정규논문 (Regular Paper)

방송공학회논문지 제19권 제1호, 2014년 1월 (JBE Vol. 19, No. 1, January 2014)

<http://dx.doi.org/10.5909/JBE.2014.19.1.83>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

손 자세 인식을 이용한 MPEG-U 기반 향상된 사용자 상호작용 인터페이스 시스템

한 국 회^{a)}, 이 인 재^{b)}, 최 해 철^{a)*}

MPEG-U based Advanced User Interaction Interface System Using Hand Posture Recognition

Gukhee Han^{a)}, Injae Lee^{b)}, and Haechul Choi^{a)*}

요 약

최근 손과 손가락을 인식하는 기술은 HCI(human computer interaction)에서 자연스럽게 친숙한 환경을 제공하기 위한 기술로 주목 받고 있다. 본 논문에서는 깊이 카메라를 이용하여 손과 손가락의 모양을 검출 및 인식하는 방법을 제안하고, 그 인식 결과를 활용하여 다양한 기기와 상호연동 할 수 있는 MPEG-U 기반 향상된 사용자 상호작용 인터페이스 시스템을 제안한다. 제안하는 시스템은 깊이 카메라를 이용하여 손을 검출한 후, 손목의 위치를 찾아 최소 손 영역을 검출한다. 이어서 검출된 최소 손 영역으로부터 손가락 끝점을 검출 한 후, 최소 손 영역의 중심점과 손가락 끝점간의 뼈대를 만든다. 이렇게 만든 뼈대의 길이와 인접 뼈대간의 각도차를 분석하여 손가락을 판별한다. 또한, 제안하는 시스템은 사용자가 MPEG-U에서 정의하는 다양한 심벌들을 손 자세로 취하였을 때 제안 방법을 이용하여 손 자세를 인식하고, 인식 결과를 상호연동 가능한 MPEG-U 스키마 구조로 표현한다. 실험에서는 다양한 환경에서 제안하는 손 자세 인식 방법의 성능을 보인다. 또한, 제안 시스템의 상호연동성을 보이기 위해 인식 결과를 MPEG-U part2 표준에 맞는 XML 문서로 표현하고, MPEG-U 참조 소프트웨어를 이용하여 그 표현 결과에 대한 표준 부합성을 검증한다.

Abstract

Hand posture recognition is an important technique to enable a natural and familiar interface in HCI(human computer interaction) field. In this paper, we introduce a hand posture recognition method by using a depth camera. Moreover, the hand posture recognition method is incorporated with MPEG-U based advanced user interaction (AUI) interface system, which can provide a natural interface with a variety of devices. The proposed method initially detects positions and lengths of all fingers opened and then it recognizes hand posture from pose of one or two hands and the number of fingers folded when user takes a gesture representing a pattern of AUI data format specified in the MPEG-U part 2. The AUI interface system represents user's hand posture as compliant MPEG-U schema structure. Experimental results show performance of the hand posture recognition and it is verified that the AUI interface system is compatible with the MPEG-U standard.

Keyword : MPEG-U, Depth camera, Hand and finger recognition, AUI(Advanced User Interaction), User interface

1. 서론

최근 HCI(human computer interaction)분야에서는 인간과 기계 사이의 인터페이스를 위하여 의사소통을 주고받는 상호작용에 대해 많은 연구가 진행되고 있다. HCI를 위한 기존 인터페이스 방법은 키보드나 마우스와 같은 스위치 기반 혹은 포인팅 디바이스들을 이용한 방법이다. 하지만 이러한 방법은 인간의 의사소통과 같은 자연스러운 방법이 아니다. 자연스러운 상호작용을 이루기 위해서는 인간의 의사소통 체계와 같은 특징을 지녀야 한다. 인간은 의사소통 수단으로 음성을 많이 이용하지만, 음성과 함께 손의 자세나 동작을 많이 이용한다. 기존의 연구에 따르면 사람은 80% 이상의 정보를 시각을 통해서 획득하므로 시각적 인터페이스는 다양한 인터페이스 방법 중에서 사람들에게 가장 익숙하고 편리한 방법 중 하나이다. 그림 1은 시각적 인터페이스를 이용한 응용기기의 한 예를 보여주고 있다. 즉,

그림 1. 시각적 인터페이스를 이용한 응용기기의 예
Fig. 1. An example of the application utilizing a visual interface

사용자가 동작 인식 카메라를 통해 손으로 원을 그렸을 때, 입력 값이 TV에서 정의된 특정 이벤트를 호출하는 모습이다. 본 연구에서는 자연스러운 상호작용을 위해 손 자세를 인식하여 다양한 기기와 상호연동 할 수 있는 인터페이스 시스템을 제안한다.

