Multimodal Gumdo Game: The Whole Body Interaction with an Intelligent Cyber Fencer

Jungwon Yoon[1], Sehwan Kim[2], Jeha Ryu[1], and Woontack Woo[2]

[1]Dept. of Mechatronics, K-JIST
[2]Dept. of Information & Communications, K-JIST
Kwangju, 500-712, Korea
{garden, skim, ryu, wwoo}@kjist.ac.kr

Abstract. This paper presents an immersive multimodal Gumdo simulation game that allows a user to experience the whole body interaction with an intelligent cyber fencer. The proposed system consists of three modules: (i) a nondistracting multimodal interface with 3D vision and speech (ii) an intelligent cyber fencer and (iii) an immersive feedback by a big screen and sound. Firstly, the multimodal interface allows a user to move around and to shout without distracting the user. Secondly, an intelligent cyber fencer provides the user with intelligent interactions by perception and reaction modules that are created by the analysis of real Gumdo game. Finally, an immersive audio-visual feedback helps a user experience an immersive interaction. The proposed interactive system with an intelligent fencer is designed to satisfy comfortable interface, perceptual intelligence, and natural interaction (*I-cubed*) and enhance the life-like impression of fighting actions. The suggested system can be applied to various applications such as education, art, and exercise.

1 Introduction

Nowadays, Virtual Reality (VR) technology is flourishing with the rapid development of high power computer and related technologies. There are wide-range of VR applications such as training, education, entertainment, engineering, medical operation, teleoperation, etc. Especially, edutainment applications are very popular and marketable immediately. However, the lack of natural interface is a main bottleneck of bringing them into widespread use. Therefore, it is necessary to make a natural multimodal interface for the VR applications. In order to enhance the effects of immersive experiences in VR-based edutainment systems, the systems should be *I-cubed*, i.e. the systems have comfortable *interface*, perceptual (or emotional) *intelligence* and natural *interaction* [1].

A number of researchers have reported on interactive systems with autonomous agents. Especially, for the whole body interaction of a user with a virtual environment, the players [2-3] are merged into a virtual environment by head-mounted displays, magnetic sensors and data gloves. Even though the whole body interaction with a

Y.-C. Chen, L.-W. Chang, and C.-T. Hsu (Eds.): PCM 2002, LNCS 2532, pp. 1088-1095, 2002.

virtual environment in these systems is possible, these systems still have limitations in providing immersive interactions because the complicated facilities have to be worn or attached on the body and then connected to computers with wires, which tends to distract users from experiencing immersion. The ALIVE system [4] and the KidsRoom [5] used a 2D vision interface in order to extract the user's actions. Note, however, that the 2D vision-based systems in both ALIVE and KidsRoom have limitations in exploiting 3D visual information. Gavrila et al. tried to identify the whole body posture by analyzing multiple-view frames [6]. However, it has limitations in applying it to real-time interactive systems because the posture is analyzed in an interactive post-processing phase. Accordingly, there are only a few *I-cubed* systems providing excitement to a user with the whole body interaction via a comfortable interface and an autonomous agent.

In this paper, we present an immersive multimodal Gumdo simulation game that allows a user to experience the whole body interaction with an intelligent cyber fencer. The proposed system consists of three modules: (i) a comfortable multimodal interface with 3D vision and speech (ii) an intelligent cyber fencer and (iii) an immersive feedback by screen and sound. After taking everything into consideration, the proposed system provides the user with an immersive Gumdo experience with the whole body movement. This paper is organized as follows: In chapter 2, we describe in detail the proposed Gumdo game system in terms of three components; a multimodal interface, fencer intelligence and audiovisual feedbacks. Experimental results and discussions are followed in chapters 3 and 4, respectively.

