

# SpaceSensor: I-NEXT 를 위한 감정인터페이스\*

홍동표, 우운택  
광주과학기술원 정보통신공학과 U-VR 연구실  
{dhong, wwoo}@kjist.ac.kr

## SpaceSensor: Emotional Interface for I-NEXT

Dongpyo Hong, Woontack Woo  
KJIST U-VR Lab.

### 요약

본 논문에서는 3 차원 시각 기술에 기반한 실시간 사용자 감정 인터페이스를 제안한다. 미래형 컴퓨팅환경에서는 보다 자연스럽고 편안한 인간-컴퓨터 상호작용기술뿐만 아니라, 사용자의 의도나 감정정보를 네트워크를 통해서 공유할 수 있는 새로운 형태의 사용자 인터페이스 기술 개발이 필요하다. 제안된 3 차원 시각 기반 실시간 사용자 감정 인터페이스는 블루 스크린과 같은 특별한 장치없이 자연스러운 배경으로부터 사용자 정보만을 분리하는 사용자 분리 방법을 사용한다. 그리고, 사용자 분리 방법을 통해서 추출된 사용자 정보와 멀티-뷰 카메라를 통해서 획득된 깊이 맵을 이용하여 사용자의 동적인 제스처를 트래킹할 수 있는 공간센서를 구축한다. 구축된 공간센서를 통해서 사용자의 감정을 인식할 수 있는 요소(공간, 시간, 힘)들을 추출한 뒤, 추출된 요소들을 신경망을 이용하여 사용자의 감정을 인식한다. 제안된 방법은 기존의 2 차원 시각 기반 사용자 인터페이스의 한계를 극복하고, 실시간 사용자 제스처 트래킹을 통한 사용자의 감정 인식의 계산 복잡성을 해소할 수 있다.

Keyword : User Segmentation, SpaceSensor, Gesture Tracking, Emotion Recognition

### 1. 소개

미래형 컴퓨팅 환경은 언제, 어디서나, 누구와도 접속할 수 있는 "접속을 통한 정보공유의 시대"가 될 것이다[1][2]. 이처럼 새로운 패러다임이 요구되는 미래형 컴퓨팅환경에서는 보다 자연스럽고 편안한 인간-컴퓨터 상호작용기술뿐만 아니라, 사용자의 의도나 감정정보를 네트워크를 통해서 공유할 수 있는 사용자 인터페이스 기술 개발이 필요하다.

이러한 요구에도 불구하고 기존의 사용자 인터페이스에 대한 연구와 개발은 키보드와 마우스와 같이 2 차원적인 한계를 벗어나지 못하였다. 이러

한 기존의 2 차원적인 인터페이스는 컴퓨터와 상호작용을 하려는 3 차원 공간속의 사용자들에게는 부자연스러운 상호작용을 제공하는 단점이 있다. 뿐만 아니라, 사용자의 의도나 감정을 파악하는데도 어려움이 있다. 따라서 전통적인 사용자 인터페이스인 키보드, 마우스등을 대체하는 시각(제스처), 음성 및 센서등에 기반한 새로운 형태의 사용자 인터페이스들에 관한 연구가 진행되어왔다. 특히, 시각 기반 사용자 인터페이스의 장점은 음성이나 센서에 기반한 사용자 인터페이스들에 비해 비교적 보정과정이 용이 하며, 사용자와 시스템간의 자연스러운 상호작용뿐만 아니라 공간적

\* 본 연구는 ICU 디지털 미디어랩의 지원에 의해 수행됨

제약을 해소하는 장점이 있다.

