Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks

Jae Young Choi, Wesley De Neve, Konstantinos N. Plataniotis, *Senior Member, IEEE*, and Yong Man Ro, *Senior Member, IEEE*

Abstract—Using face annotation for effective management of personal photos in online social networks (OSNs) is currently of considerable practical interest. In this paper, we propose a novel collaborative face recognition (FR) framework, improving the accuracy of face annotation by effectively making use of multiple FR engines available in an OSN. Our collaborative FR framework consists of two major parts: selection of FR engines and merging (or fusion) of multiple FR results. The selection of FR engines aims at determining 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, we exploit both social network context in an OSN and social context in personal photo collections. In addition, to take advantage of the availability of multiple FR results retrieved from the selected FR engines, we devise two effective solutions for merging FR results, adopting traditional techniques for combining multiple classifier results. Experiments were conducted using 547 991 personal photos collected from an existing OSN. Our 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. Further, we demonstrate that our collaborative FR framework has a low computational cost and comes with a design that is suited for deployment in a decentralized OSN.

Index Terms—Collaboration, face annotation, face recognition, online social network, personal photos, social context.

I. INTRODUCTION

ONLINE social networks (OSNs) such as Facebook [1] and MySpace [2] are frequently used for sharing and managing personal photo and video collections [3], [4]. The act of labeling identities (i.e., names of individuals or subjects) on personal photos is called *face annotation* or *name tagging* [5]. This feature is of considerable practical interest for OSNs thanks to its high commercialization potential [6], [7].

Most 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 [8], [9]. As reported in [10], the accuracy of face detection algorithms has improved considerably over the last decade. On the other hand, traditional FR solutions still come with a low accuracy when dealing with personal photos, due to severe variations in illumination, pose, and spatial resolution [6], [11].

Most existing FR systems have been developed using a centralized FR approach [12]. 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 (e.g., subject identification or subject verification [11]). 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 nature [4], [7], [13] (e.g., personalized content search and recommendation). 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 [6], [7], 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 [3], [4], [14]. This facilitates creating a collaborative FR framework by sharing FR engines between a user of the OSN and his/her list of contacts.

This paper proposes a novel *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. The following two key issues are addressed in this paper:

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) [16] and the social context in personal photo collections (e.g., the number of times that a subject appears in a particular photo collection) [8], [15]. To this end, we construct a weighted social graph model (SGM) for several OSN members. Appropriate FR engines are then selected using the SGM. Further, to merge the FR results returned by the selected FR engines, two solutions are proposed, adopting traditional techniques for combining multiple classifier results.

Using 547 991 personal photos collected from an existing OSN, extensive and comparative experiments have been conducted. In particular, the effectiveness of our collaborative FR framework has been successfully tested on challenging and real-life face images with severe variation in illumination, expression, and resolution (see Fig. 4). Our experiments show that the proposed collaborative FR framework is able to significantly improve the accuracy of face annotation for such challenging face images, compared to conventional FR approaches that only make use of a single FR engine (with accuracy gains of up to 30% for particular cases). Further, we demonstrate that our collaborative FR framework has a low computational cost and is suited for deployment in a decentralized OSN.

This paper is organized as follows. In Section II, we review related work. The novel contribution of our work is outlined as well. Section III presents an overview of the proposed collaborative FR framework. Section IV describes how to construct a social graph model by means of personal photo collections. This section also explains our approach towards selecting suitable FR engines. Section V addresses the problem of merging multiple FR results, returned by different FR engines, into a single FR result. Section VI subsequently discusses the experimental methodology used in this paper. In Section VII, we present experimental results that demonstrate the effectiveness and the efficiency of our collaborative FR framework. Conclusions and directions for future research are presented in Section VIII.

II. Related Work

Several studies have recently been presented that discuss the use of social network context in an OSN for improving the effectiveness of multimedia content annotation and retrieval. The authors of [4] propose a system that 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. In [16], the authors discuss event-based image annotation, making 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 [13], 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.

Several research efforts have demonstrated that social context in personal photo collections can be used for improving the accuracy of face annotation [8], [15]. In [8], the authors treat 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, the authors construct a language probability model for every photo in order to estimate the probability of occurrence of each subject in the photo collections considered. In [15], likelihood 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.

Few research efforts have thus far been dedicated to address the problem of face annotation in personal photo collections shared on OSNs [6], [7]. In [7], 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. In [6], the authors demonstrate that 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.

Compared to previous work, the novelty of our work is as follows.

1) 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.

2) 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.
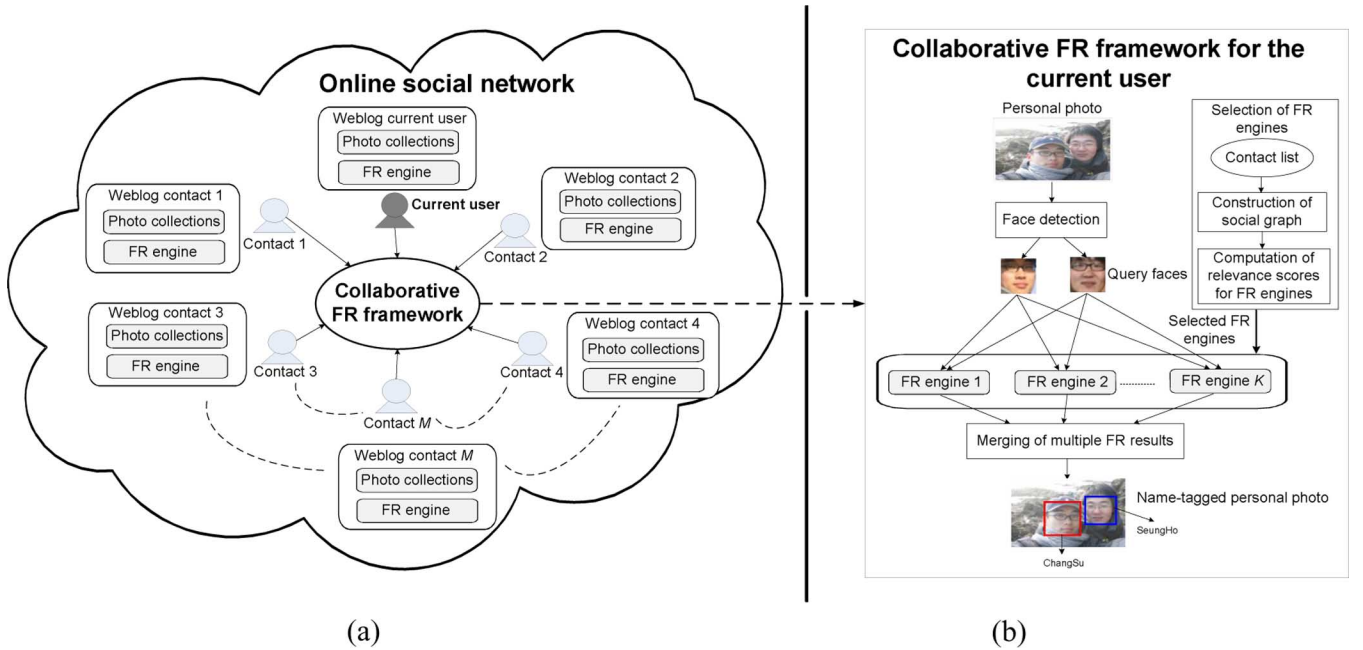
Fig. 1.   Proposed collaborative FR framework in an OSN. (a) High-level visualization. (b) Detailed visualization.

III. OVERVIEW

Fig. 1 visualizes the construction of our collaborative FR framework for a particular OSN member, further referred to as the "current user". As shown in Fig. 1(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.

Fig. 1(b) illustrates that our collaborative FR framework consists of two major parts: selection of suitable FR engines and merging of multiple FR results. To select K suitable FR engines out of $M + 1$ FR engines, we construct a social graph model (SGM) that represents the social relationships between the different contacts considered. In our research, an SGM is created by utilizing the personal photo collections shared in the collaborative FR framework. Based on the constructed 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. In order 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. The selection of suitable FR engines is discussed in Section IV, whereas Section V explains how to merge multiple FR results.

IV. SELECTION OF FR ENGINES BASED ON SOCIAL CONTEXT

Our method for selecting suitable FR engines is based on the presence of *social context* [15] in personal photos collections.

Social context refers to the strong tendency that users often take photos together with friends, family members, or co-workers.

The use of social context for selecting suitable FR engines is motivated by two reasons. First, as reported in [6], [8], and [15], social context is strongly consistent in a typical collection of personal photos. Therefore, query face images extracted from the photos of the current user are likely to belong to close contacts of the current user. Second, it is likely that 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. Consequently, by taking advantage of social context, the chance increases that FR engines are selected that are able to correctly recognize query face images. The proposed method for selecting expert FR engines will be described in more detail in Sections IV-A and B.[1]

A. Construction of a Social Graph Model

This section discusses the construction of a social graph model that allows for selecting suitable FR engines (see Fig. 2). To that end, we quantify 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.

