Haven't we met before? — A Realistic Memory Assistance System to Remind You of The Person in Front of You

Masakazu Iwamura, Kai Kunze, Yuya Kato, Yuzuko Utsumi, Koichi Kise
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Naka, Sakai, Osaka 599-8531, Japan
{masa,yuzuko,kise}@cs.osakafu-u.ac.jp, {kai,katou}@m.cs.osakafu-u.ac.jp

ABSTRACT

This paper presents a perceived real-time system for memory augmentation. We propose a realistic approach to realize a memory assistance system, focusing on retrieving the person in front of you. The proposed system is capable of fully automatic indexing and is scalable in the database size. We utilize face recognition to show the user previous encounters with the person they are currently looking at. The system works fast (under 200 ms, perceived real time) with a decent database size (45 videos of 24 people). We also provide evidence in terms of an online questionnaire that our proposed memory augmentation system is useful and would be worn by most of the participants if it can be implemented in an unobtrusive way.

Author Keywords

Memory augmentation; memory assistance system; face recognition; wearable;

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Due to an aging population, more and more people will suffer from cognitive decline and memory loss in one or the other way. The most severe and socially unacceptable case would be when forgetting a person who recognizes you. You may have broken out in a cold sweat in meeting a person who recognizes you but you cannot remember who the person is or what the person was like. A possible way to cope with the situation is to directly ask the person "I don't remember you. Where did we meet?" or "Haven't we met before?" However, it is not always acceptable to do it, as it has a risk to lose important human relations. Another possible way is to pretend that you remember the person; you would try to keep talking about harmless topics, with looking for a cue to remind you of the person. This would lead to a success in some cases but not always. Yet another way is to use a *memory assistance system* which would tell who the person is or what the person was like.

AH '14, Mar 07-09 2014, Kobe, Japan.
ACM 978-1-4503-2761-9/14/03.
http://dx.doi.org/10.1145/2582051.2582083

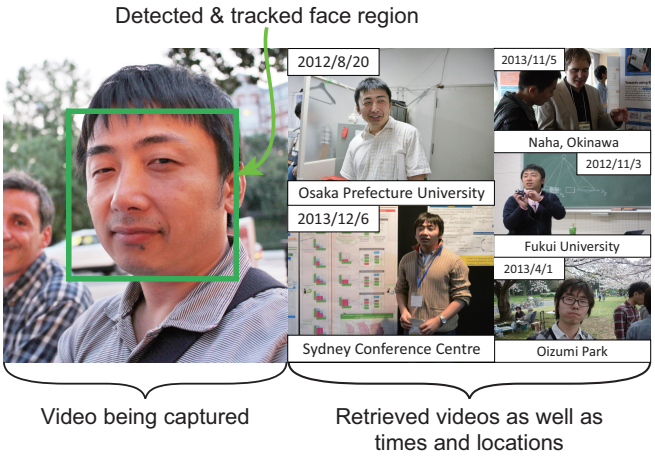

Figure 1: Pseudo screenshot of the proposed system.

The memory assistance system is a realization of *memory augmentation* that has been researched as a killer application of wearable computers [4, 3]. Its general notion is that a wearable camera records everything that happens on the user and the user retrieves something required later (e.g., [10, 2, 11]). Things that can be retrieved include documents, people, actions and locations.

Although the idea of the memory augmentation is clear and useful, it is based on ideal assumptions; all the things that can be retrieved are automatically indexed; a recognition technique can almost perfectly and quickly retrieve the correct things regardless of the size of indexed data. Unfortunately, the reality is far different from the ideal. Thus, it is very important to find a compromise between the ideal and reality.

In this paper, we propose a realistic approach to realize a memory assistance system, especially focusing on retrieving the person in front of you. A pseudo screenshot of the proposed system is shown in Fig. 1. On the left side, the video being captured with a wearable camera is displayed. The green rectangle is the face region that is detected in the first frame and tracked in the succeeding frames. The features of the face are passed to a face recognition method. Then, as shown on the right side, it retrieves several (five in the example) videos that can include the person. In addition to videos,

times and locations (the nearest landmark) the user met them are displayed. The main features of the proposed approach are as follows. (1) Videos of the people the user meets are automatically segmented and indexed in the system; no manual intervention is required. (2) The face recognition technique is scalable (fast and accurate to retrieve from a large database). (3) The system can cope with misrecognition; multiple videos that covers the correct answers are displayed simultaneously.

