Tracking and Interaction Based on Hybrid Sensing for Virtual Environments

Dongsik Jo, Yongwan Kim, Eunji Cho, Daehwan Kim, Ki-Hong Kim, and Gil-Haeng Lee

We present a method for tracking and interaction based on hybrid sensing for virtual environments. The proposed method is applied to motion tracking of whole areas, including the user's occlusion space, for a high-precision interaction. For real-time motion tracking surrounding a user, we estimate each joint position in the human body using a combination of a depth sensor and a wand-type physical user interface, which is necessary to convert gyroscope and acceleration values into positional data. Additionally, we construct virtual contents and evaluate the validity of results related to hybrid sensing-based whole-body tracking of human motion methods used to compensate for the occluded areas.

Keywords: Virtual reality, tracking, interaction, depth sensor, physical user interface (PUI), hybrid tracking.

I. Introduction

Virtual reality systems have been widely and successfully used for the purposes of training and usability evaluations in industry. In addition, as an essential factor for interaction under virtual environments, tracking technology placed on body parts, including the hands, is being consistently developed [1], [2]. Markerless-based tracking with depth sensors is useful for user interaction in virtual reality systems and is helpful in the effort to overcome the inconvenience of wired interface devices and attached markers used in tracking in magnetic and optical methods. Moreover, this supports solutions for such problems as poor lighting conditions and slow refresh rates [3]. However, a depth-based method incurs an occlusion problem away from sensing areas where the user cannot obtain depth information of human joints covered by the torso, such as when their arms are behind their body.

Wren and others introduced a real-time tracking system of the human body using the color image of the video input, which allowed a meaningful, interactive-rate interpretation of the human form [4]. Liu and others proposed a sensing method of multiperson motions based on the segmentation of multiview videos with the purpose of interaction [5]. Recently, research on depth images supporting markerless tracking and high-performance algorithms have widely increased in number. Plagemann and others presented a system that provides the detection of human body parts from depth images and an estimation of the 3D locations of the hands, feet, head, and so on [6]. Shotton and others proposed a method for estimating 3D positions from a single depth camera that can calculate a depth map with infrared light, not with temporal information but by training a dataset of body parts [7]. Khoshelham presented a measurement method for the performance of depth data obtained from a depth sensor [8]. Additionally, many researchers have made efforts to overcome occlusion problems owing to a limitation of camera viewpoints and missed sensing areas from the user's body. Kwak and others suggested an algorithm for tracking an occluded object in video sequences through learning [9]. In [10], an occlusion-handling method that estimates the motion of adjacent body parts based on the iterative closest point and particle filter was presented. However, these previous studies were not able to robustly track all human joints including occlusion spaces with a single depth sensor,

Manuscript received Apr. 24, 2012; revised Nov. 15, 2012; accepted Dec. 7, 2012.

This work was supported by the R&D program of Ministry of Culture, Sports and Tourism (MCST) and Korea Evaluation Institute of Industrial Technology (KEIT) (10039923, Development of Live4D contents platform technology based on expansion of realistic experiential space).

Dongsik Jo (phone: +82 42 860 1847, dongsik@etri.re.kr), Yongwan Kim (ywkim@etri.re.kr), Ki-Hong Kim (kimgh@etri.re.kr), and Gil-Haeng Lee (ghlee@etri.re.kr) are with the Creative Content Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Eunji Cho (smcm07@postech.ac.kr) is with the Intelligent Media Lab., POSTECH, Pohang, Rep. of Korea.

Daehwan Kim (daehwank@etri.re.kr) was with the Intelligent Media Lab., POSTECH, and is now with the Creative Content Research Laboratory, ETRI, Daejeon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.13.0212.0170

owing to the sensor being out of sight and sensing limitations.

In this letter, we present a tracking method using hybrid sensing based on a single depth camera and a physical user interface (PUI) with the purpose of interaction under virtual environments, which is capable of finding real-time human motion including occluded spaces. In terms of the PUI, which is used for the tracking of occluded spaces of body parts where these areas cannot obtain the needed information from a depth camera, we estimate positional data by mathematically converting angular velocities obtained from the sensor. We also detect occluded joints in real-time and estimate the positions of other joints, beginning with the hand using PUI with an inverse kinematics (IK) algorithm.

