

실감형 인터랙티브 환경의 인터페이스 연구

경동욱* · 한은정** · 양종렬*** · 정기철****

1. 서 론

컴퓨터에 의해 촉발된 디지털 기반의 정보과학 기술은 그 발전 속도가 가속화됨에 따라서, 최근에 이르러서는 과학기술분야로부터 확장되어 영상, 영화, 게임, 미디어아트 등의 사회, 문화적인 분야로 영향력을 넓히게 되었다. 이러한 다양한 분야에 응용될 수 있는 기반 기술로써 **실감형 인터랙티브 환경**에 관한 연구는 사용자의 다양한 움직임 정보를 입력으로 하는 실제감 있는 영상 및 정보를 제공하기 위한 환경을 말하며, 영화 '마이너리티 리포트'에서 톰크루즈가 대형 디스플레이에서 입력 장치를 양손을 사용하여 원하는 정보를 검색했던 장면처럼, 현실에서도 대형 화면 앞에서 사용자의 제스처를 입력장치로 현실감 있는 영상을 제공하는 연구를 말한다.

실감형 인터랙티브 환경은 크게 두 가지의 인터페이스로 나눌 수 있다. 첫번째는 현실감 있는 영상을 제공하기 위한 실감형 디스플레이 기술이

며, 두번째는 사용자의 직관적인 입력 방법인 제스처 정보를 인식하는 기술이다.

실감형 디스플레이 기술은 사용자에게 현실감을 제공하기 위해 많은 연구가 진행되어 왔다. 개인을 위한 대중적인 디스플레이는 LCD/CRT 모니터와, 증강현실(Augmented Reality)과 가상현실(Virtual Reality)에서 많이 사용하는 HMD(Head Mounted Display)가 있으며, 동시에 다수의 사용자에게 영상을 제공하기 위한 디스플레이 방법으로 프로젝터를 이용한 프로젝션 시스템, 입체 영상을 제공하기 위한 3D 디스플레이 등이 있다.

실감형 인터랙티브 환경에서의 직관적인 입력 방법으로 사용자의 제스처 인식 기법을 주로 사용한다. 사용자의 제스처 인식을 위한 연구로써, 일반적으로 1대의 카메라를 이용하여 2D상에서 제스처 인식을 수행하는 연구[1-3]와, 센서를 몸의 중요 부위에 부착하여 사람의 움직임을 모션캡처하는 연구[4-5] 등이 있다. 하지만 2D 제스처 인식은 카메라의 시점에 비독립적인 이유로 인해서 정확한 인식이 어려운 단점이 있으며, 모션캡처 장비는 센서를 중요 부위에 항상 착용해야 하는 불편함과 고비용 문제로 인해서 사용자에게 편의성을 제공하기 어렵다. 이러한 이유로 최근의 제스처 인식에서는 다시점 카메라를 이용하여 사용자 모양의 3차원 정보인 3D 볼륨(3D volume)을 실시간으로 생성하여 보다 정확한 인식을 수행하

※ 교신저자(Corresponding Author) : 정기철, 주소 : 서울특별시 동작구 상도5동 1-1 송실대학교(156-743), 전화 : 02) 828-7260, FAX : 02)822-3622, E-mail : kcjung@ssu.ac.kr

* 송실대학교 IT대학 미디어학부 미디어 공학(박사과정) (E-mail : kiki227@ssu.ac.kr)

** 송실대학교 IT대학 미디어학부 콘텐츠 공학(박사과정) (E-mail : hanej@ssu.ac.kr)

*** 송실대학교 IT대학 미디어학부 미디어 공학(박사과정) (E-mail : yjyhorse@ssu.ac.kr)

**** 송실대학교 IT대학 미디어학부 교수

고자 하는 연구가 많이 진행되고 있다.

본 논문은 실감형 인터랙티브 환경에서 사용되는 디스플레이 기술과 제스처 인식 기술에 대한 기술 동향을 소개하고자 한다.

