

핸드제스처를 이용한 원격미팅 자료 인터페이스

송제훈*, 최기호**, 김종원**, 이용구***

Hand Gesture based Manipulation of Meeting Data in Teleconference

Je-Hoon Song*, KiHo Choi**, JongWon Kim** and Yong-Gu Lee***

ABSTRACT

Teleconferences have been used in business sectors to reduce traveling costs. Traditionally, specialized telephones that enabled multiparty conversations were used. With the introduction of high speed networks, we now have high definition videos that add more realism in the presence of counterparts who could be thousands of miles away. This paper presents a new technology that adds even more realism by telecommunicating with hand gestures. This technology is part of a teleconference system named SMS (Smart Meeting Space). In SMS, a person can use hand gestures to manipulate meeting data that could be in the form of text, audio, video or 3D shapes. For detecting hand gestures, a machine learning algorithm called SVM (Support Vector Machine) has been used. For the prototype system, a 3D interaction environment has been implemented with OpenGL™, where a 3D human skull model can be grasped and moved in 6-DOF during a remote conversation between distant persons.

Key words : Hand gesture recognition, Hand user interface, Teleconference, Smart multi-party collaboration system, Tiled display

1. 서론 및 문헌조사

데이터글러브가 발명된 이후, 이를 응용한 기술들이 속속 등장하고 있다. 데이터글러브란 장갑을 착용한 사람의 손과 컴퓨터간의 정보교환을 위한 입출력 장치들을 가리키는 말이다^[1]. 이것은 손 동작 데이터를 획득하는 기능만을 가지는 태입과 컴퓨터가 제시한 촉감정보를 사용자가 느낄 수 있도록 피드백 기능을 추가한 태입이 있다. 리시버 및 트랜스미터를 이용하여 공간상의 절대위치를 감지하는 구조를 가진 데이터글러브들이 상대적으로 고가인 반면, 가속도계를 이용하여 손의 상대적인 움직임을 감지하는 구조의 저렴한 데이터글러브도 개발되었다. 스킨리라고 명명된 이 장치는 착용하는 입력장치를 목표로 개발되었으며 손가락 동작을 이용하여 키보드 입력이 가능하다^[2].

기존의 연구자들은 이러한 데이터글러브로부터 획득된 사람의 손동작 정보를 이용하여 유용한 일들을 할 수 있음을 알았고, 이와 관련하여 가상환경 속 물체들과의 정교한 상호작용 및 수화번역 등의 용도로 이용할 수 있음을 보였다. 가상환경에서 상호작용을 위해 데이터글러브를 이용하는 연구에서는, 양자화된 핸드제스처들을 일종의 패턴 집합으로 고려하였다. 여기서는 Artificial Neural Networks(ANN)을 이용하여 특정 사용자의 손에 적용된 분류를 수행하였다^[3]. 인공신경망 이외에 Hidden Markov Models(HMM)을 이용하여 수화(American Sign Language)를 인식하는 연구에서는 유연장갑의 일종인 CyberGlove(R)와 Flock of Birds(R)라는 모션트래킹을 이용하여 핸드제스처를 수집하였다^[4]. 사실적인 핸드모델 및 3D Object에 대한 실시간 충돌검사 및 포획(Grasp)을 수행한 연구도 발견되었다^[5]. 이 연구에서는 사용자에게 실감을 주기 위한 목적으로 기하계산에 기반한 정교한 충돌검사를 사용한다. 이러한 기존의 연구들에는 상이한 두 가지의 관점이 발견되는데, 그것은 각각 다음과 같다.

*학생회원, 광주과학기술원 정보기전공학부

**광주과학기술원 정보기전공학부

***교신저자, 정회원, 광주과학기술원 정보기전공학부

- 논문투고일: 2006. 12. 28

- 심사완료일: 2007. 03. 19

(1) 손 데이터는 가시 데이터이다. 따라서 이것은 3D 데이터와의 충돌검사를 수행하기 위해서 컨벡스럴에 기반한 충돌검사를 수행한다. 그리고 충돌 위치들을 계산하여 정교한 상호작용을 가능하게 한다. 그리고 이것은 햅틱 장비들을 제어하기 위해서 활용될 수 있다. 사용자는 가상의 환경 속에서 재구성된 자신의 손과 물체들이 상호작용하는 모습을 사실감 있게 관찰할 수 있고, 포스피드백 기능을 이용 가능한 경우에는 컴퓨터의 계산에 의해 만들어진 역감도 느낄 수 있다.

(2) 손 데이터는 사용자의 의사 데이터이다. 손 동작은 언어 대신에 사람의 의사를 표현하기 위해서 만들어지는 것이며, 데이터글러브를 이용해서 획득한 손 데이터는 사용자의 의사표현을 의미한다. 따라서 각각의 의사표현에 해당하는 손 데이터들을 정의해야 한다. 그것을 해석하기 위해서 손 데이터들은 일종의 패턴집합으로 고려되며, 패턴의 분류를 위해서 머신러닝 기반의 알고리즘들을 이용한다. 컴퓨터가 사람의 동작을 분류할 수 있다면, 이를 이용해서 여러 가지 명령을 수행하도록 할 수 있을 것이다.

한편, 점점 가속화되는 IT(Information Technology) 기술로 사람들의 공간에 대한 개념을 근본적으로 바꾸어 놓고 있으며 다른 국가에 위치한 동료와 실시간으로 정보교환을 할 수 있게 하였다. 최근의 광대역 네트워크 상용화는 음성 및 영상을 포함한 고용량의 데이터를 실시간으로 전송할 수 있게 해주었다. 이러한 기술적 배경으로, 연구자들은 광대역 네트워크 채널을 이용하여 원격지의 사람들과 효율적인 회의를 진행할 수 있는 방법들을 모색하고 있다. 현재의 연구 이슈는 텔레컨퍼런스 공간에서의 정보공유, 인터랙션 그리고 상황인지(Context Awareness) 등이 있다. 정보의 공유는 원격지의 사람들과 효율적인 회의를 진행하기 위해 자료 화면 및 회의 상황에 대한 비디오 정보를 컨퍼런스 그룹간에 전송하는 문제이다. 연구자들은 로컬의 참여자들과 비디오 정보를 공유하는 초대형/고해상도 가시화 장치 구현을 위해서 SAGE(Scalable Adaptive Graphics Environment)⁶⁾와 SAGE에서 비압축 비디오를 전송할 수 있는 SVC(Scalable Video Consumer)로 공유문제를 해결하려고 시도하였다⁷⁾. 또한 원격지의 동료들과 가상의 공간에서 만나고 공동으로 문서를 만들도록 하는 기초적인 연구도 있었는데⁸⁾, 물리적으로 분산되어 위치하는 동료들과 공유하는 가상의 공간에서 공동의 문서 작업을 하기 위한 연구였다.

텔레컨퍼런스는 참여자들이 상호 공유하는 가시화

데이터들과 인터랙션할 수 있는 환경을 제공해야 한다. 예를 들어 레이저포인터를 이용하는 연구⁹⁾에서는 스크린의 특정 위치를 가리키는 레이저포인터를 추적하고, 포인터 장치에 추가된 버튼과의 조합으로 기존의 마우스와 유사한 기능을 구현하였고, 벽에 설치된 대형 가시화 장치에 표현된 정보들을 용이하게 조작하기 위해 고안되었다.

