PERSON-IDENTITY-BASED CLUSTERING FOR DIGITAL PHOTO ALBUMING

[1]Seungji Yang, [1]Kyong Sok Seo, [2]Sang Kyun Kim and [1*]Yong Man Ro

[1]Image and Video Systems Lab., Information and Communications University, Korea
[2]Computing Lab., Digital Research Center, Samsung Institute of Technology, Korea
E-mail: *yro@icu.ac.kr

ABSTRACT

In this paper, we propose an automated photo clustering method based on person-identity. The proposed method uses face and cloth information together as person identities. Photos are segmented based on situation so that the cloth information is meaningful cue to identify the person. The experiments performed with MPEG-7 VCE3 database and results showed the clustering performance with the proposed person-identity was increased by use of cloth information as well as face information.

1. INTRODUCTION

Recently, the number of digital camera users is rapidly increasing. So people are likely to store large amounts of photos in their storages with the proliferation of digital camera. Usually people enjoy seeing their photos and recalling memories, or sometimes, they want to share the photos with their friends or relatives [1]. There is thus a strong demand for automated or semi-automated tools to catalog the large amounts of photo collection since the manual cataloging by user seems to be quite time-consuming [2]. Of lots of functionalities of digital photo album, management of photos based on person identity is known to be fundamental and essential.

The person-identity based clustering means that photos of similar persons are classified, i.e., photo series are automatically clustered based on person so that one can easily browse clustered photos and share them with people who are taken with.

To cluster the persons in the photo, face recognition is necessary. Many researches have been worked in order to detect or recognize the faces automatically from images in area of intelligent surveillance system, automatic gate control system, and face search system [3].

In general, photos taken from non professional would contain inconsistent backgrounds with illumination variation. Moreover camera control operations such as zooming and flashing make the photos be more difficult to be analyzed. Especially, the face detection and recognition from the home photos are difficult due to various poses and environments. So it is not enough to detect person in home photos with face information only.

In this paper, we propose a photo clustering method based on person-identity. Cloth information as well as the face information of the person is considered to identify person. We used MPEG-7 advanced face recognition descriptor (AFRD) for the face description, and color and texture descriptors for cloth descriptions.

This paper is composed of 4 sections. In section 2, the proposed person identification method is described. And experimental results are included in section 3. Finally, we conclude the paper in Section 4.

2. METHOD

The person-identity based photo clustering consists of two parts: one is situation-based clustering and the other is person identification. Figure 1 shows overall procedure of the proposed person-identity based photo clustering.

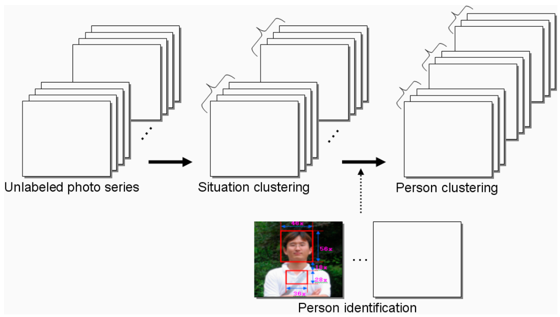

Figure 1: Overall procedure of person-identity based photo clustering

2.1. Situation-based Clustering

Home photos usually have complex background with a variety of illumination levels [2]. Also, characters on the photos wear the same clothes during finite period of time, e.g., a day. The same group of people is more likely to be taken together with the finite duration.

In this paper, 'situation' is considered as a sequence of photos that were taken in a fixed place or near the place with relatively close time period. So a group of neighboring photos in terms of taken-time may have similar situation. A group of photos associated with the same situation often contain similar background or similar semantic concepts. Thus the situation can be useful as a fundamental element of semantics for photo clustering or indexing.

The situation is also essential especially for the person-identity based photo clustering. Photos taken in the same situation may contain similar characters as well as similar background. In usual case, the clothes on which characters put may not be changed in same situation. In simple way, the situation can be a day by taken-time.

The cloth information becomes meaningful within a situation. So, the situation and cloth information is useful on top of face features to identify persons on photos.

In this paper, we found situation change boundary by detecting both visual change and time gap between photos [7]. For similarity matching to detect the situation change boundary, first, photo sequences are sorted in taken time order. And then, the similarity matching is performed with two photos neighboring in time. The time similarity (D_{time}) between the current $(j)^{th}$ photo and the previous $(j-1)^{th}$ photo are measured as [7],

$$D_{time}(i) = \frac{\log\{\mathbf{F}_{time}(i) - \mathbf{F}_{time}(i-1) + C_{time}\}}{D_{time\_max}}, \qquad (1)$$

where the taken-time feature $\mathbf{F}_{time}$ is obtained from Exif header of photo data, log(•) is a time scale function, C_{time} is a constant to avoid zero for input of the scale function. D_{time_max} is maximum time difference. The value of time similarity distance is scaled so that it is less sensitive to the large time difference. The time similarity distance at the same situation is insensitive.

