멀티모달 인터페이스(3차원 시각과 음성)를 이용한 지능적 가상검객과의 전신 검도게임

[A Full Body Gumdo Game with an Intelligent Cyber Fencer using Multi-modal(3D Vision and Speech) Interface]

윤 정 원 † 김 세 환 †† 류 제 하 ††† 우 운 택 ††††
(Jungwon Yoon) (Sehwan Kim) (Jeha Ryu) (Woontack Woo)

요 약 본 논문에서는 멀티모달(multi-modal) 인터페이스를 통해 지능적 가상검객과 체감형 검도게임을 할 수 있는 시스템을 제안한다. 제안된 검도게임 시스템은 멀티모달 인터페이스(시각과 청각), 인공지능(AI), 피드백(스크린과 사운드) 등 크게 세 가지 모듈로 구성된다. 첫 번째로, 멀티모달 인터페이스는 시각기반 3차원 인터페이스를 이용하여 사용자가 자유롭게 3차원 공간에서 움직일 수 있도록 하고, 음성기반 인터페이스를 이용하여 사용자가 현실감 있는 검도게임을 즐길 수 있도록 한다. 두 번째, 인공지능은 가상검객에게 멀티모달 인터페이스에서 입력되는 시각과 음성를 인식하여 가상검객의 반응을 유도한다. 마지막으로, 대형 스크린과 스피커를 통한 시청각 피드백은 체감형 상호작용을 통하여 사용자가 몰입감을 느끼며 검도게임을 경험할 수 있도록 한다. 따라서 제안된 시스템은 전신의 움직임으로 사용자에게 몰입감의 검도게임을 제공한다. 제안된 시스템은 오락 외에 교육, 운동, 예술행위 등 다양한 분야에 적용될 수 있다.

키워드 : 멀티모달 인터페이스, 3차원 시각기반 인터페이스, 음성인식, 검도게임, 지능의 가상검객, 전신의 상호작용, 몰입감의 교육-오락 시스템

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. First, the multimodal interface with 3D vision and speech allows a user to move around and to shout without distracting the user. Second, 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 by a big screen and sound effects helps a user experience an immersive interaction. The proposed system thus provides the user with an immersive Gumdo experience with the whole body movement. The suggested system can be applied to various applications such as education, exercise, art performance, etc.

Key words : multimodal interface, 3D vision-based interface, speech recognition, Gumdo game, Intelligent Cyber Fencer, the whole body interaction, immersive edutainment system

· 본 연구는 광주과학기술원의 기관고유사업에 의하여 지원되었음
† 비 회 원 : 광주과학기술원 기전공학과
garden@kjist.ac.kr
†† 비 회 원 : 광주과학기술원 정보통신공학과
skim@kjist.ac.kr
††† 비 회 원 : 광주과학기술원 기전공학과 교수
ryu@kjist.ac.kr
†††† 종신회원 : 광주과학기술원 정보통신공학과 교수
wwoo@kjist.ac.kr
논문접수 : 2002년 11월 28일
심사완료 : 2003년 5월 13일

1. 서 론

가상현실(Virual Reality:VR)은 컴퓨터에 의해 사람이 인공적으로 창조된 세계에 실재하고 있다고 느끼도록 하는 환영을 만드는 기술로 컴퓨터, 이와 관련된 기술산업의 빠른 성장으로 대중에게 널리 알려지고 있다. 따라서 인간의 감각 입력을 사용한 인터페이스를 사용함으로써 사용자와 가상환경 사이에 효율적일 뿐만 아니라 자연스러운 상호작용을 증진시킬 수 있다. 가상현

실의 응용 분야는 오락, 훈련, 교육, 공학, 수술, 원격조종 등으로 그 범위가 점차 확대되고 있으며 그 가운데 오락은 가장 인기 있고 넓은 시장을 확보하고 있다. 그러나 현재까지 대부분의 시스템은 제한된 인터페이스의 데스크 탑용 PC에 한정되어 있어 사용자가 가상환경에서 몰입감의 경험을 증진시키기 위해서는 자연스런 멀티모달 인터페이스를 사용한 오락용 시스템을 제작하는 것이 필요하다. 이상적인 가상현실 엔터테이먼트 시스템에서 몰입감을 가지고 오락을 즐기기 위해 사용자를 가상환경과 연결하는 인터페이스 시스템은 편안한 인터페이스(Interface), 지각 혹은 감정을 표현하는 지능(Intelligence), 자연스런 상호작용(Interaction)의 3가지 구성요소를 포함하고 있어야 한다.

지금까지 에이전트와 사용자가 상호작용하는 엔터테이먼트 시스템에 대한 많은 연구들이 있었다. Bates 등의 Woggle world는 전통적인 애니메이션과 행동 모델을 결합함으로써 사용자가 가상공간 상에서 에이전트와 상호작용하는 시스템을 구성하였다[1]. 그러나 에이전트가 비교적 복잡한 상호작용을 제공하더라도 마우스가 유일한 인터페이스 방법으로 사용자와 편안하고 자연스런 상호작용을 제공하는데 한계를 가진다. Fisher 등이 제안한 시스템은 사용자가 애니메이트된 에이전트와 실시간으로 안경(goggle)을 착용하고 상호작용 하도록 하였다[2]. Yoon 등은 Sydney K9.0으로 명명된 "clicker training"을 사용하여 훈련된 캐릭터를 제작하여 Dog-Ear라는 실세계의 음향 데이타를 가상 생명체(creature)의 커널 인식시스템에 합치는 모듈을 포함시켰다[3]. 이러한 시스템들은 제한된 인터페이스의 사용으로 사용자 신체의 일 부분만이 가상환경과의 상호작용에 참여할 수 있도록 한다. 사용자와 가상환경간의 전신(whole body)의 상호작용을 위해서 Emering은 파이트 훈련게임에서 행동인식 알고리즘(Multi-level Action Recognition Algorithm)을 제시하고 이를 평가하였다[4,5]. 위 시스템에서 사용자와 컴퓨터의 인터페이스를 위한 입력 센서로 자기(magnetic) 모션 캡쳐 시스템이 사용되었다. Molet가 제안한 가상 테니스 게임에서는 사용자가 HMD(Head Mounted Display)와 자기 센서 및 데이타 글러브(data glove) 등을 이용하여 가상환경과 상호작용하였다[6]. 비록 이러한 시스템들에서는 사용자가 가상환경과 전신의 상호작용이 가능하지만 글러브, 안경, 헬멧 등을 인터페이스로 사용함으로써 사용자의 행동을 부자연스럽게 하여 몰입감의 상호작용을 제공하는데 한계를 지닌다. 반면, ALIVE 시스템에서는 가상 생명체가 사용자의 실질적인 위치, 몸의 포즈, 손 제스처(hand gestures) 등의 정보와 인터페이스 할 수 있도록 시각 시스템을 사용하였다[7]. 그러나 ALIVE 시스템에서는 가상 아바타의 행동이 사용자의 행동에 크게 영향을 미치지 못하기 때문에 제한된 상호작용만이 가능하며 또한 시각 입력만 이용하여 시각, 청각의 멀티모달 인터페이스를 이용하는 것과 비교해 사용자에게 자연스런 상호작용을 제공하기 어렵다. KidsRoom은 센서 출력과 개인과 집단 행동들을 인식하기 위해서 이야기(the story)에서 유도된 의미(contextual) 정보를 결합하였다[8]. 위 시스템에서 시각 인터페이스는 사람을 추적하고, 움직임을 감지하고, 사람들의 행동들을 인식한다. 하지만 위 시스템에 사용된 시각 시스템은 2차원 기반으로 3차원 물체의 충돌을 감지하는 데에는 한계를 지닌다. Gavrila 등은 다중카메라 뷰 프레임을 분석하여 전신의 posture를 확인하려고 시도했지만 위 목적은 실시간이 아니라 post-processing phase로 실행되었다[9]. 현재까지 보다 편리한 3차원 인터페이스로 사용자를 제한함이 없이 지능을 가진 아바타와 전신의 상호작용으로 사용자에게 즐거움과 흥미를 제공하는 시스템은 극히 드물다.

