generalization abilities of face identifiers. We propose a new criterion for evaluating the generalization ability of a classifier from human perspective. Then, by using the criterion, new classifiers with better generalization abilities can be learned. Note that in our method, no further training examples have to be collected and the learning process remains to base on a fixed collection of training examples. Nevertheless, we do increase the number of training examples by composing new prototypes from the collected ones and will show that human perception can help further in refining the synthesized identifiers even a fixed set of training examples is used.

2. Appearance Morphing Curves

Think about a set $\prod$ consisting of all possible appearances of human faces; or more frankly speaking, all possible images of cropped single faces. In this paper, we only consider frontal or near-frontal faces, but our method can be extended to including non-frontal faces. Without lost of generality, let us consider all the facial images whose sizes are normalized to $m \times n$, and thus each image is a mn-dimensional vector in the appearance space. In essence, the set $\prod$ forms a manifold in the appearance space, which is referred to as the *face manifold* in this paper. Assume that we have M persons in the collected face database, and let $\mathbf{X}_i = \{\mathrm{x}_i^c \mid c = 1, \ldots, N_i\}$ be a set of N_i facial images belong to person i, $i = 1, .. M$. Denote that $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2 \cup \ldots \cup \mathbf{X}_M$. The training vectors contained in $\mathbf{X}$ can then be treated as a set of finite instances obtained by sub-sampling the face manifold.

For any two faces x_i^{c1} and x_j^{c2} on the face manifold, let us think over that if we move along a smooth curve lying on the face manifold from x_i^{c1} to x_j^{c2}. Since the curve is contained in the face manifold, all the images in association with this curve shall appear as faces. Along this curve, the facial images are gradually changed from the i-th person to the j-th person. In addition to i and j, the curve may also go across faces which look like other persons. Generation of an ordered set of seamlessly varying images from one object to another is referred to as 'morphing' in the computer graphics (CG) community. Thus, we call such a curve an *appearance-morphing curve* (AMC). Note that an AMC from x_i^{c1} to x_j^{c2} is not unique. In general, there may be infinite AMCs from one face to another.

Generating a morphing sequence between two facial images is a mature technique in CG. In the latter section, we will show how to employ a morphing face sequence to evaluate the generalization ability of a face-based person identity classifier. However, since current image morphing techniques in CG usually require the users to select sufficient many point-to-point correspondences between two images, or convincing morphing results may not be obtained. This increases a considerable amount of human labor. In the work of this paper, to reduce manual manipulation effort, we use a linearization principle to approximate the face appearance manifold. Using linear approximations to represent appearance manifolds is not a new idea in computer vision. For example, the so-called feature line [3] connects two prototypes with the same identity (i.e., $i = j$ in the above case) with a line passing to both of them, and the nearest-feature-line classifier has been shown better than the nearest-neighbor classifier for face recognition. In this paper, the feature-line concept is modified and extended to including cross-identity cases. Let

$$L(i, c_1; j, c_2) = \{\mathrm{x}_i^{c1} + \mu(\mathrm{x}_j^{c2} - \mathrm{x}_i^{c1}) \mid 0 \leq \mu \leq 1\}; \qquad (1)$$

that is, L is a line segment connecting x_i^{c1} and x_j^{c2} in the appearance space. Such a simple linearization process approximates the manifold in an easily tractable manner when all the pairs of training examples have been connected. L is also referred to as an *approximated AMC* in this paper since it approximates morphing curves from x_i^{c1} to x_j^{c2} when $i\neq j$. The approximation is likely to, unavoidably, generate blurred facial images sometimes. Nevertheless, to our experience, the images still look like faces reasonable to some extent when the sizes of the images are small enough. Some examples are shown in Figure 1. On the other hand, although the images generated by approximated AMCs may not be of high quality, we will show that our method (that will be introduced in detail in the following sections) is still effective in improving face recognition performances of current techniques, and is expected to perform better if more accurate AMCs can be generated by using image morphing techniques.

