Consumer Image Retrieval by Estimating Relation Tree From Family Photo Collections

Tong Zhang, Hui Chao, Chris Willis, Dan Tretter

Hewlett-Packard Laboratories
1501 Page Mill Rd. Palo Alto, CA 94304, USA
{tong.zhang; hui.chao; chris.willis; dan.tretter}@hp.com

ABSTRACT
In this paper, we propose an approach to automatically estimate relationship among people in a family image collection based on results from face analyses technologies including automated face recognition and clustering, demographic assessment, and face similarity measurement, as well as contextual information such as people co-appearance, people's relative positions in photos and image timestamps. As the result, a relation tree can be estimated which provides important semantic information regarding people involved in a photo collection and has numerous applications in photo sharing and browsing, social networking, etc. The methods for deriving and integrating information from photos and the process for estimating a relation tree are described. Experimental results on two typical consumer photo collections and examples of using these results in consumer image retrieval are presented.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; I.4.9 [**Image Processing and Computer Vision**]: Applications

Keywords
Consumer image retrieval, family relation tree, face analysis, face clustering, demographic assessment, and social networking.

1. INTRODUCTION
People and their relationships within a family and its associated social circle often can be figured out by analyzing family photo collections. In this paper, we propose an approach and workflow to automatically reveal major characters and their relations in an image dataset, and especially, estimate a relation tree to display the family and social net.

To achieve this goal, we use information obtained by employing multiple automated face analysis technologies, including face recognition and clustering (i.e. grouping photos based on face), demographic assessment of face (e.g. age and gender) and face similarity measurement (i.e. finding people who look alike); as well as contextual information such as people's co-appearances, people's relative positions in photos and timestamps; plus expert knowledge summarized from experiments and observations on consumer image collections. Integrating all the above, a relation tree may be estimated. An example is shown in Figure 1 where members of the nuclear family (husband, wife and kid), the extended family (grand-parents, siblings, cousins, etc.), relatives and friends are identified and placed in order of their relationships. For instance, the nuclear family is put in the middle. Relatives of the husband are put on the right side; while relatives of the wife are put on the left side. All couples are put next to each other, and children are put right under their parents.

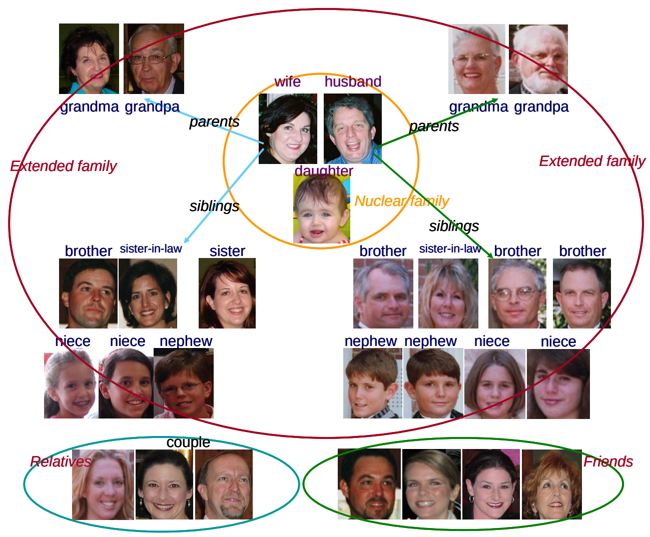


Figure 1: An example of relation tree derived from a family photo collection.

While there have been researches on extracting social relations from other types of media, especially text information such as emails [1], prior work on discovering relationships from consumer photos based on image content analysis is still rare. Golder [2] presented preliminary investigation on measuring social closeness in consumer photos, where a weighted graph was formed based on people co-appearance information. A similar approach was taken in [3]; however, a graph clustering algorithm was used to detect social clusters embedded in photo collections. Comparing with such existing work, our proposed approach aims at a bigger scope, integrating information from multiple aspects of image analyses and revealing more details of social relationships among the people involved in an image collection.

The estimated relation tree can be useful in many ways. First, it provides important semantic information for organizing, browsing and retrieving consumer images. One exemplar application in photo browsing is illustrated in Figure 2, where photos are filtered and sorted according to people's relations, so that photos of family members, relatives (near and far), friends (close and casual), and people in associated groups (soccer, work, etc.) can be easily retrieved and labeled. Further, this information may be used to generate automatic recommendations in selecting pictures for making photo products such as photo books, greeting cards, calendars, as well as slideshows, videos, electronic story-telling, etc. Moreover, the relation tree may have other social and business values. For example, the structure of a family and the demographic information of its members are definitely critical to vendors in making personalized advertisement.

