한국컴퓨터정보학회 하계학술대회 논문집 제18권 제2호 (2010. 7)

실시간 핸드 제스처 추적 및 인식

하정요°, 김계영**, 최형일*
°*숭실대학교 대학원 미디어학과
**숭실대학교 대학원 컴퓨터학과
e-mail: {saint, gykim11, hic}@ssu.ac.kr

Real-Time Hand Gesture Tracking & Recognition

Jeong-Yo Ha°, Gye-young Kim**, Hyung-il Choi*
°*Dept. of Media, Soongsil University
**Dept. of Computer, Soongsil University

● 요 약 ●

본 논문에서는 컴퓨터 비전에 기반을 둔 방법으로 실시간으로 사람의 손의 모양을 인식하는 알고리즘을 제안한다. 기본적인 전처리 과정과 피부 값의 검출을 통해서 사용자의 피부색상을 검출한 후 팔 영역과 얼굴영역을 제거하고, 손 영역만 검출한 뒤 손의 무게중심을 구한다. 그 후에 손의 궤적을 추적하기위해 칼만필터를 이용하였으며, 손의 모양을 인식하기 위한 방법으로 Hidden Markov Model을 이용하여 사용자의 손 모양 6가지를 학습한 후 인식하였다. 실험을 통하여 제안한 방법의 효과를 입증하였다.

키워드: Hand gesture, Skin color, Recognition, Kalman Filter, HMM(Hidden Markov Model)

I. 서론

최근 컴퓨터 하드웨어와 네트워크 기술의 발전에 의해 인간과 컴퓨터 간에 상호간의 의사소통기술을 더욱 필요로 하고 있다. 영상과 음성을 처리하는 멀티미디어 기술 및 마우스, 키보드 등과 같은 사용도구의 발전에 의해 매우 편리하게 되었으나 궁극적으로는 사람과 사람 사이의 의사소통 수준에 이르러야한다. 이를 위해 음성과 시각 정보에 기반을 둔 사용자 접속 인터페이스에 관한 시스템들이 현재 활발하게 개발되어 이미 상용화되고 있으며, 이에 따라 제스처 또한 사용자 접속의 한 파트로서 활발히 연구가 진행되고 있다. 본 논문에서는 카메라를 이용하여 실시간으로 손의 제스처를 인식하는 시스템을 기술하며 논문의 구성은 크게 3단계로 구성하였다. 먼저 첫 번째 단계인 손 영역 추출단계에서는 후속 단계인 특징 추출을 위해 입력받은 영상으로부터 기본적인 전처리과정을 수행한다. RGB 컬러모델을 YCbCr 컬러모델로 변환 후에 스킨컬러를 추출하여 손 영역만을 분리해낸다. 두 번째 단계인 손 영역 추적단계에서는 손 영역의 무게중심을 추출하고 칼만 필터를 이용하여 추적한다. 그리고 마지막으로 손 영역 인식단계에서는 추출한 손을 HMM방법을 이용하여 학습시킨 뒤 최종적으로 손동작을 인식한다. (그림 1)은 본 논문에서 제안하는 제스처 인식과정을 보여준다.

