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, individuals with neurodegenerative diseases, or other traumas. One issue with Visal Lifelogs is the large volume of image data generated. In previous work, we split a day's worth of images into more manageable segments, i.e. into distinct events or activities. However, each event could 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, Keyframe Selection

1. INTRODUCTION

An initial report from Microsoft has demonstrated how the SenseCam can be used in research to assist people with short term memory loss [3]. The most widespread neurodegenerative diseases are Alzheimer's and Parkinson's. 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 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, where a user wears the camera around their neck and the camera takes pictures continuously throughout the day. However, the management of the larger volume of image data generated by such devices remains a challenging problem. In previous work, we successfully split each day's worth of images into distinct events or activities [1, 2].

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 [5]. 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 [4]. 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, thus helping users learn about the conditions and variables which affect *their* physical health.

One of the issues associated with visual recording of a lifelog is selecting single images, or keyframes, which appropriately represent real life events. Techniques to address this will require the use of image processing algorithms such as SIFT. The SIFT descriptor is a gradient orientation histogram robust to illumination and viewpoint changes. In [1], SIFT descriptors are used to detect important *settings* in SenseCam images. In [2], MPEG-7 features were used in conjuction with sensor data to structure collections of SenseCam images into events. Given that each event typically consists of up to 80-100 images, in this paper we introduce a novel concept for selecting a representative image through the use of SIFT features.

3. APPROACH AND PRELIMINARY RESULTS

In this investigation, we examined examples of 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 [2]. 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 watching TV 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 their kitchen, looking in a cupboard or fridge, walking away, but then returning to the original location for another look. This process may be repeated several times in a typical kitchen activity like preparing a meal.

Given the low quality of SenseCam images [1], we determine the quality of each SenseCam image based on its contrast and salience properties [2]. Given an *image quality* score for each 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 has been completed, SIFT features are extracted from the remaining images in each event. In order to match features between images, the distance ratio test was used [1]. To examine whether a point from the 1^{st} image has a match in the 2^{nd}, it's two most similar descriptors in the 2^{nd} 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, results were compared to the more traditional approach of selecting the middle images as the keyframe in SenseCam events (see Table 1). For each of the three types of scene, two users selected five examples of each scene from their collections. This gave a total of 30 different scenes, consisting of 3,179 images. After removal of low-quality images, the SIFT keypoints from 2,178 images were analysed from which to select a keyframe image for each event. For this preliminary work, the results were qualitatively analysed by both users. For the *Static Scene*, our approach selects images captured by taking the entire scene into consideration (as opposed to, for example, an image showing just a small portion of a computer screen). There is no one definitive image that should be a keyframe image in the *Random Scene*. In general though, for images from the *Random Scene* or *Static Scene*, there was little difference in the performance of both techniques. Generally, the quality or semantic meaning of the selected keyframe influenced which approach was deemed most successful. 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 semantically more meaningful to the user than that selected using the middle keyframe. Both images show the user on his regular walk along the river side, however, Figure 1(b) shows detail of a particular fishing event which occurred on the river. Figure 1(a) could have been taken on any day as there is nothing in the image to tie it to this particular event.

Approach	Static Scene	Random Scene	Return Scene
Middle Keyframe	0	1	1
New Approach	1	3	5

Table 1: This table show's the number of event's which were judged to be better using the two approaches for each type of scene. There were 30 scene's in total; 10 for each type.

(a) (b)

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

4. CONCLUSIONS

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 techniques. In addition, we also intend to explore the benefits this technology may offer in improving the quality of life of those with memory difficulties or those requiring in-home care.

Acknowledgments

The research leading to this paper was supported by the Irish Research Concil for Science, Engineering, and Technology; Microsoft Research under grant 2007-056; and Science Foundation Ireland under grant number 03/IN.3/I361.

5. REFERENCES

[1] 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.

[2] 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.

[3] 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 on Ubiquitous Computing*, September 2006.

[4] 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.

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