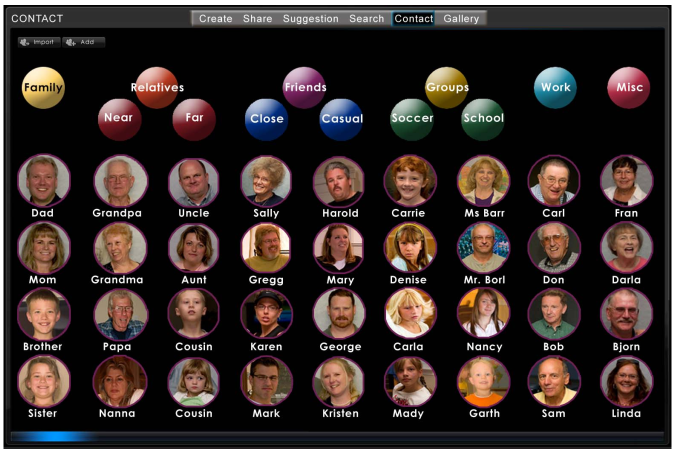


Figure 2: Browsing photos based on people's relations.

The rest of the paper is organized as follows. In Section 2, major cues for deriving people's relations and the methods for extracting them are introduced. The procedure of estimating the relation tree is then described in Section 3. Experimental results of applying the proposed approach on two typical family photo collections are presented in Section 4, followed by conclusions and future work in Section 5.

2. EXTRACTING CLUES FROM IMAGES FOR DERIVING PEOPLE'S RELATIONS

To estimate people's relations, the following methods and clues are employed in this work: face recognition and clustering; demographic estimation; face similarity measurement; co-appearances of people in photos; people's relative positions in the photos; and timestamps of the photos. Details are described below.

2.1 Automatic Face Clustering

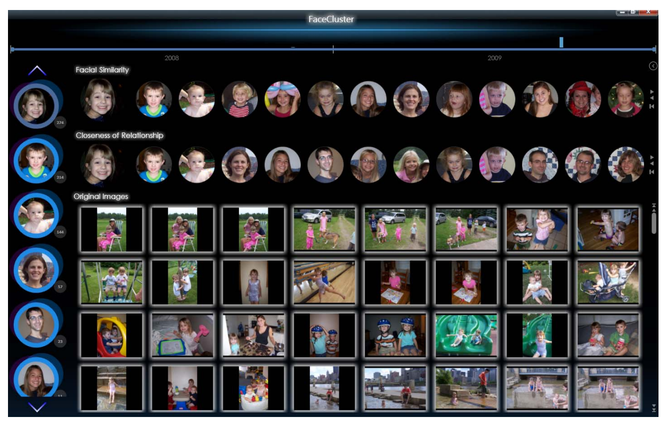
First of all, significant people (or called major characters) in a photo collection need to be identified. On top of an advanced face recognition engine built in our prior work, we have developed a face clustering technology which automatically divides a photo collection into a number of clusters, with each cluster containing photos of one particular person [5, 6]. As the result, major clusters (that is, those having a relatively large number of photos) corresponding to frequently appearing people may be deemed as containing major characters of the collection.

The face-based clustering of photos consists of three steps:

(1) Facial feature extraction and similarity computation. First, for each picture in the dataset, face regions are detected using an Ada-Boosting detector with Harr-like features. The number of false alarms is reduced with the help of a skin-color filter. Next, one facial feature vector is extracted from each detected face, which is a combination of Gabor features and LBP (Linear Binary Pattern) features, with the dimension reduced by PCA and LDA. Then a similarity matrix is obtained by computing the distance between every pair of face feature vectors so that each item in the similarity matrix indicates the similarity between a pair of faces.

(2) Semi-supervised face clustering. We take the agglomerative clustering algorithm. It starts with an initial partition where each face forms a singleton cluster; then the two most similar clusters are selected and merged into one cluster. This merging operation repeats until the similarity between the two closest clusters falls below a stopping threshold. Assuming that faces simultaneously appearing in one photo cannot belong to the same cluster (called CANNOT-links), we employ a semi-supervised clustering method in which CANNOT-links are used to control the procedure of clustering together with the similarity measures. It is especially helpful in separating similar looking people such as siblings.

(3) Face modeling and cluster consolidation. With the above step, a big cluster is formed for each frequently appearing person; but there might also be some smaller clusters of the same person that are not included in the dominant cluster due to relatively low similarity scores. Each large enough cluster is then modeled as a pattern class, and small clusters are matched with these classes in a supervised classification manner. If the recognition succeeds with one class, the small cluster is merged into the corresponding big cluster. The K-nearest neighbor classifier is used here.



On the left: automatically generated face clusters. On the right: first row – clusters similar to selected cluster in face; second row – clusters close to selected cluster in co-appearance. Lower-right panel: images in selected cluster.

Figure 3: Automatic face clustering in a consumer photo set.

Face clustering result of a family photo collection containing 591 pictures is illustrated in Figure 3 (this is the CMU dataset [4]). On the left shows major clusters that were obtained, ranked in order of cluster size. The top five clusters (containing 274, 214, 144, 57 and 23 faces, respectively) belong to the three kids (in the order of age), followed by the wife, and then the husband of the family, respectively, which is typical of family photo collections.