II. Hybrid Sensing-Based Tracking

Figure 1 shows an analysis of occlusion areas through static and dynamic motion using depth sensing. First, we consider a virtual avatar-based detection of occluded joints as the inputs of the distance between a virtual depth camera and the user, the user's body size, joint position on the skeleton, field of view, and view volume of the depth camera under static motion. We also test a situation of dynamic motion with motion velocity and the user's posture, which is the angle between the direction of the depth camera and the gazing direction of the user. As a result, the fast velocity of the user and the hidden areas of the human joints cannot be detected, which turns out to match our initial assumption. In particular, we also find that the position of either hand cannot be tracked when the two hands overlap.

Figure 2 shows our PUI, which can print out real-time raw data on gyroscope and acceleration values for tracking occlusion areas behind the depth sensors and uses verification software for accuracy and performance between a virtual and real PUI. However, because the PUI with a general configuration of 3-axis accelerations and 3-axis angular velocities cannot acquire the absolute position, we transfer from PUI outputs to positional values using a mathematical approach for estimations through a double integral of acceleration. We obtain the hand position of the occlusion spaces using the values of acceleration and execute a noise reduction of the original value and an elimination of gravity elements for accurate datasets. Figure 3 shows signal graphs of the original acceleration, noise reduction using the original information, and the elimination of gravity elements.

To extract motion data from our PUI controller for noise reduction in the first steps, we execute a moving average method including the next number following the original subset in the series [11]. For instance, the estimation equation

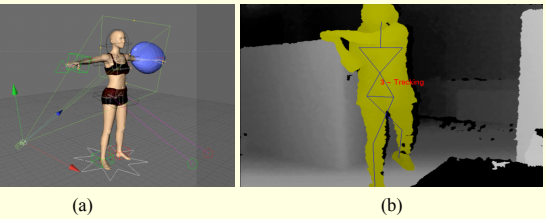

Fig. 1. Analysis of occlusion areas through (a) static and (b) dynamic motions using depth sensing. In the case of dynamic motion, when angle between depth camera and user's hand is nearly 90°, we find that positions of arm connecting hand cannot be estimated.

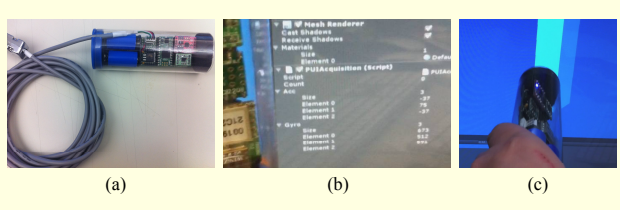

Fig. 2. (a) PUI attached to gyroscope and acceleration sensors, (b) real-time outputs of gyroscope and acceleration values of PUI, and (c) test platform for verifying movement of virtual thick stick shaped like real PUI.

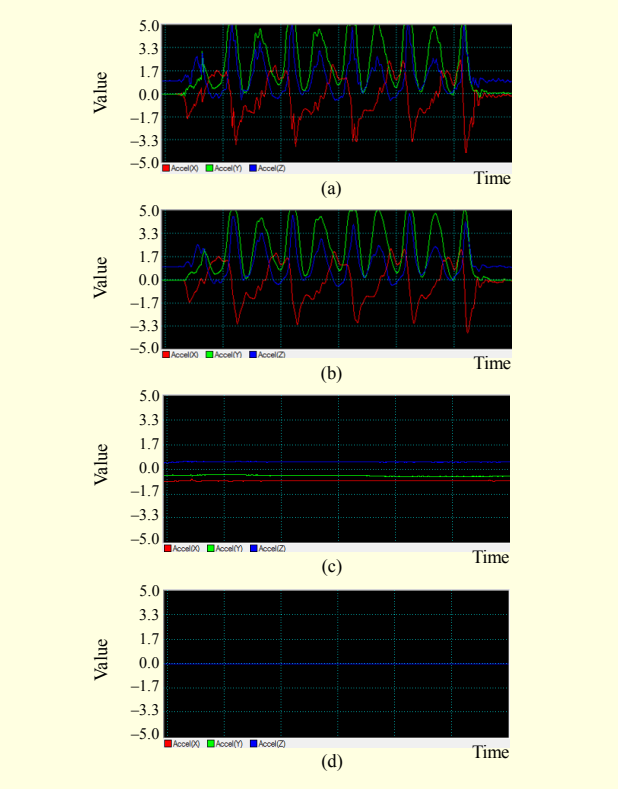

Fig. 3. (a) Original values of acceleration with time axis, (b) signal after noise reduction, (c) values of acceleration including gravity elements, and (d) signal after elimination of gravity elements.

Algorithm Motion_Tracking.