2. 실감형 디스플레이의 기술동향

최근의 디스플레이 연구는 사용자에게 사실적이고 실감나는 영상을 제공하기 위해 활발히 연구되고 있으며, 그 중에서 가상현실 및 증강현실에서 많이 사용하는 HMD, 다수 사용자를 위한 프로젝션 시스템인 타일드 디스플레이(Tiled Display)와 CAVE, 그리고 입체 영상을 제공하기 위한 3D 디스플레이 등이 있다.

2.1 HMD(Head Mounted Display)

HMD는 실제 세계를 배경으로 증강현실과 가상현실 등에서 실제 느낌을 주기 위해 많이 사용되는 개인용 디스플레이이다. Emori[6]의 연구는 현실세계에서 마커 및 2D 매트릭스 코드(그림 1(b))를 사용하여 실제 종이책을 보는 것과 같은 효과(그림 1(c))를 제공하며 (그림 1(a)), Satoh[7]의 연구는 고정된 카메라와 HMD에 부착된 마커를 이용하여 사용자의 시점에 맞는 3D 영상을 제공하는 연구로써(그림 1(d)), 두 연구는 HMD에서의 사실감을 높이기 위한 연구이다. 또한 실감형 게임(augmented game)을 위한 몰입형 디스플레이 환경을 제공하기 위한 연구로써, Szalavari[8] 등은 실제 환경에서 두 명 이상이 가상의 마작패를 이용하여 게임을 진행하는 협동적인 실감형 게임 *마작(Mah-Jongg)*을 소개하였다. 이와 같이 HMD는 개인 사용자에게 현실감 있는 영상을 제공하지만, 다수의 사용자에게 동시에 동일한 환경을 제공하기 어렵고, 착용이 불편하다는 단점이 있다.

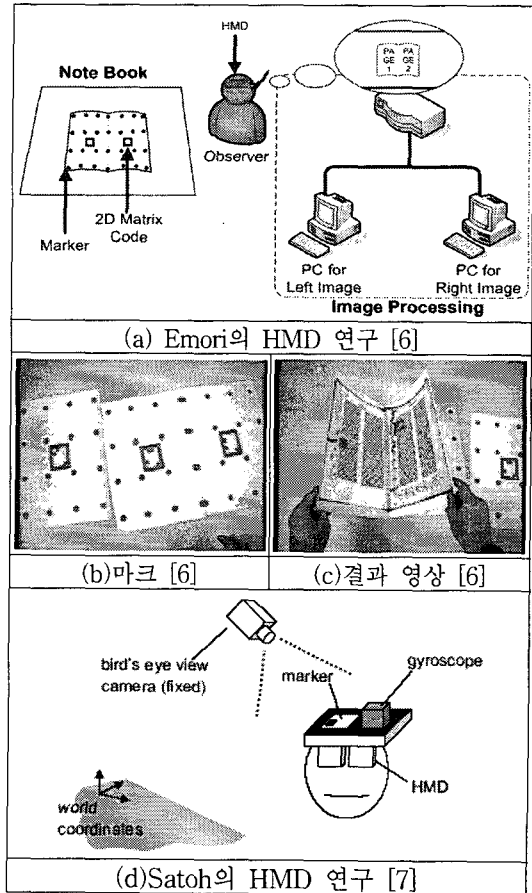


그림 1. 실감형 디스플레이로서의 HMD 연구

2.2 타일드 디스플레이(Tiled Display)

타일드 디스플레이는 여러 대의 프로젝터를 전방 또는 후방에서, 평면 또는 곡면 스크린에 투사하여 고해상도의 대형화면을 상대적으로 저렴한 가격으로 제공하는 시스템이다. 타일드 디스플레이는 개인용 디스플레이가 아닌 다수의 사용자에게 동시에 동일한 영상을 제공할 수 있다는 점과, 사용자가 장치의 착용없이 자연스런 영상을 제공할 수 있다는 장점이 있다. 실감형 디스플레이를 위해서 Raskar[9] 등은 곡면 스크린 상에서의 타일드 디스플레이 기법을 연구하였다 (그림 2).