RELATED WORK

The memory assistance systems and memory augmentation have been researched for long. We review related researches.

One of pioneer works on human memory aid is *forget-me-not* [10]; a tablet device is designed to log the user's activity, and later, the user can retrieve a lost document, somebody's name and so on. This is based on communication between electronic devices, and does not intend to recognize and index human behavior itself.

Lifelog-type memory augmentation systems are proposed. *SenseCam visual diary* [11] collects a lifelog-type data created by a wearable device that automatically captures images and records sensor data. *Personal Lifelogs* [2] record life experiences even over a lifetime; collected data include visual data similar to the SenseCam, documents, emails and SMS messages along with context data of time and location. Archived data in both researches are for reviewing in offline. Thus, their purposes are different from ours.

Memory assistance systems focusing on face are proposed; they consist of a wearable camera and head-mounted display. *Who are you?* [12] is aimed at helping to remind the user of the person by showing the name of the person. So as to realize a real-time processing, a coarse-to-fine recognition scheme is proposed. In order to display the name, however, the recognizable people have to be registered in advance with manual intervention. Thus, it is not practical enough. *Smart Glasses* [9] also has a similar purpose; it shows the name of the person to remind the user of the person. In order to avoid manual intervention to index people (faces and names), it is supposed that the user is already connected with the people on social networks. A similar idea, while this is not related to a wearable device, is suggested in [1] that it is possible to tie an offline (physical) face with an online face (available on a social network) and even obtain the social security number from the offline face. These approaches, to use online face data, are interesting. However, online data are not always available and not always useful for reminding the person and the situation they met before. The closest idea to the proposed system in the literature is *Visual Augmented Memory* [4]. This research is aimed at helping to remind the user of the person by showing a previous video. The paper proposes a fully automated indexing and video retrieval similar to ours; the retrieved video can provide information about who (from the face), where (from the background of the video), when (mainly from timestamp) and what (from any visible actions). However, the paper does not mention its technical aspects; how to realize the automatic indexing and cope with large data and misrecognition.

Other researches in the related topics are as follows. To help patients with a disease who have difficulty in remembering faces, an eye-wear is proposed [15]. *Ubiquitous Memories*, so-called a memory externalization system, associates memories with physical objects [7].

SYSTEM OVERVIEW

The proposed system consists of a wearable camera and a head-mounted display installed in a laptop computer. The wearable camera is set up to capture the frontal view of the user. In the current implementation, the largest face in the captured image is detected by a face detection method [14]. This starts both indexing and retrieval processes.

In indexing process, the detected face is tracked in the succeeding frames using a tracking method [5] until tracking fails. Then, the image sequence is regarded as a video. The video is indexed with local features, each of which describes a local region of an image, extracted from the face region of every frame in the video. We used PCA-SIFT feature [8] as the local feature. Usually more than one hundred of features are obtained from a face region (in the experiment, about 140 features were obtained).

In retrieval process, the detected face is also tracked and features are extracted from it as well as indexing process. The detected features (hereafter, called query features) are used for retrieving videos using a scalable face recognition method [13]; in [13], a face recognition experiment with up to 10 million face images revealed that the method has desirable properties in recognition accuracy and computation time; keeping recognition accuracy, computation time of the face recognition method increased in sublinear unlike other methods. The recognition method is based on local features and voting. The procedure is as follows. (1) For each query feature, the nearest feature to the query feature is found in the database; for finding the nearest feature, the state-of-the-art approximate nearest neighbor search method [6] is used. (2) Weighted votes for the corresponding video to the found feature are cast; the weight is determined by the distance between the query feature and the nearest one. (3) The above process is executed for all query features, and the videos ranked at the fifth or above are selected as retrieval results. With more than one query frame, votes are simply accumulated.

EVALUATIONS

Questionnaire

A feature of the proposed system is to provide videos of previous encounters without their names. Although this makes the system completely human intervention free, one might doubt its effectiveness. To tackle how useful people might find our system, we setup an online questionnaire including seven questions and asked people over social networks to contribute.

We obtained 153 respondents. Answers are summarized in Table 1. Table 1(a) is the age distribution of respondents. It shows most of them were 20's. Thus the answers of the questionnaire were biased towards young people who do not

Table 1: Questions and answers of questionnaire.