따라서 본 논문에서는 몰입감의 오락 시스템에서 필요로 하는 특성들을 만족하도록 오락용 VR 검도 게임에 대한 새로운 프레임웍을 제안한다. 사용자가 충분한 몰입감을 느끼며 오락을 즐기도록 VR 시스템은 사용자가 지능적 에이전트와 전신의 상호작용이 가능하도록 지원되어야 하며, 따라서 이를 위하여 제안된 시스템은 그림 1과 같이 다음의 3가지 모듈로 구성된다; (a) 사용하기 편한 멀티모달(3차원 시각과 청각) 인터페이스, (b) 인공지능(Artificial Intelligence:AI), (c) 시각 및 청각 피드백에 의한 몰입감. 첫 번째, 멀티모달 인터페이스는 사용자가 전선으로 연결되지 않고 자유롭게 움직일 수 있도록 무선 기반의 3차원 시각과 음성 인터페이스를 사용하였다. 즉, 멀티뷰 카메라를 이용하여 사람의 신체와 검의 3차원의 정보를 계산하고 음성 인터페이스는 무선 마이크를 통해 획득한 음성 신호를 이용해 사용자의 의도를 해석한다. 두 번째, 인공지능은 멀티모달 인터페이스로부터의 정보를 분석하여 가상검객이 지능을 가지고 사용자와의 상호작용이 가능하도록 한다. 마지막으로, 사용자가 가상검객과 대결하는 동안 현실감을 제공하기 위해 가상검객의 반응이 대형 스크린을 통해 시각적으로 사용자에게 피드백되고 충돌 상황 및 가상검객의 구호가 스피커를 통해 사용자에게 청각적으로 피드백됨으로써 사용자가 몰입감을 가지고 시합을 경험할 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 멀티모달 인터페이스, 가상검객의 지능 및 시청각 피드백으로 구성된 제안된 검도게임 시스템을 설명하고, 3장에서는 시스템의 구현 및 실험을 통한 성능평가를 수행하고, 결론 및 추후 과제를 4장에서 제시하였다.

2. 검도 게임 시스템

검도는 죽도(대나무 검)와 가벼운 보호장구를 착용하고 경기하는 펜싱 스포츠의 일종이다. 검객은 목표 위치인 머리, 허리, 손목에 방어 장비를 착용하고 점수를 얻기 위해 죽도로 상대방의 목표 위치를 타격해야 한다. 목표 위치의 어느 한 부분을 타격하면 점수를 얻게 되며, 이 때 타격 위치를 검객이 동시에 외쳐야 하며 타격 위치를 두 번 먼저 타격하는 검객이 승리한다.

실제의 검도시합을 가상현실 시스템으로 적절하게 표현하고 가상검객과 상호작용하기 위해 넓은 움직임 범위가 필요하며 사람검객의 의도를 파악하고 대결하는 동안 몰입감을 증대시키기 위해 음성을 포함한 멀티모달 인터페이스가 제공되어야 한다. 또한 사람검객은 지능을 가진 가상검객과의 대결을 통해서 흥미를 가지고 시합을 즐길 수 있어야 한다. 마지막으로 시각, 청각 및 촉각의 피드백이 실시간으로 시합의 상황에 따라 사람검객에게 주어져야 한다. 촉각 피드백을 제외하고 이상의 조건들을 만족하는 제안된 검도게임 시스템이 그림 1에 나타나 있다. 촉각 피드백(Haptic feedback)은 넓은 작업공간에 제한을 줄 수 있는 무거운 로봇 시스템을 필요로 하기 때문에 본 시스템에서는 고려되지 않았다.

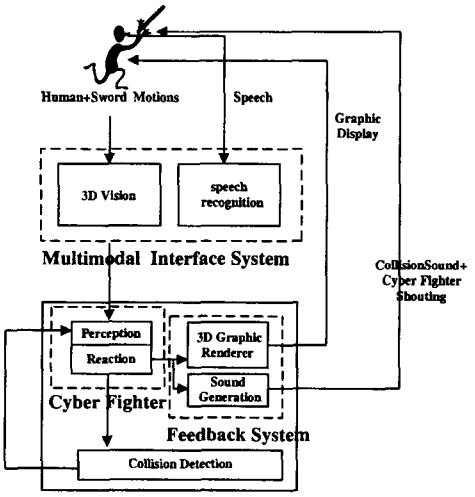


그림 1 검도게임 시스템 블럭 다이어그램

2.1 멀티모달 인터페이스(3차원 시각 및 음성인식)

3차원 공간상의 정보를 얻기 위해 넓은 작업공간의 사용 및 사용자의 움직임을 방해하지 않는 시각 인터페이스를 선택하였다[10]. 3차원 공간상에서 검과 사람의 위치에 대한 정보를 얻기 위해 입력영상에서 검을 든 사용자를 배경에서 분리하고 분리된 전경영상(foreground image)에서 검의 끝에 위치한 두 개의 색깔마커(color marker)를 분할해 그 중심점과 검의 각을 계산한다. 다음으로 주어진 깊이 정보를 통해 분할된 사람의 중심점 및 사람축에 대한 회전을 계산한다.

전경을 구하기 위해 색깔, 에지, 모션, 양안차(disparity) 정보를 복합적으로 이용한 강건한 분할기법을 사용한다. 먼저, 현재 장면의 smooth 양안차를 계산하고 색깔, 에지, 모션, 양안차를 비교하여 정적 배경 장면에서 움직이는 대상물을 분리한다. 관심 있는 대상물만 분리하기 위해 대상물이 제한된 양안차 범위, 즉 깊이를 가지고 있고 위 대상물의 양안차가 부드럽게 변한다는 가정 하에 깊이 정보를 포함하고 있는 분리된 대상물을 추적한다.

