

# 멀티뷰 카메라 기반 실시간 사용자 움직임 추적 알고리즘과 사용자 움직임에 기반한 감성 표현 시스템 구현\*

홍동표\*\* · 우운택\*\*

차 례

1. 서론
2. 멀티뷰 카메라 기반 사용자 움직임 추적 알고리즘
  - 2.1 사용자 분리 알고리즘
  - 2.2 사용자 움직임 추적을 위한 알고리즘
3. 사용자 움직임 기반 감성 표현 시스템 구현
4. 결론 및 추후과제

## 1. 서론

미래형 컴퓨팅 환경은 언제, 어디서나, 누구와도 접속할 수 있는 “접속을 통한 정보공유의 시대”가 될 것이다[1,2]. 이처럼 새로운 패러다임이 요구되는 컴퓨팅 환경에서는 사용자에게 보다 자연스럽게 편안한 인간-컴퓨터 상호 작용을 지원하기 위해서 사용자의 자연스러운 움직임을 통해서 사용자의 의도나 감정 정보를 추출할 수 있는 기술 개발이 필요

하다. 하지만, 이러한 요구에도 불구하고 기존의 인간-컴퓨터 상호 작용 지원에 관한 연구는 키보드와 마우스와 같이 2차원적인 한계를 벗어나지 못하였다. 뿐만 아니라, 사용자의 의도나 감정을 고려하지 않거나 고려하더라도 복잡하고 고가의 장비를 이용하였다. 이는 컴퓨터와 상호 작용을 하려는 3차원 공간속의 사용자들에게는 부자연스러운 상호 작용을 제공하는 단점이 있다. 따라서 전통적인 사용자 상호 작용 기법인 키보드, 마우스 등을 대체하는 시각(움직임), 음성 및 센서 등에 기반한 새로운 형태의 사용자 상호 작용에 관한 연구가 진행되어왔다. 특히, 시각 기반 사용자 상호 작용의 장점은 음성이나 센서에 기반한 사용자 상호 작용 기법들에 비해 비교적 보정 과정이 용이 하며, 사용자와 시스템간의 자연스러운 상호 작용뿐만 아니라 공간적 제약을 해소하는 장점이 있다.

일반적으로 시각 기반 사용자 상호 작용 기법에 사용되는 방법들은 마커 등을 부착하는 접촉식 방법과 배경 분리 기법 등을 이용하는 비접촉식 방법으로 구분된다[3]. 마커 등을 부착하는 접촉식 방법은 마커를 추적하여 원하는 정보만을 획득할 수 있는 이점은 있으나, 추적하고자 하는 마커가 가려지거나, 여러 개의 마커들이 동시에 존재하는 경우 추적이 어렵다는 단점이 있다[4]. 더욱이, 접촉식 방법은 사용자가 마커와 같은 것을 착용해야 하는 불편함이 있다. 이와 달리, 배경 분리 기법을 활용한 비접촉식 방법은 앞서 지적인 접촉식 방법이 갖는 한계들을 극복할 수 있다. 하지만, 이러한 장점에도 불구하고 비접촉식 시각 기반의 사용자 상호 작용 기법은 광원에 의한 간섭에는 여전히 민감하다[5-8]. 그리고, 비접촉식 시각 기반 사용자 상호 작용 기법은 사용자의 3차원적인 움직임 정보를 추적하기 위해서 실시간성이 결여된 복잡한 알고리즘을 사용하였다[9-12].