(a) Age distribution of respondents.

10's	20's	30's	40's	50's	Total
6	106	21	12	8	153

(b) **Q1: How often do you meet a person you should have met before but cannot remember?**
Answer options: (1) More than once a week, (2) Once a week, (3) Once a month, (4) Once every three months, (5) Once a half year, (6) Once a year, (7) Almost nothing.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	6	19	26	25	27	49

(c) **Q2: Do you feel very embarrassed with the situation?**
Answer options: (1) Very embarrassed, (2) Embarrassed, (3) Yes and no, (4) Little embarrassed, (5) Not embarrassed.

(1)	(2)	(3)	(4)	(5)
16	61	22	45	9

(d) **Q3: How big is the problem of asking them about previous meetings?**
Answer options: (1) Very big problem, (2) Problem, (3) Yes and no, (4) Little problem, (5) Not a problem.

(1)	(2)	(3)	(4)	(5)
15	62	20	50	6

(e) **Q4: Which is useful to remind you of the person?**
Answer options (multiple answers allowed): (1) Name, (2) Date, (3) Place, (4) Event, (5) Previous movie.

(1)	(2)	(3)	(4)	(5)
54	28	75	95	31

(f) **Q5: Would you carry a system like that with you?**
Answer options: (1) Yes, (2) If convenient, (3) If implemented in a daily-use device, (4) No.

(1)	(2)	(3)	(4)
2	32	93	26

(g) **Q6: Which is the most appropriate device for the system?**
Answer options: (1) Smartphone, (2) Glasses, (3) Wrist watch.

(1)	(2)	(3)
78	46	29

(h) **Q7: Would you wear it everyday or to specific events?**
Answer options: (1) Everyday, (2) Specific events, (3) Do not wear.

(1)	(2)	(3)
51	76	26

require a memory assistance system so much. In Tables 1(b), the result of Q1 that asks how often you meet a person you cannot remember seems to reflect the bias. However, 2/3 of respondents sometimes encountered such a situation. Tables 1(c) and 1(d) show that most of respondents thought the situation is a problem and embarrassing. Table 1(e) shows that while the name is an important key to remember a person, other keys such as the places and events of previous encounters that are provided by the proposed system are more important.

Figure 2: Examples of videos in the created dataset. The green rectangles represent detected faces.

We mentioned before asking Q5-Q7 that "we are developing a memory assistance system useful in the situation you cannot remember the person in front of you." However, we did not mention its detail. Table 1(f) shows that most of respondents desired a system implemented in a daily-use device. In Tables 1(g) and 1(h), the results of Q6 and Q7 were different from our expectations; a half of respondents thought that the most appropriate device for the system was smartphone and wanted to ware it in specific events. We guess a reason we obtained these results in Q5-Q7 is that wearable devices are still not common. When wearable devices get common, we could obtain a different result.

Experiments

The best way to evaluate the performance of the proposed method is to examine it in real use. However, this is almost impossible because according to the result of Q1 of the questionnaire, it is relatively rare to encounter such a situation for most people. In addition, the user should meet the person beforehand wearing the system.

Thus, we evaluated the performance of the proposed system with an original dataset; it consisted of 45 videos of 24 people for training, and 15 videos of 15 people for testing. Figure 2 shows some examples. Videos for training and testing were captured in different days. All the subjects were Japanese of 20's. All the videos were captured mostly indoors with a web camera (Logicool C920). Their resolution was 960×720 pixels. For more detail of the training videos, one person had four videos, three people had three videos, seven people had two video, and eighteen people had one video. Average length of the training videos was 212 frames. The length of the test videos was 5 frames. In both training and test videos, face detection [14] was applied to a frame and if a face is found, face tracking [5] was applied in the succeeding frames.

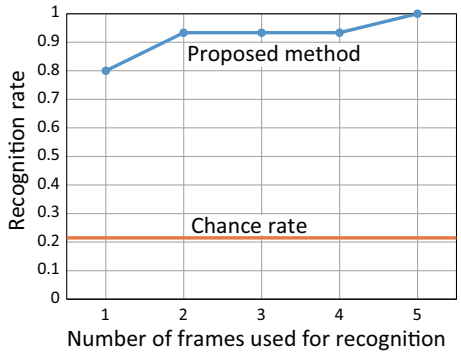

Figure 3: Recognition rates with from one to five consecutive frames of the query video.

