Improved Face Annotation in Personal Photo Shared on OSN Using Collaborative Face Reorganization

Jyoti H. JAdhav[1] Pankaj Agarkar[2]
[1,2]Department of Computer Engineering
[1,2]Pune University

Abstract— User share and access the large volume of information on social networking sites like Facebook, Flicker. Face annotation for effective management of personal photos in online social networks (OSNs) is currently of considerable practical interest. The existing OSNs only support manual face annotation, a task that can be considered time-consuming and labor-intensive. We propose a collaborative face recognition (FR) framework, to improving the accuracy of face annotation. The accuracy of face annotation is improved by effectively making use of multiple FR engines available in an OSN. The collaborative FR framework consists of two major parts: selection of FR engines and merging (or fusion) of multiple FR results. In selection of FR engines it determines a set of personalized FR engines that are suitable for recognizing query face images belonging to a particular member of the OSN. For this purpose, it uses both social network context in an OSN and social context in personal photo collections. There are two solutions for merging FR results, traditional techniques for combining multiple classifier results. Experiments were conducted using 547 991 personal photos collected from an existing OSN. The results demonstrate that the proposed collaborative FR method is able to significantly improve the accuracy of face annotation, compared to conventional FR approaches that only make use of a single FR engine. Also the collaborative FR framework has a low computational cost and comes with a design that is suited for deployment in a decentralized OSN

Key words: Collaboration, face annotation, face reorganization, online social network, social context

I. Introduction

Face annotation is a relatively new topic in the field of face detection and recognition. ONLINE social networks (OSNs) such as Facebook [1] and MySpace [2] are frequently used for sharing and managing personal photo and video collections. Social networking forms an important part of online activities of Web users. Web sites such as Facebook, MySpace and Orkut have millions of users using them every day. Face recognition presents a challenging problem in the field of image analysis and computer vision, and it has received a great deal of attention over the last few years because of its applications in various domains. Mining web facial images on the internet has emerged as a promising paradigm towards auto face annotation. Content-based image retrieval (CBIR) systems require users to query images by their low-level visual content. This not only makes it hard for users to formulate queries, but also can lead to unsatisfied retrieval results. Because of this, Image annotation is introduced. The aim of image annotation is to automatically assign keywords to images, so image retrieval users are able to query images by Keywords and automatically detect human faces from a photo image and further name the faces with the corresponding human names.

Now days large amount of photos shared by users are human facial images and it is freely available in World Wide Web (WWW). Some of these facial images are tagged properly. Due to the significant increase of the amount of photos, a strong need has been emerged for automatic indexing. The most important and common entries for indexing personal photos are "who", "where", and "when" in that order. Since most people usually organize collection of photos on the basis of some particular persons of interest (e.g., photos including their friends) finding "who" on personal photos is one of the most promising applications.

A. *Face Recognization:*

Biometric-based technologies include identification based on Physiological characteristics and behavioral traits. Face recognition appears to offer several advantages over other Biometric method; facial images can be easily obtained with a couple of inexpensive fixed cameras. Good face recognition algorithms and appropriate preprocessing of the images can compensate for noise and slight variations in orientation, scale and illumination.

Face recognition is used for two primary tasks:

1) *Verification (One-To-One Matching):*

When presented with a face image of an unknown individual along with a claim of identity, ascertaining whether the individual is who he/she claims to be.

2) *Identification (One-To-Many Matching):*

Given an image of an unknown individual, determining that person's identity by comparing (possibly after encoding) that image with a database of (possibly encoded) images of known individuals.

There are numerous application areas in which face recognition can be exploited for these two purposes, a few of which are outlined below.

1) Security (access control to buildings, airports/seaports, ATM machines and border checkpoints.
2) Computer/ network security email authentication on multimedia workstations.
3) Criminal justice systems (mug-shot/booking systems, post-event analysis, forensics).
4) Image database investigations (searching image databases of licensed drivers, benefit recipients, missing children, immigrants and police bookings).
5) Video indexing.

In addition to these applications, the underlying techniques in the current face recognition technology have also been modified and used for related applications such as gender classification, expression recognition and facial feature recognition and tracking; each of these has its utility in various domains: Face recognition is also being used in conjunction with other biometrics such as speech, iris, fingerprint, ear and gait recognition in order to enhance the recognition performance of these methods.