본 논문에서는 기존의 시각 기반 사용자 상호 작용 기법의 단점을 극복할 수 있는 3차원 시각 기술을 이용한 멀티뷰 카메라 기반 실시간 사용자 움직임 추적 알고리즘과 이를 응용한 감성 표현 시스템을 제안한

\* 본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소육성사업의 연구결과로 수행되었음

\*\* 광주과학기술원 정보통신공학과 U-VR 연구실 / {dhong, wwoo}@gist.ac.kr

다. 제안된 알고리즘은 블루 스크린과 같은 특별한 장치 없이도 자연스러운 배경으로부터 사용자 정보만을 분리하는 사용자 분리 방법을 사용한다[13]. 그리고, 사용자 분리 방법을 통해서 추출된 사용자 정보와 멀티뷰 카메라를 통해서 획득된 깊이 정보를 이용하여 사용자의 자연스러운 동작을 추적할 수 있도록 사용자의 주변 공간에 보이지 않는 박스 형태의 공간 센서를 이용한다[14]. 제안된 박스 형태의 공간센서는 사용자가 자신의 자연스러운 움직임을 통해서 자신의 의도를 표현 할 수 있도록 한다. 제안된 방법은 기존의 2차원 시각 기반 사용자 움직임 시스템의 한계를 극복하고, 실시간 사용자 움직임 추적의 계산 복잡성을 해소할 수 있다. 제안된 시스템의 유용성을 검증하기 위해서, 우리는 Network 환경에서 3차원 시각 정보를 사용하여 사용자의 문화적 감정을 상호 작용적으로 표현할 수 있는 “I-NEXT: An Interactive Networked Expression Experience Testbed”를 구축하였다[15]. 구축된 “I-NEXT”는 제안된 시스템을 활용하여 가상공간에서 네트워크를 통해 다자간의 상호 작용이 가능한 시스템이다.

본 논문의 구성은 다음과 같다. 2장에서는 제안된 사용자 분리 알고리즘, 공간센서 디자인 그리고 사용자 움직임 추적에 대한 방법을 소개한다. 3장에서는 제안된 시스템을 통한 실험 결과를 기술하고, 4장에서는 제안된 방법에 대한 결론과 향후 과제에 대해서 언급 한다

## 2. 멀티뷰 카메라 기반 사용자 움직임 추적 알고리즘

### 2.1 사용자 분리 알고리즘

제안된 시스템의 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 장치 없이도 일반적인 배경으로부터 사용자만을 분리할 수 있다. 그림 1은 제안된 공간센서에서 사용하는 배경분리 기법에 대한 알고리즘이다.