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. 2(b)]. Then, let $\mathbf{S}_{l_{user}} = \{l_m\}_{m=1}^M$ be a set consisting of M different identity labels. These identity labels correspond to M OSN members that are *contacts* of the current user. Note that

---

[1]In the remainder of this paper, we assume that the FR engine of the current user is always selected for collaborative FR. This assumption is justified by the strong tendency that the most popular individual in a typical personal photo collection is the owner of the photo collection (i.e., the current user), as reported in [6]. Consequently, in the remainder of this paper, we focus on selecting a subset of expert FR engines that belong to contacts of the current user.
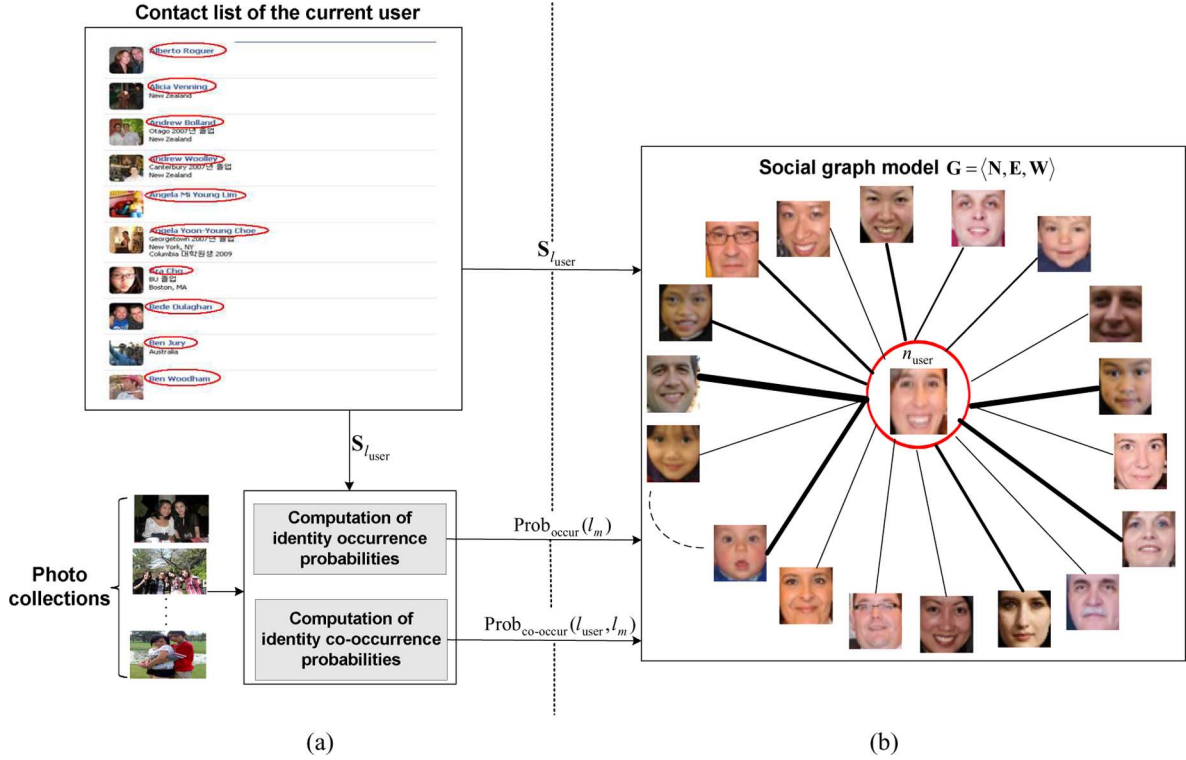
Fig. 2. Construction of a social graph model using a contact list and personal photo collections. The thickness of the lines in Fig. 2(b) represents the strength of the social relationship between the current user and a contact. The larger the weight, the thicker the edge, and the closer the current user and his/her contact.

l_m denotes the identity label of the mth contact and, without any loss of generality, $l_m \neq l_n$ if $m \neq n$.

In this work, a social graph is represented by a weighted graph as follows:

$$\mathbf{G} = \langle \mathbf{N}, \mathbf{E}, \mathbf{W} \rangle \qquad (1)$$

where $\mathbf{N} = \{n_m | m = 1, \ldots, M\} \cup \{n_{\text{user}}\}$ is a set of nodes that includes the current user n_{user} and his/her contacts, $\mathbf{E} = \{e_m | m = 1, \ldots, M\}$ 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 $\mathbf{W}$ represents the strength of the social relationship associated with e_m.

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

$$\text{prob}_{\text{occur}}(l_m) = \frac{\sum_{P \in \mathbf{P}_{\text{user}}} \delta_1(l_m, P)}{|\mathbf{P}_{\text{user}}|}, \text{ for } l_m \in \mathbf{S}_{l_{\text{user}}} \quad (2)$$

where $\mathbf{P}_{\text{user}}$ denotes the entire collection of photos owned by the current user, $|\cdot|$ denotes the cardinality of a set, and $\delta_1(l_m, P)$ 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:

$$\text{prob}_{\text{co-occur}}(l_{\text{user}}, l_m) = \frac{\sum_{P \in \mathbf{P}_{\text{OSN}}} \delta_2(l_{\text{user}}, l_m, P)}{|\mathbf{P}_{\text{OSN}}|},$$
$$\text{for } l_m \in \mathbf{S}_{l_{\text{user}}} \quad (3)$$

where $\mathbf{P}_{\text{OSN}}$ 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 $\delta_2(l_{\text{user}}, l_m, P)$ is a pairwise 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_m is computed as follows:

$$w_m = \exp(\text{prob}_{\text{occur}}(l_m) + \text{prob}_{\text{co-occur}}(l_{\text{user}}, l_m)). \quad (4)$$

In (4), the use of an exponential function leads to a high weighting value when $\text{prob}_{\text{occur}}(l_m)$ and $\text{prob}_{\text{co-occur}}(l_{\text{user}}, l_m)$ both have high values. Section IV-B describes how the w_m values are used to select relevant FR engines.

B. Selection of Face Recognition Engines

We denote the FR engine of the mth contact of the current user as $\mathbf{\Omega}_m$. In order to select suitable FR engines, we need to rank the FR engines $\mathbf{\Omega}_m$ according to their ability to recognize a particular query face image. To this end, we make use of the strength of the social relationship between the current user and the mth contact of the current user, represented by w_m [see (4)]. Specifically, w_m is used to represent the *relevance score* of the mth FR engine (i.e., the belief that an FR engine $\mathbf{\Omega}_m$ is able to correctly recognize a given query face image).

When the M FR engines $\mathbf{\Omega}_m (m = 1, \ldots, M)$ have been ranked according to w_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
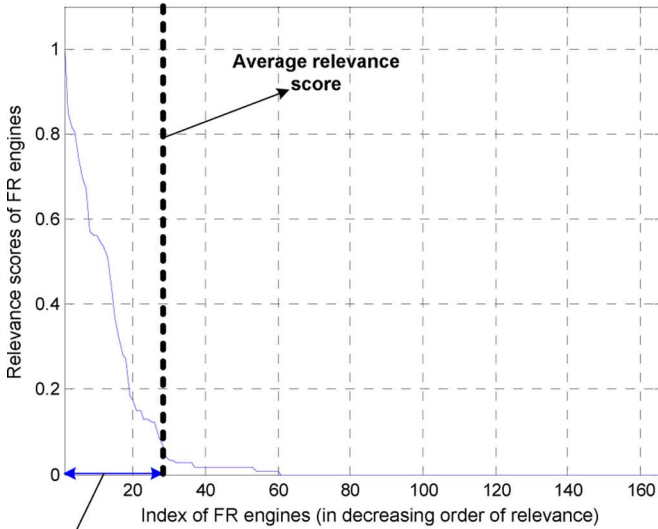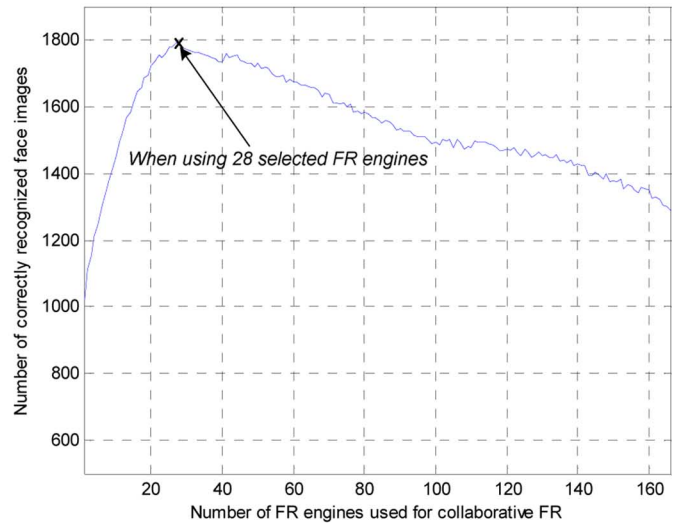
Fig. 3. Effectiveness of selecting FR engines using the proposed method. (a) Distribution of the relevance scores. (b) Distribution of the number of correctly recognized face images. Note that normalized relevance scores are used.

