

멀티미디어 콘텐츠 제어를 위한 사용자 경험 기반 동작 인식 기술

김윤식[†], 박상윤^{**}, 옥수열^{***}, 이석환^{****}, 이응주^{*****}

요 약

본 논문에서는 다양한 멀티미디어 매체의 제어 및 인터랙션을 위하여 별도의 입력장치 없이 사용자의 경험기반의 동작 인식을 통하여 멀티미디어 콘텐츠를 제어 하는 알고리즘을 제안하였다. 본 논문에서 제안한 사용자 경험 기반 멀티미디어 콘텐츠 제어를 위한 휴먼 동작 인식 방법은 먼저, 카메라로부터 입력받은 영상을 조명의 변화에 크게 영향을 받지 않는 YCbCr 컬러 영역으로 변환하여 피부색 추출과 모폴로지에 의한 잡음제거, Boundary Energy 및 Depth 영상을 이용하여 손 영역을 검출하였다. 검출된 손 영상에서 PCA 알고리즘을 이용하여 손 모양을 인식하고 차영상 및 모멘트 이론을 이용하여 손의 중심점 검출 및 궤적을 획득한 후, 손의 궤적을 시간을 기준으로 8분할하여 8방향 체인코드를 이용하여 심볼화하였다. 심볼화 된 정보로부터 HMM 알고리즘을 이용하여 손동작을 인식, 사용자의 동작 인식을 통하여 멀티미디어 콘텐츠를 제어하도록 하였다. 제안한 알고리즘을 실험에 적용한 결과 손 영역 검출은 94.25%, 손 모양 인식은 92.6%, 손 동작 인식은 85.86%, 얼굴 검출은 89.58%의 성능을 나타내었으며 이를 기반으로 컴퓨터 환경에서 생성 구축된 영상, 음성, 동영상, MP3, e-book 등과 같은 다양한 콘텐츠들을 동작인식만으로 제어할 수 있도록 하였다.

Human Gesture Recognition Technology Based on User Experience for Multimedia Contents Control

Yun-Sik Kim[†], Sang-Yun Park^{**}, Soo-Yol Ok^{***}, Suk-hwan Lee^{****}, Eung-Joo Lee^{*****}

ABSTRACT

In this paper, a series of algorithms are proposed for controlling different kinds of multimedia contents and realizing interact between human and computer by using single input device. Human gesture recognition based on NUI is presented firstly in my paper. Since the image information we get it from camera is not sensitive for further processing, we transform it to YCbCr color space, and then morphological processing algorithm is used to delete unuseful noise. Boundary Energy and depth information is extracted for hand detection. After we receive the image of hand detection, PCA algorithm is used to recognize hand posture, difference image and moment method are used to detect hand centroid and extract trajectory of hand movement. 8 direction codes are defined for quantifying gesture trajectory, so the symbol value will be affirmed. Furthermore, HMM algorithm is used for hand gesture recognition based on the symbol value. According to series of methods we presented, we can control multimedia contents by using human gesture recognition. Through large numbers of experiments, the algorithms we presented have satisfying performance, hand detection rate is up to 94.25%, gesture recognition rate exceed 92.6%, hand posture recognition rate can achieve 85.86%, and face detection rate is up to 89.58%. According to these experiment results, we can control many kinds of multimedia contents on computer effectively, such as video player, MP3, e-book and so on.

Key words: Human Gesture Recognition(휴먼 동작 인식), HMM, PCA, Multimedia Contents

※ 교신저자(Corresponding Author) : 이응주, 주소 : 부산광역시 남구 용당동 535번지 동명대학교 정보통신공학과 (608-711), 전화 : 051)629-3700, FAX : 051)629-1143, E-mail : ejlee@tu.ac.kr

접수일 : 2011년 12월 29일, 수정일 : 2012년 7월 16일

완료일 : 2012년 8월 10일

[†] 준회원, 동명대학교 정보통신공학과
(E-mail : radio025@naver.com)

^{**} 정회원, 동명대학교 정보통신공학과
(E-mail : co412nan@naver.com)

^{***} 종신회원, 동명대학교 게임공학과
(E-mail : sooyol@tu.ac.kr)

^{****} 정회원, 동명대학교 정보보호학과
(E-mail : skylee@tu.ac.kr)

^{*****} 종신회원, 동명대학교 정보통신공학과

1. 서 론

최근 정보통신 기술의 급속한 발전과 3DTV, IPTV, 스마트폰 등 다양한 멀티미디어 기기들이 나타남에 따라 기존의 키보드, 마우스, 디스플레이와 같은 입출력시스템들의 한계를 뛰어 넘어서 인간의 경험적 동작 행위들을 카메라나 센서 등을 통하여 직접 인식하여 기기들을 동작, 제어하는 인간과 기기와의 상호 교감을 위한 쌍방향 인터페이스 수단으로 활용하고자 하는 HCI, NUI 관련 연구들이 활발히 진행되고 있다[1-8]. 특히, 이와 같은 연구의 흐름 중에서 컴퓨터, 스마트 기기, 디스플레이 매체, 게임분야, 재활복지서비스 분야, 콘텐츠 등의 다양한 분야에 있어서, 인간의 신체 행위 동작과 같은 동작행위 정보를 검출, 추적, 인식하여 사용자 중심의 새로운 인터페이스 수단으로 하여 기기를 동작 제어하는 연구에 대한 필요성이 빠르게 확산 진행되고 있다.