B. Face Recognition Techniques:

Face recognition techniques can be broadly divided into three categories: methods that operate on intensity images, those that deal with video sequences, and those that require other sensory data such as infra-red imagery.

1) Face Recognition From Intensity Images:

Face recognition methods for intensity images fall into two main categories: Feature-based and Holistic. Feature-based approaches first process the input image to identify and extract (and measure) distinctive facial features such as the eyes, mouth, nose, etc., and then compute the geometric relationships among those facial points, thus reducing the input facial image to a vector of geometric features and Holistic approaches attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face.

2) Face Recognition From Video Sequences:

A video-based face recognition system typically consists of three modules: one for detecting the face; a second one for tracking it; and a third one for recognizing it. Most of these systems choose a few good frames and then apply one of the recognition techniques for intensity images to those frames in order to identify the individual.

C. Search Based Face Annotation (Sbfa):

Now a day's search-based face annotation plays a vital role. Specifically, given a user-uploaded facial image for annotation, the search-based face annotation scheme firstly retrieves a short list of top-K most similar facial images from a large scale web facial image database, and then annotates the query facial image by mining the labels associated with the top-K similar facial images. In general, the search-based face annotation scheme has to tackle two main challenges.

1) Efficiently retrieving the top-K most similar facial images from a large facial image database given a query facial image.
2) Effectively exploit the shortlist of Candidate facial images and their weak labels for naming the Faces automatically.

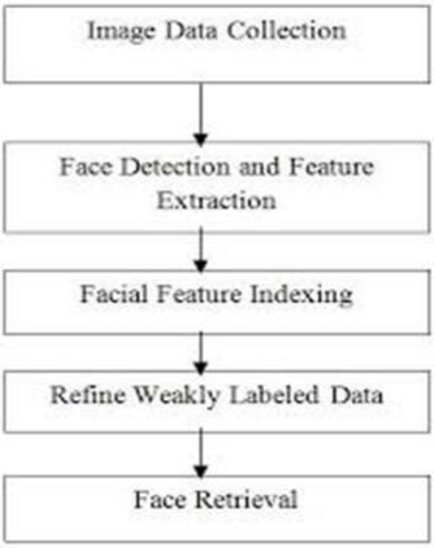

Fig. 1: Search Based Face Annotation

The SBFA consists of following steps.
Step 1: Facial image data collection
Step 2: Face detection and facial feature extraction
Step 3: High-dimensional facial feature indexing
Step 4: Learning to refine weakly labeled data
Step 5: Similar face retrieval

D. Face Annotation on Photos:

To index and retrieve personal photos based on an understanding of "who" is in the photos, annotation (or tagging) of faces is essential. However, manual face annotation by users is a time-consuming and inconsistent task that often imposes significant restrictions on exact browsing through personal photos containing their interesting persons. As an alternative, automatic face annotation solutions have been proposed. So far, conventional FR (Face Recognition) technologies have been used as main part to index people appearing in personal photos. FR techniques can take benefits to improve annotation accuracy by taking into account context information. In addition, in contrast to previous research in this field, method requires no training data labeled by hand from photos. From a practical point of view, it is highly desirable in most cases with a shortage of labeled data.

Three representative subspace FR methods are adopted as FR framework in the following: Principal Component Analysis (PCA or "eigenfaces"), Fisher-Linear Discriminate Analysis (FLDA or "fisher-faces), and Bayesian ("Probabilistic Eigen space"). Also, feature and measurement-level fusion strategies are used to efficiently take advantages of multiple facial features per person. In contrast to other FR based applications (e.g. Surveillance security and law enforcement), annotation of faces on personal photos can gain beneficial properties from time and space contextual information due to the following facts:

1) A sequence of photos taken in close proximity of time has relatively stationary visual context.
2) One would tend to take several pictures in a fixed place.

The act of labeling identities (i.e., names of individuals or subjects) on personal photos is called face annotation or name tagging. This feature is of considerable practical interest for OSNs. The existing OSNs only support manual face annotation, a task that can be considered time-consuming and labor-intensive, especially given the observation that the number of personal photos shared on OSNs continues to grow at a fast pace. To eliminate the need for manual face annotation, computer-based face detection and face recognition (FR) should be integrated into an automatic face annotation system. Traditional FR solutions still come with a low accuracy when dealing with personal photos, due to severe variations in illumination, pose, and spatial resolution.