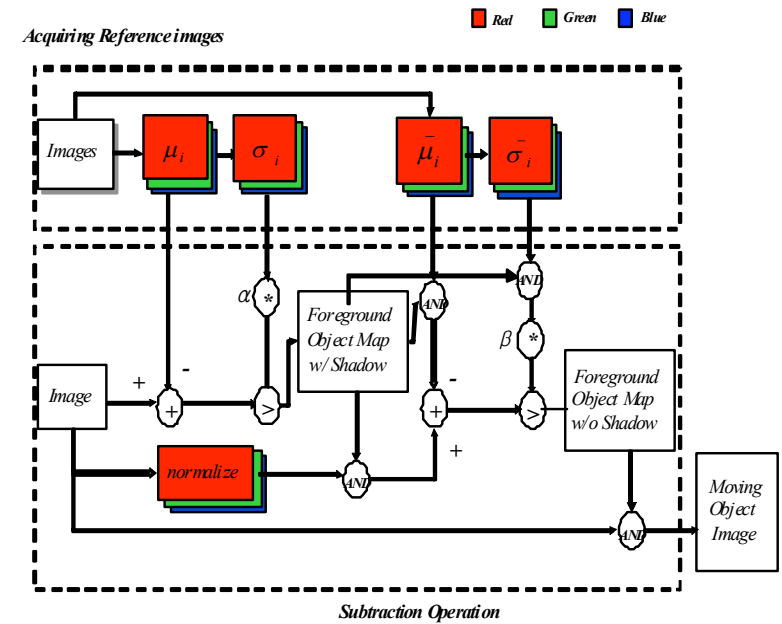


그림 1. 제안된 사용자 분리 알고리즘

일반적으로 객체 분리 알고리즘에는 블루 스크린과 같이 이미 알고 있는 배경의 컬러 정보를 이용하여 객체를 분리하는 방법, 깊이 맵의 차이를 이용하여 분리하는 방법, 그리고 특정 시간동안 배경 정보를 학습한 참조 영상과 현재 영상과의 차이를 이용한 방법 등이 사용되고 있다. 특히, 컬러 정보를 이용하여 객체를 분리할 때에는 빛의 영향에 강건한 컬러 모델을 사용하는 것이 용이하다. 정규화된 RGB 컬러 공간에서는 색의 밝기에 대한 정보가 없기 때문에, 그림자와 같은 광원의 영향을 최소화

화할 수 있다. 따라서, 제안된 사용자 분리 알고리즘에서는 그림 1에서 나타난 것처럼 RGB 컬러 공간과 정규화된 RGB 컬러 공간을 함께 사용하였다.

### 2.2 사용자 움직임 추적을 위한 알고리즘

제안된 공간센서는 배경으로부터 분리된 사용자의 3차원 정보를 이용하여 사용자에게 자연스러운 움직임을 제공하면서도 정확한 사용자의 움직임을 추적하도록 디자인되었다[14]. 제안된 공간센서의 디자인은 공간센서를 사용자를 중심으로 동적으로 위치시킴으로써 사용자에게 자연스러운 움직임을 제공한다. 그리고, 사용자를 중심으로 사용자의 행동반경 내에 공간센서를 위치시킴으로써 정확하게 사용자의 움직임을 감지한다. 제안된 공간센서 구현을 위해서 사용자의 중심점( $H_c = \{H_x, H_y, H_z\}$ )을 다음 식과 같이 구한다.

$$H_c = \frac{1}{N_j} \sum_{i=1}^{N_j} SUD_{j,i} \quad , j = \{x, y, z\} \quad (1)$$

위 식에서  $SUD_j$  (Segmented User Depth 정보)는 배경으로부터 분리된 사용자의 3차원 좌표 정보를 나타낸다.  $N_j$ 는 3차원 좌표 정보에서 각 좌표축에 해당하는 점의 개수이다.

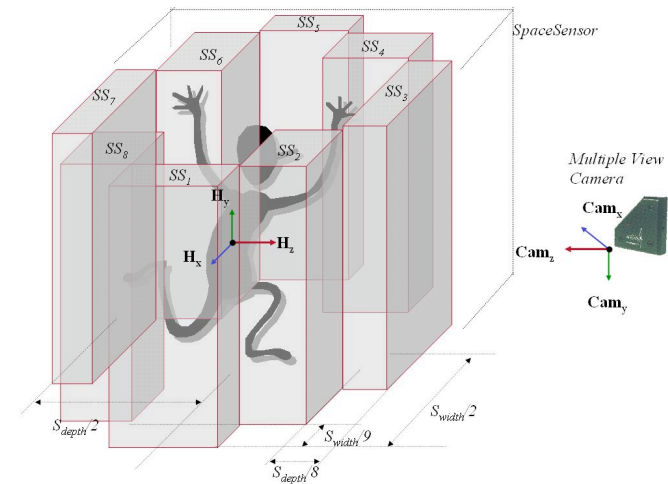


그림 2 제안된 공간센서들의 위치

식(1)을 통해서, 우리는 공간센서 구축에 필요한 변수를 다음과 같이 획득한다.

$$SS_{width} = SS_{height} = SS_{depth} = SUD_{\max\{y_i\}} - SUD_{\min\{y_i\}} \quad (2)$$

위 식에서  $SS_{width}$ ,  $SS_{height}$  그리고  $SS_{depth}$ 는 각각 사용자의 중심점을 기준으로 한 공간센서의 넓이, 높이 그리고 깊이를 나타낸다.  $SUD_{\max\{y_i\}}$ 와  $SUD_{\min\{y_i\}}$ 는 사용자가 차지하는 영역의 최고 높이 점 좌표와 최저 높이 점 좌표를 나타낸다. 식(2)에서  $SS_{width}$ ,  $SS_{height}$  그리고  $SS_{depth}$ 를 같도록 설정한 것은 Leonardo Davinci의 “Vitruvian Man”(http://www.aiwaz.net/Leonardo)를 참조하였다.