따라서 본 논문에서는 기존의 입력 방식에서 벗어난 자유 형태의 인터페이스로서 휴먼 동작 인식을 분석해서 이를 컴퓨터와의 인터페이스로 활용, 사용자와 컴퓨터 간의 인터랙션을 통해 멀티미디어 콘텐츠 제어 등에 활용하는 “NUI 기반 멀티미디어 콘텐츠 제어를 위한 휴먼 동작 인식 기술”에 대한 연구를 수행하고자 한다.

본 논문에서는 휴먼동작 인식을 하기 위해서는 크게 5단계로 분류 한다. 첫 번째 단계에서는 YCbCr 컬러 모델과 Boundary Energy, Depth 영상을 이용하여 손 영역을 검출하고, 두 번째 단계에서는 검출된 손 영역을 이용하여 PCA 알고리즘을 적용하여 손 모양을 인식한다. 세 번째 단계에서는 검출된 손 영역에서 Moment와 차영상을 이용하여 손 중심과 궤적을 검출하고, 네 번째 단계에서는 손의 궤적을 8방향 체인코드로 심볼화하여 HMM 알고리즘에 적용하여 손 모양을 인식하도록 하였다. 그리고 마지막 단계에서는 제안된 휴먼 동작 인식 기술을 멀티미디어 콘텐츠 제어에 적용을 하였다. 본 논문의 마지막에서는 다섯 가지의 손 모양 인식과 다섯 가지의 손동작을 인식하여 각각의 멀티미디어 콘텐츠에 적용한 결과 및 인식성능에 대하여 분석하였다.

2. 멀티미디어 콘텐츠 제어를 위한 휴먼 동작 인식 기술

본 논문에서는 멀티미디어 콘텐츠 제어를 위해서

휴먼동작 인식 방법을 다음과 같이 제안한다. 피부색 후보 영역을 검출하기 위해 YCbCr 컬러 스페이스와 모폴로지를 이용하고, 검출된 피부영역에서 Depth 영상과 Boundary Energy를 이용하여 손 영역을 검출하고 손 모양 인식과 손동작 인식을 수행한다. 손 모양 인식을 기존의 PCA 알고리즘으로 학습된 데이터를 이용하여 5가지의 손모양(1-5)을 인식한다.

손동작 인식은 두 단계로 나뉜다. 첫 번째 단계는 손의 중심점을 Moment를 이용하여 구하고, 차영상을 이용하여 중심점의 궤적을 찾는다. 두 번째 단계로는 중심점의 궤적을 8방향 체인코드를 이용하여 심볼화 시키고, 심볼된 코드값을 HMM 알고리즘을 이용하여 5가지의 손동작을 인식 하게 된다. 제안한 휴먼동작 인식의 처리 과정은 다음과 같다.

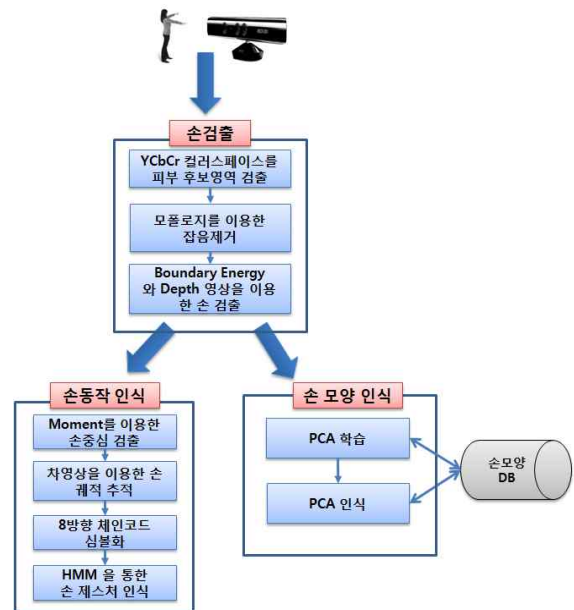


그림 1. 제안된 휴먼 동작 인식 알고리즘 흐름도

2.1 컬러기반 손 영역 검출

손 영역 검출은 손 모양 인식과 손 동작 인식을 위한 전처리 과정으로 정확하게 손을 추출하는 것이 중요하다. 손 영역과 배경을 정확히 분류함으로써 손 모양 및 동작 인식이라든지 기타 손의 특징요소를 정확 하게 검출 할 수 있다. 손 영역을 검출하기 위해 컬러영상을 RGB좌표계에서 YCbCr로 변환하여 피부색이 될 수 있는 후보영역을 검출하고, 손 영역 검출의 개선을 위한 잡음 제거로서 후보영역을 이진화하여 Opening을 이용하여 기타 노이즈를 제거 한다.

