

동작 캡처 기술

강 호 석

삼성 전자 소프트웨어 센터

I. 서 론

최근 들어 컴퓨터의 성능이 급격히 향상되고, 3차원 게임이나 interactive application, 인터넷을 사용한 다자 참여형 가상 세계에 대한 응용이 급증하면서, 가상 환경에 대한 연구가 많이 진행되고 있다. 그 동안 가상환경의 사용자에게 가상환경에 완전히 몰입할 수 있도록 멋있고 사실적인 이미지를 보여주기 위해서, 원하는 화면을 효과적으로 표현하는 scene description 분야와 이 description을 실제 이미지로 변환하는 rendering 분야의 연구가 진행되어 왔다. 그 중 rendering 분야는 다양한 shading algorithm과 ray tracing 기법, hardware의 고성능화 등에 힘입어 많은 발전을 이루어 왔지만, scene description 분야는 아직도 많은 연구가 필요한 상태이다. 특히 가상환경에서 사용자를 대신해서 등장하는 캐릭터 (avatar 혹은 virtual human이라고 부름)들의 움직임을 묘사하는 기술은 아직 초보적인 단계이다.

현장감과 사실감을 주기 위해서는 변화하는 image를 1초에 적어도 10 frame 이상, 24 frame이나 30 frame씩 연속적으로 보여 주어야 하기 때문에, 움직이는 화면을 적절히 기술하는 것이 아주 중요하다. 특히 사람이나 동물과 같은 다관절체의 캐릭터는 표현해야 할 움직임의 양이 많고, 더욱이 자연스러운 동작을 기술해야 하므로 무척 어려운 문제이다. 이러한 움직임을 기술하기 위해서 key framing, kinematics, dynamics, constraints, functional modeling, space-time optimization 등의 기법을 사용하는 연구가 계속 진행되고 있으나

아직 실시간에 사용자가 원하는 만큼의 자연스러운 동작을 만들어 내지는 못하고 있다.

최근 이런 문제들을 해결하기 위한 새로운 접근법으로 동작 캡처를 이용하는 연구들이 진행되고 있다. 동작 캡처는 실제 사람의 움직임을 추적하여 얻어낸 데이터를 컴퓨터 내부에 모델링된 캐릭터에 적용하여 유사한 움직임을 얻어내는 기술이다. 이를 위하여 자기 센서나 광학 센서들을 몸에 부착하고 동작을 하면서 각 센서의 위치와 각도를 기록하는데, 이렇게 동작 캡처로 만든 데이터가 있으면, 원하는 때에 기록된 동작을 재현할 수 있어서 복잡한 캐릭터의 동작을 자연스럽게 효율적으로 구현할 수 있다.

본 논문에서는 동작 캡처와 이를 위한 다양한 장비, 동작 캡처를 사용한 animation 기법과 실제 활용 예, 기술적인 문제점들에 대하여 간단히 설명한다.

II. 동작 캡처

동작 캡처는 물체의 움직임을 추적하여 얻어낸 위치 데이터를 컴퓨터 내부에 모델링된 캐릭터에 적용하여 유사한 동작을 하도록 하는 기술이다. 즉 실제 공간에서 사람이나 동물의 몸, 얼굴 표정, 카메라, 조명 등의 위치와 방향을 측정해서 컴퓨터에서 사용할 수 있는 형식으로 기록을 한 후, 기록된 데이터로 물체들의 움직임을 제어하여 원하는 장면을 컴퓨터로 만든다. 이 방법은 영화나 컴퓨터 게임, 가상 환경 등에서 사용할 동작을 미리 기록

하여 만들거나, 실시간에 캐릭터를 움직이는 performance animation에 사용된다. 이 기술을 사용하면 자연스러운 동작 데이터를 비교적 빠른 시간에 손쉽게 얻을 수 있고, 일단 캡처한 동작 데이터는 데이터베이스화하여 나중에 재 사용하는 것이 가능하므로, 전 세계적으로 급성장하고 있다.

