2000年度 大韓電子工學會 夏季綜合學術大會 論文集 第 23 卷 第1號 2000/6

연속 영상에서 학습 효과를 이용한 제스처 인식

이 현 주(李 泫 周), 이 칠 우(李 七 雨)
전남대학교 컴퓨터공학과
전화 : (062) 530-0258 / 팩스 : (062) 530-1809

Gesture Recognition using Training-effect on image sequences

Hyun Ju Lee, Chil Woo Lee
School of Computer Engineering Chonnam National University
E-mail : leehj98@daum.net, leecw@chonnam.ac.kr

Abstract

Human frequently communicate non-linguistic information with gesture. So, we must develop efficient and fast gesture recognition algorithms for more natural human-computer interaction. However, it is difficult to recognize gesture automatically because human's body is three dimensional object with very complex structure. In this paper, we suggest a method which is able to detect key frames and frame changes, and to classify image sequence into some gesture groups. Gesture is classifiable according to moving part of body. First, we detect some frames that motion areas are changed abruptly and save those frames as key frames, and then use the frames to classify sequences. We symbolize each image of classified sequence using Principal Component Analysis(PCA) and clustering algorithm since it is better to use fewer components for representation of gestures. Symbols are used as the input symbols for the Hidden Markov Model(HMM) and recognized as a gesture with probability calculation.

I. 서론

컴퓨터 기술의 발달과 함께 정보 시스템이 복잡하게 되면서 인간과 정보 시스템 사이에 자연스럽게 정보를 교환할 수 있는 지적 시스템에 관한 관심이 날로 커지고 있다. 인간은 일상 생활에서 제스처, 표정과 같은 비언어적인 수단을 이용하여 수많은 정보를 전달한다. 따라서 자연스럽고 지적인 인터페이스를 구축하기 위해서는 제스처와 같은 비언어적인 통신 수단에 대한 연구가 매우 중요하다. 최근에 들어, 대규모 비디오 데이터베이스의 구축, 감시 시스템, 고압축 통신 시스템의 구축을 위해 제스처 인식에 관한 연구가 활발히 진행되고 있다.

제스처를 인식한다는 것은 인체 각 부위가 시간축에 대해 어떠한 형상 변화를 가지는가를 자동으로 알아내는 것을 의미한다. 그러나 인체는 매우 복잡한 3차원 관절 구조를 지니고 있어서 자동으로 제스처를 인식하는 것은 매우 어렵다.

본 논문에서는 연속적인 영상 시퀀스를 몇 개의 제스처 그룹으로 분류하기 위해 신체의 움직임이 갑자기 변하는 키 프레임을 자동으로 검출한다. 그리고 각각의 제스처 시퀀스들은 주성분 분석법(Principal Component Analysis)에 의해 심볼(symbol)로 형상화되어진다. 마지막으로 은닉 마르코프 모델(Hidden Markov Model: HMM)에 의해서 인식되어진다. 전체 시스템 구성도는 그림 1에서 보여주는 것과 같다.

II. 영상 시퀀스의 자동 분할

1. 세그멘테이션

전경 영역(foreground region)을 배경으로부터 분리하기 위해서는 먼저 배경 모델을 생성해야 한다. 배경 모델(background model : BM)은 전경 영역을 포함하지 않은 영상 시퀀스로부터 계산되어지는 것으로 식(1)과 같이 표현되어진다.

$$BM = \{ M(x, t), N(x, t), D(x, t) \} \quad (1)$$

여기서 M(x, t)는 화소 x가 시간 t에 의해서 갖는 최소 밝기값, N(x, t)는 화소 x가 시간 t에 의해서 갖는 최대 밝기값을 나타낸다. D(x, t)는 화소 x가 가질 수 있는 최대 밝기 차이값을 나타낸다.

전경 영역은 식 (2)에 의해서 결정되어진다[1]. 즉 식 (2)를 만족하는 화소 x는 모두 전경 영역으로 세그멘테이션된다.

$$| M(x, t) - I(x, t) | > D(x, t) + C \quad \text{or}$$
$$| N(x, t) - I(x, t) | > D(x, t) + C \qquad (2)$$

여기서 I(x, t)는 입력 영상이고 C는 상수값이다.

