

Efficient Propagation for Face Annotation in Family Albums

Lei Zhang, Yuxiao Hu, Mingjing Li, Weiying Ma, Hongjiang Zhang

Microsoft Research Asia

49 Zhichun Road, Beijing 100080, China

+86-10-62617711

{leizhang, i-yuxhu, mjli, wyma, hjzhang}@microsoft.com

ABSTRACT

In this paper, we propose and investigate a new user scenario for face annotation, in which users are allowed to multi-select a group of photographs and assign names to these photographs. The system will then attempt to propagate names from photograph level to face level, i.e. to infer the correspondence between name and face. Given the face similarity measure which combines methodologies from face recognition and content-based image retrieval, we formulate name propagation as an optimization problem. We define the objective function as the sum of similarities between each pair of faces of the same individual in different photographs, and propose an iterative optimization algorithm to infer the optimal correspondence. To make the propagation result reliable, a reject scheme is adopted to reject those with low confidence scores. Furthermore, we investigate the combination and alternation of browsing mode for propagation and viewer mode for annotation, so that each mode can benefit from additional inputs from the other mode. The experimental evaluation has been conducted within a typical family album of over one thousand photographs and the results show that the proposed approach is effective and efficient in automated face annotation in family albums.

Categories and Subject Descriptors

I.4.8 [Image Processing and Computer Vision]: Scene Analysis - *Object recognition*; I.5.4 [Pattern Recognition]: Applications - *Computer vision*.

General Terms

Algorithms, Management, Experimentation.

Keywords

Face annotation, propagation, content-based image retrieval, face recognition.

1. INTRODUCTION

The rapid development of digital cameras has significantly

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'04, October 10–16, 2004, New York, New York, USA.

Copyright 2004 ACM 1-58113-893-8/04/0010...\$5.00.

increased accumulation rate of family photographs. Unfortunately, though there are many commercial products available, annotating semantic content of photographs required in organizing photographs, a tedious task, is still left to users. Motivated by the need to automatically organize and retrieve images from large collection, content-based image retrieval (CBIR) has been studied for over decades. Partially because it targets at solving the general image retrieval problems, CBIR research efforts have not resulted in effective and practical solutions to automatic family photograph organization and management.

For typical family albums, besides *when* and *where*, the other most common method for finding and organizing photographs is by subject matter, i.e. what is in the photograph? By far the most common thing that users take photos of is the people in our lives. However, although people are very adept at identifying and recognizing faces, this has remained a very challenging problem in the world of computing. Though efficient and robust face detection algorithms [6, 16, 18] have become available, the effectiveness of available face recognition algorithms is still limited to images of mug shots in which faces are mostly in frontal and with reasonably homogenous lighting conditions and small variations in facial expressions [3, 12, 22]. To overcome the difficulties in face recognition, a Bayesian approach to face annotation in family albums was proposed in [21]. This work integrated methodologies from both content-based image retrieval and face recognition and the face similarity measure is defined as *maximum a posteriori* (MAP) estimation. In this system, if a user desires to label a face in a photograph, he/she needs to open that photograph and moves the mouse onto that face. The system will generate a candidate name list for that face. Then the user can either select one of the names to annotate the face, or set a new name for that face. One major disadvantage of the work presented in this system is that it requires users to view photographs one by one.

In this paper, we present our work to extend and improve the previous work [21]. To further simplify the face annotation, we target at a new user scenario. Rather than open each photograph, a user can look at the thumbnails to see who is in each photograph, uses his/her mouse to multi-select all of the photographs with a particular individual that he/she wants to label. Compared to the scenario in [21], this interface further reduce the users' labeling efforts, because the function of browsing thumbnails becomes common in most photograph management systems, and the user works on photographs level, instead of face level. Thus users can label a group of photographs in a batch way.

Having names associated with a number of photographs, the proposed system will attempt to propagate names from photograph level to face level, i.e. to infer which face in a photograph corresponds to the name associated with this photograph. This problem is similar to relevance feedback problem or keywords propagation problem in region-based image retrieval. However, this problem does not require a user to first submit a query. Because users have already been able to search photographs by names associated with them, it is not necessary for users to infer the correspondence between face and name. However, correspondence inference can facilitate to improve later annotation accuracy.

Given the face similarity proposed in [21] for family albums, we formulate the propagation problem as an optimization problem. We define the objective function as the sum of similarities between each pair of faces of the same individual in different photographs. Maximization of the objective function leads to the optimal solution for name propagation. To make the system more effective in annotation and propagation, we further take advantages from users' interactive operations between annotation and propagation. That is, the name propagation accuracy can be improved if some faces have annotated before, and similarly, the annotation accuracy can be improved if some photographs have been associated with names.

The rest of this paper is organized as follows. Section 2 briefly reviews the related works on both face recognition and content-based image retrieval. Section 3 presents the proposed framework, with a detailed description of the proposed algorithms. In Section 4, we describe the experiment setting and the evaluation result of the proposed algorithm. Thereafter, we will give concluding remarks in Section 5.