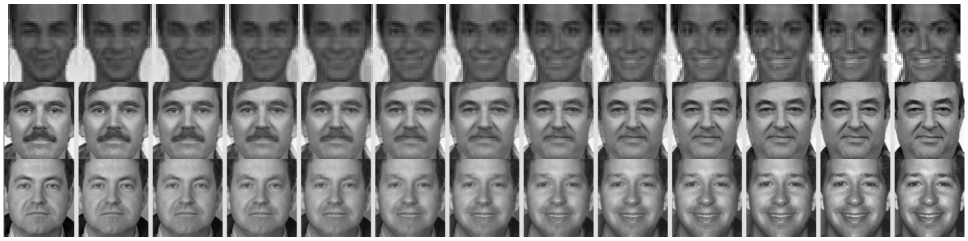

Figure 1. Each row shows a continuously-varying facial image sequence in association with an (linearly) approximated AMC starting from one person's face to another's.

3. Human-Perception-Assisted Assessment of a Face Identifier

3.1. Human-perception-based Segmentations of Appearance Morphing Curves

Consider an approximated AMC $\boldsymbol{L}_{ij}$ consisting of a set of continuously varying facial images from one person (say, person x_i) to another person (say x_j). We can treat faces along with $\boldsymbol{L}_{ij}$ as consisting of a set of incessantly changed appearances (or rough appearances) of faces. In essence, faces varying from x_i to x_j along $\boldsymbol{L}_{ij}$ are getting similar to x_j and dissimilar to x_i, as shown by the examples in Figures 1 and 2. In such a sequence of images mapped by the points lying on $\boldsymbol{L}_{ij}$, a region (of points) contained in $\boldsymbol{L}_{ij}$ associates to faces not very similar to either person i or person j. They are, in effect, new faces generated by interpolating x_i and x_j. We call this region the *ambiguous region* of $\boldsymbol{L}_{ij}$, and note that sometimes the lengths of ambiguous regions are small. Except to the ambiguous region, the other points in $\boldsymbol{L}_{ij}$ can be divided into two regions, faces look alike to person i and faces look alike to person j. To sum up, to human perception, $\boldsymbol{L}_{ij}$ can thus be divided into three regions, where two of them relate to the faces similar to x_i and x_j, respectively, and the other associates with the ambiguous faces (see Figure 2).

3.2. Human Perception-based Assessment of Generalization Ability

Let us consider a two-class classification problem at first. Based on the training data set of the *i*-th and the *j*-th persons, $\mathbf{X}_{i;j} = \mathbf{X}_i \cup \mathbf{X}_j$, assume that $\Psi(\cdot)$ is a classifier designed to discriminate person *i* and person *j* from their faces with $\Psi(x)>0$ being person *i* and $\Psi(x)<0$ being person *j*. Let $\Psi_B = \{x \mid \Psi(x) = 0\}$ be the classification boundary of Ψ. In practice, there are many methods that can be used to learn a classifier Ψ. Among them, what we care of is the generalization abilities of the generated classifier, i.e., the classification error rates to novel data not yet seen. In the past, many statistics-based principles or evaluation criterions have been proposed to assess the generalization ability of a classifier. For example, methods based on margin maximization, cross-validation (CV) or leave-one-out (loo) errors were widely adopted. However, they may or may not reflect the real generalization ability in an unknown test environment in practice. In this paper, instead of formulating a mathematical measure, we focus on the following problem – are the classifiers designed consistent with human perceptions to some extent? Since human perception is very accurate in face recognition, a classifier that is not coherent with human perception is regarded as suspect in its generalization ability.

