Preprint
UCRL-JC-141131

Feedback From Video From Virtual Reality Navigation

L. V. Tsap

This article was submitted to
Second International Workshop on Digital and Computational Video
Tampa, FL
February 9, 2001

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

October 27, 2000

FEEDBACK FROM VIDEO FOR VIRTUAL REALITY NAVIGATION[1]

Leonid V. Tsap

Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
University of California
Livermore, CA 94551, U.S.A.
e-mail: tsap1@llnl.gov

Abstract

Important preconditions for wide acceptance of virtual reality (VR) systems include their comfort, ease and naturalness to use. Most existing trackers suffer from discomfort-related issues. For example, body-based trackers (hand controllers, joysticks, helmet attachments, etc.) restrict spontaneity and naturalness of motion, while ground-based devices (e.g., hand controllers) limit the workspace by literally binding an operator to the ground. There are similar problems with controls. This paper describes using real-time video with registered depth information (from a commercially available camera) for virtual reality navigation. Camera-based setup can replace cumbersome trackers. The method includes selective depth processing for increased speed, and a robust skin-color segmentation for accounting illumination variations.

I. MOTIVATION AND NEW APPROACH

In virtual reality (VR) systems, the human operator is connected to a computer that simulates a real or an imaginary environment and enables the person to perform real tasks in such environment [1]. Examples include flight simulators, virtual product design, learning environments, telerobotics, telemedicine, games, etc. An important part of any VR system is position tracking and mapping [1], where tracking means the real-time position and orientation estimation of a moving object, and mapping means surface measurements. Depending on the application, the importance of tracking certain body parts is fundamental to its success, for instance, head and eye tracking for visual displays, hand and arm tracking for haptic interfaces, body surface mapping for videoconferencing, etc.). Tracking human body parts is also vital in the context of controlling computer-generated (or remote) objects (for example, in telerobotics).

Most existing trackers suffer from discomfort-related issues. For instance, mechanical trackers are inexpensive and reasonably accurate. However, body-based trackers (hand controllers, joysticks, helmet attachments, etc.) restrict spontaneity and naturalness of motion, while ground-based devices (e.g., hand controllers) limit the workspace by literally binding an operator to the ground. Another disadvantage is the limited number of degrees of freedom (DOFs) that can be measured (when multiple limbs are moving).

New Approach

To address these, problems, a video-based tracking can be used. It is known that using optical sensors for position tracking and mapping would address these issues. However, most approaches still employed body markers for easy recognition, some used structured light systems and lasers, but did not give passive light systems proper consideration.

Augmented-reality systems are being studied in medicine [1] to present surgeons with visual displays that combine dynamically changing views of the patient during the operation with previously collected detailed graphical data from various tests and scans. For example, a surgeon's view of a brain surgery can be overlaid with images from earlier CAT scans and real-time ultrasound. In a case like this most trackers would certainly restrict surgeon's hand sensitivity. Using a set of tiny cameras with real-time application-specific software responsible for the data analysis and display would be much more beneficial.

Similarly, traditional controls can also be replaced with gestures. Conventional controls include 3- and 6- dimensional mice/trackball/joystick devices. They differ from their desktop equivalents by having extra buttons and wheels that are used to control not just the XY translation of a cursor, but its Z dimension and rotations in all three directions. This paper shows that tracking and control hardware can be replaced with optical analysis of human gestures.

Previous Work

But what is the best way to include optical data into the system? And, most importantly, what data is needed? Traditional approaches to tracking typically relied on segmentation of the intensity data, using motion or appearance data. A majority of the methods began by segmenting the human body from the background. Wren *et al.* [2] achieved segmentation by classifying pixels into one of several models, including a static world and a dynamic user represented by gaussian blobs. Yang and Ahuja [3] used skin color and the geometry of palm and face regions for segmentation stages of their system. Yacoob and Black proposed parameterized representation of human movement [4]. Cutler and Davis [5] segmented the motion and computed a moving objects self-similarity (including human motion experiments). A review by Pavlovic *et al.* [6] addressed main components and directions in gesture recognition research for HCI.

Most gesture-tracking and recognition applications could certainly benefit from including depth data and having more information recovered from a scene. Until recently, however, it was not feasible because of the speed and cost considerations. Recent availability of less expensive, faster range data makes it a feasible additional source of information for tracking. Oda *et al.* [7] reported application of a real-time range to virtual reality which utilized comparison of the depth information in real and synthetic data. In addition to efficient range processing, the proposed method also deals with the major shortcoming of color-based localization methodologies variability of the skin-color classification results under different illumination conditions.