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 our collaborative FR framework, an FR engine is selected if its associated relevance score is higher than the average relevance score $\sum_{m=1}^{M} w_m / M$.

To validate the effectiveness of the proposed method for selecting FR engines, we conducted an experiment using four different test beds and four corresponding sets of query face images (see Sections VI-A and B for a detailed description of the test beds and the query sets used, respectively). Since similar results were obtained for the four test beds and the four query sets, we only present the results obtained for "Testbed1" (see Table II) and the "Q1" (see Table IV) query set. We computed relevance scores for 165 different FR engines. Each FR engine belongs to a contact of the current user associated with "Testbed1". In addition, we calculated the number of correctly recognized face images in the "Q1" query set, varying the number of FR engines used for collaborative FR. Note that the "Q1" query set contains a total of 2188 face images.

As shown in Fig. 3(a), the relationship between the relevance scores and the FR engines suitable for collaborative FR follows a power law distribution [18]. As such, most of the relevance scores, except for the relevance scores obtained for the dominant FR engines (i.e., the FR engines in the head of the power law distribution), are lower than the average relevance score [represented as a dotted line in Fig. 3(a)]. In addition, as can be seen in Fig. 3(b), the number of correctly recognized face images reaches a maximum when 28 FR engines are used for per-

forming collaborative FR (as selected by the proposed method). Consequently, the results in Fig. 3(a) and (b) demonstrate that most of the face images can be correctly recognized by using a small fraction of all available FR engines, and that these FR engines can be effectively selected by the proposed method, taking advantage of social context information.

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 [20]. Such an observation holds especially true for OSNs that have been implemented in a decentralized way, a topic that is currently of high interest [14]. Therefore, we only consider fusion of multiple classifier results at measurement level and at decision level [19].

In Sections V-A and B, we introduce two different solutions for merging multiple FR results. Before discussing these techniques in more detail, we first introduce a common mathematical notation. Let $\{\mathbf{\Omega}_k\}_{k=1}^{K}$ be a set containing K personalized FR engines that have been selected using the method described in Section IV-B. Note that, in general FR systems, it is reasonable to assume that $\mathbf{\Omega}_m$ consists of two major components [17], [20]: a face feature extractor and an associated classifier. Also, let $\mathbf{Q}$ and $\{\mathbf{T}^{(n)}\}_{n=1}^{G}$ be a query face image and a target set composed of G different face images of G distinct individuals.

In addition, we define a function $\ell(\cdot)$ that returns the identity label for a given input face image. Finally, we assume that all classifiers assigned to $\mathbf{\Omega}_k$ make use of nearest neighbor (NN) classification (e.g., by using the Euclidean distance metric).

A. Fusion Using a Bayesian Decision Rule

To combine multiple FR results at measurement level, we propose to make 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 [22]. 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 [23].

To perform collaborative FR, $\mathbf{Q}$ and $\mathbf{T}^{(n)}$ are independently and simultaneously submitted to K different FR engines. Then, let $\mathbf{q}_k$ and $\mathbf{t}_k^{(n)}$ be a feature vector extracted from $\mathbf{Q}$ and $\mathbf{T}^{(n)}$, respectively, by using the face feature extractor of the FR engine $\mathbf{\Omega}_k$. Also, let us denote the dissimilarity value between $\mathbf{q}_k$ and $\mathbf{t}_k^{(n)}$ in the feature subspace as $d_k^{(n)}$. Here, $d_k^{(n)}$ can be computed by using the NN classifier assigned to $\mathbf{\Omega}_k$. Note that the total number of produced distance scores is equal to $G \cdot K$.

Distance scores calculated by different NN classifiers may not be comparable due to the use of personalized FR engines. Therefore, as discussed above, we need to map the incomparable distance scores onto a common representation that takes the form of *a posteriori* probabilities. To obtain an *a posteriori* probability related to $d_k^{(n)}$, we first convert the distance scores into corresponding confidence values using a sigmoid activation function [22]. This can be expressed as follows:

$$c_k^{(n)} = \frac{1}{1 + \exp\left(d_k^{(n)}\right)}. \tag{5}$$

In (5), it should be emphasized that $d_k^{(n)}(1 \leq n \leq G)$ for a particular k must be normalized to have zero mean and unit standard deviation[2] prior to the computation of the confidence value $c_k^{(n)}$. A sum normalization method is subsequently employed to compute an *a posteriori* probability:

$$F_k^{(n)} = \text{prob}(\ell(\mathbf{Q}) = \ell(\mathbf{T}^{(n)})|\mathbf{\Omega}_k, \mathbf{Q}) = \frac{c_k^{(n)}}{\sum\limits_{n=1}^{G} c_k^{(n)}} \tag{6}$$

where $0 \leq F_k^{(n)} \leq 1$. As such, $F_k^{(n)}$ represents the probability that the identity label of $\mathbf{T}^{(n)}$ is assigned to that of $\mathbf{Q}$, assuming that $\mathbf{Q}$ is forwarded to $\mathbf{\Omega}_k$ for FR purposes. In the following, we refer to $F_k^{(n)}$ as the *FR result score* retrieved from $\mathbf{\Omega}_k$.

Given multiple FR result scores $F_k^{(n)}(k = 1, \ldots, K)$, a common practice in conventional measurement-level fusion methods is to simply combine $F_k^{(n)}$ into a single FR result score. However, this approach has the disadvantage that FR engines with a low reliability (for a given query face image

[2]Distance score normalization techniques are described in more detail in [24]. In our paper, the popular "z-score" technique is used to normalize distance scores.

Age	Gender	No. of contacts	No. of years during which each volunteer has maintained his/her weblog
28	Female	165	7 years
29	Male	118	4 years
30	Female	170	6 years
27	Male	84	8 years

$\mathbf{Q}$) contribute FR result scores to the final FR result score that have the same importance as the FR result scores contributed by highly reliable FR engines. Therefore, we assign a weight to each $F_k^{(n)}$ that takes into account the relevance score of the associated FR engine. The rationale behind this weighting scheme is that FR engines with high relevance scores are expected to be highly trained for a given query face image $\mathbf{Q}$ (thanks to the use of social context information to steer the selection of appropriate FR engines). Then, the *weighted FR result score* $\hat{F}_k^{(n)}$ for $F_k^{(n)}$ can be defined as follows:

$$\hat{F}_k^{(n)} = F_k^{(n)} + \alpha \cdot F_k^{(n)} \cdot R_k \tag{7}$$

where $R_k = (w_k - w_{\min})/(w_{\max} - w_{\min})$, $w_{\max}$, and $w_{\min}$ denote the maximum and minimum of all values $w_k(k = 1, \ldots, K)$, and the parameter $\alpha(0 \leq \alpha \leq 1)$ reflects the importance of R_k relative to $F_k^{(n)}$. Note that the importance of R_k becomes higher as α increases. Thus, when $\alpha = 0$, only the FR results $F_k^{(n)}$ are used during the fusion process. In (7), by properly adjusting the value of α, we can increase the importance of $F_k^{(n)}$ produced by FR engines with high relevance scores. On the other hand, we can decrease the importance of $F_k^{(n)}$ produced by FR engines with low relevance scores.

To merge the $\hat{F}_k^{(n)}$ computed by (7), the sum rule is used:

$$CF^{(n)} = \sum_{k=1}^{K} \hat{F}_k^{(n)}, \quad n = 1, \ldots, G. \tag{8}$$

The sum rule allows for optimal fusion at measurement level, compared to other rules such as the product and median rule [23]. Finally, to perform face annotation on $\mathbf{Q}$, the identity label of $\mathbf{Q}$ is determined by choosing the identity label of $\mathbf{T}^{(n)}$ that achieves the highest value for $CF^{(n)}$:

$$\ell(\mathbf{Q}) = \ell(\mathbf{T}^{(n^*)}) \text{ and } n^* = \arg\max_{n=1}^{G} CF^{(n)}. \tag{9}$$

B. Fusion Using Confidence-Based Majority Voting

This section discusses the use of confidence-based majority voting (CMVF) for the purpose of decision-level fusion of multiple FR results [23]. 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.

In the proposed fusion strategy, individual identity labels and corresponding confidence values are separately computed by matching each query feature $\mathbf{q}_k$ against a set of target features $\{\mathbf{t}_k^{(n)}\}_{n=1}^{G}$, given that the kth FR engine $\mathbf{\Omega}_k$ is used. To be more specific, based on the dissimilarity scores $d_k^{(n)}$ computed between $\mathbf{q}_k$ and $\mathbf{t}_k^{(n)}$, we calculate the number of votes for a par-

TABLE II
DETAILED DESCRIPTION OF THE FOUR DIFFERENT TEST BEDS USED

Test bed ID	Test bed 1	Test bed 2	Test bed 3	Test bed 4
Total no. of photos	251,211	109,021	117,772	69,987
No. of different photo collections	166	119	171	85
Avg. no. of photos in a collection	1,311	821	532	765

ticular identity label (i.e., $\ell(\mathbf{T}^{(n)})$) and the confidence values of these votes.