The face clustering method has been tested on over 6000 photos from six family collections labeled with ground truth, and it was found that the clustering performance is quite robust. That is, similar results were obtained for different datasets. Within each cluster, almost all the photos contain the face of one same person, which means the precision rate is 100% or close to it. While one person may have multiple clusters (usually one large cluster and a few small ones), on average around 70% of a major character's photos are grouped in his/her major cluster (i.e. the recall rate). And based on cluster similarity, different clusters of one person can be easily retrieved and merged, reaching an even higher recall rate of 80~90%. In particular, the algorithm proved to be able to cluster photos spanning a long time period (e.g. 10~20 years). More details can be found in [5] and [6].

2.2 Demographic Estimation

Demographic assessment is done within each face cluster, which estimates the person's gender and age. A gender classifier was trained with over 10,000 faces of different ethnicities and various poses, illuminations and expressions from consumer images [7]. Each detected face is first aligned to a common coordinate system with 88 feature points on the face and then warped to a 32×32 rectangle face image. A probabilistic boosting tree (PBT) is trained for the classification. PBT is a divide-and-conquer tree with soft probability boundaries. In this tree structure, each node is a strong classifier trained by an LUT-based Real AdaBoost. Simple Haar-like features of the normalized face are exploited for the AdaBoost classifier. In constructing the boosting tree, the confidence output of a parent node is used to split the training set to sub-trees. The classification boundary is learned step by step. On a test set of 1200 faces from consumer photos, the method achieved a correct classification rate of 85%.

We also developed an age classification algorithm through which people are partitioned into four age categories [8]: baby (0 to 1 year old); child (2 to 16), adult (17 to 50), and senior (above 50). Gabor features [9] are employed to represent the face and linear discriminant analysis (LDA) is used [10] to form the classifier. In the training phase, thousands of frontal or near frontal face images were collected from consumer images with various illuminations and facial expressions. Each face is labeled with an estimated age between 0-80 due to lack of exact age information. Then, Gabor features of 3 scales and 4 orientations are extracted from each face. This amounts to 12 convolved face images, of which only magnitude images are used as the raw features. PCA is applied for dimension reduction. In the classification stage, a fuzzy LDA method is applied. To cope with age ambiguity, an age membership function $\mu_i(x)$ is defined as in Figure 4, where:

$$0 \le \mu_i(x) \le 1, \quad i \in \{1,2,3,4\}, \quad 0 \le x, \quad \sum_{i=1}^{4} \mu_i(x) = 1$$

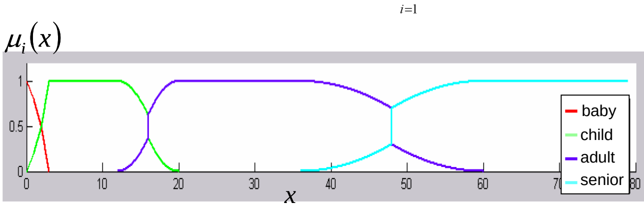


Figure 4: A fuzzy age membership function.

It describes to what extent a face with an estimated age of x is a member of the i^{th} age group. Note that the age ambiguity always happens at the boundary of two adjacent age groups, and each age can belong to at most two age groups. With the age membership function $\mu_i(x)$ defined, LDA is conducted using $\mu_i(x)$ as the class weighting value. This age estimation method achieved an average accuracy rate of 99% on the training set containing 5408 faces, and an average accuracy rate of 92% on the testing set containing faces of 57 babies, 350 children, 492 adults and 79 seniors.

While demographic estimation is not perfect for each individual face, when conducted on a big enough face cluster where reliable statistics of estimation results can be obtained, the cluster-wise result will be much more accurate. Shown in Table 1 is the demographic estimation result on the 14 largest clusters (corresponding to the 14 members of the nuclear family and extended family) of a family photo set.

Table 1: Demographic Estimation of Face Clusters

Face clusters	Ground truth	Detected face	Gender estimation		Age estimation			
			M	F	B	C	A	S
No.1	F, B/C	436	178	258	94	274	61	7
No.2	F, B/C	266	109	157	88	162	14	2
No.3	F, A	247	37	210	2	30	186	29
No.4	M, A	215	171	44	4	44	115	52
No.5	M, C	86	46	40	12	71	2	1
No.6	F, A	72	13	59	1	1	59	11
No.7	M, S	65	49	16	1	7	11	46
No.8	M, S	62	41	21	2	1	3	56
No.9	F, S	57	17	40	0	6	11	40
No.10	M, A/S	40	22	18	1	4	8	27
No.11	F, A	33	2	31	0	0	33	0
No.12	F, S	30	4	26	0	4	4	22
No.13	M, A	18	16	2	0	4	12	2
No.14	M, C	17	13	4	1	13	3	0

In the table, the number of faces classified into each gender or age group within each face cluster is listed, where M, F indicate male and female, respectively; B, C, A and S indicate baby, child, adult and senior, respectively. Since the photos span a period of several years, some subjects may belong to multiple age groups such as B/C and A/S. The numbers in red indicate the gender/age group estimated for the cluster from majority vote, and the numbers in orange indicate the second dominant age group if available. As can be seen, while there are mistakes in estimating gender and age for faces in each cluster (especially, it tends to make mistakes in recognizing the gender of babies and young children, which is understandable), cluster wise, all the classification results match with the ground truth correctly.