일반적으로 시각 기반 사용자 인터페이스에 사용되는 방법들은 마커등을 부착하는 접촉식 방법과 배경 분리 기법등을 이용하는 비접촉식 방법으로 구분된다[3]. 마커등을 부착하는 접촉식 방법은 마커를 트래킹하므로서 원하는 정보만을 획득할 수 있는 잇점은 있으나, 트래킹하고자 하는 마커가 가려지거나 여러 개의 마커들을 동시에 트래킹하는데 어렵다는 단점이 있다[4]. 뿐만아니라, 접촉식 방법은 사용자에게 마커와 같은 것을 착용하게 하는 불편함을 제공한다. 하지만, 배경 분리 기법등을 활용한 비접촉식 방법은 앞서 지적한 접촉식 방법이 갖는 한계들을 극복하는 장점이 있다. 하지만, 이러한 장점에도 불구하고 비접촉식 시각 기반의 사용자 인터페이스는 광원에의한 간섭에는 여전히 민감하다. 이러한 문제점을 해결하고자 많은 연구가 진행되고 있다[5][6][7][8]. 그리고 비접촉식 시각기반 사용자 인터페이스에서 사용자의 3차원적인 움직임 정보를 트래킹하여 사용자의 의도나 감정정보를 인식하기 위해서 실시간성이 결여된 복잡한 알고리즘을 사용하였다[9][10][11][12].

본 논문에서는 3 차원 시각 기술에 기반한 실시간 사용자 감정 인터페이스를 위한 사용자 감정 인식 방법을 제안한다. 제안된 3 차원 시각 기반 실시간 사용자 감정 인식 방법은 블루 스크린과 같은 특별한 장치없이 자연스러운 배경으로부터 사용자 정보만을 분리하는 사용자 분리 방법을 사용한다[13]. 그리고, 사용자 분리 방법을 통해서 추출된 사용자 정보와 멀티-뷰 카메라를 통해서 획득된 깊이 맵을 이용하여 사용자의 동적인 제스처를 트래킹할 수 있는 공간센서를 구축한다[14]. 구축된 공간센서를 통해서 사용자의 감정을 인식할 수 있는 요소(힘, 공간, 시간)를 추출한 뒤, 추출된 요소들을 신경망을 이용하여 사용자의 감정을 인식한다[15]. 제안된 방법은 2 차원 시각 기술의 한계를 극복하고, 실시간 사용자 제스처 트래킹의 계산 복잡성과 사용자 감정 인식의 어려움을 해소할 수 있다.

이와 같은 필요성을 실험하기위해서, 우리는 Network 환경에서 3 차원 시각정보를 사용하여 사

용자의 문화적 감정을 상호작용적으로 표현할 수 있는 "I-NEXT: An Interactive Networked Expression Experience Testbed" 를 구축하였다[16]. 구축된 "I-NEXT"는 개인화된 감정 인터페이스를 활용하여 구축된 가상공간에서 네트워크를 통해서 다자간의 상호작용이 가능한 시스템이다.

본 논문의 구성은 다음과 같다. 2 장에서는 제안된 사용자 분리 알고리즘, 공간센서 디자인, 제스처 트래킹, 그리고 사용자의 감정 인식 방법에 대해서 자세히 소개한다. 3 장에서는 제안된 사용자 감정 인식 방법을 통한 실험 결과를 기술하고, 4 장에서는 제안된 방법에 대한 결론과 향후 과제에 대해서 언급한다

## 2. SpaceSensor: 3 차원 시각기반 실시간 사용자 감정 인터페이스

### 2.1 사용자 분리

제안된 3 차원 시각기반 사용자 감정 인터페이스인 공간센서는 그림자와 같은 광원의 간섭에 의한 정확한 사용자 분리의 어려움을 극복하는 사용자 분리 알고리즘을 사용한다[13]. 또한, 사용된 사용자 분리 알고리즘은 블루 스크린과 같은 특별한 장치없이 일반적인 배경으로부터 사용자만을 분리할 수 있다. 그림 1 은 제안된 공간센서에서 사용하는 배경분리 기법에 대한 알고리즘이다.