노이즈 제거 후 Labeling 하여 각각의 오브젝트를 분리하고, depth 데이터와 Boundary Energy를 이용하여 손 검출을 하게 된다.

2.1.1 YCbCr정보를 이용한 피부색 검출

카메라로부터 입력받은 영상에서 YCbCr정보를 이용하여 피부색의 영역을 추출한다. YCbCr 컬러 모델은 휘도(Y)와 색차신호(Cb, Cr)로 구성되어 있으며, YCbCr 컬러 모델은 피부색 영역이 HSI 색공간 영역과 RGB 색공간의 피부색 영역보다 조밀하다는 것을 실험 영상으로부터 알 수 있었다.

다음과 같이 피부 영역을 검출한다.

$$\text{if}(Y > 128) \text{ then } \begin{cases} \theta_1 = -2 + (256 - Y)/16 \\ \theta_2 = 20 - (256 - Y)/16 \\ \theta_3 = 6 \\ \theta_4 = -8 \end{cases} \quad (1)$$

$$\text{if}(Y \leq 128) \text{ then } \begin{cases} Cr \geq -2(Cb + 24) \\ Cr \geq -(Cb + 17) \\ Cr \geq -4(Cb + 32) \\ Cr \geq 2.5(Cb + \theta_1) \\ Cr \geq 0.5(\theta_1 + Cb) \\ Cr \geq -2(\theta_4 + Cb) \end{cases} \quad (2)$$

식 (1)를 이용하여 임계값을 구하고 식 (2)를 모두 만족하는 범위에 속하는 영역을 피부색 영역으로 추출한다.

2.1.2 검출 개선을 위한 잡음 제거

YCbCr 컬러스페이스를 이용하여 검출된 피부영역은 조명 및 배경등 잡음이 많기 때문에 컬러영상을 이진화 영상으로 전환하고, Opening 기법을 이용하여 잡음을 제거한다. 레이블링 기법을 이용하여 손이라고 판단할 수 있는 오브젝트 추출을 한다.

2.1.3 Boundary Energy와 depth data를 이용한 손 영역 검출

물체의 면적과 둘레는 기하학적인 형태의 두 가지 중요한 속성이다. 하지만 이 두 가지 속성만으로 기하학적인 형태를 분류하는 것은 어렵다. 일반적으로 같은 면적을 가진 물체라도 둘레의 길이는 같지 않다. 이를 구분하기 위해 ‘원형도’라는 기하학적인 형태를 나타내는 특성을 사용한다. 만일 기하학적인 형태가 원이라고 하면 ‘원형도’는 가장 작은 값을 가진다. 그리고 물체가 복잡한 에지를 가진다면 그 값은 커질 것이다. ‘원형도’의 값은 에지의 복잡도를 나타

낸다. ‘원형도’의 일반적인 식은 아래 식(3)과 같다.

$$C = P^2/A \quad (3)$$

P는 물체 영역의 둘레 길이이며 A는 면적을 나타낸다. 만약 물체가 원이라면 $4\pi r$ 가 된다. 그리고 물체가 더욱 복잡한 에지를 갖는다면 그 값은 더 커질 것이다. 원형도(C)와 에지 복잡도는 서로 비례하는 관계를 갖는다. 이 원형도, 즉 에지 복잡도를 boundary energy라고 한다. 물체의 둘레의 길이가 P 이고 시작점에서 임의의 에지점까지의 거리를 p 라고 가정하자. 어떤 점에서든지 에지는 곡률 반지름r(p) 을 갖는다. 에지점에서의 탄젠트 값은 원의 반지름이다. 따라서 p 점에서의 곡률 함수는 식(4)와 같다.

$$K(p) = 1/r(p) \quad (4)$$

K(p) 함수는 P 주기를 갖는 주기 함수이다. 식(5)에 의해 boundary 길이의 average energy가 계산되어 진다.

$$E = \frac{1}{P} \int_0^P |K(p)|^2 dp \quad (5)$$

원형의 경우는 가장 작은 boundary energy 값을 갖게 된다.

$$E_0 = \left(\frac{2\pi}{P}\right)^2 = \left(\frac{1}{R}\right)^2 \quad (6)$$

R은 원의 반지름이며 boundary energy 값은 에지 복잡도를 사람이 인지하는 것과 유사하게 나타낼 수 있다.

지금까지의 분석으로 사람의 손 영역은 영상의 많은 부분을 차지하고 또한 복잡한 에지를 가지고 있다. 따라서 이 두 가지 특징을 식(7)을 이용하여 증강시킬 수 있다.

$$\omega = a^{E^*A} \quad (a > 1) \quad (7)$$

여기서 a는 증강 계수이며, 그 결과 값을 조정할 수 있다. 각각의 후보 영역에 적용하여 ω 값을 구한다. 그리고 큰 값을 갖는 2개의 영역을 추출한다.