Furthermore, photo timestamps may help to obtain more accurate estimation of the current age of a person – the time information is particularly crucial when a person's cluster consists of pictures spanning a long period of time. From both age classification and timestamps, if the rough time range can be found during which a person's dominant estimated age category changed from baby to child, from child to adult, and so on, then an approximation of the person's age can be made. In Figure 5, age estimation results of 266 face images, taken within the years 2005 ~2009, of a person are shown. We can see that pictures taken before May 2006 are mostly classified as "baby"; pictures taken after March 2007 are mostly classified as "child"; and pictures taken between these two times are basically split between the two groups with photos in the baby group becoming sparse along time, which indicates that the person changed from baby to child during this period of time. Therefore, the person can be estimated to be a child of around four-year old in 2009. We expect that a finer age classification based on facial features (i.e. divide age into more categories) may generate more accurate estimation of a person's current age.

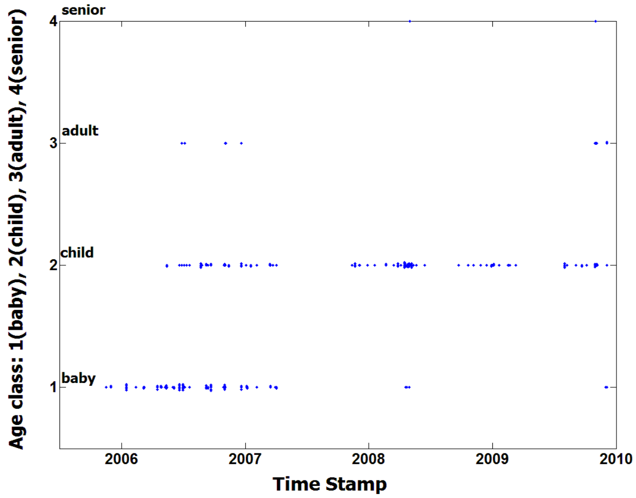


Figure 5: Estimating a person's age based on timestamps.

2.3 Face Similarity Measurement

Once face clusters are obtained, the similarity between each pair of clusters can be calculated using a pre-defined cluster similarity measure. In this work, we use the K-nearest neighbor method. That is, each cluster is represented by all of its member faces with which every face in another cluster is compared, and the distance of the two clusters is determined by the average distance of the K nearest neighbors. Then, for one cluster, other clusters that are similar to it can be ranked according to the similarity measure. Since one cluster contains photos of one person, it shows people who look similar to the selected person.

Figure 6: Face similarity between clusters.

It has been found that clusters of family members with blood relations (e.g. siblings, parents and kids) tend to be similar and rank high on each other's list, which helps to find social relations of people. Illustrated in Figure 6 is one example where one man's cluster is selected from the cluster list shown on the left side of the interface. On the first row of the right side, it shows clusters that are similar to this cluster measured by face similarity. It could be figured out that among the top ranked clusters, there are those clusters belonging to this person's siblings and father based on the fact that they do look like each other and they appeared together in family reunion photos which are shown in the lower-right panel of the interface.

2.4 People's Co-appearances in Photos

People co-appearance in photos is another important information which is a sign of who and who are close to each other. For instance, while a man may look most similar to his parents or siblings, he may be closest to his wife and appear most often with her. Referring again to Figure 6, on the second row of the right side, it shows people who appear most often in photos with the selected person, and the top ranked ones are the man's three kids and his wife, even though none of them appears in the first row. Therefore, co-appearance and facial similarity are compensating information in estimating people's relations.

Figure 7: People's co-appearances indicate relationships.

Besides indicating the closeness between people, co-appearances may disclose more relation information. Figure 7 shows a subset of photos taken during a wedding. Apparently, (a) and (b) are photos of family members with the newlywed, and the couple in the middle are the groom and the bride. Then, in (c) and (d) are photos of the new couple with two senior couples, respectively. These two senior couples also appear next to the new couple in

(a) and (b), respectively. It could be figured out that they are parents of the new couple. Next, there is the photo of one of the senior couples with only the groom in (e), and there is the photo of the man in this senior couple with the groom in (f). Now it is obvious that appearing in (d) and (e) are the groom's parents because it is natural that the groom takes a photo with his parents or his father during the wedding, while it would be weird if the groom takes pictures with his in-laws without the bride. Then, in (b) must be family members of the groom, including his parents, siblings, siblings' spouses, nephews/nieces, etc.; while in (a) and (c) are folks of the bride.