검의 3차원 위치를 추적하기 위해서는 먼저 배경으로부터 분할된 전경영상으로부터 검의 끝에 각각 빨간색과 녹색의 색깔마커를 부착해 마커들을 분할하였다. 만일 매칭 점을 검의 양쪽 끝에 위치한 색깔마커의 중심점으로 잡는다면 그림 2에 보인 것과 같이 색깔분할에 의해 구해진 두개의 3차원 중심점에 기인해 검의 위치를 예측할 수 있다. 검의 방위각(azimuth angle) α와 고도각(elevation angle) β는 다음과 같이 구한다.

$$L_1 = \sqrt{(x_r - x_g)^2 + (y_r - y_g)^2 + (z_r - z_s)^2} \quad (1)$$

$$\alpha = \tan^{-1}((x_r - x_g)/(z_r - z_g)) \quad (2)$$

$$\beta = \sin^{-1}((y_r - y_g)/L_1) \quad (3)$$

위 식에서, (xr, yr, zr)은 빨간색에 대한 3차원 위치,

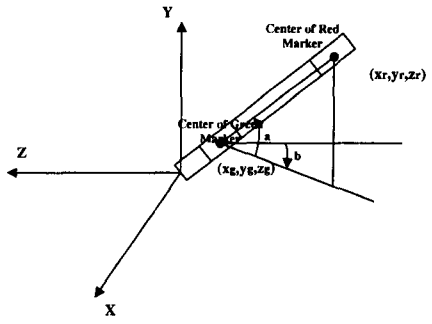


그림 2 검에 부착된 색깔마커

그리고(x_g, y_g, z_g)는 녹색에 대한 3차원 위치를 나타내고, L_1은 두 마커간의 길이를 나타낸다.

분할된 전경영상에서 색깔마커들을 분할한 후 사람의 위치정보는 전경영상의 중심점을 이용하여 예측 가능하고, 또한 전경영상의 영상 모멘트를 사용하여 사람검객의 축에 대한 회전각을 구할 수 있다[11]. 영상 모멘트는 전체 영상 정보에 유용한 요약정보를 제공하며 모멘트는 전체 화소에 대한 합을 포함하여 적은 화소변화에 대해 강건하다는 장점이 있다. 그림 3에서 전경 영상의 중심점에서 y축 단면적을 포함하는 특성함수는 다음과 같이 정의된다.

$$b(z,x)=\begin{bmatrix}1 & \text{for points for objects}\\ 0 & \text{for background points}\end{bmatrix} \quad (4)$$

그리고 이차까지의 모멘트 $M_{(i,j)}(i,j=0,1,2)$는 다음 식 (5)와 같이 구해진다.

$$\begin{aligned} M_{00}&=\sum_z\sum_x b(z,x), \quad M_{11}=\sum_z\sum_x zxb(z,x)\\ M_{10}&=\sum_z\sum_x zb(z,x), \quad M_{01}=\sum_z\sum_x xb(z,x)\\ M_{20}&=\sum_z\sum_x z^2b(z,x), \quad M_{02}=\sum_z\sum_x x^2b(z,x)\end{aligned} \quad (5)$$

이차 모멘트로 불리는 상수 a, b, c는 식 (6)과 같이 정의되고 여기서 x_c, y_c, z_c는 전경의 중심 좌표 값이다.

$$a=\frac{M_{10}}{M_{00}}-z_c^2, \quad b=2\left(\frac{M_{11}}{M_{00}}-z_cx_c\right), \quad c=\frac{M_{02}}{M_{00}}-x_c^2 \quad (6)$$

마지막으로, 사람검객의 축에 관한 각 θ는 식 (7)에 의해 구해진다.

$$\theta=\frac{\arctan(b,(a-c))}{2} \quad (7)$$

이상의 3차원 시각 인터페이스에서 구해진 검 및 사용자의 위치 정보는 음성인식의 결과와 함께 가상검객의 커널로 전달된다. 그러나 실시간으로 움직이는 객체의 3차원 좌표를 얻는 것이 쉬운 문제는 아니다. 제한된 컴퓨터 성능의 사용으로 움직이는 객체의 3차원 좌표를 얻기 위해 배경으로부터 객체를 영상분할하고 3차원 좌표를 구하는 수식을 적용하는 데는 많은 계산량으로 인해 시간지연이 발생하기 때문이다. 이로 인해 계산 시간 동안 3차원 공간상의 실제 객체는 다른 위치로 이동하지만 계산된 객체의 위치는 현시점(t_i)의 객체의 위치가 아닌 이미 지난 시간(t_{i-1})의 위치를 계산하게 된다. 따라서 빠른 움직임의 죽도 사이의 충돌을 검출하는 데에는 많은 오차를 야기시키므로 이러한 오차를 보상하기 위해 t_{i-1} 시점의 객체의 위치를 정확히 반영하여야 한다.

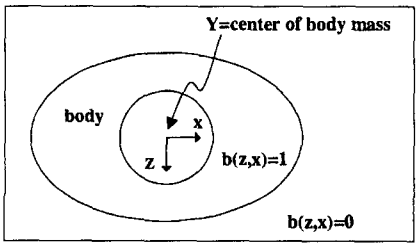


그림 3 사람의 중심점(Y=yc)에서의 Z-X

그림 4는 카메라로부터 획득된 영상을 영상분할(segmentation)하고 3차원 위치 값을 계산하는데 소요되는 시간과 랜더링 시간 등을 고려하여, 이러한 과정들이 완료된 시점에서의 실제 위치와 추정된 위치에 대해 각각 객체의 위치를 비교해 나타낸 것이다. 즉, t_i시점에서 p_i인 3차원 공간에 객체가 위치하는 실제 위치를 실선으로 나타냈으며, 제안한 방법을 사용하여 t_{i-1} 이전의 정보로부터 추정된 위치를 점선으로 나타내었다. 추정된 객체의 위치는 다음과 같은 수식을 통해 얻을 수 있다.

$$p_i'=p_{i-1}+\alpha v_{i-1}\times\Delta t_{i-1}+f(e_{i-1}),\ 0<\alpha<1 \quad (8)$$

위 식에서, p_i'는 t_i시점에서 추정된 객체의 위치, p_i는 t_i시점에서 실제 객체의 위치, v_{i-1}은 t_{i-1}에서 t_i시점까지 객체가 움직이는 속도, 그리고 $\triangle t_{i-1}$는 t_{i-1}에서 t_i시점까지 객체가 움직인 시간을 나타낸다. 그리고 α는 t_i 시점에서의 객체의 위치를 추정하기 위해 t_{i-2}부터 t_{i-1}까지의 시간 간격에 대한 비를 결정하는 계수로 0에서 1의 사이의 값을 갖는다. 마지막으로, $f(e)$는 검의 방향이 바뀌는 순간에 발생하는 오차를 보상해 주는 항으로 실제 값과 추정된 값 사이의 오차에 대한 함수로 표현된다. 즉, 객체의 움직임이 단일 방향이거나 방향의 변화가 심하지 않은 경우에는 그 위치를 적절히 추정하는 것이 가능하지만, 검도의 경우와 같이 검의 방향이 갑자기 바뀌는 경우에는 이를 처리하지 못하는 경우가 발생하기 때문에 $f(e)$를 사용하여 이를 보상하였다.

