NON-INTRUSIVE TRACKING OF MULTIPLE USERS IN A SPATIALLY IMMERSIVE DISPLAY

Jiyoung Park, Seon-Min Rhee
Dept. of Computer Science & Engineering, Ewha Womans University, Seoul, Korea

Myoung-Hee Kim
Dept. of Computer Science & Engineering, Ewha Womans University, Seoul, Korea
Center for Computer Graphics and Virtual Reality, Ewha Womans University, Seoul, Korea

Keywords: multi-user tracking, visual tracking, immersive display.

Abstract: We present a novel vision-based system for tracking multiple users in a spatially immersive display. Without requiring them to wear any markers or other devices, we can detect and track the heads of several participants. In a projection-based display environment, the lighting conditions make it difficult to extract silhouettes or shape features from acquired images. Using a separate IR lighting and stereo camera system solves the problem, and makes background subtraction simple and fast. We start by finding general location of the users' heads in each image, from the silhouettes and projection histogram of the foreground regions. These points are used to create search areas, one in each image of a stereo pair. By cross-correlation between the search areas, corresponding points in each image are identified, and these are used to determine an accurate 3D location on the head. Finally, the search areas in consecutive frames are correlated to maintain the identification of the users over time. Experimental results demonstrate the viability of the proposed system.

1 INTRODUCTION

User tracking has been studied extensively in the context of many applications, such as surveillance and virtual reality (VR). In an immersive VR environment, tracking is essential to provide the user with the correct view and accurate interactions with virtual objects. Although there are many user tracking methods, non-intrusive tracking is the most desirable since it offers the user a more natural and comfortable VR environment. Spatially immersive displays are now widely used and their scale is increasing, so that the tracking of multiple users is becoming crucial. But relatively little work has been done on non-intrusive multiple user tracking in immersive environments. While Vorozcovs et al. (2005) recently developed an optical tracking method for a spatially immersive display, it can only follow a single user and requires a special device to be worn on the user's head. There are two main reasons for the lack of work on user tracking in large VR environments. First, a static background cannot be guaranteed because the image projected on the display screens also results in continuously varying illumination of the users, which is disastrous for robust feature extraction. Secondly, possible camera set-ups are very restricted because the camera must not occlude the projection screen. One option is to attach cameras just above the screens. But this configuration does not allow general feature-based tracking methods to provide stable results, because important features of the face or body of a user are not always seen in the resulting image sequences.

In this paper, we introduce a vision-based approach for detecting and tracking multiple users in a spatially immersive display. A generic CAVE™ system was used, with no additional custom hardware. The lack of intrusive devices encourages intuitive interaction between multiple users and maximizes the capacity of collaborative environments. We believe it will motivate the

Park J., Rhee S. and Kim M. (2006).
NON-INTRUSIVE TRACKING OF MULTIPLE USERS IN A SPATIALLY IMMERSIVE DISPLAY.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 464-467
DOI: 10.5220/0001372204640467

enhancement of existing VR applications, and the use of VR technology in new areas of practical significance.

2 FOREGROUND EXTRACTION IN A SPATIALLY IMMERSIVE DISPLAY

Our method of tracking users in an immersive display is based on the generation of infrared images which are not affected by the time-varying images on the screens, which are in the visible system. This allows us to extract the foreground of the scene (i.e. the users in the VR environment) and we can then track the position of their heads by employing computer vision techniques.

2.1 Hardware Setup

We use a four-sided CAVE™-like environment (with a front, left, right and bottom screen, each measuring 2.4m x 2.4m) for our experiments. We located two IR lights and cameras on top of the front screen as depicted in Fig. 1. The directions of the two lights are set differently to illuminate the top and bottom part of a body in the environment, as evenly as possible. A band-pass filter is attached to each camera so that it only receives IR light. A non-IR reflecting curtain in the entry to the environment improves the IR segmentation results. We used Computar IR75 lamps and Point Grey Dragonfly grayscale video cameras with an IEEE-1394 interface.