Problem Definition. Detection and tracking of human joints including the occluded spaces.
Input. Sensing information from both a depth camera and a physical user interface.
Output. Positional datasets of human joints.
Approach.
1) Install a depth camera, hold a PUI.
2) Start an application, gather sensor data.
3) Repeat steps 3.1 to 3.4 until the application exists.
  3.1) Detect occluded joints by confidence of depth information.
  3.2) Estimate the position of hand grabbing the PUI if there are occluded joints: convert from angular velocity to positional datasets *(handpos.x,handpos.y,handpos.z)*
  3.3) Estimate the position of elbow and shoulder using IK.
  3.4) Print out the position data of all joints *(jointpos.x, jointpos.y, jointpos.z)*

Fig. 4. Algorithm for hybrid sensing-based tracking.

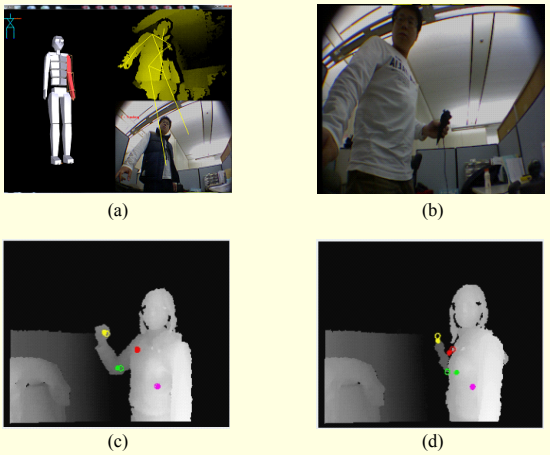

Fig. 5. (a) Results of hybrid sensing-based tracking: detection of occluded areas of human body parts, where the red signs of 3D biped model indicate occluded areas; (b) hybrid sensing-based tracking with depth sensor and PUI; (c) estimated position of hand, elbow, and shoulder using IK when joints are not occluded (unfilled circles indicate positions estimated by depth information and filled circles indicate positions estimated by IK); and (d) estimated position when some joints are partially occluded.

of tracking is as follows:

$$M_t = (p_t + p_{t-1} + p_{t-2} + ... + p_{t-n-1}) / n.$$

Here, M_t is the recalculated value from the moving average method, P_t represents an original value from a point in time, and n is the number of frames adopted for the samples. We use five frames to consider the unique characteristics of continuous datasets and eliminate the gravity elements from the values of the original acceleration. However, since double integration with the acceleration of the PUI produces calculus errors through cumulative effects, we use an integration of the angular velocity to overcome the problem of estimation. The prediction equation of the hand position with angular velocity is derived from

$$\theta_t \approx \int V_t dt\ ,\ \ \theta_t = \theta_{t-1} + \Delta t \times V_{t-1} + (\Delta t \times (V_t - V_{t-1})) / 2\ ,$$

$$Hand_{x\_t} = Hand_{x\_t-1} + W_x \times \Delta\theta_t\ .$$

Here, θ_t represents the angle value of the PUI at moment t, V_t is the value of angular velocity, $\Delta t \times V_{t-1}$ represents the size of the increased areas from the point of time t–1, $(\Delta t \times (V_t - V_{t-1}))/2$ represents the size of errors generated during integration using the trapezoidal method, $Hand_{x\_t}$ is the estimated hand position of the current frame, $Hand_{x\_t-1}$ is the hand position regarding the previous frame, and W_x represents the weight of the distance related to the angle.

Figure 4 shows our algorithm for hybrid sensing-based tracking including occluded areas using a depth camera and a PUI. Although a depth camera can detect the positions of most human joints, this has a limitation regarding self-occlusion. For example, the elbow and shoulder positions occluded by the body cannot be estimated precisely. For this reason, using the confidence values of each joint, we detect whether the joint is occluded. When areas are occluded, the position of the occluded hand, which is located on the opposite side of the depth camera, is calculated using the angular velocity of the PUI. The position of the occluded elbow or shoulder is also then estimated from the end effector, which is the occluded hand, using the method of damped least squares of IK [12].

Figure 5 shows the results of hybrid sensing-based tracking using a depth camera and a PUI. As a result, we estimate the positional datasets for joints within 2 cm using the IK algorithm and for the hand within 9 cm using hybrid sensing, which is the value converted into distance when comparing the coordinate values of depth and the estimated values of the PUI.