손을 검출하거나 손 자세를 인식하는 기존의 많은 연구들은 2차원 영상 데이터를 사용한다. 카메라의 폭넓은 보급으로 인해 하나의 카메라를 사용하는 2차원 영상 데이터는 경제적이란 장점은 있으나, 현실적으로 다양한 환경의 변화에 민감할 수 있다. 이러한 한계를 극복하기 위해 최근에는 3차원 영상 데이터를 사용하여 손과 손가락을 검출하여 인식하는 연구가 활발히 진행되고 있다. 3차원 영상 데이터는 2차원 영상 데이터에 깊이 정보가 추가된 영상 데이터이다. 깊이 정보를 획득하기 위해서는 두 대 이상의 일반 카메라를 결합하거나, 깊이 센서를 활용하는 방법들이 있다. 최근에는 깊이 카메라가 범용화 되고 있고, 스마트 기기에 깊이 카메라를 탑재해 제품을 출시할 양상이 보이고 있다. 따라서 본 연구에서는 환경변화에 강인한 인식을 위하여 깊이 정보를 추가적으로 이용함으로써 정확한 인식이 가능한 3차원 영상 데이터를 활용하는 연구를 제안한다. 3차원 영상 데이터는 깊이 센서 기반인 Microsoft사의 KINECT XBOX 360을 사용한다.

사용자 인터페이스 시스템에서 고려해야 할 중요한 점 중 하나는 기기들 간의 상호 연동성이다. 시선 추적, 동작 인식, 손 자세 인식 등과 같은 자연스러운 상호작용을 위한 방법은 편리한 기능을 제공할 수 있지만, 제품을 출시하는 회사마다 서로 다른 AUI(advanced user interaction) 인터페이스 데이터 포맷을 정의하고 있기 때문에 각각의 제품들은 서로 다른 회사의 장치들과 연동되기 어렵다. 이에 따라 ISO/IEC JTC1/SC29/WG11 MPEG(Moving Picture Experts Group)에서는 AUI 인터페이스 데이터 포맷을 표준화하고 있으며, 그 표준화 프로젝트를 ‘MPEG-U part2: AUI Interface’로 명명하였다^[1]. MPEG-U part2의 목적은 AUI 인터페이스에서 표준으로 정의된 일관성을 갖는 데이터 포맷을 제공함으로써 장치들 간에 효과적인 상호연동 및 인터페이스 방법을 지원하는 것이다^[2]. MPEG-U part 2에서는 앞서 기술한 AUI 인터페이스 데이터 포맷들을

a) 한밭대학교 정보통신전문대학원 멀티미디어공학과 (Multimedia Engineering, Hanbat National University)

b) 한국전자통신연구원 방송통신융합연구부실감방송미디어연구부 (Broadcasting & Telecommunications Media Research Department, ETRI)

✉ Corresponding Author : 최해철(Haechul Choi)
E-mail: choihc@hanbat.ac.kr
Tel: +82-42-821-1149

※ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2013R1A1A1010344) 및 미래창조과학부가 지원한 2013년 정보통신·방송(ICT) 연구개발사업의 연구결과로 수행되었음. (KCA-2013-(13-912-03-001))

Manuscript received October 28, 2013 Revised December 30, 2013
Accepted December 30, 2013

XML 기반의 스키마(schema) 구조에 따라 정의하고 있다. XML은 이미 충분한 인프라를 갖추고 있으며 인터넷 환경에서 XML 특유의 유연성을 바탕으로 사물에 대한 상황정보를 효과적으로 표현할 수 있다. 또한 정형화된 형식에 맞추어 정보를 교환함으로써 객체간의 정보 전달을 위한 매체로 이용할 수 있다. 따라서 본 연구에서는 손 자세 인식을 활용하여 사용자와 기계와의 상호작용을 위한 인터페이스뿐만 아니라 이중 기계간 상호연동 가능한 AUI 인터페이스 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 시스템과 관련된 연구에 대해 소개하고, 3장에서는 손 자세 인식을 이용한 MPEG-U 기반 향상된 사용자 인터페이스 시스템 구조에 대해서 설명한다. 4장에서는 실험을 통하여 손과 손가락 인식의 강인성을 보여주고, MPEG-U 기반 향상된 사용자 상호작용 인터페이스 시스템의 타당성을 검증한다. 마지막으로 5장에서 결론을 내린다.

II. 관련 연구

본 장에서는 제안하는 시스템과 관련된 손과 손가락을 검출하는 방법을 소개하고, MPEG-U에 대한 내용을 소개한다.

1. 손 및 손가락 검출 방법

손과 손가락을 검출하는 기술은 HCI에서 자연스럽고 친숙한 환경을 제공하기 위한 중요한 기술이다. 손과 손가락을 검출하는 방법은 그 동안 다양한 연구가 있었다. 손을 검출하는 방법에는 사용하는 방법에 따라서 네 종류로 분류될 수 있다. 첫 번째로는 특정한 보조 도구를 착용하는 방법^[3,4], 두 번째는 색상 정보를 이용하는 방법^[5,6,7], 세 번째는 3차원 손 모델을 이용하는 방법^[8], 마지막으로 3차원 영상 데이터베이스를 이용한 방법^[9]이 있다. Robert Y. Wang^[3]과 R. Lockton^[4]는 그림 2와 같이 특정한 장갑을 끼거나 손목에 컬러 밴드를 착용하여 손을 검출한다. 이처럼 특정한 보조 도구를 착용하여 손을 검출하는 방법은 정확한 손 영

그림 2. 특정한 특징을 지닌 장갑을 끼거나 손목에 컬러밴드를 착용하여 손을 검출하는 방법^[3,4]

Fig. 2. Hand detecting methods using a specific glove or a wrist cover band^[3,4]