제안된 시스템의 음성 입력은 지능적 가상검객의 인

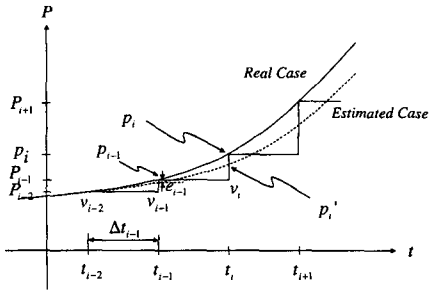


그림 4 시간 지연 보상

식시스템에 사용되며 세 단어, 즉, 머리, 허리, 손목을 분류하는 것을 목적으로 한다. 세 단어를 구분하기 위해 음성 특징 추출(feature extraction)을 이용해 실시간으로 구분하였다. 특징 추출을 위해 음성 파라메터로 LPC (Linear Prediction Coding) 켑스트럼(Cepstrum) 계수[12]를 이용하였고, 인식기로 Euclidean distance를 이용해서 패턴을 비교하였다. 켑스트럼은 음성신호 스펙트럼의 로그 값의 푸리에 변환으로 정의할 수 있고 음성을 전극(all-pole) 필터 모델로 근사화 하면 LPC 계수로부터 선형 전개에 의해 켑스트럼 계수를 얻을 수 있다. LPC 계수로부터 추출된 켑스트럼 계수를 LPC 켑스트럼이라고 하고, LPC 켑스트럼 계수를 구하기 위해 해밍 윈도우(Hamming Window)와 Durbin 알고리즘을 이용해 LPC계수를 구하고 구한 계수를 LPC 켑스트럼 계수로 변환하였다. 따라서, 머리, 허리, 손목의 고립 단어에 대한 패턴의 인식을 위해 각 단어에 대한 켑스트럼 계수를 미리 저장하고, 무선 마이크로 음성을 실시간으로 입력시 위의 세 단어의 켑스트럼과 현재 들어온 단어의 켑스트럼 값의 Euclidean distance가 최소가 되는 단어를 선택하였다. 또한, Euclidean distance가 어느 이상 값을 초과하는 경우에는 원하는 단어가 들어오지 않는 것으로 해석하였다. 무선 마이크로부터의 음성신호는 22,050Hz의 속도로 8bit의 데이타로 들어오며 512 샘플의 윈도우를 가진다. 음성의 시작점을 감지하기 위해 문턱치 값을 두어 2개의 윈도우 프레임이 연속으로 넘으면 음성 입력의 시작위치로 간주하였다.

2.2 가상검객의 인식 및 반응

가상검객은 인식시스템과 반응시스템으로 구성된다. 인식시스템(perception system)은 내외부 입력으로부터 주위 환경을 인식한다. 반응 시스템은 욕구(drive)와 상태(state)로 구성된 동기 시스템(motivation system), 실행할 수 있는 행동 시스템(behavior system), 동작 집합(set of activity)을 선택할 수 있는 모터 시스템(motor system)으로 구성된다.

가상검객의 인식 시스템은 사람검객의 음성인식으로 실시간 음성 입력에 대응할 수 있고 사람위치 및 검의 방향 정보에 의해 사람의 행동을 인식할 수 있다. 사람의 행동을 인식하기 위해, 인식을 위한 사람의 몸과 검의 기초 행동형태는(거리, 방향)으로 이루어져 있다. '거리'는 사람 중심점과 상대검객 중심점과의 거리(혹은 실제 검과 목표지점과의 최단 거리)를 나타내며 '방향'은 이동 방향을 나타내는 벡터로 표현한다. 따라서, 가상검객은 사람검객의 몸과 검의 운동을 각각 인식함으로써 전체 사람 검객의 기술을 인식하게 된다. 예를 들어, '머리 치기'의 경우 사람 몸의 '전방 이동' 행동과 '전방 타격'의 검의 행동으로 나뉘게 되고, 각 몸과 검의 기초 행동을 Boolean 식으로 인식한다. 즉, '앞머리 치기'=(타격이내 거리, 전방 이동)과 (검과 가상검객의 머리와의 거리, 전방타격)으로 인식한다. 또한, 대결 동안에 현재의 승률, 경기장 내에서의 가상검객의 위치와 같은 가상환경 상황은 내부 센서를 통해 획득된다.

동기 시스템은 욕구와 상태로 구성되는데, 상태는 검객의 감정 및 행동과 관련되어 있고 욕구는 사람검객과의 대결 동안 가상검객의 행동의 욕망(desire)을 나타낸다. 검도 시스템에서 욕구 시스템은 검도의 중요한 세 가지 요소들[심(mind), 기(spirit), 력(power)]로 구성되어 있다(그림 5). 심은 침착성과 분별력과 같은 가상검객의 정적인 상태를 나타내며, 인식 시스템에서 전송되는 정보를 습득할 수 있는 능력을 나타낸다. 가상검객의 심이 부족하면 검객은 상대방 검객의 공격에 대한 방어 능력이 줄어들고 두 검객 사이의 거리 및 검객의 상태에 따라 그 값이 좌우된다. 기는 검객의 싸우고자 하는 의지(will), 즉 동적인 검객의 상태를 나타낸다. 만일 검객이 높은 기를 가지고 있다면 공격에 더 치중하게 되며 대결에 보다 적극성을 보일 것이다. 그러나, 사람검객의 함성(shouting)은 가상검객의 기를 감소시키는 역할을 한다. 력은 공격할 수 있는 능력을 나타내며 상대검객에 대한 공격과 방어에 대한 민첩성 및 템포를 나타낸다. 력은 시간에 따른 두 검의 충돌 횟수에 따라 영향을 받게된다. 상태는 두려움(fear)으로 구성되고 가상검객의 두려움의 정도는 우월한 체격의 사람 검객을 만나든지 가상검객의 예상치 못한 행동에 대해 증가한다. 실제 상태는 검객의 행동뿐만 아니라 감정표현에 영향을 주지만 실제 검도 시합 시, 검객의 감정표현이 나타나지 않으므로 현재 감정표현을 위한 행동은 구현되지 않았다. 위에서 설명한 동기 시스템에서 동기의 욕구는 사람검객의 행동과 가상환경의 상황에 따라 그 값들이 변하여 가상검객이 어떤 행동을 취해야 할지를 결정한

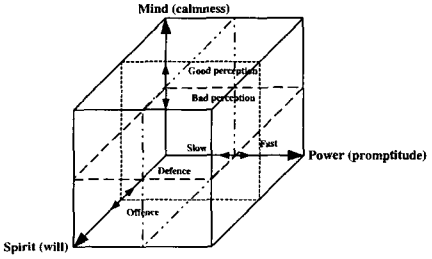


그림 5 욕구(drive) 시스템

다. 예를 들어, 게임의 시작 초기에 가상검객이 체격의 우월성을 보인다면 두려움이 증가한 상태로 게임이 시작된다. 이 두려움은 인식력을 결정하는 가상검객의 심에 영향을 미치게 되며 사람검객의 함성은 가상검객이 방어 모드를 취하도록 한다. 하지만, 사람검객의 동작이 느려지던가, 함성이 줄어들면 가상검객은 증가된 기로 사람검객에 접근하게 되며 증가된 기는 가상검객이 공격 모드를 취하도록 한다. 반면에, 두 검 사이의 충돌이 증가하면 가상검객의 력이 줄어들기 때문에 사람검객의 공격에 취약성을 드려낸다. 따라서, 가상검객의 움직임이 둔해지게 된다. 그러나, 사람검객이 가상검객의 목표 지점을 타격하면 가상검객은 사람검객의 공격에 보다 조심스럽게 반응하게 된다.