2.5 People's Relative Positions in Photos

People's relative positions in photos may reveal their relations as well. For instance, couples usually stand or sit next to each other, and parents usually hold their own kids, in group photos. Faces that are close to each other indicate intimate relations such as husband and wife, lovers, parent/grandparent and kid, sisters and so on. In wedding photos, when folks of both sides are present, the bride's relatives usually stand at the side next to the bride, while the groom's relatives stand at the other side. Some example pictures are shown in Figure 8 where ground truths of people's relations are marked which validate these observations.

Figure 8: People's positions in photos indicate relations.

Even though the above rules may not be followed in every photo, they are generally true when applied to statistics obtained from face clusters. For example, for a female relative who appears in a number of photos, the male who is next to her in many of the photos is most probably her husband or boyfriend, especially if they appear together in different events, or if their faces are close to each other in some of the photos. Similarly, if one or more kids appear in photos next to one or both of a couple more frequently than to others, and especially if there are photos containing only the kid(s) with the couple, then most likely the couple are parents of the kid(s).

2.6 Photo Timestamps

Besides helping estimating a person's current age as described in Section 2.2, timestamps are useful in judging how close people are with each other as well. For example, by timestamps, we can tell whether someone appears in the collection just during one event, or in different events, or he appears frequently, which indicates how closely the person is associated with the owner (or the family) of the photo collection. Also, from whether a group of people (i.e. two or more people) co-appear just in one event, or they co-appear together at different times or in different events, a measure can be obtained as to the closeness of these people.

Furthermore, for a photo collection spanning many years, time-stamps reveal the dynamics and evolvement of people's relations, as well as events that happened to the family. For example, the addition or lost of an important family member can be discovered when viewing a dynamic family tree that changes along time.

3. THE PROCEDURE FOR ESTIMATING THE FAMILY RELATION TREE

The procedure for estimating people's relationships and deriving the family relation tree has the following five steps.

3.1 Gathering Information from Photos

First of all, all kinds of information useful for estimating people's relations, as introduced in the above section, are extracted from photos. Automatic face clustering is conducted on a family photo collection. Major clusters are identified and major characters are located. If user inputs are available, the clusters may be adjusted with operations such as merging and splitting [6], or even labeled with names. A similarity matrix of clusters is computed in which each item represents the similarity score between two clusters. Based on this matrix, for each major cluster, a list of the top ten clusters that are most similar to it can be generated.

The gender classifier is applied to each face in the major clusters, and through majority vote each cluster is labeled as either male or female. The age classifier is also applied and a plot is obtained for each major cluster similar to the one shown in Figure 5. If only one age group is dominant in the plot and votes for other age groups are negligible, then the cluster is labeled with that age group, i.e. baby, child, adult or senior. If there are two or more age groups that have a significant number of votes, then the over-lapping period(s) of consecutive age groups on the plot is located from which an approximation of the current age can be made. The cluster is labeled with this approximated age.

For every two major clusters, images that are in both clusters contain co-appearance of the two corresponding characters. A co-appearance matrix is built in which each element represents co-appearances of a pair of clusters, which is an array consisting of paths of images that are in both clusters. For each image containing more than one major character (i.e. the image belongs to more than one major cluster), the corresponding face locations in the image (labeled with cluster number) are recorded in the order of left to right, top to bottom.

Timestamps of all the pictures are extracted and recorded. Next, pictures are clustered into events based on time as follows: sort timestamps in order of time, and intervals between consecutive timestamps are computed. Suppose there are N intervals, sort the interval values from small to large in the array interval[k], k=1~N.

Set: $i = N * \alpha$, $high\_bar = \text{interval}[N - i - 1]$, $low\_bar = \text{interval}[i]$, then: $threshold = high\_bar + (high\_bar - low\_bar) * \beta$, where α and β are empirically determined parameters. Use the threshold to check against each interval value along time, and whenever an interval is bigger than the threshold, an event boundary is set at that interval. Finally, photos are labeled with event indexes.

3.2 Identifying the Nuclear Family

To identify the nuclear family (i.e. husband, wife and kids) of the photo collection owner, the following steps are taken:

(1) Among all the major clusters (i.e. clusters having more than *M* faces, *M* is empirically selected, e.g. *M*=3), find most significant clusters that may contain nuclear family members.

(a) For clusters labeled as "baby" or "child", group them and analyze the distribution of cluster size. In one example, there are 12 clusters in the group. The size distribution of the clusters is [614 447 48 37 21 8 6 5 3 3 3 3].

(b) To reveal the statistical significance among the sizes of the clusters, the Gaussian Mixture model is used with the number of mixtures to be 3, assuming there are three subgroups of clusters in the model (i.e. nuclear family kids clusters, relatives or close friends and others). The above distribution yields an index as [1 1 2 2 2 3 3 3 3 3 3 3] with GMM analysis.

(c) Candidates of kids of the nuclear family are the ones in the group with largest clusters, in this case, clusters labeled with index=1. Similarly, candidates of kids as relatives or close friends are those in the group with second largest clusters, in this case, clusters labeled with index=2.

(d) In a similar way, candidate adult members in the nuclear family, and candidate adult members as relatives or close friends are found among the clusters labeled as "adult" or "senior".