II. ALGORITHM AND EXPERIMENTAL RESULTS

Figure 1 shows that optical trackers/controls can be used instead of or in addition to other control and tracking elements. Combination of these two functions points to the efficiency of the approach in addition to the comfort issues already discussed. Imaging data used in the system includes registered color and depth images. Algorithm involves the following steps (see [8] for details):

- noise removal,
- region classification by color information,
- range processing of moving skin-colored regions, and
- motion/ manipulation/ action computation.

The following experiments involve application of the algorithm to color and range image sequences of gestures. Triclops color stereo vision system (manufactured by Point Grey Research, Vancouver, Canada) is used to capture these sequences. The module connects to a single-processor Pentium III PC. Range information is recovered in real time from a correlation-based trinocular stereo algorithm. Typical color and range images produced by the stereo vision system are shown in Figure 2. Closer objects appear lighter in the range data, except for the darkest areas (for instance, some hair and far wall regions), indicating that no correspondence was found during the stereo matching process.

As a result of applying the first two steps of the algorithm, skin regions are selected (shown as rectangular enclosing boxes in Figure 3(top row)). Figure 3(bottom row) shows depth computation for chosen regions

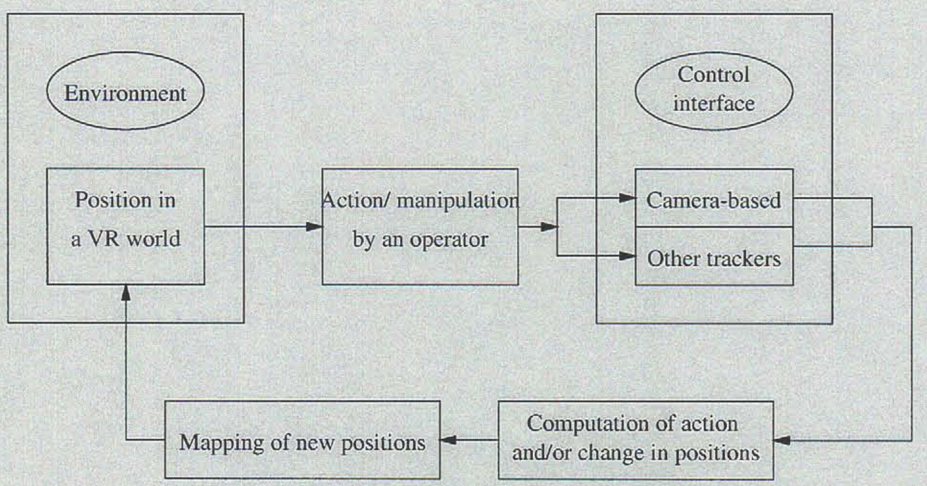

Fig. 1. Integration of video processing into a VR system.

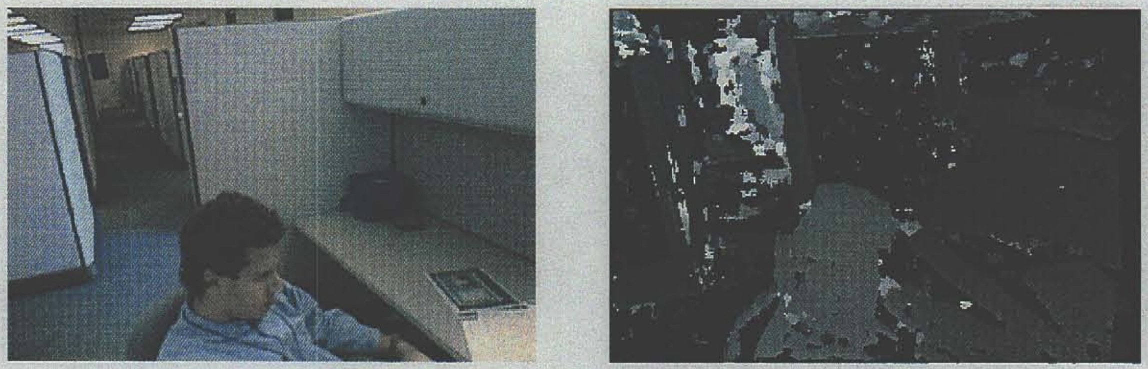

Fig. 2. Typical color and depth images produced by the stereo vision system.