행동 시스템은 주어진 인식, 동기 시스템에 기초해 적절한 행동을 실행하고 기초적인 기술과 움직임으로 구성된 모터 시스템에 기초 행동을 지시한다. 행동 시스템은 두 가지 형태로 구성된다. 첫 번째는 동기 시스템의 욕구 및 상태에 대한 일련의 행동이고, 두 번째는 공격의 반사작용에 의한 반응(reflective) 작용이다. 일반 행동은 사람검객을 공격하고자 하는 공격행동들과 사람의 공격을 방어하는 방어행동들로 나눠진다. 반사작용에 의한 행동은 가상검객의 의지와 관계없이 수행하는 행동으로 사람 검객이 머리를 타격하려고 할 경우, 가상검객은 뒤로 몸을 피하거나 검으로 막는 등의 반사작용에 의한 행동을 실행한다. 또한, 음성 입력이 활성화되면 공격을 방어하려는 반사행동이 실행된다. 모든 행동 시스템은 계층적(hierhiercal) 구조로 이뤄지고 각 움직임은 주어진 조건에서 부드럽게 연결되어야 한다. 그러므로, 모든 가능한 검의 움직임 또한 계층적으로 구성되고 각 행동의 계층적 네트웍은 인식과 동기 시스템에 따라 선택된다.

모터 시스템은 가상검객이 기술을 실행하고 그것을 스크린에 디스플레이 하도록 하는 수단으로써 실행되며 적절한 기술의 구현을 위해 실시간 움직임을 보간한다. 검객의 기본 행동을 신체 운동과 검 운동으로 구분할 수 있고 신체 운동은 정지, 전방, 후방, 좌, 우 등의 운동으로 정의되고, 검 운동은 전 자세와 동일, 기본 자세로 정지, 전방, 후방, 좌, 우 머리 타격, 좌 우 허리 타격으로 구성된다. 그림 6은 이상에서 설명된 가상검객 행동의 개요도를 나타낸다. 화살표는 구성체 사이의 정보 흐름을 나타낸다.

2.3 시청각 피드백(visual and audio feedback)

시합 동안에 사람 검객에게 시각적인 피드백을 주기 위해 OpenGL API를 사용하여 가상환경을 구현하였다. 사람의 기하학적 모델은 구형 타입의 머리 및 실린더 타입의 상체, 하체로 구성되며 사람과 검 모델의 움직임은 시각 인터페이스에 의해 조절된다. 시각 인터페이스에서 구해진 사람 중심점의 x, y 좌표는 사람 모델과 검의 x, y 병진 운동을 조작하고 시각 인터페이스로부터 구해진 사람 축에 대한 z축의 회전은 사람 모델과 검을 z축으로 회전시킨다. 두 개의 색깔 마커 중심점을 이용해서 구해진 검의 각은 피치(pitch)와 요(yaw)의 움직임으로 나눠진다. 검의 피치 움직임 는 그림 7에 보인 것 같이 어깨, 팔목, 손목의 연속된 회전을 만들고 위 회전들은 검의 피치 각과 관련되고 그 조건식 (9)는 다음과 같다.

$$\phi = c_1\alpha_1 = c_2\alpha_2 = c_3\alpha_3 \tag{9}$$

위 식에서, c_1, c_2, c_2는 실제적인 팔의 움직임을 구성

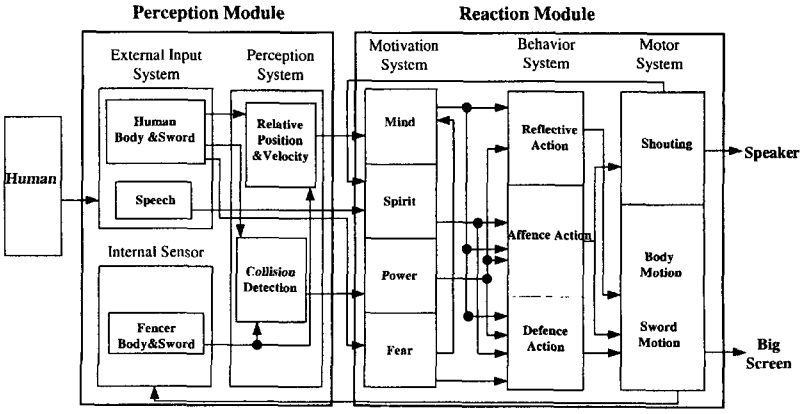


그림 6 가상검객의 행동 개요도

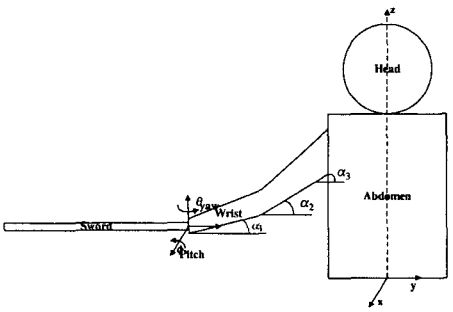


그림 7 가상 사람/검과 실제 검 움직임 사이의 움직임 매핑

하기 위해 실험적으로 미리 정해진 상수이다. 검의 요 움직임은 사람 모델의 움직임에 영향을 주지 않는다.

가상검객의 의도를 파악하고 대결시의 충돌상황을 사용자에게 피드백시켜 몰입감을 증대시키기 위해 청각 피드백이 사용된다. 충돌 검출 피드백을 위해 사람 모델의 타격 위치인 머리, 허리, 손목 및 검과 상대 검의 충돌로 전체 충돌 검출이 결정되는데, 두 검의 충돌 검출은 검 사이의 최단 거리를 계산함으로써 검출될 수 있다. 즉, 두 검 사이의 충돌은 다음 조건식 (10)으로 구할 수 있다.

$$D < 2R_{sword} \tag{10}$$

위 식에서, D는 검을 축으로 하는 선의 최단 거리로 표현되고 R_{sword}는 검의 반경이다. 허리와 손목의 형상은 간단한 실린더로 모델링되었기 때문에 충돌 검출은 비슷한 방법으로 계산할 수 있으며 머리의 형상은 구로 모델링되어 검과 머리의 충돌 검출은 선과 구의 충돌로 구현된다. 제안된 검도시스템에 사용된 단순한 형상에 기초한 충돌 검출은 다각형(Polyogn) 기반의 충돌검출 알고리즘과 비교해 계산량을 현저하게 줄일 수 있다 [13,14]. 현재 검도 게임 시스템에서는 실제 검이 가상검과 충돌하더라도 사용자가 충돌하는 힘을 느끼지 못하므로 상대 검을 통과하여 상대 목표지점을 타격할 수도 있다. 따라서, 검의 충돌 직후의 상대목표 타격은 무효로 처리된다.

그림 8은 제안된 검도 게임에서의 피드백 절차를 보여준다. 가상환경은 충돌 검출, 그래픽 랜더링, AI, 그래픽 디스플레이 그리고 충돌 사운드로 구성된다. 시각과 음성 인터페이스를 통해, 사람과 검의 위치는 가상환경으로 전송되고 이 정보는 그래픽 랜더링, 충돌 검출을 갱신하고 가상 검객의 커널에 따라 가상검객의 행동을 생성한다.

