특집

제스쳐 인식기술 및 응용

閔 丙 雨
시스템 工學硏究所

I. 서 언

제스쳐는 말의 표현이나 전달 효과를 높이기 위해 하는 몸짓, 손짓, 표정 따위라고 사전에 정의되어 있듯이 사람 사이의 의사 전달에서 제스쳐는 보조적인 수단으로서 매우 큰 비중을 차지하고 있으며, 청각 장애자를 위한 수화(sign language)의 경우는 제스쳐가 유일한 의사소통의 수단으로 활용되기도 한다. 최근에 사람과 컴퓨터와의 접속기술(HCI : human-computer interaction)은 인공지능 기술을 주축으로 여러 관련 기술이 접목되어 사람들이 대화하는 것과 유사한 기술의 개발에 중점을 두고 있다. 따라서, 자연스러운 환경에서 대화가 이루어 질 수 있도록 하기 위한 시각화(visualization), 대화의 과정이 유연하며 적응력을 갖추기 위한 지능화(intelligent), 각 사용자의 수준과 취향을 맞추어 나가기 위한 개인화(personalization)에 촛점이 맞추어져 연구되고 있으며, 입력 정보도 음성정보, 필기문자 정보, 제스쳐 및 얼굴영상 정보 등으로 다양화 되고 있다[1].

본 논문에서는 의사 소통의 보조수단으로 활용되는 제스쳐를 컴퓨터와 사람과의 사용자 접속기술로서 활용하기 위한 연구들에 대해 기술 현황과 응용에 대해서 살펴보고자 한다. 제스쳐의 동작 부위는 몸 전체라고 할 수 있으나, 아직은 머신비젼이 지니는 기술의 한계성 때문에 간단한 손 제스쳐를 인식하여, 이를 사용자 접속기술로서 활용하기 위한 연구가 주를 이루고 있으며 본 논문도 여기에 촛점을 맞추어 작성하였다. 여러가지 이유로 내용이 불충분한 부분이 많을 것으로 생각되며, 자료의 대부분은 95년 6월에 스위스 쮸리히에서 개최되었던 제1회 얼굴 및 제스쳐 인식 국제 학술회의(intrenational workshop on automatic face-and gesture-recognition)의 논문집에서 발췌하였다[2]. 구성은 1 장의 서언에 이어 2 장에서는 제스쳐의 모델링에 대해, 그리고 3 장에서는 제스쳐 인식 및 응용에 대해서 살펴보았다. 4 장에서는 제스쳐 합성에 대해 알아본 뒤에, 5 장에서 글을 맺는다.

II. 제스쳐의 모델링

S. Thieffry가 "모든 제스쳐는 정신적인 사고의 물리적인 표현이다"라고 말한 것처럼 제스쳐는 우리의 일상 생활에서 보편적 의미를 전달하는 중요한 수단이다. 또한, 제스쳐는 각 지역별로 문화적 특성에 의해 많은 차이가 있어 서로 다른 표현방식을 갖기 때문에, 제스쳐와 관련된 기술 개발은 문화적 환경에 바탕을 두어야 함을 알 수 있다. 지금까지 제스쳐의 모델링에 관한 연구는 인지과학적 측면에서 사람의 시각인식 기능이 어떻게 신체의 움직임을 이해하는가에 대한 연구와, 주로 손 제스쳐 영상을 컴퓨터 상에서 해석하기 위해 어떻게 모델화 할 것인가에 관한 연구가 있어 왔다.

1. 몸 제스쳐(body gesture) 모델링

사람의 몸동작을 인식하는 연구는 인지과학적 측면에서 접근이 시작되었다. Johanson은 1973년에 모션을 인식하는 실험을 행하기 위해 암실에서 인체의 연결부에 작은 전구들을 몸의 윤곽선 부분에 부착하고 움직였을 때, 사람이 이를 어떻게 인지할 것인가에 관하여 조사하였다[3]. 결과는 매우 극적인 면을 보여 주었다. 움직이지 않은 상태에서는 작은 광원들의 집합으로 보였으나, 움직였을 때는 인체의 움직임이 쉽게 인지되었다. 이 실험은 후에 MLD(moving light display) 구조분야의 초석이 되었다. 이러한 연구를 바탕으로 하여 Herman은 사람의 몸통을 16 개의 구성요소와 각 구성요소를 연결하는 13 개의 연결부로 분리하고 이를 2차원으로 사상(mapping)하여 신체언어를 해석하는 연구를 수행하였다. 1980년 경에 O' Rourke와 Badler는 컴퓨터에서 인공적으로 생성하는 이미지에 의해 사람의 움직임을 현실감 있게 보이게 하기 위한 연구를 수행하여, 신체의 움직임을 기록에 의해 비젼시스템에서 이를 재현할 수 있게 하는 중요한 개념을 발전 시켰다. 그리고, Lee와 Chen은 하나의 시각으로 부터 사람 몸의 모든 연결 부분의 삼차원 위치를 재구성하는 연구를 수행하였으며, Tsukiyama와 Shirai는 복도에 설치된 카메라로 부터 입력되는 실세계의 영상으로 부터 복도를 통과하는 사람의 형체를 포착하여 추적하는 시스템을 개발하였다. Akita는 많은 제약 조건을 두기는 하였으나 키 프레임(key frame)과 키 피쳐(key feature)를 이용하여 신체의 움직임을 분석을 할 수 있는 시스템을 개발하였으며, Leung과 Yang은 신체 움직임의 연속 프레임에서 변화가 발생하여 움직이는 부분과 변화가 없이 움직이는 부분(log)을 인식하여 동작을 이해하는 Log-Tracker 시스템을 개발하였다[4].