2. RELATED WORKS

In this section, we briefly review related works of face recognition and content-based image retrieval, and describe the relationship between the proposed approach and previous works.

Face recognition has been extensively studied for over twenty years. However, although many of the published algorithms have demonstrated excellent recognition results, often with error rates of less than ten percent, they still fall short from the desired accuracy and robustness of commercial applications, including family photo albums. Variations in pose, illumination, and expression are the bottleneck. The face recognition vendor test (FRVT) 2002 [12] also reports that pose and outdoor lighting still remain challenging even in the most commercially successful face recognition systems. For example, for the best face recognition systems, the recognition rate for faces captured outdoors, at a false accept rate of 1%, was only 50% [12]. However, variations in pose and illumination are very common in family albums. It is impractical to assume faces are frontal and in homogenous lighting in photographs taken by consumers.

Although the recognition accuracy can be improved by view-based methods or 3D face recognition approaches, it is generally

very difficult for such approaches to recover the pose when input faces are in exaggerated illumination and expression. Consequently, such approaches are still in research stage, not practical for real applications in family albums.

A variation from face recognition is the problem of similar face retrieval. Gudivada et al. proposed a framework to retrieve face images from face database based on a set of semantic attributes [5]. The focus of that work was on the retrieval method and face database was used merely as a prototype database for testing the algorithm. Baker built a mug-shot search system that adopts CBIR techniques in searching for faces based on eigenfaces in the face database [1]. Satoh et al. developed a system called *Namt-It* to associate faces and names in news videos [14]. They employed the eigenface-based method to evaluate face similarity, and the co-occurrence factors to associate names and faces. Navarrete et al. proposed an interactive face retrieval approach using self-organizing maps [11] to search a face without having an explicit image of it but only its remembrance.

Since the early 1990's, content-based image retrieval has been an active research topic [13]. Most of the early researches had been focused on algorithms of using low-level visual features, such as color, texture and shape, to define similarities between images. But it is extremely difficult to find a mapping between the high-level semantics and low-level features in images. Although relevance feedback, an interactive learning mechanism, is introduced to bridge the large gap, the improvement is still limited yet it requires many user interactions. Recently, there is an emerging trend on image annotation [2, 8] to propagate keywords to images or regions. Partially because they target at solving the general image retrieval problems, such research efforts have not resulted in effective tools useful in automatic family photo organization and management.

In recent years, with the rapid development of digital image acquisition technologies, there is a growing strong need of digital photo album tools to help users organize and manage digital photographs. The MyPhotos system [15] searches for images in the photo album by keywords and visual features. It also provides functions to detect faces in images, but no automated face annotation functions. To provide a more robust solution, content-based image retrieval techniques were introduced to face annotation in [4] and [21]. However, the work in [4] is primarily focused on the CBIR technologies related to feature extraction and similarity measure, whereas face recognition technologies are not well integrated. In the work [21], the face annotation was reformulated from a pure recognition problem to a problem of similar face search and annotation and a solution is proposed by integrating content-based image retrieval and face recognition algorithms in a Bayesian framework.

In the work presented in this paper, we propose a new approach to supporting broader user scenarios. We have formulated batch face annotation as an optimization problem and developed a solution to it. Face annotation will be more effective by taking advantages from alternations between annotation and propagation.