Let us first define a *FR result score* outputted by each $\mathbf{\Omega}_k$:

$$F_k^{(n)} = \Delta\left(\mathbf{q}_k, \mathbf{t}_k^{(n)}\right) = \begin{cases} 1, & \text{if } n = \arg\min_{i=1}^{G} d_k^{(i)} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where $\Delta\left(\mathbf{q}_k, \mathbf{t}_k^{(n)}\right)$ is an indicator function that returns one when the minimum of the dissimilarity values $d_k^{(i)}$ between $\mathbf{q}_k$ and $\mathbf{t}_k^{(i)}(i = 1, \ldots, G)$—computed using the classifier assigned to $\mathbf{\Omega}_k$—is achieved at $i = n$, and returns zero otherwise. Then, let $N_{\text{vote}}^{(n)}$ be the total number of votes given to the nth target identity label, received from all FR engines $\mathbf{\Omega}_k(k = 1, \ldots, K)$:

$$N_{\text{vote}}^{(n)} = \sum_{k=1}^{K} F_k^{(n)}, \quad n = 1, \ldots, G. \quad (11)$$

To determine the confidence value that corresponds with $N_{\text{vote}}^{(n)}$, the $d_k^{(n)}$ values for a particular k are first normalized to have zero mean and unit standard deviation. The normalized distance values are subsequently mapped onto values in the confidence domain using a soft-max function [43]:

$$c_k^{(n)} = \frac{\exp\left(-d_k^{(n)}\right)}{\sum_{n=1}^{G} \exp\left(-d_k^{(n)}\right)}. \quad (12)$$

In (12), the confidence value $c_k^{(n)}$ measures the degree to which the vote for an identity label of $\mathbf{Q}$ (received from $\mathbf{\Omega}_k$) is equal to the correct vote for the true identity label of $\mathbf{Q}$. Using R_k defined in Section V-A and $c_k^{(n)}$ defined in (12), the total confidence value associated with $N_{\text{vote}}^{(n)}$ is computed as follows:

$$C_{\text{conf}}^{(n)} = \sum_{k=1}^{K} \Delta\left(\mathbf{q}_k, \mathbf{t}_k^{(n)}\right) \cdot \hat{c}_k^{(n)} \quad (13)$$

where

$$\hat{c}_k^{(n)} = c_k^{(n)} + \alpha \cdot c_k^{(n)} \cdot R_k \quad (14)$$

and $0 \leq \alpha \leq 1$. In (13), $C_{\text{conf}}^{(n)}$ is the sum of the *weighted confidence values* $\hat{c}_k^{(n)}$ corresponding to the vote for the nth target identity label and where the vote has been received from $\mathbf{\Omega}_k$.

Finally, the target identity label that achieves the highest combined value of $N_{\text{vote}}^{(n)}$ and $C_{\text{conf}}^{(n)}$ is used to annotate $\mathbf{Q}$:

$$\ell(\mathbf{Q}) = \ell(\mathbf{T}^{(n^*)}) \text{ and } n^* = \arg\max_{n=1}^{G}\left(N_{\text{vote}}^{(n)} \cdot C_{\text{conf}}^{(n)}\right). \quad (15)$$
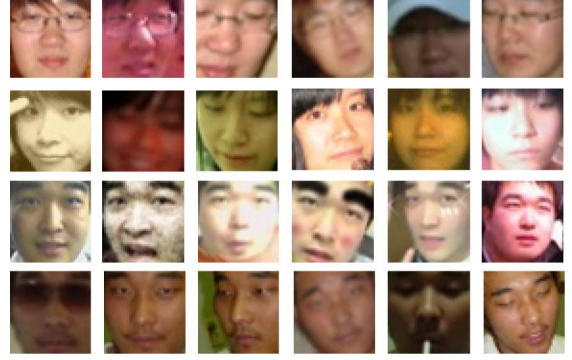

Fig. 4. Example face images, obtained after face detection and preprocessing. Each row contains face images of the same subject.

VI. EXPERIMENTAL METHODOLOGY

An experimental study was conducted to investigate both the effectiveness and the efficiency of the proposed collaborative FR framework. A total of 547 991 personal photos were retrieved from Cyworld [25], an online social network that has its roots in Korea. This made it easier to find volunteers willing to participate in our study. The collected personal photos were used to construct four different test beds. Sections VI-A–D describe the four different test beds and our evaluation methodology.

A. Construction of Test Beds and Ground Truth

To retrieve personal photos from "Cyworld", we relied on the cooperation of four volunteers whose age is ranging from 28 to 30. The four volunteers, acting as current users (see Fig. 2), are all active members of Cyworld. Further, 537 other individuals, present in the contact lists of the four volunteers, agreed to contribute their photos to our experimental study. Table I provides more information about the four volunteers who participated in our experimental study.

We collected all photos from the weblog of each volunteer, as well as all photos posted on the weblogs of the contacts of each volunteer. As a result, we constructed a test bed for each volunteer. Table II provides more details regarding the four test beds used in our experiments. As an example, "Test bed 1" contains one photo collection that was acquired from the current user, whereas the remaining 165 photo collections were obtained from the 165 contacts of the current user.

Given a particular test bed, we constructed a corresponding ground truth composed of manually labeled face images. For doing so, *Viola-Jones* face detection algorithm [26] was applied to all photos in each test bed. In practice, users prefer to annotate known individuals, such as friends and family members (as reported in [6] and [8]). For this reason, we manually labeled detected face images of individuals who appear at least ten times

TABLE III
INFORMATION ABOUT THE GROUND TRUTH CREATED FOR EACH TEST BED, INDEPENDENTLY CONSTRUCTED FOR EACH PHOTO COLLECTION. THE TERM "NAMED INDIVIDUAL(S)" REFERS TO SUBJECTS WITH IDENTITY LABELS MANUALLY ASSIGNED DURING THE CREATION OF THE GROUND TRUTH

Ground truth ID	G1	G2	G3	G4
No. of named individuals	2,510	1,834	2,607	1,302
No. of photos containing named individuals	188,422	81,211	94,297	59,753
No. of detected face images	213,363	94,452	104,408	64,412
Average no. of photos per named individual	66	45	38	51

TABLE IV
SUMMARIZING STATISTICS FOR THE FOUR DIFFERENT QUERY SETS

Query set ID	Q1	Q2	Q3	Q4
Total no. of target subjects to be annotated	35	64	57	42
No. of target subjects enrolled in a contact list	28	47	45	34
Total no. of face images of target subjects	2,188	4,152	2,967	3,365
No. of face images of target subjects enrolled in a contact list	1,994	3,311	2,178	3,133
Avg. no. of face images per target subject	61	43	54	74
No. of face images belonging to the volunteer	321	142	211	289

in each photo collection used. Also, all of the detected face images were individually rotated and rescaled to 86×86 pixels using the center coordinates of the eyes. That way, eye centers were placed on known pixel locations (as recommended by the FERET protocol [27]). The center coordinates of the eyes were automatically determined using the eye detection algorithm proposed in [28].

Fig. 4 shows a number of example face images that are part of the ground truth. It should be clear that the face images used in our experiments are significantly challenging for face recognition purposes, due to severe variations in terms of illumination and pose, the use of a low spatial resolution, and the presence of heavy make-up and occlusions. In addition, Table III provides detailed information about the constructed ground truth datasets.

B. Construction of Target and Query Face Image Sets

Using the ground truth datasets described in Table III, we constructed corresponding sets of target and query face images in order to evaluate the face annotation accuracy of the proposed collaborative FR framework. As face annotation needs to be performed on photos posted on the weblog of the current user, different target and query sets were created using the photo collection owned by each volunteer (i.e., the current user). More specifically, we constructed four different target sets, and each target set contains a single face image for each subject who appears at least ten times in the photo collection owned by the current user. In addition, we separately constructed four different query sets using the four corresponding ground truth datasets. Each query set consists of unseen face images that belong to subjects who are enrolled in the corresponding target set. Statistics for the four different query sets are listed in Table IV.

Table IV allows making the following common observations for the four different query sets:
- as much as 90% of the face images contained in each query set belong to individuals that are enrolled in the contact list of the volunteer;
- face images of the volunteer (i.e., the current user) are the most dominant in all four query sets;

- the number of target subjects included in each query set is significantly smaller than the total number of subjects in a contact list (as shown in Table I).

C. Construction of FR Engines

As mentioned in Section I, we assume that the current user and each contact of the current user make use of a personalized FR engine. Hence, to evaluate the performance of our collaborative FR framework, we created several training sets, allowing for the construction of a personalized FR engine for each OSN member (see Table II). For instance, each of the 166 different photo collections in "Test bed 1" was used to create a corresponding training set, resulting in a total of 166 different training sets. These 166 different training sets were, in turn, used for constructing 166 independent FR engines (one FR engine for each OSN member in "Test bed 1"). To simulate a realistic collaborative FR framework (i.e., a collaborative FR framework consisting of personalized FR engines that are able to recognize a small group of individuals), we selected the 15 most frequently appearing subjects in each photo collection, always including the owner of the weblog. In addition, 21 face images per selected subject were used on average to construct an FR engine.