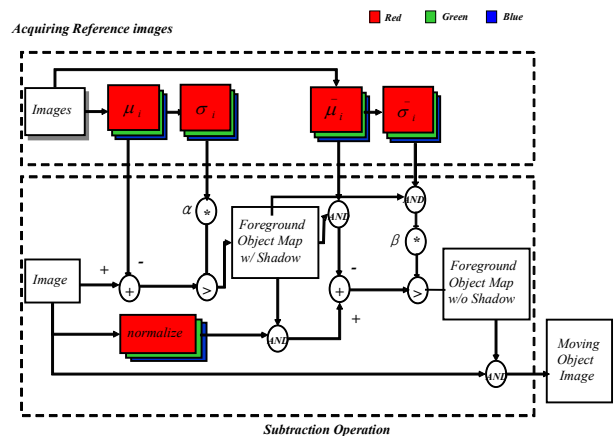


그림 1. 제안된 사용자 분리 알고리즘

## 2.2 공간센서 디자인과 제스처 트래킹

제안된 공간센서는 배경으로부터 분리된 사용자의 3 차원 정보를 이용하여 사용자에게 자연스러운 움직임을 제공하면서도 정확한 사용자의 제스처를 트래킹하도록 디자인되었다[14]. 제안된 공간센서의 디자인은 공간센서를 사용자를 중심으로 동적으로 위치시킴으로서 사용자에게 자연스러운 움직임을 제공한다. 그리고 사용자를 중심으로 사용자의 행동반경 내에 공간센서를 위치시킴으로서 정확하게 사용자의 제스처를 감지한다. 정확한 사용자의 제스처 감지와 사용자 제스처의 3 차원 위치 정보를 활용하기 위해서 사용자의 행동 반경을 기준으로 모든 공간센서가 비교적 동일한 공간을 차지하도록 공간을 그림 2 처럼 분할한다.

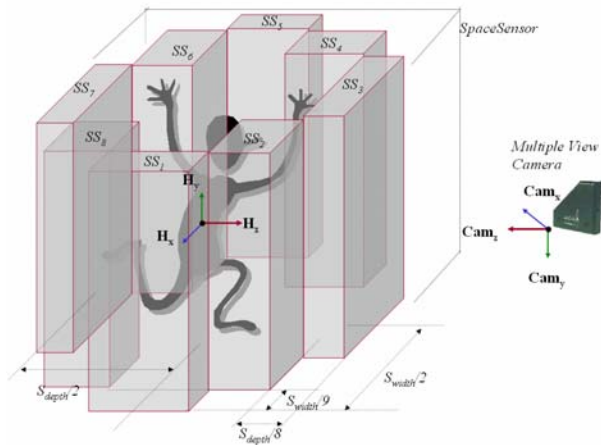


그림 2. 제안된 공간센서들의 위치

그림 2 에서 보여진 것처럼, 제안된 공간센서는 사용자의 제스처를 보다 쉽게 하기위해서 공간센서를 8 개의 영역으로 나누고, 각 영역(SS<sub>0</sub>,...,SS<sub>7</sub>)에 상태를 갖도록 한다. 따라서, 공간센서의 상태(SS)는  $SS = \{SS_0, \dots, SS_7\}$  와 같이 표현할 수 있다. 만약 사용자가 공간센서의 특정 영역을 터치하게되면, 터치된 영역의 공간센서의 상태는 1 이 되고, 그렇지 않은 영역의 공간센서의 상태는 0 이 된다. 따라서, 제안된 공간센서는 기존의 시각기반 트래킹 기술과는 달리 사용자의 제스처를 트래킹하기위해서 공간센서의 상태가 1 인 것만을 추출한 뒤, 추출된 공간센서의 영역에서 사용자가 터치한 지점의 3 차원 좌표를 그림 3 과 같이 구한다.

제안된 공간센서는 사용자의 상,하,좌,우 공간뿐

만 아니라 전,후를 사용할 수 있기때문에 기존의 2 차원 영상에 기반한 시각기반 시스템들이 갖는 한계를 극복하고 보다 다양한 사용자의 제스처를 트래킹할 수 있다. 또한, 제안된 공간센서는 사용자의 제스처를 쉽게 트래킹하기위해서 3 차원 박스에 기반한 센서를 이용하였다. 사용자 주변에 3 차원 박스를 많이 위치시키면 사용자의 움직임을 보다 정확하게 트래킹할 수 있지만, 계산 복잡성은 증가하게된다. 그리고, 공간센서의 상태 정보를 이용하여 부가적인 정보도 추출할 수 있다. 예를 들면, 현재 사용자의 움직임 정도(속도), 개인 공간 활용 정도(넓게 혹은 좁게), 그리고 움직임의 무게감 (가속도<sup>1</sup>) 등을 추출할 수 있다. 따라서, 제안한 공간센서는 사용자로부터 복잡한 센서나 기기를 제거할 뿐만아니라 새로운 형태의 사용자 인터페이스로 활용 가능하다.

