Enhancing Person Annotation for Personal Photo Management Applications

Saman H. Cooray and Noel E. O'Connor
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Ireland
Email: coorays@eeng.dcu.ie

***Abstract*—This paper addresses a sub-problem of the broad annotation problem, namely "person annotation", associated with personal digital photo management and investigates approaches to enhancing person annotation in personal photo management applications. We study a number of approaches to enhance the performance of semi-automatic person annotation using real-life personal photo collections as the test data. To this end, face and body-patch features are employed to describe the appearance of a person as a means to more effectively capture the identities of re-appearing people in personal photo archives. Experiments are carried out to identify a suitable initial annotation method, compare the performances of event-constrained person matching with global person matching, and the effect of the size of initial annotation on the overall performance of person annotation in real-life personal photo archives. The evaluation results, presented in terms of H-Hit rate figures, illustrate that using event-constrained person matching with event-based initial annotation proves to be a better performing approach than global person matching for person annotation in personal photo archives. Results also clearly demonstrate the nature of compromise that needs to be made when annotating large photo collections in terms of accuracy against user-interaction.**

***Keywords*-personal photo management; person annotation; face recognition; content-based descriptors**

I. Introduction

Personal digital photos are the typical photos taken by an average consumer to record some events of special significance in their lives. They differ from those taken by professional photographers who generally work for commercial purposes with loosely associated context. In general, most personal photo users expect a lifetime store of their photographs, possibly depicting several hundreds of important events during a period of a few decades.

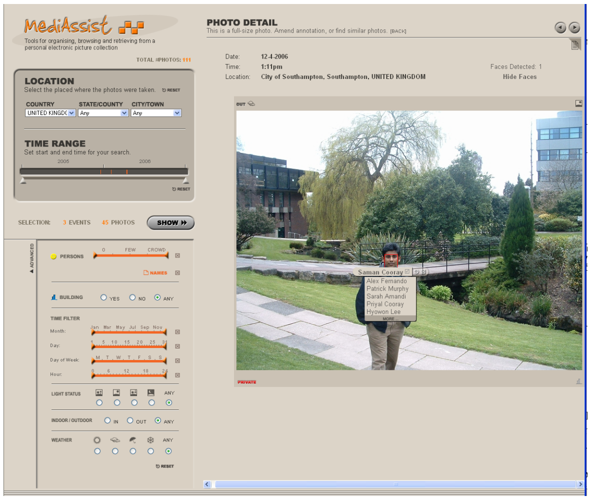
The nature of personal photography is changing as the use of digital cameras becomes increasingly pervasive. With emerging advanced technologies and falling prices of digital cameras the task of picture-taking has become much easier and more enjoyable for typical home users. As a result, they are taking more digital photographs than ever before, leading to significant increases in the size of their photo collections. Despite such a dramatic change in the perspectives of users, the lack of technology for automatically organising large personal photo archives remains a crucial drawback in digital photography.

Personal photo management systems are generally designed along the 4 dimensions, namely *when*, *where*, *what* and *who*. Utilizing time and location information of the photos alone has, however, shown to be of limited use in photo management. Conversely, technologies for effectively describing "what and who" in photo archives are just beginning to emerge. Recent user studies, such as [10], [3], suggest that identifying "who is in the photo" is one of the most important requirements in personal photo management systems. Addressing this problem, there has recently been significant research interest in technologies for supporting effective photo management [1], [6], [7], [9], [12], [10], [3]. While some approaches exploit content-based technologies alone, other research proposals indicate that using combined content and context features of the photos may be preferable to person annotation in personal digital photo archives. However, the existing approaches have not so far addressed the issues in relation to identifying an effective initial annotation technique and more importantly a suitable method to select the content and context features for person recognition in personal photo archives.