위치와 방향을 측정하는 방법으로는 기계적 방식, 음향학적 방식, 관성 방식, DGPS(Differential Global Positioning System), 자기 방식과 광학 방식 등이 사용되는데, 동작 캡처에는 주로 자기 방식이나 광학 방식이 사용된다. Analogus사¹⁾의 Animotion Suit (그림 1 (a))나 Digital Image Design사²⁾의 Monkey2는 mechanical linkage를 움직여 potentiometer의 변화량을 측정하는 기계적 방식을 사용한다. 음파를 발사한 후 삼각측량법으로 위치를 계산하는 음향학적 방식은 이제는 거의 사용되지 않는다^[1]. 최근 들어 개발된 Angularis Inertial Technologies사³⁾의 VR360은 자이로스코프와 가속도계를 이용한 관성 추적장치를 사용하여 방향을 아주 정밀하게 측정할 수 있다. 4개의 위성과 2개 이상의 GPS 수신기를 사용하는 DGPS로는 약 60cm이내의 오차를 가진 위치 정보를 얻을 수 있다.

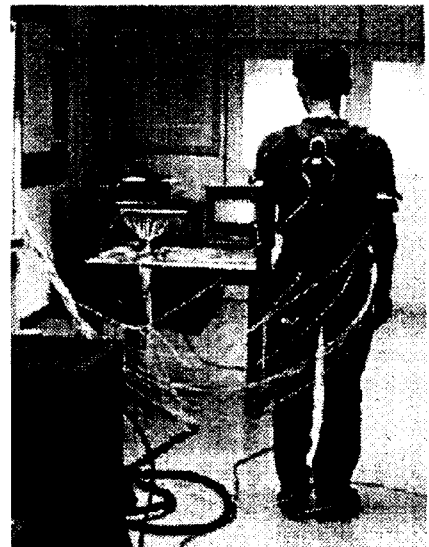
자기 방식은 자기장의 변화를 사용하여 동작 데이터를 측정하는 방식이다. 즉 행위자의 관절부분에 x, y, z 방향의 자장을 측정하는 수신기를 부착하고, x, y, z 방향의 직교 코일을 내장한 발신기에서 발생하는 자장 속에서 몸을 움직이면서, 수신기에 유도된 전류로부터 수신기의 3차원 위치나 방향 등의 데이터를 계산한다. 이 방식은 3차원 좌표 데이터의 흐름을 실시간에 출력해 주므로 실시간 응용에 매우 유리하다. 그러나 사람의 몸에 수신기를 부착하기 위한 많은 케이블이 연결되어 (tethering) 동작을 제한하는 단점이 있다 (그림 1 (b)). 최근에는 이런 줄 엉킴 문제를 해결하기 위하여 무선 형태의 시스템도 개발이 되었다.

광학 방식은 두 대 이상의 카메라를 사용하여

이미지를 수학적으로 처리하여 위치 데이터를 측정하는 방식이다. 많은 수의 작은 반사 marker를 행위자의 관절에 부착하고 (그림 1 (c)) 두 대 이상의 고품질 비디오 카메라를 사용하여, marker의 다른 배경 등은 필터링을 하여 삭제하면서 행위자의 동작을 녹화한다. 각각의 이미지에서 비디



(a) Analogus사의 animotion Suit

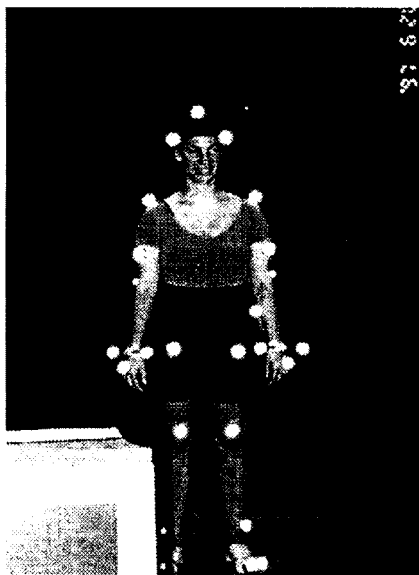


(b) Ascension Technology사의 Motion Star

1) <http://www.analogus.com>

2) <http://www.didi.com>

3) 미국 전화 617-621-1563



(c) Motion Analysis사의 Expert Vision

(그림 1) 동작캡처 장비를 행위자가 착용한 모습
 오 프로세서가 각 marker에 대한 2차원 좌표 데이터들을 계산한 후, 그 값들로부터 각 marker에 대한 3차원 운동 궤적을 계산한다. 이 방식의 가장 큰 단점은 동작할 때 marker가 때때로 가려질 수

있다는 점이다. 이 문제는 전후 frame 이미지의 marker를 참조하여 계산한다. 그러나 이런 데이터 보정과 geometry processing을 하는 post processing 시간이 오래 걸리므로, 실시간의 응용에는 사용할 수 없다.