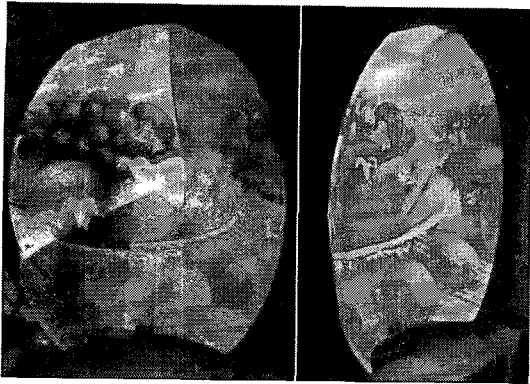


그림 2. 곡면형 타일드 디스플레이 (9)

곡면형 타일드 디스플레이에서의 각 프로젝션 영상의 정렬 및 기하보정은 선형 왜곡을 가지는 평면 디스플레이에 비해 비선형 왜곡으로 인한 많은 연산량과 수행 시간을 필요로 하는 문제점이 있다(그림 3).

이와 같은 곡면형 타일드 디스플레이에서의 문제를 해결하기 위해서, 프로젝터의 프로젝션되는 각도에 따라 카메라를 함께 배치하여 수작업 없이 자동으로 정확한 보정을 수행 하기위한 연구가 진행되고 있다. 그 예로 Jeroen [10]등의 연구는 그림 4(a)와 같이 프로젝션 되는 위치에 카메라를 함께 사용함으로써 곡면에서 정교한 보정이 가능하게 하는 연구이다. 이와 유사한 연구로써, Moriya [11]등은 같은 수의 프로젝터와 카메라를 배치하여 곡면 스크린에서 정확한 보정을 위한

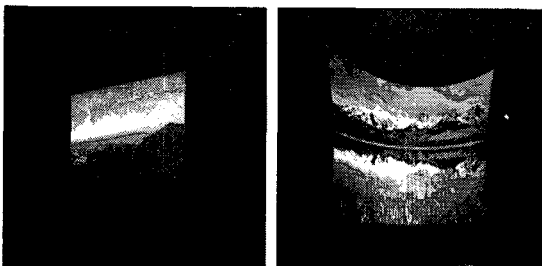
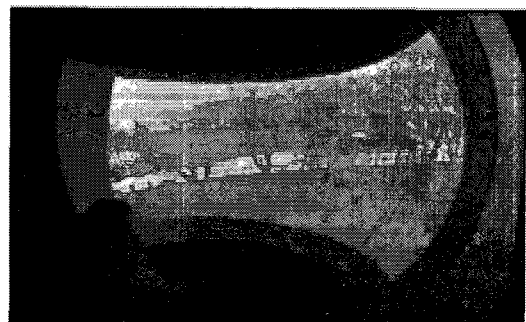
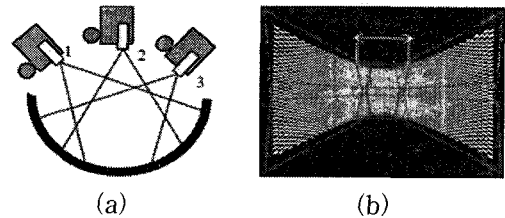


그림 3. 평면 및 곡면에서의 프로젝션의 왜곡

연구를 수행하였다. 그림 4(b)는 카메라에서 보이는 영상으로써 중심은 적은 왜곡을 보여주지만 외곽으로 갈수록 왜곡이 심해진다. 이러한 문제를 해결하기 위해서 카메라를 프로젝터와 동일한 수를 사용하여 각각 프로젝션 되는 영역을 각각의 카메라를 이용하여 보정함으로써 문제를 해결하는 연구이다(그림 4(c)).