정확한 사용자의 움직임을 감지와 사용자 움직임의 3차원 위치 정보를 활용하기 위해서 사용자의 행동 반경을 기준으로 모든 공간센서가 비교적 동일한 공간을 차지하도록 공간을 그림 2처럼 분할한다. 그림 2에서

보여진 것처럼, 제안된 공간센서는 사용자의 움직임을 보다 쉽게 하기위 해서 공간센서를 8개의 영역으로 나누고, 각 영역(SS(0, ..., 7))이 상태를 갖도록 한다. 따라서, 공간센서의 상태(SS)는  $SS = \{ss_0, \dots, ss_7\}$  와 같이 표현할 수 있다. 만약 사용자가 공간센서의 특정 영역을 터치하게 되면, 터치된 영역의 공간센서의 상태는 1이 되고, 그렇지 않은 영역의 공간센서의 상태는 0이 된다. 따라서, 제안된 공간센서는 기존의 시각 기반 추적 기술과는 달리 사용자의 움직임을 추적하기 위해서 공간센서의 상태가 1인 것만을 추출한 뒤, 추출된 공간센서의 영역에서 사용자가 터치한 지점의 3차원 좌표를 다음 식과 같이 계산된다.

$$P(x, y, z) = \left( \frac{1}{N_x} \sum_{i=1}^{N_x} x_i, \frac{1}{N_y} \sum_{i=1}^{N_y} y_i, \frac{1}{N_z} \sum_{i=1}^{N_z} z_i \right) \quad (3)$$

위 식에서  $P(x, y, z)$  는 공간센서에 사용자가 터치한 위치의 좌표를 나타낸다.  $N_x, N_y$  그리고  $N_z$  는 공간센서에서 사용자 터치한 점의 개수이다.

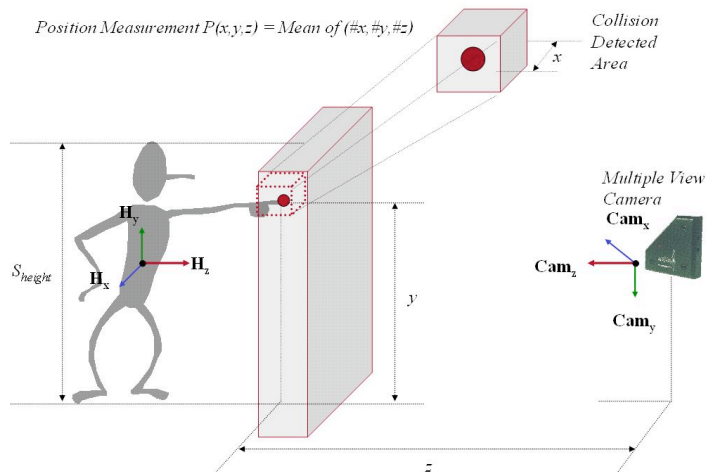


그림 3. 제안된 공간센서에서의 움직임 추적

제안된 공간센서는 사용자의 상, 하, 좌, 우 공간뿐만 아니라 전, 후를 사용할 수 있기 때문에 기존의 2차원 영상에 기반한 시각 기반 시스템들이 갖는 한계를 극복한다. 또한, 공간을 3차원 박스를 이용하여 분할하고 사용자의 움직임을 추적한다. 이때 3차원 박스의 갯수는 움직임 추적의 정확성과 계산 복잡성을 고려하여 결정한다. 움직임 정보는 사용자 대신 공간센서의 상태 정보를 추적하여 추정한다. 따라서, 제안한 공간센서는 사용자로부터 복잡한 센서나 기기를 제거할 뿐만 아니라 새로운 형태의 사용자 인터페이스로 활용 가능하다. 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.