2. 손 제스쳐(hand gesture) 모델링

손 동작은 의도하는 바를(예 : 지시, 거절, 이해 등) 실현하기 위한 동작이다. 처음의 의도로 부터 동작을 마칠 때 까지 제스쳐는 공간과 시간 상에서 특징적인 동적 패턴을 따른다. Kendon은 하나의 제스쳐를 세개의 동작 즉, 준비(preparation), 스토로크(stroke) 그리고 복귀(retraction)로 구분하였다. 이러한 구분은 언제 어디에서나 적용될 수 있는 것으로 어떻한 특정한 손동작을 기술할 수 있다. Quek은 손동작의 패턴에 바탕을 둔 손동작의 분리(segmentation)을 위해서 다음과 같은 규칙을 개발하였다[5]. 첫째, 제스쳐는 정지 상태로 부터 느린 속도로 움직이기 시작하여, 속도가 점차적으로 증가되는 스토로크의 상태를 지나, 다시 정지 상태로 되면서 끝을 맺는다. 둘째, 손은 스토로크 단계에서 특정한 모양이 가정된다. 셋째, 정지 상태에서의 저속의 움직임은 제스쳐가 아니다. 넷째, 손동작에 의해 만들어 지는 제스쳐는 일정한 공간 영역에서 형성된다. 다섯째, 정적인 제스쳐는 인식을 위해 일정한 시간이 요구된다. 여섯째, 반복적인 움직임도 제스쳐가 될 수 있다. 또한 그는 몇가지의 예외는 있지만 손가락의 움직임은 손이 정지 상태에 있을 때 의미를 갖음을 제안하였다. 정적인 상태에서의 손동작은 특정의 손가락, 엄지손가락, 손바닥에 의해 구성되는 손의 모양에 의해 결정된다. 그리고, 동적인 손 동작은 처음과 마지막 스토로크의 손 모양과 일반적인 스토로크 동작에 의해 이루어 진다. 손 모양은 움직이는 동안 변

할 수 있으나, 그 변화는 어떠한 제스쳐도 포함하지 않기 때문에 무시될 수 있다. 과거에 개발되었던 제스쳐 분석 시스템의 대부분은 단순하고도 정적인 모델을 사용하였으나, 최근은 더 복합적인 모델을 통하여 제스쳐의 모든 가능성을 탐구하고 있다.

III. 손 제스쳐 인식과 응용

위에서의 기술한 바와 같이 손 제스쳐의 인식은 단순하지 않음을 알 수 있다. 제스쳐 인식은 시간적 및 공간적인 분석을 포함해야 하며, 이에 더하여 아무리 단순한 제스쳐라도 손의 영상을 배경의 움직임 또는 정지하고 있는 물체와 분리하는 것도 매우 어려운 작업이다. 또한, 궁극적으로는 손, 가상의 물체 그리고 음성사이의 상호작용도 허용되어야 우리가 좀더 지능적으로 느낄 수 있는 사용자 접속이 가능하다. 손 제스쳐를 분석하는 데는 크게 세가지 부류 즉, 데이타 글로브를 낀 상태의 제스쳐 인식(glove-based), 자연스런 제스쳐의 인식(vision-based), 필기 상태의 제스쳐(drawing gesture) 인식으로 분류할 수 있으며, 각각에 대해 상세한 내용을 살펴 보았다.