Each FR engine consists of a face feature extractor and a corresponding classifier. When constructing an FR engine, the training step typically focuses on creating a face feature extractor, rather than putting the focus on creating an associated classifier [30]. Hence, in our experiments, the training set created for constructing an FR engine was used to train its associated feature extractor. To this end, the following six feature extraction algorithms were used to construct six different types of FR engines: eigenfeature regularization and extraction (ERE) [31], marginal Fisher analysis (MFA) [32], kernel direct discriminant analysis (KDDA) [33], Bayesian [34], Fisher linear discriminant analysis (FLDA) [35], and PCA [36]. The use of six different types of FR engines allows testing the effectiveness and stability of a heterogeneous collaborative FR framework.

Note that a radial basis function (RBF) was adopted as the kernel function for implementing KDDA [33]. Also note that, for MFA, the parameters associated with the number of nearest neighbor samples (used to construct the intrinsic and penalty
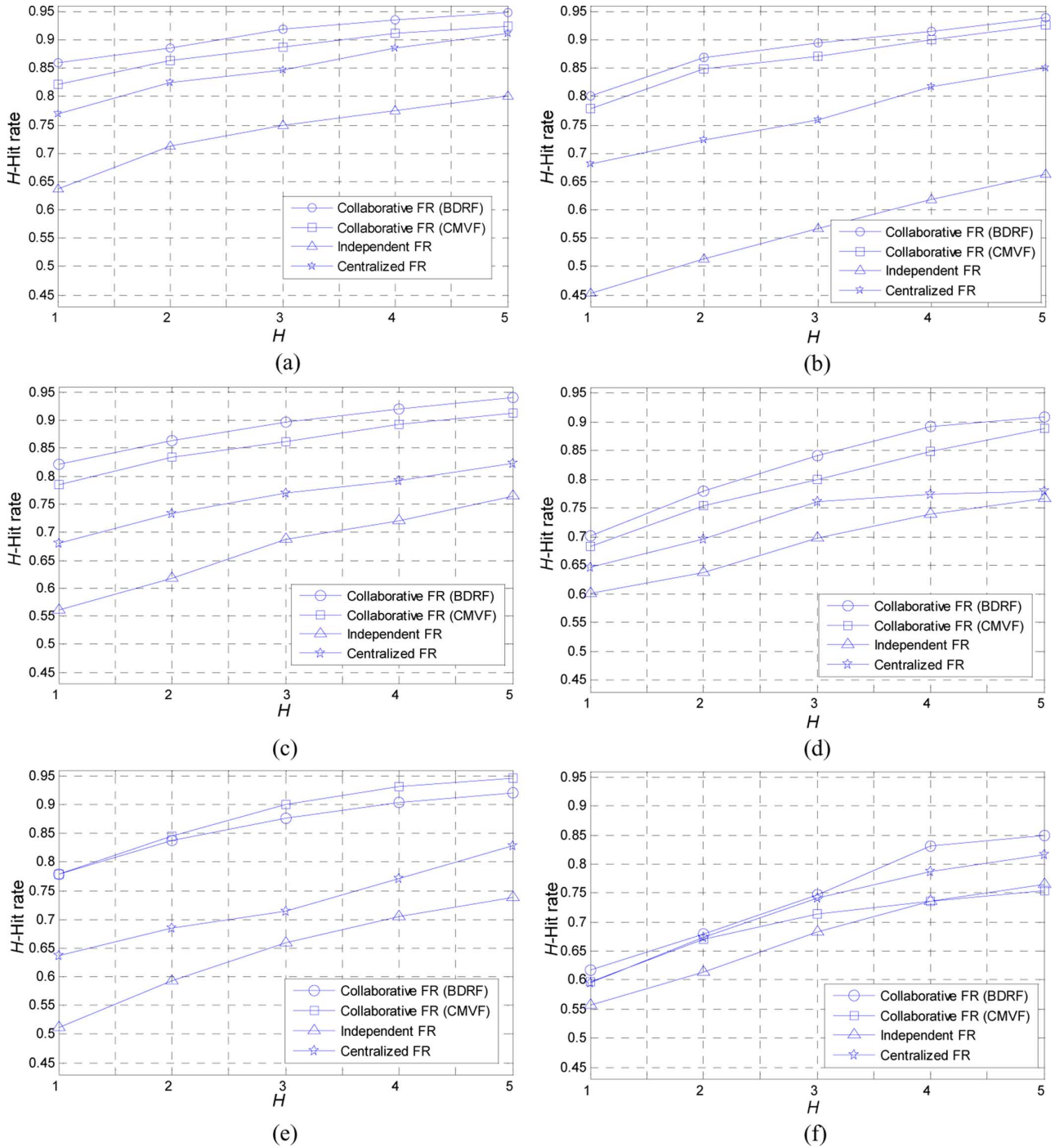
Fig. 5. Comparison of H-Hit rates. (a) ERE. (b) MFA. (c) KDDA. (d) Bayesian. (e) FLDA. (f) PCA.

graphs) were determined using the approach as suggested in [32]. As for the nearest neighbor classifiers, the Euclidean distance was used for MFA and KDDA, whereas the cosine distance measure was used for ERE (as recommended in [31]). Finally, the Mahalanobis distance [37] was employed for PCA and the Bayesian approach to FR.

D. Evaluation Method

The face annotation accuracy of two non-collaborative FR frameworks, denoted as centralized and independent FR, is referred to as the baseline performance in our comparison. Specif-

ically, for the case of centralized FR, all face images from all training sets—which are used to construct a set of FR engines in our collaborative FR framework—were combined into a single global training set. The performance of the face feature extractor trained using this global training set serves as the baseline face annotation performance. For the case of independent FR, face annotation accuracy is measured by averaging the annotation performance of all personalized FR engines used to perform collaborative FR. In order to guarantee a fair comparison, the target and query face image sets are the same in both the non-collaborative and collaborative FR frameworks.

In our experiments, the H-Hit rate [38] was adopted to measure the accuracy of subject identification, whereas the F-*measure* was used to measure the performance of subject-based photo retrieval. The F-*measure* in our experiment is defined as follows:

$$F\text{-}measure = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \qquad (16)$$

where

$$\text{precision} = \frac{1}{G} \sum_{n=1}^{G} \frac{N_{\text{correct}}^{(n)}}{N_{\text{retrieval}}^{(n)}} \text{ and}$$

$$\text{recall} = \frac{1}{G} \sum_{n=1}^{G} \frac{N_{\text{correct}}^{(n)}}{N_{\text{ground}}^{(n)}} \qquad (17)$$

and G is the total number of target subjects, $N_{\text{retrieval}}^{(n)}$ is the number of retrieved photos annotated with identity label n, $N_{\text{correct}}^{(n)}$ is the number of photos correctly annotated with identity label n, and $N_{\text{ground}}^{(n)}$ is the number of photos annotated with identity label n in the ground truth.

VII. Experimental Results

Four experiments have been carried out to investigate the effectiveness and the efficiency of the proposed collaborative FR framework, focusing on the face annotation accuracy related to FR. As such, face detection error rates were excluded from our performance evaluation. However, errors related to the computation of eye coordinates in the preprocessing procedure, which may cause a mismatch between the estimated and the true eye positions, were not excluded.

From an application point-of-view, two collaborative FR scenarios can be identified:

1) "Collaborative FR scenario 1": the same type of FR engines is used.
2) "Collaborative FR scenario 2": different types of FR engines are used.

In Sections VII-A and B, the face annotation accuracy of the proposed collaborative FR framework is evaluated with respect to the two aforementioned collaborative FR scenarios.

A. Evaluation of Collaborative FR Scenario 1

This section assesses the face annotation accuracy of the proposed collaborative FR framework when using the same type of FR engine. Specifically, each FR engine uses the same type of face feature extractor and the same type of NN-based classifier. All results were obtained after having performed the proposed FR engine selection step. As a result, for the two collaborative FR methods using either BDRF or CMVF to merge multiple FR results, "28", "47", "52", and "34" FR engines were selected from a total of "166", "119", "171", and "85" FR engines for the four query sets Q1, Q2, Q3, and Q4, respectively.

The face annotation accuracy was separately evaluated for the four query sets. All results were then averaged over the four different query sets. When implementing BDRF and CMVF, a good compromise was found by setting α [shown in (7) and (14)] to 0.4. However, our experimental results demonstrated that varying these weights in the range of $[0.2, 0.9]$ did not
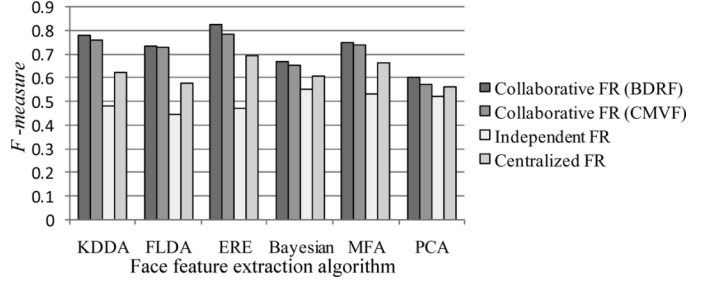

Fig. 6. Comparison of F-*measure* values.

significantly alter the performance of the two fusion strategies used.