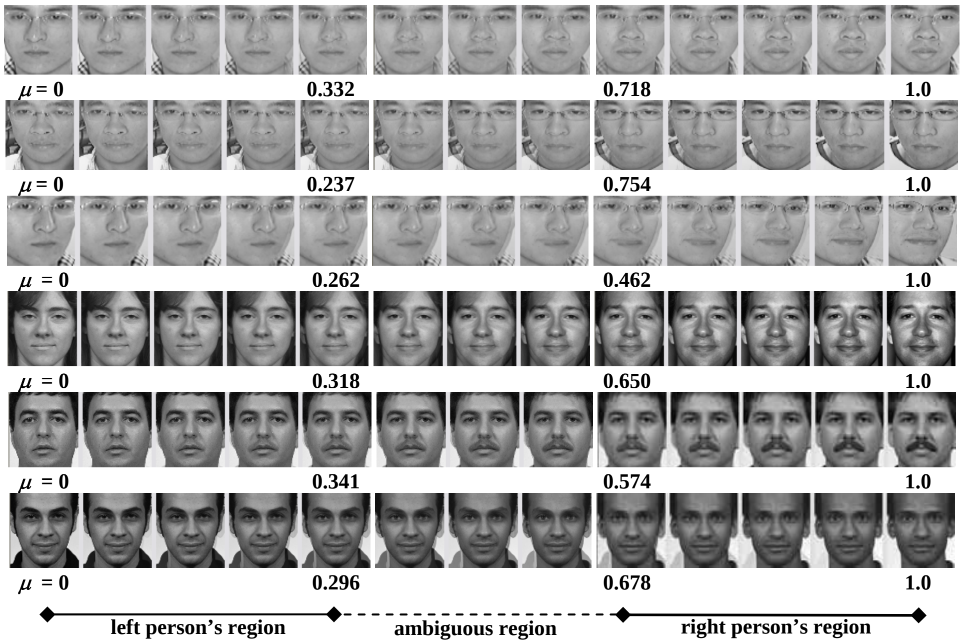

Figure 2. Segment facial image sequences of approximated AMCs. Each row of images of an approximated AMC are generated from $\mu = 0$ to $\mu = 1$ by equation (1). Below the last row of images, the values $\mu = 0.296$ and 0.678 indicate that the ambiguous region being chosen is from $\mu = 0.296$ to $\mu = 0.678$. Other rows follow the same interpretation.

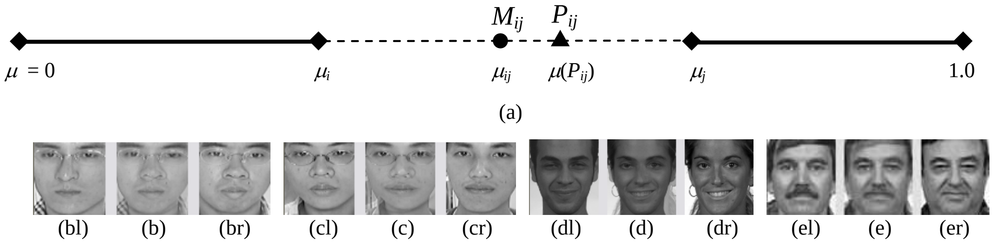

Figure 3. (a) An AMC from $\mu = 0$ to $\mu = 1$, where M_{ij} (corresponded by μ_{ij}) is the middle point of the ambiguous region, and μ_i and μ_j correspond to the left and right boundary points of the ambiguous region, respectively. (b), (c), (d), (e) are the novel faces corresponding to the middle points of the ambiguous regions of the AMCs generated for the image pairs (bl, br), (cl, cr), (dl, dr), and (el, er), respectively.

In principle, the classifier is expected to separate the training data sets $\mathbf{X}_i$ and $\mathbf{X}_j$ on two opposite sides. Hence, for some training vector x_i (contained in $\mathbf{X}_i$) and x_j (contained in $\mathbf{X}_j$), we would anticipate that the classification boundary Ψ_{B} intersects $\boldsymbol{L}_{ij}$ in some point P_{ij} because the AMC (or approximated AMC) continuously varies from a prototype x_i to another prototype x_j. By thinking of this, the following question is worth considered:

> *Where is a reasonable location of P_{ij} in the AMC for a classifier in harmony with human perception*?