Depth 데이터를 얻기 위해서 MS사에서 개발하여 판매중인 kinect 카메라 센서를 이용하였다. 이 센서의 특징은 Color 영상과 Depth 영상 두 가지를 제공한다. 아래 그림 2(a)은 Kinect 센서의 depth 영상이다. 카메라에 가까운 객체일수록 Depth(깊이값)가 작고, 카메라에서의 먼 객체는 depth가 클 것이다. depth 값의 표현은 8비트(0~255)로 구성 되어 있기

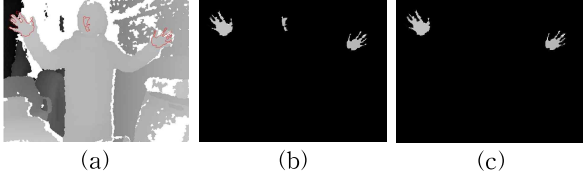


그림 2. depth 영상 결과; (a) depth 영상, (b) 후보 영역 추출, (c) depth 오브젝트 2개 추출 결과.

때문에 depth값이 작다면 검은색(0)에 가까우며, depth 값이 크다면 흰색(255)에 가까울 것이다. 피부 색의 후보영역과 같은 depth 영상의 영역의 데이터를 확인하여, depth값이 작은 2개영역을 추출한다.

Boundary Energy에서 2개의 오브젝트와 Depth 영상에서 2개의 오브젝트를 이용하여 2개의 오브젝트가 같고, 또한 거리 차이가 거의 없을 경우 손이라고 판단한다. 실험환경에서는 얼굴보다 손이 앞에 있도록 실험을 하였다.

2.2 주성분 방법을 이용한 PCA 알고리즘

주성분 분석법(PCA, Principal Component Analysis)는 다차원의 값들을 최대한 멀리 떨어져 있는 것은 멀리 떨어져 있고, 가까운 것은 가까워 보이도록 2차원이나 3차원, 혹은 다른 차원으로 투영하는 기법으로, 다차원 특징 벡터로 이루어진 데이터에 대하여 높은 차원에서의 정보를 유지하면서 낮은 차원으로 차원을 축소시키는 다변량 데이터 처리 방법 중의 하나이다. PCA의 또 다른 장점은 데이터의 패턴을 찾고 그 데이터를 압축할 수 있다. PCA는 학습 및 인식의 두 단계로 분리 되어진다.

2.2.1 학습

인식을 하기 위해서 사전에 Database의 손 영상을 기반으로 하여 새로운 데이터가 입력 되었을때 비교를 하기 때문에 학습데이터가 필요하다.

$$T = \begin{pmatrix} T_1 (100 \times 100 = 10000) \\ T_2 (100 \times 100 = 10000) \\ T_3 (100 \times 100 = 10000) \\ T_4 (100 \times 100 = 10000) \\ \dots \\ T_N (100 \times 100 = 10000) \end{pmatrix} \in R^{N \times 10000} \quad (8)$$

PCA는 공분산을 이용하기 때문에 공분산 matrix를 만들어야 한다. 공분산 matrix를 만들기 위해서는 각 이미지들의 평균을 다음과 같은 식을 통해 구한다.

$$\bar{T} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{10000} T_{ij} \quad (9)$$

앞서 말한 것과 같이 얻은 data에서 평균영상의 차를 구한다. 그의 결과는 다음과 같다.

$$\Phi_T = [T - \bar{T}]^T = \begin{pmatrix} T_1 - \bar{T} (100 \times 100 = 10000) \\ T_2 - \bar{T} (100 \times 100 = 10000) \\ T_3 - \bar{T} (100 \times 100 = 10000) \\ \dots \\ T_N - \bar{T} (100 \times 100 = 10000) \end{pmatrix}^T \in R^{10000 \times N} \quad (10)$$

공분산 행렬을 구하기 위해 위의 수식과 그것의 전치행렬을 곱한다. 이때 주의할 점은 차원이 10,000개이기 때문에 공분산 행렬의 차원 또한 10,000개나 나와야 한다. 그러므로 공분산 행렬은 다음과 같은 식을 통해 10,000*10,000의 크기를 갖는 행렬을 얻을 수 있다.

$$C = \frac{1}{N} \Phi_T \Phi_T^T \in R^{10000 \times 10000} \quad (11)$$

PCA의 알고리즘 중에 가장 중요한 부분은 고유 벡터 (또는 고유 손)를 찾는 과정이다. 손 영상의 각 픽셀이 손공간상에서 구성된다고 할 때, 고유 손이란 손 공간을 구성하는 기저벡터에 해당하여 모든 인식 후보 손의 공통적인 특징을 가장 잘 나타내는 기저 벡터를 의미한다. 즉, 고유 손은 모든 인식 후보 손 영상의 평균 손 영상과 각 후보 손 영상의 차 벡터로 공분산 행렬의 고유벡터에 해당한다. 공분산 행렬의 고유치는 평균 손 영상에 대한 분산의 정도를 나타내기 때문에 가장 큰 고유치에 대응하는 고유벡터로 이루어진 고유 손이 가장 비슷한 손이 되며, 고유치가 작아질수록 손 특성이 작아진다.