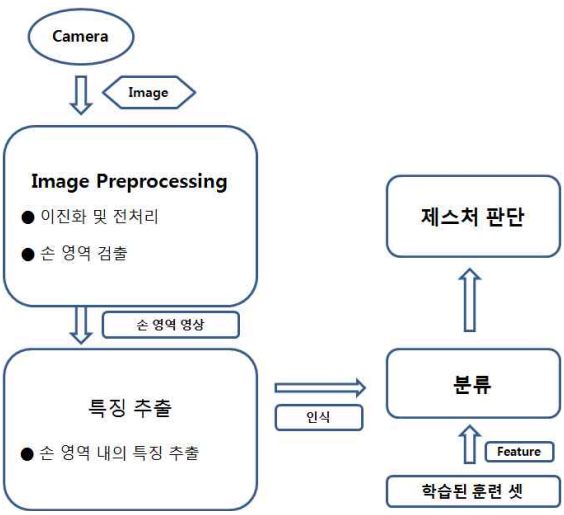


그림 1. 시스템 구조
Fig. 1. System Architecture

II. 손 영역 추출

1. 피부색 추출

사람의 시각이 물체를 구분할 수 있는 여러 특징 중 색상과 모양이 있다. 본 논문에서는 색상정보를 사용하여 손 영역을 획득한다. 최근에는 YCbCr 컬러공간이 RGB 컬러공간에 비해 유용한 에너지 압축을 지원한다는 사실에 기반을 두어 영상 데이터를 처리하거나 압축하는데 휘도-채도 기반컬러공간(luminace-chromonance-based colorspace)을 주로 사용한다. 본 논문에서는 YCbCr 컬러모델을 이용하여 피부색상을 검출하였다. RGB 컬러모델에서 YCbCr 컬러모델로의 변환은 식(1)과 같은 연산과정을 통해 얻을 수 있다.

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.2989R + 0.5866G + 0.1145B \\ -0.16874R - 0.33126G + 0.500B \\ 0.5000R - 0.41869G - 0.0816B \end{pmatrix} \quad (1)$$

피부영역 검출 시 밝기 성분인 Y값이 조명에 대해 영향을 많이 받기 때문에 피부 색상과 직접적으로 관련이 있는 색도 성분인 Cb, Cr 값만을 사용하게 된다. Chai등이 제안한 피부색 참조 맵을 이용하여 식(2)의 피부색 범위를 충족하는 피부색상을 모델링하여 피부영역을 검출한다[1].

$$F(x,y) = \begin{Bmatrix} 1(white) & \text{if}(77 \le C_b \le 127) \\ & \cap (133 \le C_r \le 173) \\ 0(black) & Otherwise \end{Bmatrix} \quad (2)$$

2. 얼굴검출 및 제거

피부색을 추출하는 과정에 손 영역과 함께 추출된 얼굴영역을 제거하기위한 방법으로 에이다부스트(Adaboost)를 이용한 얼굴 검출 알고리즘을 사용하였다. 에이다부스트 알고리즘은 많은 특징 집합으로부터 약한 분류기(Weak classifier)를 구성하고 이를 선택적으로 조합하여 복잡한 패턴을 분류할 수 있는 강한 분류기(Strong classifier)를 학습하는 방법으로 학습된 강한 분류기를 단계적으로 배열하여 얼굴 분류를 하였다.

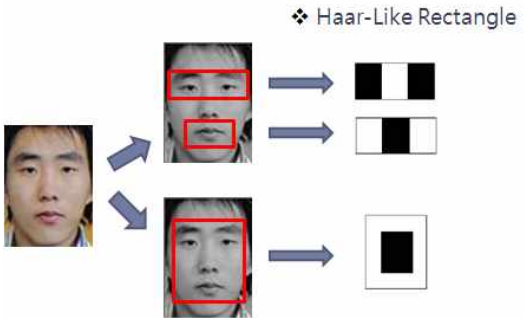
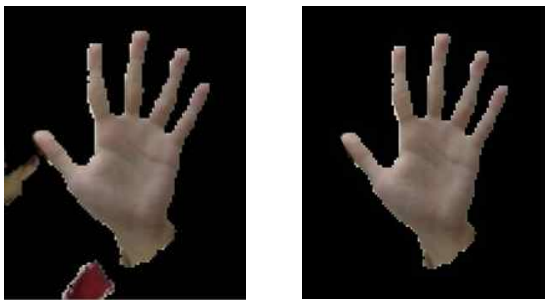


그림 2. Harr-Like를 이용한 얼굴영역추출
Fig. 2. Using Harr-Like face region extract

(그림 2)와 같이 얼굴 영역 내에서 간단하고 객체 간의 구별능력이 있는 특징들을 사용하여 해결하였다.