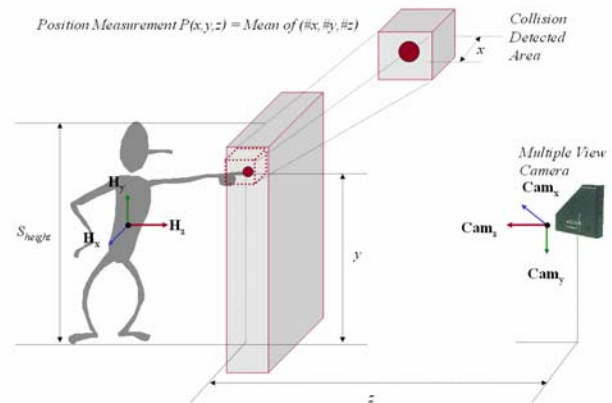


그림 3. 제안된 공간센서에서의 제스처 트래킹

## 2.3 사용자의 제스처를 통한 감정 인식

기존의 시각 기반 사용자 감정 인식 기법들은 2 차원 영상 기반의 기법들이었다 [9]-[12]. 2 차원 영상에 기반한 사용자 감정 인식 기법의 취약점은 위치나 거리에 따라 같은 동작으로 인식되기 쉽다는 단점이 있다. 뿐만아니라, 위치에 강건한 특징점을 추출할 수 있다고 하더라도 2 차원 영상의 제약때문에 만족할 만한 인식율을 갖기는 어렵다.

제안된 공간센서는 루돌프 라반이 제안한 움직

<sup>1</sup> 뉴턴의 운동 제 2 법칙  $F = ma$  에 따르면, 힘은 가속도와 비례한다.

임 이론(Effort Theory)에 기반하여 사용자의 제스처를 정량적으로 분석하여 제스처의 기본 요소인 시간(Time), 공간(Space), 무게(Weight)로 사용자의 제스처를 추출한다[15]. 그림 4 는 라반이 제안한 움직임의 요소를 조합하여 8 가지 기본 행동과 이에 해당하는 사용자의 감정 맵핑을 나타내는 그림이다.

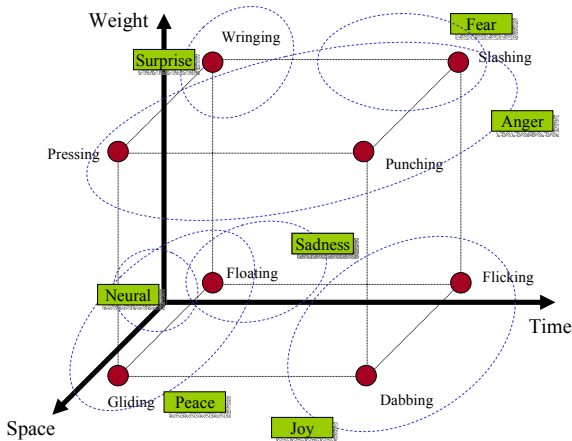


그림 4. 사용자 제스처의 8가지 기본 동작과 감정 맵핑

그림 4 에서 보여진 것처럼, 사용자의 감정은 8 가지의 기본동작의 조합으로 표현된다. 즉, 사용자는 자신의 연속적인 동작을 통해서 자신의 감정을 표현다. 예를 들면, 현대 무용에 있어서 무용수의 제스처는 7 가지의 의미로 분석되기도 한다[17]. 따라서, 본 논문에서는 표 1 과 같이 사용자의 감정을 사용자의 8 가지 기본 동작으로 구분하였다.