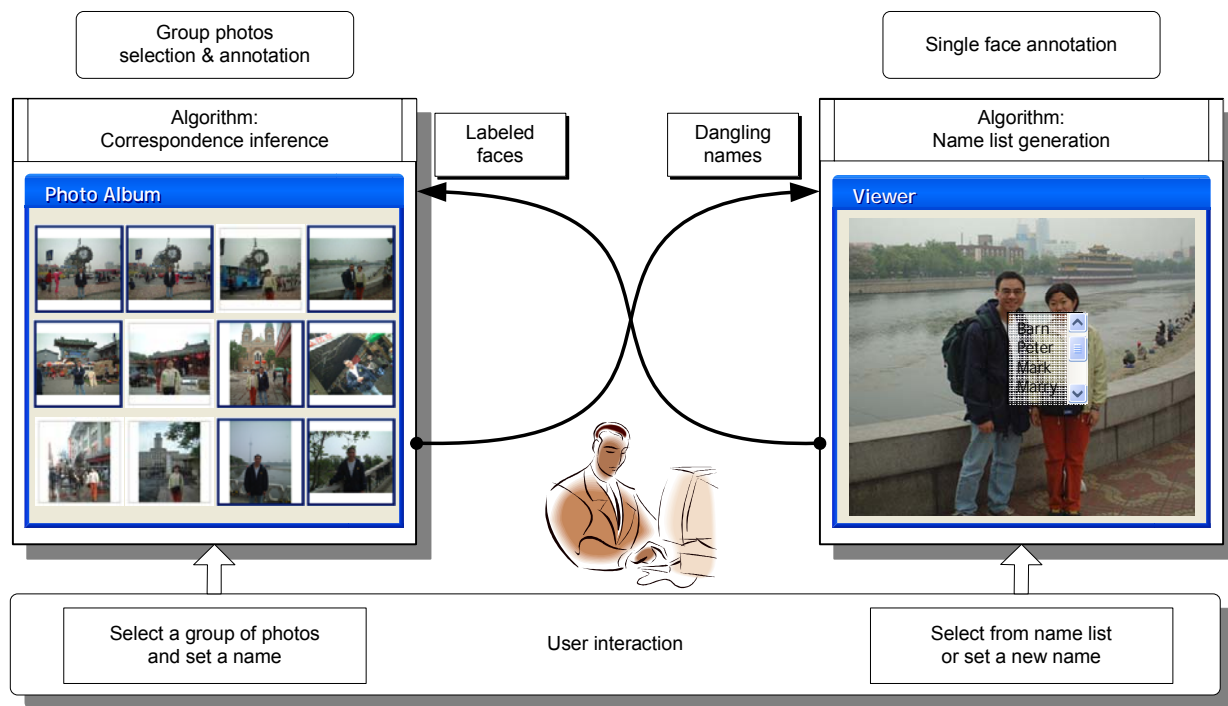


Figure 1. Overview of face annotation and propagation

3. Efficient face annotation and propagation

3.1 Overview

In [21], face annotation was conducted by users on image viewer mode. That is, if a user desires to label a face, he/she moves mouse onto that face, and a candidate name list will popup to provide a recommendation, as shown in the right block in Fig.1. The user can either select a name from the name list, or set a new name for that face. Algorithm wise, the system tries to generate the candidate name list from historical labeling results.

In contrast, in the proposed new approach, the browsing mode is supported for face annotation and name propagation. As shown in the left block in Fig. 1, a user can select a number of photographs in the browsing mode and then perform a one-click assignment of a name to all of those photographs. Algorithm wise, the system tries to infer the correspondence between name and face, i.e. propagate the name from image level to face level.

Traditionally, thumbnails are generated by directly down-sampling the original images. However, thumbnails generated by this approach are difficult to recognize, especially when the thumbnails are very small. To overcome such limitation, we employ a technique called “smart thumbnail” to facilitate browsing experiences [9, 17]. With the detected faces and attention areas, the system automatically generates smart thumbnails. Initially, the system display traditional down-sampled thumbnails in browsing mode. When a user moves mouse onto an image, the system automatically switches the thumbnail to a smart thumbnail, displaying the most informative part of the image, i.e. face areas, to the user. If the mouse pointer hovers on the image for few seconds, the system will automatically animate the image with an optimal browsing path to

let users browse through the different parts of an image. In this way, users will rarely suffer from recognizing individuals from small thumbnails.

With viewer mode and browsing mode, face annotation is conducted naturally and efficiently. However, the process could be more efficient if we integrate the two, because users may frequently alternate between viewer mode and browsing mode. For name list generation algorithm in viewer mode, there have some additional inputs from browsing mode, i.e. dangling names on images. For name propagation algorithm in browsing mode, there also have some additional inputs from viewer mode, i.e. labeled faces in some images. Taking into account these additional inputs, face annotation will be more efficient. In addition, given the face similarity measure function, the system also provides the function of similar face retrieval. In this way, users are allowed to search similar faces by specifying either a face or a name and then annotate multiple faces in a batch way.

3.2 Face Representation

As current face recognition algorithms are not robust enough for face annotation in family album systems, in the proposed face annotation framework, contextual features are incorporated in addition to those used in classical face recognition algorithms, as proposed in [21]. Therefore, each face will be represented by both facial appearance feature and contextual feature.

Following the same approach as in [21], facial appearance features are extracted based on face detection result. Because face appearance features are most reliable when extracted from frontal faces, a texture-constrained active shape model [19] is applied to determine if an input face is in frontal view [21]. If the face is not in frontal, the face appearance feature is treated as a missing

feature. After this process, each face is geometrically aligned into the standard normalized form to remove the variations in translation, scale, in-plane rotation and slight out-of-plane rotation. Then facial appearance features are extracted from normalized gray face images.

The contextual features are extracted from the extended face region as in [21]. Compared to contextual features used in [21], we add color moment feature in LUV color space to compensate the lack of global color feature. By dividing the extended face region into 2×1 blocks, local regional features are extracted to capture the structural information of body patches. As in [21], we also restrict that the date difference between two photographs must be within two days when comparing their contextual similarity.

3.3 Similarity Measure

In [21], face similarity is defined as *maximum a posteriori* (MAP) estimation.

Let $F = \{f_i | i = 1, \dots, N_f\}$ denote the feature set, where each feature f_i is a vector corresponding to a specific feature described in the previous section.