1. 글로브(glove-based) 신호에 의한 분석

데이타 글로브를 낀 상태의 손 제스쳐 인식은 1970 년대 후반 부터 시작되었다. 글로브는 손가락을 구부리거나 핀 상태의 손 모양을 전기적 신호로 변환하기 위해 기계적 혹은 광(optical) 기구가 부착되어 있다. 지금까지 널리 사용되는 글로브 중의 하나는 VPL 연구소에서 개발된 *DataGlove*로서 굴곡을 감지하기 위해서는 광섬유 기술을 사용하고, 위치 추적을 위해서는 마그네틱 센서를 활용한다. *DataGlove*가 허용하는 자유도는 총 16 개로서 10 개는 굴곡에 관한 것이고 나머지 6 개는 위치 정보에 대한 것이다. 이를 이용한 시스템들을 살펴보면, Sturman 등은 가상환경 인터페이스에 *DataGlove*를 활용하였다. 이들은 손의 동작기능을 button, valuator, locator, pickup device로 분류하였다. Cordella등은 손 제스쳐, 음성, 소리, 스트레오 스코프 그래픽, 머리 움직임을 종합한 복수 사용자 가상세계(a multi user *Virtual World*)에 활용하였다. 이 시스템은 두 사용자가 실시간으로 생성하고, 잡을 수 있으며, 때릴 수도 있는 유연 물체를 포함하는 방을 가상적으로 만드는 데 사용하였다. Fels와 Hinton은 *Glove-Talk*라는 시스템을 개발하여 사용자의 손 제스쳐와 음성 합성기 사이의 접속기로 활용하였으며, 신경망을 이용하여 203 개의 제스쳐를 단어사전에 대응시킴으로서 완전한 단어로 정합시켰다. 그리고, 글로브에 기초한 시스템 중 유명한 하나는 Baudel등이 개발한 *Charade*가 있다[6]. 이 시스템에서는 하이퍼텍스트 프레젠테이션 용의 브라우징 제어기로 손 제스쳐를 활용하였다. 실시간으로 동작되고 16 개의 제스쳐 명령어를 인식하며, 모든 명령어는 세가지 상태 즉, 시작, 동작 그리고 끝의 상태로서 구성된다. 다른 명령어들 사이의 구분은 시작 및 동작상태에 기초한다. Su와 Furata 등은 *VPA*(virtual panel architecture)를 고안하였다. *VPA*는 모든 물리적 판넬 및 컴퓨터에 기초한 판넬의 요소들을 결합하여 사용자들이 가상 버튼을 누르거나, 가상 슬라이더를 움직이고, 가상의 스크린을 포인팅하는 기능을 갖고 있으며, 서버는 어떤 제스쳐를 나타내기 위해서 6 개 까지의 시간적 및 공간적 정보를 사용하였다. Wang과 Cannon은 손 제스쳐에 의해 로보트를 훈련하고 명령을 내릴 수 있는 *Virtual End-Effector*라는 포인팅 시스템을 개발하였다. 골격선 변환에 기초한 신경망을 사용하였으며 표면검사에 이 시스템을 이용하였다. Figueiredo 등은 *GIVEN*(Gesture-based Interaction in Virtual Environments) 으로 명명된 시스템을 개발하였으며, 사용자가 가상의 물체를 정확히 쥐기위한 접속기술을 제공하였다. 또한 이들은 Dataglove에 촉각 센서를 활용함으로써 가상공간에서 좀 더 현실감있는 환경을 제공할 수 있다고 제안하였다. 또한, W. Krueger 등은 진 일보된 개념에 의한 *Response Workbench* 시스템을 개발하였다. 이 시스템은 현실의 작업대에서 가상의 물체들을 지시하고 도구들을 제어할 수 있는 가상 작업환경을 제공

하여, 동일한 프로젝트를 수행하는 사용자들 사이에 협동적인 작업을 증진시킬 수 있음을 보여 주었다.

2. 시각에 기반한(Vision-Based) 분석

시각에 기반한 손 제스쳐는 사람과 컴퓨터와의 제스쳐에 의한 접속기술로서 일반적이며, 궁극적으로 추구하는 사용자 접속기술이다. 이는 주위의 정보를 인지할 수 있는 가장 자연스런 방법이기 때문이다. 그러나, 아직은 머신비젼이 지니고 있는 기술적인 한계성 때문에 만족할 만한 수준의 기술개발이 매우 어려운 상황이다. 몇가지 다른 접근 방법이 지금까지 실험되어 왔으며, 가장 일반적인 방법은 한개 또는 두개의 비디오 카메라를 사용하여 약간은 가상적인 환경하에서 사람에 대한 시각정보를 얻고 필요한 제스쳐를 추출한다. 그러나, 이 방법은 몇가지의 심각한 문제에 직면하게 된다. 즉, 복합적인 환경에서 움직이는 손의 분리, 환경과 관련하여 손 위치의 추적, 그리고 손 모양의 인식등이 이에 해당된다. 이러한 부담을 줄이기 위해 몇 몇 시스템들은 능동적 혹은 수동적 의미에서의 표식을 사용하거나, 표식이 있는 장갑을 사용하기도 한다. 또 다른 경우에는 제한적인 환경을 설정하는 시스템 즉, 일정한 배경, 매우 제한적인 제스쳐 어휘, 아주 단순한 정적 모양에서의 분석을 행하는 경우가 많다. 따라서, 시각에 기반한 시스템들을 분류하면 3 차원 손 모양에 기초한 분석, 표식 또는 표식을 부착한 장갑에 기초한 분석, 영상의 특성에 기초한 분석 등이 있으며 표 1에서 각 시스템들의 특성들을 보여 주고있다.