(2) Within the candidate nuclear family members obtained in the above step, find the adult or senior male and the adult or senior female who appear most often in the collection, and co-appear with each other most frequently (photos including the two people exclusively and photos in which the two people are next to each other are given higher weights; photos in which the two people's faces are close to each other are given even higher weights). They are determined as husband and wife of the nuclear family.

(3) It allows for the situations that one or both of the husband and wife are missing from the photo collection. That is, it is possible that no couple satisfying the above criteria can be found in the candidates. In such case, the adult/senior male or female whose cluster is among the largest and who co-appears most often with the top candidate family kid(s) (again, photos including the two people exclusively, photos in which the two people are next to each other, and photos in which the two people's faces are close to each other are given higher weights) is determined as the parent of the nuclear family. If such a person cannot be found, then no parent is identified for the nuclear family.

(4) According to estimated age of the husband and the wife (when available), a range of possible age of their children is determined.

(5) Among candidate nuclear family kids who are within the estimated scope of age, find those who co-appear frequently with the husband and/or wife and co-appear frequently with each other (photos of kids with the husband and/or wife exclusively, photos of siblings exclusively, group photos in which kids are right next to the husband and/or wife are given higher weights). They are determined as children of the nuclear family.

3.3 Identifying the Extended Family

Now, group all the candidates of relatives and friends (adults and kids), as well as candidates of nuclear family members who are not selected in the previous step. Then members of the extended family, including parents, siblings, siblings-in-law, nephews and nieces of the husband and wife of the nuclear family, are found among this group of candidates based on co-appearance, positions in photos, gender, age and face similarity as follows.

(1) Based on the estimated age of the husband and the wife (when available), scopes of possible age of their parents on each side are determined. Among the candidates labeled as "senior", identify males and females who co-appear most often with nuclear family members (photos in which he/she is right next to at least one of the nuclear family members and photos in which he/she holds one of the nuclear family kids are given higher weights) and compute a co-appearance score for each of them. Meanwhile, the face similarity score is calculated for each candidate senior with the husband and the wife, respectively.

(2) To identify parents of the husband, find a senior male and a senior female who are in the determined age group, have the highest co-appearance scores with the nuclear family and the highest scores of face similarity with the husband – both scores satisfying predefined thresholds. Photos containing co-appearance of the two, especially those in which the two have close positions, further confirm their relation as a senior couple. Photos where one or both of them co-appear with the husband alone, or in close positions with the husband in group photos, further confirm they are from the husband's side. Similar analysis is used to identify parents of the wife.

(3) It allows for the situations that parents on one or both sides are divorced and/or remarried; or one or more parents passed away. For example, it is possible that one of the parents is intimate with a non-parent in photos (i.e. frequent co-appearances and close positions) who might be a step-parent.

(4) Within the group of candidates, siblings of the husband and the wife are identified as those who appear together with nuclear family members as well as with parents of the husband or wife in a number of photos (photos taken at different events are given higher weights), and who have high face similarity with either the husband or the wife. The significant-other of a sibling is identified as the one who co-appears and is in close position with the sibling in photos. Kids of siblings are those who are in photos together with identified grandparents; and/or who have large amount of co-appearance with nuclear family kids; and/or who appear together with an identified sibling and/or sibling-in-law.

3.4 Identifying Relatives and Friends

Other major characters are determined as relatives or friends of the family. People are identified as relatives if they co-appear not only with the nuclear family, but also with the extended family members in photos of different events or in a relatively large amount of photos in one event. For instance, if the husband's parents are together with a newlywed couple in some wedding photos, then the couple might be relatives of the family from the husband's side. Relationships such as husband-wife, parents-kids and siblings can be estimated for these people. Based on gender, age, co-appearance, positions in event photos and face similarity, relatives such as cousins, aunts and uncles may be determined. The rest major characters are determined as friends. Closeness with friends might be judged by frequency of co-appearance and positions in photos. Co-appearances at the same event are given lower weights than co-appearances at different events.

3.5 Displaying the Family Relation Tree

Once major characters in a photo collection are found and their roles are determined, a relation tree can be set up and displayed. Figure 1 shows one way of displaying the relations. Apparently, it can be an open research issue as to the design of an informative, interactive and artistic-looking family relation tree, as well as the way for the user to correct and annotate the resulting tree.

4. EXPERIMENTAL RESULTS

We tested the proposed approach on two typical family photo sets which were labeled with ground truth. In one collection, there are 1647 photos spanning over nine years (2000~2009). Table 2 lists the size, estimated age and gender of major clusters (those having four or more photos) generated with our automated face clustering technology. Among these, five are baby/child clusters with size [447 274 92 18 11]; others are adult/senior clusters with size [254 224 74 69 65 58 43 34 30 18 9 9 9 9 8 8 8 8 7 7 7 6 5]. With GMM analysis, the clusters of kids are classified as [1 1 2 3 3]; the adult clusters are grouped as [1 1 2 2 2 2 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4]. Obviously, the top two kid clusters and top two adult clusters are candidate nuclear family members. To be safe, we also include clusters labeled with index=2 in the candidate group.