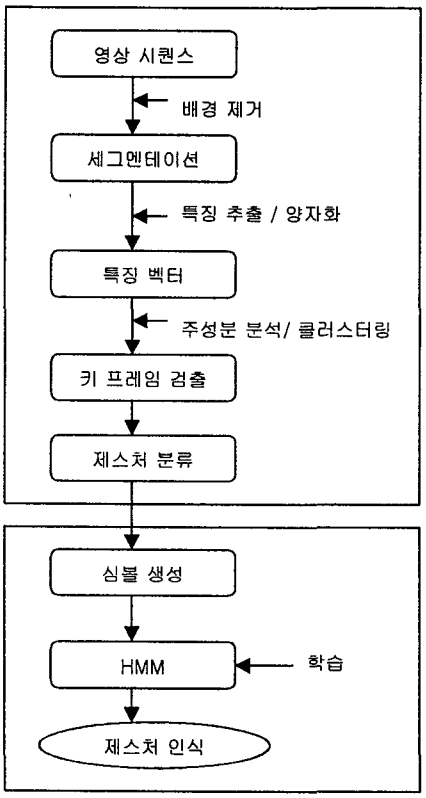


그림 1. 시스템 전체 구성도

2. 특징 추출

제스처들을 분류하기 위해 사용한 특징값은 신체 영역의 가로축 길이(FeretX)와 세로축 길이(FeretY)의 비를 나타내는 페렛비(Feret_ratio), 무게 중심의 x좌표, 무게 중심의 y좌표, 조밀성(Compactness), 모멘트의 주축, 모멘트 주축의 수직인 축으로 총 6가지이다.

이들 특징값은 그림 2에서 보여주는 것처럼 입력 영상이 n개가 들어왔을 때, 식 (3)에 의해 구해진 각 영상에 대해 계산되어진다.

$$I_t, I_t - I_{t+1}, I_t - I_{t+2}, \ldots\ldots I_t - I_{t+n-1} \qquad (3)$$

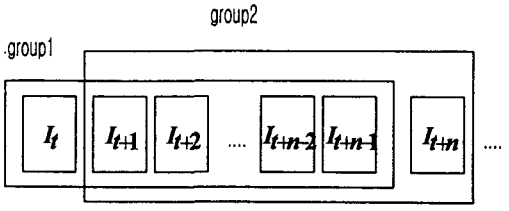


그림 2. 시간에 따른 영상 그룹화

즉, 시간에 따른 움직임의 변화량을 특징값에 의해서 형상화한 것이다. 특징의 집합을 x라 하고 식 (4)과 같이 표현한다.

$$x = [x_1, x_2, x_3, \ldots, x_N]^T \qquad (4)$$

여기서 N은 n개의 영상으로 구성된 group의 수로 전체 영상 시퀀스의 수가 T일 때 N은 T- n + 1의 값과 같다. 그리고 $x_i (i = 1, \ldots, N)$은 n × 6개의 특징 값으로 구성된다. 그러나 특징의 집합을 그대로 사용하게 되면 수치적으로 동일한 단위를 가지고 있지 않기 때문에 양자화 과정을 거쳐야한다.

3. 주성분 분석

2절과 같이 특징 집합을 이용하여 신체의 전체적인 외관 특징들을 표현할 수 있는 저차원 벡터공간, 즉 파라메트릭 고유공간을 생성한다. 고유공간을 계산하기 위해서는 먼저 모든 특징 벡터에서 평균 벡터를 구하여 각 특징들과의 차를 구한다. 평균 벡터 c와 새로운 특징 집합 X를 식 (5)와 식 (6)과 같이 나타낸다[6].

$$c = (1/N) \sum_{i=1}^{N} x_i \qquad (5)$$

$$X \triangleq [x_1 - c, x_2 - c, x_3 - c, \ldots, x_N - c]^T \qquad (6)$$

고유공간을 구하기 위해서는 M × N의 크기를 지닌 특징 집합 X를 식 (6)과 같이 계산하고 식 (7)을 만족하는 고유벡터를 구하면 된다[6]. 즉, 공분산 행렬 Q에 대한 고유치 λ와 고유벡터 e를 구한다.

$$Q \triangleq XX^T \qquad (7)$$
$$\lambda_i e_i = Q e_i \qquad (8)$$

여기서 M은 n개의 영상으로부터 계산된 특징의 수이고 N은 전체 영상의 개수를 나타내는 정수이다.

고유치 분해를 위하여 특이치 분해(singular value decomposition)을 이용한다. 특이치 분해를 이용하면 특징 집합 X의 공분산 행렬에 대한 고유벡터를 쉽게 얻을 수 있다[6]. 이제 얻어진 고유공간에 평균 벡터 c에서 뺀 특징 집합 x를 모두 식 (9)을 이용하여 투영시킨다.

$$m_i = [e_1, e_2, e_3, \ldots, e_k]^T (x_i - c) \quad (9)$$

4. 키 프레임 생성

고유 공간을 계산하기 위해 이용된 특징 벡터 x_1과 x_2를 고유 공간에 투영시켜 얻은 점들이 각각 m_1과 m_2라면 이 두 점 사이의 거리가 가까울수록 비슷한 제스처이다. 따라서 인접하는 특징 벡터 x_1과 x_2가 고유 공간상에서 동떨어져 있다면, x_2는 다른 제스처 시퀀스의 시작을 나타내는 키 프레임으로 분류될 수 있다. 키 프레임에 의해 분류되어진 각 제스처 시퀀스들은 하나의 심볼 시퀀스로 변환되고 은닉 마르코프 모델의 입력으로 사용된다.