By introducing two classes of face variations, intra-personal variations Ω_I and inter-personal variations Ω_E [10], the similarity between two faces is defined as:

$$S(F_1, F_2) = \frac{\prod_{j=1}^{N_f} p(\Delta f_j | \Omega_I) p(\Omega_I)}{\prod_{i=1}^{N_f} p(\Delta f_i | \Omega_I) p(\Omega_I) + \prod_{i=1}^{N_f} p(\Delta f_i | \Omega_E) p(\Omega_E)} \quad (1)$$

where $p(\Omega_I)$ and $p(\Omega_E)$ are the *a priors*, $\Delta f_i = (f_{i1} - f_{i2})$, and $p(\Delta f_i | \Omega_I)$ and $p(\Delta f_i | \Omega_E)$ are the likelihoods for a given difference Δf_i .

This similarity function integrates multiple features into a Bayesian framework. In case there are missing features, marginal probability is used so that samples which have missing features can be compared with those having the full feature set to ensure a non-biased decision [21].

Based on this similarity measure, name candidates for a given unknown face can be derived by statistical learning approaches, such as K nearest neighborhood algorithm.

3.4 Batch Annotation and name propagation

Candidate name list generation greatly reduces users' efforts in face annotation and empowers users in indexing and searching their albums. However, rather than opening each photograph, it is more convenient for users to look at thumbnails to see who is in each photograph and perform a one-click assignment of a name to multiple photographs. In most cases, users browse their albums in the order of folder, which is closely related to time and event. Beyond this, the browsing could be more effective by means of similar face retrieval. In either browsing mode, batch annotation can be conducted and then name propagation is employed to propagate names from image to face.

3.4.1 Problem statement and formulation

Given the face similarity measure defined in Eq.1, we formulate name propagation as an optimization problem.

Assume a user selected N photographs, denoted by $I = \{I_1, I_2, \dots, I_N\}$ and assign a name ϑ , e.g. "Peter", to these N photographs, which means that each photograph I_i contains ϑ . Let $F^i = \{F_1^i, \dots, F_{C_i}^i\}$ be the faces (individuals) in photograph I_i . The name propagation problem is to select N faces denoted by $\Theta = \{f_1, f_2, \dots, f_N\}$ from N photographs (one face from one photograph), and assign the name ϑ to these N faces.

We define the objective function as the sum of similarities between each pair of faces of the same individual in different photographs, because it is supposed that these photographs share the same property in term of specified individual, and the faces of the individual should be similar with each other. Therefore, the objective function can be formulated as:

$$Sim(f_1, f_2, \dots, f_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N S(f_i, f_j) \quad (2)$$

Maximization of the objective function leads to the optimal solution for name propagation.

$$\{f_1, f_2, \dots, f_N\} = \arg \max_{f_1, f_2, \dots, f_N} Sim(f_1, f_2, \dots, f_N) \quad (3)$$

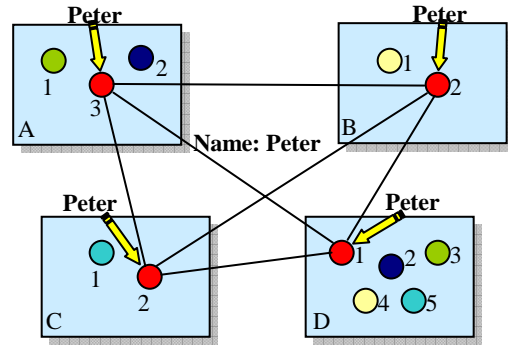


Figure 2. An illustration of optimal solution for name propagation

For example, as illustrated in Fig.2, a user multi-selected four photographs and assigned a name "Peter" to these photographs. That is, the user informs the system that each one of the four photographs contains "Peter". Based on the similarity measure between two faces, the system tries to infer the correspondence between face and the name "Peter". Totally, there are $3 \times 2 \times 2 \times 5 = 60$ possible solutions for face selection from each photograph. In Fig.2, because A3, B2, C2, D1 are similar with each other (illustrated by red color), they are assigned with the name "Peter".

It is worth noting that in some cases the objective function may not be able to result in a correct solution. For example, if there are two individuals jointly appeared in the user selected photographs, there is no way to infer the correct solution unless the user annotates anyone of the two individuals in the photographs. In this case, rejection scheme will be adopted to avoid wrong propagation.

3.4.2 Iterative optimization algorithm

It is an NP-hard problem to find the optimal solution as defined in Eq.3 for name propagation. A brute force approach is to enumerate all the possible solutions and select the optimal solution in terms of maximal similarity defined in Eq.2. However, the enumerative approach will result in combination explosion

problem. For example, if a user selected 20 photographs and there are two faces in each photograph, the number of possible solution will be $2^{20}=1,048,576$.

To avoid combination explosion, we propose an iterative approach to this optimization problem. The algorithm first selects an initial solution, and then iteratively and greedily adjusts faces in each photograph to obtain a solution with larger similarity, until the increasing of similarity approaches to be stable. The algorithm is detailed as follows:

1. Given N photographs $\{I_1, I_2, \dots, I_N\}$, where each photograph I_i contains face set $F^i = \{F_1^i, \dots, F_{C_i}^i\}$, and similarity measure $S(F_1, F_2)$.
2. Initialization: select N faces from N photographs (one face from one photograph) as an initial solution $\Theta = \{f_1, f_2, \dots, f_N\}$, and set initial similarity sum $Sim_Old = 0$.
3. For $t = 1, \dots, N$
 In photograph I_t , i.e. face set F^t , select f_t as

$$f_t = \arg \max_{i \in F^t} \sum_{j=1, \dots, N \cap j \neq t} S(f_i, f_j) \quad (4)$$
4. Calculate new similarity sum Sim :

$$Sim(f_1, f_2, \dots, f_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N S(f_i, f_j) \quad (5)$$
5. if $Sim - Sim_Old > \epsilon$, where ϵ is the pre-determined convergence threshold, then set $Sim_Old = Sim$ and goto 3.
6. Otherwise, exit and output solution $\{f_1, f_2, \dots, f_N\}$.