이 단계까지가 손인식을 위한 PCA의 학습 단계가 끝나게 된다. 또한, 고유값과 고유 벡터 에 대한 공식은 다음과 같다.

$$C u_k = \lambda_k u_k (u_i^T u_j = 0, \forall i \neq j, u_k \in 3600 \times 1, k = 1 \sim K) \quad (12)$$

$$C [u_1 \ u_2 \ \dots \ u_k] = [\lambda_1 u_1 \ \lambda_2 u_2 \ \lambda_3 u_3 \ \dots \ \lambda_k u_k] \quad (13)$$

$$= [u_1 \ u_2 \ \dots \ u_k] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_k \end{bmatrix} \quad (14)$$

2.2.2. 손 인식

고유 손은 다른 영상을 표현하는데 있어서 기저벡터가 된다. Database에 있는 사진들도 각각의 고유 손에 대한 weight vector를 갖고 있게 된다.

$$w_k = u_k^T \Phi, k=1, 2, \dots, K \tag{15}$$

Database에 저장되어 있는 각각의 사진들에 대해 다음과 같이 weight vector를 구한다.

$$\Omega_N = [w_1 \ w_2 \ \dots \ w_k] \tag{16}$$

새로운 입력 영상이 들어오게 되면 Database에 존재하는 sample data의 차 영상과 고유벡터를 통해 weight vector를 구한다. 입력된 weight vector와 Database의 사진에 대한 weight vector가 존재하기 때문에 이 두 개 벡터의 차가 최소로 되는 것의 사진을 선택하면 된다.

$$\Omega_{input} = [w_1 \ w_2 \ \dots \ w_k] \tag{17}$$

$$\Omega_N = [w_1 \ w_2 \ \dots \ w_k] \tag{18}$$

이때 Ω_N 과 Ω_{input} 의 벡터의 거리를 계산을 하는데 있어서 여러 가지 방법이 있는데 대표적으로 Euclidean distance 많이 쓴다. 두 점 $p1(x_1, y_1)$ 과 $p2(x_2, y_2)$ 가 있다고 가정 할 때 식(19)와 같이 사용하여 벡터의 거리를 계산한다.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{19}$$

2.3 HMM 기반 손동작 인식 알고리즘

제안한 알고리즘의 손 동작을 인식하는 과정은 다음과 같이 수행된다. 첫 번째 검출된 손 영역에서 모멘트를 이용하여 중심점을 검출하고, 현재 프레임과 이전 프레임을 비교하여 중심점의 이동이 있을지 추적을 하게 된다. 추적이 끝난 후에 궤적을 시간의 기준으로 하여 8구간으로 분류하고, 8 방향 체인코드를 이용하여 심볼값을 구한다. 심볼값을 HMM알고리즘을 이용하여 손동작을 인식하게 된다.

2.3.1 Moment와 차영상을 이용한 손중심 궤적 추적

손바닥 영역은 상대적으로 크고 원과 비슷한 모양을 가지며 손가락은 상대적으로 작고 긴 사각형 모양의 형태학적인 특징을 가지고 있다. 매우 유용하고 실용적인 형태를 기술하는 집합은 모멘트 이론을 기

반으로 한다. 손의 형태와 손바닥의 중심, 그리고 손의 방향을 기술하는데 모멘트 이론을 적용 할 수 있다[5]. Moment를 이용하여 손 중심점을 검출한 후 이전 프레임과 현재 프레임의 차영상을 이용하여 손의 중심의 이동경로를 점검한다. 검출된 손 영역에서 움직임이 발생하게 되면 현재 지점에서 발생된 지점의 위치값을 기록하는 방식으로 움직임이 발생하지 않을 때까지 기록하게 된다.

2.3.2 8방향 체인코드를 이용한 궤적 심볼화

손의 궤적을 8방향 체인코드에 적용시키기 위해 시간을 기준으로 하여 8구간으로 양자화 하였다. 8구간의 영역의 Start Point 와 End Point 구한다.

손의 궤적을 8 구간으로 등분하고, 각각의 영역의 Start Point 와 End Point를 이용하면 두점 사이의 기울기, 두점 사이의 거리, 각도를 구할수 있다. 그림 3은 Start Point와 End Point 두 점 사이의 관계를 나타내고 있다.

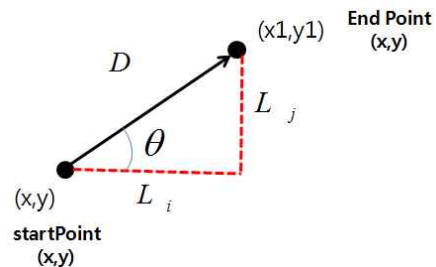


그림 3. Start Point 와 End Point

$$L_i = (x_1 - x) \tag{20}$$

$$L_j = (y_1 - y) \tag{21}$$

$$\theta = \text{atan}\left(\frac{L_j}{L_i}\right) \tag{22}$$

시간에 따른 손 중심점의 이동 방향은 궤적의 의미 해석에 중요한 정보를 제공한다. 일반적으로 체인코드는 연속된 두 점 들간의 방향성을 표현하는데 가장 널리 이용되며 방향을 몇 단계로 세분화하느냐에 따라 4-방향 체인코드 8-방향 체인코드 12-방향 체인 코드 등으로 확대 가능하다. 4-방향 체인 코드는 두 점간의 각도가 90도 이내로 변하지 않으면 같은 코드를 할당하며, 8-방향 코드는 두 점간의 각도가 45도 이내로 변하지 않으면 같은 코드를 할당한다. 본 논문에서는 가장 일반적인 8-방향 체인 코드

를 사용하였으며 방향에 따른 심볼값을 표 1에 정의하였다.