2.3 CAVE 시스템

CAVE는 룸 형식의 몰입형 가상 환경으로 3면 이상의 평면으로 둘러싸인 형태로 영상을 제공함으로써, 보다 현실감 있는 환경을 제공할 수 있는 프로젝션 기반의 가상현실 시스템으로 개발 되었다[12-13]. 그림 5는 실제 CAVE의 환경과 영상의 결과를 보여주고 있다. 실제적으로 가장 몰입형 환경으로써 사용자에게 보다 사실감 있는 환경을



(c)

그림 4. 곡면 스크린의 타일드 디스플레이: (a) 프로젝터와 카메라의 배치도 (10), (b) 곡면스크린에서의 왜곡 정도 (11), (c) 결과 영상 (11)

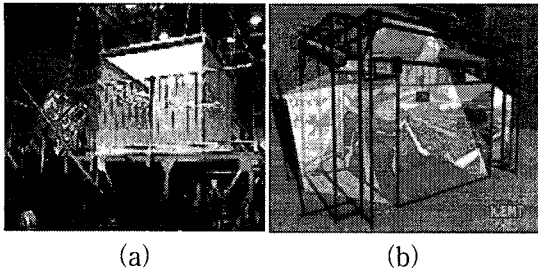


그림 5. CAVE 시스템: (a) CAVE 시스템의 환경, (b) CAVE의 결과 영상

제공하고 있지만, 다수의 사람들에게 제공하기 위해서는 많은 공간과 많은 장비가 필요하다는 단점이 있다.

2.4 3D 디스플레이 시스템

3D 디스플레이는 입체 영상 시스템으로써, 실제에 가까운 입체 영상을 제공하는 디스플레이를 말하며[14], 눈으로 분리되어 들어오는 좌·우안 영상을 두뇌에서 합성하여 3D 영상이 보이는 양안 시차 3D 방식(Stereoscopic 3D)으로써, 기존의 안경을 착용하던 방법에서 보다 발전되어 안경 없이도 입체 영상을 볼 수 있는 방식이 연구되고 있다(그림 6). 양안 시차 방식은 안경식과 무안경식으로 나눌 수 있으며, 90년대 이후에 평판 디스플레이의 개발로 무안경식 3D의 개발이 활발히 진행되고 있다.

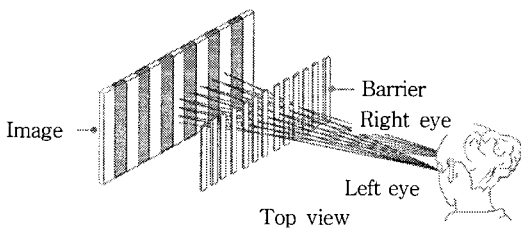


그림 6. 3D 디스플레이의 개념 [14]

3. 인터랙티브환경의 인터페이스기술동향

인터랙티브 환경에서의 인터페이스로 가장 직관적인 방법 중 하나인 사용자의 제스처나 모션이 주로 사용된다. 사용자의 제스처나 모션을 사용하는 인터페이스 방법은 실제 생활과 유사한 환경을 제공해주는 장점으로 인하여 인터랙티브 환경에서 중요한 연구 분야로 간주되며, 보다 편리하고 정확한 인식을 위해서 많은 연구가 진행되고 있다.

기존의 연구는 주로 2D 기반의 제스처 인식 연구와, 사용자의 몸에 부착된 센서의 위치를 감지/인식하는 모션캡처 연구, 그리고 사용자의 움직임 을 실시간으로 3D 볼륨 데이터로 생성하여 모션을 인식하는 연구로 구성된다. 기존의 2D 인식 연구는 1대의 카메라를 사용하여 얼굴 인식(face recognition), 게이트 인식(gait recognition), 제스처 인식(gesture recognition) 등을 수행하지만 깊이 정보의 부재 및 카메라의 시점에 자유롭지 못한 단점으로 인해 자연스런 인터랙티브 환경의 인터페이스로는 적합하지 않다. 또한 센서를 부착하여 모션을 인식하는 제스처 인식 연구는 정확한 제스처 및 모션을 인식하지만, 사용자가 미리 장비를 착용해야 하는 불편함과 고비용의 단점을 가지고 있다. 3D 볼륨을 이용한 연구는 다시점 카메라를 이용하여 사용자의 3D 볼륨을 실시간으로 생성하여 정확한 제스처 인식을 함으로써 기존의 문제점을 해결하기 위해 연구된 방법이다.

