3-D Representation of Video Sequences Acquired from Ubiquitous Sensor Environment

Shoichiro Iwasawa* **Yusuke Otaka†*** **Yasuyuki Sumi†*** **Sadanori Ito*** **Kenji Mase‡***

*ATR Media Information Science Laboratories
†Graduate School of Informatics, Kyoto University
‡Information Technology Center, Nagoya University

mailto:shoichiro.iwasawa@acm.org

ABSTRACT

Keywords
3-D, spatio-temporal, collage, wearable, ubiquitous

INTRODUCTION

With the recent spread of home videos, video data recorded in our everyday life and social events have increased. In order to better utilize such video data, we aim to make 3-D virtual spaces, using the video images as input, where we can re-experience and share our daily experiences and important events with others. The 3-D spaces will enable us to share personal experiences and knowledge at museums, heritage sites, and educational settings. Most existing works [1] for making 3-D virtual spaces have focused on exact reproduction of the modeled objects and space. However, the resulting built spaces tend to be homogenized, making it difficult for virtual visitors to the spaces to find the "scent" needed for truly re-experiencing the scene.

OUR APPROACH

We simply employ a spatio-temporal collage representation for this purpose. The word collage is generally understood as an artistic composition made from various materials applied to a surface, or an assembly of diverse fragments. Photo collage uses photo images as materials or fragments. We assume a test-bed environment [2,3] that is fully equipped with ubiquitous video cameras, which include wearable cameras attached to persons and to ceilings or walls. The test bed can capture and store video sequences from multiple viewpoints we refer to as POV (point of view). The wearable cameras can acquire user attention or intentions visually. Then the proposed spatio-temporal video collage feeds video sequences as the materials. We can "collage" two video frames with different timestamps at the same. Furthermore, in making the collage we have no limitation on the alignment of video frames, regardless of the original viewpoint used in taking the sequences.

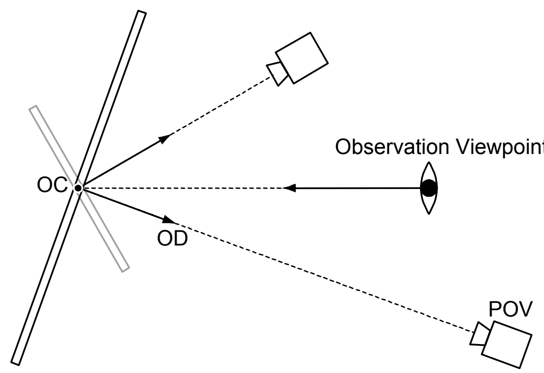

Figure 1: An observation of proposed method

Implementation of "Collage"

For every identified object from the ubiquitous sensor technology (described below), the right place (Object Coordinate; OC) where it should exist and the right direction (Object Direction; OD) where it should turn to are known in advance by using the in-house infrared ID tag sensing technology [2]. Accordingly, images taken from each video frame can be aligned at the OC, and the size of an image is scaled by using the magnification factor linearly corresponding to the distance between the POV and OC. The opacity of each image is defined based on the scalar product of the POV vector and an arbitrary observation view vector. Actually, our implementation uses only horizontal components of the OD, POV, and observation view vectors. Disregarding vertical components makes total operations easier and simpler.

We assume that it is effective for past and future images to be rendered at the same when we observe the specified time frame at a glance. A bigger time lag from the current time specified by the observer would give more transparency to an image. Sometimes an image coordinate, or OC of the object, is not fixed because the object is moving during the time frame, such as when a person is the object.

EXPERIMENT

The proposed ubiquitous sensor technology covers the entire experimental environment. This technology includes wearable sensor sets for users and fixed sensor sets that are distributed throughout the indoor environment. Both sensor sets are capable of continuous video recoding and tracking of infrared ID tags. The ID tracking can acquire a tag's ID and its coordinates in an image plane. The wearable set is equipped with an in-house local positioning system (LPS). LPS is used for locating the position and rotation of the user wearing the sensor set as he or she moves. The LPS output provides the essential OC and OD of the camera on the wearer's set. Currently, the positions and rotations of other sensor sets must be obtained manually.