Figure 3. Iterative optimization algorithm

In each iteration, the algorithm greedily searches for new solution and guarantees that objective similarity sum is monotonously increased. Thus the algorithm can at least reach to local optima.

However, the algorithm greatly reduces the computational complexity from $O(C_1 * C_2 * \dots * C_N)$ to $O(m(C_1 + C_2 + \dots + C_N))$, where m is the iteration number. In practice, the system can flexibly select enumerative approach or iterative approach in accordance with the estimation of computational complexity.

3.4.3 Rejection scheme

Users usually expect that propagation results are correct. However, such propagation scheme may occasionally result in wrong propagation result. As it is generally impractical to let users make a double check of the propagation results, it is necessary to design a rejection scheme to reject those propagation results with low confidence scores.

After the optimal solution is obtained, the system will calculate confidence score for each face as follows:

$$Conf_i = 1 - \max_{j \in F^i \cap j \neq i} S(f_i, f_j)$$

Taking into account the most similar face in the same photograph that f_i is located in, the confidence score actually reflects the uniqueness of face f_i in the photograph. If $Conf_i < T_{reject}$, a pre-determined threshold for rejection, then the name will not be propagated to this face, and instead, the name is dangled on image level.

Nevertheless, dangling names are still useful for both photograph search and further face annotation. For photograph search, as names have been associated with photographs, it is straightforward to find these photographs by the associated names. For face annotation in viewer mode, if there is a name associated with the photograph in which a user want to label a face, the candidate name list generation will be more accurate, as the dangling name is actually a strong prior.

3.5 Alternate Annotation and Propagation

Given both viewer mode and browsing mode, users may frequently and smoothly alternate between these two modes. Taking into account additional inputs, such as dangling names or labeled faces, face annotation could be more efficient.

3.5.1 From annotation to propagation

For name propagation algorithm in browsing mode, there have some additional inputs from viewer mode, i.e. labeled faces in some images. For example, among the photographs selected by a user, there may be some faces being labeled before. In this case, the iterative optimization algorithm is adapted to utilize such additional information.

Let P^+ be the positive set of faces labeled with the same name as ϑ and let P^- be the negative set of faces labeled with different name to ϑ . Without losing generality, either P^+ or P^- can be empty. The iterative optimization algorithm in Fig.3 is modified according to the following rules:

1. Faces in P^+ are actually part of the solution in $\Theta = \{f_1, f_2, \dots, f_N\}$. We fix this part of faces, and only change faces $\Theta \setminus P^+$ in the corresponding photographs in step 3.
2. Eq.4 is modified to include the influence from P^- .

$$f_t = \arg \max_{i \in F^t \cap i \notin P^+} \left(\sum_{j=1, \dots, N \cap j \neq t} S(f_i, f_j) - \sum_{j \in P^-} S(f_i, f_j) \right)$$

3. Eq.5 is also modified to include the influence from P^- .

$$Sim(f_1, f_2, \dots, f_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N S(f_i, f_j) - \sum_{i=1}^N \sum_{j \in P^-} S(f_i, f_j)$$

With these modifications, the optimal solution will be close to P^+ while simultaneously far away from P^- .

3.5.2 From propagation to annotation

For name list generation algorithm in viewer mode, there have some additional inputs from browsing mode, i.e. dangling names on images. For example, in the photograph in which a user desires to label a face, a name has been associated with the photograph, yet not been propagated to a face due to low confidence. To utilize dangling names, we adopt an ad hoc strategy by adjusting the prior of dangling name P in photographs I_i as follows:

$$p(P) = \max(p(P), \frac{1}{|F^i|})$$

where $p(P)$ is the prior of individual P estimated from historical labeling results, $\frac{1}{|F^i|}$ is actually the prior of the individual P that

will appear in this photograph, given that $|F^i|$ is the number of faces in photographs I_i . In this way, the dangling name will be

moved forward in the candidate name list and annotation accuracy is improved.

For example, the prior of “Peter” is estimated to be only 10% from previous labeling results and one name “Peter” is dangling on photograph (a) in Fig.4. Obviously, given this photograph and the dangling name, the prior that face A being “Peter” is 33%. We will take 33% as the prior probability of “Peter” to generate candidate name list for unknown face A.