The answer to the above question is very simple – we would expect that the point P_{ij} lies in the ambiguous region of $\boldsymbol{L}_{ij}$. The reason is that, if P_{ij} lies in the region in association with the facial images which look likes person *i* (or equivalently, P_{ij} is itself corresponding to a face looks alike to person *i*), then the classification boundary Ψ_{B} passes through the region in association with the *i*-th person and divides this region into two separate parts, where one part consists of facial images which look like person *i* but are miss-classified as person *j*. The above argument is applicable to person *j* too.

Even if the classifier boundary has passed through the ambiguous region, the appropriate location of *P* inside this region is still an issue worthy of being addressed. Assume that the ambiguous region is from $\mu = \mu_i$ to $\mu = \mu_j$ in (1). Denote $\mu_{ij} = (\mu_i + \mu_j)/2$, which corresponds to the middle point M_{ij} of the ambiguous region, as illustrated in Figure 3. We would expect that P_{ij} is as close to M_{ij} as possible, and thus the following measure is minimized:

$$d_{ij} = \frac{|\,\mu(P_{ij}) - \mu_{ij}\,|}{l_{ij}/2}, \tag{2}$$

where $l_{ij} = |\mu_i - \mu_j|$ is the width of the ambiguous region, and $\mu(P_{ij})$ is the interpolation coefficient (i.e., the μ-value) in association with P_{ij} in (1). Denote that $r_{ij} = 1 - d_{ij}$. Note that r_{ij} can be treated as a normalized distance from P_{ij} to the boundary of the ambiguous region (measured in the μ-domain) since $r_{ij} = 2min[|\mu_i - \mu(P_{ij})|, |\mu_j - \mu(P_{ij})|]/l_{ij}$. Hence, in some sense, r_{ij} can be thought of as a 'margin' based on human perception. The closer is the intersection point to the middle point of the ambiguous region, the larger is the 'margin' of the classifier synthesized. Consider all the AMCs, we have

$$\frac{1}{\|\Omega\|}\sum_{L_{ij}\in\Omega} r_{ij}, \tag{3}$$

where Ω is the set composed of all the approximated AMCs obtained from pairs of training examples of different labels in $\mathbf{X}_{i;j}$, and $\|\Omega\|$ is its cardinality. We call (3) the *margin from human perception* (MHP). Similar to the margins defined in statistics-based learning, a classifier has a larger MHP is expected to has a better generalization performance, and our experimental results (that will be shown in Section 5) also support this.

Remark: (1) Except to the original training data $\mathbf{X}_{i;j}$, all the other facial images mapped by the points in $\boldsymbol{L}_{ij}$ can be treated as novel data because they are not contained in the original training data set. We thus can treat equation (3) a new criterion, benefiting from human perception, to evaluate the generalization ability of an AFR classifier based on the unseen test data in association with AMCs.

(2) The above arguments were made by two-class case. Nevertheless, it can be easily generalized to multi-class case.

(3) We would like to emphasize that, although our criterion reflects the generalization ability to some extent, it still (like all the statistics-based criterions) suffers from the problem happened when the training and the testing data distributions are not consistent enough. Nevertheless, our criterion at least reflects the generalization ability to the novel data in association with AMCs.

3.3. Evaluation of the Generalization Ability for AFR – An Example using SVM

Among the existing classifier-learning methods, kernel-based methods have shown their efficiencies for many applications [6][8]. The most well known kernel-based method is the support vector machine (SVM), which has been employed for face recognition and has shown its promising performances [2]. Although SVM might not be the best for AFR among kernel-based methods [6], it is a representative method widely adopted. In this paper, we investigate the generalization ability of SVM from human perspective. Nevertheless, our method can be used for other learning methods as well.

