

특징 벡터 보정 기반의 헤드 제스처 인식

전인자^U 최현일 이필규
인하대학교 전자계산공학과
g1991252@inhavision.inha.ac.kr

Head Gesture Recognition Technique based on Mean Acceleration Measure(MAM)

In-Ja Jun^U Hyun-Il Choi Phill-Kyu Rhee
Dept. of Computer Science & Engineering, Inha University

요 약

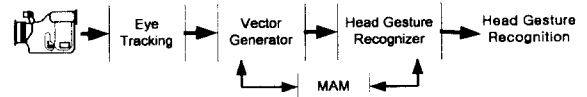
본 논문에서는 MAM을 이용한 특징 벡터의 보정을 기반으로 하는 헤드 제스처 인식에 관해 기술한다. 제안된 시스템은 얼굴 움직임 검출 모듈과 눈 영역 추적 모듈, 미 측정된 벡터 보정 모듈, 측정된 제스처에 대한 인식모듈로 구성된다. 신경망과 모자이크 이미지를 이용하여 얼굴 영역을 검출하고, 이 영역에서 눈 영역을 검출한다. 만약 눈의 쌍이 검출되지 않는다면 시스템은 특징 벡터 보정(MAM)을 수행하여 손실된 정보를 예측한다. 검출된 눈 영역은 정규화된 벡터로 변경된다. 이 벡터의 분산을 이용하여 긍정, 부정, 중립의 제스처를 판단한다. 제스처의 인식은 직접 관측, 이중 HMM, 삼중 HMM을 사용한 다중 인식기를 이용한다.

1. 서 론

제스처는 인간의 의도를 손, 몸, 얼굴 등을 이용하여 표현하는 의사표현 방법이다. 손이나 헤드 제스처를 컴퓨터와 인간 사이의 인터페이스로 사용하려는 연구가 계속 이루어지고 있다[1]. 본 논문에서는 MAM을 이용한 특징 벡터 보정을 기반으로 하는 은닉 마르코프 모델(Hidden Markov Models:HMM)을 이용한 다중 인식기를 통해 사람의 연속되는 머리 움직임 정보에 대하여 대한 긍정, 부정, 중립의 제스처를 판단한다. 실시간 환경에서 Bala, Talmi, Liul[2]는 얼굴과 눈을 검출하고 추적하는 알고리즘을 제안하였으며 속도 향상을 위해 피부색을 사용하였다. Morimoto 와 Yacoub[3]은 은닉 마르코프 모델을 사용하여 헤드 제스처를 인식하는 방법을 제안하였다. Cascia, Isidoro, Sclaroff[4]는 3차원 머리 움직임 추적을 위한 방법을 제안하였으며, 3차원 표면 모델의 텍스처 맵에서의 영상 정합의 방법으로 추적을 수행하였다.

얼굴영상에서 눈 영역의 측정이 불가능할 경우 현재 입력된 영상에서 찾고자 하는 특징벡터의 값은 예측할 수 없다는 문제점이 발생한다. 본 논문에서는 이러한 문제점을 해결하기 위하여 영상의 위치를 예측(Prediction)하여 벡터의 위치를 예측할 수 있는 특징 벡터 보정 기반의 다중 인식기를 이용한 헤드 제스처 인식 시스템에 대해 서술한다. 제안된 시스템은 그림1과 같이 구성된다.

본 논문의 구성은 다음과 같다. 2 장에서는 제안된 눈 영역 추적 방법 및 MAM을 이용하여 미 측정된 눈 영역의 예측에 관하여 기술하며, 3 장에서는 본 논문에서 사용된 헤드 제스처 인식에 대해 논한다. 4 장에서는 실험 결과를, 마지막으로 5 장에서는 결론을 내린다.



[그림 1. 제안된 시스템 구조]

2. 눈 영역 추적

얼굴 영역 검출은 모자이크(mosaic)와 역-전달 신경망(back propagation neural network)[5]을 사용한다. 비디오 카메라에서 입력되는 영상은 잡음이 많기 때문에 히스토그램 정규화[6], 모자이크[7]를 사용함으로써 잡음이 제거된 영상을 획득할 수 있다.