Given the taken-time similarity, the similarity matching between the $(j)^{th}$ and $(j-1)^{th}$ photos is performed with the content-based features and their importance [7]. It can be written as

$$D_{total}(j) = \exp\left[D_{time}(j) \times \sum_{k=0}^{K}\{w_k(j) \times D_k(j)\}\right], \qquad (2)$$

where $D_k(i)$ is similarity distance defined as $d_k\{\mathbf{F}_k(j) - \mathbf{F}_k(j-1)\}$ in which d(•) is similarity matching function for the k^{th} content-based features. Using the exponential function, the similarity value at smaller value of feature difference is reduced while it is expanded at higher value. The w_k is an importance value for the k^{th} feature. It is adaptive to the visual semantics of photo [7].

Provided the time and visual similarity distances, situation change boundary is detected by peak detection as shown in Fig. 2 [7]. The detected situation are represented as,

$$\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \cdots, \mathbf{s}_I\},$$
$$\text{where } \mathbf{s}_i = \mathbf{s}_i\big|_{(j,j+k)} = \{p_j, p_{j+1}, p_{j+2}, \cdots, p_{j+k}\}, \qquad (3)$$

where $\mathbf{S}$ is situation cluster set, $\mathbf{s}_i$ is a situation cluster which has k photos, and p_j is a photo belong to the situation cluster.

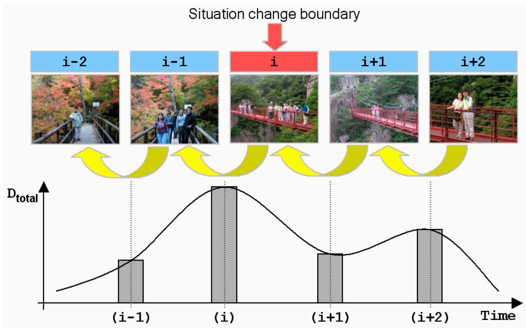

Figure 2: Situation change boundary detection

2.2 Person Identification

Within a situation, face and cloth features of photos are measured for detecting person identification. In this paper, it is assumed that the face is reliably detected. We employed MPEG-7 AFRD face detection method..

After face detection, normalized face is obtained by scaling the gray scale of the face image into 56 lines with 46 pixels in each line. In a similar way, normalized cloth of person is obtained by scaling the original cloth image into 28 lines with 36 pixels in each line. The cloth region is defined as 18 lines below the face region so as to reduce effect from the person's pose.

Given the situation $\mathbf{s}_i$, the identification features of the p_j photo are described as,

$$\mathbf{F}_j\big|_{s_i} = \{\mathbf{F}_{face}, \mathbf{F}_{cloth}\}, \qquad (4)$$

where $\mathbf{F}_{face}$ and $\mathbf{F}_{cloth}$ are face feature and cloth feature, respectively. The cloth usually has texture patterns and colors. So the $\mathbf{F}_{cloth}$ can be described with two types of visual features: color and texture features, represented by $\{\mathbf{F}_{texture}, \mathbf{F}_{color}\}$.

Next, the similarity distances are measured among the cloth features in the situation $\mathbf{s}_i$. It can be written as

$$D_{cloth}(j, j+k) = \left\langle \sum_{f \in F_{color}} \{w_f \times D_f(j, j+k)\} \right\rangle \times \left\langle \sum_{f \in F_{texture}} \{w_f \times D_f(j, j+k)\} \right\rangle, \qquad (5)$$

where <•> is normalization function for differently-scaled features. The similarity distances of color and texture are multiplied since they are assumed to be independent of each other.

The similarity distances are measured among the face features in the situation $\mathbf{s}_i$ as

$$D_{face}(j, j+k) = \sum_{f \in F_{face}} \{ w_f \times D_f(j, j+k) \} \cdot \qquad (6)$$

The total similarity distances among the photos in the situation $\mathbf{s}_i$ is as follows.

$$D_{total}(j, j+k) = \langle D_{cloth}(j, j+k) \rangle + \langle D_{face}(j, j+k) \rangle \cdot \qquad (7)$$

Given the total similarity distance, photos in a situation are clustered into several person groups. The similarity of a photo from other photos in a situation is measured by a rank threshold. As comparing among person groups for all situation groups, we merge several person groups that are composed of the same person. Finally, person-identity based groups are made as

$$\mathbf{g}_l = \left[\{ p_j \} \middle| rank(D_{total}(l, j)) \le K,\ 1 \le j \le I, \forall j \right], \qquad (8)$$

where rank(•) is a rank function to determine the rank of the similarity distance values and K is the rank threshold indicating whether p_j belong to the person group $\mathbf{g}_l$.