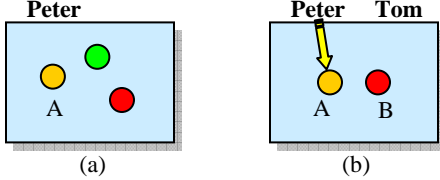


Figure 4. Annotation with dangling names

More importantly, multiple dangling names are also useful for annotation, because any face being annotated will reduce uncertainties in the corresponding photograph. For example, in photograph (b) in Fig.4, there are two dangling names associated with the photograph. Once face A is annotated as “Peter”, the name of face B is immediately determined to be “Tom”.

4. EXPERIMENTS

In this section, we present experimental evaluations of the proposed face annotation and propagation framework with a large set of family photographs.

4.1 Experiment Design

The key of experiment design is to simulate users’ interactions with system. In most cases, users browse their albums in the order of folder, which is closely related to time and event. During browsing process, they may multi-select a group of photographs and assign a name to these photographs. Occasionally, they view some of photos in viewer mode and annotate some faces in the photographs. Based on such scenarios, we designed two experiments to evaluate the performance of proposed algorithm and framework.

Experiment 1: We conduct the first experiment by simulating users’ browsing behaviors to evaluate the performance of batch annotation, or, name propagation. From a group of photographs which are continuous in terms of timeline or event, we randomly select a number of photographs, each of which contains a same individual. Then name propagation algorithm is conducted to infer correspondence between name and face. We repeat such simulation for M times and calculate propagation accuracy on average. With such experiment, we compare the performance between enumerative and iterative optimization approach, and draw receiver operator curves (ROC) to evaluate the rejection scheme.

Experiment 2: The second experiment is similar to the first experiment. However, we will evaluate how much propagation performance can be improved by part of annotation result, i.e. labeled faces, as discussed in section 3.5.1. We assume in each group of photographs selected by users, there are $|P^+|$ positive faces and $|P^-|$ negative faces being labeled already, and then evaluate the performance of name propagation. In this paper,

negative faces refer to faces annotated with names different with the assigned name ϑ to this group of photographs.

4.2 Test Dataset

The test data set we used for the performance evaluation is a typical family album, consisting of over one thousand photographs taken by digital cameras. Therefore, accurate time when each photograph was taken can be extracted from the EXIF header stored in image files. The photographs in the album were taken during 12 months, and in a variety of scenes such as birthday party, wedding, family gathering, and sightseeing.

Among so many individuals appeared in the photographs, we manually labeled more than 20 most frequently appeared individuals as the ground truth, while ignore other individuals who appeared in less than three photographs.

4.3 Performance Evaluation

In the first experiment, we did 500 times of simulations. The simulations vary from different events and individuals across the family album. The average propagation accuracy and total time for 500 times of simulations on a common PC with a 2.7GHz CPU are shown in Table 1. It can be seen clearly that iterative optimization algorithm greatly reduces computational complexity of enumerative approach from 798 seconds to 2 seconds, with comparable propagation accuracy.

Table 1. Comparison between enumerative and iterative optimization approaches

	Enumerative	Iterative
Total Time	798 sec	2 sec
Accuracy	81.3%	78.6%

To demonstrate the effectiveness of the proposed objective function for name propagation, we randomly select ten simulations, and list the optimized objective function values in Eq.2 for different approaches in Table 2. As shown in Table 2, enumerative solutions have the largest objective function value in each simulation, and in most cases the objective function values in both ground-truth solutions and iterative solutions consist with those in enumerative solutions. This table shows that the name propagation is well formulated and the definition of objective function is reasonable.

Table 2. Optimized objective function value comparison

	Photos	Ground-truth Solution	Enumerative Solution	Iterative Solution
1	5	8.689263	8.689263	8.689263
2	6	13.178927	13.178927	13.178927
3	9	31.195576	31.195576	31.195576
4	5	8.660696	8.660696	8.660696
5	7	17.438576	17.564528	17.560561
6	5	8.909868	8.909868	8.909868
7	13	67.413310	67.413310	67.413310
8	6	13.195004	13.257750	13.184114
9	7	17.941117	17.941117	17.941117
10	7	17.926214	18.028435	17.963342

To obtain high propagation accuracy, rejection scheme is evaluated and ROC curves are drawn in Fig.5. The figure clearly

shows that propagation accuracy is increased with the increasing of rejection rate. When rejection rate is 30%, the propagation precision is about 90%. When rejection rate is 43%, the propagation precision is about 95%, which means that users only need to make corrections in a very small number of photographs.

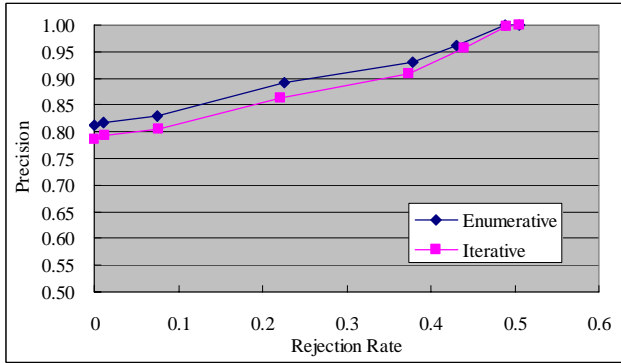


Figure 5. Performance comparison of enumerative and iterative optimization approaches with rejection

After the first experiment, we conducted the second experiment to evaluate the performance improvement if there are additional inputs from annotation to propagation.