III. HMM을 이용한 제스처 인식

1. 은닉 마르코프 모델

은닉 마르코프 과정은 다음의 5개 요소로 정의 가능하다.

S : 상태의 유한 집합 ; $S=\{s_i\}$

Y : 출력 심볼의 집합

A : 상태 천이 확률의 집합 ; $A=\{a_{ij}\}$

a_{ij} : 상태 s_i에서 s_j로 천이할 확률

$\sum_j a_{ij}=1$

B : 출력 확률의 집합 ; $B=\{b_{ij}(k)\}$

b_{ij} (k) : 상태 s_i에서 s_j 로 천이할 때 심볼 k를 출력할 확률

$\sum_k b_{ij}(k)=1$

Π : 초기 상태 확률의 집합 ; $\Pi=\{\pi_i\}$

π_i : 초기 상태가 s_i일 확률

$\sum_j \pi_j=1$

2. 제스처 모델의 학습

은닉 마르코프 모델의 각 제스처 모델(π, A, B)은 Baum-Welch 알고리듬에 의해서 추정되어지고 식(10)-식(11)에 의해서 계산된다[2].

$$\xi_t(i,j) = \frac{P(q_t=i,\ q_{t+1}=j,\ Y|\lambda)}{P(Y|\lambda)} = \frac{a_t(i)\ a_{ij}\ b_j(y_{t+1})\ \beta_{t+1}(j)}{P(Y|\lambda)}$$

$$= \frac{a_t(i)\ a_{ij}\ b_j(y_{t+1})\ \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} a_t(i)\ a_{ij}\ b_j(y_{t+1})\beta_{t+1}(j)} \quad (10)$$

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \quad (11)$$

여기서 $\xi_t(i,j)$ 는 시간 t에서는 상태 i, 시간 t+1에서는 상태 j일 확률이고 $\gamma_t(i)$는 전체 관측 시퀀스와 λ가 주어졌을 때, 시간 t에서 상태 i일 확률을 나타낸다. 식 (10), 식 (11)를 이용하여 제스처 모델은 식 (12), 식 (13), 식 (14)으로 추정되어진다.

$$\overline{\pi_j} = \gamma_1(i) \quad (12)$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (13)$$

$$\overline{b_j(k)} = \frac{\sum_{t=1,\ s.t.\ o_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \quad (14)$$

3. 제스처 인식

관측된 심볼 시퀀스(Y)가 주어지면 모델 λ_i에 대해서, forward - backward 알고리즘을 이용하여 확률을 계산한다. foward - backward 알고리즘은 forward 변수인 $\alpha_t(i)$와 backward 변수인 $\beta_t(i)$를 이용하여 식 (15)와 같이 구한다.

$$P(Y|\lambda_i) = \sum_i \sum_j \alpha_t(i)\ a_{ij}\ b_{ij}(y_{t+1})\beta_{t+1}(j) \quad (15)$$

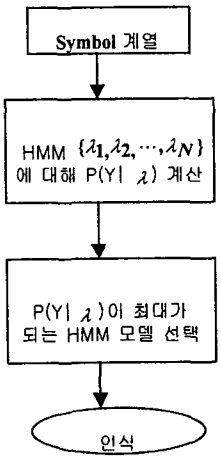
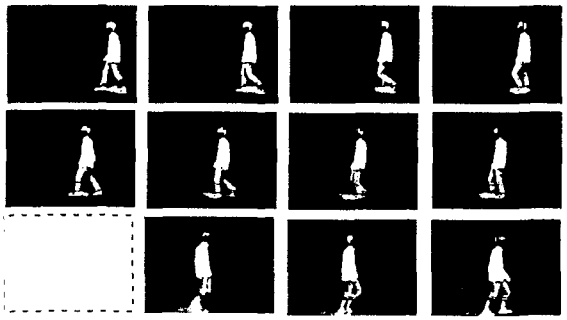