Figure 3 shows overall procedure of the person-identity based photo clustering algorithm.

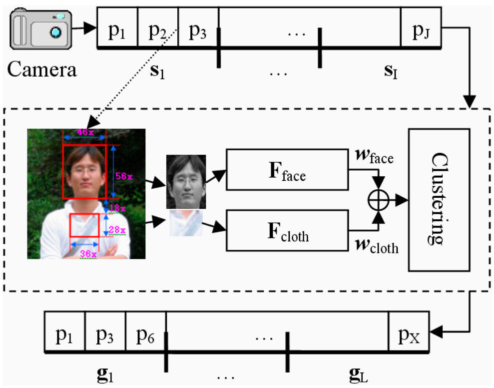

Figure 3: Person-identity based photo clustering algorithm

3. EXPERIMENTS

In order to verify the proposed person-identity based photo clustering method, experiment has been performed with MPEG-7 data set for visual core experiment 3 (VCE-3) [8]. The data set is composed of 1385 photos. Using an automatic face detection tool developed by SAIT [6], 1819 faces were proposed as ground truth to the VCE-3. It was assumed that the face features of the 1385 photo data set are enough to be described by AFRD. Also we assumed that the photo has no cloth information if faces are located at margin of photo or two faces are too close to extract cloth information. With the criteria, the cloth information was detected from 1615 out of 1819 facial images.

For low-level feature, 3 MPEG-7 descriptors have been utilized; they are advanced face recognition descriptor (AFRD), homogeneous texture descriptor (HTD), and color structure descriptor (CSD) with illumination invariant color descriptor (IICD).

For the proposed person-identity based clustering, the facial query photos were sorted by taken-time in ascending order. They were divided into several situation groups by using the proposed situation clustering method.

To evaluate the performance, recall and precision with rank threshold were used. The recall and precision are defined as,

$$P_{recall} = S/K \text{ and } P_{precision} = S/T, \qquad (9)$$

where K is rank threshold which is the number of top ranks to be evaluated. T is the number of ground truth, and S is the number of the ground truth which is found in the K retrieved images.

Figure 4 shows average performance with the variation of the rank threshold, where the rank threshold was set to multiple of the number of ground truth. In this experiment, the rank threshold was determined as the value at equal point of recall and precision. As shown in Fig. 4, recall was equal to precision when the rank threshold K is same as the number of ground truth.

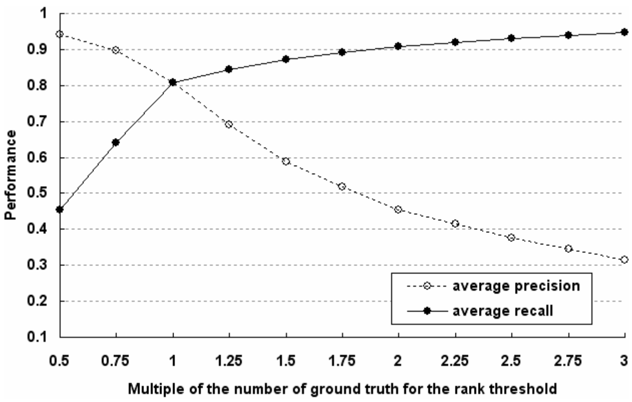

Figure 4: Average performance with the variation of the rank threshold

Given a situation group, performance of person-identity based photo clustering was measured by averaging the performance over all person groups in situation group. Overall performance was measured by averaging the performance over all situation groups. The average performance $p_{avg}(n)$ for the $(n)^{th}$ situation group is measured as

$$p_{avg}(n) = \sum_k \{ p(n,k) \times g(n,k) \} \Big/ \sum_k g(n,k), \qquad (10)$$

where n is the $(n)^{th}$ situation group and k is the $(k)^{th}$ person group. p(n, k) is the performance of the $(k)^{th}$ person group in the $(n)^{th}$ situation group. g(n, k) is the number of the ground truth of the $(k)^{th}$ person group in the $(n)^{th}$ situation group. The performance of each person group is weighted by g(n, k).

The overall performance $p_{overall}$ is measured over all situation groups as

$$p_{overall} = \sum_{n=1}^{R} \left\{ p_{avg}(n) \times h(n) \right\} \Big/ \sum_{n=1}^{R} h(n) , \qquad (11)$$

where R is the total number of situation group and h(n) is the number of ground truth of the $(n)^{th}$ situation group. The performance of each situation group is also weighted by the number of ground truth, g(n, k).