### 3. 사용자 움직임 기반 감성 표현 시스템 구현

그림4에서는 사용자 분리과정을 통해서 추출된 영상 정보와 멀티-뷰 카메라로부터 획득한 영상정보를 합성하여 사용자만의 3차원 정보를 추출하는 과정을 보여준다.

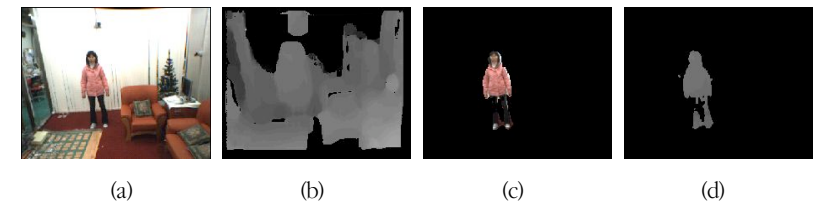


그림 4. 사용자의 3차원 정보 추출 과정

(a) 현재 영상 (b) 현재 영상의 깊이 맵 (c) 분리된 사용자 이미지 (d) 분리된 사용자의 깊이 맵

그림 4(b)에서 보듯이, 멀티-뷰 카메라로부터 획득한 깊이 맵만을 이

1) 뉴턴의 운동 제2법칙  $F=ma$ 에 따르면, 힘은 가속도와 비례 한다.

용하여 사용자의 3차원 정보만을 추출하기는 어렵다. 사용자뿐만 아니라 배경에 대한 정보도 함께 제공되기 때문이다. 하지만, 그림 4(d)에서 보듯이, 사용자 분리 방법과 깊이 맵을 함께 이용하면 사용자에 대한 3차원 정보만을 보다 정확하게 추출할 수 있다. 따라서, 사용자에 대한 3차원 정보 추출은 사용자 분리 알고리즘의 강건함과 비례하게 된다. 물론, 깊이 맵의 정확성과도 관련이 있지만, 본 실험에서는 멀티-뷰 카메라에서 제공하는 알고리즘과 깊이 맵을 사용하였다. 그림 5는 본 시스템에 사용된 멀티-뷰 카메라의 거리에 따른 정확도를 측정한 것이다.

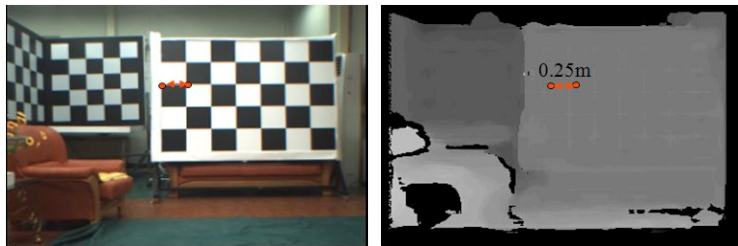


그림 5. 시스템에 사용된 멀티뷰 카메라의 정확도 측정

그림 5에서와 같이, 테스트 패턴의 격자간 간격은 0.25m이고, 카메라를 2m에서 5m까지 이동하면서 측정된 에러는 최소 0.02m, 최대 0.07m이다. 제안된 시스템은 사용자의 정확한 동작을 추적하는 방법 대신 사용자의 전체적인 움직임을 추적하는 시스템이기 때문에, 측정된 정도의 정확도로도 충분히 활용가능하다.

다음 그림은 사용자의 움직임에 따라 동적으로 생성되는 공간센서를 보여준다. 그리고, 사용자 주변에 실제로 공간센서가 증강되어 있음을 보여주기 위해서, 사용자 분리과정을 통해서 획득한 사용자의 3차원 정보를 가상 공간에 역투영 시켰다.