〈표 1〉 시각에 기반한 제스쳐 인식 시스템

개 발 자	응용 분야	특 징	카메라	속 도
S. Ahmad	Hand tracking	Finger detection model	Mono	10–30 fps
S. Ahmad	Posture recognition	Gaussian NN	Mono	n.av.
R. Cipolla	Gesture recognition	Marked fingertips	Mono	25 fps
K. Cho	Posture recognition	Local shape property learning	Mono	10 fps
T. Darrel	Gesture recognition	Set-of-views model	Mono	n.av.
J. Davis	Hand Tracking	Fingertip detection model	Mono	4 fps
J. Davis	Gesture recognition	Marked fingertips	Mono	n.av.
A. Downton	Limb tracking	3D cylinderical limb model	Mono	n.av.
M. Etoh	Hand tracking	3D cylindrical hand model	Stereo	n.av.
M. Fukumoto	Pointing	Finger detection and virtual projection origion	Stereo	real-time
W. Freeman	Gesture recognition	Orientation histograms	Mono	n.av.
C. Kervrann	Hand tracking	Stochastic deformable model	Mono	n.av.
R. Kjeldsen	Gesture recognition	ALVINN	Mono	n.av.
M. Krueger	Object manipulation	Silhouette	Mono	30 fps
J. Kuch	Gesture recognition	3D NURBS hand model	Mono	3–30 s/frame
J. Lee	Gesture recognition	3D hand skeleton model	Mono	40–80 min/frame
C. Maggioni	Gesture recognition	Marked glove	Mono	25 fps
J. Rehg	Gesture recognition	3D cylindrical hand model	Stereo	10 fps
J. Schlenzig	Gesture recognition	Zernike moments	Mono	1/2 fps
J. Segen	Gesture recognition	Shilutte edges	n.av.	real-time
T. Starner	Gesture recognition	HMM image geometry parameters	Mono	5 fps
A. Torgie	Pointing	Marked glove	Stereo	30 fps

(1) 3 차원 모델에 기초한 분석

손 제스쳐 인식에 사용되어 왔던 하나의 방법은 손의 3 차원 모델을 구축하여 활용하는 것이다. 모델은 카메라에 의해 얻어진 손의 이미지로서 손 바닥의 방향 및 관절의 연결각도 등에 일치하는 파라메터들을 구하여, 이들을 제스쳐 분류에 사용하였다. Downton과 Drouet는 사람의 사지(limb)를 추적할 수 있는 시스템을 개발하였다. 영상은 단순한 배경에서 하나의 카메라를 통하여 얻어지고, 모델과 이미지와의 정합은 원통 모델의 원근 투영법 및 영상으로 부터 구해진 가장자리에 비교된다. 이 시스템의 출력은 수화를 위해 사용될 수 있는 관절에서의 동적인 연결각도이다. 일본 ATR연구소의 Etoh 등은 일반화된 원통형 모델을 사용하였다. 그들은 스테레오 카메라를 사용하여 물체를 계층적이며 일반화된 원통의 집합으로 분할하는 알고리즘을 개발하였다. 윤곽선과 축적-공간 상의 확장 테크닉을 활용하여 국부적 윤곽선 상의 점들이 구해진다. 추출된 점들은 원통형과 연관된 축 및 윤곽선들을 구하는 데 사용되어진다. 이 시스템은 사람 손을 모델화하는 데 활용하였으나, 특별한 제스쳐 분석은 행하지 않았다. Regh와 Kanade는 DigitEyes로 불리는 제스쳐 인식 시스템을 개발하였다. 이 시스템은 3 차원 원통형 동적모델의 사람 손을 사용하며, 27 개의 자유도를 가지고 설계되어 있다. 손 가락 끝과 관절부가 모델정합의 특성으로 취해지고 제한적인 배경에서 에지에 기반한 분석방법으로 구해진다. 특성 추적 및 모델-변수의 평가를 위해 Gauss-Newton의 특성오차 최소화 방법을 사용하였다. Lee와 Kunii 등은 27 개의 자유도를 갖으며 3 차원적인 손의 골격선에 기초한 손 제스쳐 분석시스템을 개발하였다[7]. 이 들은 모델의 탐색공간을 줄이기 위해 5 개의 주요 제한조건들을 통합하였으며, 모델 정합을 단순화하기 위해 표식이 있는 장갑을 사용하기도 하였다. 이 모델은 반복적인 형태로 16 개의 미국수화언어(ASL : American sign language) 심볼을 사용하였으며, 계산 시간이 비교적 많이 소요된다. Kuch는 300 개 점의 NURBS(non-uniform rational B-spline)에 기초한 모델을 만들었다. 이 모델은 26 개의 자유도를 갖으며 이는 실제의 손 운동에서 6 가지의 제한 조건을 수용하는 수준이다. 이 시스템은 특정한 사람에게 맞추어 하나의 카메라에서 입력되는 꽤 긴 연속 이미지로 부터 복잡한 손 제스쳐를 추적하는 데 사용되며 가상총(virtual gun), 수화언어 추적 또는 손 제스쳐 합성에 활용될 수 있다.