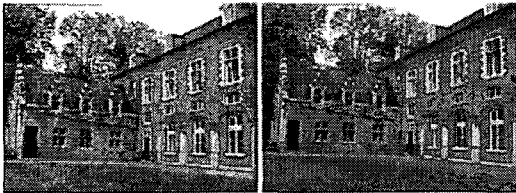
3D 볼륨을 이용한 모션인식 연구는 크게 두 가지 세부 연구 분야로 구분할 수 있으며, 이는 다시점 카메라를 이용한 3D 볼륨 데이터 생성 부분과 생성된 3D 모델을 이용한 모션인식 연구이다.

3.1 3D 볼륨 생성 연구

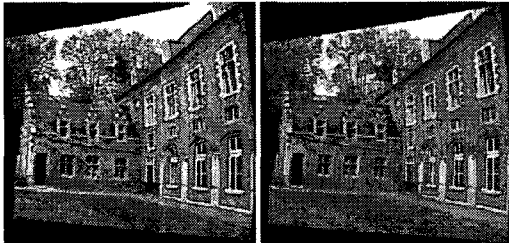
카메라 기반 3D 볼륨 생성 연구는 크게 3가지로 나눌 수 있다. 첫 번째로 움직임 기반 모델링(Shape from Motion), 두 번째로 음영 기반 모델링(Shape from Shading), 그리고 세 번째로 실루엣 기반 모델링(Shape from Silhouette)이다.

움직임 기반 모델링 기법은 정적 객체의 주위에 카메라를 이동시켜 가면서 촬영한 영상 내 특징점을 이용하여, 영상 간 대응 관계를 계산하여 3차원 모델을 얻는 방법으로써 Pollefeys의 연구[15] 등이 있다(그림 7).

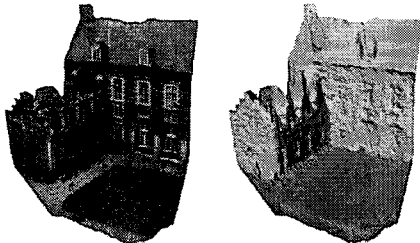
음영 기반 모델링 기법[16]은 영상 내 음영 정보를 이용한 3D 모델을 생성하는 방법이다 (그림 8).



Step1. Match points over the whole image sequence.



Step2. Initialize the structure and motion recovery



Step3. For every additional view
Step4. Refine the structure and motion through bundle adjustment.

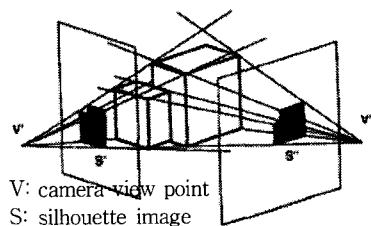
그림 7. 움직임 기반의 모델링 과정 (15)



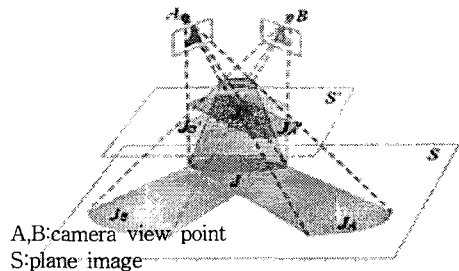
그림 8. 음영 기반을 이용한 복원된 3D 모델 (16)

이 두 가지 기법은 사용자의 제스처나 모션인식을 위해 실시간으로 3D 볼륨을 생성할 수 없다는 단점이 있다. 그 이유는 움직임 기반 모델링 기법은 고정된 객체(rigid object)만을 가지고 3D 볼륨을 생성하기 때문에 사람과 같이 움직이는 객체에는 사용할 수 없으며, 음영 기반 모델링 기법도 마찬가지로 움직이는 객체에 다양한 광원 조건을 적용하기 힘들기 때문이다.