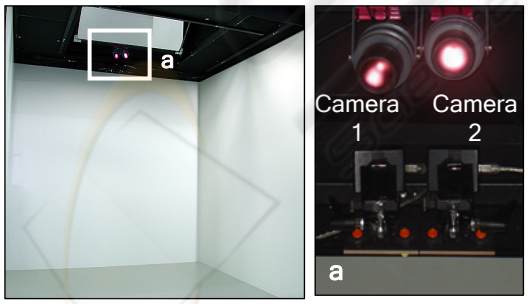

Figure 1: Hardware setup.

2.2 Foreground Segmentation using IR

Having eliminated the changing background by the use of IR, the grayscale images from the cameras are input to a background subtraction method optimized for infrared light (based on Matusik 2001). First, a sequence of n frames of the static background is recorded and the mean and standard deviation of each pixel in the image are calculated. Then the intensity of each pixel is compared with the mean value at every frame. If the modulus of the difference between them is greater than a predefined multiple k of the standard deviation, the pixel is classified as foreground. In a post-processing stage, a median filter is used for noise removal and the resulting silhouette is smoothed. An example of an extracted silhouette is shown in Fig. 2.

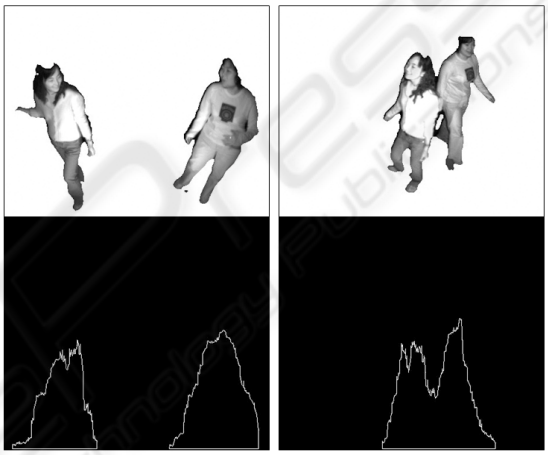

Figure 2: Examples of extracted silhouette(first row) and projection histogram.

3 MULTI-USER HEAD TRACKING

Our method of detecting multiple users from infrared reflective images was inspired by 'Hydra' (Haritaoglu et al. 1999), which tracks several people using silhouettes and a projection histogram of the foreground regions.

At every frame, we extract the foreground regions by subtracting background, and then generate a projection histogram from the regions we have detected. A vertical projection histogram is subsequently computed by projecting the binary foreground region on to the horizontal axis, as depicted in Fig. 2. Combining the silhouette and histogram from each image of a stereo pair, we are able to identify each user and locate their head position in three dimensions. At present, the system has to be told the number of users before the detection process starts.

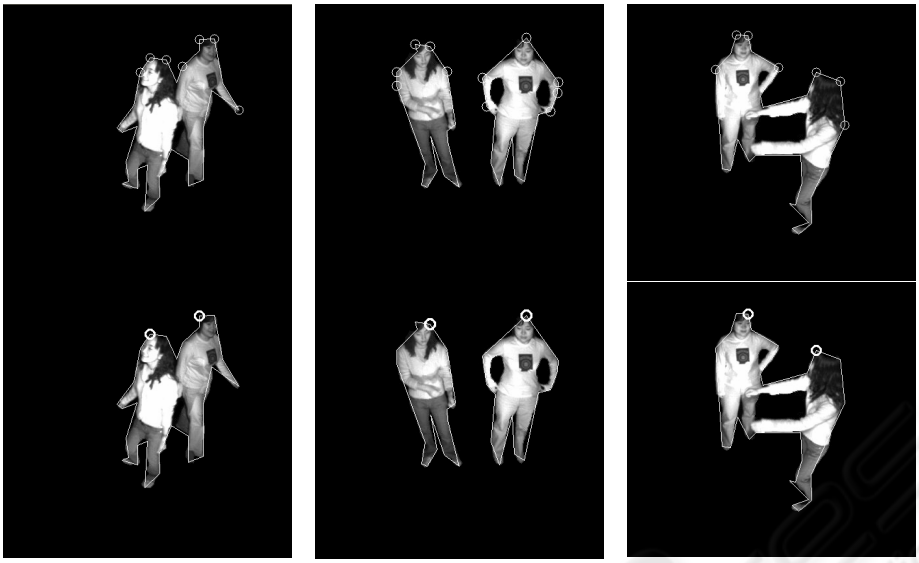

Figure 3: An example of candidate heads(first row) and detected head points.