또한 로컬 및 리모트에 위치한 그룹들에는 굉장히 많은 수의 사람들이 참여할 수 있는데, 이때 현재 발언하고 있는 사람이 누구인지, 어디에 위치하고 있는지를 인지하고 알려주는 기능이 강하게 요구될 것이다. 이는 상황인지와 관련한 이슈이며, Yoshimi¹⁰⁾에 의해서 연구된 바 있다.

본 논문에서는 앞에서 설명한 텔레컨퍼런스의 최근 이슈들 중 하나인 인터랙션에 관해 다룬다. 또한 데이터 글러브를 텔레컨퍼런스에서의 인터페이스 하드웨어로 제안하고, 이를 통해서 텔레컨퍼런스에서 공유되는 가시화 데이터의 조작에 이용하고자 한다.

앞에서 설명한 바와 같이, 획득되는 손 데이터는 가시 데이터의 특성과 의사 데이터의 특성을 모두 가지고 있지만 텔레컨퍼런스를 위해서는 의사 데이터의 관점이 좀더 고려된다. 의사 데이터의 관점이 상대적으로 낮은 계산복잡도를 가진 시스템의 구현이 가능하며, 텔레컨퍼런스에서는 상호작용의 실감 구현보다는 그 효율성이 중요하기 때문이다.

2. 논문의 목표

이 논문의 목표는 텔레컨퍼런스 룸의 가시화 장치에 디스플레이 된 자료 화면들과의 상호작용을 위해 회의 참여자가 자신의 자연스러운 손 동작을 이용할 수 있도록 하는 것이다. 이를 위해서 HUI(Hand User Interface)와 4×4 행렬연산에 기반한 3D 데이터의 조작방법을 정의하였다.

2.1 지능형 다자간 협업환경

광대역 네트워크 환경은 원격지의 동료들과 고용량, 고화질의 가시화 데이터들을 실시간으로 교환할 수 있게 하였다. 네트워크를 통해 전송된 데이터들은 텔레컨퍼런스 룸에 설치된 타일드 디스플레이 장치를 통해서 가시화되며, 이를 원격지의 동료들과 공유할 수 있다. 타일드 디스플레이 장치는 초고해상도의 그래픽스 데이터들을 가시화할 수 있으며, 리모트에 위치한 그룹들과의 협업에 이용된다.

이 논문에서는 네트워크 기반의 가시화 하드웨어들

을 이용하는 텔레컨퍼런스 시스템을 “지능형 다자간 협업 환경”으로 명명하였고, 이를 SMS라 칭하도록 하였다. 초고해상도 가시화 시스템은 광대역 네트워크와 타일드 디스플레이를 이용하여 구성한다. 타일링되어 있는 모니터들은 2개가 1쌍이 되어 하나의 컴퓨터에 의해서 통제된다. SMS에서는 그래픽스 데이터의 디스플레이를 위해 이러한 타일드 디스플레이 시스템을 활용하고 있고, SW라 칭하고 있다. 회의에 참여하는 사람들은 SW에 가시화된 고화질의 자료화면들을 보면서 로컬과 리모트SMS에 위치한 동료들과 효율적인 회의를 진행할 수 있다.

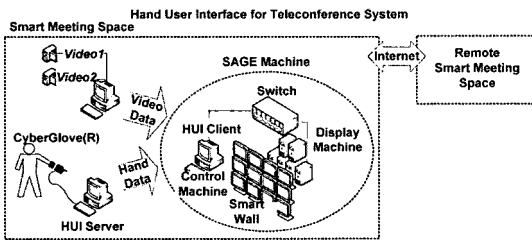


Fig. 1. 타일드 디스플레이를 이용한 SMS의 구성.

이러한 SMS를 구성하기 위한 타일드 디스플레이 장치의 구성을 Fig. 1에서 도시하였다. 본 논문에서는 6×4 크기의 타일드 디스플레이를 구성하기 위해서 24개의 LCD 모니터들과 12개의 PC가 사용되었다. 12개의 컴퓨터들은 네트워크 스위치에 연결되어 있으며 제어 머신에서 생성한 고해상도 그래픽스 데이터를 수신하여 자신이 통제하는 LCD 모니터에 이미지 데이터를 렌더링한다. 제어 머신이 생성한 그래픽스 데이터의 통제를 위해서 내부적으로 SAGE가 이용된다. 이러한 SMS는 리모트에 위치한 SMS와 인터넷을 통해서 연결되며 SMS에 의해서 공유되는 데이터들은 로컬 및 리모트SMS에 위치한 동료들에 의해서 조작된다.

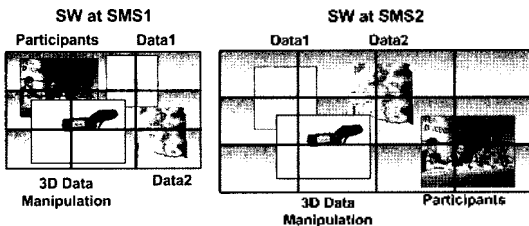


Fig. 2. 스마트미팅공간에서 회의를 진행하는 화면.

Fig. 2는 원격 및 리모트에서 동시에 텔레컨퍼런스를 진행할 때의 SW를 나타내었다. SMS1과 SMS2는

분리적으로 떨어져 있는 위치에 존재한다. 회의에 참여하는 사람들은 SW를 통해서 다른 위치에 존재하는 동료들의 모습을 볼 수 있고 관련 데이터를 조작할 수 있다. 회의가 시작되면, SW에는 회의에 필요한 여러 개의 자료화면이 가시화되어 있는 상태가 된다.

이제 남아있는 문제는 로컬 및 리모트에 위치한 동료들이 공유되는 그래픽스 데이터들을 용이하게 조작하도록 돕는 것이다. 회의 참여자들은 회의에 필요한 여러 자료화면들을 보다 편리하고 자연스런 방식으로 조작하기를 원할 것이다. 이것은 인터페이스와 관련이 있는 사항이다. 키보드 및 마우스가 기본적인 인터페이스로 이용될 수 있을 것이다. 왜냐하면, 텔레컨퍼런스 시스템이 근본적으로는 컴퓨터의 제어에 의해서 실행되고 있기 때문이다. 하지만 참여자들이 컴퓨터의 존재를 알고 있을 필요는 없다. 참여자들에게 보다 중요한 것은 원활한 컨퍼런스의 진행이기 때문이다.

참여자들의 손에 착용된 장치를 이용하여 직관적인 방식으로 가시화 장치에 디스플레이 된 데이터들을 조작하도록 한다면, 회의에 참여하는 사람들을 더욱 회의에 집중할 수 있을 것이다. 이러한 배경으로 핸드 유저인터페이스(HUI)의 개발을 착수하였다.