실시간 3D 볼륨을 생성하기 위한 방법으로 많이 사용하는 방법은 실루엣 기반 모델링 기법으로써, 다수의 카메라로부터 영상을 획득한 후, 추출된 실루엣을 바탕으로 객체의 visual hull을 생성함으로써 3D 볼륨을 생성하는 기법이다. 실루엣 기



(a) 실루엣의 투사 영역을 이용한 방법 [17]



(b) 다중 플레인(plane) 기반의 방법 [18]

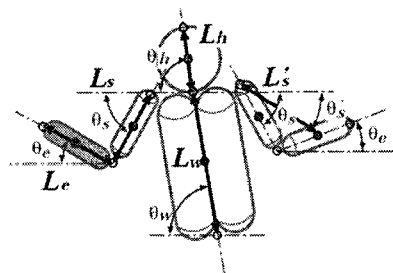
그림 9. 3D 볼륨 생성에서 2가지 구현 방법

반 3D 볼륨 생성의 구현은 크게 두 가지로 나뉜다. 첫 번째는 각 카메라의 실루엣 영상을 바탕으로 각 실루엣 영역을 투사하여 교차된 영역을 3D 볼륨으로 생성하는 방법[17]과 (그림 9(a)), 두 번째는 병렬처리를 위해서 Z축 복셀(voxel) 단위만큼의 플레인 생성 및 교차영역을 추출함으로써 3D 볼륨을 생성하는 방법[18]이다 (그림 9(b)).

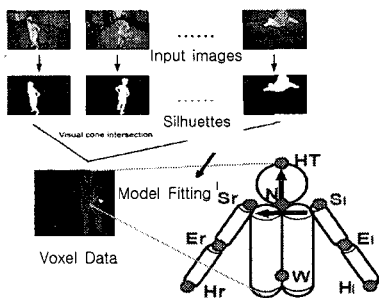
3.2 3D 볼륨을 이용한 모션인식

생성된 3D 볼륨을 이용한 모션인식 연구는 초기모델(입력되는 객체의 기본 정보)을 이용하는 방법과, 초기모델 없이 실시간으로 3D 볼륨을 이용하여 초기모델을 생성 및 적용하여 모션을 인식하는 방법으로 나눌 수 있다.

초기모델을 기반으로 모션 캡처를 수행하는 연



(a)



(b)

그림 10. 3D 볼륨을 이용한 모션 캡처 수행과정 (19): (a) 초기 모델. (b) 수행 과정

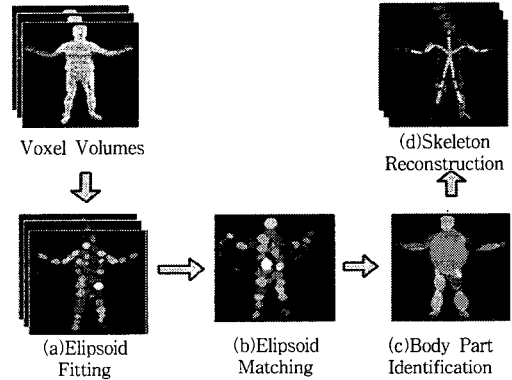


그림 11. 전체 처리과정 (20)

구는 다음과 같다[19-21]. Sumi의 연구[19]는 플레인 기반으로 3D 볼륨을 생성하여 상반신만을 초기 모델에 적용하여 모션 파라미터를 예측한다. 그림 10(a)는 초기 상반신의 모델로써 각 관절과 관절간의 각도 및 위치 등을 표현하며, 그림 10(b)는 3D 볼륨 생성을 위한 환경 및 초기모델과의 적합도 계산을 통해서 적용되는 과정을 보여준다.