Table 2: Major Face Clusters Formed for a Family Photo Set

Cluster No.	1	2	3	4	5	6	7	8	9	
Cluster size	447	274	254	224	92	74	69	65	58	
Est. age	C	C	A	A	C	A	S	S	S	
Est. gender	F	F	F	M	M	F	M	M	F	
Cluster No.	10	11	12	13	14	15	16	17	18	
Cluster size	43	34	30	18	18	11	9	9	9	
Est. age	S	A	S	A	C	C	A	A	A	
Est. gender	M	F	F	M	M	M	M	M	F	
Cluster No.	19	20	21	22	23	24	25	26	27	28
Cluster size	9	8	8	8	8	7	7	7	6	5
Est. age	S	S	S	A	S	S	A	A	S	A
Est. gender	F	F	F	F	F	F	F	F	F	F

(C – Child; A – Adult; S – Senior; M – Male; F – Female)

Within the group, cluster No.3 is the biggest female adult cluster, and cluster No.4 is the biggest male adult cluster. They are at the first place on each other's co-appearance ranking list, having 87 co-appearances spanning the years 2000~2009, among which 27 are photos containing only the two, 39 are photos containing the two and one or two kids in the candidate family kids group, and 14 are group photos in which they are next to each other. There are four photos in which their faces are very close to each other. Therefore, it can be determined that No.3 and No.4 are wife and husband of the family, respectively.

Among the candidate family kids, clusters No.1 and No.2 are the largest clusters in the photo set. They are ranked highest on each other's co-appearance list, followed by the wife, and then the husband. Especially, there are 121 photos containing exclusively one or both kids with one or both of the husband and wife. In contrast, while there are 92 photos in cluster No.5, there is just one co-occurrence with the husband in a group photo, and only a few with the wife. Thus, No.1 and No.2 are nuclear family kids.

Clusters No.7 and No.8 are the biggest senior male clusters. Both have many co-appearances with nuclear family members (35 and 41, respectively). Especially, each has a number of photos holding one or both of the family kids. No.7 has the highest face similarity with the husband. He has 15 photos with the husband alone, while he appears with the wife only in group photos. On the contrary, No.8 ranks high in both face similarity and co-appearance with the wife. Apparently, No.7 and No.8 are fathers of the husband and the wife, respectively. Similarly, it can be decided that No.9 and No.12 are mothers of the husband and the wife, respectively. High co-appearance rankings between No.7 and No.9, and those between No.8 and No.12 further confirm this estimation.

Cluster No.6 is an adult female. On her lists, the husband ranks 1st in face similarity; cluster No.10 ranks highest in co-appearance, followed by the husband. There are 16 photos of her co-appearing with nuclear family members, some of which include parents of the husband. There are also 12 photos of her appearing alone with one or two parents of the husband, and two photos including only her and the husband. Obviously No.6 is a sister of the husband. There are 24 photos of No.6 and No.10 being next to each other, and in close positions in some of them, so No.10 is found to be the significant-other of No.6. No.14 was held by No.6 or No.7 (the husband's father) in some photos, and he co-appears with both nuclear family kids. He was hence determined to be the kid of No.6. Similarly, No.13 and No.11 are recognized as brother and sister-in-law of the wife, respectively; and No.5 is their kid.

Cluster No.19 is a senior female who has several exclusive photos with the husband's father. She also has photos in which she holds a family kid or the husband's nephew. Thus, she is a close relative or friend of the husband's father. Considering the fact that there are no photos of the husband's mother after the year 2003; and there are no photos of No.19 before the year 2006, it is possible that No.19 is the current significant-other of the husband's father.

Clusters No.16-17, No.20-26 are nine people that appear in group photos with the nuclear family and the husband's sister's family. Most photos were taken in one event, and the rest from another event. Thus, these are relatives or friends of the husband's family. Clusters No.18, No.28 and the wife are together in some photos, thus, the three are relatives or friends. Cluster No.18 also appears in a few photos involving other family members, so she seems to be closer to the family than No.28. Cluster No.15 is a child having co-appearances with one family kid. She was also held by or next to No.18 in a couple of photos, indicating they are related. No.27 is in photos with family members taken at two events, so might be a relative or friend of the family.

The relations estimated above are shown in Figure 9, which match with ground truth rather accurately except for the case of No.27 – instead of being a family relative or friend, No.27 is actually the portrait of Chairman Mao in Beijing's TianAnMen Square. It happened that the family took quite some photos there at different times. Also, the estimated gender is wrong for No.15, No.20 and No.27, showing that our gender estimation may not be reliable for little kids and seniors, particularly when the cluster is small.