표 1과 표 2에 자기 방식과 광학 방식 시스템의 장단점과 사양을 비교하였다. 자기 방식을 사용하는 대표적인 시스템은 미국의 Ascension Technology사⁴⁾의 FOB, MotionStar와 Polhemus 사⁵⁾의 FASTRAK, ULTRATRAK PRO 등이다. 광학 방식을 사용하는 시스템은 미국의 Motion Analysis사⁶⁾의 ExpertVision, AdaptiveOptics사⁷⁾의 MultiTrax, Oxford Metrics사⁸⁾의 Vicon 370E 등이 대표적이다.

그림 2에 보인 Ascension Technology사의 최신 제품인 MotionStar는 ERT(Extended Range Transmitter), ERC(Extended Range Controller), 여러 개의 수신기 card, 그리고 486DX2-66MHz cpu card를 장착한 Computing Unit 등의 module 과 특수 control을 위한 부가적인 keyboard 및 mouse, monitor로 구성된다. 하나의 Computing Unit에 최대 17개의 수신기 card를 부착할 수 있

(표 1) 자기 방식 시스템과 광학 방식 시스템의 비교

	자기 시스템	광학 시스템
장점	실시간 처리가 가능하다.	자유로운 동작이 가능하다.
	가격이 상대적으로 싸다.	가격의 증가 없이 센서의 수를 늘릴 수 있다.
		캡처 범위가 넓다.
		빠른 동작을 캡처할 수 있다.
단점	금속의 영향을 받는다.	조명의 영향을 받는다.
	Noise가 많고 calibration이 힘들다.	데이터를 처리하고 보정하기 위한 post-processing이 필요하므로 실시간 처리가 힘들다.
	캡처 범위가 좁다.	가격이 비싸다.
용도	TV, 광고, animation, PA	움직임이 격렬한 격투기 등의 동작 캡처
	동작의 범위가 제한된 게임	골프, 테니스, 야구 등 움직임이 빠른 프로스포츠 분야의 bio-mechanical 연구
	산업, VR 시스템, LBE	이동 범위가 넓은 응용 분야

4) <http://www.ascension-tech.com>
 5) <http://www.polhemus.com>
 6) <http://www.motionanalysis.com>

7) <http://www.aoainc.com>
 8) 미국 지사 310-903-1219

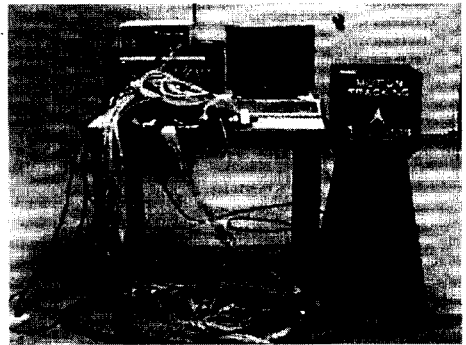
〈표 2〉 대표적인 동작 캡처 시스템들의 사양

시스템	MotionStar	UltraTrak Pro	ExpertVision
방식	Pulsed DC 자기장	AC 자기장	광학
범위	10 ft	15 ft	25 ft
위치 정밀도	0.6인치 RMS at 10 ft	3 인치 RMS at 15 ft	
방향 정밀도	1 도 RMS at 10 ft	3 도 RMS at 15 ft	
최대 센서수	120	32	100
Data rate (fps)	144	120 for 16 receivers, 60 for 32 receivers	240
가격	약 \$ 40,000		\$ 80,000 ~ \$ 250,000
비고	각 센서 당 추가비용 약 \$ 2,500		가격에 상관없이 센서의 수를 늘릴 수 있음

고, 여러 대의 Computing Unit들을 서로 연결할 수도 있다. 기존의 FOB 제품과 구별되는 특징은, 개개의 수신기 신호 처리 unit을 Computing Unit 안에 장착할 수 있는 확장 card 형태로 제공하여 설치가 용이하고, Ethernet interface option을 제공하여 Ascension 사의 BIRNET protocol을 통한 client-server mode로 동작할 수 있다는 점이다. FOB와 MotionStar 모두 x, y, z 위치 좌표와 Euler angle, matrix, quaternion 등을 출력할 수 있다.

이런 자기 방식의 장비를 사용하려면, 원칙적으로는 자장의 영향을 받지 않는 약 5m x 5m x 5m 크기의 공간이 필요하다. 시스템을 설치할 때 송신기와 수신기간의 calibration 기술과, computer와 장비간의 interface 기술이 필요하다. 일단 시스템이 설치된 후에는 여러 가지 noise source, 즉 장비 자체, 동작 실행자, 주변 환경 등으로부터 발생하는 noise와 error를 적절히 filtering 해내는 기술이 요구된다.