2.2 SMS를 위한 상호작용 환경

이 논문에서는 SMS의 참여자들이 회의진행에 필요한 자료화면을 그들의 손으로 제어할 수 있도록 개발된 핸드유저인터페이스를 소개한다. SMS의 인터페이스에 해당하는 HUI는 SW에 표현된 윈도우를 제어하기 위해서 데이터글러브를 착용한 사람의 손 데이터를 이용한다. Fig. 3은 참여자가 자신의 손을 이용하여 SW에 가시화된 윈도우를 제어하는 모습이다. 참여자는 3D 데이터의 6DOF 제어 수행 및 윈도우에 대한 2D 제어를 수행할 수 있다.

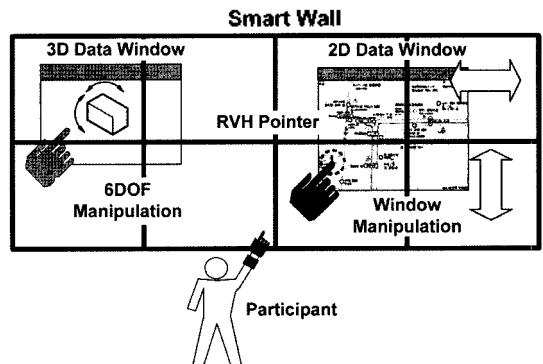


Fig. 3. 2D 및 3D 자료화면을 제어하기 위한 HUI.

3. 제안된 핸드유저인터페이스의 특징

본 논문에서는 가상현실을 위한 실감위주의 연구가 아니라 텔레컨퍼런스에서 제시되는 자료데이터의 제어 목적이기 때문에 상호작용의 효율성에 그 초점을 맞춘다. 상호작용의 실감이 증가하면 컴퓨팅자원의 소모는 그만큼 증가하게 되기 때문에 실시간으로 진행되는 SMS에서의 회의는 효율적이지 못할 수 있다. 컴퓨팅자원은, 리모드SMS 및 로컬SMS의 영상을 실시간으로 인코딩 및 디코딩 하는 데에도 적지 않게 이용될 것이기 때문이다. 따라서 HUI가 필요로 하는 컴퓨팅자원을 최소화시키는 동시에, 윈도우와 3D 데이터를 일관적으로 제어할 수 있는 상호작용 방법이 정의되어야 한다. 상호작용의 방법은 최소한 다음의 두 가지 주요 요건을 충족시켜야 한다.

(1) 2D 및 3D 데이터를 담은 윈도우는 그 크기 및 위치의 제어가 필요하다.

(2) 윈도우내의 3D 데이터에 대해서는 크기, 위치, 방향이 모두 고려된 6DOF의 제어가 수행되어야 한다.

이러한 목적들을 달성하기 위해서 본 논문에서는 단순화시킨 경계구(Boundary Sphere)에 기반한 접촉판별 모델과 핸드 제스처의 인지에 기반한 상호작용 모델을 제안한다. 이 방법은 정교하지는 않지만 계산복잡도가 작기 때문에 SMS 환경에서는 더 적합하다.

3.1 단순화시킨 접촉판별 모델

3D대상물체와의 접촉판별을 위해서는 각각의 3D 대상물체들을 담을 수 있는 경계구의 중심과 반경이 계산되어야 한다. 상호작용을 위해서 Reconstructed Virtual Hand(RVH)의 검지 끝이 경계구의 내부에 있는가를 판별한다. 이것은 Fig. 4에 도시되어 있다. Fig. 4(a)의 경우에는 경계구1의 내부에 RVH포인터가 존재하는 경우이다. 이 경우에는 경계구1에 대한 3D 물체를 “하이라이드” 한다. Fig. 4(b)의 경우는 경계구 1과 경계구2가 동시에 RVH 포인터를 포함하는 경우이다. 이 경우에는 경계구의 크기가 작은 쪽이 선택의 우선순위를 가지도록 한다. 이것은 경계구의 크기가 큰 쪽이 작은 경계구를 완전히 포함하는 경우를 고려한 것이다. 한편, 물체의 모양이 둥글지 않은 경우에 대해서는 단일의 구가 아니라 다수의 구를 이용하여 물체를 표현함에 의해서 인터랙션 오차를 줄일 수 있을 것이다.

그 이외의 3D 대상체를 선택하기 어려운 경우는 핸드제스처에 의해서 선택을 순환하도록 한다.

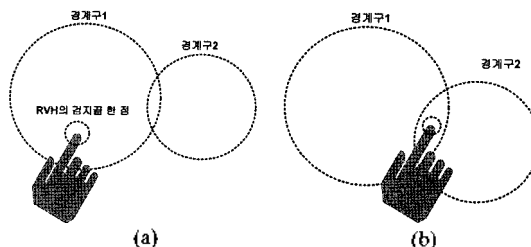


Fig. 4. RVH 포인터가 (a) 경계구1에만 포함되는 경우, (b) 경계구1, 2에 동시에 포함되는 경우.

Fig. 4(b)의 상황에서 사용자가 “Exchange” 제스처를 취하게 되면, 경계구1과 경계구2가 번갈아가며 선택한다.

2D 윈도우의 제어에 대해서도 비슷한 방식을 이용한다. Fig. 5을 보면, 자료화면1과 자료화면2, 그리고 RVH 포인터를 도시하였다. 개개의 자료윈도우들은 직사각형의 모양을 가지게 되고, 이들 직사각형이 윈도우의 경계상자로 이용된다. 이 경우에는 윈도우의 depth가 작은 것(앞쪽)이 선택의 우선순위를 가지게 되며 “Exchange” 제스처를 취했을 경우 약간의 시간차이를 가지고 자료화면에 대한 선택을 순환한다.

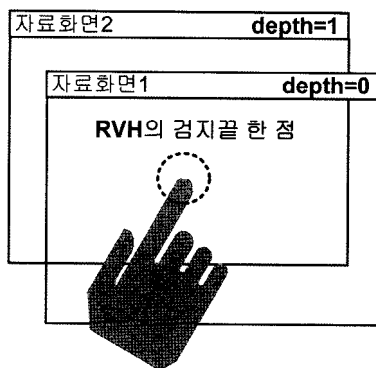


Fig. 5. 자료화면 윈도우와의 접촉판별.

SMS는 내부적으로 그래픽스 데이터를 통제하는 SAGE를 이용하여, 개개의 그래픽스 데이터에 대해서 크기 및 위치, 그리고 depth 속성을 부여할 수 있다. 이러한 depth 속성을 나타내는 정수 값을 선택의 우선순위를 부여하기 위해서 이용한다. 사용자의 핸드제스처가 “Grasp”인 경우, 우선순위에 의해 결정된 윈도우 또는 3D 자료화면 내의 3D 데이터를 제어한다. 특별히, “Grasp”에 대해서는 3D 데이터에 대한 6자유도 제어, 그리고 윈도우에 대한 2D 위치제어를 수행한다.