In the second experiment, the simulations are the same as in the first experiment, except that in each simulation, we assume there are $|P^+|$ positive faces and $|P^-|$ negative faces being annotated already. As shown in Table 3, name propagation is apparently improved by additional inputs from either positive or negative faces being annotated, because the more the additional inputs, the less the uncertainties of the optimal solution are in a group of photographs. From the experimental result, it can be seen that negative faces also contribute to optimal solutions because they eliminate ambiguities.

Table 3. Performance improvement of propagation with additional inputs from annotation

$ P^+ , P^- $	0,0	1,0	2,0	0,1	0,2	1,1
Accuracy	0.786	0.867	0.946	0.826	0.839	0.902

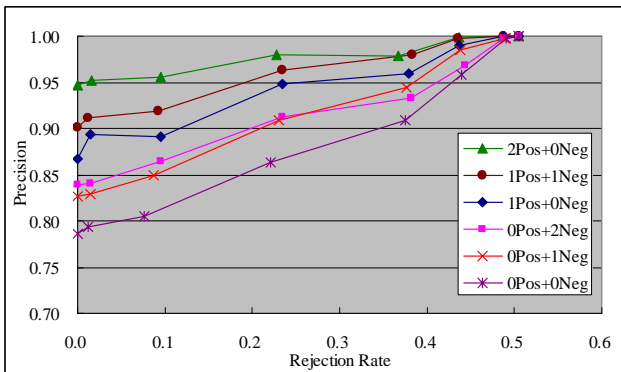


Figure 6. Performance comparison of different additional inputs with rejection. $mPos+nNeg$ denotes m positive faces and n negative faces.

The precision vs. rejection rate of propagation based on different additional inputs from annotation are also illustrated in Fig. 6. Because additional inputs from annotation are considered in the

propagation algorithm, comparable precision rate can be achieved at lower rejection rate. For example, given one positive face being annotated, precision of 95% is obtained when the rejection rate is about 22%, while it requires the rejection rate of about 43% without any additional inputs from annotation.

4.4 Discussion

The success of the proposed scenario and solution can be attributed to three aspects. First, the similarity measure integrates both facial feature and contextual feature, which well captures the characteristics of family albums. Second, the name propagation is well formulated as an optimization problem. The detailed experimental results about maximized objective function values show consistencies among ground-truth solution, enumerative solution, and iterative solution. The iterative optimization is computationally efficient and has comparable performance, which makes name propagation scenario practical. Third, propagation in browsing mode and annotation in viewer mode complement each other and provide flexibilities for users switching between two modes smoothly and naturally.

It is worth noting of an implementation issue in the system. As mentioned in section 3.4, if there are two individuals jointly appeared in the user selected photographs, there is no way to infer the correct solution, as such case conflicts with the assumptions of the algorithm. To deal with such ambiguous case, we adopt a heuristic strategy to efficiently exploit users' interaction. As the iterative algorithm is very fast, whenever a user annotates a face or a group of photographs with a name, the system immediately collect photographs related to that name and taken within the nearby time slot, and performs the group annotation with additional inputs from single annotations. Therefore, if the system infers a wrong result for such ambiguous case, users can conveniently correct the result and immediately see the improved result.

Another issue needed to be further investigated is how to deal with missing faces caused by inaccurate face detection. In the photographs multi-selected by a user, if there is a face not detected by face detector, yet it is just what the user desires to annotate, the propagation algorithm may lead to a wrong prediction in this photograph. Obviously, unless the user manually specifies the location of the missing face, the system will never propagate the name to this missing face. A seemingly better solution is to reject propagation and hold the name at image level. In this paper, we adopt a reject scheme based on the uniqueness of the face within the photograph, as discussed in Section 3.4.3. However, such reject scheme cannot deal with missing faces. Instead of leaving this issue to user interface design, a better way is to consider the dissimilarity between each pair of faces in the optimal solution, and reject the outlier. We will address this issue in the practical system.

We also conducted an informal user study. We asked two users to label 15 typical family albums, each of which contains 50-200 photographs. The users can freely choose functions of single face annotation, similar face retrieval and group face annotation in the prototype system to annotate faces in the albums. In the prototype system, there is a preview panel to display the selected image in large size, and thus there is no explicit boundary between viewer mode and browsing mode. For most cases, the users choose group face annotation to label faces, and then they correct wrong propagation results if there is any. Both users feel that it is very

convenient that the system provides both single and group face annotation, and group face annotation apparently accelerates the labeling process.

5. CONCLUSION AND FUTURE WORK

We have proposed in this paper a new approach to name propagation problem for face annotation in family photo albums. The name propagation is formulated as an optimization problem and the objective function is defined as the sum of similarities between each pair of faces of the same individual in different photographs. Based on such formulation, we proposed an iterative optimization algorithm to infer the optimal correspondence between name and face. The experimental evaluations show that the proposed scenario and solution are practical and efficient for automated face annotation in home photo albums.