2 Description of the Proposed Gumdo Simulation System

Gumdo is one of the fencing sports with a bamboo sword and light protective armor. Fencers wear protective equipments covering target areas; head, wrists and abdomen. To make a valid cut, a fencer must strike on the target areas of the opponent with a bamboo sword, while shouting the name of the target areas. The one who strikes the target areas twice among three rounds becomes a winner. In order to properly simulate an immersive Gumdo simulation in a VR setting, first of all a wide range of motion should be covered by a vision interface for detecting a real fencer's full body motion. In addition, a reliable speech interface is needed to express and understand the intension of a fencer. Next, for exciting interaction with a cyber fencer, some realistic intelligence should be provided to the cyber fencer. Finally, a user should experience full immersion with visual, auditory and haptic feedback that can display fighting situation in real-time. Figure 1 shows the block diagram of the proposed Gumdo game system that incorporates all requirements except haptic feedback. Haptic feedback is not considered because it requires a heavy robotic system that may restrict wider full body motion.

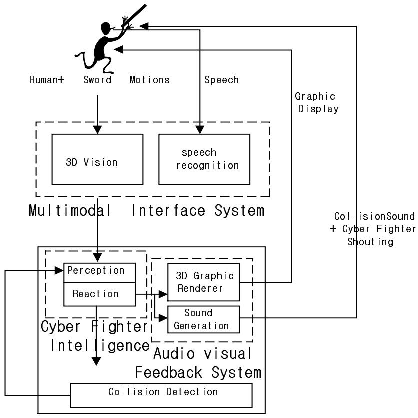


Fig. 1. Block diagram of Gumdo system

2.1 Multimodal Interface (3D Vision & Speech Recognition)

We adopt a non-contact vision-based 3D interface, exploiting depth information without distracting the user while tracking the user in 3D space. To track the movement of the user and the sword in 3D space, we first segment moving foreground from static background after estimating depth information [7]. Next, we separate the sword from the segmented foreground by exploiting two colored markers located at the end points of the sword. Finally, we estimate the line of the sword and the center of the user to track the movements of both of them. Using the moments of the foreground object, the orientation angle of the body about the z-axis is calculated.

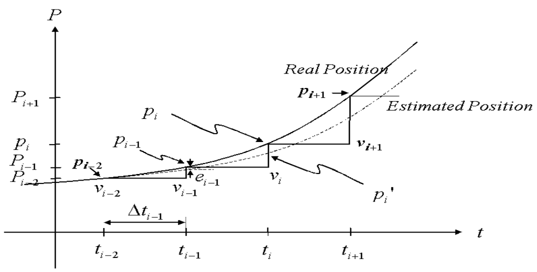


Fig. 2. Compensation of time delay

In general, the real-time tracking of a moving sword in 3D is not easy because some processing time is required to segment and then to estimate its 3D posture [8]. The time-delay due to the processing time has a serious influence on obtaining an accurate 3D position of the sword especially when the movement of the sword is faster than the frame rate of the multiview camera. Therefore, a scheme for compensating for the time delay is proposed. In Fig. 2, the real and estimated positions of the sword are compared. The real 3D position trajectory is represented as a solid line while the estimated one as a dotted line. Based on the past positions prior to t_i, the estimated position of the sword is obtained by:

$$\boldsymbol{p_i}' = \boldsymbol{p}_{i-1} + \alpha \boldsymbol{v}_{i-1} \times \Delta \boldsymbol{t}_{i-1} + \boldsymbol{f}(\boldsymbol{e}_{i-1}),\ 0 < \alpha < 1\ . \quad (1)$$

where p_i' denotes the estimated position of the sword at time t_i, p_{i-1} is the real position of the sword at t_{i-1}, v_{i-1} is the velocity of the sword from t_{i-2} to t_{i-1}, and Δt_{i-1} the time duration from t_{i-2} to t_{i-1}. The constant α represents a scale factor for determining the ratio of time duration from t_{i-2} to t_{i-1} to estimate the position of the sword at t_i ($0<\alpha<1$). The last term $f(e_{i-1})$ is used to compensate for the estimation errors. This function is expressed as a nonlinear function of the error between real and estimated positions. The nonlinear compensator function is usually necessary because the direction of the sword changes abruptly in Gumdo simulation.