표 1. 각도값 과 해당 심볼

| θ (Angle) | Symbol |
|---|--------|
| $157.5 < \theta < 202.51$ | 0 |
| $112.5 < \theta < 157.51$ | 1 |
| $67.5 < \theta < 112.51$ | 2 |
| $22.5 < \theta < 67.51$ | 3 |
| $0 < \theta < 360.1$ $0 < \theta < 22.51$ | 4 |
| $292.5 < \theta < 337.51$ | 5 |
| $247.5 < \theta < 292.51$ | 6 |
| $202.5 < \theta < 247.51$ | 7 |

손의 궤적을 시간으로 분배한 후, 각각의 영역의 Start Point 와 End Point를 이용하여 길이 값 및 각도를 구하여 심볼값에 적용한 결과는 그림 4와 같다.

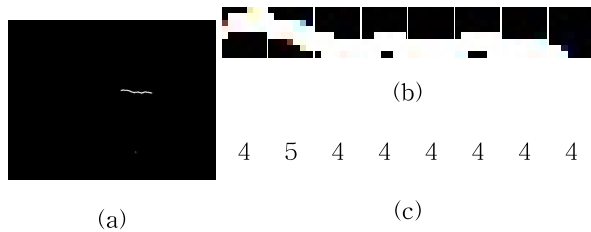


그림 4. 궤적 분할 및 심볼값 (a) 손 중심 궤적, (b) 8분할, (c) 심볼 값.

2.3.3 HMM을 이용한 손 동작 인식

HMM알고리즘을 이용하여 인식할 손동작은 그림 5와 같이 총 5개의 손 동작을 인식하게 된다. 5번째 손 동작은 depth 영상을 이용하여 손 중심점의 z값의 변화를 이용하여 HMM알고리즘에 적용시킨다[9].

| | | | | |
|-----------------|------------------|---------------|-----------------|------------------|
| | | | | front |
| Gesture1 (left) | Gesture2 (right) | Gesture3 (up) | Gesture4 (down) | Gesture5 (click) |

그림 5. 인식할 손동작의 종류

3. 제안된 휴먼 동작 인식 기술 기반 멀티미디어 콘텐츠 제어

3.1 멀티미디어 콘텐츠 구성 및 시스템 동작

본 논문에서 제안한 휴먼 동작 인식 기술을 이용하여 멀티미디어 콘텐츠를 제어하는 것은 3가지 시스템으로 구성되어 진다.

휴먼 동작 인식에서는 메인 프로그램에서 영상을 입력받아 양손의 검출과 얼굴검출 영상 및 휴먼동작 인식 데이터를 메인 프로그램으로 보내준다. 메인 프로그램에서는 휴먼 동작 인식으로부터 검출된 영상 및 데이터를 전송받아 현재 활성화된 콘텐츠를 확인하여 콘텐츠에 명령 데이터를 전달하게 된다. 콘텐츠에서는 현재 활성화된 데이터를 메인 프로그램에 전달하며, 메인 프로그램으로부터 콘텐츠 제어 데이터를 전송받아 각각의 콘텐츠를 실행하게 된다. 멀티미디어 콘텐츠의 시스템도는 그림6과 같다.

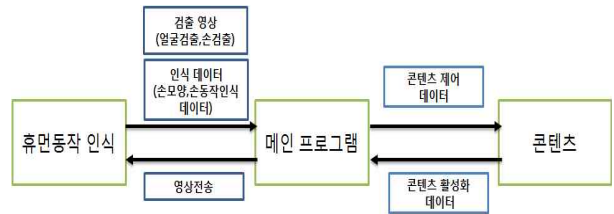


그림 6. 멀티미디어 시스템도

3.2 메인 콘텐츠 프로그램

실험을 위한 메인 프로그램은 그림 7과 같이 4 단계로 분류할 수 있다. 첫번째는 카메라로부터 입력 받은 2D 영상과 Depth영상을 이용하여 얼굴 검출 및 양손 검출 그리고 손 중심점을 구하여 디스플레이 한다. 두 번째는 첫 번째 단계에서 검출된 영상을 이용하여 손동작(1-5) 인식, 손모양 (up, down, right,



그림 7. 메인 콘텐츠 프로그램

left, click)을 인식하여 디스플레이한다. 세 번째는 현재 멀티미디어 콘텐츠 종류를 보여주고 있다. 네 번째는 선택된 콘텐츠를 미리보기 한다.

3.3 휴먼 동작 인식 기반 콘텐츠 제어

이 장에서는 휴먼 동작을 인식 하여 콘텐츠를 제어 하는 부분이다. 콘텐츠의 실행 및 종료 명령인 Click을 제외하고, 실제적으로 총 4개의 손동작 인식을 이용하여 제어하게 된다.