E. Centralized FR Approach:

Most existing FR systems have been developed using a centralized FR approach. Existing social networking services are centralized and the companies providing the services have the sole authority to control all the data of the users. It is not a trivial task for a user to reuse his own data, including his social network, messages with friends and photos among other applications, as there are not many robust mechanisms to port all the data from one platform to another. This includes traditional FR application domains such as video surveillance and national security.

A centralized FR system relies on a single FR engine for the purpose of performing FR operations. The operations like subject identification or subject verification.

F. Collaborative FR Framework:

In OSNs, however, we believe that the use of multiple FR engines—belonging to members with close social relationships—can improve the accuracy of face annotation for the following two reasons.

– OSN applications typically have a personalized. Therefore, it can be expected that the weblog of each user will be equipped with a personalized FR engine that is specialized in recognizing a small set of individuals, such as the owner of the FR engine and the family members and friends of the owner. Consequently, personalized FR engines are expected to locally produce high-quality FR results for query face images detected in the personal photos of their respective owners.
– Current OSNs enable a user to maintain a list of contacts and to share image and video content with other OSN members. This facilitates creating a collaborative FR framework by sharing FR engines between a user of the OSN and his/her list of contacts.

In the proposed system collaborative FR framework that is able to effectively make use of a set of personalized FR engines that are available in an OSN, improving the accuracy of face annotation.

System is divided into two parts

1) The selection of expert FR engines that are able to recognize query face images
2) The merging of multiple FR results, originating from different FR engines, into a single FR result.

To select suitable FR engines, we make use of

– The social network context in an OSN (e.g., information about the connections between OSN members)
– The social context in personal photo collections (e.g., the number of times that a subject appears in a particular photo collection).

After that we construct a weighted social graph model (SGM) for several OSN members. Appropriate FR engines are then selected using the SGM. To merge the FR results returned by the selected FR engines, two solutions are proposed, adopting traditional techniques for combining multiple classifier results.

G. Uniqueness Of The Proposed Work:

– The face annotation methods reviewed above only make use of a single FR engine. In contrast, we propose a novel approach that makes use of multiple FR engines in a collaborative way, taking advantage of the social connections in an online social network.
– Previous face annotation methods utilize social context in personal photo collections or social network context as complementary evidence when determining the identity of faces on photos. Specifically, social context in personal photo collections and social network context are combined with baseline FR results. In the proposed method, however, we use social context for the effective selection of suitable FR engines. That way, we can increase the possibility of selecting FR engines that have been trained with a high number of face images, and where these training images are more likely to contain the identity of the query face image to be annotated.

II. Related Work

Several presentations uses social network context in an OSN for improving the effectiveness of multimedia content annotation and retrieval.

In [3] a system makes use of the social connections of a user in an OSN in order to facilitate the discovery of items of interest. In particular, search results are ranked to highlight recently posted items by contacts in the OSN, assuming that these items are of particular interest to the user who issued the search query.

The event-based image annotation, make use of both personal and social network context. The key idea in this work is that members of an OSN have strongly correlated real-world activities when they are friends, family members, or co-workers. By computing the correlation between the personal context models of the OSN members, the accuracy of event-based image annotation can be significantly improved.

In [5], to personalize image search results, a tag-based query only retrieves images that were either posted by people listed in the contact list of the user who issued the query, or that were posted by contacts of the contacts in question. The basic assumption in this work is that all users in a particular contact list tend to have common tagging behavior and common image interests.

System treats the annotation of personal photos as a stochastic process, using a time function that takes as domain the set of all people appearing in a photo collection. In this work, they construct a language probability model for every photo in order to estimate the probability of occurrence of each subject in the photo collections considered.

The scores are computed by taking into account the popularity of each subject and the co-occurrence statistics of pairs of individuals. The two aforementioned approaches show that likelihood scores can be effectively used to produce a limited set of candidate names for subjects appearing in a particular photo. However, a significant number of manually labeled photos are needed in order to build a reliable social context model.

A face annotation method based on incremental learning is proposed. This face annotation method shares identity information among members of an OSN that are connected to each other. The authors also discuss the differences between traditional FR systems and FR systems designed to operate in an OSN. In particular, they suggest that an FR engine customized for each member in an OSN is expected to be the most accurate for annotating faces in his/her own personal photo collections.

