웹캠 이미지에서 손동작 인식 알고리즘

최철승*, 오갑석**
*동명대학교 정보통신공학과
**동명대학교 전자공학과
e-mail:ccsccs1004@nate.com

Gesture Recognition Algorithmfrom the Webcam Images

Chul-Seung Choi*, Kab-Suk Oh**
*Information & Communications Engineering, Tomyoung University
**Department of Electronics Engineering, Tomyoung University

요 약

최근 무선통신 기술의 발당과 인터넷의 발달로 웹캠을 활용하여 얼굴인식, 몸동작 인식, 제스처 인식 등의 연구가 활발히 진행되고 있다. 본 연구에서는 웹캠으로부터 입력되는 손동작 동영상을 처리하여 수화 패턴 신호를 인식하는 손동작 인식 알고리즘을 제안한다. 제안방법은 프레임간의 차 연산을 이용하여 움직이는 오브젝트를 추출하고, YCbCr로 변환하여 손 모양 영역을 추출하여 특정 크기의 정규화 영상으로 변화하며, 정규화 영상의 오브젝트의 무게중심점을 기준으로 원의 반지름을 결정하고 원을 탐색하여 손가락의 펴짐과 굽힘에 대한 패턴 비교를 통하여 손 모양을 인식한다.

1. 서론

현대 사회에는 선천적 혹은 후천적으로 농아인으로 살아가는 사람이 많아졌으며, 이들의 사회활동 참여도 활발해 지고 있다. 농아인의 대화 방법은 수화를 비롯한 지화, 표정, 제스처 등이 사용된다. 이들이 사용하는 수화는 함축적인 의미를 갖고 있어 일반인들도 쉽게 배울 수 있다.

한편 IT 기술이 발달함과 동시에 생활이 윤택해짐에 따라 한 가정에 다수의 서로 다른 가전제품용 리모컨이 있어 때로는 혼란스럽고, 리모컨을 찾기 못할 경우 수동으로 가동해야하는 불편함을 감수해야 한다. 또한 하체가 불편하여 이동이 불가능할 경우 리모컨이 멀리 떨어져 있으면 가전제품을 가동하기가 불가능한때가 많다. 이러한 불편함을 해소하고자 본 연구에서는 웹캠 이미지에서 손동작 인식 알고리즘을 제안하고 가전제품이 구동되는 유저인터페이스 시스템에 적용하고자한다. 제안 시스템은 웹캠으로부터 영상획득 및 변환, 움직이는 오브젝트 추출, 손 모양 영역 추출, 손동작 구분 등의 알고리즘을 통하여 손동작의 의미를 분석하고, 무선으로 리모컨에 데이터를 전송하여 해당 가전제품을 구동한다. TV ON, TV OFF, 채널변경 등의 시나리오를 통하여 제안 시스템의 유효성을 검증한다.

2. 손동작 인식 알고리즘

손동작 인식 알고리즘은 영상획득 및 변환, 움직이는 오브젝트 추출, 손 모양 영역 추출, 손동작 구분 등으로 구성하였다.

2.1 영상획득 및 변환

시스템 처음 동작 시 영상처리에 사용될 배경 영상이 존재하지 않기 때문에 웹캠으로부터 처음 들어오는 현재 프레임 24비트 컬러영상 $\sum f_{xy}(t)$를 식(1)을 이용하여 8비트 그레이 영상 $\sum g_{xy}(t)$로 변환하고 [1], 이를 배경 영상 $\sum bg_{xy}$로 지정한다. 여기서 x는 현재 프레임의 가로크기, y는 세로크기이며, $0 \le g_{xy}(t) \le 255$ 이다.

$$\sum g_{xy} = (R + G + B)/3 \tag{1}$$

2.2 움직이는 오브젝트 추출 알고리즘

오브젝트의 움직임 검출을 위해 위와 같은 방법으로 다음 프레임 영상 $\sum g_{xy}(t)$를 획득하여 [그림 1]과 같이 이전 프레임 영상 $\sum g_{xy}(t-1)$과의 차영상 $\sum gs_{xy}(t)$를 구한다.