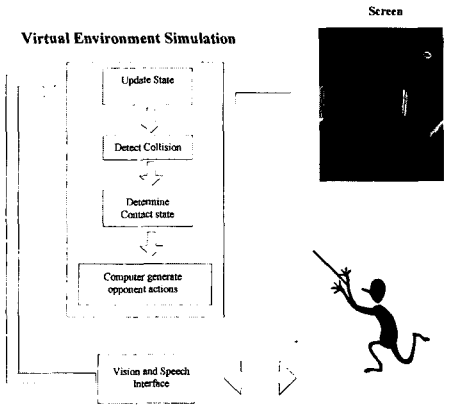


그림 8 피드백 행동의 절차

3. 시스템 구현 및 실험

그림 9는 멀티모달 인터페이스, 인공지능 그리고 시청각 피드백을 사용하여 구현한 검도 게임 시스템을 나타낸다. 사람검객이 가상검객을 정면에서 볼 수 있도록 가상환경이 일인칭 시점(view point)으로 선택되었다. 검과 사람을 효율적으로 검출할 수 있도록 카메라는 사용자의 중심점을 바라보도록 위치시켰고, 대형 스크린은 사용자가 현실감을 느끼도록 정면에 위치시켰다. 그리고 음성인식을 위해 사용자가 무선 마이크를 착용하도록 하였다. 사용자가 가상검객과 대결하는 동안 현실감을 제공하기 위해 OpenGL로 모델링된 가상검객의 반응이 대형 스크린을 통해 시각적으로 사용자에게 피드백되고, 충돌 상황 및 가상검객의 구호가 스피커를 통해 사용자에게 청각적으로 피드백된다. 사운드 피드백을 위해서는 검과 검 사이의 충돌시의 소리 및 머리, 허리, 손목 같은 함성 소리 등을 미리 녹음하여 상황에 맞게 재현하였다.

이를 기반으로 사용자와 가상검객과의 대련은 다음과 같은 과정으로 진행되는데, i) 시각 인터페이스로부터 사용자와 실제 검의 위치를 얻고, ii) 실제 검과 사람의

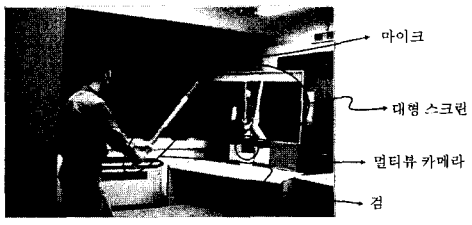


그림 9 시각 청각 피드백의 몰입감의 검도 게임

좌표를 가상환경의 좌표로 전환한다. iii) 충돌 검출 실행 및 사람 검객의 행동을 인식한 후, iv) AI 과정으로 가상환경에서 가상검객의 반응 행동을 결정하고, 최종적으로 v) 대형 스크린에 가상검객의 반응을 디스플레이하고 충돌사운드를 피드백한다.

제안된 시스템은 Dual-Pentimum III 1.8GHZ의 컴퓨터에 multi-thread로 구현되었다. 계산 속도에 있어 시각 인터페이스에 의해 계산된 검출 가능 대상물의 최대 속도는 9Hz였다. 이 결과는 멀티뷰 영상에 의한 삼차원 양안차의 계산량 때문이다. 이러한 사람의 추적 성능은 가상검객과 시합을 하기에는 충분한 편이나, 검의 속도는 대결하기에는 조금 느린 편이다. 표 1은 제안된 검도 시스템의 실험으로 평가된 성능을 보여주고 있다.

표 1 검도게임의 성능

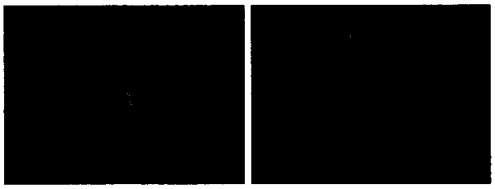
측정 변수	측정값
대결 가능공간	2m (width) × 4m (height)
칼의 측정분해능	5 deg (pitch), 5 deg (yaw)
몸의 측정 분해능	0.05m
시스템 대역폭	최대 9Hz
음성인식	3 Words Recognition

그림 10은 검도 게임을 위한 사용자의 분할 결과들을 보여준다. 그림 10(a)는 확률 정보를 이용하여 배경에서 객체를 분리한 분할 영상, 그림 10(b)는 대응되는 양안차 영상을 나타낸다. 실험 결과에 따르면 제안된 복합 분할 기법은 효과적으로 사용자를 블루스크린(blue screen) 뿐만 아니라 일반 배경의 영상에서도 분리해 낸다는 것을 알 수 있었다[15-17]. 그림 11은 색깔 영상분할의 결과를 보여준다. 그림 11(a)는 배경에서 대상물체가 분리된 전경에서 색깔 정보에 기반해서 마커를 다시 분리한 영상이고 그림 11(b)는 대응되는 양안차 영상이다.

제안된 시각 시스템의 정밀도를 확인하기 위해 모션트랙커(motion tracker) 기반의 센서와 비교하였고 그림12에 나타낸 것과 같이 시각 기반의 시스템이 모션트랙커 기반의 시스템과 비교해 넓은 행동 반경을 요하는 검도 시스템에 적합한 것으로 나타났다. 그림 12(a)는 가로축을 기준점으로부터 실제 대상물까지의 거리로 잡고 세로축을 각각 카메라에서 실제 대상물까지의 측정된 거리 및 FASTRAK receiver에서 transmitter까지의 측정된 거리로 잡았을 때의 결과를 보여주고 있다. 그림 12(b)에서는 각각의 경우 거리에 대한 측정 오차를 보여주고 있다. 거리가 2m 이후에서는 거리가 증가함에 따라 시각 시스템의 정밀도가 모션 트랙커보다 우월한 것을 볼 수 있다. FASTRAK의 사양에서는 기준 receiver가 기준 transmitter의 76cm 안에 위치할 때만 사양의

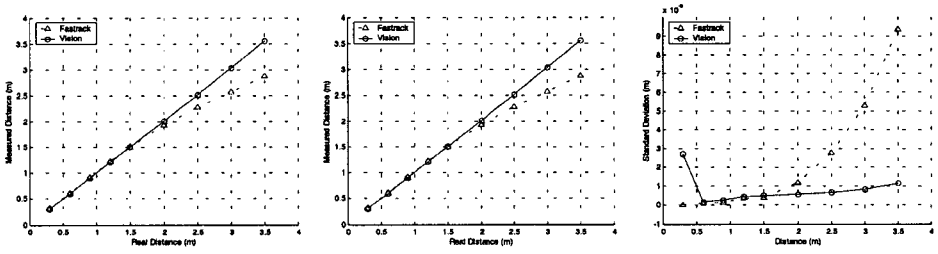
(a) 사용자 분할 (b) 대응되는 양안차 영상

그림 10 사용자 분할



(a) 칼러 분할 (b) 대응되는 양안차 영상

그림 11 색깔 분할



(a) 대상물까지의 실제거리(m) (b) 거리 오차(m) (c) 표준편차(m)

그림 12 모션 트랙커(FASTRAK) 시스템과 시각 인터페이스의 정밀도 성능 비교

정확도를 제공한다고 제시하고 있다[18]. 그림 12(c)에 보여진 것처럼 트랙커 기반의 경우 표준편차가 시각 기반의 시스템과 비교해 거리가 증가함에 따라 악화되는 것을 볼 수 있다. 그리고 그림 12(c)에서 양안차 계산 범위에 대한 한계로 검객과 멀티뷰 카메라의 거리가 0.3 m 이내일 때 다른 값과 비교해 오차가 크게 증가하였으나 그 외의 부분에서는 상대적으로 잘 동작하였다. 또한, 일반적으로 시각 기반 시스템이 외부환경의 조명 때문에 강건(robust)하지 못하다는 단점이 있는데 반해, 제안된 시각 시스템은 normalized color space를 채택함으로써 일반적인 실내 조명 조건의 외부 환경에 크게 영향을 받지 않고 잘 작동하였다.