In this paper, we study a number of approaches for semi-automatic person annotation using real-life personal photo collections as the test data. The person annotation task is designed in such a manner that having provided a user annotated set of photos, which we term the "initial annotation set" in this paper, the system automatically suggests a list name candidates for each person to be annotated in sub-sequent annotations (see Figure 1 for an example semi-automatic person annotation system). Key contributions from this research include identifying a suitable initial annotation method and proving a potentially important fact that event-constrained person matching is a better approach than global person matching for person annotation. Additionally, this experimental study demonstrates the nature of compromise that needs to be made when annotating large photo collections in terms of accuracy against user-interaction.

II. Related Work

A significant number of approaches for person annotation have been proposed in the literature, leading to varying degrees of success in personal photo management. While some approaches have explored methods for person anno-



MediAssist

Tools for organising, browsing and retrieving from a personal electronic picture collection

TOTAL #PHOTOS: 111

LOCATION

Select the placed where the photos were taken.

COUNTRY: UNITED KINGDO

STATE/COUNTY: Any

CITY/TOWN: Any

TIME RANGE

Set start and end time for your search.

SELECTION: 3 EVENTS 45 PHOTOS

BUILDING ○ YES ○ NO ● ANY

LIGHT STATUS ○ ○ ○ ○ ● ANY

INDOOR / OUTDOOR ○ IN ○ OUT ● ANY

WEATHER ○ ○ ○ ○ ● ANY

PHOTO DETAIL

This is a full-size photo. Amend annotation, or find similar photos. [BACK]

Date: 12-4-2006

Time: 1:11pm

Location: City of Southampton, Southampton, UNITED KINGDOM

Faces Detected: 1

Saman Cooray ☑

Alex Fernando
Patrick Murphy
Sarah Amandi
Priyal Cooray
Hyowon Lee

Figure 1: Semi-automatic person annotation.

tation using content and context-based technologies, other authors have also demonstrated the importance of employing intuitive user interface and visualisation technologies for effective annotation of the photos.

In the semi-automatic annotation prototype system proposed by Zhang *et al.* [11], content-based features such as face and body-patch features, integrated into a Bayesian framework, are used for similarity matching. Person annotation is performed in a manner similar to ours where a candidate list of names for each person to be annotated is automatically prompted by the system.

Addressing the inherent difficulties in employing content-based technologies, Shneiderman and Kang [9] proposed the PhotoFinder system as a direct annotation method based on drag and drop of text labels. Annotation efficiency is improved by supporting the feature of bulk annotation where a number of instances of the same person can be selected and promptly annotated.

Employing only context information to annotate people in photos, Naaman *et al.* [6] proposed a semi-automatic person annotation system using temporal, spatial and social context features. As the user continues to annotate photos, name suggestions are prompted by the system for un-annotated persons based on the patterns of re-occurrence and co-occurrence of previously annotated persons in different locations and events, thereby leveraging the context available from photo metadata and user input.

Addressing the labor intensive drawback in most semi-automatic annotation systems, Zhao *et al.* [12] proposed a technique for annotating clusters of people that are created based on evidence from face, body, and context information. Photos are first clustered into events based on time and location context information. Within each event, the body information is clustered and then combined with face recognition results using a graphical model. Finally, the clusters with high confidence values of face recognition and context probabilities are identified as the ones belonging to a specific person.

Considering the lack of robustness in employing face recognition technologies to this problem, Suh and Bederson [10] carried out a series of user studies using the SAPHARI (Semi-Automatic PHoto Annotation and Recognition Interface) system to investigate the effectiveness/usability of semi-automatic annotation techniques and novel zoomable interfaces. Participants were provided with two different types of interfaces: a semi-automatic and a manual annotation interface. To study the effectiveness of person annotation, they compared person-based annotation using clothing recognition with manual annotation. However, there was only a little improvement from the semi-automatic annotation with clothing-based person recognition compared to the manual annotation.

In the user study carried out by Cui *et al.* [3] using the EasyAlbum semi-automatic photo annotation system, both face recognition and clothing features were employed to recognise people for person annotation. They also applied novel techniques for cluster annotation, contextual re-ranking and adhoc annotation through innovative user-interface technique, illustrating that such features can be used in combination with content-based technologies to enhance the performance of person annotation. The authors used Adobe Photoshop Elements 4.0 as a baseline to compare the performance of EasyAlbum and showed that EasyAlbum clearly outperformed Adobe Photoshop Elements 4.0 in both large and small size photo-collection management.