(a) 이전 프레임 (b) 현재 프레임 (c) 차 영상

[그림 1] 움직이는 물체 검출을 위한 차 연산

본 시스템은 일반적인 웹캠을 사용하기 때문에 미세한 빛의 변화에도 컬러 값이 변하게 되므로 표본 실험을 통하여 이진화 임계값을 25로 설정 하였다. 결과 영상 $\sum gs_{xy}(t)$에서 임계값 이상의 값을 갖는 픽셀이 20개를 초과할 경우 움직임이 있다고 판단하고, 20개 미만일 경우 정지되었다고 판단한다. 만약 연속하여 100 프레임동안 움직임이 없을 경우 현재 프레임 영상 $\sum g_{xy}(t)$를 배경 프레임 영상 $\sum bg_{xy}$로 지정 한다. 이는 카메라를 이동 또는 시간 경과에 따라 배경 영상을 개선하기 위함이다.

만약 움직임이 검출된 다음 정지 한 시점의 현재 프레임 영상 $\sum g_{xy}(t)$와 배경 프레임 영상 $\sum bg_{xy}$의 차 영상 $\sum sg_{xy}$를 구하고, 이진화를 통하여 오브젝트 이진 영상 $\sum bsg_{xy}$를 구한다. 배경 영상과 오브젝트 영상의 예를 [그림 2]에 보인다.

(a) 배경 영상 (b) 오브젝트영상 (c) 이진 영상

[그림 2] 배경 영상과 손 영상의 예

2.3 손 모양 영역 추출 알고리즘

배경 영상과 오브젝트 검출 영상으로부터 구해진 오브젝트 이진 영상 $\sum bsg_{xy}$의 픽셀 값이 0인 위치에 해당하는 컬러 영상 $\sum f_{xy}(t)$의 픽셀 값을 0으로 설정하면 [그림 3]의 (a)와 같이 배경이 제거된 컬러 영상 오브젝트가 추출된다. 컬러 영상에서 피부색 영역을 검출하기 위하여 식(2)을 이용하여 YCbCr 영역으로 변환하고[1], 피부색 영역을 추출하는 식(3)을 이용하여 이진화된 손 모양 영역 영상 $\sum bi_{xy}$를 구하면 [그림 3]의 (b)와 같은 손 모양 영역이 추출된다.

$$\begin{aligned} Y &= 0.2999*R+0.587*G+0.114*B \\ Cb &= -0.1687*R-0.3313*G+0.50*B+128.0 \\ Cr &= 0.50*R-0.4187*G-0.0813*B+128.0 \end{aligned} \quad (2)$$

$$bi_{xy} = \begin{cases} 1, & (77 \le Cb \le 127 \text{ and } 133 \le Cr \le 173) \\ 0, & Otherwise \end{cases} \quad (3)$$

(a) 컬러 영상 오브젝트 (b) 손 모양 이진 영상

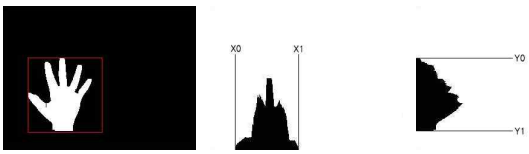
[그림 3] 손 모양 이진 영상 추출

[그림 3]의 (b)와 같이 추출된 손 모양 이진 영상 $\sum bi_{xy}$에 열림연산을 적용하여 미세한 잡음을 제거하고 손 모양 영역에 대해 정규화 한다. 120x120 크기의 손 영역 정규화 방법은 다음과 같다.

Step 1 : [그림 4]의 (a)와 같이 이진화된 손 모양 영상에 X축과 Y축으로 투영한 히스토그램을 [그림 4]의 (b) 및 [그림 4]의 (c)와 같이 구한다.

Step 2 : 시작 좌표 (X0, Y0)를 구한다.

Step 3 : X0에서 X1까지의 길이와 Y0에서 Y1까지의 길이 중 더 긴 것을 가로와 세로 길이로 지정하여 정사각형의 좌표를 구한다.



(a) 손 모양 영상 (b) X축 투영 (c) Y축 투영

[그림 4] X축과 Y축으로 투영한 히스토그램

Step 4 : 손 영역을 모두 포함하는 정사각형 영역을 확대 또는 축소하여 120x120 크기의 정규화 영상 $\sum bn_{xy}$를 [그림 5]와 같이 구한다.