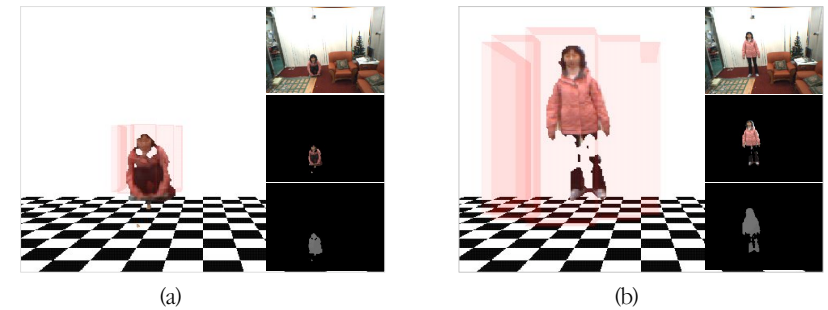


그림 6. 사용자 자세에 따른 동적 공간센서 생성

그림 6에서 보듯이, 사용자의 자세에 따라서 공간센서는 동적으로 그 사용자의 중심점 및 공간센서 생성에 필요한 변수를 추출하여 사용자 주변에 증강된다. 제안된 공간센서는 기존의 방법들과는 달리 사용자에 따른 별도의 보정 과정이 필요 없다. 따라서, 제안된 공간센서의 이러한 특성을 이용하여 사용자 움직임에 대한 수치적 해석이 가능하다. 예를 들면, 사용자가 그림 6(a)와 6(b)와 같은 자세를 취했다고 가정한다면, 그림 6(b)의 사용자는 그림 6(a)보다 훨씬 더 많은 공간을 차지하고 보다 활동적이라는 것을 공간센서의 부피를 계산하여 알 수 있다.

그림 7에서는 사용자의 움직임을 공간센서가 실시간에 추적하고 있음을 보여주는 그림이다.

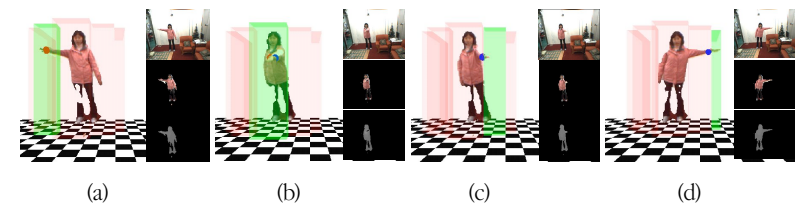


그림 7. 실시간 사용자 움직임 추적

그림 7에서 보듯이, 사용자 주변에 증강된 공간센서는 모든 방향에 대한 사용자의 움직임을 추적할 수 있을 뿐만 아니라, 동적으로 사용자의 중심을 추적해서 공간센서로부터 사용자의 움직임 요소를 실시간으로

추출함을 보여준다. 제안된 공간센서는 2차원 시각 기술의 한계를 극복하고 실시간으로 사용자의 움직임을 추적할 수 있도록 디자인 되었다.

그림 8은 “I-NEXT”에 사용되어진 배경화면을 보여준다. I-NEXT는 사용자의 움직임과 자세를 실시간으로 분석하여 가상 공간에 사용자의 움직임을 반영한 것을 보여주는 그림이다.