(2) 표식 또는 표식있는 장갑 사용에 의한 분석

기하학적인 모델로서의 사람 손은 비교적 복잡한 형태로서 카메라로 부터 입력되는 손의 형태를 찾기란 그리 쉬운 일이 아니다. 이러한 점을 극복하기 위해서 표식을 사용하며, 표식은 일반적으로 손가락 끝에 붙인다. 표식은 영상 히스토그램 분석을 통해 찾기 쉬운 칼라 정보를 갖고 있어, 표식이 찾아지고 추적되면 몇가지의 분류 방법을 사용하여 제스쳐가 인식될 수 있다. Torige와 Kono는 손 움직임의 방향을 인지할 수 있는 제스쳐 인식시스템을 고안하였다[8]. 스트레오스코프 카메라 및 손끝, 손목, 팔꿈치, 어깨에 칼라 표식기가 부착된 검은 장갑을 사용하였으며; 3 차원 손가락 위치 및 모션의 매개변수를 계산함으로써 로보트 조종기 제어에 활용하였다. Davis와 Shah가 개발한 시스템은 일정한 배경을 갖는 여러개의 프래임에서 손끝을 추적함으로써 모션의 괘적을 알아낸다. 각 제스쳐는 start-end 벡터의 집합으로 모델링되며, 4 fps(frames per second)에서 이 시스템은 7 개의 정의된 손 제스쳐의 성공적인 시간 분할을 수행한다. 또한 이들은 원통형 손가락 모델에 기초한 손가락 끝 인식을 위한 기술을 개발하여 손 제스쳐를 3 차원 적으로 해석하였다. Maggioni는 특수한 표식을 갖는 장갑, 즉 2 개의 중심점이 다른 칼라 원으로 표시된 장갑에 의해 실험을 하였으며, 이 시스템을 스웨덴 컴퓨터 과학 연구소의 DIVE(distributed virtual environment)의 한 모듈로 활용하고 있다. Cipolla등은 손동작 및 위치를 결정하기 위해서 손가락 끝 부분에 표식이 있는 장갑을 사용하였다. 모션 시차(parallax)에서 이들은 손의 변환과 회전을 계산하였으며 시점이나 가상물체의 변화를 생성시키는 피드백 시스템에 이를 적용하였다.