2.4 손동작 구분 알고리즘

본 연구에서는 손가락의 펴짐 여부를 확인할 수 있는 한 가지 방법으로 손 모양 영역의 무게중심점을 기준으로 하는 원을 고려한다. 무게중심점을 기준으로 다양한 크기의 원을 그릴 수 있으며, 적절한 반지름을 정하고 그 둘레를 탐색하면 손가락이 펴져 있거나 굽혀져 있음을 판단할 수 있다. 정규화 영상 $\sum bn_{xy}$으로부터 무게중심을 구하는 방법은 식(4)와 같다.

$$x_0 = \frac{\sum bn_{xy} * x}{\sum bn_{xy}}, \quad y_0 = \frac{\sum bn_{xy} * y}{\sum bn_{xy}} \qquad (4)$$

여기서 $\sum bn_{xy}$는 1 또는 0의 값을 갖고, $0 \le x, y \le 119$ 이며, x_0은 x축 좌표, y_0은 y축 좌표이다.

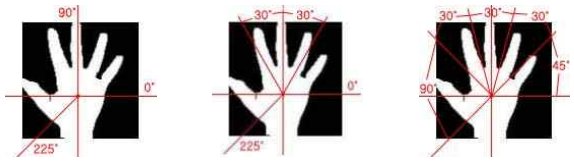
식(3)의 방법으로 구해진 무게중심점을 손 모양 영역에 표시하면 [그림 5]와 같다.

[그림 5] 손 모양 영역의 무게중심

무게중심점에서 모든 손가락을 구분할 수 있는 원의 반지름은 다음과 같이 결정 하였다. 정규화된 120x120 사이즈의 영상이기 때문에 최대 반지름(r1)은 60이 된다. 정규화된 손 모양 영상에서 손가락이 모두 침식되어 없어지고 손바닥 영역만 남을 때 까지 침식연산을 수행한다. 다음으로 침식된 손바닥 영역의 X축 좌표의 최대값과 최소값의 차(r2)를 구하고, 식(5)를 이용하여 반지름 r을 구한다.

$$r = \frac{r1 + r2}{2} \qquad (5)$$

일반적으로 중지를 천정과 수직이 되도록 하였을 경우 손가락은 [그림 6]의 (a)와 같이 무게중심점을 기준으로 0°~225° 사이에 존재한다. 이를 자세히 보면 90° 부근에 중지가 존재하고, 0°~90° 사이에 소지와 약지가 존재하며, 90°~225° 사이에 검지와 엄지가 존재한다. 그리고 [그림 6]의 (b)와 같이 검지와 약지의 위치가 중지로부터 약 30°정도 떨어져 있는 것을 볼 수 있다. 이를 이용하여 [그림 6]의 (c)와 같이 90°를 기준으로 왼쪽 15°, 오른쪽 15° 위치에 중지가 존재하고, 이를 중심으로 왼쪽 30°에는 검지, 오른쪽 30°에는 약지, 나머지 위치에는 엄지와 소지가 존재하는 것을 알 수 있다. 이러한 원리를 이용하여 손가락 탐색영역을 설정하고 해당 영역을 탐색하면 특정 손가락의 펴짐 여부를 판단할 수 있다. [표 1]에 손가락 위치 각도를 나타내었다.



(a) 손가락 위치 (b) 손가락 간격 (c) 손가락 각도

[그림 6] 손가락의 각도

[표 1] 손가락의 각도

손가락	손가락 위치 각도(°)
엄지	136~225
검지	106~135
중지	76~105
약지	46~75
소지	0~45

원을 따라 0°~225° 사이를 일주하며, 식(6)을 이용하여 해당 손가락 각도에 1의 값이 5개 이상 검출될 경우 해당 손가락이 펴져 있다고 판단하고, 그렇지 않을 경우 손가락이 펴지지 않은 것으로 판단한다.

$$\begin{aligned} x &= |cx + (r * \cos(\theta))| \\ y &= |cy + (r * \sin(\theta))| \end{aligned} \qquad (6)$$

여기서 cx와 cy는 무게중심의 좌표이고, r은 원의 반지름, θ는 각도를 나타낸다.

판단한 내용을 바탕으로 엄지손가락부터 차례대로 5자리를 부여하여 모두 펴졌을 경우 11111, 모두 펴지지 않았을 경우 00000과 같은 패턴을 구한다. 여기서 패턴의 각 자리들은 손가락의 위치를 나타내며 펴졌으면 1, 펴지지 않았으면 0을 나타낸다. 이 손가락 위치에 대한 패턴들의 의미는 [표 2]에 나타내었다.