3.2 핸드 제스처의 인식을 위한 SUPPORT VECTOR MACHINE

가장 널리 쓰이는 두 개의 신경망 알고리즘은 MLP (Multilayer Perceptron)과 SVM(Support Vector Machine)이다. 본 연구에서는 분류(Classification) 작업에 대해 더 나은 성능을 보이며 학습모드에서 훨씬 빠른 속도를 가진다고 알려져 있는 SVM이 핸드제스처의 학습 및 분류에 이용되었다^[12,13]. SVM은 분류속도에 대해서도 만족스러운 성능을 보인다. Pentium(R) 4 3.2GHz PC에서의 실험결과, 144개의 특징을 가진 4종류의 핸드데이터를 평균 5 ms 이내에 분류할 수 있었다. 단순화된 접촉판별 및 SVM에 기반한 제스처 인지방법은 실감위주의 상호작용에 비해서 낮은 프로세서 점유율과 빠른 속도를 가진다. Fig. 6은 SVM 뉴럴네트워크의 구조를 도시한 것이다. 여기서 $\phi_k(x)$ 는 입력벡터들을 더 높은 차원의 데이터로 확장시키기 위한 비선형 변환 함수이다. 여기서 입력벡터는 사용자의 핸드제스처를 표현하는 벡터이다.

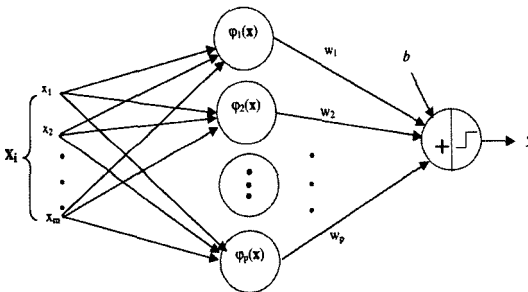


Fig. 6. SVM 뉴럴네트워크의 구조.

이러한 SVM 뉴럴네트워크의 출력은 다음과 같이 표현된다. 여기서 b 는 바이어스 항이다.

$$y(x) = \sum_{k=1}^p w_k \phi_k(x) + b$$

SVM이 MLP와 구별되는 주요특징은 $\phi_k(x)$ 에 의해서 입력벡터들이 선형분류 가능한 형태로 변환되어 단층의 뉴럴네트워크를 이용할 수 있다는 것이다. w_k 와 b 는 학습집합의 분류에러를 최소화하기 위해 클래스들의 중간을 가로지르는 하이퍼플레인 되도록 최적화된다^[14]. 최적화된 w_k 와 b 를 이용하여 핸드제스처에 대한 분류작업이 수행될 수 있는 것이다. 본 논문에서는 SVM의 소프트웨어 라이브러리 패키지인 SVMTorch^[13]를 핸드제스처의 학습 및 분류에 이용하였다. HUI는 내부적으로 SVM에 의한 학습 및 분류

기능을 가지고 있다. 데이터글러브를 착용한 사람은 HUI를 통해서 자신의 핸드제스처를 정의하고 분류하도록 할 수 있다.

4. 핸드유저인터페이스의 구현 및 검증

CyberGlove(R)는 그것을 착용한 사람의 손 동작 정보를 6자유도 데이터로 획득할 수 있게 해주는 장치이다. 이러한 CyberGlove(R)는 Fig. 7에 도시하였다. Fig. 8의 (a)는 벤딩모션을 감지하는 플렉스 센서가 위치하는 부분이다. 22개의 플렉스 센서가 개개의 손가락 및 각각의 마디, 그리고 손등 부위에 부착되어 있으며 손의 움직임에 민감하게 반응한다. 플렉스 센서를 이용하면, 손등 및 개개의 손가락 마디들의 수신기(c)에 대한 상대좌표를 얻을 수 있다. (b)는 전자기파 신호를 송출하는 송신기이다. (c)는 송신기(b)에서 송출되는 전자기파 신호를 이용하여 송신기(b)를 원점으로 한 절대좌표를 추적하도록 한다. (a), (b), (c)의 조합에 의해서 손의 모든 부위의 위치는 (b)를 원점으로 한 절대좌표로 계산된다.

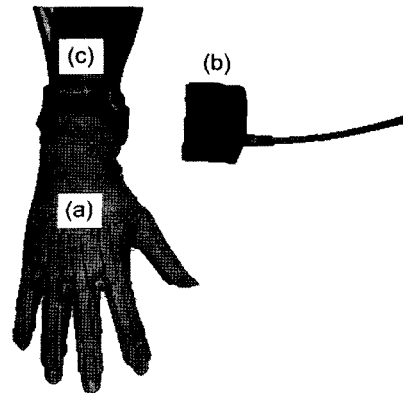


Fig. 7. (a)손의 벤딩 모션을 감지부, (b)손의 6자유도 트래킹을 위한 전자기파 송신기, (c)전자기와 송신기(b)에서 발생하는 신호를 수신하는 수신기. 수신된 신호를 이용하여 손목의 6자유도 좌표를 계산하게 된다.

4.1 가상 손의 재구성

이 논문에서는 손 동작을 의사데이터의 관점으로 이용하고 있지만, 재구성된 손 동작을 스크린을 통해 보임으로서 상호작용의 편의를 돕고자 한다.

HUI에서 처리하는 데이터는 16개의 4×4 행렬인데, 이는 사용자의 손 모양을 재구성하기 위한 것이다. HUI를 이용하는 사용자는 자신의 손 모양을 흉내내는 재구성된 가상 손(Reconstructed Virtual Hand)

을 보면서 SW와 상호작용할 수 있다. 이러한 4×4 행렬은 Fig. 8(a)에서 보인 바와 같이 플렉스 센서의 벤딩 모션 값과 수신기에 의해서 생성된 절대좌표에 의해서 계산한다. 플렉스 센서는 Fig. 8(b)에서 보인 바와 같이 손가락 마디와 마디 사이의 각도를 이용하여 상대 좌표를 생성할 수 있게 한다. 손의 개개의 관절에 대한 벤딩 모션을 감지 함으로서 각 관절에 대한 상대적인 위치를 측정한다.

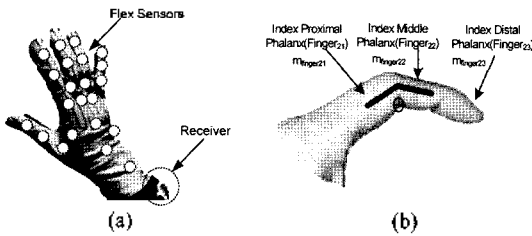


Fig. 8. (a) 각 센서의 위치, (b) 벤딩 모션의 검출.