In-house local positioning system

The pair of an upward-directional ID tracker and a set of ID tags on the ceiling is a major component of our LPS. The positions of all ceiling ID tags are surveyed carefully, so a tracker can identify each tag by using its ID. In the area of computer vision, it is a basically simple and easy problem if you can observe some of the tags. Our LPS can estimate position and rotation when the ID tracker can detect at least four ceiling ID tags at the same time. We employ the POSIT [4,5] algorithm developed by Davis as the core estimation method.

Setup and results

All experiments utilize 60 LPS ceiling ID tags, which were aligned in a lattice order with spacing of 20 cm. Three subjective users equipped with wearable sets and another subjective user with only an ID tag participated in the 10-minute experiment. The user with only an ID tag was considered to have fixed OC and OD throughout the experiment.

After the capturing session, we observed captured interactions between the four subjects using a walkthrough type of viewpoint. Figure 1 shows a snapshot of a generated observation. In this figure, each subject is indicated as a black tube. All rendered images were taken from three wearable set's cameras. The subject marked as a red circle was indicated as the subject who received the visual attention of the other two subjects based on ID tracker data. Consequently, the rendered images are obtained by two cameras. The far left tube has no image at all, implying the subject was not attracting others' attention. From the experiment, the proposed method seems to often represent or highlight a subject's visual attention or a brief intention.

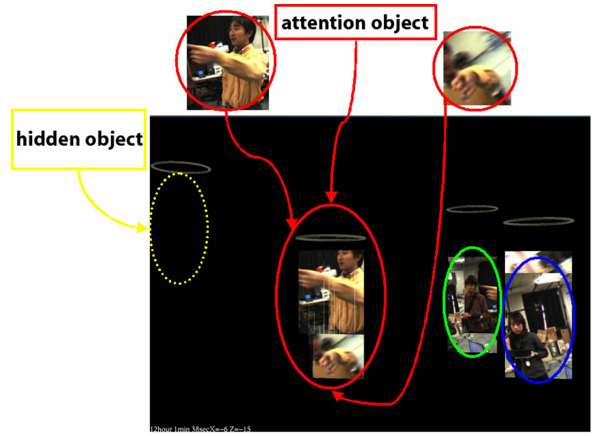

Figure 2: An observation of proposed method

CONCLUSIONS

A spatio-temporal image collage from video sequences is proposed for observing and re-experiencing past events. The 3-D collage representation we have created provides easier identification of subjects' visual attention. We also implemented an infrared ID tag/tracker system based on a local positioning system (LPS) for estimating users' positions and rotations. The LPS ceiling tags are so cheap that we can fully cover large areas. Currently, our LPS gives fairly rough information that is still far from precise. Improving the accuracy of LPS is one of our most important future works. In the near future, LPS for outdoor environments is expected to become available for a wide variety of applications.

ACKNOWLEDGMENTS

This research was supported in part by the National Institute of Information and Communications Technology.

REFERENCES

1. Ishida, T.: Digital City Kyoto: Social Information Infrastructure for Everyday Life. *Communications of the ACM*, Vol. 45, No. 7, pp. 76-81, 2002.
2. Sumi, Y., Ito, S., Matsuguchi, T., Fels, S. and Mase, K.: Collaborative Capturing and Interpretation of Experiences, *Proceedings of Pervasive 2004 Workshop on Memory and Sharing of Experiences* (Linz/Vienna Austria, April 2004), pp. 1-7.
3. Sumi, Y., Matsuguchi, T., Ito, S., Fels, S. and Mase, K.: Collaborative capturing of interactions by multiple sensors, *Ubicomp 2003 Adjunct Proceedings*, pp. 193-194, Oct. 2003.
4. D. DeMenthon, D. Oberkampf and L.S. Davis: Iterative pose estimation using coplanar feature points, *CVGIP: Image Understanding*, Vol. 63, No. 3, pp. 495–511, 1995.
5. D. DeMenthon, and L.S. Davis: Model-based object pose in 25 lines of code, *International Journal of Computer Vision*, Vol. 15, pp. 123–141, 1995.