(a) 얼굴 및 노이즈 제거 전 (b) 얼굴 및 노이즈 제거 후

그림 3. 얼굴 제거 및 노이즈 처리 영상
Fig. 3. Face elimination & Noise processing image

(그림 3)은 피부색 검출을 통하여 함께 추출된 얼굴영역을 제거하고, 손 영역을 분할시킨 후 노이즈의 제거를 위하여 영상으로부터 모폴로지 연산(Closing)을 추가로 수행하였다.

3. 손 영역 무게중심 추출

손 영역 추출 후에 손의 무게중심을 구하기 위하여 수평·수직 프로젝션을 수행하였다. 이진 영상을 수평방향으로 검색하여 흰색의 픽셀 값의 개수를 카운트하여 수평히스토그램을 만들고, 수직방향으로 검색하여 흰색인 픽셀 값의 개수를 카운트하여 수직히스토그램을 만든다. 손 영역 바깥 부분에 노이즈는 임계값 g를 주어 잡음을 제거하였다. 다음의 (그림 4)는 손 영역의 프로젝션 이미지이며, 손 영역의 모든 픽셀들의 합은 식(3)과 같이 나타낼 수 있다.

$$A = \sum_{i=x1}^{x2} \sum_{j=y1}^{y2} B(i,j) \quad (3)$$

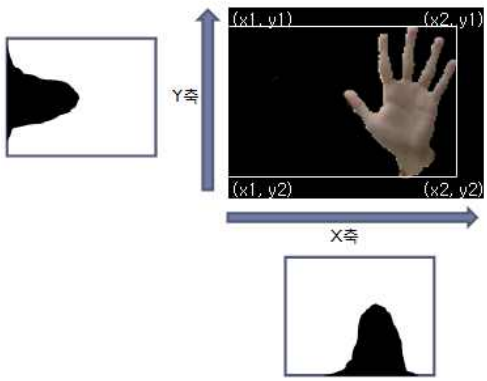


그림 4. x, y축 프로젝션 영상
Fig. 4. x, y projection image

식(3)의 A는 이진 영상에서의 수평영역 $_{x1}$에서 $_{x2}$까지 이고, $_{y1}$에서 $_{y2}$영역까지의 수직영역의 합을 나타낸다. 이진 영상에서 각 화소에 대한 밝기를 그 점에 대한 무게로 간주하면, 면적의 중심의 무게중심과 같아지게 되고 식(4)와 같이 나타낼 수 있으며, 무게중심좌표는 (C_x, C_y)가 된다.

$$C_x = \frac{\sum_{i=x1}^{x2}\sum_{j=y1}^{y2} jB(i,j)}{A}, \quad C_y = \frac{\sum_{i=x1}^{x2}\sum_{j=y1}^{y2} iB(i,j)}{A} \quad (4)$$

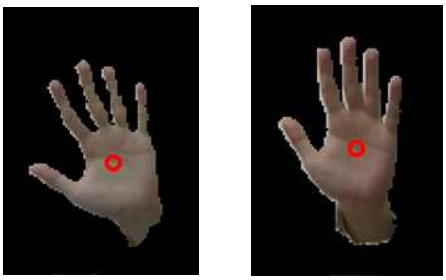


그림 5. 손 영역의 무게중심 추출영상
Fig. 5. Hand Region centroid extract image

위의 (그림 5)는 영상에서 프로젝션 후 손 영역의 무게중심을 추출한 영상이다.