(3) 영상의 특성에 의한 분석

개발된 손 제스쳐 인식의 몇 가지는 손 모양 영상과 관련된 특성추출에 기초한다. 분석되는 특성은 기본적인 기하학적 특성으로 부터 복합적인 특성 분석에 이르기 까지 넓은 범위가 있다. 접근 방법들의 가장 일반적인 특성은 그들이 실제 손모양에서의 연결각도와 같은 매개변수 평가에 귀착되지는 않는다. 이러한 분석을 사용하는 시스템들은 단순한 손의 추적이나 복합적인 제스쳐 분류의 모두에 사용될 수 있다. Darrel과 Pentland는 손 제스쳐를 모델화하기 위해 각 시점(view)에서 손 모양들의 집합을 사용하는 시스템을 개발하였다[9]. 각 손 제스쳐는 다른 시점에서 자신의 영상 집합으로 나타내 지며 이들은 후에 시간보정 및 DTP (dynamic time warpping) 방법을 사용하여 영상 시퀀스에 정합시킨다. 특수한 목적의 정합 하드웨어를 사용함으로써 10 fps의 속도가 가능하다. Segen은 간단한 그림자로 부터 영상 매개변수를 추출하는 경계선에 기반한 인식기술을 개발하였으며, 이 시스템은 실시간으로 10 개의 구별되는 손 모양을 식별할 수 있다. Adam과 Tresp는 영상의 일부가 없거나, 불분명한 입력의 경우도 구분할 수 있는 연구를 하였다. 그들은 이 문제를 해결하기 위해 Bayesian 방법에 폐쇄된 형태의 Gaussian신경망의 근사방법을 제안하였다. 이 신경망은 손끝 좌표와 손의 무게 중심에 의해 기술되는 손 모양을 구분한다. Starner와 Pentland는 ASL(American Sign Lengnage)인식을 위해 입력영상의 기하학적 매개변수를 사용했다. 기하학적 매개변수는 일정한 색깔의 손 이미지로 부터 추출되고, 은닉 마코프 모델의 5 가지 상태 위상이 제스쳐 구분을 위해 선택되며 약 85 %의 인식 결과가 보고되었다. M. Krueger의 *VIDEOPL-ACE, VIDEODESK* 및 *VIDEOTOUCH*는 사용자 손의 그림자를 기하학적으로 분석하는 데에 기초한 특수한 가상 세계 시스템이다. 이 시스템은 사용자의 이미지를 분석하고, 사용자의 몸의 부위를 확인하고, 사용자가 손을 이용하여 물체를 가리키거나 수정할 수 있다. Schlenzig등은 영상 특성으로 Zernike 모멘트를 사용하여 비교적 제한된 환경에서 하나의 카메라에 의해 입력된 영상에서 특성을 추출하였다. 특성 추출후에 은닉 마코프 모델에 의해 손 제스쳐 영상을 인식하여 원격지 로보트의 제어에 활용하였다. Kjeldsen은 컴퓨터 윈도를 제어하기 위한 손 제스쳐 시스템을 고안하였다. 단일 카메라에서 영상 히스토그램에 의한 분리에 기초하였으며, ALVINN 신경망 모델을 사용하여 인식하였다. 손에 의한 포인팅 및 몇 가지의 제어용 제스쳐가 표준 마우스 입력을 대체하는 실험을 하였다. 최근 Freeman과 Roth는 국부적으로 각 방향의 히스토그램을 사용하여 단순하고 빠르며 조명에 덜 민감한 시스템을 개발하였고, 컴퓨터 그래픽에서의 제어 모듈로서 접목을 시도하였다.

(4) 기타의 시각기반에 의한 분석

지금까지 소개한 방법들 이외에도 여러 참신한 기법들이 활용되고 있다. 이러한 기법들은 제스쳐 인식을 위해 특별히 적용될 수 있음은 물론 일반적인 인식에도 성공적으로 활용될 수 있다. NTT 인간 접속 연구실의 Fukumoto등은 3 차원 포인팅 및 음성 명령어를 도와줄 수 있는 시스템을 고안하였다[10]. 두개의 카메라를 사용하며 사용자의 어깨 근처에 위치한다고 가정하는 가상투영원점(VPO : virtual projection origin)을 사용한다. VPO는 손가락의 위치 정보와 함께 포인팅 방향을 결정하는 데 활용된다. 개발된 기술은 음성 명령어의 동기화 기술과 함께 프레젠테이션 시스템, 비디오 브라우저, *Space Writer*에 활용하였다. 국부적 형태 특성과 결합된 손 모양 모델링 방법이 Cho와 Dunn에 의해 제안되었다. 이들은 국부적인 정보와 대량의 클래스로 여러가지 경우의 수를 학습할 수 있는 특성기반 학습 알고리즘을 사용하였다. 직선성분이 국부적 형태의 선택된 에지분석으로 부터 추출되며, 서로 다른 5 가지의 손 모양의 분류를 수행할 수 있다. 또한, Kervrann과 Heitz는 비정형 동적 모델의 변형을 모델링하고 훈련시킬 수 있는 일반적인 구조를 발표하였으며, 프로토타입 형태의 통계적으로 기술된 변형을 사용하여 최적화에 기초를 두고 각 모형을 평가할 수 있다. 이 시스템은 복합적인 배경에서 손의 추적을 위해 활용된다.