역을 검출할 수 있는 장점이 있지만, 사용자가 항상 보조 도구를 착용해야하는 불편한 단점이 있다. Anastasios Roussos^[5], Yuh-Rau Wang^[6], Xintao Li^[7]등은 피부색과 같은 색상 정보를 추출하여 손을 검출한다. 이러한 방법들은 하나의 카메라로 검출할 수 있어 경제적이라는 장점이 있지만, 복잡한 배경 및 조명 변화에 취약하다는 단점이 있다. V. Athitsos^[8]는 그림 3과 같이 3차원 손 모델을 이용하여 손을 검출한다. 이 방법은 입력된 손과 준비된 3차원 손 모델을 비교하여 손을 검출한다. 이처럼 3차원 손 모델을 이용하여 손을 검출하는 방법은 색상 정보만을 이용하여 손을 검출하는 방법보다는 정확한 손 검출 및 손의 자세를 인식할 수 있는 장점이 있으나, 3차원 모델링을 위해 많은

그림 3. 3차원 손 모델을 이용하여 손을 검출하는 방법 [8]

Fig. 3. Hand detecting method using a three-dimensional hand model [8]

그림 4. 3차원 영상 데이터를 이용하여 손을 검출하는 방법^[9]Fig. 4. Hand detecting method using three-dimensional video data^[9]

노력이 필요하다. Candescent NUI에서 제공하는 오픈소스^[9]는 그림 4와 같이 3차원 영상 데이터를 이용하여 손을 검출한다. 이 방법은 깊이 정보를 활용하기 때문에 손을 정확하고, 빠르게 검출할 수 있는 장점과 복잡한 배경 및 조명 변화에 강인하고, 계산량이 적은 장점이 있다. 하지만, 3차원 영상 데이터를 획득할 수 있는 장비가 필요하다는 단점이 있다. 본 연구에서는 선행 연구들의 장점과 단점을 기반으로 다양한 환경변화에 따른 손 검출의 강인성과 깊이 카메라의 범용화에 따라 특정 장비가 필요하지 않는 3차원 영상 데이터를 이용하여 손을 검출한다.

손 자세를 정확히 인식하기 위해서는 손가락도 구별하여 인식할 필요가 있다. 기존 손가락을 검출하는 다양한 연구에는 기울기 정보를 기반으로 손가락 사이의 굴곡 정보를 이용한 검출 방법^[10], Convex hull 기반의 convexity defects를 이용한 방법^[11], 에지 정보를 이용해 손가락 형상 모델과의 확률적 매칭을 통한 검출 방법^[12], 손바닥 포즈 정보를 기반으로 화소의 끊어진 횟수를 세는 방법^[13] 등이 있다. S. Malik^[10]는 그림 5와 같이 기울기 정보를 기반으로 손가락

그림 5. 기울기 정보를 기반으로 손가락 사이의 굴곡 정보를 이용하여 손가락을 검출하는 방법 [10]

Fig. 5. Finger detecting method using the curve information between fingers based on the gradient information [10]

사이의 굴곡 정보를 이용하여 손가락을 검출한다. 이 방법은 검출된 손의 외곽선에서 곡률의 변화가 가장 크게 일어나는 점을 손가락 끝점으로 판별한다. 기울기 차이가 뚜렷한 형상에 대해서는 좋은 결과를 보여주지만, 크기 변화에 민감하며, 손 이외의 손목, 팔꿈치 등 부가적인 영역이 포함되면 다수의 오 검출 결과를 보인다. C. Davatzikos^[11]는 그림 6과 같이 Convex hull 기반의 convexity defects를 이용하여 손가락을 검출한다. 이 방법은 인식된 손의 경계 정보를 이용하여 Convex Defects 계산을 통해 손가락 끝 후보 지점을 추출한다. 손가락 끝 후보 지점을 손가락 첫 마디가 가지는 타원형의 형태학적 특성을 토대로 개선된 타원 근사화 모델과의 적합도를 평가하는 추가적인 후처리 과정을 거쳐 최종적인 손가락 끝점을 검출한다. 이 방법은 기울기 정보를 기반으로 손가락 사이의 굴곡 정보를 이용한

그림 6. Convex Hull 기반의 Convexity Defects를 이용하여 손가락을 검출하는 방법^[11]Fig. 6. Finger detecting method using the Convex Hull based Convexity Defects^[11]

검출 방법보다는 계산량이 적고, 비교적 잡음에 강인하지만, 정확한 손가락의 추정 및 해당 손가락의 방향 정보를 획득하기 위한 추가적인 후처리 과정이 필요하다. I. Oikonomidis^[12]는 그림 7과 같이 에지 정보를 이용해 손가락 형상 모델과의 확률적 매칭을 통하여 손가락을 검출한다. 이 방법은 검출된 손과 준비된 손가락 형상 모델과의 확률적 매칭을 통해 손가락을 검출하는 방법이다. 정확한 손가락 검출이 가능하지만, 계산량이 많고, 잡음에 약한 단점이 있다. J. Choi^[13]는 그림 8과 같이 손바닥 포즈 정보를 기반으로 화소의 끊어진 횟수를 세어 손가락을 검출한다. 이 방법은 검출된 손을 이진화한 다음 손목부터 손가락 끝 방향으로 진행하면서 진행방향과 수직인 방향으로 화소의 끊어진 횟수를 센다. 화소 값이 1에서 0으로 변한 다음 다시 1로 변하는 횟수를 세어 손가락 끝점을 검출한다. 계산량이 적지만 다양한 환경 변화에 취약하다는 단점이 있다. 본 연구에서는 손 자세 인식을 이용한 AUI가 실시간으로 동작이 가능하고, 다양한 환경변화에 강인한 Convex hull 기반의 convexity defects를 이용하여 손가락을 검출한다.