그림 13은 영상분할, 3차원 위치 계산 그리고 랜더링 시간 등을 고려하여 3차원 공간상에서 검이 움직이는 속도를 식 (8)을 사용하여 보상한 결과를 나타낸 것으로, 사용자가 카메라로부터 2.2m 떨어진 거리에 위치하여 검을 흔들면서 실험하였다.

그림 13(a),(b),(c)는 카메라의 샘플링률보다 느리게 움직이는 객체에 대해, 실제의 경우, 보상되지 않은 경우 그리고 보상된 경우에 대해 검이 위치하는 좌표를 나타낸다. 즉, 실선은 현시점에서 객체가 위치 해야할 3차원 공간상의 좌표를 나타낸 실제적인 좌표, 삼각형은 영상분할, 3차원 위치 계산 등의 시간지연 요소로 인해 얻어지는 현시점에서의 좌표로 오차를 갖는 객체의 3차원 위치값이다. 마지막으로, 원형은 이와 같은 오차를 보상해서 얻은 결과를 나타낸 것이다. 그림 13(d),(e),(f)에서 보여지는 것처럼 x, y, z 각 방향으로 각각 대응되는 오차를 통해 제안된 방법이 효과적임을 알 수 있다.

제안된 시스템의 음성 입력에서 머리, 허리, 손목의 세 단어를 구분해 내기 위해 각각의 단어에 대한 켑스트럼 계수를 구하였다. 세 단어를 구분하기 위해 단어의 첫 번째 음절 "머", "허", "손" 만 구분하면 가능하기 때문에 실제 표본 켑스트럼 계수를 각 단어의 첫자만을 이용하여 구하였고 이때 각 단어의 첫자만을 인식하기 위해 8개의 윈도우를 사용하였다. 따라서, 하나의 윈도우는 12개의 켑스트럼 계수를 가지며, 전체 96차수의 켑스트럼 계수의 특징 벡터가 된다. 그림 14에 세 단어 인식을 위한 각각의 첫 문자 켑스트럼 계수를 나타내고 있다. 입력 음성이 문턱치 값을 넘으면 그 시점을 기준으로 주어진 시간 윈도우(time window)동안 데이타를 획득하여 LPC켑스트럼 계수를 구하였고, 그림 14에서 구해진 켑스트럼 계수와의 Euclidean 거리를 구해 가장 적은 값을 구했다. 주의할 점은 실제 검도 시합에서는 타격하고자 하는 위치와 검객의 함성이 순간적으로 같아야 하지만 현실적으로 동시에 충돌 검출과 음성인식을 측정하는 것은 실시간 움직임의 충돌과 음성의 불규칙적인 간격의 차이 때문에 어렵다는 것이다. 따라서 음성인식은 가상 검객에게 부가적인 정보를 주는 용도로만 사용되었다.

음성 인식에 관하여 세 단어 각각의 음성 인식률을 실험하였다. 표 2에 음성신호와 각 Euclidean 거리의 예를 보여주고 있다. 손목은 다른 두 단어와 비교해 틀

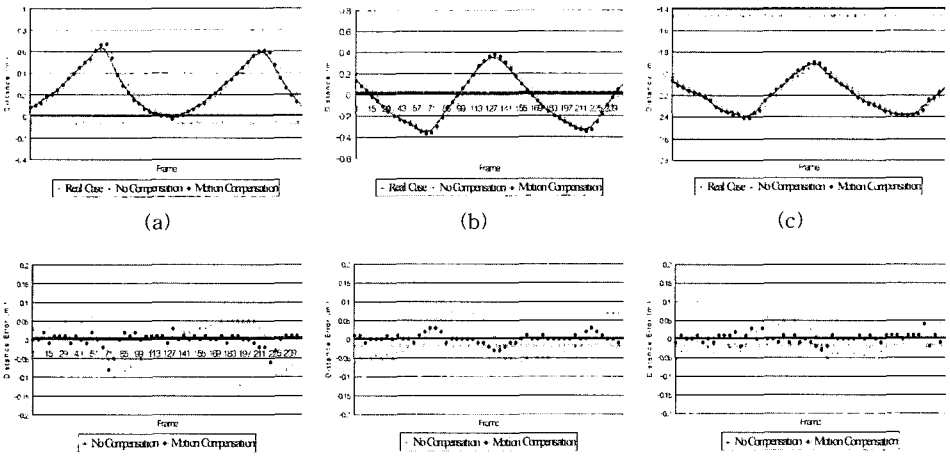
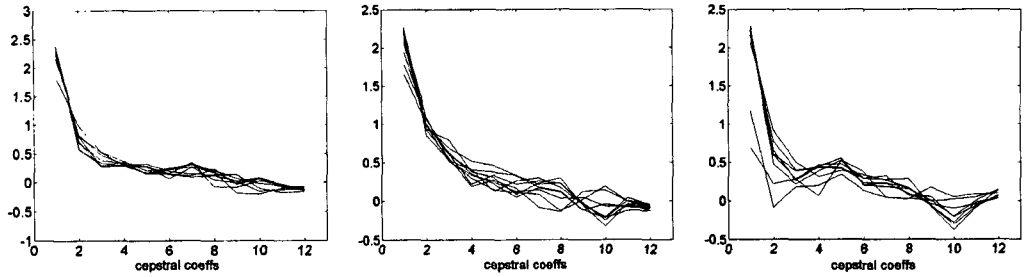


그림 13 검의 x, y, z 각 방향으로의 움직임 보상 결과, (a),(b),(c) 각각 x, y, z방향으로의 보상 결과 (d),(e),(f) 보상이 된 경우와 안된 경우의 실제 검의 좌표와의 오차



(a) "머" (b) "허" (c) "손"

그림 14 세단어의 켑스트럼 계수

표 2 패턴 음성신호와 새로운 음성신호의 켑스트럼계수의 Euclidean 거리

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

런 음운 구조를 가지기 때문에 80%의 음성인식 성공률을 보였다. 반면, 머리, 허리는 앞의 두 자음을 제외하고는 동일한 음운을 갖기 때문에 인식률이 상대적으로 낮았다(평균 60%). 전체적으로 사용자가 실험에 의해 훈련되었을 경우, 음성 인식 성공률은 70% 이상이었다. 비록 음성인식이 실패하더라도 사용자의 구호에 의한 가상 검객의 행동 변화를 사용자는 예상할 수 없기 때문에 게임을 하는 데는 크게 문제를 야기시키지는 않았다.