표 1. 8가지 기본동작에 따른 사용자 감정 구분

감정	동작
Peace/Love	Floating, Gliding
Joy/Happiness	Gliding, Dabbling, Flicking
Surprise	Wringing
Anger	Pressing, Punching, Slashing
Sadness	Floating
Fear	Wringing, Slashing
Neural	-

본 논문에서는 사용자의 제스처에 기반한 감정 분석을 위해서 시간, 공간, 무게를 다음과 같이 정의하고 사용자 움직임의 요소로서 추출한다.

공간(Space): 개인적 공간의 크기

시간(Time): 주어진 시간에 터치된 박스들의 개수

무게(Weight): 주어진 시간에 터치된 서로 다른 박스들의 개수

기존의 시각기반 사용자 감정 혹은 제스처 인식 기법에서는 사용자 특징점을 트래킹하고 분석해야 어려운 점이 있다. 하지만, 제안된 공간센서는 위에서 정의된 사용자의 움직임 요소들을 간단히 터치된 공간센서만을 트래킹하여 추출 할 수 있다.

그림 5 는 신경회로망을 이용한 감정분석의 기본 구조를 나타낸다. 사용자의 동적인 제스처를 분석하기 위해서는 현재의 동작 정보뿐만 아니라, 이전의 움직임 정보(어떻게 현재 위치에 도달하게 되었지를 추정할 수 있는 정보)도 고려해야 한다. 제안된 신경회로망에서의 입력단은 앞에서 추출한 요소들이다. 그리고, 입력단에서 사용된 시간 딜레이는 앞서 언급한 사용자의 이전 움직임과 관련된 요소를 포함하기 위함이다. 한편, 출력단에서의 시간 딜레이는 일시적으로 잘못 판단한 요소를 제거함으로써 감정 분석의 일관성을 유지하기 위한 것이다. 즉, 연속적인 동작에 포함된 무의미한 동작의 경우도, 연속적인 동작에 따른 감정 요소로 분류된다.

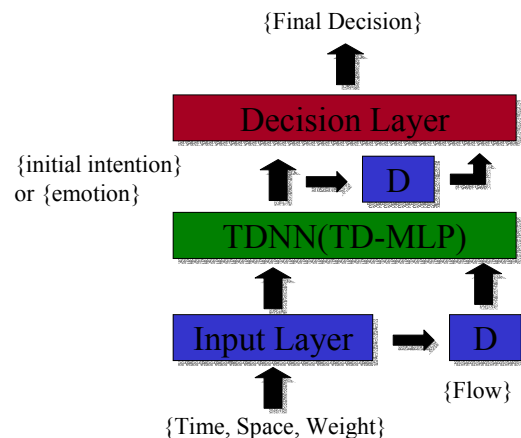


그림 5. 신경회로망을 이용한 감정분석

### 3. 실험결과

그림 6에서는 사용자 분리과정을 통해서 추출된 영상 정보와 다중-뷰 카메라로부터 획득한 영상정보를 합성하여 사용자만의 3 차원 정보를 추출하는 과정을 보여준다.

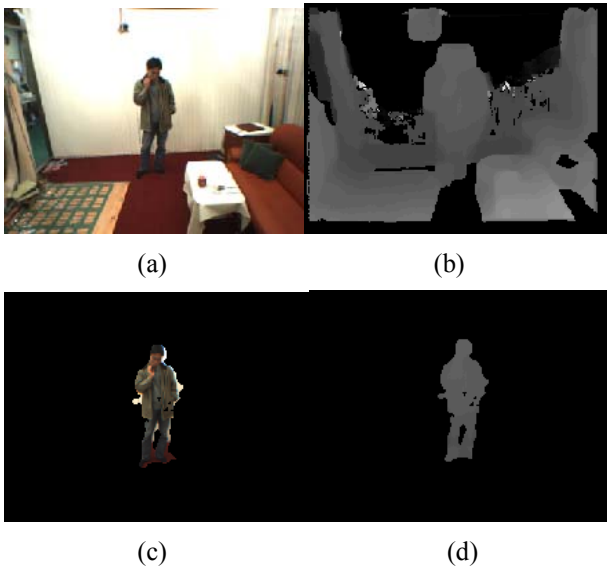


그림 6. 사용자의 3차원 정보 추출 과정. (a) 현재 영상 (b) 현재 영상의 깊이 맵 (c) 분리된 사용자 이미지 (d) 분리된 사용자의 깊이 맵