The dataset used in the evaluation has ten people. Each has 30 facial images. We used the LibSVM [8] to train the 45 two-class classifiers in association with all the pairs of the ten. The model parameters of SVM have been appropriately chosen by model selection. Then, a multi-class classifier (with 10 classes) was composed according to the one-against-one strategy. The 10-fold CV accuracy obtained from this experiment is 99.7%, which turns to be very high and we also found that comparable high values can be achieved by many model parameters in a flat region in the parameter space. This phenomenon reveals that appropriate model parameters are difficult to be chosen for this dataset, and the yield generalization performance may not be satisfied.

To evaluate the SVM classifier trained by MHP, we have to segment each AMC into person-*i*, person-*j*, and ambiguous regions manually. In the current experiment, we have not segmented all the approximated AMCs yet. Instead, to reduce human labor, we generated a subset of segmentations of the AMCs. How to choose an appropriate subset is also an interesting issue.

Since a SVM classifier has already been trained, we can employ it for the choice. We have to find pairs of training vectors, where each pair contains two training vectors having different labels. Intuitively, a pair that is commendable to be chosen is better to have both of its training vectors close to the classification boundary in the feature space. Note that the classification boundary of SVM in the high-dimensional feature space is a plane. For each support vector of the SVM, say x_i, in the two-class classifier mentioned above, we can compute its mirror point with respect to this plane. However, this mirror point may not own a pre-image in the low-dimensional input space. We thus find an approximate mirror point that has the pre-image x_j which is also a training vector instead. Details about finding approximated mirror points can be found in [1]. Then, the ambiguous region of the AMC of (x_i, x_j) was manually segmented.

By choosing pairs of training examples in this way, a total of 1929 AMCs were segmented. These AMCs serve as a base for an approximate evaluation of the MHP defined in (3). The approximate MHP of the trained SVM classifier is 0.762. Ideally, we hope the MHP to be as close to one as possible. Compared with the CV accuracy (99.7%) that is difficult to be refined further, the MHP of the trained SVM still has space to be improved since it is not very close to one yet. In the next section, we will introduce how to increase the MHP by employing the manual segmentation results. In addition, we will also show experimentally that the generalization performance can be raised while the MHP is increased.

4. Human-Perception-Assisted Classifier Learning

Remember that, except to the ambiguous region, the other points in an AMC are grouped into two regions corresponding to the *i*-th and the *j*-th persons, respectively. To use this information for improving the classifier performance for face identification, we choose one facial image in *i*-th person's region and the other in *j*-th person's region, and call these two facial images *extended prototypes* corresponding to the pair (x_i, x_j). In particular, we choose the extended prototypes to be the two facial images corresponding to the boundary points of the ambiguous region, since they are supposed to carry critical information for identifying persons *i* and *j*. By including the two extended prototypes for each AMC, an *augmented face database* is generated. Then, we train a classifier for the augmented face database instead of the original database. In addition to the collected training examples, the extended prototypes are new ones inserted via human perception. Since the extended prototypes are closer to the middle point of the ambiguous region than x_i and x_j, the MHP is expected to being increased. Moreover, since the extended prototypes are all novel faces in addition to the original training examples, making them correctly categorized is highly possible to yield a classifier with better generalization ability as well.

In our experiments, we remained to use the SVM to train the new classifier based on the augmented face database. The principle of additionally employing the extended prototypes for improving facial-image classifier performances can be used as well when other learning methods are adopted. We call this a *performance-lifting-by-human-perception* (PL-HP) trick since the extended prototypes provided by a convincing source (here, human judge) are used to raise the AFR performance.

Further Remark: (1) The prototypes extended are based on the human segmentations for the AMCs. They are highly related to human perception, which may thus be affected by subjective opinions of different people. Currently, we have not investigated on how different subject opinions affect the assessment results yet. Nevertheless, note that a normal person's face-identification ability is still regarded as better than machine's. Hence, by referring the subject opinion of a common person, the face recognition accuracy is even expected to being raised by using the PL-HP trick.