The system recognizes three Korean words: “Meo-Ri (HEAD)”, “Heo-Ri (ABDOMEN)” and “Son-Mok (WRIST)”. The speech inputs in the proposed Gumdo simulation are used for the perception module of the intelligent cyber fencer. The speech inputs are classified into three words based on extracted speech features, LPC (Linear Prediction Coding) cepstrum coefficients [9]. Finally, the positions of both the sword and the user will be transferred to the cyber fencer kernel, combined with the results of speech recognition.

2.2 Cyber Fencer Intelligence

A cyber fencer must have some intelligence for realistic and exciting game with a real-person fencer. The intelligent cyber fencer has perception and reaction modules. The perception module perceives its environment information by internal sensors and external inputs while the reaction module consists of motivation, behavior and motor modules. Fig. 3 shows the schematic diagram of the cyber fencer action. Arrows represent information flows between components.

The perception module should manage intelligently audio-visual inputs from the multimodal interface. The basic actions of the user and the sword can be perceived by observing the motion of the user and the sword in terms of distance and direction as well as by hearing the shout sound. The behavior of the user can be recognized by the body action followed by the sword action. Also, the situation of virtual environment during fighting will be detected by internal sensors such as the percentage of victories and the current position of the cyber fencer inside a fighting field.

The motivation module consists of state and drive. The state refers to the feeling of the cyber fencer and is associated with the action of the cyber fencer. The drive represents the desire of the cyber fencer while fighting with the user. The drive consists of three important elements of Gumdo; *mind, sprit and power*. The *mind* reflects the static condition of the cyber fencer such as the calmness and discernment and shows the ability to recognize the information transferred from the perception module. The *spirit* represents the will of the fencer, i.e. the dynamic conditions of the cyber fencer.

The *power* represents the ability to attack, i.e. the promptitude and tempo of the attack and defense of the fencer. The state is composed of fear that shows the current feeling of the fencer. The fear level of the cyber fencer increases when he meets the human fencer with physical superiority or he encounters unexpected actions from the human fencer.

Given the perception and motivation inputs, the behavior module sends relevant actions to the motor module. The behavior module has two kinds of actions; general and reflective actions. The general action is a set of behaviors to achieve goals. The reflective action is a set of behaviors to protect the attack of the user.

The motor module enables the cyber fencer to match with the user as well as to display the cyber fencer on the screen. The motor module, controlling both the body and the sword motions of the cyber fencer, does real-time motion interpolation to execute Gumdo simulation.

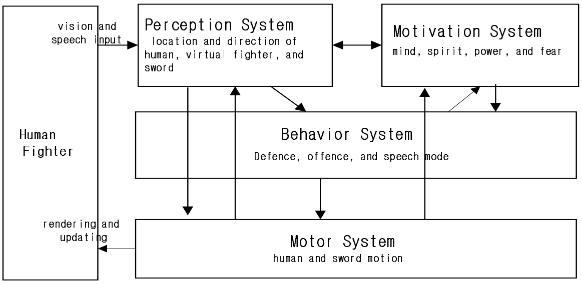


Fig. 3. Schematic Diagram of a Cyber Fencer Action. (adapted from [10])

2.3 Audio-Visual Feedback

For giving immersive feeding to the user, audio-visual feedback is provided with a big screen and sound effects. The geometric model of the cyber fencer consists of a spherical head, a cylindrical upper body, and cylindrical upper and lower arms. The motions of the fencer and the sword models are controlled by a vision interface. For sound effects, we record and play sound from collision impact between two swords, between the sword and one of the targets, fencer's shouting voices, and referee voices that declare start and end of the fighting and that indicate who is the winner after striking. To properly display the collision situation, we detect the collision between two swords or between a sword and one of the target areas: head, wrist and abdomen by considering as collisions between a line and a line or a line and a sphere. In our Gumdo simulation, since the real sword collides with the virtual sword, the sword collision cannot physically stop the motion of the real sword due to the absence of the force feedback. Therefore, the striking passing through the target areas is considered as an invalid striking.