초기모델을 사용하지 않고 실시간으로 생성하여 적용하는 연구로써, Aguiar [20]는 초기 모델을 사용하지 않고, 인체의 중요부분을 식별하여 모션 캡처 파라미터를 추출한다(그림 11). 이와 같은 연구는 3D 볼륨을 이용해서 모션 파라미터를 획득하기 위한 연구로써, 정확한 모션 데이터를 제공할 수 있다.

4. 결론

실감형 인터랙티브 환경은 사용자에게 보다 실감있고 편리한 환경을 제공하기위한 연구로써, 디스플레이 및 인터랙티브 환경을 위한 많은 연구가 진행되었고 앞으로도 많은 발전을 이루어질 것이다. 본 논문은 실감형 인터랙티브 환경을 제공하기 위한 두 가지 필수 기술인 디스플레이 및 제스처 인식을 위한 기존 연구의 조사에 관한 논문이

며, 이러한 분야에 대한 많은 연구가 진행될 필요가 있음을 보였다.

참 고 문 헌

- [1] 정기철, 한은정, 강현, “비디오 게임 인터페이스를 위한 인식 기반 제스처 분할,” 멀티미디어학 회논문지, Vol. 8, No. 9, pp. 1177~1186, 2005.
- [2] Chan Wah Ng, Ranganath, S. “Gesture Recognition via Pose Classification,” Proc. 15th International Conference on Pattern Recognition, Vol. 3, pp. 699~704, 2000.
- [3] A. Elgammal, V. Shet, Y. Yacoob, L. S. Davis, “Gesture Recognition using a Probabilistic Framework for Pose Matching” Proc. 7th International Conference on Control, Automation, Robotics and Vision, pp. 2~5, 2002.
- [4] 이지형, 이찬수, 박찬중, 김동현, “인체계측학 정보를 이용한 분신의 실시간 제스처 인식,” 한국 정보처리학회 1997년 추계학술대회, Vol. 4, No. 2 pp. 576~579, 1997.
- [5] V. Pavlovic, R. Sharma, and T. Huang, “Gestural Interface to a Visual Computing Environment for Molecular Biologists,” Proc. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, 1996.
- [6] Emori, M., Saito, H., “Texture Overlay onto Deformable Surface using HMD,” Proc. Virtual Reality, pp. 221~222, 2004.
- [7] Satoh, K., Uchiyama, S., Yamamoto, H., “A Head Tracking Method using Bird’s-eye view Camera and Gyroscope,” Proc. 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 202~211, 2004.
- [8] Z. Szalavari, E. Eckstein and M. Gervautz, “Collaborative Gaming in Augmented Reality,” Proc. ACM Symposium on Virtual Reality Software and Technology, pp. 195~204, 1998.
- [9] Ramesh Raskar, Jeroen van Baar, Paul A. Beardsley, Thomas Willwacher, Srinivas Rao, Clifton Forlines, “iLamps: Geometrically Aware and Self-configuring Projectors,” ACM Trans. Graph. pp. 809~818, 2003.
- [10] Van Baar J., Willwacher T., Rao S., Raskar R., “Seamless Multi-projector Display on Curved Screens,” Proc. Eurographics Workshop on Virtual Environments, 2003.
- [11] Toshio Moriya, Fumiko Beniyama, Kei Utsugi, Tsuyoshi Minakawa, Haruo Takeda, Kenji Ando, “Multi-camera and Multi-projector based Seamless Live Image Display System,” Proc. 10th International Conference on Multimedia Modelling, pp. 265~272, 2004.
- [12] Cruz-Neira, Sandin, DeFanti, “The Design and Implementation of the CAVE,” Proc. Surround-Screen Projection-Based Virtual Reality, 1993.
- [13] Czernuszenko, Pape, Sandin, DeFanti, Dawe, Brown, “The ImmersaDesk and Infinity Wall Projection-Based Virtual Reality Display,” Computer Graphics, 1997.
- [14] 한국디스플레이연구조합, <http://www.edirak.or.kr>
- [15] Marc Pollefeys, Luc J. Van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, Reinhard Koch, “Visual Modeling with a Hand-Held Camera,” International Journal of Computer Vision Vol. 59, No. 3, pp. 207~232, 2004.
- [16] R. Zhang, P. Tsai, J.E. Cryer and Mubarak Shah, “Shape from Shading: A Survey,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 21, No. 8, pp. 690~706, 1999.
- [17] A. Laurentini, “The Visual Hull Concept for Silhouette based Image Understanding,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2 pp. 150~162, 1994.
- [18] Takashi Matsuyama, Xiaojun Wu, Takeshi Takai, Shohei Nobuhara, “Real-time 3D Shape Reconstruction, Dynamic 3D Mesh Deformation, and High Fidelity Visualization for 3D Video,”