그림 3. HMM의 인식과정

IV. 실험 및 결론

실험에 사용한 영상의 크기는 320 × 240을 사용하였고, 9개의 영상을 그룹화 시켰다. 그림 4는 오른쪽에서 왼쪽으로 걸어가다가 돌아서 반대로 걸어가는 영상 시퀀스를 3개의 고유벡터 즉 3차원 고유공간에 투영한 것이다. 그리고, 그림 4(a)-(e)는 사람이 걸어가다가 방향을 바꾸기 위해 돌아서는 영상에 해당하는 것으로 고유공간상에서 걷는 제스처와 동떨어져 존재함을 확인할 수 있다. 그림 5는 걸어가다가 앉아서 다리운동을 하는 제스처 영상 시퀀스를 그림 4와 같이 고유공간상에 투영시킨 것이다. 그 결과 걷는 제스처 그룹(그림 5(a)-(b)), 서 있다가 앉는 제스처 그룹(그림 5(c)-(d)), 다리 운동을 하는 제스처 그룹(그림 5(e)-(h))으로 분류되어진다. 지금 현재는 단순한 제스처 시퀀스를 사용하고 있지만, 앞으로는 보다 복잡한 영상 시퀀스까지 확장하여 실험할 계획이다.



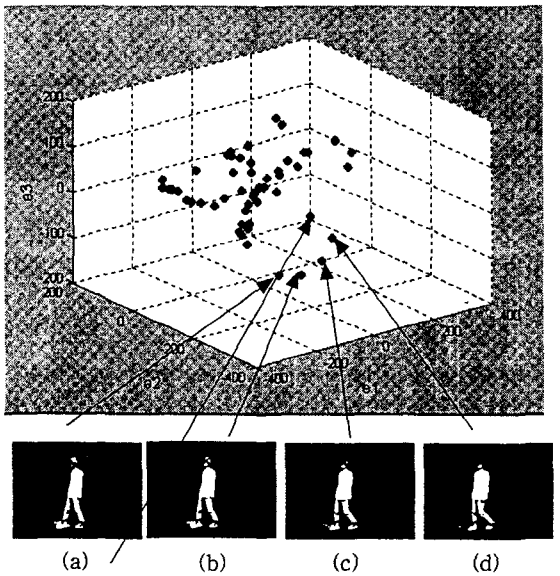


그림 4. 걷다가 돌아서 다시 걷는 제스처

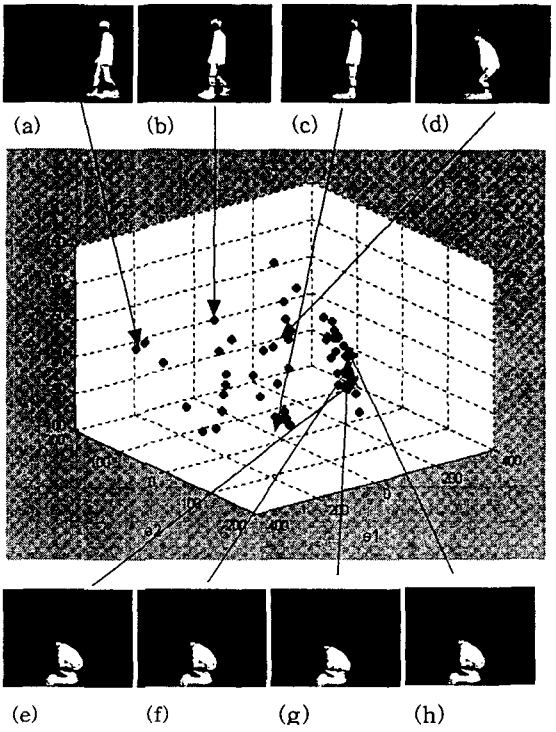


그림 5. 걷기, 앉기, 다리 운동

참고문헌

[1] Ismail Haritaoglu, David Harwood and Larry S. Davis, "W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People", International Conference on Face and Gesture Recognition, 1998, pp. 14-16

[2] Yoshio IWAI, Tadashi HATA, and Masahiko YACHIDA, "Gesture Recognition based on Subspace Method and Hidden Markov Model", IEEE, 1997, pp. 960-966

[3] Ismail Haritaoglu, Ross Cutler, David Harwood and Larry S. Davis, "Backpack: Detection of People Carrying Objects Using Silhouettes", IEEE International Conference on Computer Vision (ICCV), 1999

[4] Takahiro Watanabe and Masahiko Yachida, "Real Time Recognition of Gesture and Gesture Degree Information Using Multi Input Image Sequence", ICPR, 1998

[5] Shigeyoshi Hiratsuka, Kohtaro Ohba, Hikaru Inooka, Shinya Kajikawa, and Kazuo Tanie, "Stable Gesture Verification in Eigen Space", LAPR Workshop on Machine Vision Applicaion, 1998, 17-19

[6] 이용재, 이칠우, "외관 기반의 파라메트릭 고유 공간을 이용한 물체인식", 정보과학회, 1999