4. 결론 및 추후 과제

본 논문에서는 멀티모달 인터페이스(시각 및 음성기반 인터페이스), 지능 에이전트, 몰입형 상호작용을 위한 피드백(시각 및 음성)으로 구성된 단순하고 강건한 3차원 검도 게임시스템을 제안하였다. 제안된 시스템은 가상환경에서 가상검객과 사용자를 방해함이 없이 전신의 상호작용을 가능하게 하였고 비교적 느린 시각 시스템의 속도(9Hz)를 보상하기 위해 외삽에 의한 검의 위치를 예측하였다. 실험 결과 제안된 시각 시스템은 일반적인 움직임 센서 시스템인 모션 트랙커 시스템에 비해 보다 넓은 작업공간에서 사용 가능함을 보였고 큰 작업공간(2m×2m)에서 사람의 위치를 추적하는데 큰 무리가 없는 정밀도(5cm 이내)를 가졌다. 다음으로 무선에 의한 사용자의 음성을 인식하여 가상 검객의 지능에 사용함으로써 사용자의 게임에 대한 몰입감을 증대시켰다. 마지막으로, 사용자와 가상검객 사이에 실제감 및 몰입감을 증가시키도록 멀티모달의 정보를 가지는 지능의 가상검객을 구현하였다. 현재까지 각 부분의 통합은 사용자로 하여금 온몸의 운동 및 음성을 사용하도록 유도하여 상대 검객과 흥미진진한 시합을 할 수 있다고 생각된다. 추후에는 검과 검, 또는 검과 타격 목표지점이 충돌할 때의 힘을 경험할 수 있는 검도 시스템을 개발할 예정이다. 시각과 음성인식이 결합된 힘반영(force feedback)은 검도게임에 완전한 몰입감을 제공할 것이다. 또한 3 차원 가상환경에 실감 영상의(photo-realistic) 아바타를 삽입할 경우보다 현실적인 경험을 제공할 것이다.

참 고 문 헌

[1] J. Bates, J. Altucher, A. Hauptman A, M. Kantrwoitz, A. Loyall, K. Murakami, P. Olbrich P, Z. Popovic Z, W. Reilly, P. Sengers P, W. Welch, P. Weyharauch, and A. Witkin, "Edge of Intention," *SIGGRAPH 93 Visual Proceedings, Machine Culture, ACM SIGGRAPH*, pp. 113-114, 1993.

[2] F. Fisher, M. Girard, S. Amkraut, and Menagerie, "Tomorrow's Realities," *SIGGRAPH-93 Visual Proceeding, ACM SIGGRAPH 1993*, pp. 212-213, 1993.

[3] I.S Y Yoon, R. Burke, B. Blumberg, and G. Schneider, "Interactive Training for Synthetic Characters," submitted to *AAAI 2000*.

[4] L. Emering, R. Boulic, and D. Thalmann, "Interacting with Virtual Humans through Body Actions," *IEEE Computer Graphics and Appli*

cations, Vol.18, No1, pp.8-11, 1998.
[5] L. Emering, R. Boulic R, S. Balcisoy, and D. Thalmann D, "Real-Time Interactions with Virtual Agents Driven by Human Action Identification," *First ACM Conf. on Autonomous Agents'97*, Marina Del Rey, pp.476-477, 1997.
[6] T. Molet, A. Aubel., T. Çapin, S. Carion., E. Lee, N. M. Thalmann, H. Noser, I. Pandzic, G. Sannier, and D. Thalmann, "ANYONE FOR TENNIS?," *Presence*, Vol. 8, No. 2, pp. 140-156, April 1999.
[7] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, "The ALIVE System: Full-body Interaction with Autonomous Agents," *In Proc. Computer Animation*, Geneva, Switzerland, IEEE Computer Society Press, Los Alamitos, California, ISBN 0-8186-7062-2, pp. 11-18, 1995.
[8] F. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campell, Y. Ivanov, A. Schutte, and A. Wilson, "The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment," *Presence*, Vol. 8, N0. 4, pp. 369-393, Aug.1999.
[9] Gavrila, L.S. Davis, "3D Model-Based Tracking of Humans in Action: A Multi-View Approach," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp 73-80, SanFrancisco, USA, June 1996.
[10] A. Mulder, Human movement tracking technology, Simon Fraser University: Technical Report 94-1, 1994.
[11] B. K. P. Horn. Robot vision. MIT Press, 1986.
[12] Shuzo Saito, "Fundamentals of Speech Signal Processing," Academic Press, 1985
[13] Bruce F. Naylor. *A Tutorial On Binary Space Partitioning Trees*. Computer Games Developer conference Proceedings, pp 433-457, 1998.
[14] S. Gottschalk, M. Lin, D. Manocha, "OBB-Tree: A Hierarchical Structure for Rapid Interference Detection," *SIGGRAPH 1996*, pp.171-180, 1996.
[15] W. Woo and Y. Iwadate, "Object-oriented hybrid segmentation using stereo images," *in Proc. SPIE VCIP*, pp. 487-495, Jan. 2000.
[16] W. Woo, N. Kim, and Y. Iwadate, "Object segmentation for z-keying using stereo images," *in Proc. WCC*, pp. 1249-1253, Aug. 2000.
[17] N. Kim, W. Woo, and M. Tadenuma, "Photo-realistic 3d virtual environment using multiview video," *in Proc. SPIE VCIP*, Jan. 2001.
[18] Polhemus, http://www.polhemus.com/ftrakds.htm.

윤 정 원
1998년 전북대학교 정밀기계공학과(공학사). 2000년 광주과학기술원 기전공학과(공학석사). 2000년~현재 동 대학원 기전공학과 박사과정 관심분야는 HCI, Rehabilitation Robot, Man Machine Interface,

김 세 환
1998년 서울시립대학교 전자공학과(공학사). 2000년 광주과학기술원 정보통신공학과(공학석사) 2000년~현재 광주과학기술원 정보통신공학과(박사과정). 관심분야는 Virtual/Mixed Reality, 3D Vision, HCI, Wearable Computing

류 제 하
1982년 서울대학교 기계공학과(공학사) 1984년 KAIST 기계공학과(공학석사) 1991년 The University of Iowa, 기계공학과(공학박사). 1992년~1994년 United Defense LP 선임 연구원. 1995년~현재 광주과학기술원 기전공학과 교수. 관심분야는 병렬로보트 기구학/동력학/제어/Calibration, Haptic Device for VR interface

우 운 택
1989년 경북 대학교 전자공학과(공학사) 1991년 포항공과대학교 전기전자공학과(공학석사). 1998년 University of Southern California, Electrical Engineering-Systems(공학박사) 1991년~1992년 삼성종합기술연구소 연구원. 1999년~2001년 ATR MIC Lab., Japan, 초빙 연구원. 2001년~현재 광주과학기술원 정보통신공학과 조교수 관심분야는 3D Vision, vision-based HCI, Networked Mixed Reality, Ubiquitous Computing, Wearable Computing