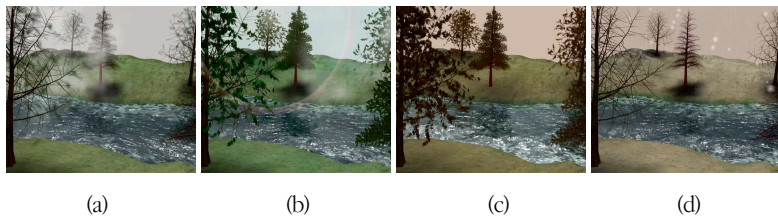


그림 8. 사용자의 움직임과 자세에 따른 가상공간의 변화  
(a) 봄 (b) 여름 (c) 가을 (d) 겨울

특히, 그림 9에서 볼 수 있듯이, 사용자는 자신의 움직임을 자연스럽게 표현하면, 공간센서는 이러한 사용자의 움직임을 실시간으로 추적하여 사용자가 차지하는 물리적 공간을 수치적으로 계산한다. 그리고, 사용자의 움직임과 자세에 대한 히스토리-버퍼를 비교하여 그림 8에서 보여준 가상 공간을 변화시킨다. 사용자의 움직임과 자세와 관련해서 히스토리-버퍼를 사용하는 것은 사람의 움직임이나 자세가 아주 미세하게 변화하는 것에 대한 영향을 최소화하기 위함이다. 그림 8의 작품에서 볼 수 있듯이, 사용자는 자신의 움직임을 통해서 가상공간에 자신의 의사를 표현할 수 있다.

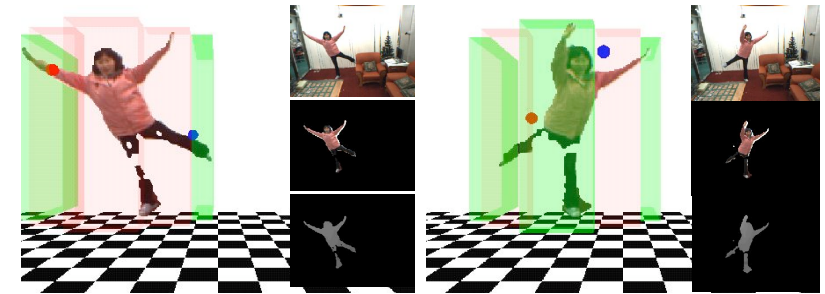


그림 9. 사용자의 움직임과 자세에 따른 공간센서의 변화

#### 4. 결론 및 추후과제

본 논문에서는 3차원 시각 기술을 이용한 멀티-뷰 카메라 기반 실시간 사용자 움직임 추적 알고리즘과 이를 활용한 감성 표현 시스템을 제안하였다. 특히, 제안된 공간센서는 2차원 시각 기술의 한계를 극복하고, 실시간 사용자 움직임 추적의 계산 복잡성을 해소할 수 있다. 하지만, 제안된 공간센서는 사용자의 의도나 의사를 보다 정확하게 표현하기 위해서 보다 정확한 사용자 분리 기법과 깊이 추정 기법이 요구된다. 추후 과제로는 제안된 공간센서를 통해 추출된 사용자 움직임을 분석하여 사용자의 감성 정보를 보다 다양하게 표현할 수 있는 기법에 관한 연구가 필요하다.

## ■ 참고문헌

- [1] M. Weiser, "The Computer for the 21st Century", Scientific American, pp. 94-104, 1991.
- [2] S. Jang, S. Lee and W. Woo, "Research activities on Smart Environment", Magazine of the KITE, vol. 28, pp. 1359-1371, 2001.
- [3] W. Woo, N. Kim, K. Wong and M. Tadenuma, "Sketch on Dynamic Gesture Tracking and Analysis Exploiting Vision-based 3D Interface", In Proceeding of SPIE PW-EI-VCIP'01, vol. 4310, pp. 656-666, 2001.
- [4] Kida, K., Ihara, M., Shiwa, S., Ishibashi, S., "Motion tracking method for the CAVETM system", In Proceeding of 5th International Conference on Signal Processing, vol. 2, pp. 859-862, 2000.
- [5] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection", Proceeding in IEEE ICCV'99 FRAME-RATE Workshop, 1999.
- [6] Ahmed Elgammal, David Harwood, and Larry Davis, "Non-parametric Model for Background Subtraction", 6th European Conference on Computer Vision, 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, 1992.
- [9] Soren Lenman, Lars Bretzner, Bjorn Thuresson, "Computer Vision Based Hand Gesture Interfaces for Human-Computer Interaction", Technical report TRITANA-D0209, June 2002.
- [10] William T. Freeman, and Craig D. Weissman, "Television control by hand gestures", IEEE International Workshop on Automatic Face and Gesture Recognition, Zurich, June, 1995.
- [11] Atsushi Nishikawa, Akio Ohnishi, Fumio Miyazaki, "Description and Recognition of Human Gestures Based on the Transition of Curvature from Motion Images", In Face and Gesture Recognition, pp. 552-557, 1998.
- [12] Markus Kohler. "System Architecture and Techniques for Gesture Recognition in Unconstraint Environments", International Conference on Virtual Systems and Multimedia (VSMM'97), pp. 137-146, 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 Conference on Artificial Reality and Telexistence (ICAT03), pp. 95-99, 2003.
- [15] 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 MultiMedia (VSMM'03), pp. 455-462, 2003.