영역 추적 알고리즘을 통해 추정된 얼굴 영역에서 눈의 쌍을 추적한다. 눈 추적이 수행된 후 검출된 눈 영역은 x 축과 y 축의 정규화 된 벡터로 변경된다. 눈 영역 검출 시간을 단축하기 위해 눈의 쌍에 가까운 영역에서 수행된다. 만약 눈의 쌍이 검출되지 않으면 특징 벡터 보정(MAM) 기법을 이용하여 유실된 벡터를 예측한다.

2.1 얼굴 영역 검출

얼굴 영역이 성공적으로 검출되면 영상 분할, 라벨링, 이진화 [8]등의 과정을 수행하여 불필요한 정보를 제거한후 눈 영역의 후보를 추출하게 된다.

2.2 눈 영역 검출

눈 영역의 정보를 통해 눈의 쌍이 검출되며, 왼쪽 눈과 오른쪽 눈의 거리정보는 중요한 기하학적 요소로 사용된다. 눈의 기하학적 정보에 기반하여 눈을 검출하며, 검출된 눈 영역 중 가장 결과가 좋은 영역을 선택한다.

2.3 특징 벡터 보정 : MAM(Mean Acceleration Measure)

특징벡터의 추출에 실패했을 경우 헤드 제스처 변화의 연속성을 보장하기 위해 이전 프레임을 참조한다. 제안된 방법은 MAM(Mean Acceleration Measure) 을 이용하여 시간 t 에서 특징 벡터 \mathbf{p}_t 를 추정한다. t 부터 $t+k$, $k=1, 2, \dots, n$ 까지의 n 는 윈도우 스케일 요소 k 와 속도 \mathbf{d} 를 포함하며 이는 식 (1)와 같이 나타내어 진다.

$$a_{(t,t+k)} = \frac{1}{t+k} \sum_{i=t}^k \mathbf{d}_i \quad \text{식 (1)}$$

여기서 \mathbf{d} 는 다음과 같이 정의된다.

$$\mathbf{d}_t = \mathbf{p}_t - \mathbf{p}_{t-1} \quad \text{식 (2)}$$

a 를 통해 손실된 특징벡터 \mathbf{p}_t 를 식 (3)와 같이 추측한다.

$$\mathbf{p}_t = \mathbf{p}_{t-1} + a_{(0,t)} \quad \text{식 (3)}$$

헤드 제스처의 특징 벡터열은 제스처가 긍정일 때 위, 아래로, 부정일 때 좌, 우로 움직이기 때문에 사인 형태의 패턴을 보인다. 만약 제안된 MAM 방법이 이러한 패턴에 적용된다면 추측된 \mathbf{p}_t 는 정확한 값을 가지지 않게 된다. 이러한 문제를 해결하기 위해 특징 벡터열의 변곡점 \mathbf{p}_{\max} 가 필요하다. 변곡점 \mathbf{p}_{\max} 를 기준으로 좌측 부분의 가속도와 우측 부분의 속도는 서로 다르기 때문에 다시 계산되어야 할 필요가 있다. 미측정된 프레임의 수에 따라 아래와 같이 \mathbf{p}_{\max} 가 계산되어 진다.

$$\mathbf{p}_{\max} = \begin{cases} \text{if } a_{(t,t+k)} < 0 \text{ then } \mathbf{p}_{\max} \\ \text{if } a_{(t,t+k)} \geq 0 \text{ then continue} \end{cases} \quad \text{식 (4)}$$

실험결과 k 의 값은 평균적으로 $k < 3$ 이었고, 얼굴 영역 추출을 재수행 하지 않기 때문에 계산 속도의 향상 효과를 얻을 수 있었다.

3. 헤드 제스처 인식

눈 영역 추적이 수행된 후 눈 영역의 좌표는 헤드 제스처 인식에 적합한 벡터로 변경된다.

3.1 헤드 제스처 패턴

긍정은 머리를 끄덕이는 것이며 부정은 고개를 좌우로 흔드는 것, 중립은 긍정, 부정을 제외한 나머지로 정의된다.