그림 6(b)에서 보듯이, 멀티-뷰 카메라로부터 획득한 깊이 맵만을 이용하여 사용자의 3 차원 정보를 추출하기는 어렵다. 사용자뿐만 아니라 배경에 대한 정보도 함께 제공되기 때문이다. 하지만, 그림 6(d)에서 보듯이, 사용자 분리 방법과 깊이 맵을 함께 이용하면 사용자에게 대한 3 차원 정보만을 보다 정확하게 추출할 수 있다. 따라서, 사용자에게 대한 3 차원 정보 추출은 사용자 분리 알고리즘의 강건함과 비례하게 된다<sup>2</sup>.

다음 그림은 제안된 공간센서를 사용자 주변에 증강시킨것을 보여준다. 그림 7 에서 보듯이 사용자 주변에 실제로 공간센서가 증강되어있음 보여 주기 위해서, 사용자 분리과정을 통해서 획득한 사용자의 3 차원 정보를 가상 공간에 백-프로젝션

시켰다.

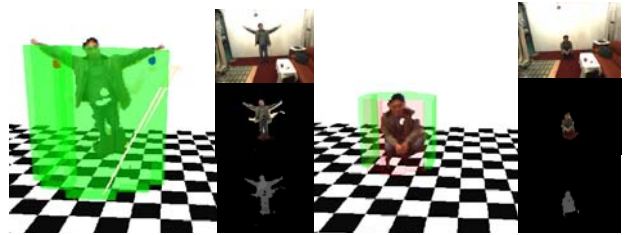


그림 7. 사용자의 움직임 요소 추출

그림 7 에서 보였듯이, 사용자 주변에 증강된 공간센서는 모든 방향에 대한 사용자의 제스처를 트래킹할 수 있을 뿐만 아니라, 동적으로 사용자의 중심을 트래킹해서 공간센서로부터 사용자의 움직임 요소를 실시간으로 추출함을 보여준다. 따라서 제안된 공간센서는 2 차원 시각기술의 한계를 극복하고 실시간으로 사용자의 제스처를 트래킹할 수 있도록 디자인 되었다.

그림 8 은 추출된 사용자의 감정 상태를 가상 공간의 작품에 반영한 결과를 보여준다. 본 작품은 “I-NEXT”에 올려진 작품이다.

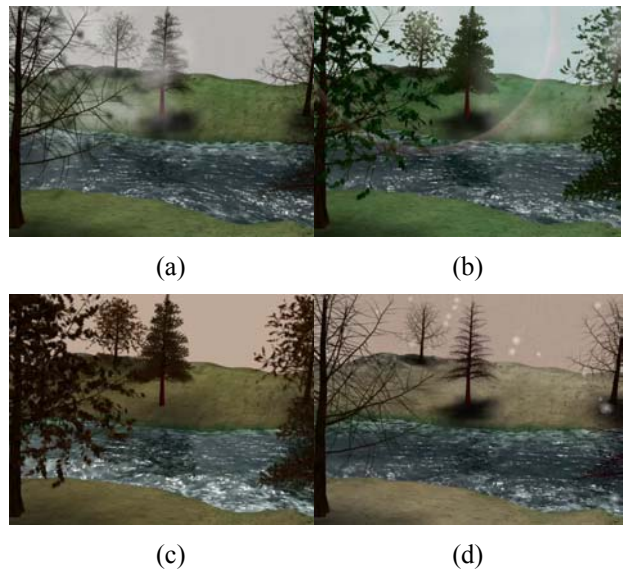


그림 8. 사용자의 감정 정보에 따른 가상공간의 변화. (a) 봄 (b) 여름 (c) 가을 (d) 겨울

그림에서 볼 수 있듯이, 사용자의 감정이 부정적인 요소가 많아지면 점점 차가운 계절로 변화됨을 나타내고, 이와 반대로 긍정적인 요소가 많아지면 점점 따뜻한 계절로 변화됨을 나타낸다.