Computer Vision and Image Understanding, Vol. 96, No. 3, pp. 393~434, 2004.

[19] Kazuhiko Sumi, Koichi Tanaka, Takashi Matsuyama, "Measurement of Human Concentration with Multiple Cameras," Proc. 9th International Conference on Knowledge-Based Intelligent Information & Engineering Systems, pp. 129~135, 2005.

[20] de Aguiar, E., Theobalt, C., Magnor, M., Theisel, H., Seidel, H.-P., "M³: Marker-free Model Reconstruction and Motion Tracking from 3D Voxel Data," Proc. 12th Pacific Conference on Computer Graphics and Applications, pp. 101~110, 2004.

[21] Cheung G, Baker S, Kanade T, "Shape-from-silhouette of Articulated Objects and Its Use for Human Body Kinematics Estimation and Motion Capture," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 77~84, 2003.



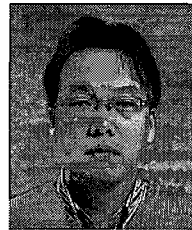
경 동 욱

- 2002년 동의대학교 산업공학과(공학사)
- 2005년 부산대학교 컴퓨터공학과(공학석사)
- 2005년~현재 숭실대학교 IT대학 미디어학부 미디어 공학(박사과정)
- 관심분야 : HCI, 영상처리/컴퓨터 비전, 하드웨어 설계, 인공지능, 정보보호



한 은 정

- 2003년 신라대학교 국어국문학과(인문학사)
- 2005년 동서대학교 소프트웨어대학원 소프트웨어학과(공학석사)
- 2005년~현재 숭실대학교 IT 대학 미디어학부 콘텐츠 공학(박사과정)
- 관심분야 : 모바일 콘텐츠, E-Learning, M-Learning, HCI, CBIR, 온톨로지



양 종 열

- 1993년 경북대학교 컴퓨터공학과(공학사)
- 1995년 경북대학교 대학원 컴퓨터공학과(공학석사)
- 1997년 포항공대 정보통신 연구소 위촉 연구원
- 2005년~현재 숭실대학교 IT대학 미디어학부 미디어 공학(박사과정)
- 2005년~현재 ㈜ 에이티솔루션 연구원
- 관심 분야 : 게임, AI, 영상처리



정 기 철

- 1994년 경북대학교 컴퓨터공학과(공학사)
 - 1996년 경북대학교 대학원 컴퓨터 공학과(공학석사)
 - 1999년 *Intelligent User Interfaces group at DFKI (The German Research Center for Artificial Intelligence GmbH), Germany*, 방문연구원
 - 2000년 *Machine Understanding Division, Electro Technical Laboratory in Japan*, 방문연구원
 - 2000년 경북대학교 대학원 컴퓨터공학과(공학박사)
 - 2000년~2002년 *PRIP lab, Michigan State University* Postdoc
 - 2003년~현재 숭실대학교 IT대학 미디어학부 교수
 - 관심분야 : HCI, 콘텐츠공학, 인터랙티브 게임, 영상처리/컴퓨터 비전, 증강현실, 인공지능
-
-