Keyframe Detection in Visual Lifelogs

Michael Blighe, Aiden Doherty, Alan F. Smeaton and Noel E. O'Connor
Centre for Digital Video Processing, Adaptive Information Cluster
Dublin City University, Ireland
{blighem, oconnorn}@eeng.dcu.ie, {adoherty, alan.smeaton}@computing.dcu.ie

ABSTRACT

The SenseCam is a wearable camera that passively captures images. Therefore, it requires no conscious effort by a user in taking a photo. A Visual Diary from such a source could prove to be a valuable tool in assisting the elderly, individual's with neurodegenerative diseases, or other traumas. One issue with Visal Lifelog's is the large volume of data generated (approximately 650,000 images captured per year). In previous work, we split a days worth of images into more manageable segments, i.e. into distinct events or activities. However, an event will still consist of 80-100 images, thus, in this paper we propose a novel approach to selecting the key images within an event using a combination of MPEG-7 and Scale Invariant Feature Transform (SIFT) features.

Keywords

Visual Diary, Health Management, SIFT

1. INTRODUCTION

An initial report from Microsoft has demonstrated how the SenseCam can be used in order to assist people with short term memory loss [6]. The most widespread neurodegenerative diseases are Alzheimer's and Parkinson's [1]. Alzheimer's disease is an irreversible neurodegenerative disorder that progressively degrades the brains ability to maintain normal executive, attention, and memory functions. A treatment that could delay the onset of Alzheimer's by 5 years would reduce the number of sufferers by 50% in 50 years.

Besides assisting with neurodegenerative diseases, there is a growing belief that technology can be used to address the growing problem of in-home care for the elderly. We believe that the use of a Visual Diary could lead to significant improvements in the health and quality of life of elderly people within their own homes. We propose the use of passive capture devices, such as the SenseCam, to assist in this area. A user wears the camera around their neck and the camera takes pictures continuously throughout the day. However, the management of the increased volume of data generated by such devices remains a challenging problem. In previous work, we split a day's worth of images into distinct events or activities [3, 2, 5]; and addressing the health context in particular, we detect events which are of most interest to the user and to health practitioners [4].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...$5.00.

2. RELATED RESEARCH

In health management, an analysis of behavioural factors plays a critical role. For example, previous research has introduced photography into diabetes self-management routines to help patients make their behaviours explicit and to work with physicians to see possible correlations between self medication and long-term health [8]. Past approaches have manually collected images, however, using SenseCam this process could be automated, giving a better understanding of how to improve the diagnoses and treatment of illnesses that are highly influenced by behavioural routines. In addition, many home monitoring technologies have been proposed to detect health crises, support aging-in-place, and improve medical care [7]. The potential costs, and fears over breaches of privacy amongst health professionals and members of the public, mean that these technologies have had a limited impact to date. However, there is some evidence that these systems may be more readily adopted if they are developed as tools for personalised use which help the users learn about the conditions and variables which affect their physical health.

The SIFT descriptor is a gradient orientation histogram robust to illumination and viewpoint changes. In [3], SIFT descriptors are used to detect important *settings* in SenseCam images. In [2, 5], MPEG-7 features were used in conjuction with sensor data to structure collections of SenseCam images into events. Given that an event typically consists of 80-100 images, in this paper we introduce a novel concept for selecting a representative image through the use of SIFT features.

3. EXPERIMENTAL RESULTS

In this preliminary investigation, we examined three typical events taken from the collections of two SenseCam users. Details of how images are segmented into these distinct events can be found in [4]. The three scenarios investigated are known as *Static Scene*, *Random Scene* and *Return Scene*. A *Static Scene* is one where the user is relatively stationary whilst wearing the SenseCam. An example of such a scene is when the user is sitting working at their computer for an extended period of time. There may be some small movements to the left or right, but essentially the scene remains the same. A *Random Scene* occurs where the user is walking around wearing the camera. All the images taken are of random objects and places, depending on where the user is going. A *Return Scene* is one where the user is in one location, moves temporarily to somewhere else, but then returns to the original location. An example could be someone walking around a museum, looking at an artifact, walking away, but then returning to the original artifact for another look. This process could repeat multiple times. For each of these three scenes, the user's selected five examples of each scene from their collections. This gave a total of 30 different scenes, consisting of 2,178 images, for these experiments.