## | 국문 초록 |

## 멀티-뷰 카메라 기반 실시간 사용자 움직임 추적 알고리즘과 사용자 움직임에 기반한 감성 표현 시스템 구현

홍동표 · 우운택

본 논문에서는 멀티-뷰 카메라에 기반한 실시간 사용자 움직임 추적 알고리즘과 이를 활용한 사용자 움직임 기반 감성 표현 시스템을 제안한다. 사용자에게 보다 자연스럽고 편안한 인간-컴퓨터 상호 작용을 지원하기 위해서는 사용자의 자연스러운 움직임을 통해서 사용자의 의도나 감정을 추출할 수 있는 기술이 필요하다. 제안된 방법은 사용자의 움직임을 통해서 사용자의 의도나 감정 상태를 분석하기에 용이한 멀티-뷰 카메라 기반 사용자 움직임 추적 알고리즘이다. 특히, 제안된 방법에서는 사용자의 움직임을 추적하기 위해서 블루 스크린과 같은 특별한 장치 없이 자연스러운 배경으로부터 사용자 정보만을 분리할 수 있는 사용자 분리 방법을 사용한다. 그리고, 사용자 분리 방법을 통해서 추출된 사용자 정보와 멀티-뷰 카메라를 통해서 획득된 깊이 정보를 이용하여 사용자의 자연스러운 움직임을 추적할 수 있도록 사용자 주변의 공간에 보이지 않는 박스 형태의 공간센서를 활용한다. 또한, 제안된 방법을 활용하여 사용자의 움직임을 통해 사용자의 감성을 가상 공간에 표현할 수 있는 시스템을 구현하였다. 제안된 시스템은 기존의 2차원 시각 기반 사용자 움직임 추적 시스템의 한계를 극복하고, 실시간 사용자 움직임 추적의 계산 복잡성을 해소 할 수 있다.

| 주제어 : 사용자 분리, 공간 센서, 움직임 추적

## | Abstract |

## A Multiple View Camera-based Real-time Gesture Tracking Algorithm and Emotional Expression System

Dongpyo Hong · Woontack Woo

In this paper, we propose a user motion tracking algorithm by using a multiple view camera, and emotional expression system by tracking user's gestures in a personal space. In order to provide more natural and comfortable human-computer interactions to a user, it is essential to extract intentions or emotions of the user through his or her natural movements. Thus, the proposed system can support to analyze intentions or emotions of the user by tracking his or her motions in the personal space. To track motions of the user precisely, the proposed system utilizes a user segmentation method that can extract only the user from a natural background without special devices like blue screen. In addition, it exploits invisible SpaceSensor that can be constructed from segmented results and disparity (depth) map in order to track user's motions. The proposed system can overcome the previous 2D vision-based user motion tracking system and can resolve the complexity of real-time motion tracking algorithms.

| keyword : User Segmentation, SpaceSensor, Gesture Tracking