3. 실험 및 고찰

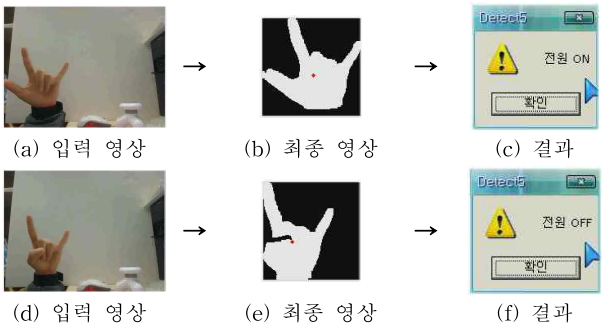
실험에 사용된 시스템은 Intel Core2Duo E7400 2.8Ghz 프로세서, 2GB RAM Windows XP SP3 등으로 구성된 PC에 영상처리를 위해 Microsoft사의 Visual C++ 6.0과 OpenCV 1.1pre1a를 채택 하였고, 웹캠은 Logitech사의 QuickCam Pro 9000 모델로 Carl Zeiss® 렌즈 사용, 자동초점 시스템, 동영상 캡쳐 최대 200만 화소, 정지 영상 캡쳐 최대 800만 화소, 초당 최대 30프레임의 속도 및 24비트 트루 컬러의 색상 심도를 갖는다. 실험용 영상은 웹캠을 이용하여 동영상으로 입력 받으며, 크기가 320x240

pixel 이고 24비트 트루 컬러인 영상을 초당 최대 30프레임의 속도로 입력 받는다. 제안한 알고리즘의 성능 평가를 위해 실험대상 인원 10명에 대해 제약 조건인 긴팔을 착용하게 하고, 손바닥을 정면으로 보이게 하여 똑바로 세우고 실험 하였다.

[표 2] 손가락의 패턴과 의미

입력영상	손 모양 영역 영상	패턴	의미
		00000	숫자 0
		01000	숫자 1
		01100	숫자 2
		01110	숫자 3
		01111	숫자 4
		10000	숫자 5
		11000	숫자 6
		11100	숫자 7
		11110	숫자 8
		11111	숫자 9
		11001	전원 ON
		01001	전원 OFF

전원 ON의 경우 패턴 인식 시 바로 동작하고, 숫자 패턴의 경우 숫자 패턴을 인식 후 전원 OFF 패턴을 인식하면 최종적으로 채널을 변경한다. 전원 OFF의 경우 숫자 패턴 인식 없이 전원 OFF패턴을 인식 하면 동작한다. 실험 시나리오는 ① 전원 ON → 전원 OFF, ② 전원 ON → 채널1 변경 → 채널5 변경 → 전원 OFF이며, [그림 7]에 실험 시나리오 일부를 나타내었다. 실험실 내에서의 실험 시나리오는 안정적으로 동작하는 것을 확인할 수 있었다.



(a) 입력 영상 (b) 최종 영상 (c) 결과

(d) 입력 영상 (e) 최종 영상 (f) 결과

[그림 7] 실험 시나리오

4. 결론

본 논문에서는 수화를 기반으로 한 간단한 손동작으로 리모컨을 제어하기 위하여 웹캠으로부터 영상 획득 및 변환, 움직이는 오브젝트 추출, 손 모양 영역 추출, 손동작 구분 등의 알고리즘을 통하여 손동작의 의미를 부여하고, 무선으로 리모컨에 데이터를 전송하여 해당 가전제품을 구동하는 방법을 제안하였다. 12개의 패턴에 적용한 실험에서 전원, 채널변경 신호 전달이 가능함을 확인하였다. 리모컨 제작 및 시스템적용은 향후 과제로 남긴다.

참고문헌

[1] 정성환, 이문호, 「오픈소스 CxImage를 이용한 Visual C++ 디지털영상처리」, 홍릉과학출판사 2006.

[2] Schettini, R., & Gasparini, F. Skin segmentation using multiple thresholding. In Internet Imaging VII, IS&T/SPIE (pp.60610F-1-60610F-8). SPIE. 2006

[3] 이원주, 「웹캠을 이용한 동적 제스쳐 인식 기반의 감성 메신저 구현 및 성능 분석」, 한국컴퓨터정보학회 논문지, 제15권, 제7호, pp.75~81, 2010.