2. MPEG-U

MPEG 표준은 모든 오디오비주얼 데이터를 객체 기반으로 처리하고 표현할 수 있도록 하며, 자연 영상/오디오뿐만 아니라 컴퓨터 합성영상, 합성 오디오 등을 포함하는 멀티미디어 압축 및 처리, 전송을 지원하는 국제 표준이다. 대화형 멀티미디어 응용 분야, 게임, 인터넷 멀티미디어 서비스, Mobile 통신 등에 활용 가능하며, 응용에 대한 기술 개발이 본격적으로 시작되고 있다. MPEG은 객체 기반 오디오비주얼 데이터 압축 규격뿐만 아니라 객체 표현 및 조작에 관한 전반적인 표준도 다루고 있다. 그 중 MPEG-U는 리치 미디어 사용자 인터페이스를 위하여, 단말간 위젯을 표현하고 처리할 수 있는 상호 호환적인 위젯 표현 포맷과 위젯 패키징 포맷을 정의한다. MPEG-U part2에서는 UID(user interaction device)와 다양한 객체(비디오, 오디오, 2D 그래픽 객체, 애니메이션 등)로 표현되는 장면 기술(scene description) 사이에서 서로의 상호연동을 돕는다. 그림 9는 UID와 장면 기술을 이용한 MPEG-U part 2 시스템의 데이

그림 7. 에지 정보를 이용하여 손가락 형상 모델과의 확률적 매칭을 통한 손가락 검출 방법 [12]
Fig. 7. Finger detecting method using the probable matching with the edge based finger phase model^[12]

그림 8. 손바닥 포즈 정보를 기반으로 화소의 끊어진 횟수를 세어 손가락을 검출하는 방법 [13]
Fig. 8. Finger detecting method based on the connectivity between the hand segments [13]

터 흐름을 보이고 있다. 그림 9의 하단부분에 위치한 UID (모션 센서, 멀티 터치 인터페이스, 지능형 카메라, 밴드 센서)에서 사용자의 물리적 입력 정보를 감지하면, 사용자 인터페이스 입력부에서 데이터를 받아 MPEG-U에서 정의하는 하나의 데이터 타입으로 표현한다. 표 1은 MPEG-U part 2가 정의하는 데이터 타입들과 종류별로 구분한 6 가지 패턴을 보여주고 있다. 패턴은 기하학적 패턴(Geometric Pattern), 상징적 패턴(Symbolic Pattern), 터치 패턴(Touch Pattern), 손의 자세(Hand Posture Pattern), 그리고 손동작(Hand Gesture Pattern)과 혼합 패턴(Composite Pattern)으

그림 9. MPEG-U part2의 시스템 구성 [2]

Fig. 9. MPEG-U part2 system architecture [2]

표 1. MPEG-U part 2에서 정의하는 AUI 인터페이스 데이터 포맷
Table 1. AUI interface data format specified in the MPEG-U part 2

Pattern	Type
Geometric	Point, Line, Rect, Arc, Circle
Symbolic	Victory, Heart, Rock, Scissors, Paper, Okay
Hand Posture	OpenPalm, Fist, Pointing, Thumb-Up, Thumb-Down, Grab
Hand Gesture	Push, Pull, Slap left, Slap right, Slap top, Slap bottom, Circle Clockwise, Circle anti-Clockwise, Waving, Check
Touch	Tap, DoubleTap, Press, Drag, Rotate, Flick
Composite	

로 나뉜다. 현재 MPEG-U는 part3 Conformance and reference software를 준비 하고 있으며 가상세계와 실세계간의 인터페이스를 정의하는 MPEG-V와도 함께 연구 되고 있다. 본 연구에서는 다양기기와 상호연동하기 위하여 MPEG-U를 따르는 인터페이스 시스템을 제안한다.

III. 손 자세 인식 기반 MPEG-U 사용자 인터페이스 시스템

제안하는 시스템은 깊이 카메라인 KINECT를 이용하여 손 자세 인식에 의한 MPEG-U 사용자 인터페이스 시스템

그림 10. MPEG-U 기반 사용자 인터페이스 시스템 구성

Fig. 10. MPEG-U based user interface system structure

이다. 제안하는 시스템의 구성은 그림 10과 같으며, 크게 두 가지 모듈(손 자세 인식부, MPEG-U XML 생성부)로 구성된다. 손 자세 인식부는 다시 검출부와 인식부로 나뉜다. 검출부는 Candescant NUI에서 제공하는 오픈소스를 이용하여 손 영역을 검출한다. 검출된 전체 손 영역 중 최소 손 영역을 검출하기 위하여 손목 지점을 찾아 손목 지점에서 손가락 끝점까지의 최소 손 영역을 포함한 최소 사각형을 검출한다. 다음으로 손가락 검출을 위한 최소 손 영역의 중심점을 구하고, Convex Hull 기반의 Convexity Defects를 이용하여 손가락 끝점을 검출한다. 인식부는 검출된 결과를 바탕으로 각 손을 왼손과 오른손으로, 각 손가락을 엄지, 검지, 중지, 약지, 소지로 판별하기 위해 제안하는 알고리즘을 적용하여 판별하고, 손의 자세와 각 손가락의 퍼짐을 판단하여 최종적인 손 자세를 인식한다. 다음으로 MPEG-U XML 생성 부는 인식된 손 자세를 상호연동 가능한 데이터 포맷으로 표현하기 위해 MPEG-U part2 표준에 부합하는 특징을 지닌 포즈를 취하여 MPEG-U part2 표준에 맞게 XML 문서로 파싱한다.