OSNs are beneficial for automatically labeling identities in photo collections. For example, it is possible to obtain a high number of tagged photos for a specific individual using the entire OSN. Further, an annotation method is proposed that uses a conditional random field (CRF) model to incorporate social network context into an automatic FR system. Specifically, by using already labeled photos available in the OSN, identity occurrence and co-occurrence statistics are combined with baseline FR scores to improve the accuracy of face annotation.

III. OVERVIEW

Fig. visualizes the construction of our collaborative FR framework for a particular OSN member, further referred to as the "current user".

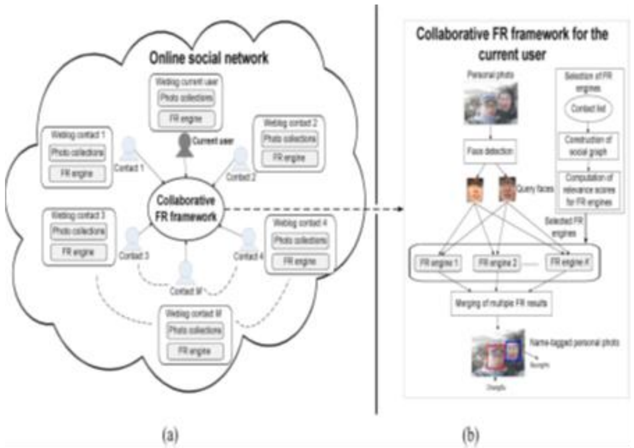

Fig. 2: Proposed collaborative FR framework in an OSN. (a) High-level visualization. (b) Detailed visualization.

In Fig. 2(a), the collaborative FR framework for the current user is constructed using M+1 different FR engines: one FR engine belongs to the current user, while M FR engines belong to M different contacts of the current user. We assume that photo collections and FR engines can be shared within the collaborative FR framework.

A. Phases Of Collaborative FR Framework:

Fig. 2(b) illustrates that our collaborative FR framework consists of two major parts:

– Selection of suitable FR engines and
– Merging of multiple FR results.

K suitable FR engines were selected out of M+1 FR engines. To select K suitable FR engines construct a social graph model (SGM) that represents the social relationships between the different contacts considered. An SGM is created by utilizing the personal photo collections shared in the collaborative FR framework. Based on the SGM, a relevance score is computed for each FR engine. K FR engines are then selected using the relevance scores computed for the FR engines. Next, the query face images detected in the photos of the current user are simultaneously forwarded to the selected K FR engines. To merge the FR results returned by the different FR engines, two solutions can be developed, both adopting traditional techniques for combining multiple classifier results. A key property of the two solutions is that they are able to simultaneously account for both the relevance scores computed for the selected FR engines and the FR result scores.

IV. SELECTION OF FR ENGINES BASED ON SOCIAL CONTEXT

Selecting suitable FR engines is based on the presence of social context in personal photos collections. The use of social context for selecting suitable FR engines is motivated by two reasons.

First social context is strongly consistent in collection of personal photos. Therefore, query face images extracted from the photos of the current user.

Second, each FR engine has been trained with a high number of training face images and corresponding name tags that belong to close contacts of the owner of the FR engine.

A. Construction of A Social Graph Model:

The Social Graph Model represents the social relationship between the current user and the OSN members in his/her contact list by making use of the identity occurrence and the co-occurrence probabilities of individuals in personal photo collections. Below figure shows the construction of a social graph model that allows for selecting suitable FR engines.

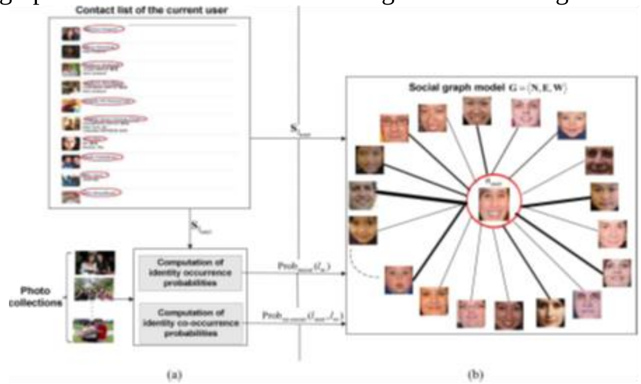

Fig. 3: Construction of social graph using contact list and personal photo collections.