According to the recent research studies, it is clear that improved techniques for semi-automatic person annotation are undoubtedly useful for dealing with exponentially growing personal photo collections in the future. In this respect, we investigate a number of approaches in relation to choosing a suitable initial annotation method and identifying a person matching technique using content and context features for person annotation in personal photo archives. To the best of our knowledge, these issues have not been clearly addressed in the literature, which we investigate using real-life personal photo collections.

III. Content-Based Descriptors

Figure 2 depicts the technique adopted for person recognition using face and body-patch descriptors in this paper. Body-patch segments are extracted relative to automatic face detection results as described in [2]. We employ three content-based descriptors in fusion for person recognition. Based on our experimental results reported elsewhere, the MPEG-7 scalable colour descriptor (SCD) [5] and MPEG-7 homogeneous texture descriptor (HTD) [5] are used for body-patch matching and the local binary pattern (LBP) [8]

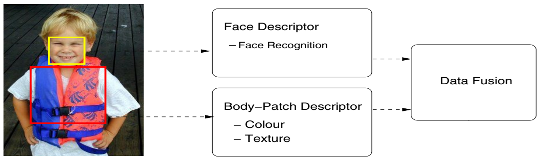


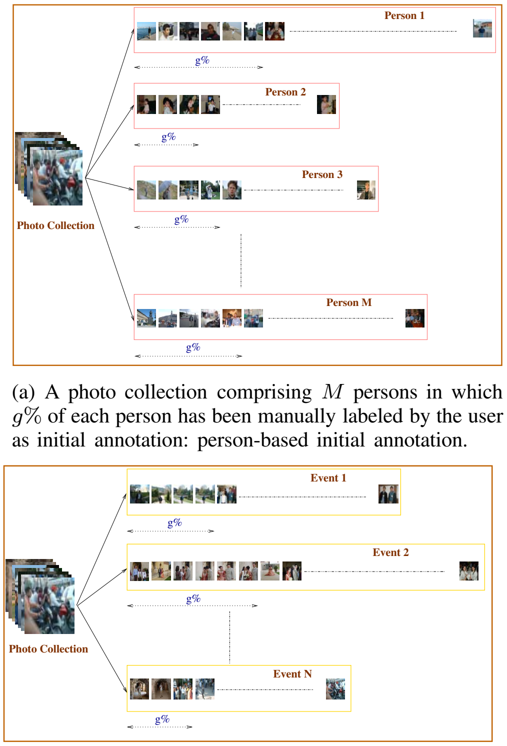
Figure 2: Person recognition in digital photo archives.

descriptor is used for face matching in this experimental study.

IV. Experimental Tasks

In this experimental study, a performance evaluation of person annotation is carried out using an *incremental* annotation model, where knowledge from all the previous annotations is applied when recognising people in subsequent annotations thereby improving the coverage of person identities as the annotation process progresses. We carry out the following experimental tasks to examine the performance of event-based against person-based initial annotation, event-constrained against global person matching, and the effect of the size of initial annotation set. To define events, we follow a similar approach to that proposed in [4] where an event is said to happen if a series of photos was taken in a relatively short period of time without leaving significant gaps between any two photos.

- Performance analysis of event-based and person-based initial annotation: In a semi-automatic person annotation framework, the user is generally expected to enrol some numbers of each distinct person in the collection so that they can be used for recognising people in the other photos. The main task of a person annotation system is then to recognise all the remaining people in the collection with minimal user intervention. This experiment is devised to identify the best method of initial annotation out of two possible scenarios, i.e. event-based and person-based (see Figure 3). In event-based initial annotation, the user provides the labels for randomly selected people by manually annotating g% of each event in the collection. In person-based initial annotation, the user initially labels g% of each distinct person in the collection;
- Performance analysis of event-constrained and global person matching: Experiments are conducted to analyse the performance of person annotation using person matching based on combined face and body-patch descriptors applied within events, termed event-constrained person matching, and using only face recognition applied within and across events, termed global person matching, in this paper;
- Effect of the size of initial annotation set: Each annotation experiment is performed for different sizes of initial annotation sets, i.e. g%, provided by the user, in order to examine the performance variation of person annotation against the level of initial annotation.