Fig. 8(b)에서는 Finger₂₁와 Finger₂₂에 대한 벤딩 모션 측정을 도시하였다. Finger₂₁과 Finger₂₂는 각각 손의 “Index Proximal Phalanx”와 “Index Middle Phalanx”에 해당하는 부위이다. 측정된 각도를 이용하여 Finger₂₂의 Finger₂₁에 대한 상대적인 운동을 감지하고 이를 Finger₂₂의 Finger₂₁에 대한 4×4 행렬로 구성한다. 손의 각 부분은 손목, 손바닥, 엄지, 검지, 중지, 약지, 그리고 소지로 구분될 수 있다. 이들 각 부분은 Wrist, Palm, Thumb_k, Finger_{1k}, Finger_{2k}, Finger_{3k} 그리고 Finger_{4k}로 명명한다. 여기서 인덱스는 k=1, 2, 3의 값을 가지는 정수인데 손바닥에서 가까울수록 작은 인덱스를 갖는다. 여기서 M은 손의 각 부분에 대한 ATM(Absolute Transformation Matrix)이고, m은 손의 각 부분에 대한 RTM(Relative Transformation Matrix)이다. 예를 들어 Fig. 8(b)의 m_{Finger23}은 m_{Finger22}에 대한 상대적인 모션을 표현하는 RTM이고, m_{Finger22}은 m_{Finger21}에 대한 상대적인 모션을 표현하는 RTM이다. m_{palm}은 m_{wrist}에 대한 RTM이지만, 예외적으로 m_{wrist}은 손목에 부착된 6자유도 수신기로부터 얻어진 ATM이다. 손의 각 부분의 ATM을 얻기 위해서는, 예를 들어 다음과 같은 행렬 곱셈연산이 필요하다.

$$M_{thumb1} = m_{wrist} \cdot m_{palm} \cdot m_{thumb1}$$

Thumb₁의 ATM을 얻기 위해서는 m_{wrist}, m_{palm} 그리고 m_{thumb1}의 행렬곱 연산이 필요하다. ATM들을 얻는 방법을 다음과 같이 정리하였다.

| Joint | Matrix combination |
|----------------------|--|
| Wrist | $M_{wrist} = m_{wrist}$ |
| Palm | $M_{palm} = m_{wrist} \cdot m_{palm}$ |
| Thumb ₁ | $M_{thumb1} = m_{wrist} \cdot m_{palm} \cdot m_{thumb1}$ |
| Thumb ₂ | $M_{thumb2} = m_{wrist} \cdot m_{palm} \cdot m_{thumb1} \cdot m_{thumb2}$ |
| Thumb ₃ | $M_{thumb3} = m_{wrist} \cdot m_{palm} \cdot m_{thumb1} \cdot m_{thumb2} \cdot m_{thumb3}$ |
| Finger ₂₁ | $M_{finger21} = m_{wrist} \cdot m_{palm} \cdot m_{finger21}$ |
| Finger ₂₂ | $M_{finger22} = m_{wrist} \cdot m_{palm} \cdot m_{finger21} \cdot m_{finger22}$ |
| Finger ₂₃ | $M_{finger23} = m_{wrist} \cdot m_{palm} \cdot m_{finger21} \cdot m_{finger22} \cdot m_{finger23}$ |
| ... | ... |

손의 나머지 부분에 대한 ATM도 동일한 방법을 이용하여 생성할 수 있다. 이러한 방법으로 계산된 ATM들을 이용하여 CyberGlove(R)을 착용한 사람의 손을 재구성할 수 있다. 손의 각 부위에 대한 3차원 위치정보는 4×4 행렬에 의해서 완전히 표현될 수 있기 때문이다. 손의 각 부분들을 표현하는 17개의 ATM들은 손의 상태 그 자체를 의미한다. 이들 ATM들은 손이 현재 어떤 위치에서 어떤 제스처를 가지고 있는지 알 수 있는 데이터이기 때문에 제스처 인식을 위한 데이터로도 이용할 것이다.

Fig. 9(a)는 OpenGL™로 표현한 재구성된 손으로 각 부위의 RTM들이 표시되어 있다. Fig. 9(b)는 Cyber Glove(R)를 착용한 사람의 손 데이터를 획득하고 이를 갱신하여 실시간으로 손동작을 흉내 내도록 작성한 프로그램이다. 3차원으로 모델링 된 손의 각 블록들은 핸드데이터로부터 가공한 ATM들을 이용한다.

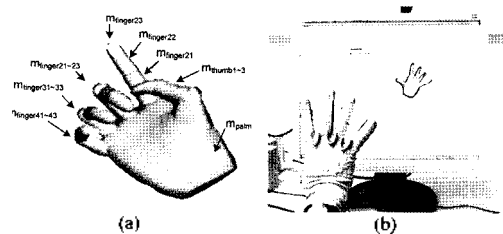


Fig. 9. (a) OpenGL™을 이용한 RVH 및 손의 각 부위에 대한 RTM들, (b) CyberGlove(R)과 이를 흉내 내는 RVH의 모습.

ATM들을 이용하여 생성한 RVH는 3D 데이터와 상호작용할 때 가시화되며, 장갑을 착용한 손의 움직임을 잘 표현해 줄 수 있음을 확인하였다.

부가사항으로는, 손가락들의 현재 상태를 표현하는 ATM들은 3.2절에서 설명한 SVM 뉴럴네트워크의 입력으로 사용하기 위해서 하나의 1차원 벡터로 변형된

다. 예를 들어 1개의 손가락이 제스처의 인식을 위해 이용된다면 3개의 4×4 행렬이 사용될 것이고 48개의 배정밀도 실수로 구성된 벡터 데이터(×)로 변형된다.

4.2 제스처들의 기능 정의 및 제스처 인식

ATM들이 가지고 있는 배정도 형의 실수 데이터들은 HUI 내부의 SVM 함수들에 의해 이용된다. 이 논문에서는 제스처의 인식을 위해 엄지, 검지 그리고 중지의 3개 손가락에 대한 데이터를 이용한다. 3개의 손가락에 대한 제스처 정보와 손의 3차원 위치 추적 정보를 이용하여 3D데이터의 6DOF 제어 및 윈도우의 2D제어를 충분히 수행할 수 있기 때문이다. 3개의 손가락으로부터 만들어진 9개의 4×4 행렬은 144개의 실수 데이터를 가지며, 이 데이터들은 HUI의 SVM 함수들에 의해서 이용된다. HUI가 손 동작을 이용하기 위해 내부적으로 실행하는 SVM관련 함수는 SVM_Learn()과 SVM_Test()이다.

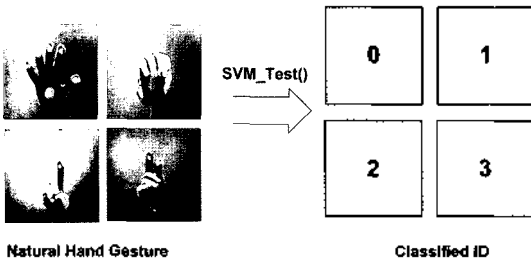


Fig. 10. 핸드 제스처 및 분류 수치.