3 Experiment Results

Fig. 4 shows the proposed Gumdo simulation system consisting of multimodal interface, artificial intelligence and audio-visual feedback. The proposed system was implemented in a Workstation with Pentium III Xeon Dual 1GHz CPUs.

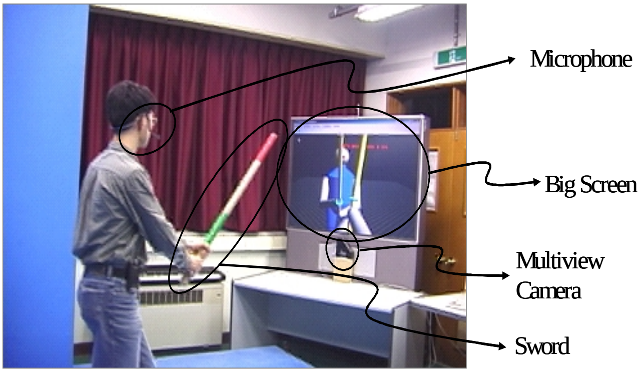


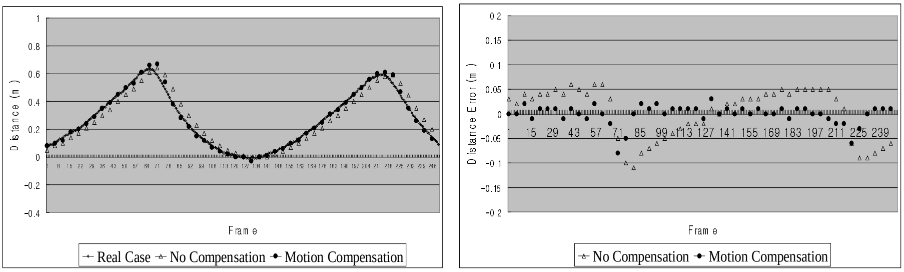
Fig. 4. An immersive Gumdo game with audio-visual feedback

Table 1 shows system specifications of the Gumdo simulation game. The movement range is 2m x 4m and the resolution of the vision system is 5cm/disparity in the range of 3-5m. The speed of the Gumdo simulation game is about 5 Hz due to the time consumption in calculating 3D disparity from multiview images. The system recognize 3 words, “Meo-Ri, Heo-Ri , Son-Mok”, except the noise.

Table 1. Specifications of the Gumdo system

Parameters	Measurements
Fighting Area	2m (width) * 4m (depth)
Sword Resolution	5 deg (pitch), 5 deg (yaw)
Body Resolution	0.05 m
Bandwidth of System	5 Hz
Speech Recognition	3 words Recognition

In the experiments, we captured a moving sword, which is about 220 cm away from a multiview camera. Through the experiments, we observed that the tracking performance of the user body is not sufficient to enjoy a real fighting with the cyber fencer in the virtual environment since the tracking speed (5Hz) of the sword is a little slow. To overcome the hardware limitations, we adopted a simple but effective motion compensation technique. In Fig. 5, we show the motion compensation results considering the processing time-delay. The errors are reduced substantially as shown in Fig. 5 (b).



(a) (b)

Fig. 5. Motion compensation results with the sword moving in x direction, (a) compensation results in x direction, (b) position errors between real case and compensated or uncompensated cases

In the Gumdo simulation, it was enough just to consider the first phoneme of each word, that is, “Meo”, “Heo” and “Son” to differentiate three words. Thus, we just used the first phoneme of each word to get cepstrum coefficients for speech samples. For speech recognition, we checked the recognition rates for each of the three words. The speech signal and each Euclidean distance are shown as an example in Table 2. The “Son-Mok” has over 90% success recognition rate because its phonemic structure is so different from the others. Even though there are very different intensities of the “Son-Mok” speech input, the recognition algorithm worked well. However, because “Meo-Ri” and “Heo-Ri” had similar sounds, we had about 80% success recognition rate. When users were trained by experiments, the average success recognition rate was over 90%.