1. 손 자세 인식부

1.1 검출부

KINECT는 적외선을 쏘아 반사되어 돌아오는 시간을 감지 센서가 측정하여 깊이 정보를 획득한다^[14]. KINECT를 사용하기 위해서는 PC와 호환을 가능하게 해주는 3D 동작 인식 미들웨어(3D motion recognized middleware)가 필요하다. KINECT를 위해서 주로 이용하는 미들웨어로는 마이크로소프트사의 KINECT SDK와 프라임센스 OpenNI 등이 있다. Candescant NUI는 이러한 미들웨어를 활용하여 손과 손가락을 인식하는 오픈소스를 제공한다. Candescant NUI는 3차원 영상 데이터를 활용하여 손을 정확하고, 빠르게 검출할 수 있는 장점과 복잡한 배경 및 조명 변화에 강인하고, 계산량이 적다는 장점이 있다. 제안하는 시스템의 손 및 손가락 인식부는 KINECT로부터 입력 받은 깊이 정보를 Candescant NUI의 오픈 소스^[9]를 이용해 그림 11의 (a)와 같이 손 영역을 검출한다. 이렇게 검출된 전체 손 영역 중 최소 손 영역을 검출하기 위하여 다음의 과정을 수행

그림 11. (a) Candescant NUI에서 제공하는 오픈소스를 이용하여 손을 검출한 결과, (b) Line fitting algorithm을 적용하여 손목 지점을 검출한 결과, (c) 최소 손 영역을 포함한 최소 사각형을 검출한 결과, (d) Distance transform algorithm을 적용하여 최소 손 영역의 중심점을 검출한 결과, (e) Convex hull 기반의 convexity defects를 이용하여 손가락을 검출한 결과
Fig. 11. (a) result of the hand detection using the Candescant NUI, (b) result of the wrist detection by applying the line fitting algorithm, (c) result of detecting the least square of hand region, (d) result of detecting the center of mass of the least square applying the distance transform algorithm, (e) result of finger detecting using the Convex Hull based Convexity Defects

한다. 먼저, 최소 자승법을 이용한 라인 피팅^[15]을 적용하여 전체 손 영역을 대표하는 하나의 직선을 얻는다. 이 직선의 위치 및 기울기를 이용하여 손 영역의 외곽선과 만나는 두 점들의 거리를 구한 후, 이 거리를 비교하여 일정해지기 시작하는 부분을 손목 지점이라 결정짓는다. 그림 11의 (b)는 전체 손 영역을 대표하는 하나의 직선을 나타내어 손목 지점을 나타낸 그림이다. 손목 지점을 찾은 다음으로 손목 지점부터 손가락 끝 지점까지의 최소 손 영역을 포함한 최소

의 사각형을 구한다. 그림 11의 (c)는 손목 지점부터 손가락 끝점까지의 최소 손 영역을 포함한 최소 사각형을 나타낸 그림이다. 다음으로 최소 손 영역의 중심점을 거리 변환 알고리즘^[16]을 적용하여 찾는다. 중심점은 추후 손가락 검출 및 인식을 위해서 필요하며, MPEG-U AUI 인터페이스의 데이터 포맷에 손의 position 값으로 자주 활용 된다. 그림 11의 (d)는 거리 변환 알고리즘을 적용하여 최소 손 영역의 중심점을 나타낸 그림이다. 다음으로 Convex hull 기반의 convexity defects^[11]를 이용하여 손가락 끝점을 검출한다. 이 방법은 정확한 손가락의 추정 및 해당 손가락의 방향 정보를 얻을 수 있지만, 추가적인 후처리 과정이 반드시 필요하다. 추가적인 후처리 과정은 다음과 같다. 손가락 첫 마디가 가지는 타원형의 형태학적 특성을 토대로 개선된 타원 근사 화 모델과의 적합도를 평가한다. 이 결과를 바탕으로 각 손가락 끝점과 중심점과의 벡터 각도 차 정보를 이용하여 최종적인 손가락 끝점을 검출한다. 그림 11의 (e)에서 붉은 점이 최종적으로 검출된 손가락 끝점이다.