(moving hand in this experiment). Selective depth estimation is necessary for faster processing (see [8] for specific details). Computed 3-D coordinates of hand positions are transfered to a VR system (currently the transfer is not in real time). Figure 4 shows applying computed hand motion to the virtual hand (view from above).

III. SUMMARY AND FUTURE WORK

One of the main goals of the VR system is immersion of the user in the synthesized environment with a strong sense of presence and interaction. Important preconditions for wide acceptance of virtual reality (VR) systems include their comfort, ease and naturalness to use. Most existing trackers suffer from discomfort-related issues. For example, body-based trackers (hand controllers, joysticks, helmet attachments, etc.) restrict spontaneity and naturalness of motion, while ground-based devices (e.g., hand controllers) limit the workspace by literally binding an operator to the ground. There are similar problems with controls. This paper describes using real-time video with registered depth information (from a commercially available camera) for virtual reality navigation. Camera-based setup can replace cumbersome trackers. The method includes selective depth processing for increased speed, and a robust skin-color segmentation for accounting illumination variations. The proposed approach was successfully tested for basic hand gestures. It will be applied to a larger set of gestures responsible for control of a virtual environment.

ACKNOWLEDGMENTS

This was done as a part of SAVAnTS scalable visualization project. I would like to thank members of

Fig. 3. Tracking of skin-color regions (top row) and depth computation (bottom row) for zoom gesture (hand moving towards the camera).

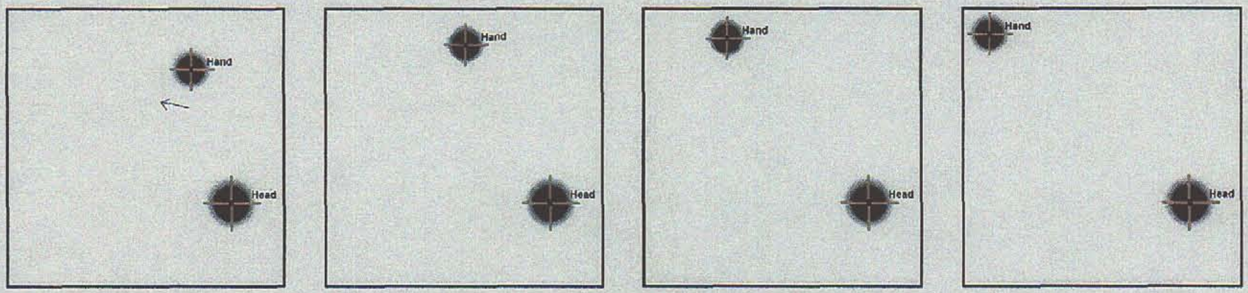

Fig. 4. Applying computed hand motion to the virtual hand (view from above).

Visualization Group of the Center for Applied Scientific Computing at the LLNL for their support and suggestions; and Dmitry Goldgof and Min Shin at the Vision Lab (USF) for assistance with the face tracker.

REFERENCES

[1] N. I. Durlach and A. S. Mavor, editors. *Virtual Reality: Scientific and Technological Challenges.* National Academy Press, Washington, DC, 1995.

[2] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on PAMI*, 19(7):780–785, July 1997.

[3] M.-H. Yang and N. Ahuja. Recognizing hand gestures using motion trajectories. In *Proceedings of IEEE CS Conference on Computer Vision and Pattern Recognition*, volume 1, pages 466–472, Fort Collins, CO, June 1999.

[4] Y. Yacoob and M. J. Black. Parameterized modeling and recognition of activities. *Journal of Computer Vision and Image Understanding*, 73(2):232–247, 1999.

[5] R. Cutler and L. Davis. Real-time periodic motion detection, analysis, and applications. In *Proceedings of IEEE CS Conference on Computer Vision and Pattern Recognition*, volume 2, pages 326–332, Fort Collins, CO, June 1999.

[6] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. on PAMI*, 19(7):677–695, July 1997.

[7] K. Oda, M. Tanaka, A. Yoshida, H. Kano, and T. Kanade. A video-rate stereo machine and its application to virtual reality. In *Proceedings of ISPRS '96*, 1999.

[8] L. V. Tsap. Real-time local range on-demand for tracking gestures. In *Proceedings of IEEE Workshop on on Human Modeling, Analysis and Synthesis*, Hilton Head Island, SC, June 2000.

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.