Face regions were resized to 256 $\times$ 256 pixels and then PCA-SIFT feature [8] was extracted from the face regions.

In the experiment, a laptop computer with a 4-core CPU (Intel Core i7 2600K 3.4GHz) and 8GB memory was used. Computation times required for face detection, face tracking, feature extraction and face recognition were 109 ms, 61.7 ms, 161 ms and less than 1 ms, respectively.

Recognition accuracy was measured by the following criterion; a query is regarded to be successfully recognized if at least one video corresponding to the person in the query video is included in the five retrieved videos. Figure 3 shows recognition rates with one to five consecutive frames of the query video. It shows that only with one frame, 80% of query videos were successfully recognized. In comparison with the chance rate of 21.5%, the accuracy was significantly high. The accuracy increased with more query frames, and with five frames it became 100%. This result shows the effectiveness of the proposed system.

CONCLUSION

We present a system to augment human memory by showing videos, times and locations of previous encounters of the person in front of the user. Both speed and accuracy are more than acceptable to run the system. We are in the tight time boundary of retrieving relevant person specific information, so that other people might not recognize that the user has a hard time remembering them.

According to the results of the questionnaire, the places and events of previous encounters provided by the proposed system are more important than the name. This supports the effectiveness of the proposed system.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI #25240028.

REFERENCES

1. Acquisti, A., Gross, R., and Stutzman, F. Faces of facebook:privacy in the age of augmented reality. In *BlackHat USA* (2011).
2. Chen, Y., and Jones, G. J. F. Augmenting human memory using personal lifelogs. In *Proc. 1st Augmented Human* (Apr. 2010).
3. DeVaul, R. W., Pentland, A., and Corey, V. R. The memory glasses: Subliminal vs.overt memory support with imperfect information. In *Proc. 7th ISWC* (Oct. 2003), 146–153.
4. Farringdon, J., and Oni, V. Visual augmented memory (vam). In *Proc. 14th ISWC* (Oct. 2000), 167–168.
5. Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In *Proc. 12th European Conference on Computer Vision* (2012).
6. Iwamura, M., Sato, T., and Kise, K. What is the most efficient way to select nearest neighbor candidates for fast approximate nearest neighbor search? In *Proc. 14th International Conference on Computer Vision* (Dec. 2013).
7. Kawamura, T., Fukuhara, T., Takeda, H., Kono, Y., and Kidode, M. Ubiquitous memories: a memory externalization system using physical objects. *Personal and Ubiquitous Computing 11*, 4 (Apr. 2007), 287–298.
8. Ke, Y., and Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2 (June 2004), 506–513.
9. Kurze, M., and Roselius, A. Smart glasses linking real live an social network's contacts by face recognition. In *Proc. 2nd Augmented Human* (Mar. 2011).
10. Lamming, M., and Flynn, M. "forget-me-not" — intimate computing in support of human memory. In *Proc. '94 FRIEND21* (Feb. 1994).
11. Lee, H., Smeaton, A. F., O'Connor, N. E., Jones, G., Blighe, M., Byrne, D., Doherty, A., and Gurrin, C. Constructing a sensecam visual diary as a media process. *Multimedia Systems 14*, 6 (2008), 341–349.
12. Utsumi, Y., Kato, Y., Kunze, K., Iwamura, M., and Kise, K. Who are you? — a wearable face recognition system to support human memory. In *Proc. 4th Augmented Human* (Mar. 2013), 150–153.
13. Utsumi, Y., Sakano, Y., Maekawa, K., Iwamura, M., and Kise, K. Fast face recognition with large-scale database based on local feature and approximate nearest neighbor search. Tech. Rep. 2013-CVIM-186 (4), Information Processing Society of Japan (IPSJ), Mar. 2013. http://id.nii.ac.jp/1001/00090867/. In Japanese.
14. Viola, P., and Jones, M. J. Robust real-time face detection. *International Journal of Computer Vision 57*, 2 (May 2004), 137–154.
15. Wang, X., Zhao, X., Prakash, V., Shi, W., and Gnawali, O. Computerized-eyewear based face recognition system for improving social lives of prosopagnosics. In *Proc. 7th PervasiveHealth* (May 2013), 77–80.