1.2 인식부

인식부는 검출된 손과 손가락을 바탕으로 양 손을 왼손과 오른손으로 판별하고, 각 손가락을 엄지, 검지, 중지, 약지, 소지로 판별한다. 손 판별은 검출된 최소 손 영역이 한 손일 때와 양 손일 때의 상황을 고려하여 판별한다. 양 손이라면 위상학적 판단으로 왼쪽에 있는 영역을 오른손, 오른쪽에 있는 영역을 왼손이라 판별한다. 손가락 판별은 그림 12와 같은 제안하는 손가락 뼈대 판별 알고리즘을 적용한다. 먼저 검출된 손 영역의 개수를 파악한다. 손 영역의 개수가 2개 일 때는 중심점은 2개, 손가락 끝점은 10개로 초기화 하고, 손 영역의 개수가 1개 일 때는 중심점은 1개, 손가락 끝점은 5개로 초기화 한다. 이후 다음과 같이 손의 뼈대를 만든다. 뼈대는 중심점을 $C(x_c, y_c)$ 라 하고, 각 손가락 끝점을 $F_i(x_i, y_i)$, ($i=0\sim4$)라 두면 각 손가락 뼈대 L_i 는 C 와 F_i 를 연결하는 직선으로 정의할 수 있다. 그리고 L_i 의 길이를 $Length_L_i$ 로 정의한다. 각 뼈대의 $Length_L_i$ 를 중 가장 짧은 뼈대 2개를 구하여 엄지 후보 뼈대를 찾고, 그

엄지 뼈대 후보를 Ls_1 , Ls_2 라 한다. 엄지 뼈대 후보 중 임계치보다 작은 뼈대를 최종 엄지 뼈대라 판별하고, 임계치보다 작거나 혹은 임계치보다 클 경우 인접 뼈대 간의 각도를 이용해 최종 엄지 뼈대를 판별한다. Ls_1 의 인접 뼈대 간

그림 12. 제안하는 손가락 뼈대 판별 알고리즘

Fig. 12. Proposed finger skeleton classification algorithm

각도를 Angle_s1이라 하고, Ls2의 인접 뼈대 간 각도를 Angle_s2라 한다면 최대 각도를 갖는 뼈대를 최종 엄지 뼈대로 판별한다. 엄지 뼈대를 판별한 후, 나머지 손가락의 뼈대는 엄지와 가장 가까이 위치한 뼈대를 차례로 검지, 중지, 약지, 소지 뼈대로 판별한다. 이후 판별된 손가락 뼈대를 이용하여 손가락 끝점이 해당 뼈대 위에 위치해 있거나 해당 뼈대의 근사한 위치에 있을 경우 그 손가락 끝점을 해당 뼈대의 손가락 끝점으로 판별한다. 실시간 응용을 위해서 매 입력 시 새로운 뼈대를 생성하지 않고, 검출부만 수행하여 획득된 최소 사각형을 이전 시점의 최소 사각형과 비교하여 이동 및 회전 변환 계수를 구하고, 이를 이전 시점에서 생성된 뼈대에 적용함으로써 뼈대 생성을 위한 복잡도를 감소시킨다. 다음으로 손의 자세와 각 손가락의 퍼짐을 판단하여 최종적인 손 자세를 인식한다.

2. MPEG-U XML 생성부

MPEG-U XML 생성부는 MPEG-U part2에서 정의한 스키마 구조에 따라서 MPEG-U part2 표준을 따르는 XML 문서를 생성한다. MPEG-U part2는 모든 데이터 타입에 대

한 스키마 구조를 XSD 문서로 지원한다. 만약 사용자가 인식된 손과 손가락을 이용해 가위(Scissors)라는 포즈를 취하게 되면, MPEG-U XML 생성부에서는 가위 타입의 스키마 구조를 이용하여 해당 XML문서를 생성한다. 문서 안에는 가위 타입의 정보가 포함되어 있으며, 손 및 손가락 인식부에서 처리된 파라미터까지 MPEG-U part2 표준에 맞는 X,Y,Z 좌표로 나뉘어 기술한다. 그림 13은 MPEG-U part2 표준에 부합하는 특징을 지닌 포즈에 대하여 MPEG-U part2 표준에 맞게 XML 문서로 파싱한 그림이다.

IV. 실험

본 장에서는 실험을 통하여 Convex Hull 기반의 Convexity Defects를 이용한 방법으로 손가락 끝점을 나타내고, 제안하는 손가락 판별 알고리즘의 인식 결과를 보인다. 또한, 다양한 환경에서도 손 자세 인식이 가능한 점을 실험을 통하여 검증한다. 마지막으로 제안하는 시스템을 MPEG-U part2 참조 소프트웨어로 실험하여 상호연동을 위한 손 및 손가락 인식 기반 MPEG-U 사용자 인터페이스 시스템의 표준 부합성을 검증한다.

1. 손가락 판별 알고리즘 검증

제안하는 시스템은 손을 왼손과 오른손으로 인식하고, 손가락을 엄지, 검지, 중지, 약지, 소지로 인식한다. 실험에서 10개의 손가락을 모두 뻗을 때와 몇 개의 손가락을 뻗을 때, 그리고 이동 및 회전을 하여도 강인하게 인식되는 손 자세를 보인다. 그림 14의 (a)는 10개의 손가락 모두를 뻗을 때 판별된 손과 손가락을 나타낸 실험 결과이고, 그림 14의 (b)는 몇 개의 손가락만 뻗을 때 판별된 손과 손가락을 나타낸 실험 결과이다. 그림 14의 (c)는 이동 및 회전 변화를 주어 사용자가 손 자세를 취했을 때 손과 손가락을 판별하는 실험 결과이다. 그림에서 손 영역의 0과 1은 각각 왼손과 오른손을 나타내며, 손가락 부근의 0~4는 각각 엄지, 검지, 중지, 약지, 소지를 나타낸다.