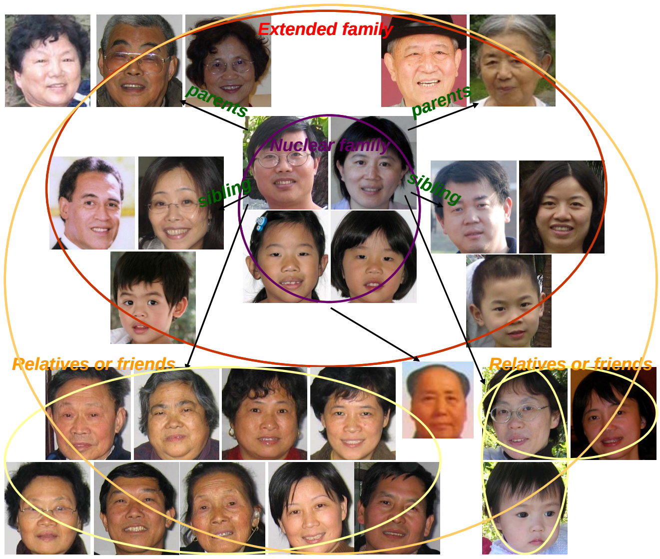


Figure 9: Estimated relation tree from one family photo set.

The other photo set contains 950 photos, and there are 28 clusters that have 4 or more photos. The derived relation tree is shown in Figure 1. It was straight forward to figure out the nuclear family members who are the top three clusters (wife, husband and baby, respectively), as well as extended family on the wife's side since they appear frequently in the dataset. However, it was a little tricky to find out folks on the husband's side because many of them show up just in one event (i.e. the couple's wedding). For instance, there are only four photos of the husband's father, but co-occurrence and relative position information in these photos with the husband and the husband's mother (luckily she was easily found) help our approach recognize him. One adult male cluster has a number of co-appearances with the wife's sister including two exclusive ones, while he is also in photos with the husband's family, especially one containing just the husband's mother and him in the wedding event. Plus, the person and the husband are highest on each other's face similarity ranking. Based on the principle that blood relation estimation is more reliable than significant-other relation estimation, the person is assigned as a sibling of the husband by our approach.

The estimated relation tree provides crucial semantic information for consumer image retrieval. For example, it makes possible to retrieve photos with text queries such as "photos of the nuclear family", "photos of little Amy with paternal grandparents" and "photos of Andy and Emily with cousins". Figure 10 lists images retrieved with the query "baby with both maternal grandparents".

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an approach to automatically extract people's relations and especially derive a tree structure of major characters from a family photo collection which provide critical semantic clues for consumer image indexing and retrieval. We use information derived with multiple face analysis technologies and different kinds of contextual cues, as well as expert knowledge obtained from experiments and observations, to formulate a work-flow of estimating a relation tree from an image collection. It has been proved to be effective tested on different datasets.

Figure 10: Example of consumer image retrieval.

There is still plenty to do in developing a more robust and useful system for estimating family relation tree. First, we will train the workflow with more varieties of family photo collections so that it can adapt to various cases of family relations. Second, more machine learning elements will be included into the workflow. Also, a finer classifier of age will be developed so as to estimate a person's age more accurately. Moreover, user annotations, when available, may include keywords indicating name, place, relation and so on that can be integrated into our workflow.

6. REFERENCES

[1] R. Rowe, G. Creamer, S. Hershkop and S. Stolfo. Automated social hierarchy detection through email network analysis. *Proc. of the 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis*, pages 109-117, San Jose, California, USA, 2007.

[2] S. Golder. Measuring social networks with digital photograph collections. *Proceedings of the 19th ACM Conference on Hypertext and Hypermedia*, pages 43-47, June 2008.

[3] P. Wu and D. Tretter. Close & Closer: social cluster and closeness from photo collections. *Proceedings of ACM Conf. on Multimedia*, pages 709-712, Beijing, Oct. 2009.

[4] A. C. Gallagher and T. Chen. Using context to recognize people in consumer images. *IPSJ Trans. on Computer Vision and Applications*, 1:115-126, 2009.

[5] L. Gu, T. Zhang and X. Ding. Clustering consumer photos based on face recognition. *Proceedings of IEEE Conf. on Multimedia and Expo*, pages 1998-2001, Beijing, July 2007.

[6] T. Zhang, J. Xiao, D. Wen and X. Ding. Face based image navigation and search. *Proceedings of ACM Conf. on Multimedia*, pages 597-600, Beijing, Oct. 2009.

[7] W. Gao and H. Ai. Face gender classification on consumer images in a multiethnic environment. *Proceedings of the 3rd IAPR International Conference on Biometrics*, University of Sassari, Italy, June 2-5, 2009.

[8] F. Gao, H. Ai. Face age classification on consumer images with Gabor feature and fuzzy LDA method. *Proceedings of the 3rd IAPR International Conference on Biometrics*, University of Sassari, Italy, June 2-5, 2009.

[9] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Trans. on Image Processing*, 11(4):467-476, April 2002.

[10] P. Belhumeur, J. Hespanha, *et al.* Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *Proc. European Conf. on Computer Vision*, pages 45-58, 1996.