3.2. 헤드 제스처 인식 기법

3.2.1 직접 관측에 의한 전처리

직접관측에 의한 인식에서는 정규화된 x 축 벡터나 y 축 벡터를 이용하여 긍정과 부정의 제스처를 개략적으로 판단하는데 사용된다. y 축과 x 축의 평균을 비교함으로써 어떤 좌표축의 평균이 높은지 판단할 수 있다. 높은 평균을 가진 축은 사인 형태의 패턴을 가지고 있다. 어떤 패턴이 사인 형태의 패턴인지 관측함으로써 긍정과 부정을 쉽게 알 수 있다.

3.2.2 이중 HMM을 이용한 인식

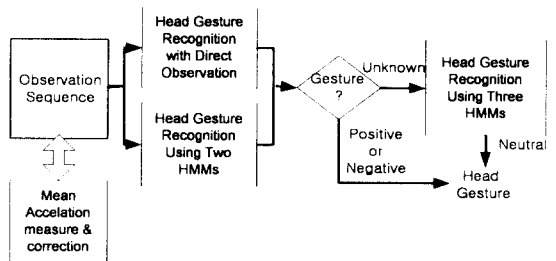
HMM으로의 입력은 정규화된 벡터이며 이는 Baum-Welch[11] 절차에 의해 학습되어 진다. 첫번째 HMM 는 불규칙한 선형 패턴을 인식하며 두 번째 HMM 는 사인 형태의 패턴을 인식한다. x 축에서 사인 형태의 패턴이 존재하고 y 축에서 불규칙한 선형 패턴을 가지고 있다면 시스템은 사용자가 부정의 제스처를 취했다고 인식한다.

3.2.3 삼중 HMM을 이용한 인식

직관적인 방법과 이중 HMM방법들은 중립 제스처를 인식하기가 매우 어렵다. 중립상태를 정확히 인식할 수 있는 다른 방법으로 양쪽 눈 영역의 중점의 벡터를 사용하여 설계되었다. 얼굴 움직임은 8 방향을 가진 벡터 코드로 변경된다. 비디오 영상으로부터 추출된 8 방향 코드는 벡터 코드로 변경되며, 이 벡터 코드는 사용자의 제스처를 인식하기 위한 HMM 으로 입력된다.

본 논문에서는 헤드 제스처를 인식하기 위해 4 개의 상태와 9 개의 심볼로 HMM 를 구성하였다.

3.3 MAM을 기반으로 한 다중 인식기를 이용한 헤드 제스처 인식기의 구성



[그림 2. MAM을 기반으로 한 다중시스템의 구성도]

앞에서 서술된 직접관측, 이중 HMM, 삼중 HMM의 방법을 결합하여 그림 2에서는 다중 인식기 시스템의 구조를 보이고 있다. 직접 관측에 의한 전처리, 이중 HMM, 삼중 HMM을 이용한 인식 등의 방법은 독립적으로 동작되며, 처음 두 가지 방법으로는 중립 제스처를 인식하기 어렵기 때문에 세 번째 방법을 이용하여 중립 제스처를 학습시키고 이에 대한 결과가 출력된다. 만약 처음 두 방법의 결과가 동일하다면 인식 결과가 출력 되고, 그렇지 않으면 세 번째 방법을 통해 헤드 제스처를 인식한다

4. 실험 결과

실험은 비디오 카메라로부터 획득된 연속적인 영상에서 수행되었다. 제안된 알고리즘은 펜티엄 400 Mhz 를 탑재한 PC 에서 Windows 98 을 기반으로 하여 visual C++ 6.0 을 사용하여 구현되었다. 영상의 크기는 320 x 240 이며 256 단계의 Gray 영상이다.

눈 영역 검출은 26 명으로부터 획득된 410 개의 영상으로 수행되었다. 실험 결과에서 안경과 조명은 고려되었으며 안경과 조명에 따라 분류되었다. HMM 을 사용한 헤드 제스처 인식을 실현하기 위해 사용된 학습 데이터로서 6 명을 사용하였으며 각각은 긍정, 부정, 중립 제스처를 학습시키기 위해 사용된다. 총 90 개의 데이터를 사용하여 제안된 시스템을 테스트하였다. 표 1 에서는 MAM 을 사용하였을때의 상태를 실험한 세가지 방법의 결과를 보이고 있다. 표 2 에서는 결합된 시스템의 결과를 보이고 있다.