Given the low quality of SenseCam images [2], we determine the quality of each and every SenseCam image based on its contrast and salience properties [4]. Given an *image quality* score for each and every image we then apply the Kapur adaptive thresholding technique to select a subset of images from the event that are of a sufficiently high *image quality*.

Once this process had been completed, the SIFT features were extracted from the remaining images in each event. In order to match features between images, the distance ratio test was used [3]. To examine whether a point from the 1st image has a match in the 2nd, it's two most similar descriptors in the 2nd image are found. If the ratio of the nearest distance to the second nearest distance is less than 0.7, a match is declared. The number of matches between an image and all other images in the event are summed, and then the average number of matches is calculated. The image which has the highest average is deemed to be the most similar to all other images in the event and, hence, is selected as the keyframe for that event.

In order to evaluate this approach, the results were compared to another approach to selecting a keyframe from SenseCam events. This approach involved simply selecting the middle image from each event. For this preliminary work, these results were qualitatively analysed by both users. For images from the *Random Scene* or *Static Scene*, there was little difference in the performance of both techniques. It's difficult to say whether a selected keyframe image is correct or incorrect in the *Random Scene*. For the *Static Scene*, we looked for images which captured the entire scene in question (as opposed to, for example, an image showing just a small portion of a computer screen). However, the *Return Scene* did produce a discrepancy between our approach and simply selecting the middle image from an event. An example from one scene is shown in Figures 1(a) & 1(b). Both images were taken in a similar location, however, the image selected using our approach is more semantically meaningful to the user than that selected using the middle keyframe. Both images show the user walking along a river side, however, Figure 1(b) shows detail of a particular event which occurred on the river. Figure 1(a) could have been taken on any day - there is nothing in the image to tie it to this particular event.

4. CONCLUSIONS

This work is at a preliminary stage and much more work remains to be done. Numerous strategies exist for selecting representative images from a collection and we intend to perform a much more detailed set of experiments to compare our approach to other approaches from the literature. In addition, we also intend to explore how this technology can be used to improve the quality of life of those with memory difficulties or requiring in-home care.

(a)

(b)

Figure 1: (a) Middle image selected; (b) Image selected using new approach

Acknowledgments

The research leading to this paper was supported by the European Commission under contract FP6-027026 (K-Space), Microsoft Research and Science Foundation Ireland under grant number 03/IN.3/I361.

5. REFERENCES

[1] M. Beal, A. Lang, and A. Ludolph. *Neurodegenerative Diseases: Neurobiology, Pathogenesis and Therapeutics.* Cambridge Press, July 2005.

[2] M. Blighe, H. L. Borgne, and N. O'Connor. Exploiting context information to aid landmark detection in sensecam images. In *2nd International Workshop on Exploiting Context Histories in Smart Environments - Infrastructures and Design (ECHISE)*, September 2006.

[3] M. Blighe, N. O'Connor, H. Rehatschek, and G. Kienast. Identifying different settings in a visual diary. In *9th International Workshop on Image Analysis for Multimedia Interactive Services*, May 2008.

[4] A. Doherty, D. Byrne, A. Smeaton, G. Jones, and M. Hughes. Investigating keyframe selection methods in the novel domain of passively captured visual lifelogs. In *ACM International Conference on Image and Video Retrieval*, Niagara Falls, Canada, July 2008.

[5] A. Doherty, A. Smeaton, K. Lee, and D. Ellis. Multimodal segmentation of lifelog data. In *RIAO 2007 - Large-Scale Semantic Access to Content (Text, Image, Video and Sound)*, Pittsburgh, PA, USA, 30 May - 1 June 2007 2007.

[6] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. Sensecam: A retrospective memory aid. In *Eighth International Conference of Ubiquitous Computing*, September 2006.

[7] S. Intille, K. Larson, J. Beaudin, E. M. Tapia, P. Kaushik, J. Nawyn, and T. McLeish. The placelab: a live-in laboratory for pervasive computing research. In *Pervasive 2005 Video Program*, May 2005.

[8] B. Smith, J. Frost, and M. A. R. Sudhakar. Improving diabetes self-management with glucometers and digital photography. *Personal and Ubiquitous Computing*, 2006.