3. 필기 제스쳐(drawing gesture) 분석

필기 제스쳐는 컴퓨터에 명령을 주기 위해 연속적인 스토로크로서 구성되는 제스쳐를 의미한다. 이것은 입력 기구로서 스타일러스나 마우스를 포함하며 클릭이나 드래그 같은 행위를 뛰어넘어 사용범위를 확장할 수 있다. 그리고, 필기 제스쳐 인식은 온 라인 필기체 문자인식 범위도 포함할 수 있으나, 이 분야의 연구는 문자인식 분야에서 그동안 여러 경로를 통해 많은 연구결과가 발표되었기 때문에 여기서는 포함하지 않는다. Rubin은 *GRANDMA*(gesture recognizers automated in a novel direct manipulation architecture)라는 필기 제스쳐 응용 툴을 고안하였다[11]. 이 시스템은 입력기구로서 복수의 손가락을 사용할 수 있는 Sensor Frame으로 스크린 상의 물체를 두 개의 손가락 제스쳐로 회전, 변환, 확대 및 축소할 수 있다. 또한, 그는 *GSCORE*라 부르는 악보편집 시스템을 고안하였다. Yang, Xu 등은 필기 제스쳐 인식을 위한 은닉 마코프 모델을 개발하여, 이를 9개의 필기 숫자를 인식하는 데 활용하였다.

4. 기타의 제스쳐 인식 기술

지금까지 설명한 방법들 이외에도 사람과 컴퓨터 사이의 접속을 위한 여러가지 제스쳐 인식 기술들이 개발되어 졌다. 이중에서 장차 매우 유망해 보이는 연구의 하나가 근전도(EMG : electromyograms)에 의한 인식 기술이다. EMG의 장점은 분해능이 다른 방법에 비해 매우 높다는 점이다. Putman과 Knapp은 EMG 신호에 신경망에 기초한 제스쳐 인식기술을 접목하여, 실시간으로 그래픽 유저 인터페이스를 제어하는 데 활용하였다.

IV. 제스쳐 합성

제스쳐 합성은 지금까지 컴퓨터 그래픽이나 애니메이션 기술에 의존하여 매우 제한된 연구 개발이 수행되어 왔다. 그러나, 화상전화 시스템과 같은 새로운 영역의 출현은 손 제스쳐의 분석은 물론 합성기술도 요구하게 되었다. 가상환경에서의 제스쳐 분석 시스템들의 대부분은 매우 단순한 3차원 손 모델을 활용하며, 일반적으로 원통형 근사(cylindrical approximation)를 한다. 이러한 모델들의 운동역학은 대부분 무시되고, 가상공간 상에서 전체 모델의 움직임은 제스쳐 분석 단계에서 도출되는 몇 가지 추정 매개변수에 의존한다. 이러한 모델들은 실험실 환경에서는 활용이 가능하나, 다음 세대의 가상 환경에 활용되기 위해서는 성능 향상이 필요하다. 컴퓨터 애니메이션은 매우 높은 현실감을 요구한다. Magnenat-Thalmann 등이 개발한 *HUMAN FACTORY* 프로젝트에서 가상 배우의 움직임을 애니메이션 하는 데 세세한 기술 개발이 이루어 졌다[12]. 즉, 연결부를 둥글게 하고 근육을 현실감이 있을 정도로 부풀게 하며, 피부 변형의 표현을 매우 자연스럽게 처리함으로써 쥐는 동작을 현실감 있게 보여 주었다. 또한, 이들은 *HUMANOID* 프로젝트에서 쥐는 동작의 분석과 모델링을 수행중에 있다. Kuch는 3 차원 NURBS 모델에 기초하여 매우 자연 스럽게 손의 움직임을 합성할 수 있는 시스템을 개발하였다. 이들이 개발한 손 모양의 모델은 6 개의 운동역학적인 제한 요소를 포함하며, 각 사용자의 손 모양을 나타내기 위해 보정도 가능하다.

V. 결 언

가상의 세계가 우리 앞에 현실로서 다가오고 있다. 지금까지의 사용자 접속기구로 부터 탈피하고 새로운 가상환경을 다루고 탐색하기 위한 기술 수요는 사람과 컴퓨터와의 접속기술에서 새로운 방법들을 요구하게 되었다. 이러한 관점에서 제스쳐의 사용은 사람이 자연 환경에서 대화하고 접속하는 일반적인 방법에 영향을 받는다. 이 논문에서 주로 제스쳐에 기반한 모델링, 분석, 합성에 사용되는 여러 방법들에 대해 살펴보았다. 최근의 사용자 접속기술에서 사용되는 손 제스쳐 모델은 손 모양의 평가 뿐만 아니라 손 제스쳐가 고유하게 가지