동작 캡처 시스템은 아직은 상당히 비싼 편이지만, PC를 사용하여 몸 동작이나 얼굴 표정 캡처를 하는 저가형 시스템들도 개발되고 있다. 또한 진보된 컴퓨터 비전 및 새로운 기술이 개발되어 기존의 광학 방식의 단점이 보완되고 있다. 예를 들어,



〈그림 2〉 Ascension Technology사의 자기방식 동작 캡처 장비인 MotionStar로 좌측 상단에 송신기 Controller와 Computing Unit, 하단에 케이블로 연결된 수신기들, 중앙에 모니터와 키보드, 우측에 있는 것이 송신기이다.

Lawrence Livermore National Laboratory⁹⁾의 CyberSight는 동작 캡처를 하면서 실시간에 움직이는 물체의 3차원 모델도 구성할 수 있다. 이 시스템은 두 대의 CCD 비디오 카메라로부터 입력된 스테레오 이미지로부터 3차원 데이터를 추출한다. 아직 연구단계에 있으나, 이 기술을 이용하면 실제 사람과 똑같이 움직이는 digital clone을 생성할 수 있다. CompInt¹⁰⁾의 Photo4D-Pro도 두 대 이상의 카메라로부터 입력된 비디오를 이용해 marker를

9) <http://www.llnl.gov>

10) <http://www.igs.net/compint>

사용하지 않고 동작 캡처를 할 수 있다. 여기서는 사용자가 처음 frame에 원하는 feature point들을 표시하면, 그 feature point들을 연속된 frame에서 추적하고 3차원 데이터를 추출한다^[2]. 이 외에 Northern Digital사^[11]의 OPTOTRAK은 active marker (적외선 발광 diode)를 이용하여 실시간에 동작 캡처가 가능하다고 한다.

또한 Acclaim Motion Capture Studios^[12]나 BioVision^[13], 프랑스의 Medialab^[14] 등에서는 동작 캡처 시스템과 스튜디오를 대여하거나 동작 캡처를 대행해주며, Viewpoint Datalabs^[15]나 Boston Dynamics사^[16] 등에서는 걷는 동작, 뛰는 동작, 물건을 던지는 동작 등의 기본적인 동작 데이터를 판매하기도 한다.

III. 동작 캡처의 활용

Performance Animation (PA, 혹은 motion-capture animation) 이란, 행위자의 동작을 캡처하여 이를 실시간에 캐릭터의 animation으로 표현하는 것을 말한다. 이렇게 하면 행위자와 animator가 피드백을 주고받을 수 있어서 동작을 즉시 수정할 수 있고, 이런 실시간 animation을 녹화하거나 생방송으로 방영하는 것이 가능하다.

이를 위하여 동작 캡처 장비나 얼굴 표정 캡처 장비, data glove 등을 사용하여 행위자의 동작을 캡처 하는 동시에 고성능 그래픽 workstation을 사용하여 이를 캐릭터의 움직임으로 display한다. 이때 보조자(puppeteer, 꼭두각시 조정자)가 페달, joystick 등을 사용하여 눈동자와 입술모양 등을 조종하기도 한다. 이렇게 캐릭터를 animate 하면서, 컴퓨터 그래픽으로 만든 다양한 배경이나

video와도 합성을 할 수 있는데, 이 모든 과정을 총괄해 주는 SW 시스템으로는 프랑스 Medialab의 CLOVIS PA, Protozoa^[17]의 ALIVE, SimGraphics^[18]의 VActor 등이 있다. 이를 사용하면 생방송이나 놀이동산, 가상세계, virtual studio 등에서 가상 캐릭터가 실제 사람들과 상호작용 할 수 있고, 많은 양의 animation을 효율적으로 빨리 제작할 수 있다.