(a) A photo collection comprising M persons in which $g\%$ of each person has been manually labeled by the user as initial annotation: person-based initial annotation.

(b) A photo collection comprising N events in which $g\%$ of each event has been manually labeled by the user as initial annotation: event-based initial annotation.

Figure 3: Two scenarios of initial annotation.

V. Performance Evaluation

Based on the pre-determined optimal configuration for person recognition combining face and body-patch descriptors, we evaluate the performances of the proposed approaches in this section. Performance evaluation of the person annotation model is carried out using the H-Hit rate criterion, which was originally proposed by Chen *et al.* [1]. A Hit is said to happen if the true name of the person is included in the predicted name-list. Assuming that the entire collection is divided into two sub-collections: training (C_1) and test (C_2) with N_1 and N_2 persons in them, H-Hit defines the prediction accuracy for a given query with H indicating the number of candidates in the list:

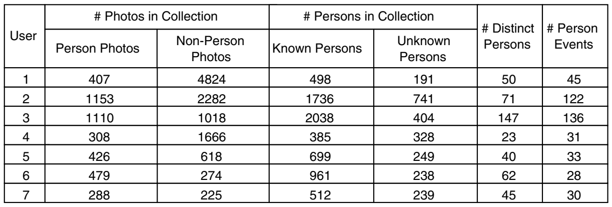
$$H - Hit = \frac{1}{N_2} \sum_{f \in C_2} hit_{H,C_1}(f) \tag{1}$$

where $hit_{H,C_1}(f)$ is 1 if f is included in the suggested list of H names taken from C_1, and 0 otherwise.

A. Test Data

Performance evaluation of the above-described approaches is carried out using personal photo collections belonging to 7 different users of the MediAssist system [2]. These photo collections comprise different types of events, such as birthday parties, meetings, family gatherings, graduation ceremonies and weddings.

Table I presents a statistical description of these photo collections for the 7 users (user 1 to user 7) in descending order of the total number of photos (face photos+non-face photos) each user has in his/her collection. Table I also describes the characteristics of the photo collection that each user has donated in terms of the number of photos that contain people (Face Photos) and that do not contain people (Non-face Photos), the number of known and unknown faces (Known Faces, Unknown Faces) in each collection, the number of distinct faces (Distinct Faces), and the number of person events (Person Events). The number of distinct faces in a collection corresponds to the number of known people that possess unique identities whereas the number of person events corresponds to the number of events formed using the photos that contain people in them.



User	# Photos in Collection		# Persons in Collection		# Distinct Persons	# Person Events
	Person Photos	Non-Person Photos	Known Persons	Unknown Persons		
1	407	4824	498	191	50	45
2	1153	2282	1736	741	71	122
3	1110	1018	2038	404	147	136
4	308	1666	385	328	23	31
5	426	618	699	249	40	33
6	479	274	961	238	62	28
7	288	225	512	239	45	30

Table I: A statistical description of the test data.