Table 2. The Euclidean distance from the cepstrum coefficient

Euclidean distance / Pattern Speech	“Meo-Ri”	“Heo-Ri”	“Son-Mok”
“Meo-Ri”	1.9633	2.3052	8.4024
“Heo-Ri”	2.3798	1.6915	6.8680
“Son-Mok”	8.1364	7.8384	2.4082

4 Conclusions and Future Work

In this paper, we proposed a simple immersive Gumdo simulation game that consists of a multimodal interface (3D vision and sound), an intelligent cyber fencer based on perception and reaction modules, and an audio-visual feedback during the fighting between a user and a cyber fencer. The proposed system allows the whole body interaction with the cyber fencer without distracting the user. The proposed vision system is able to have a wider motion range (2m (width) * 4m (depth)). In addition, we added the speech modality using a wireless microphone to the user to enjoy a natural Gumdo simulation and to interact with the cyber fencer. The proposed interactive system with an intelligent fencer can improve interactions between the user and the cyber fencer in a virtual environment and can enhance the life-like impression of fighting actions. The 5Hz speed of Gumdo game may be slow for applications requiring fast motion. This slow speed may be increased by; higher sampling rate (currently 5Hz) by superior PC, higher rate of communication, use of compensation scheme using extra interpolation, and so on. A remaining challenge is to develop the Gumdo simulation exploiting a force feedback. The force feedback combined with the vision and speech will provide a Gumdo fighting with a full immersion. A photo-realistic avatar in the 3D virtual environment will also provide more realistic experience. In addition, a networked VR game with a real person will be another challenging task.

Acknowledgement. This research was supported by the BK21 research fund from the MOE, Korea.

References

1. W. Woo, N. Kim, K. Wong and M. Tadenuma, " Sketch on Dynamic Gesture Tracking and Analysis Exploiting Vision-based 3D Interface," *in Proc. SPIE PW-EI-VCIP'01*, vol. 4310, pp. 656-666, Jan. 2001.
2. L. Emering, R. Boulic, D. Thalmann, Interacting with Virtual Humans through Body Actions,IEEE Computer Graphics and Applications,1998 , Vol.18, No1, pp8-11.
3. T. Molet, A. Aubel, T. Çapin, S. Carion, E. Lee, N. M. Thalmann, H. Noser, I. Pandzic, G. Sannier, D. Thalmann, "ANYONE FOR TENNIS? ", Presence, Vol. 8, No. 2, pp.140-156, April 1999,
4. Maes P., T. Darrell, B. Blumberg, A. Pentland, 1995. *The ALIVE System: Full-body Interaction with Autonomous Agents.* In Proc. Computer Animation'95, pp 11-18, 1995.
5. F. Bobick, S. S. Intille, J. W. Davis, F. Baird, C. S. Pinhanez, L. W. Campell, Y. A. Ivanov, A. Schutte, A. Wilson, "The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment ", Presence, Vol. 8, no. 4, pp.369-393, Aug. 1999,
6. Gavrila, L.S. Davis, 1996. *3D Model-Based Tracking of Humans in Action: A Multi-View Approac*h. Proc. of IEEE Conf. on CVPR, pp 73-80, June 1996
7. W. Woo and Y. Iwadate, "Object-oriented hybrid segmentation using stereo images," *in Proc. SPIE VCIP*, pp. 487-495, Jan. 2000.
8. S. Kim and W. Woo, "3D Movement Tracking with Asynchronous Multi-cameras for Interactive Systems", *in Proc. SPIE PW-EI-VCIP'02*, vol. 4671, pp. 502-512, Jan. 20-25, 2002.
9. Shuzo Saito, " Fundamentals of Speech Signal Processing", Academic Press, 1985
10. S-Y Yoon, R. C. Burke, B. M. Blumberg, G. E. Schneider, "Interactive Training for Synthetic Characters", *AAAI 2000*.