<sup>2</sup> 물론, 깊이 맵의 정확성과도 관련이 있지만, 본 실험에서는 멀티-뷰 카메라에서 제공하는 알고리즘과 깊이 맵을 사용하였다.

#### 4. 결론 및 추후과제

본 논문에서는 3 차원 시각 기술에 기반한 실시간 사용자 감정 인터페이스를 위한 사용자 감정 인식 방법을 제안하였다. 제안된 공간센서는 2 차원 시각 기술의 한계를 극복하고, 실시간 사용자 제스처 트래킹의 계산 복잡성과 사용자 감정 인식의 어려움을 해소할 수 있다. 하지만, 제안된 공간센서는 사용자의 감정을 인식하기 위해서 보다 정확한 사용자 분리 기법과 깊이 추정 기법이 요구된다. 뿐만 아니라, 사용자의 다양한 제스처에 따른 감정 표현에 대한 데이터 수집과 분석이 필요하다.

#### 참고문헌

- [1] M. Weiser, "The Computer for the 21st Century," Scientific American, pp. 94-104, Sep. 1991
- [2] S. Jang, S. Lee and W. Woo, "Research activities on Smart Environment," Magazine of the KITE, vol. 28, pp.1359-1371, Dec. 2001.
- [3] 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
- [4] Kida, K., Ihara, M., Shiwa, S., Ishibashi, S., "Motion tracking method for the CAVETM system", Signal Processing Proceedings, 2000 WCCC-ICSP 2000. 5th International Conference on , 859 -862 vol.2, 2000
- [5] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," Proc. IEEE ICCV'99 FRAME-RATE Workshop, Kerkyra, Greece, September 1999
- [6] Ahmed Elgammal, David Harwood, and Larry Davis, "Non-parametric Model for Background Subtraction," 6th European Conference on Computer Vision, Dublin, Ireland, June/July 2000
- [7] A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis "Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance", Proceedings of the IEEE, July 2002.
- [8] C. Kim, W. Woo, and H. Jeong, "Determination of Optical Flow by Stochastic Model," Journal of the Korea Information Science Society (KISS), vol.19, no.6, pp.581-594, Nov., 1992
- [9] Soren Lenman, Lars Bretzner, Bjorn Thuresson, "Computer Vision Based Hand Gesture Interfaces for Human-Computer Interaction," Technical report TRITANA-D0209, CID-report, June 2002
- [10] William T. Freeman, and Craig D. Weissman, "Television control by hand gestures", IEEE Intl. Wrkshp. on Automatic Face and Gesture Recognition, Zurich, June, 1995
- [11] A.Camurri, R.Trocca, G.Volpe, "Interactive Systems Design: A KANSEI-based Approach", Proc. NIME2002, Dublin, May 2002
- [12] Markus Kohler. "System Architecture and Techniques for Gesture Recognition in Unconstraint Environments", In Nadia Magnenat Thalmann, editor, International Conference on Virtual Systems and Multimedia VSMM'97, pages 137-- 146, University of Geneva, Switzerland, September 10--12th 1997.
- [13] D. Hong, W. Woo, "A Background Subtraction for a Vision based User Interface", Pacific-Rim Conference on Multimedia (PCM2003),PCM03-CD Proceeding, pp. 1B3.3, 2003
- [14] D.Hong, W.Woo, "SpaceSensor:Real-Time Gesture Tracking for I-NEXT," *International Conferece on Artificial Reality and Telexisitence(ICAT03)*, paper ISSN 1345-1278, pp. 95-99, 2003.
- [15] R. Laban, "Modern Educational Dance", Trans-Atlantic Publications, Inc., 1988(1963).
- [16] D.Hong,W.Lee, J.Jeong, J.Kim, W.Woo, "I-NEXT: An Interactive Networked Expression eXperience Testbed," *Ninth International Conference on Virtual Systems and MutilMedia(VSMM03)*, pp. 455-462, 2003.
- [17] R. Suzuki, Y.Iwadate, M.Inoue, and W.Woo, "MIDAS: MIC Interactive Dance System," in Proc. IEEE SMC, pp.751-756, 2000