그림 13. 사용자가 Rect 자세를 취했을 때 파싱된 MPEG-U part2 Geometric Pattern의 Rect XML

Fig. 13. Rect XML of MPEG-U part2 geometric pattern parsed when a user takes the Rect posture

그림 14. 제안하는 손가락 판별 알고리즘으로 인식한 결과 (a) 10개의 손가락을 모두 폈을 때의 결과, (b) 몇 개의 손가락만 폈을 때 결과, (c) 이동 및 회전변화에서의 인식 결과

Fig. 14. Hand posture recognition result of the proposed method (a) result when all the 10 fingers are spread out, (b) result when a few fingers are unfolded, (c) result under translation and rotation of hand

2. 다양한 환경에서의 손 자세 인식 검증

제안하는 시스템은 3차원 영상 데이터를 활용하기 때문에 다양한 환경 변화에 강인하다. 3차원 영상 데이터를 활용하여 손을 검출하는 Candescent NUI는 다양한 환경 변화에 강인하고, 계산량이 적다. 실험에서 피부색과 유사한 배경 및 조명 변화에서 제안 방법의 성능을 보인다. 그림 15의 (a)는 밝기 값이 큰 밝은 배경에서 인식되는 손 자세 인식 결과를 나타낸 그림이고, 그림 15의 (b)는 밝기 값이 낮은 어두운 배경에서 인식되는 손 자세 인식 결과를 나타낸 그림이다. 그림 15의 (c)는 피부색과 유사한 배경에서 인식되는 손 자세를 나타낸 그림이다. 그림에서와 같이 제안 방법이 조명이나 환경 변화에 강인함을 알 수 있다.

그림 15. (a) 밝기 값이 큰 밝은 배경에서의 손 자세 인식 결과, (b) 밝기 값이 낮은 어두운 배경에서의 손 자세 인식 결과, (c) 피부색과 같은 배경에서의 손 자세 인식 결과

Fig. 15. (a) Recognition result of the hand posture under the bright background, (b) recognition result of the hand posture under the dark background, (c) recognition result of the hand posture under the background having the similar color with the human skin

3. 손 자세 인식에 의한 MPEG-U 사용자 인터페이스 시스템 검증

MPEG-U part2에서 정의하는 6가지 패턴들 중 Geometric Pattern, Symbolic Pattern 대한 타입을 실험하여 제안하는 시스템을 검증한다. 실험에 쓰일 타입들은 Geometric Pattern의 Rect와 Symbolic Pattern의 Scissors 이다. Geo-

그림 16. 손 및 손가락의 포즈로 표현한 (a) Rect, (b) Scissors
Fig. 16. (a) Hand posture of Rect, (b) Hand posture of Scissors

그림 17. 파싱된 MPEG-U part2 표준에 맞는 XML 문서 (a) Rect, (b) Scissors
Fig. 17. XML document in accordance with the MPEG-U part 2 standard (a) Rect, (b) Scissors

metric Pattern의 Rect는 Top Left Position과 Bottom Left Position, Top Right Position, Bottom Right Position의 데이터가 필요하고, Symbolic Pattern의 Scissors는 Position과 Size의 데이터가 필요하다. 제안하는 시스템에서 Rect와 Scissors를 표현하기 위해 Rect는 구분 되어진 양 손의 엄지와 검지를 활용하고, Scissors는 한 손의 엄지와 검지, 손 영역의 크기를 활용한다. 그림 16은 인식된 손 자세의 포즈로 Rect와 Scissors를 표현한 그림이고, 그림 17은 Rect의 포즈 정보와 Scissors의 포즈 정보를 MPEG-U part2 표준에 맞는 XML 문서로 파싱한 그림이다. 파싱된 XML 문서의 타당성 검증을 위해 MPEG-U part2 참조 소프트웨어로 검증하였다. 먼저, 파싱된 XML문서를 MPEG-U part2 참조 소프트웨어에 그림18, 19의 (a)와 같이 전달한다. 이후, 언마셜(Unmarshal) 작업을 통해 각 해당되는 좌표 값이 그림 18, 19의 (b)와 같이 추출하고, 그 추출된 좌표 값을 사용자

그림 18. Scissors의 XML 문서를 입력 받아 좌표로의 출력과 GUI로의 출력
Fig. 18. X-Y coordinate output and GUI output from the XML document of Scissors

그림 19. Rect의 XML 문서를 입력 받아 좌표로의 출력과 GUI로의 출력
Fig. 19. X-Y coordinate output and GUI output from the XML document of Rect

상호작용 인터페이스 시스템의 GUI로 그림 18, 19의 (c)와 같이 출력한다. 이처럼 MPEG-U 표준에 맞는 XML 문서를 다른 기기에 전달하고, 다른 기기는 표준에 적합한 XML 파서를 갖추고 있다면 동일한 결과를 얻을 수 있다.