4. 실험결과

4.1 손 검출 결과

본 논문의 실험에서는 손 검출 알고리즘의 효율성을 평가하기 위해서 다양한 손의 위치에 따른 실험 영상을 사용하였다. 손 영역 검출 인식을 위해 배경이 단순한 환경과 복잡한 환경 각각 200장의 영상을 기반으로 하여 실험을 하였다. 표 2는 손 영역 검출 결과이다.

표 2. 손 영역 검출 결과

| 배 경 | 영상 갯수 | 인식 | 오인식 | 인식률 |
|--------|-------|-----|-----|-------|
| 단순한 배경 | 200 | 193 | 7 | 96.5% |
| 복잡한 배경 | 200 | 182 | 18 | 92% |

4.2 손 모양 인식 결과

PCA 학습으로 인해 DB 에는 훈련영상의 데이터가 저장되어 있다. 검출된 양손의 영역중 오른쪽 손을 이용하여 PCA에 알맞은 영상으로 정규화 시킨다. 정규화된 손의 영상은 유클리디언 거리를 계산하고, 제일 가까운 영역의 영상의 값을 읽어내어, 손 모양 인식을 하게 된다.

손 모양 인식을 손의 영역이 검출이 되었다는 가정 아래 실험을 하였다. 손 모양 인식 결과는 다음과 같다.

표 3. 손 모양 인식 결과

| 손 모양 | 영상 수 | 인식 | 오 인식 | 인식률 |
|------|------|----|------|-----|
| 1 | 100 | 94 | 6 | 94% |
| 2 | 100 | 96 | 4 | 96% |
| 3 | 100 | 89 | 11 | 89% |
| 4 | 100 | 93 | 7 | 93% |
| 5 | 100 | 91 | 9 | 91% |

4.3 손동작 인식 제어 결과

손동작을 인식하기 위해 손 중심을 통하여 손의 궤적을 인식하고, 그 궤적을 8방향 체인 코드를 이용하여 심볼화 시키고 , 심볼값을 이용하여 HMM알고리즘을 이용하여 손동작 인식을 하였다.

표 4. 체인코드 심볼값

| 손동작 인식 | 체인코드 심볼 | | | | | | | |
|--------|---------|---|---|---|---|---|---|---|
| right | | | | | | | | |
| | 4 | 4 | 4 | 4 | 5 | 4 | 4 | 4 |
| left | | | | | | | | |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| up | | | | | | | | |
| | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| down | | | | | | | | |
| | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |

표 5는 5종류의 손동작 인식 결과이다. 그리고 실험에 사용한 영상은 손 검출 및 손 영역의 정확한 중심점이 검출되어진 영상을 이용하였다.

표 5. 손동작 인식 결과

| 손 모양 | 영상 수 | 인식 | 오 인식 | 인식률 |
|-------|------|-----|------|-------|
| Right | 121 | 101 | 10 | 83.5% |
| Left | 105 | 97 | 8 | 86.7% |
| Up | 117 | 105 | 12 | 89.7% |
| Down | 123 | 107 | 16 | 87.0% |
| Click | 132 | 108 | 14 | 82.4% |

4.4 멀티미디어 콘텐츠 제어 결과

표 6. 멀티미디어 콘텐츠 인식 결과 - 동영상 플레이어

| 손동작 | 영상수 | 인식 | 오인식 | 인식률 |
|-------|-----|-----|-----|-------|
| up | 134 | 115 | 19 | 85.8% |
| down | 108 | 91 | 17 | 84.3% |
| right | 115 | 97 | 18 | 84.3% |
| left | 129 | 105 | 24 | 81.4% |

표 7. 멀티미디어 콘텐츠 인식 결과 - 이북

| 손동작 | 영상수 | 인식 | 오인식 | 인식률 |
|-------|-----|-----|-----|-------|
| up | 122 | 104 | 18 | 85.2% |
| down | 121 | 102 | 19 | 84.3% |
| right | 119 | 101 | 18 | 84.9% |
| left | 130 | 106 | 24 | 81.5% |

표 8. 멀티미디어 콘텐츠 인식 결과 - MP3 플레이어

| 손동작 | 영상수 | 인식 | 오인식 | 인식률 |
|-------|-----|-----|-----|-------|
| up | 123 | 106 | 17 | 86.1% |
| down | 113 | 96 | 17 | 84.9% |
| right | 121 | 101 | 20 | 83.5% |
| left | 129 | 105 | 24 | 81.3% |

5. 결 론

일반적으로 멀티미디어 콘텐츠를 제어 하는 방법에는 스마트 TV에서는 리모콘, 컴퓨터에서는 마우스 및 키보드가 일반적이다. 그러나 다양한 멀티미디어 기기의 등장과 하드웨어적인 복잡성, 인간중심의 편리성 등을 고려하여 본 논문에서는 별도의 입력장치 없이 손, 머리 동작과 같은 인간의 동작 정보를 인식하여 컴퓨터상의 다양한 멀티미디어 콘텐츠를 제어하는 알고리즘을 제안하였다. 리모콘과 같은 버튼은 대부분 한 버튼에 한가지의 기능을 넣을 수 있고, 버튼이 무한정 많아지면 휴대성이나 복잡한 기능에 의한 조작에 어려움이 있을 수 있으나, 휴먼 동작 인식 방법은 동작을 정의하기에 따라 그 수가 무한으로 늘어날 수 있으므로 그 응용이 무한해 질 수 있다.