SVM_Learn()은 손 데이터들을 획득하고 이를 바탕으로 학습집합을 생성 및 학습을 진행한다. 학습에 필요한 시간은 Pentium(R) 4 3.2GHz PC에서 5초 정도가 소요되었다. 이것은 144개의 특징과 4종류의 제스처를 이용하여 학습을 진행한 결과이다. SVM_Test()는 학습 이후에 주어지는 데이터들을 분류하는 기능을 수행한다. Fig. 10에는 4개의 제스처와 각각 부여된 분류-아이디를 도시하였다. 각각의 제스처들은 유용한 기능을 가질 수 있도록 프로그램이 가능하다 여기서서는 제스처들이 다음 테이블과 같은 기능을 갖도록 정의할 것이다.

| ID | 기능 | 설명 |
|----|----------|------------------------------|
| 0 | Normal | - |
| 1 | Grasp | 3D데이터의 6-DOF제어, 윈도우의 2D 위치제어 |
| 2 | Exchange | 2D, 3D 데이터의 순환선택 |
| 3 | Resize | 윈도우의 크기조정 |

“Normal”에서는 특별한 기능이 지정되지 않는다. “Grasp”에서는 3D 데이터에 대한 6자유도 제어 및 그래픽데이터 윈도우의 2D 위치 제어를 위해서 사용한다. CyberGlove(R)를 착용한 사람은 손 모양을 Classified ID 1과 같이 취한 후에 3D 데이터 윈도우 내에서는 3D 데이터를 제어하고, 윈도우의 외부에서는 그래픽데이터 윈도우의 크기를 조절할 수 있다. “Exchange”에서는 현재 선택하고자 하는 3D 데이터들 혹은 그래픽데이터 윈도우들에 대한 선택을 약간의 시간차를 가지고 순환하도록 하는 기능이다. “Exchange”에 대해서는 5장의 초반부에 관련 사항이 기술되어 있다. “Resize”에서는 그래픽스 데이터 윈도우의 크기를 조정하기 위한 기능이다. 윈도우의 위, 아래, 왼쪽 그리고 오른쪽 경계근방에서 Classified ID 3의 동작을 취한 후에 손을 이동시키면, 그 크기가 조정되도록 프로그램되었다.

3D 데이터의 제어를 위한 “Grasp” 기능에 대한 알고리즘은 Fig. 11과 같이 구현된다. 손 상태의 검출은 SVM_Test() 함수의 실행결과로 얻어진 ID 값의 Queuing에 의해서 구현된다. 큐에 삽입된 2개의 ID 값들을 검출함에 의해서 GRASP-START, GRASP-ING, GRASP-END와 같은 상태들이 결정되는 것이다. 예를 들어, 큐의 상태가 {ID_{normal}, ID_{grasp}}이라면 손 상태는 GRASP-START로 할당하고, 큐의 상태가 {ID_{grasp}, ID_{grasp}}이라면, GRASP-ING로 할당한다. 큐의 상태가 {ID_{grasp}, ID_{normal}}이라면, GRASP-END를

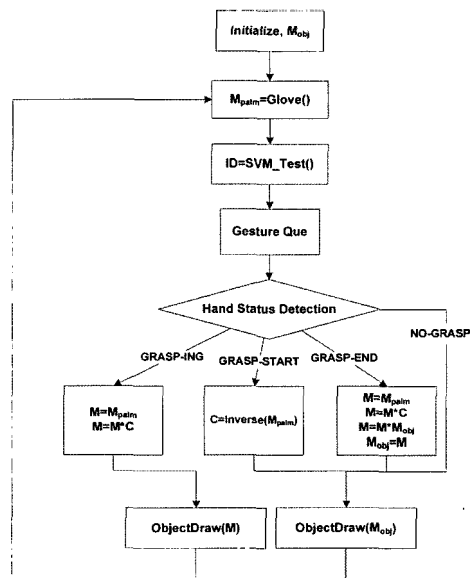


Fig. 11. 3D 모델의 6DOF 제어 알고리즘.

손의 상태로 할당한다. NO-GRASP에서는 3D 물체의 초기 위치가 유지된다. 사용자가 Grasp 제스처를 취하면, GRASP-START 상태가 만들어진다. 여기서는 M_{palm} 의 역행렬을 변수C에 할당한다. 그 다음에는 GRASP-ING 상태가 만들어지고, $M=M_{palm}$, 그리고 $M=M*C$ 연산을 차례로 수행한다. 이 연산에 의해서 3D물체의 위치는 M_{palm} 을 따른 것이다.

사용자가 GRASP 제스처를 해제했을 경우에는 GRASP-END 상태가 만들어지고, M_{obj} 를 업데이트시킨다. 그 이후에는 NO-GRASP 상태가 만들어지고, 3D 물체는 제어가 해제된다. 이 알고리즘과 사용자의 손을 흉내 내는 RVH에 의해서 글러브를 착용한 사람은 마치 3D 데이터가 자신의 손에 의해서 제어되는 듯한 느낌을 받게 된다.

4.3 제한된 방법을 이용한 인터랙션의 검증

Fig. 12는 이러한 방식으로 물체에 대한 접촉판별을 수행하는 것이다. RVH의 중지 끝 한 점이 물체의 경계구에 포함되었을 때 물체와의 인터랙션을 시작할 수 있다. (b)는 인터랙션이 가능한 상태임을 사용자에게 알리기 위해서 RVH가 하이라이트 된 것이다. 하이라이트 된 상태에 “Grasp” 동작을 취하여 선택된 물체를 6 DOF로 제어할 수 있다.

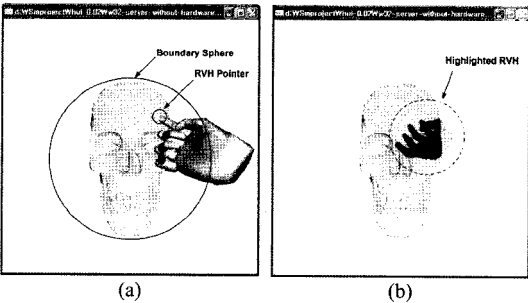


Fig. 12. (a)경계구를 이용한 접촉판별, (b)검지의 끝이 경계구에 들어갈 경우는 접촉으로 해석하고, RVH를 하이라이트 한다.

4.4 인터페이스의 실행 성능

앞서 제시했던 두 개의 방법, 경계구에 기반한 접촉판별과 제스처 인지에 기반한 가상환경과의 상호작용이 그렇지 않은 경우에 대해서 얼마나 CPU 자원을 절약할 수 있는지 알아보았다. 그러기 위해서 CyberGlove(R)에서 제공하는 VHT(Virtual Hand Toolkit)^[15]을 이용하여 동일한 기능을 하는 소프트웨어를 구현하였다. Fig. 13에서는 실시간으로 측정되는 자원소비량을 보였다.

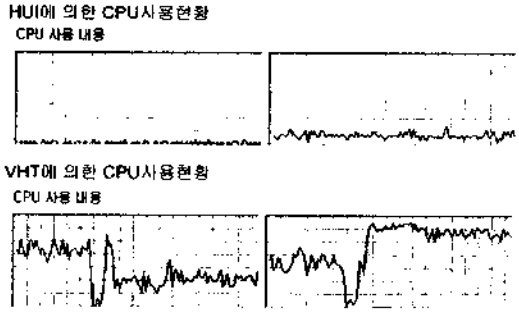


Fig. 13. CPU자원 소비량비교, 위쪽은 HUI에 의한 것이고 아래쪽은 VHT에 의한 것이다.