3.1 Head Detection using Silhouettes and a Projection Histogram

Location of the users' heads begins with an approximated outline of the foreground boundary (Fig. 3). Due to the structure of the human body, the silhouette boundary will have salient points corresponding to body parts such as the head, hands and feet. We expect heads to be at the top of the foreground area, and to have a higher value than the mean in the projection histogram.

To detect the head, we check the convexity of each pair of adjacent line segments on the contours. If a junction is convex, and its vertical position is higher than the centroid of the whole area, it is classified as a candidate head. When all candidate heads have been found, we use the projection histogram to make a final choice of final head points. If more heads are detected than the number specified, we prune surplus heads in the order of their histogram value. Most 'false' heads turn out to be hands..

3.2 Stereo Correspondence and Tracking

Next, the spatial location of heads is calculated. In each stereo pair, we should now have two corresponding points to the top of each user's head. However, because these points were determined from the approximated silhouette, the exact three-dimensional head position cannot be obtained by using them as corresponding points. Instead, we need to find corresponding points in the original images of the stereo pair. We therefore create a search area around each head point which was finally determined in the last step. And we set a mask on the search area and cross-correlate the mask regions of the same size in each search area. Moving two masks in each whole search area and computing cross-correlation coefficient value, we select the most similar from each image regions, and make the centers of the mask areas the corresponding points.

Finally, in order to identify each user in every frame, we analyze the correlation of each search area in the current frame with those in the previous frame. The two consecutive search areas with the highest correlation are assumed to correspond to the same person.

4 EXPERIMENTAL RESULTS

The result of tracking two users in a VR environment is shown in Fig. 4. In this sequence, one user moves from left to right across the other user. Corresponding points on the two users' heads are found accurately, and tracking copes well with the users' movements.

Our tracking method runs at 20-25Hz on a PC with dual 2.8GHz Xeon processors. The Intel OpenCV library was used to handle 640x480 resolution grayscale images. Execution times depend on the number of participants, and the size of the

Figure 4: Examples of stereo correspondence and tracking : the view of camera#1(first row) and camera#2 in Fig. 1.

search area, and the mask used in cross-correlation. We set the size of the search area to 30×30 pixels and the mask to 20×20 pixels in this experiment.

5 CONCLUSIONS

We have proposed a vision-based tracking method for multiple users in a spatially immersive display. It requires no markers and detection and tracking are robust and fast.

However, our system cannot guarantee perfect tracking results if some of the users are totally occluded. In a future enhancement of our system, we plan to use depth information to separate and identify users correctly.

In addition, we are looking for ways of recognizing body parts more effectively in acquired images. Because the viewpoint of the cameras is high, the images are foreshortened, which makes it difficult to extract and identify body parts accurately. One possible solution is to warp the acquired images, making them more similar to those that would be acquired from a lower camera position, and thus showing the body more clearly.

Eventually, we intend to develop a VR application which provides collect images of the virtual scene to each user, based on the viewing positions obtained from our system.

6 ACKNOWLEDGEMENTS

This work was supported by the Korean Ministry of Information and Communication under the Information Technology Research Center (ITRC) Program.

REFERENCES

Haritaoglu, I., Harwood, D., Davis, L. S., 1999. Hydra: Multiple people detection and tracking using silhouettes. In *IEEE International Workshop on Visual Surveillance*. IEEE Computer Society Press.

Matusik, W., 2001. *Image-based visual hulls*. Master's thesis, Massachusetts Institute of Technology.

Vorozcovs, A., Hogue, A., Stuerzlinger, W., 2005. The Hedgehog: a novel optical tracking method for spatially immersive displays. In *IEEE VR 2005*. IEEE Computer Society Press.