B. Experiments and Results

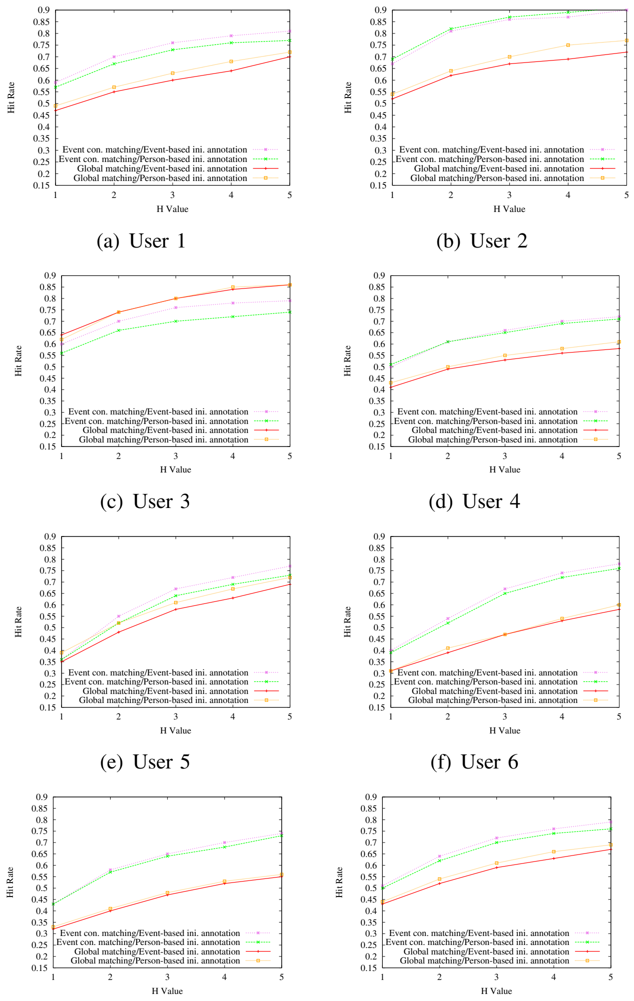
In this experimental study, we measure the performance of person annotation using two different initial annotation methods, two different person matching methods, i.e. event-constrained and global, and several differently sized initial annotation sets from 10% to 70% at intervals of 10. Results are presented for each individual user, i.e. User 1 to 7, in Figure 4(a) - 4(g), and also as an average measure of all the 7 users in Figure 4(h), with a 30% initial annotation set of each user as a reasonable choice. We use the nearest-neighbor classifier to infer the identity of all remaining persons in the collection. The hit-rate figures are computed by comparing the classification result with the true label of the person. As the annotation process continues, knowledge from all the previous annotations is applied to recognise people for subsequent annotations.

Concerning the results given in Figure 4(a) - 4(g), it can be seen that event-constrained person matching proves to be a better approach than global person matching in 6 out of the 7 scenarios studied in this experiment. We believe the reason for discrepancy in User 3 is due to that particular photo collection, which apparently is found to consist of a rather unbalanced distribution of known identities compared to the rest of the collections. Comparing the performances of the two initial annotation methods, it can be also observed that event-based initial annotation is better than person-based initial annotation in the case of event-constrained person matching. An opposite behavior can be, however, observed in the case of global person matching, depicting that person-based initial annotation is a better approach when carrying out person annotation using global person matching. Overall, it can be clearly seen than event-constrained person matching with event-based initial annotation leads to best person annotation performance in general. The results shown in Figure 4(h), which correspond to the average of the performance figures of all 7 collections, strongly generalise the findings arising from this research.

The results given in Figure 5 illustrate the behavior of person annotation against the level of initial annotation for different values of H. The relatively small drops in the performance levels of different hit-rate figures at 10%, 30%, 50% and 70% initial annotation suggest that by restricting the level of initial annotation to a small proportion ($\leq 10\%$) but providing the user with a name-list comprising at least 3 candidate names would result in a hit-rate figure of over 0.65. However, if the initial annotation set is increased to 30%, it may be possible that a hit-rate figure of over 0.5 can be achieved even with just one name suggestion, i.e. H=1. It can be also observed that using an initial annotation set larger than 30% results in only a very little hit-rate improvement. The results, in general, demonstrate the nature of compromise that needs to be made when annotating large photo collections in terms of accuracy against user-interaction.