III. Virtual Contents Based on Hybrid Sensing and Evaluation

We measure the positional data of human joints when the user's arms are occluded. As a precise reference for tracking the data, we use optical-based NatualPoint OptiTrack rigid-body markers and an object-tracking toolkit that can track 6DOF (six degrees of freedom) objects related to the motion of the whole human body. Our system is built on a PC with a 3.07-GHz Intel Core i7 CPU, 12 gigabytes of main memory, an NVIDIA GeForce GTX 260 graphics chip, an MS Kinect depth camera, and our PUI. As a software development platform, we use the Unity 3D tool to render the virtual content, MS Kinect SDK to process human motion, and C# language to integrate the skeleton tracking with a depth camera and the movement of the

(a) (b)

Fig. 6. (a) Virtual content of baseball pitcher applying hybrid sensing and physics simulation under virtual environments and (b) interactive real-time content of virtual baseball batter.

Table 1. Evaluation using average errors in tracking accuracy.

Method	Motion	Value (mm)
Only depth camera	Pitching	235
	Hitting	446
Hybrid sensing by our approach	Pitching	32
	Hitting	83

PUI on a Microsoft Windows 7 operating system. Figure 6 shows our virtual baseball content that is operated by real-time tracking and interaction. The skeleton of the human body, which is composed of 20 joints relative to the depth sensor origin to discover the skeleton of joints of a human standing in front of the depth sensor, is continuously tracked using hybrid sensing. Table 1 reflects our quantitative evaluation, completed using the margin of error (distance) in our tracking accuracy.

IV. Conclusion

Although real-time motion tracking for interaction under virtual environments is an essential element, current problems remain, including an uncomfortable wired interface and poor sensing areas occluded behind the body. In this letter, we presented a method for whole-body tracking of human motion including occluded spaces using hybrid sensing, mixing both depth information and values from a physical user interface (PUI). We also proposed a numerical model converting from angular velocities of the PUI to positional data from a Cartesian coordinate system and an algorithm for the tracking of human joints by combining two types of sensors. Additionally, we constructed virtual contents, which we evaluated to validate the usefulness of the proposed system. We hope that our proposed system will contribute to the utilization of interaction technology with real-time motion tracking and natural gestures. Unlike other state-of-the-art systems, our proposed system can be executed not only while tracking the occlusion space but also with a combined use of a PUI and depth sensor, which can provide a more effective way to deal with interaction that depends on virtual content. Moreover, motion estimation using a gyroscope is widely applicable to camera tracking for the registration of augmented reality.

As future research, we will work toward the development of statistical models for human motion, mapping certain postures to support more accurate tracking and a rule of motion presets. We will also expand the technology, reducing the number of accumulated errors of double integration and implement virtual content related to the occlusion areas of human tracking under immersive environments, such as head-mounted and multiprojection displays.

References

[1] G.A. Lee et al., "Virtual Reality Content-Based Training for Spray Painting Tasks in the Shipbuilding Industry," *ETRI J.*, vol. 32, no. 5, Oct. 2010, pp. 695-703.

[2] D.S. Jo, U.Y. Yang, and W.H. Son, "Design Evaluation System with Visualization and Interaction of Mobile Devices Based on Virtual Reality Prototypes," *ETRI J.*, vol. 30, no. 63, Dec. 2008, pp. 757-764.

[3] V. Ganapathi et al., "Real Time Motion Capture Using a Single Time-of-Flight Camera," *Proc. ICCV*, 2010, pp. 755-762.

[4] C. Wren et al., "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, July 1997, pp. 780-785.

[5] Y. Liu et al., "Markerless Motion Capture of Interacting Characters Using Multi-view Image Segmentation," *Proc. CVPR*, 2011, pp. 1249-1256.

[6] C. Plagemann et al., "Real-Time Identification and Localization of Body Parts from Depth Images," *Proc. ICRA*, 2010, pp. 3108-3113.

[7] J. Shotton et al., "Real-Time Human Pose Recognition in Parts from Single Depth Images," *Proc. CVPR*, 2011, pp. 1297-1304.

[8] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data," *Proc. ISPRS*, 2011, pp. 29-31.

[9] S.H. Kwak et al., "Learning Occlusion with Likelihoods for Visual Tracking," *Proc. ICCV*, 2011, pp. 1551-1558.

[10] D.H. Kim and D.J. Kim, "Self-Occlusion Handling for Human Body Motion Tracking from 3D ToF Image Sequence," *Proc. ACMMM 3DVP*, 2010, pp. 57-62.

[11] Y.-L. Chou, *Statistical Analysis: With Business and Economic Applications*, 2nd ed., New York: Holt, Rinehart & Winston of Canada Ltd, 1975, Section 17.9.

[12] S.R. Buss and J.S. Kim, "Selectively Damped Least Squares for Inverse Kinematics," *J. Graphics Tool*, vol. 10, no. 3, 2005, pp. 37-49.