III. 손 영역 추적 및 인식

1. 손 영역 추적

손 색상 모델을 이용한 영역 분리를 수행하여 추출된 후보 손 영역들 중에서 전 시점의 손 영역으로 이전 프레임에서 찾아진 위치에 기반을 두어 한 이미지 프레임에서 손의 위치를 예측하기 위해 칼만 필터를 사용한다[2][3]. 칼만 필터는 가우시안 잡음을 가진 선형 동적 시스템에 대한 최적의 예측 방법을 제공하므로 동작 예측 분야에서 가장 널리 알려진 기법 중 하나로 이전 프레임에서 얻은 영상의 위치와 정보를 분석하여 현재 프레임에서의 위치를 결정하고, 다음 프레임에 움직일 위치를 예측하는 방식이다. 필터는 프로세스 상태를 어떤 시간에 추정하고, 노이즈가 있는 측정으로부터 피드백을 얻는다. (그림 6)은 칼만 필터계산 과정으로 예측단계에서의 식(5)와 식(6)을 각각 이용하여 상태 값과 에러 공분산을 예측하며, 갱신단계에서 구한 칼만게인과 실제 측정값을 이용하여 상태 값을 갱신하고, 에러 공분산을 수정한다. XP는 손의 무게중심이며, R은 Measurement 잡음 벡터, H는 측정값과 상태값을 연결시켜주는 행렬이며, Bu는 선택적인 사항인 전이행렬, Q는 process 잡음 벡터, A는 모델의 전이행렬, P는 에러공분산이다. (그림 7)은 손의 움직임의 예측결과 영상이며 빨강색원은 손의 중심이며, 초록색 원은 칼만으로 이동방향을 예측한 손의 중심이다. 이동방향에 따라 벡터를 산출하고 오른쪽으로 움직인다면, 예측되는 지점에 더욱 weight를 주어서 추적하도록 한다.

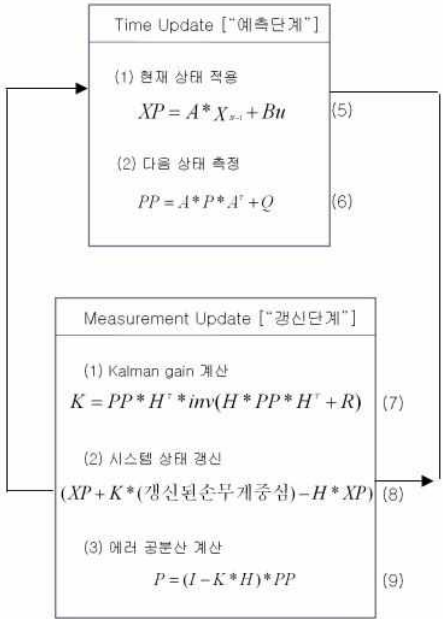


그림 6. 칼만필터 계산과정
Fig. 6. Kalman-Filter Processing

그림 7. 손 추적결과 영상
Fig. 7. Hand Tracking Result Image

2. 손 영역 인식

인식단계에서는 HMM(Hidden Markov Model)를 이용한다[4]. 본 논문에서는 먼저 각 state의 관측벡터에 대한 모델 파라미터의 초기 값을 결정하고, 다음 학습화 단계에서는 Baum-Welch 알고리즘[5]을 사용하여 데이터베이스의 각 개인의 손이미지에 대한 학습화 모델을 생성하며 인식단계에서 Forward 알고리즘[5]을 이용하여 주어진 손이미지의 학습화 모델에서 관측확률이 최대인 것을 손 동작인식을 한다. 즉 손 영역 이미지로부터 관측벡터를 추출한 후에 HMM 모델의 관측시퀀스 확률모델을 계산, 가장 높은 Likelihood 확률을 갖는 모델을 선택하고 이 모델이 인식된 손이 된다. HMM의 변수계산을 보면 먼저 측정되는 관측(Observation)에 대한 심볼(symbol)들을 정의, bj(k)의 계산 수행 후 이를 이용하여 나머지 변수 a_{ij}와 π_i를 구해야 한다. 이를 구하기 위해서 주어진 관측 시퀀스를 이용 HMM의 확률을 최대화 시키는 반복적인 업데이트 방법을 이용해야 한다[6].