VI. Conclusion

In this paper, we presented an evaluation of a number of approaches to enhancing semi-automatic person annotation in personal photo archives. We employed content-based technologies, i.e. face and body-patch descriptors, together with photo capture time as a context feature to evaluate the performance of person annotation using person matching both within and across events. Our experiments prove that using event-constrained person matching with event-based initial annotation is a more effective approach than global person matching in semi-automatic person annotation. We also analysed the variation of person annotation performance against the level of user interaction, i.e. the size of initial annotation set and the length name-list suggestion, demonstrating the nature of compromise one would have to make when annotating large photo collections. Our future works aim at examining the effectiveness of person annotation using bulk annotation methods by capitalising on the findings arisen from the research presented in this paper.



(a) User 1

(b) User 2

(c) User 3

(d) User 4

(e) User 5

(f) User 6

(g) User 7

(h) Overall: User 1 to 7

Figure 4: Person annotation results for different users.

Acknowledgment

This research is supported by Science Foundation Ireland under grant 07/CE/I1147.

References

[1] L. Chen, B. Hu, L. Zhang, M. Li, and H. Zhang. Face Annotation for Family Photo Album Management. *Intl. Journal of Image and Graphics*, 3:1–14, 2003.

[2] S. Cooray, N. O'Connor, G. Jones, N. O'Hare, and A. F. Smeaton. Identifying Person Re-occurrences for Personal Photo Management Applications. In *VIE 2006*, pages 144–149, September 2006.

[3] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang. Easy Album: An Interactive Photo Annotation System Based on Face Clustering and Re-ranking. In *CHI'07*, pages 367–376, USA, 2007.

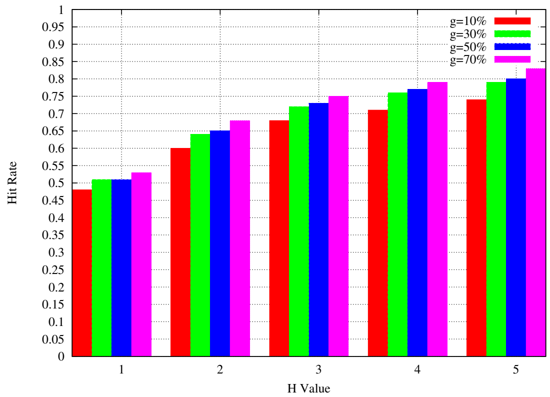


Figure 5: Comparison of hit-rate figures against H at 10%, 30%, 50% and 70% event-based initial annotation.

[4] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as Essence for Photo Browsing Through Personal Digital Libraries. In *ACM Joint Conference on Digital Libraries*, 2002.

[5] B. S. Manjunath, J.-R. Ohm, and V. V. Vasudeven. Color and Texture Descriptors. *IEEE Tran. on Circuits and Systems for Video Technology*, 11:703–715, June 2001.

[6] M. Naaman, R. B. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *JCDL'05*, pages 178–186, Denver, Colorado, USA, 2005.

[7] N. O'Hare, H. Lee, S. Cooray, C. Gurrin, G. Jones, J. Malobabic, N. O'Connor, A. F. Smeaton, and B. Uscilowski. Automatic Text Searching for Personal Photos. In 1^{st} *Intl. Conf. on Semantics And Digital Media Technology (SAMT'06)*, pages 43–44, Athens, Greece, 2006.

[8] T. Ojala, M. Pitikainen, and D. Harwood. A Comparative Study of Texture Measures with Classification based on Feature Distributions. *Pattern Recognition*, 29:51–59, 1996.

[9] B. Shneiderman and H. Kang. Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In *Intl. Conf. on Information Visualization*, pages 88–95, 2000.

[10] B. Suh and B. B. Bederson. Semi-Automatic Photo Annotation Strategies using Event based Clustering and Clothing based Person Recognition. *Interacting with Computers*, 19:524–544, 2007.

[11] L. Zhang, L. Chen, M. Li, and H. Zhang. Automated annotation of human faces in family albums. In *ACM Conference on Multimedia*, pages 335–338, Berkeley, November 2003.

[12] M. Zhao, S. Liu, T.-S. Chua, and R. Jain. Automatic Person Annotation of Family Photo Album. In *CIVR'06*, pages 163–172, USA, 2006.