따라서 본 논문에서는 카메라로부터 획득한 입력 휴먼 동작 영상을 인식하고 인식된 결과를 컴퓨터와 인터랙션 하도록 알고리즘을 구성하였다. 제안된 방법을 적용하여 실험한 결과 손영역 검출은 94.25% 손모양 인식 92.6% 손동작 인식 85.86% 인식을 보였다.

마지막으로 휴먼 동작 인식을 하기 위해서는 오브젝트의 검출이 무엇보다도 중요하다. 본 논문에서는 컬러 기반으로 하였기 때문에 빛의 세기나 배경색의 영향, 실험하는 시간대 등에 따라서, 색 영역이 변하기 때문에 실험 결과에 영향을 미친다. 또한 피부색과 유사한 배경이 겹치거나 얼굴과 손이 겹치게 되

면, 문제가 발생하게 된다. 이 문제들은 지속적인 연구를 통하여 향후 해결해야 할 과제이다.

참 고 문 헌

- [1] Li Wei and E. J. Lee, "Human Head Mouse System Based on Facial Gesture Recognition," *Journal of Korea Multimedia Society*, Vol. 10, No. 12, pp. 1591-1600, 2007.
- [2] Anh Viet Nguyen, *A Development of Non-touch Human Computer Interface Technology using Gesture Information*, Tongmyong University Master Thesis, 2008.
- [3] Bai Yu and Eung-Joo Lee, "The Hand Mouse: Hand Detection and Hand Tracking," *International Conference on Multimedia, Information Technology and its Application (MITA 2009)*, pp. 244-245, 2009.
- [4] Shang Yiting and Eung-Joo Lee, "The Human Facial Combined Expression Recognition System," *Proceedings of the Korea Multimedia Society Conference*, pp. 187, 2009.
- [5] 박상윤, 휴먼 컴퓨터 인터랙션을 위한 신경회로망 기반 손 제스처 인식, 동명대학교 박사 학위논문, 2010.
- [6] V. Athitsos and S. Sclaroff, "An Appearance-Based Framework for 3D Hand Shape Classification and Camera Viewpoint Estimation," *IEEE Conference on Automatic Face and Gesture Recognition*, pp. 45-52, 2002.
- [7] K. Fukushima, "Neural Network Model for Selective Attention in Visual Pattern Recognition and Associative Recall," *Applied Optics*, Vol. 26, No. 23, pp. 4985-4992, 1987.
- [8] P. Dayan, S. Kakade, and P.R. Montague, "Learning and Selective Attention," *Nature Neuroscience*, Vol. 3, No. 10, pp. 1218-1223, 2000.
- [9] WenKai Xu and Eung-Joo Lee, "Hand Gesture Recognition using Improved Hidden Markov Models," *Journal of Korea Multimedia Society*, Vol. 14, No. 7, pp. 866-871, 2011.



김 은 식

2009년 2월 동명대학교 정보통신공학과 공학사
2012년 2월 동명대학교 대학원 전기전자정보통신공학과 공학석사
관심분야: 영상처리, 컴퓨터비전, 패턴인식, 생체인식



이 석 환

1999년 경북대학교 전자공학과 졸업(공학사)
2001년 경북대학교 전자공학과 졸업(공학석사)
2004년 경북대학교 전자공학과 졸업(공학박사)

2005년~현재 동명대학교 정보보호학과 부교수
2010년~현재 IEEE R10 창원섹션 임원
관심분야: 워터마킹, DRM, 영상신호처리



박 상 운

1996년 2월 경성대학교 물리학과 이학사
1998년 8월 경성대학교 대학원 멀티미디어응용학과 공학석사
2011년 8월 동명대학교 대학원 전기전자정보통신공학과 공학박사

관심분야: 영상처리, 컴퓨터비전, 패턴인식, 생체인식



이 응 주

1992년 2월 경북대학교 대학원 전자공학과 공학석사
1996년 8월 경북대학교 대학원 전자공학과 공학박사
2000년 7월~2002년 2월 (주)디지털넷뱅크 대표이사

2005년 7월~2006년7월 중국 대련경공업대학교 초빙연구 교수

2005년 1월~현재 한국멀티미디어학회 이사
1997년 3월~현재 동명대학교 정보통신공학과 교수
관심분야: 영상처리, 컴퓨터비전, 생체인식



옥 수 열

1994년 2월 동아대학교 산업공학과 공학사
1998년 3월 일본 TSUKUBA대학교 이공학연구과 공학석사
2001년 7월 일본 TSUKUBA대학교 공학연구과 공학박사

2001년 8월~2004년 1월 일본 통신종합연구소(CRL) 연구원

2004년 3월~현재 동명대학교 게임공학과 부교수
관심분야: 게임 인터페이스, 게임 인공지능, 컴퓨터 그래픽스, 가상현실