5. 결론

본 논문에서는 영상으로부터 얼굴 영역 검출, 검출된 얼굴 영역을 이용하여 눈 영역 검출, 검출된 눈 영역을 이용하여 특징벡터를 생성한다. 이 추출된 특징벡터가 있다면 특징벡터를 추정하는 MAM 모듈을 수행하여 유실된 벡터를 추정하여 값을 보정한다. 생성된 특징벡터를 직접관측, 이중 HMM, 삼중 HMM 을 이용한 다중 인식기를 사용하여 이에 대한 인식을 수행하는 헤드 제스처 인식 시스템에 대하여 서술하였다. 입력 영상 중 유실된 특징벡터는 MAM 추정을 이용하여 눈 영역의 특징벡터의 값을 비교적 정확하게 유추할 수 있었다.

제안된 인식 시스템은 시각 사용자 인터페이스에 효과적인 적용이 가능 할 것이다. 제안된 시스템에서 발생한 문제점은 사용자가 얼굴을 왼쪽이나 오른쪽으로 움직이면 이는 부정으로 인식하기도 하고, 사용자가 얼굴을 위나 아래로 움직이면 긍정으로 인식하는 문제점이 있다. 이러한 문제를 해결하기 위해 얼굴의 특성과 얼굴의 시선등을 고려한 연구가 진행되고 있다.

참고 문헌

[1] P.Ekman, T.Huang, T.Sejnowski, and J. Hager. "Final Report to NSF of the planning Workshop on Facial Expression Understanding", Technical report, National Science Foundation, Human Interaction Lab., UCSF, CA 94143, 1993.

[2] Lars-Peter Bala, Kay Talmi, and Jin Liu. "Automatic Detection and Tracking of Faces and Facial Features in Video Sequences". Picture Coding Symposium 97, pp.251-256, September 1997

[3] Carlos Morimoto, Yaser Yacoob, and Larry Davis. "Recognition of head gestures using hidden markov models", IEEE Proceedings of ICPR 96, pp.461-465, 1996.

[표 1. MAM 추정을 사용한 헤드 제스처 인식 결과]

Data		Positive	Negative	Neutral	Total
Head Gesture Recognition with direct observation	The number of Data	30	30	30	90
	Success	28	28	22	78
	Recognition Rate(%)	93.3%	93.3%	73.3%	86.7%
Head Gesture Recognition using Two HMMs	The number of Data	30	30	30	90
	Success	28	28	21	77
	Recognition Rate(%)	93.3%	93.3%	70.0%	85.6%
Head Gesture Recognition using Three HMMs	The number of Data	30	30	30	90
	Success	27	27	25	79
	Recognition Rate(%)	90.0%	90.0%	83.3%	87.8%

[표 2. 다중 인식기를 이용한 헤드 제스처 인식 결과]

Data	Positive	Negative	Neutral	Total
The number of data	30	30	30	90
success	29	28	23	80
Recognition rate(%)	96.7%	93.3%	76.7%	88.9%

[4] Marco La Cascia, John Isidoro, and Stan Sclaroff. "Head Tracking via Robust Registration in Texture Map Images", IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Santa Barbara, California, pp.508-514, June, 1998.

[5] Jung-II Choi, Che-Woo La, and Phil-Kyu Rhee. "Face and Eye Location Algorithms For Visual User Interface", IEEE Signal Processing Workshop on Multimedia Signal Processing, pp.239-244, Jun 23-25, 1997, Princeton, NJ.

[6] Maher A. Sid-Ahmed. Image Processing : Theory, Algorithms, and International Journal of Computer Vision, pp. 99-111, 1992.

[7] Guanzheng Yang and Thomas S. Huang. Human Face Detection In A Complex Background. Pattern Recognition, Vol. 27, No. 1, pp.53-63, 1994.

[8] I. Pitas. Digital image processing algorithms, Prentice Hall, 1993.

[9] R. L. Lumina, G. Shapiro, and O. Zuniga. "A New Connected Components Algorithm for Virtual Memory Computers." Computer Vision, Graphics, and Image Processing, Vol. 22, pp. 287-300, 1983.

[10] L. Rabiner. A tutorial on hidden markov models and Selected applications in speech recognition, Proceedings of the IEEE, 77(2):257-285, February 1989.

[11] L. Baum, T. Petrie, S. G., and N. Weiss. "A maximization techniques occurring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat., 41(1):164-171, 1970.