However, there still have some problems that need to be further studied. For example, it would be better if the system could impose an explicit face model in the objective function to simultaneously maximize the likelihood of each face to the face model, and the sum of similarities between each pair of faces. However, due to the large face variations in family albums, it is generally difficult to automatically select samples suitable for face recognition algorithm. In the future work, we will continue this direction by incorporating face recognition to further improve the annotation performance.

6. REFERENCES

- [1] Baker E., "The Mug-Shot Search Problem - A Study of the Eigenface Metric, Search Strategies, and Interfaces in a System for Searching Facial Image Data", Ph.D Thesis, Division of Engineering and Applied Sciences, Harvard University, January, 1999.
- [2] Barnard, K., P. Duygulu, D. Forsyth, N. de Freitas, D. Blei, and M.I.Jordan, "Matching Words and Pictures", Journal of Machine Learning Research (JMLR), Special Issue on Text and Images, p.1107-1135, vol. 3, 2003.
- [3] Chellappa R., Wilson C. L. and Sirohey S., "Human and machine recognition of faces: A survey", In Proc. of IEEE, p. 705, vol.83, May, 1995.
- [4] Chen L., Hu B., Zhang L., Li M. and H.J. Zhang, "Face annotation for family photo album management", International Journal of Image and Graphics, Vol. 3, No. 1, p.1-14, 2003.
- [5] Gudivada V. N., Raghavan V. V. and Seetharaman G. S., "An Approach to Interactive Retrieval in Face Image Databases Based on Semantic Attributes", In Third Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, 1993.
- [6] Hsu R. L., Mottaleb M. A. and Jain A. K., "Face Detection In Color Images", In IEEE Trans. Pattern Analysis and Machine Intelligence, p. 969, 24(5), May, 2002
- [7] Huang J., Kumar S. R., Mitra M., Zhu W. J. and Zabih R., "Image indexing using color correlograms", In IEEE Conf. on Computer Vision and Pattern Recognition, p. 762, 1997.
- [8] Li B., Goh K.-S., and Chang E., "Confidence-based Dynamic Ensemble for Image Annotation and Semantics Discovery", in ACM International Conference on Multimedia, pp. 195-206, Berkeley, 2003.
- [9] Liu H., Xie X., Ma W.Y., Zhang H.J., "Automatic Browsing of Large Pictures on Mobile Devices", in ACM International Conference on Multimedia, Berkeley, 2003.
- [10] Moghaddam B. and Pentland A., "Probabilistic visual learning for object representation". IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), p.696-710, 1997.
- [11] Navarrete P. and Ruiz-del-Solar J., "Interactive face retrieval using self-organizing maps", In proc. Int. Joint Conf. on Neural Networks (IJCNN), Honolulu, USA, 2002
- [12] Phillips P.J., *et al.* "Face Recognition Vendor Test 2002 Evaluation Report", <<http://www.frvt.org/FRVT2002/Default.htm>>, 2003.
- [13] Rui Y., Huang T. S. and Chang S., "Image Retrieval: Current Techniques, Promising Directions and Open Issues", Journal of Visual Communication an Image Representation, p. 39, vol.10, March, 1999.
- [14] Satoh S., Nakamura Y., and Kanade T., "Name-it: Naming and detecting faces in news videos". IEEE Multimedia, 6(1), p.22-35, January-March 1999
- [15] Sun Y.F., *et al.* "MyPhotos - A System for Home Photo Management and Processing", in Proc. ACM Multimedia 2002, December, 2002
- [16] Viola P. and Jones M. J., "Robust Real-time Object Detection", Technical Report, COMPAQ Cambridge Research Laboratory, Cambridge, MA, Februray,2001
- [17] Wang M.Y., Xie X., Ma W.Y., Zhang H.J., "MobiPicture - Browsing Pictures on Mobile Devices", in ACM International Conference on Multimedia, Berkeley, 2003, demo.
- [18] Xiao R., Li M.J., Zhang H.J., "Robust Multi-Pose Face Detection in Images", *IEEE Transactions on Circuits and Systems for Video Technology (CSVT), Special Issue on Biometrics*, Volume 14, PART1, 2004.
- [19] Yan S.C., *et al.*, "Texture-Constrained Active Shape Models", In Proceedings of The First International Workshop on Generative-Model-Based Vision. Copenhagen, Denmark. May, 2002.
- [20] Yang M. H., Kriegman D. and Ahuja N., "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), p. 34, 24(1),2002
- [21] Zhang L., Chen L., Li M., Zhang H. "Automated annotation of human faces in family albums", in ACM International Conference on Multimedia, pp.335-358, Berkeley, 2003.
- [22] Zhao W., Chellappa R., Rosenfeld A. and Phillips P., "Face recognition: A literature survey", Technical Report, Maryland University, CfAR CAR-TR-948, 2000.