프랑스의 유선방송회사인 CANAL+^[19]의 자회사인 Medialab은 PA분야에서 가장 뛰어난 기술을 보유한 회사의 하나로^[3], 고품질의 캐릭터 animation을 실시간에 생방송하고, 매일 10분씩 1주에 5회 방영되는 Cyberflash (Cleo라는 이름의 가상 캐릭터가 진행하는 컴퓨터 그래픽으로 제작된 host show), 매회 23분씩 26회가 방영되는 Donkey Kong Country 시리즈 등을 제작하고 있다. Cyberflash의 제작과정을 간단히 설명하자면, 우선 목요일과 금요일에 다음 번에 방영할 1주일분의 스크립트와 스토리-보드를 작성하고, 월요일에 Cleo의 목소리만을 먼저 녹음한다. 화요일 하루 동안 PA를 사용하여 Cleo의 움직임을 기록하고, 수요일에 필요한 post processing을 하여 video를 제작하여 다음주에 방영한다. 기존의 2D animation 방식을 사용하면 여러 달이 걸려야 제작할 수 있는 50분 분량의 90000 frame (50분 * 60초 * 30장)을 5명의 staff (감독 1명, PA 엔지니어 1명, 행위자 1명, puppeteer 2명)이 5일만에 제작한다. Donkey Kong Country의 경우, 모두 11 캐릭터가 등장하는 10시간 분량의 컴퓨터 그래픽 animation인 900000 frame 중 92%를 PA로 제작한다. 즉 1회분 23분을 PA를 사용하여 약 10일만에 제작한다^[4].

Medialab에서 PA를 하는 과정은 다음과 같다. 우선 시나리오에 따라 등장 캐릭터들을 모델링한

11) <http://www.ndigital.com>
 12) <http://www.acclaimnation.com>
 13) <http://www.bio-vision.com>
 14) <http://www.demon.co.uk/mlab>
 15) <http://www.viewpoint.com>

16) <http://www.bdi.com>
 17) <http://www.protozoa.com>
 18) <http://www.SimG.com>
 19) <http://www.cpl.us.fr>

후, 상세한 스크립트를 작성하고 미리 음성을 녹음한다. 가능한 자장의 영향을 받지 않도록 넓은 공간에 나무로 사방 약 4m, 높이 1m 정도의 스테이지를 만들고, 그 위에서 미리 녹음한 음성에 맞추어서 동작캡처를 한다. 이때 행위자의 동작과 puppeteer의 control이 실시간에 캐릭터의 움직임으로 3차원 animation이 되는 것을 볼 수 있도록, 주위에 6개의 모니터를 설치한다. 스테이지 끝에 2명의 puppeteer가 있는데, 한사람은 joystick을 사용하여 눈동자의 방향을, 왼손 data glove를 사용하여 얼굴 표정을 조절하고, 다른 사람이 오른손 data glove를 사용하여 lip-sync (목소리에 맞춰서 입술을 움직이는 것)를 한다 (그림 3 (a)). 스테이지를 마주보며 두 사람이 있는데, 한사람은 PA 프로그램을 사용하는 엔지니어이고, 다른 사람은 총 감독으로 전체적인 지휘를 한다. 감독과 행위

자, puppeteer가 서로 feedback을 받으며 수정하여, 감독이 만족하는 일련의 동작이 얻어질 때까지 행위자의 동작 데이터와 puppeteer들의 control 데이터를 record하고 재생해 본다. 그 후, 기록된 데이터를 바탕으로 PA 엔지니어가 배경, 조명, 캐릭터, 의상, 카메라 앵글 등을 조정한 후 캐릭터의 동작 scene을 다시 실시간에 rendering 한다 (그림 3 (b)). 여기에 video나 image와 영상 합성을 하고, 필요한 post processing을 하여 얻어진 최종 animation을 PA의 video out option을 사용하여 video tape에 녹화한다.

IV. 기술적 문제점

동작캡처와 이를 사용한 PA 등은 가상현실이나 실시간 animation을 하기 위한 하나의 접근 방법이다. 빠른 시간에 자연스러운 동작을 제공해 주고 상호작용하며 수정할 수 있어 기존의 문제점의 일부를 해결해 주지만 이 또한 자체적으로 많은 문제점을 안고 있으므로 기존의 방법과 상호 보완적으로 사용되고 있다.

PA를 하기 위해서는 실시간에 동작 데이터를 수집하는 기술과, 실시간에 고성능 rendering을 하는 기술이 필요하다. 동작 캡처를 위한 장비의 개발, calibration, 데이터의 noise 처리 등의 기술과, 기존의 캐릭터 animation 기술들 즉 충돌 검사, 동역학, 얼굴표정, 입 모양이나 관절과 피부의 deformation 등을 처리하는 기술도 있어야 하며 초고성능의 그래픽 workstation도 있어야 한다. 원하는 캐릭터의 concept에 따라 이를 디자인하고 제대로 움직이도록 모델링 한 후, 능력 있는 행위자의 동작을 캡처하고 얼굴 표정을 처리하고 입술 모양을 맞춰 주고, 실시간에 처리할 수 없는 특수 효과와 여러 가지 제작상의 문제점을 post-production으로 처리하는 PA^[5]를 하기 위해서는 3차원 컴퓨터 그래픽 분야와 영상 제작 분야의 모든 기술이 필요하다.