$\lambda = (A, B, \pi)$

- A : state transition probability
  $\{ a_{ij} \mid a_{ij} = p(q_{t+1}=j \mid q_t=i) \}$
- B : symbol output/observation probability
  $\{ b_j(v) \mid b_j(v) = p(x=v \mid q_t=j) \}$
- π : initial state distribution probability
  $\{ \pi_i \mid \pi_i = p(q_1=i) \}$

IV. 결론

본 논문의 실험을 위하여 구성된 컴퓨터는 Intel Pentium 듀얼 코어 2.2Ghz CPU 와, 2GByte 의 메모리를 탑재한 Desktop PC 를 사용하였고, OS는 Microsoft의 Windows XP professional을 사용하였으며, Visual Studio 2008 사용하여 시스템을 제작하였다. 본 논문에서는 영상의 크기를 편의를 위해 320 x 240의 크기로 하였다.

아래의 (그림 8)은 0-5까지의 손의 데이터를 학습 후에 인식한 실험결과영상이다.

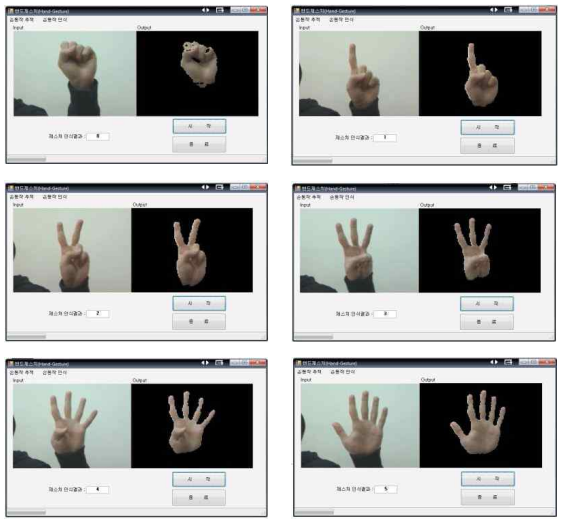


그림 8. 손 영역의 무게중심 추출영상
Fig. 8.Hand region centroid extract image

실험을 위해 연구실에 있는 10명의 학생을 대상으로 6가지 손 모양을 각 10번씩 총 100번씩 실험을 하여 제안한 손 모양 방법을 실험하였다.

다음 <표 1>은 손 모양의 인식결과와 인식률을 나타낸다. 표의 결과와 같이 평균 94%이상의 인식률을 보였다.

〈표 1〉 손 모양 인식률

	인식결과	인식률(%)
0	90	90
1	95	95
2	97	97
3	96	96
4	95	95
5	94	94
평균	94.5	94.5

본 논문에서 제안한 방법을 통하여 실시간으로 사용자의 손 모양 추적 및 인식을 보였다. 본 실험에서는 단순한 배경에서 실험을 하여 높은 인식률을 보였으나 향후실험에는 더욱 복잡한 배경에서도 손 모양을 인식할수 있는 연구가 필요하며, 손 영역을 더욱 빠르고 안정적으로 추출하기위한 연구와, 제스처의 다양한 특성을 고려한 벡터의 추출, 학습 및 인식에서 제스처와의 상관관계의 고려에 의해 이루어 질 수 있는 알고리즘 연구가 수행되어야한다.

참고문헌

[1] D. Chai and K.N. Ngan, “Face segmentation using skin-color map in videophone application.” IEEE Trans. Circuits System Video Technol. 9(4) (1999) 551-564.

[2] Greg Welch and Gary Bishop, “An Introduction to the Kalman Filter”

[3] Kalman, R. E. 1960. “A New Approach to Linear Filtering and Prediction Problems,” Transaction of the ASME—Journal of Basic Engineering, pp. 35-45 (March 1960)

[4] Rabiner, L. R., and Reisman, S., 2000, “A Study of Factors Affecting Cardiovascular Reactivity,” Proceedings of the IEEE 26th Annual Northeast, pp. 53~54.

[5] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proceeding of the IEEE, vol.77, no. 2, pp. 257~286, February 1989.

[6] 하정요, 이민호, 최형일, “HMM(Hidden Markov Model)을 이용한 핸드 제스처 인식,” Journal of Digital Contents Society, vol.10, no. 2, pp. 291~298, Jun. 2009.