Fig. 5 compares the H-Hit rates of the two proposed collaborative FR methods with the H-Hit rates obtained for the two baseline FR methods. The following observations can be made.

- Compared to independent FR, the H-Hit rates of the two collaborative FR methods are significantly better for most of the face feature extraction algorithms considered.
- Both collaborative FR methods significantly outperform centralized FR, except when PCA and Bayesian are in use.
- Compared to the two baseline FR methods, substantial improvements in H-Hit rates can be achieved by our collaborative FR methods when supervised face feature extraction algorithms are used to construct FR engines (that is, ERE, MFA, KDDA, and FLDA). The underlying reason for this is that the FR accuracy of supervised feature extraction algorithms, in general, depends largely on the presence or absence of target subjects (i.e., subjects to be recognized) in a training set, as well as the number of training samples per subject [35], [39]. Thus, the complementary effect of collaborative FR on the improvement in accuracy is significant when being able to utilize expert FR engines (i.e., FR engines that have been trained with a high number of training images).

Fig. 6 compares the F-*measure* values of the two proposed collaborative FR methods with the F-*measure* values obtained for the two baseline FR methods. The results shown in Fig. 6 are consistent with the results depicted in Fig. 5, confirming the effectiveness of our collaborative FR framework.

B. Evaluation of Collaborative FR Scenario 2

This section investigates the face annotation performance of the proposed collaborative FR framework when different types of FR engines are used. The six different face feature extractors (see Section VI-C) were randomly assigned to the FR engines using a uniform distribution. As a result, the FR engines were distributed evenly among the different types of face feature extractors. For instance, the number of face feature extractors allocated to the 28 FR engines selected to annotate query set Q1 was "5", "5", "5", "5", "4", and "4", in the order of ERE, MFA, KDDA, Bayesian, FLDA, and PCA. To guarantee stable experimental results, 20 independent runs were executed, each randomly assigning face feature extractors to the FR engines considered. Thus, all results were averaged over 20 runs.

The H-Hit rates and the F-*measure* results for the two collaborative FR methods and the independent FR method are
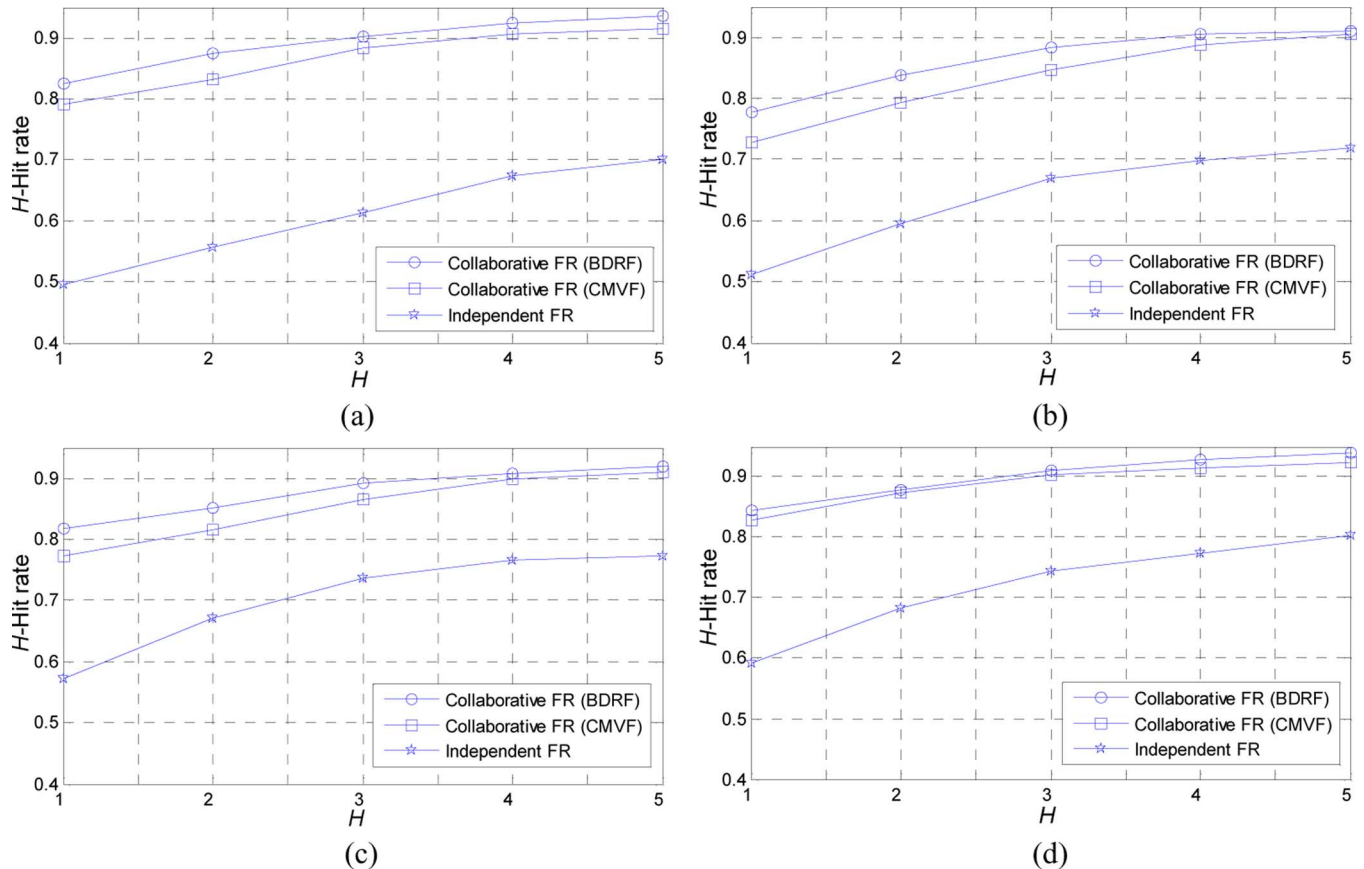
Fig. 7.   Comparison of H-Hit rates. (a) Q1. (b) Q2. (c) Q3. (d) Q4.

TABLE V
COMPARISON OF $F\text{-}measure$ VALUES

Query set	Collaborative FR (BDRF)	Collaborative FR (CMVF)	Independent FR
Q1	0.8324	0.7887	0.4820
Q2	0.6943	0.6721	0.4713
Q3	0.7299	0.7011	0.4855
Q4	0.8011	0.7834	0.5422

shown in Fig. 7 and Table V, respectively. Note that baseline performance is only measured for the case of independent FR. Centralized FR is not possible when different types of FR engines are in use. For the two collaborative FR methods using BDRF and CMVF, 28, 47, 52, and 34 FR engines were used. The value of α (for BDRF and CMVF, respectively) was set to 0.4. As can be seen in Fig. 7 and Table V, the use of collaborative FR significantly improves the face annotation accuracy compared to independent FR. For instance, for the case of collaborative FR using BDRF, the 1-Hit rate (at $H = 1$) increases with 33%, 26%, 24%, and 25% for the Q1, Q2, Q3, and Q4 query sets, respectively. In summary, our experimental results demonstrate that collaborative FR can significantly enhance the face annotation accuracy, even when different types of FR engines are used.

C. Effect of Selecting FR Engines and Weighting FR Results

In this subsection, we examine the effect of selecting suitable FR engines and the effect of weighting FR results on the overall face annotation accuracy of our collaborative FR framework. Note that these two factors are related to the exploitation of social context both in an OSN and in personal photo collections.

The experiments presented in this subsection share the following setup. When suitable FR engines are selected, 28, 47, 52, and 34 FR engines were used to perform collaborative FR or independent FR for the four query sets Q1, Q2, Q3, and Q4, respectively. On the other hand, when skipping the selection of suitable FR engines, all 166, 119, 171, and 85 FR engines were used.