(2) Employing human judge for increasing the learning performance has been used in other fields too. For example, in content-based image retrieval, query-based or active learning [7] (which chooses a query from a pool of un-labeled pre-collected training data) was used to boost the classifier performance via relevance feedback. However, this technique may not be suitable for AFR since many un-labeled facial images of real people have to be collected. The annotator may be asked a question like that "does a real person A look much more alike to B or C?" Such a question is sometimes difficult to answer, and the performance lifted may not be significant by replying this question. In our approach, instead of using facial images of real persons, an important characteristic is that we generate images of "virtual persons" along an AMC on the face manifold directly from the two persons (e.g., B and C in the above example) needing to be identified. From a sequence of successively varying facial images of the two persons, we can take advantage of the ambiguous region (which contains critical information for separating them) to increase the AFR performance.

5. Experimental Results

By employing the PL-HP trick introduced in Section 4, we have added 3858 (=1929×2) extended prototypes to the original dataset. We train the augmented facial database by using SVM, in which model selection combined with 10-fold cross validation was used for training. Then, the model parameters found was used to train the whole augmented database, and a classifier is obtained. The MHP of this classifier is 0.904, which is considerably increased from the original value (0.762).

In addition to MHP, we also care about the generalization performance of the trained classifier. To evaluate this, we have collected two other datasets (namely, D_1 and D_2) for the same ten persons, and each person has 30 facial images in both datasets too. These two datasets were gathered in different time (about a week later) from the time for gathering the original one depicted in Section 3.3. The trained classifiers were applied to the datasets D_1 and D_2, respectively. In the beginning, the error rates for D_1 and D_2 are 35% and 31%, respectively, when the original dataset was used for training. After employing the PL-HP trick, the error rates were reduced to 30.67% and 28.33% for D_1 and D_2, respectively, when the augmented database was used for training.

In another experiment, we tested a subset of the FERET [5] database. This subset contains 70 persons and each has six facial images. For each person, one of the six images was reserved for testing, and the other five were used for training. We used 5-fold cross-validation combined with model selection to train a SVM classifier. The classifier was then applied to the test data. In this experiment, the error rate for the test data was reduced from 18.57% to 12.86% by using the PL-HP trick after 8000 extended prototypes constructed from the training data has been added to the augmented database.

6. Conclusions and Future Work

In this paper, we present a systematic method on how to employ human perception for evaluating and improving face recognition performances by using the concept of AMC on the face manifold. We have also briefly introduced a procedure to choose a subset of cross-identity facial-image pairs by using support vectors and their approximated mirror points in the high-dimensional feature space, so that human labor for AMC segmentation can be lessened. Experimental results show that our method can successfully increase face identifiers' performances for AFR.

In the future, we also plan to use more learning methods for face classification, investigate the affections of subjective opinions, and employ 3D face morphing in the PL-HP trick.

References

[1] J. H. Chen and C. S. Chen, "Reducing Classification Time of SVM with Multiple Mirror Classifiers," *IEEE Trans. SMCB*, pp. 1173-1183, 2004.

[2] H. B. P. Ho, J. Wu and T. Poggio., "Face Recognition: Component-based versus Global Approaches," *Computer Vision and Image Understanding*, Vol. 91, No. 1/2, 6-21, 2003.

[3] S. Z. Lai, K. L. Chan, and C. L. Wang, "Performance Evaluation of the Nearest Feature Line Methods in Image Classification and Retrieval" *IEEE Trans. PAMI*, pp. 1335-1339, 2000.

[4] Q. Liu, R. Huang, H. Lu, and S. Ma, "Face Recognition Using Kernel Based Fisher Discriminant Analysis," in *Proc. of the Fifth IEEE Intl. Conf. on AFG,* 2002.

[5] P. J. Phillips, et. al, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing,* Vol. 16, pp 295-306, 1998.

[6] M. H. Yang, "Face Recognition Using Kernel Methods," in *Proc. of NIPS, 14*, pp. 215-220, MIT Press, 2002.

[7] C. Zhang and T. Chen, "Annotating Retrieval Database with Active Learning," in *Proc. of ICIP2003*.

[8] C. J. Lin, http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html.