VHT를 이용한 경우는, 3D 모델과의 상호작용에 28%~42%(평균 34%) 정도의 CPU 사용률을 보였으며 본 논문에서 제시한 방식을 이용했을 경우에는 2%~5%(평균 3%)의 CPU 사용률을 보였다. VHT를 이용하여 3D 모델과의 상호작용을 수행 시에는 내부적으로 충돌검사엔진이 실행되는데, 이것은 계산기하에 기반한 방식이기 때문에 CPU 자원을 상대적으로 많이 차지한다. 텔레컨퍼런스 시스템이 네트워크에 의한 통신을 기반으로 하기 때문에 사용하는 컴퓨터가 내부적으로 겪는 네트워크 처리 이벤트들로 인해 과부하를 받기 쉬운 점을 고려한다면 HUI가 좀 더 쉽게 SMS에 쉽게 적용될 수 있다. 이 결과는 HUI가 좀 더 나은 성능을 보인다는 것을 말하는 것은 아니다. 손 데이터를 보는 “의사데이터”의 관점이 상대적으로 낮은 CPU 점유율을 달성할 수 있으며 텔레컨퍼런스 응용에서는 좀 더 효율적일 수 있음을 보이는 것이다.

5. SMS로의 적용 및 검증

5.1 SMS에 적용한 핸드유저인터페이스

HUI의 하드웨어는 HUI 서버와 HUI 클라이언트로 구성된다. HUI 서버는 Fig. 14과 같이 사용자의 손에 장착된 CyberGlove(R)와 네트워크에 연결되어 있는 컴퓨터로 구성되어 있다. 핸드데이터는 CyberGlove(R)에 의해서 획득된 플렉스 센서 값과 수신기의 절대좌표 값을 이용하여 계산된 손가락 각각의 부위에 대한 16개의 4x4 행렬 데이터들과 1개의 제스처 분류 아이다로 이루어진다. 핸드데이터는 네트워크를 통해서 멀티캐스팅 방식으로 다른 HUI 클라이언트로 전송된다. 특별히, 멀티캐스팅 방식을 사용한 이유는 핸드데이터가 하나 이상의 HUI 클라이언트로 전송되는 상황을 고려한 것이다.

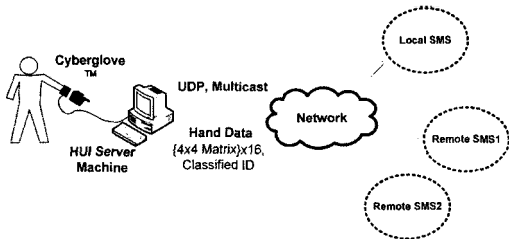


Fig. 14. HUI 서버 머신.

HUI 서버 머신에서 송신한 핸드데이터는 그 HUI 서버 머신이 위치한 로컬SMS에서 이용하지만, 물리적으로 떨어져 있는 리모트SMS로도 전송이 된다. 동일한 멀티캐스팅 IP 주소 및 포트번호를 가진 리모트 SMS는 이러한 핸드데이터를 공유할 수 있다. HUI 클라이언트가 전송 받은 핸드 데이터는 3D데이터와 윈도우의 제어를 위해 이용된다. 관련 사항을 Fig. 15에서 보였다. HUI 클라이언트 머신에서 실행되는 SAGE와 fsManager 및 fsConsole은 그래픽스 데이터 윈도우를 생성하고 디스플레이 머신으로 분배하기 위해서 실행되는 프로그램이다¹⁾. 이 프로그램들을 이용하여 SW에서 표현되는 그래픽스데이터의 로딩, 언로딩, 그리고 위치 및 크기의 제어를 수행할 수 있다.

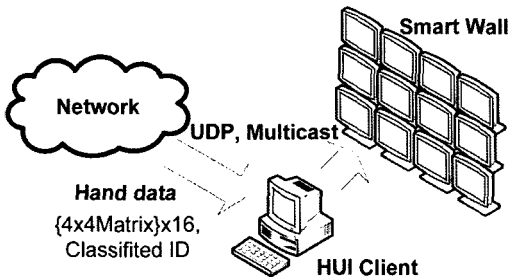


Fig. 15. HUI 클라이언트 머신.

이 프로그램들은 기본적으로 커맨드라인 인터페이스를 가지고 있지만, HUI 모드로 실행할 수 있는 기능이 덧붙여 졌다. 네트워크를 통해서 전달되는 4x4 행렬 형태의 데이터로부터 그래픽스 데이터 윈도우의 제어를 위해 (x, y) 좌표가 추출된다. 추출된 좌표는 분류 아이디(Classified ID)와 함께 SW에 디스플레이 된 윈도우들을 제어하는데 이용한다.

데이터글러브로부터 추출된 (x, y) 좌표는 SW에서의 한 점을 표현하기 위해 스케일링 된다. 이것은 기본적으로 마우스 포인터의 역할을 수행하지만, 3D 데이터 윈도우 내부로 진입하게 되면, 6 DOF 제어모드로 전환된다.

5.2 핸드유저인터페이스의 작동 검증

HUI라 명명된 핸드제스처에 의한 유저 인터페이스는, GIST(Gwangju Institute of Science and Technology) 정보기전공학부 내에 설치된 SMS의 일부 기능으로서, 현재 시험 중에 있는 기능이다. Fig. 16은 HUI를 이용하여 SMS에서 표현되는 그래픽데이터를 제어하기 위한 데모환경이다. 이 환경을 구축하기 위해서 24개의 LCD 모니터들과 12개의 컴퓨터들이 이용되었다. 이들 하드웨어는 6x4의 타일드 디스플레이를 구성하는데 이용되었는데, 구성된 타일드 디스플레이는 9600x4800의 해상도를 가진다.

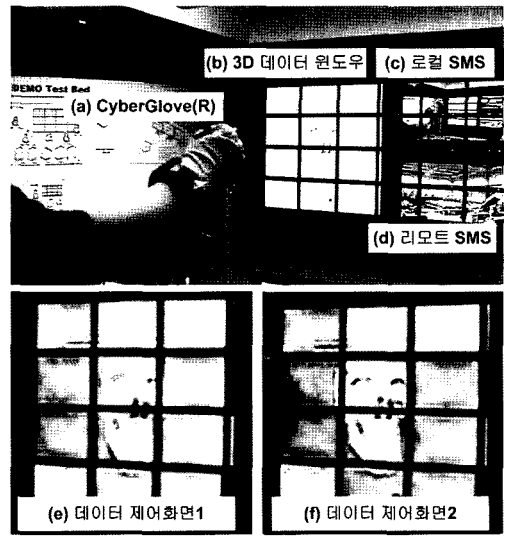


Fig. 16. SMS에서의 HUI 사용.

Fig. 16은 HUI를 SMS에 적용하고, 시연하는 사진이다. (a)는 데이터글러브의 일종인 CyberGlove(R)이고, (b)는 3D 데이터 조작 데모를 위해 OpenGL™을 이용하여 제작한 테스트 프로그램이다. (c)는 비디오 카메라를 이용하여 로컬SMS의 영상을 디스플레이 한 것이며, (d)는 리모트에 위치한 SMS의 상황을 디스플레이 한 것이다. (e)는 3차원 의료모델 데이터를 가시화하고, 데이터글러브를 이용하여 6DOF로 제어를 수행하는 모습이다. (f)는 참여자가 자신의 손을 이용하여 의료데이터(e)를 약간 회전시킨 모습이다. 리모트에 위치한 동료들은 Fig. 16에서 표현된 가시화 데이터들을 공유하고 있다. 상호 협력하는 원격 및 리모트의 동료들은 SMS를 통해서 전체적인 회의 진행 모습을 실시간으로 확인할 수 있고, 공유되는 데이터들에 접근 및 제어할 수 있었다. 무엇보다도, 참여자 자신의 손으로 제어되는 데이터들을 이용하여 보다 집중

도 높은 회의를 진행할 수 있었다.