Fig. 8 shows the impact of selecting FR engines and weighting FR results for collaborative FR scenario 1. The 1-Hit rates are averaged over the four query sets. In Fig. 8, for the case of using a merging strategy with FR engine selection, the selection of FR engines has been performed prior to merging the FR results. Also, we set $\alpha = 0.4$ when weighting FR results for both BDRF and CMVF. On the other hand, when not weighting the FR results, α was set to zero regardless of the selection of FR engines. That is, *non-weighted* FR result scores are only considered in the merging step. For comparison purposes, the 1-Hit rates obtained for independent FR are also provided. From Fig. 8(a) and (b), the following three observations can be made: 1) the exclusion of the FR engine selection step yields much worse face annotation accuracy compared to the case where the selection of FR engines is incorporated
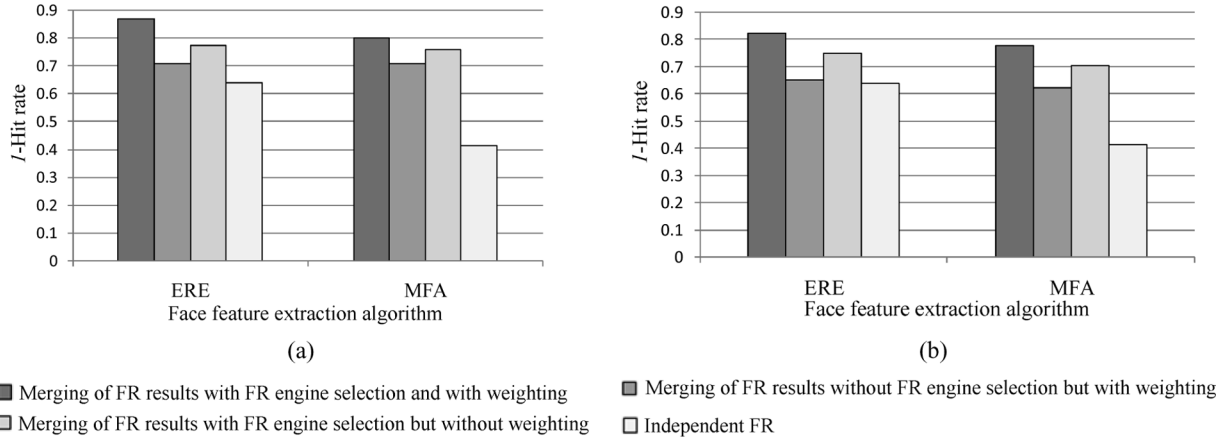
Fig. 8. Impact of selecting FR engines and weighting FR results for collaborative FR scenario 1. (a) BDRF. (b) CMVF.
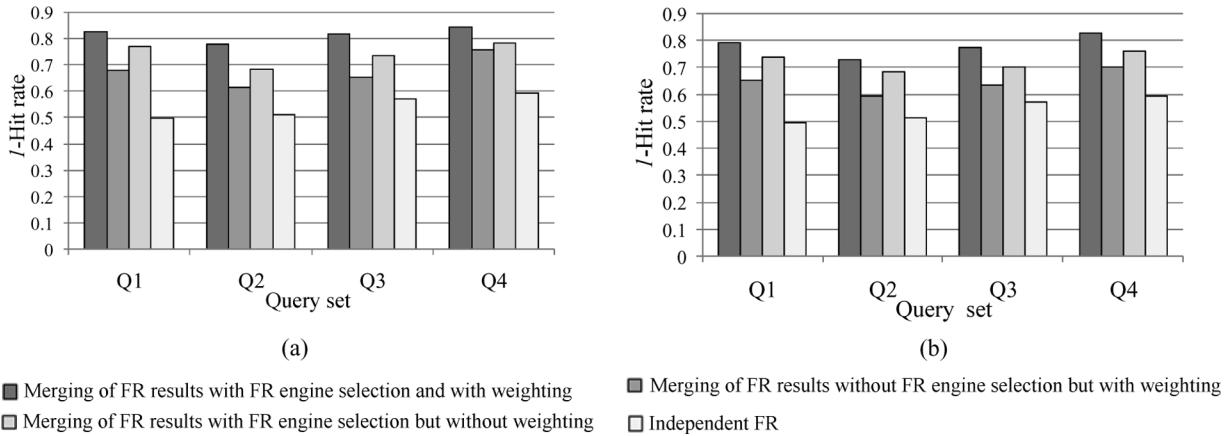

Fig. 9. Impact of selecting FR engines and weighting FR results for collaborative FR scenario 2. (a) BDRF. (b) CMVF. Note that $\alpha = 0.4$ when including the FR engine selection step, while $\alpha = 0$ when skipping the FR engine selection step.

into both merging strategies; 2) for both merging strategies, an improvement in face annotation accuracy can be achieved by imposing weights on FR results prior to their fusion; and 3) the proposed collaborative FR methods allow for a much higher accuracy than baseline independent FR, even when either the selection of FR engines or the weighting of FR results is excluded.

Fig. 9 shows the impact of selecting FR engines and weighting FR results for collaborative FR scenario 2. The different face feature extraction algorithms used and the approach to randomly assign a face feature extractor to each personalized FR engine are exactly the same as those discussed in Section VII-B. Being similar to the results in Figs. 8 and 9 makes it clear that the effect of the FR engine selection step on the face annotation accuracy is significant. In addition, weighting FR results helps improving the face annotation accuracy when making use of a fusion strategy in collaborative FR scenario 2 (for all query sets used).

D. Runtime Performance

To evaluate the efficiency of our collaborative FR framework, we measured the face annotation execution times over a dataset of 3200 face images (part of the Q2 query set) on a computer

with an Intel Pentium IV 2.4-GHz CPU. The processing time needed to select FR engines and to merge FR results was included in the total execution time. On contrary, the processing time needed to create each of the FR engines was not included in the total execution time as the construction of FR engines can be regarded as an offline process in the proposed collaborative FR framework.

Fig. 10 shows the execution time needed by centralized FR and the execution time needed by the two collaborative FR methods for annotating the query face images. As can be seen in Fig. 11, for collaborative FR using BDRF, we are able to achieve execution times of about 164, 554, and 974 s when annotating a total of 3200 face images, respectively using 10, 47, and 91 FR engines. For collaborative FR using CMVF, the resulting execution times are about 110 and 220 s when using 47 and 91 FR engines, respectively.

Compared to centralized FR, the execution time of collaborative FR highly depends upon the number of selected FR engines used when merging FR results. To discover the relationship between the execution time needed by collaborative FR and the number of FR engines used in both merging strategies, the execution time needed to annotate 3200 face images is shown in Fig. 11 for a varying number of selected FR engines. Fig. 11 illustrates that the time complexity of collaborative FR is found
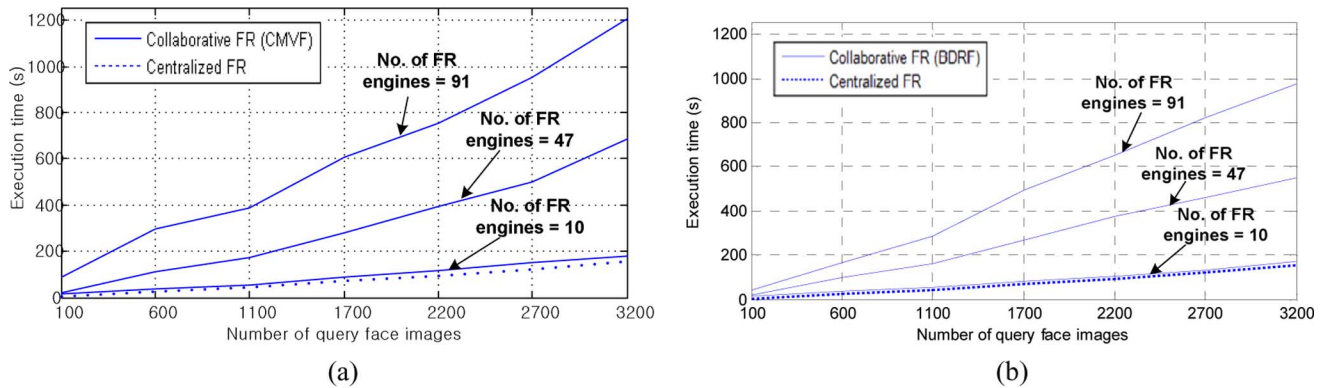
Fig. 10.   Execution times needed for annotating 100 to 3200 query face images. (a) BDRF. (b) CMVF.
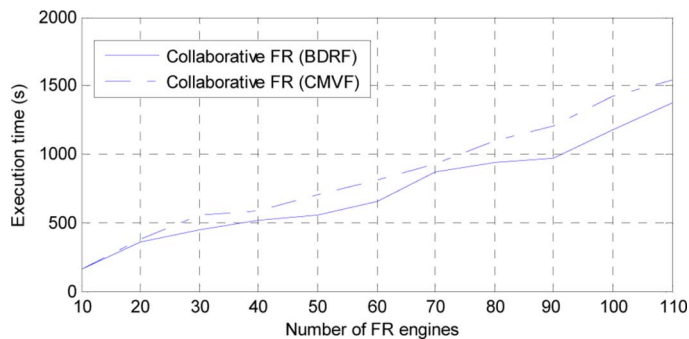

Fig. 11.   Execution time needed by collaborative FR for annotating 3200 face images with a varying number of FR engines.

to be approximately linear, that is, the time complexity of collaborative FR is of order $O(K)$, where K denotes the number of selected FR engines.

VIII. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

This paper demonstrates that the collaborative use of multiple FR engines allows improving the accuracy of face annotation for personal photo collections shared on OSNs. The improvement in face annotation accuracy can mainly be attributed to the following two factors.

1) 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.

2) 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.