그 중 동작캡처의 가장 큰 issue는 캡처한 동작



(a) 가운데의 행위자가 자기센서를 몸에 부착하고 전체 몸동작을 캡처하고, 좌, 우에 앉은 puppeteer들의 data glove와 joystick, pedal을 사용하여 얼굴 표정과 입 모양 등을 조절한다.



(b) Cyberflash의 주인공 Cleo

<그림 3> PA를 위한 동작 캡처 session의 예

데이터의 응용성이다. 즉 특정 신체구조를 가진 행위자의 특정 동작을 캡처하여 특별히 제작한 캐릭터에 적용하였는데, 과연 이 데이터를 다른 용도로도 사용할 수 있는냐는 점이다. 그렇지 못하다면 동작캡처는 그 활용가치가 떨어질 것이다. 이를 해결하기 위해서 다음과 같은 연구들이 진행되고 있다.

특정 신체 치수를 가진 행위자의 동작을 신체 치수가 많이 차이 나는 캐릭터에 적용하는 연구는 많이 진행되어 왔다. 캡처 프로그램에서도 어느 정도의 scale 문제를 지원하고, 각 관절의 각도 비율을 유지하거나, 특정 위치를 지나도록 역운동학을 사용하여 수정해 주기도 한다. 그러나 이를 근본적으로 해결하기 위해서는, 특정 동작의 특성을 추출하고 일반화한 후 그 특성을 유지하면서 변환해주는 연구가 필요하다.

미리 캡처하여 라이브러리로 구축한 동작 데이터의 수는 결국 유한하고, 사용자가 원하는 동작은 미묘한 차이를 감안하면 무한히 많기 때문에, 기존의 동작 데이터를 가공하여 원하는 자연스러운 동작을 새롭게 실시간에 만드는 연구도 필요하다. 이를 위해서는 어떤 특정 동작들을 연결하거나 합성, 변형하여 이로부터 새로운 동작을 만들 수 있어야 한다.

V. 결 론

본 논문에서는 가상환경에 등장하는 캐릭터의 움직임 묘사하는 기술 중의 하나인 동작 캡처와 각종 장비, 이들의 활용 예, 기술적인 문제점들과 향후 연구 방향에 대하여 간단히 설명하였다. 아직까지 동작캡처는 주로 특정 캐릭터의 자연스러운 특정 동작을 얻기 위하여 사용되어 왔다. 이렇게 얻어진 동작 데이터가 있을 때, 그 동작의 특성을 추출하여 일반화하고, 변형, 합성, 생성하는 이론적인 바탕이 완성되고, 여기에 사용자가 원하는 동

작을 자유롭게 표현하는 입력방법과, 특정 목표를 이루기 위하여 일련의 동작을 조합하는 동작 계획 알고리즘이 가미되면, 가상세계에서 가상 캐릭터의 자연스러운 움직임을 마음대로 조정할 수 있을 것이다.

참 고 문 헌

- [1] Motion Capture, EDGE magazine, p.58 ~ p.65, 9월 1995년
- [2] George Maestri, Making Characters Move, Digital Magic, p.31 ~ p.34, 8월, 1996년
- [3] Adele Hars, Masters of Motion Capture, Computer Graphics World, p.26 ~ p.34, 10월, 1996년
- [4] Motion Capture in Practice, SIGGRAPH 97 Course Note 1, 8월, 1997년
- [5] David Sturman, Computer Puppetry, IEEE Computer Graphics and Applications, p.38 ~ p.45, 1월, 1998년

저 자 소 개



姜 豪 爽

1962年 10月 1日生, 1985年 2月 서울대학교 전기공학과 졸업, 1988年 5月 미국 University of Florida 전기전자공학과 석사학위, 1994年 12月 미국 Purdue University 전기전자공학과 박사학위, 1995年 4月~1998年 2月 삼성전자 소프트웨어 센터 선임연구원, 주관심 분야: 컴퓨터 그래픽스, 가상현실, 실시간 Animation, Human-Computer Interaction, Multimodal Human Interface, Computer Vision 등