The thickness of the lines in Fig. 3(b) represents the strength of the social relationship between the current user and a contact. The larger the weight, the thicker the edge, and the closer the current user and his/her contact

Let l_user be the identity label or the name of the current user [this is, the subject enclosed by the solid line circle in Fig. 3(b)]. Then, let S_(l_user)={l_m }M=1 be a set consisting of M different identity labels. These identity labels correspond to M OSN members that are contacts of the current user. l_m denotes the identity label of the mth contact and, without any loss of generality, l_m±l_n if m ± n .

A social graph is represented by a weighted graph as follows:

G = <N, E, W>………………………… (1)

Where N = {n_m |m=1,…,M} {n_user} is a set of nodes that includes the current user n_user and his/her contacts, E={e_m |m=1,…,M} is a set of edges connecting the node of the current user to the node of the mth contact of the current user, and the element w_(m)in W e_m represents the strength of the social relationship associated with .

To computew_(m), we estimate the identity occurrence and co-occurrence probabilities from personal photo collections. The occurrence probability for each contact is estimated as follows:

$$prob_{ocuur}(l_m) = \frac{\sum_{P \in P_{user}} \delta 1(l_m, P)}{P_{user}}, \text{ for } l_m \in S_{l_{user}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2)$$

Where P_user denotes the entire collection of photos owned by the current user, .| denotes the cardinality of a set, and δ1(l_m,P) is an indicator function that returns one when the identity of the mth contact is manually tagged in photo P and zero otherwise. In addition, the co-occurrence probability between the current user and the mth contact is estimated as follows:

$$prob_{co-ocuur}(, l_{user}, l_m) = \frac{\sum_{P \in P_{OSN}} \delta 1(l_{user}, l_m, P)}{|P_{OSN}}, \text{ for } l_m \in S_{l_{user}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Where P_OSN denotes all photo collections in the OSN the current user has access to (this includes photo collections owned by the current user, as well as photo

collections owned by his/her contacts), and (l_user,l_m,P) is a pair wise indicator function that returns one if the current user and the mth contact of the current user have both been tagged in photo P and zero otherwise.
Using (2) and (3), w_mis computed as follows:

w_m = exp ($prob_{ocuur}(l_m)$+$prob_{co-ocuur}(l_{user}, l_m)$)....(4)

B. Selection Of Face Recognition Engines:

The FR engine of the mth contact of the current user is denoted as Ω_m . To select suitable FR engines, we need to rank the FR engines Ω_m according to their ability to recognize a particular query face image. We make use of the strength of the social relationship between the current user and the mth contact of the current user represented byw_m. Specifically, w_m is used to represent the relevance score of the mth FR engine (i.e., the belief that an FR engine Ω_m is able to correctly recognize a given query face image).

When the M FR engines Ω_m (m=1… M) have been ranked according tow_m, two solutions can be used to select suitable FR engines.

- The first solution consists of selecting the top K FR engines according to their relevance score w_m.
- The second solution consists of selecting all FR engines with a relevance score that is higher than a certain threshold value.

In practice, the first solution is not reliable as the value of K may significantly vary from photo collection to photo collection (and where these photo collections are the subject of face annotation). Indeed, for each photo collection, we have to determine an appropriate value for K by relying on a heuristic process.

Therefore, we adopt the second solution to select suitable FR engines. Specifically, in collaborative FR framework, an FR engine is selected if its associated relevance score is higher than the average relevance score∑_(m=1)^M▒〖w_m /M〗 .

V. Merging Face Recognition Results

The purpose of merging multiple FR results retrieved from different FR engines is to improve the accuracy of face annotation. Such an improvement can be accomplished by virtue of a complementary effect caused by fusing multiple classification decisions regarding the identity of a query face image. In an OSN, fusion methods operating at the level of feature extractors or features are less suited from an implementation point-of-view. Indeed, FR engines that belong to different members of the OSN may use different FR techniques. For example, some feature extractors may have been created using global face representations, whereas other feature extractors may have been created using local face representations.