고 있는 운동역학적인 지식도 포함한다. 또한, 제스쳐의 합성은 인식용 장갑이나 스타일러스등과 같은 기구를 사용하는 것 대신에 점증적으로 사용자에게 아무런 부담이 없는 컴퓨터 비젼에 의존하려는 경향이다. 그러나, 아직은 지금까지 설명된 시스템들이 대부분 실험실 환경에서 동작되고 있는 실정으로, 앞으로 *DataGlove*나 표식의 부착에 의존하는 시스템들의 실용화가 먼저 이루어 질 것으로 판단된다. 국내에서도 '90 년대 초반부터 가상현실에 대한 관심이 높아 지고 *DataGlove*의 사용이 빈번해 지면서 손 제스쳐의 신호를 가상 공간에 있는 물체에 행위를 가하는 명령어로 변환하는 응용 기술들이 실용화 되어왔다. 최근에는 감성공학 등에 관련된 프로젝트에서 이러한 유형의 연구가 활발히 진행되고 있다. 시각정보에 의한 제스쳐 인식 연구는 최근 2~3 년 전부터 시작되어 필자가 근무하고 있는 시스템 공학연구소를 비롯하여, KAIST, 숭실대, 고려대 등에서 연구를 진행하고 있으나, 아직은 기초 기술 확보를 위한 초보적인 단계에 있다. 그러나, 선진외국의 예에서 알 수 있듯이 제스쳐 인식의 연구는 미래의 기술을 이끌어 나갈 중요한 기술로 평가되어 일본에서 수행중인 RWC(real world computing) 프로젝트 등에서 매우 심도 있게 추진 중에 있으며, 금년 10 월에는 MIT대학의 multimedia lab.이 주관이 되어 제 2 회 얼굴 및 제스쳐 인식에 관한 국제학술 회의가 개최될 예정으로 국내에서도 제스쳐 인식기술 개발을 위한 체계적인 연구가 필요한 실정이다. 한편, 제스쳐 인식 기술은 기대에 비해 기술의 발전 속도는 그리 빠르지 못할 것으로 판단되며, 향후 컴퓨터 비젼 기술의 비약적인 발전 및 여러 인공지능적인 기술들의 접목이 잘 이루어 져야 사람과 컴퓨터의 접속기술에 새로운 장을 열 수 있으루 것으로 기대된다.

참 고 문 헌

〔1〕 민병우, "시각인식에 기반한 사용자 접속기술", *'95 정보처리학회 춘계학술발표 논문집*, 제2권 1호, pp. 48-51, 1995년 5월

〔2〕 T.S. Huang and V.I. Pavloic, "Hand Gesture Modeling, Analysis, and Synthesis," *Proc. of International Workshop on Automatic Face-and Gesture-Recognition*, pp. 73-79, Zurich, June 1995.

〔3〕 G. Johansson, "Visual perception motion and a model for its analysis," *Perception and Physics 14*, pp. 201-211, 1973.

〔4〕 W. long and Y.H. Yang, "Log-Tracker : an attribute-based approach to tracking human body motion," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 5, no. 3, pp. 439-458, 1991.

〔5〕 F. Quek, "Toward a vision-based hand gesture interface," *Proc. of Virtual Reality Software and Technology Conference*, Aug. 1994.

〔6〕 T. Baudel and M. Beaudouin-Lafon, "Charade : Remote control of objects using free-hand gestures," *Communication of ACM*, vol. 36, no.7, pp. 387-393, IEEE, 1993.

〔7〕 J. Lee and T.L. Kunii, "Constraint-based hand animation," *Models and techniques in computer animation*, pp. 110-127, Tokyo : Springer-Verlag, 1993.

〔8〕 A. Torgie and T. Kono, "Human-interface by recognition of human gestures with image processing recognition of gesture to specify moving direction," *Proc. of IEEE international workshop on robot and human communication*, pp. 105-110, 1993.

〔9〕 T. Darrel and A. Pentland, "Space-time gesture," *Proc. of Computer Vision and Pattern Recognition Conference*, 1993.

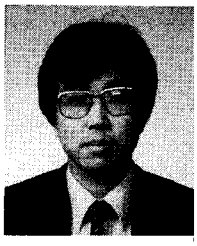
〔10〕 M. Fukumoto, Y. Suenga, and K. Mase, "Finger-pointer : Pointing interface by

image processing," *Computers and Graphics*, vol. 18, no. 5, pp. 633-642, 1994.

[11] D. Rubin, "Integrating gesture recognition and direct manipulation," *Proc. of the Summer 1991 USENIX Technical Conference*, pp. 281-298, June 1991.

[12] N. Magnenat-Thalmann, R. Laperriere, and D. Thalmann, "Joint-dependent local deformation and object grasping," *Proc. of Graphics Interface*, vol. 23, pp. 21-30, July 1989.

저 자 소 개



閔 丙 雨

1955年 3月 18日生

1979年 서울대학교 공과대학 졸업(공학사)

1993年 충북대학교 전자계산학과 졸업(이학석사)

1996年 큐슈공업대학 박사과정 수료

1982年 ~ 현재 시스템공학연구소 인공지능연구부 선임연구원

주관심 분야 : 영상인식 및 처리, 지식기반 시스템, 신경망 시스템