6. 결론 및 향후 과제

본 논문에서는 텔레컨퍼런스 시스템을 위한 핸드 유저 인터페이스의 기본골격을 제시하였다. HUI는 SW에 가시화된 데이터를 제어하기 위한 일관성있는 방법을 제공하기 위해서, 단순화한 충돌검사 모델과 제스처 인지를 이용한다. 그로 인해 인터페이스에 필요한 컴퓨팅 자원을 최소화시킬 수 있었고, 잉여 컴퓨팅 자원은 사실적인 데이터의 표현과 영상 및 음성을 실시간으로 전송하는데 이용될 수 있었다. HUI는 참여자들의 손을 이용하여 가시화된 데이터를 조작할 수 있는 방법을 제공해 주었다. 그로 인해 텔레컨퍼런스 환경에서의 편의성 및 회의 진행의 효율성 향상에 기여하였다.

이와 관련한 향후 과제로는, 제스처와 모션을 함께 고려한 인지방법의 개발이다. 현재까지는 핸드제스처라는 정지 동작만을 인지하고 이를 특정 명령으로 매핑 하는 것이지만, 보다 자연스러운 방식은 손의 움직임도 함께 고려하는 것이다. 그러면 좀더 다양한 제스처를 정의할 수 있기 때문에 SMS 환경에서의 데이터 조작성을 향상시킬 수 있을 것이다.

감사의 글

본 연구는 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅 및 네트워크원천기반기술개발사업의 지원에 의해 수행되었습니다.

참고문헌

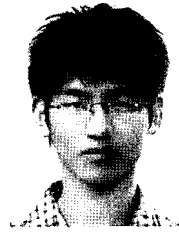
1. Wise, Samuel A., Rosen, Joseph, Fisher, Scott, Glass, Karen, Wong, Yunko, "Initial Experience with the Data Glove, a Semi-automated System for Quantification of Hand Function", *RESNA '87, Meet the Challenge. Proc. of the 10th Annu Conf on Rehabil Technol.*, pp. 259-260, 1987.
2. Kim, Y. S., Soh, B. S. and Lee, S.-G., "A New Wearable Input Device: SCURRY", *IEEE Transactions on Industrial Electronics*, Vol. 52, No. 6, pp. 1490-1499, 2005.
3. Weissmann, John, Salomon, Ralf, "Gesture Recognition for Virtual Reality Applications Using Data Gloves and Neural Networks", *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, pp. 2043-2046, 1999.

4. Wang, H., Leu, M.C., and Oz, C., "American Sign Language Recognition Using Multi-dimensional Hidden Markov Models", *Journal of Information Science and Engineering*, Vol. 22, No. 5, pp. 1109-1123, 2006.
5. Wan, H., Luo, Y., Gao, S. and Peng, Q., "Realistic Virtual Hand Modeling with Applications for Virtual Grasping", *Proceedings VRCAL 2004 - ACM SIG-GRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 81-87, 2004.
6. Renambot, L., Rao, A., Singh, R., Jeong, B., Krishnaprasad, N., Vishwanath, V., Chandrasekhar, V., Schwarz, N., Spale, A., Zhang, C., Goldman, G., Leigh, J. and Johnson, A., "SAGE: The Scalable Adaptive Graphics Environment", *Proceedings of WACE 2004*, 2004.
7. 최기호, 김재윤, Dao Le Hai, 김종원, "다자간 협업 환경에서 다중 고해상도 비디오와 그래픽스 가시화를 위한 Tiled-display 시스템 구현", *한국통신학회 (KICS) 하계종합학술발표회*, Vol. 33, pp. 282-285, 2006.
8. Toomey, Lori, Adams, Lia, Churchill, Elizabeth, "Meetings in a Virtual Space: Creating a Digital Document", *Proceedings of the Hawaii International Conference on System Sciences*, Vol. 2, pp. 236-244, 1998.
9. Zhang, L., Shi, Y., Chen, J., "Drag and Drop by Laser Pointer: Seamless Interaction with Multiple Large Displays", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4159 LNCS, pp. 12-20, 2006.
10. Bi, X., Shi, Y. and Chen, X., "uPen: A Smart Pen-like Device for Facilitating Interaction on Large Displays", *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, TABLETOP'06, art. no. 1579208, pp. 160-166, 2006.
11. Yoshimi, B. H. and Pingali, G. S., "A Multimodal Speaker Detection and Tracking System for Teleconferencing", *Proceedings of the ACM International Multimedia Conference and Exhibition*, pp. 427-428, 2002.
12. Osowski, S., Siwek, K. and Markiewicz, T., "MLP and SVM Networks - A Comparative Study", *Report - Helsinki University of Technology, Signal Processing Laboratory*, No. 46, pp. 37-40, 2004.
13. Collobert, R. and Bengio, S., "SVM-Torch: Support Vector Machines for Large-scale Regression Problems", *Journal of Machine Learning Research*, Vol. 1, No. 2, pp. 143-160, 2001.
14. Yu Hen Hu and Jenq-Neng Hwang, "Handbook of Neural Network Signal Processing", CRC Press, 2002.
15. "Virtual Hand[®] User's Guide V2.7", Immersion Corp, 2001.



송 제 훈

2005년 단국대학교 전자공학과 학사
2007년 광주과학기술원 기전공학과 석사
관심분야: Autofocus of Optical Systems,
HCI, Scientific Computation



최 기 호

2006년 전남대학교 전자공학과 학사
2007년 현재 광주과학기술원 정보통신공학과 석사과정
관심분야: Advanced Collaborative Environment Supporting High-quality Video Transmission over IP Networks, Distributed Computing and Networking, and P2P Networking



김 종 원

1987년 서울대학교 제어계측공학과 학사
1989년 서울대학교 제어계측공학과 석사
1994년 서울대학교 제어계측공학과 박사
1994년 3월~1999년 7월 공주대학교 전자공학과 조교수
1997년 8월~2001년 7월 University of Southern California 연구 조교수

1999년 12월~2000년 7월 Technology Consultant for VProtect Systems Inc.

2000년 7월~2001년 6월 Technology Consultant for Southern California Division of InterVideo Inc.

2001년 9월~현재 광주과학기술원 정보기전공학부 부교수
관심분야: Networked Media Systems and Protocols focusing "Reliable and Flexible Delivery for Integrated Media over Wired/Wireless Networks" (네트워크미디어: <http://nm.gist.ac.kr>)



이 용 구

1992년 서울대학교 기계설계학과 학사
1994년 서울대학교 기계설계학과 석사
1997년 서울대학교 기계설계학과 박사
1997년~2000년 삼성SDS 책임연구원
2000년~2003년 미국 상공부 산하 국립표준기술연구소 객원 연구원

2003년~현재 광주과학기술원 기전공학과 조교수
관심분야: Nanomanufacturing, Virtual Reality, Haptics