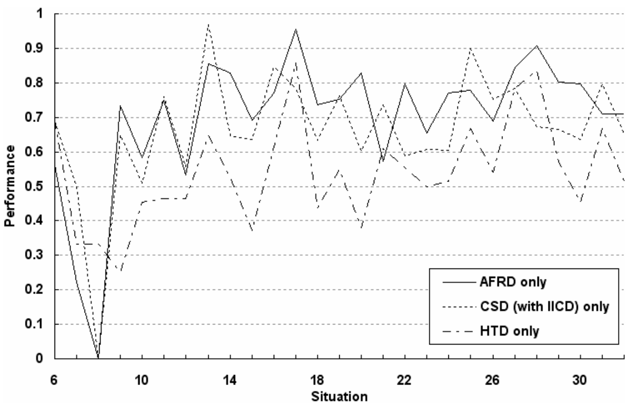

Figure 5: Performance of person identification for single feature over the 32 situation groups

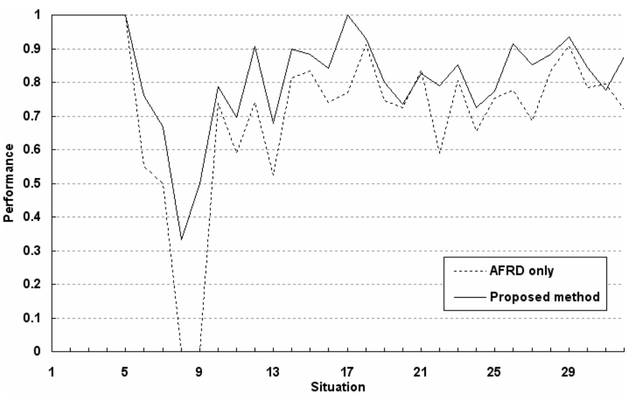

Figure 6: Performance of person identification for the proposed method over the 32 situation groups

In the experiment, 32 situation groups were generated by the proposed situation clustering from the data set. Figure 5 shows the performance of person identity for single feature over the 32 situation groups. There are three results for AFRD only, HTD only, and CSD (with IICD) only. As shown in Fig. 5, the AFRD outperformed both the CSD and HTD over most of the situation groups as far as single feature is concerned.

Next, the experiment was performed with both face and cloth features where the importance of each feature was optimized as 0.9 for $w_{CSDwithIICD}$, 0.8 for w_{HTD} and 1.5 for w_{AFRD}. Figure 6 shows the performances of person identity for the proposed method over the 32 situation groups. As shown in Fig. 6, the proposed method outperformed the AFRD over the most of the situation groups. The overall performance $p_{overall}$ over the all situation groups was increased to 0.81 in the proposed method. For comparison, the clustering performance for face information only in the entire dataset was 0.4.

5. CONCLUSION

In this paper, we proposed a photo clustering method based on person-identities. The proposed method used face and cloth information together as person identity. To utilize the cloth information, situation-based clustering was proposed. The experiment results have shown that the clustering performance would be much increased, from 71% to 80% approximately. In future work, we will focus on indexing all facial groups given the results clustered by the proposed method, where the clustered facial images can facilitate person-identity based photo indexing using multiple queries.

REFERENCES

[1] Y. Yagawa, N. Iwai, K. Yanagi, K. Kojima, and K. Matsumoto, “The Digital Album: a personal file-tainment system,” Proc. of IEEE Intl’ Conference on Multimedia Computing and Systems, pp.433-439, 1996
[2] J.C. Platt, “AutoAlbum: clustering digital photographs using probabilistic model merging,” Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries, pp.96-100, 2000.
[3] G. Pike, R. Kemp, and N. Brace, “The psychology of humane face recognition,” Proc. of IEE Colloquium on Visual Biometric, pp.11/1-11/6, 2000.
[4] K. Sobottka and I. Pitas, “Segmentation and tracking of faces in color images,” Proc. of Second Intl. Conf. on Automatic Face- and Gesture-Recognition 96, 1996
[5] B.S. Manjunath, P. Salembier, and T. Sikora, Introduction to MPEG-7, John Wiley & Sons, LTD
[6] T. Sikora, Introduction to MPEG-7, John Wiley & Sons, LTD
[7] Y. M. Ro et al, “Automated Situation Clustering of Home Photos for Digital Albuming,” Proc. of SPIE Electronic Imaging, 2005
[8] MPEG NIST, ftp://mpeg.nist.org/PhotoAlbum