V. 결 론

본 논문에서 제안하는 시스템은 깊이 카메라를 이용한 손 자세 인식에 의한 MPEG-U 기반 향상된 사용자 인터페이스 시스템이다. 깊이 카메라를 이용하여 보다 강인하게 인식되는 손과 손가락을 보였고, 손가락을 판별하는 제안하는 알고리즘은 다양한 이동 또는 회전 변화에서 강인한 판별을 할 수 있고, 접었다 펴는 동작에서도 강인한 판별을 할 수 있다. 또한, MPEG-U 기반으로 서로 다른 장면 기술 장치들을 표준 기반으로 상호 연동할 수 있다는 점과 사용

자 인터페이스 장치와 장면 기술 장치 간에 표준 기반의 인터페이스를 보였다. 제안하는 시스템은 추후 멀티모달 인터페이스 시스템에서 모달리티로서의 활용이 가능할 것이다. 또한, 스마트 TV나 스마트폰 등에서 MPEG-U를 활용하는 기술 개발을 추진한다면 더욱 다양한 인터페이스를 실현할 수 있을 것이다.

참 고 문 헌 (References)

- [1] Information technology - Rich media user interfaces - Part 2: Advanced user interaction (AUI) interfaces. - ISO/IEC 23007, Feb. 2012.
- [2] G. Han, A. Baek, and H. Choi, "MPEG-U part2 based Advanced User Interaction Interface System," Journal of Korea Contents Association, Vol. 12, No. 12, pp. 54~62, Dec. 2012.
- [3] Robert Y. Wang and Jovan Popović, "Real-Time Hand-Tracking with a Color Glove," ACM Transactions on Graphics, Vol. 28, Issue. 3, No. 63, Aug. 2009.
- [4] R. Lockton and A. Fitzgibbon, "Real-time gesture recognition using deterministic boosting," BMVC, pp. 1~10, Sep. 2002.
- [5] Anastasios Roussos, Stavros Theodorakis, Vassilis Pitsikalis and Petros Maragos, "Hand tracking and affine shape-appearance hand-shape sub-units in continuous sign language recognition," ECCV Workshop on Sign, Gesture and Activity, Hersonissos, Crete, Greece, Sep. 2010.
- [6] Yuh-Rau Wang, Wei-Hung Lin, and Ling Yang, "A novel real time hand detection based on skin-color," Consumer Electronics (ISCE), IEEE 17th International Symposium on, pp. 141~142, Jun. 2013.
- [7] Xintao Li, Can Tang, Chun Gong, Sheng Cheng and Jianwei Zhang, "Hand Segmentation Based on Skin Tone and Motion Detection with Complex Backgrounds," Chinese Intelligent Automation Conference, Springer Berlin Heidelberg, Vol. 256, pp. 105~111, Jan. 2013.
- [8] V. Argyros and S. Sclaroff, "Database indexing methods for 3D hand pose estimation," Gesture Workshop, pp. 288~299, Apr. 2003.
- [9] Candescent NUI Samples & Source code <http://candescentnui.codeplex.com/> SourceControl/latest#CCT.NUI.Visual/ClusterLayer.cs.
- [10] S. Malik, "Real-time hand tracking and finger tracking for interaction," CSC2503F Project Report, Department of Computer Science, University of Toronto, Dec. 2003.
- [11] C. Davatzikos and J. L. Prince, "Convexity analysis of active contour problems," Image Vision Computing, Vol. 17, pp. 27~36, Jan. 1999.
- [12] I. Oikonomidis, N. Kyriazis and A. Argyros, "Efficient Model-based 3D Tracking of Hand Articulations using Kinect," BMVC, pp. 101.1~101.11, Sep. 2011.
- [13] Junyeong Choi, Hanhoon Park and Jong-Il Park, "Hand shape recognition using distance transform and shape decomposition," Image Processing(ICIP), pp. 3605~3608, Sept. 2011.
- [14] KINECT. <http://en.wikipedia.org/wiki/Kinect>.

[15] Sara Taskinen and David I, "Robust estimation and inference for bi-variate line fitting in allometry," Biometrical Journal, pp. 652-672, Jun. 2011.

[16] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Distance Transforms of Sampled Functions," Theory of Computing, Vol. 8, pp. 415-428, Sep. 2012.

저 자 소 개

한 국 희

- 2012년 2월 : 한밭대학교 멀티미디어공학과 (공학사)
- 2012년 3월 ~ 현재 : 한밭대학교 정보통신전문대학원 멀티미디어공학과 (공학석사)
- 주관심분야 : 영상처리, 컴퓨터비전, UI/UX, 모바일 프로그래밍

이 인 재

- 1999년 2월 : 성균관대학교 전자공학과 (공학사)
- 2001년 2월 : 성균관대학교 전기전자및컴퓨터공학부 (공학석사)
- 2001년 1월 ~ 현재 : 한국전자통신연구원 선임연구원
- 주관심분야 : 리치미디어, 영상처리, 컴퓨터비전, UI/UX

최 해 철

- 1997년 2월 : 경북대학교 전자공학과 학사
- 1999년 2월 : 한국과학기술원 전기및전자공학과 (공학석사)
- 2004년 8월 : 한국과학기술원 전기및전자공학과 (공학박사)
- 2004년 9월 ~ 2010년 2월 : 한국전자통신연구원(ETRI) 방송미디어연구부 선임연구원
- 2010년 3월 ~ 현재 : 한밭대학교 정보통신공학과 조교수
- 주관심분야 : 영상통신, 비디오 부호화, 패턴 인식