We conclude this paper by suggesting a number of extensions and improvements. First, an increasing amount of attention is paid to the creation of *decentralized* OSNs, which are able to overcome issues related to privacy, security, and data ownership [14], [40], [41]. We believe it is likely that each user in a decentralized OSN will have a personalized FR engine. In addition, it

can be expected that these personalized FR engines are shared by members of an OSN. Hence, we believe that our collaborative FR framework is well-suited for use in decentralized OSNs.

Second, our approach for constructing an SGM relies on the availability of manually tagged photos to reliably estimate the identity occurrence and co-occurrence probabilities. In practice, however, two limitations can be identified: 1) individuals are only part of the constructed SGM when they were manually tagged in the personal photo collections; 2) our approach may not be optimal when manually tagged photos do not clearly represent the social relations between different members of the OSN (e.g., due to a limited availability of tagged photos and non-reliable name-tagging). To overcome the aforementioned limitations, social relations between members of an OSN and their strength could be modeled using other sources of information. The number of e-mail or instant messages exchanged between OSN members is for instance a rich source of information. In addition, alternative solutions exist for constructing an SGM. For example, as described in [3], we could effectively predict the SGM without using any manually tagged photos by applying unsupervised FR and graph clustering algorithms on personal photo collections.

Finally, in this work, the *Viola-Jones* face detection algorithm was used for detecting face images in personal photos. In practice, however, the accuracy of the *Viola-Jones* face detection algorithm may be problematic depending upon the targeted applications as well as the associated parameter setup, as described in [42]. Hence, more advanced face detection techniques could be used in our face annotation framework (such as techniques that are robust against severe pose variation [10]), thus allowing for more accurate face detection results.

REFERENCES

[1] [Online]. Available: http://www.facebook.com.
[2] [Online]. Available: http://www.myspace.com.
[3] P. Wu and D. Tretter, "Close & close: Social cluster and closeness from photo collections," in *Proc. ACM Int. Conf. Multimedia*, 2009.
[4] A. Gursel and S. Sen, "Improving search in social networks by agent based mining," in *Proc. Int. Joint Conf. Artificial Intelligence*, 2009.
[5] J. Zhu, S. C. H. Hoi, and M. R. Lyu, "Face annotation using transductive kernel fishder discriminant," *IEEE Trans. Multimedia*, vol. 10, no. 1, pp. 86–96, 2008.
[6] Z. Stone, T. Zickler, and T. Darrell, "Autotagging Facebook: Social network context improves photo annotation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2008.

[7] K. Choi and H. Byun, "A collaborative face recognition framework on a social network platform," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (AFGR)*, 2008.

[8] N. O'Hare and A. F. Smeaton, "Context-Aware person identification in personal photo collections," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 220–228, 2009.

[9] J. Y. Choi, W. De Neve, Y. M. Ro, and K. N. Plataniotis, "Automatic face annotation in photo collections using context-based unsupervised clustering and face information fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 10, pp. 1292–1309, Oct. 2010.

[10] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, 2002.

[11] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color face recognition for degraded face images," *IEEE Trans. Syst., Man, Cybern. B*, vol. 39, no. 5, pp. 1217–1230, 2009.

[12] 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.

[13] K. Lerman, A. Plangprasopchok, and C. Wong, "Personalizing image search results on flickr," in *Proc. Int. Joint Conf. Artificial Intelligence*, 2007.

[14] 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.

[15] M. Naaman, R. B. Yeh, H. G. Molina, and A. Paepcke, "Leveraging context to resolve identity in photo albums," in *Proc. ACM Int. Conf. ACM/IEEE-CS Joint Conf. on Digital Libraries*, 2005.

[16] B. Shevade, H. Sundaram, and L. Xie, "Modeling personal and social network context for event annotation in images," in *Proc. Int. ACM/IEEE-CS Joint Conf. Digital Libraries*, 2007.

[17] 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.

[18] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distribution in empirical data," *SIAM Rev.*, vol. 51, pp. 661–703, 2009.

[19] A. K. Jain, A. Ross, and S. Prabhaker, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004.

[20] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1886, Aug. 2009.

[21] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognit.*, vol. 39, no. 6, pp. 1746–1762, 2006.

[22] C. L. Liu, "Classifier combination based on confidence transformation," *Pattern Recognit.*, vol. 38, no. 11, pp. 11–28, 2005.

[23] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, 1998.

[24] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005.

[25] [Online]. Available: http://www.cyworld.co.kr/.

[26] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Int. Conf. CIVR*, 2001.

[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. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[28] P. Wang, M. B. Green, and Q. Ji, "Automatic eye detection and its validation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition Workshops*, 2005.

[29] [Online]. Available: http://overstated.net/2009/03/09/maintained-relationships-on-facebook.

[30] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 166–178, 2006.

[31] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, 2008.

[32] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, 2007.

[33] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 117–126, 2003.

[34] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, no. 11, pp. 1771–1782, 2000.

[35] P. N. Belhumeur, J. P. Hesphanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 7, pp. 711–720, 1997.

[36] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[37] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognit. Lett.*, vol. 25, no. 12, pp. 1421–1430, 2004.

[38] 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.

[39] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3079–3087, 2003.

[40] A. Vlachou, C. Doulkeridis, D. Mavroeidis, and M. Varzirgiannis, "Designing a peer-to-peer architecture for distributed image retrieval," in *Proc. LNCS Int. Conf. Adaptive Multimedia Retrieval: Retrieval, User, and Semantics*, 2008.

[41] M. Bender, T. Crecelius, and M. Kacimi, "Peer-to-peer information search: Semantic, social, or spiritual?," *IEEE Data Eng. Bull.*, vol. 30, no. 2, pp. 51–60, 2008.

[42] J. Lu and K. N. Plataniotis, "On conversion from color to gray-scale images for face detection," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition Workshops*, 2009.

[43] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.

Jae Young Choi received the B.S. degree from Kwangwoon University, Seoul, Korea, in 2004 and the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2008. He is currently pursuing the Ph.D. degree at KAIST.

In 2007, he worked as a research intern for the Electronics and Telecommunications Research Institute (ETRI) in Korea. In 2008, he was a visiting student researcher at the University of Toronto, Toronto, ON, Canada. His research interests include face detection/recognition/tracking, image/video indexing, pattern recognition, machine learning, the Social Web, and computer vision.

Wesley De Neve received the M.Sc. degree in computer science and the Ph.D. degree in computer science engineering from Ghent University, Ghent, Belgium, in 2002 and 2007, respectively.

He is currently working as a senior researcher for the Image and Video Systems Lab (IVY Lab), in the position of Assistant Research Professor. IVY Lab belongs to the Department of Electrical Engineering of the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. Prior to joining KAIST, he was a post-doctoral researcher at both Ghent University-IBBT in Belgium and the Information and Communications University (ICU) in Korea. His research interests and areas of publication include the coding, annotation, and adaptation of image and video content, GPU-based video processing, efficient XML processing, and the Semantic and Social Web.

Konstantinos N. (Kostas) Plataniotis (S'90–M'92–SM'03) received the B.Eng. degree in computer engineering from the University of Patras, Patras, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from the Florida Institute of Technology (Florida Tech), Melbourne, FL, in 1992 and 1994, respectively.

He is a Professor with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto (UofT) in Toronto, ON, Canada, and an Adjunct Professor with the School of Computer Science at Ryerson University, Toronto. He is the Director of the Knowledge Media Design Institute (KMDI) at the University of Toronto, and the Director of Research for UofT's Identity, Privacy and Security Institute. His research interests include biometrics, communications systems, multimedia systems, and signal and image processing.

Dr. Plataniotis is the Editor-in-Chief (2009–2011) for the IEEE SIGNAL PROCESSING LETTERS, a registered professional engineer in the province of Ontario, and a member of the Technical Chamber of Greece. He is the 2005 recipient of IEEE Canada's Outstanding Engineering Educator Award "for contributions to engineering education and inspirational guidance of graduate students" and the co-recipient of the 2006 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award for the paper entitled " Face recognition using kernel direct discriminant analysis algorithms" published in 2003.

Yong Man Ro (SM'98) received the B.S. degree from Yonsei University, Seoul, Korea, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea.

In 1987, he was a visiting researcher at Columbia University, New York, and from 1992 to 1995, he was a visiting researcher at the University of California, Irvine. He was a research fellow at the University of California, Berkeley, and a visiting professor at the University of Toronto, Toronto, ON, Canada, in 1996 and 2007, respectively. He is currently holding the position of full Professor at KAIST, where he is directing the Image and Video Systems Lab. He participated in the MPEG-7 and MPEG-21 international standardization efforts, contributing to the definition of the MPEG-7 texture descriptor, the MPEG-21 DIA visual impairment descriptors, and modality conversion. His research interests include image/video processing, multimedia adaptation, visual data mining, image/video indexing, and multimedia security.

Dr. Ro received the Young Investigator Finalist Award of the ISMRM in the USA in 1992 and the Scientist of the Year Award in Korea in 2003. He served as a TPC member of several international multimedia conferences, including IWDW, WIAMIS, AIRS, and CCNC. He was also the co-program chair of IWDW 2004.