Two different solutions for merging multiple FR results are

- Fusion Using a Bayesian Decision Rule
- Fusion Using Confidence-Based Majority Voting

Common mathematical notation

Let {Ω_k} k=1 be a set containing K personalized FR engines that have been selected. Note that, in general FR systems, it is reasonable to assume that Ω_mconsist of two major components: a face feature extractor and an associated classifier. Also, let Q and {T^((n))} n=1be a query face image and a target set composed of G different face images of G distinct individuals. A function l(.) that returns the identity label for a given input face image.

A. Fusion Using A Bayesian Decision Rule:

To combine multiple FR results at measurement level, it makes use of fusion based on a Bayesian decision rule (BDRF). This kind of fusion is suitable for converting different types of distances or confidences into a common a posteriori probability. Hence, multiple FR results originating from a set of heterogeneous FR engines can be easily combined through a Bayesian decision rule. Moreover, the use of a Bayesian decision rule allows for optimal fusion at measurement level.

To perform collaborative FR, Q and T^((n)) are independently and simultaneously submitted to K different FR engines. Then, letq_k and t_(k^((n))) be a feature vector extracted from Q and T^((n)) , respectively, by using the face feature extractor of the FR engine Ω_k. Also, let us denote the dissimilarity value between q_k and t_(k^((n))) in the feature subspace asd_(k^((n))). Here, d_(k^((n))) can be computed by using the NN classifier assigned toΩ_k. Note that the total number of produced distance scores is equal to G.K.

B. Fusion Using Confidence-Based Majority Voting:

This solution uses the confidence-based majority voting (CMVF) for the purpose of decision-level fusion of multiple FR results. CMVF has been designed to take into account both the number of votes for a particular identity label (received from the selected FR engines) and the confidence values of these votes, leading to a final FR result that is highly reliable.

VI. Conclusions

The In this paper we have discuss the collaborative use of multiple FR engines allows improving the accuracy of face annotation for personal photo collections shared on OSNs. The accuracy can be attributed to the factor that in an OSN, the number of subjects that needs to be annotated tends to be relatively small, compared to the number of subjects encountered in traditional FR applications. Thus, an FR engine that belongs to an OSN member is typically highly specialized for the task of recognizing a small group of individuals.

The accuracy can also be attributed to the factor that in an OSN, query face images have a higher chance of belonging to people closely related to the photographer. Hence, simultaneously using the FR engines assigned to these individuals can lead to improved face annotation accuracy as these FR engines are expected to locally produce high-quality FR results for the given query face images.

References

[1] [Online]. Available: http://www.facebook.com.
[2] [Online]. Available: http://www.myspace.com.
[3] Gursel and S. Sen, “Improving search in social networks by agent based mining,” in Proc. Int. Joint Conf. Artificial Intelligence, 2009.
[4] R. Tron and R. Vidal, “Distributed face recognition via consensus on SE(3),” in Proc. 8th Workshop Omnidirectional Vision, Camera Networks and Non-classical Cameras-OMNIVIS, 2008.

[5] K. Lerman, A. Plangprasopchok, and C. Wong, “Personalizing image search results on flickr,” in Proc. Int. Joint Conf. Artificial Intelligence, 2007.
[6] C. A. Yeung, L. Licaardi, K. Lu, O. Seneviratne, and T. B. Lee, “Decentralization: The future of online social networking,” in Proc. Int. Joint Conf. W3C Workshop, 2009.
[7] Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-law distribution in empirical data,” SIAM Rev., vol. 51, pp. 661–703, 2009.
[8] Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” Pattern Recognit., vol. 38, no. 12, pp. 2270–2285, 2005.
[9] Moghaddam, T. Jebara, and A. Pentland, “Bayesian face recognition,” Pattern Recognit., vol. 33, no. 11, pp. 1771–1782, 2000.
[10] M. A. Turk and A. P. Pentland, “Eigenfaces for recognition,” J. Cognitive Neurosci., vol. 3, no. 1, pp. 71–86, 1991.
[11] V. Perlibakas, “Distance measures for PCA-based face recognition,” Pattern Recognit. Lett., vol. 25, no. 12, pp. 1421–1430, 2004.
[12] L. Chen, B. Hu, L. Zhang, M. Li, and H. J. Zhang, “Face annotation for family photo album management,” Int. J. Image Graph., vol. 3, no. 1, pp. 1–14, 2003.