韓國 컴퓨터情報學會 論文誌
第9卷 第1號, 2004. 3.

2004-9-1-1-4

HMM을 이용한 지휘 동작의 인식 *

문 형 득 **, 구 자 영 ***

Recognition of Conducting Motion using HMM

Hyung-deuck Moon **, Ja-young Koo ***

요 약

본 논문은 지휘자의 지휘 동작으로부터 일련의 영상들을 추출하여 지휘자가 지휘하는 박자를 인식하는 방법을 제안하고 있다. 색상판별에 의해서 손의 위치를 감지하였으며 양자화를 통해서 그 위치를 기호화함으로써 지휘 동작을 일련의 기호로 표현하였다. 변형을 포함하는 기호열의 인식에 좋은 결과를 보이는 HMM(Hidden Markov Model)을 사용함으로써 표현된 기호열을 지휘박자로 인식하도록 하는 시스템을 구성하였다.

Abstract

In this paper, a beat recognition method from a sequence of images of conducting person was proposed. Hand position was detected using color discrimination, and symbolized by quantization. Then a motion of the conductor was represented as a sequence of symbols. HMM (Hidden Markov Model), which is excellent for recognition of sequence pattern with some level of variation, was used to recognize the sequence of symbols to be a motion for a beat.

▸ Keyword : Motion Recognition, Beat Recognition, Gesture Recognition, Hidden Markov Model(HMM)

* 본 연구는 2003년 단국대학교 교내 연구비에 의해서 연구되었음 .
, * 단국대학교 정보·컴퓨터학부

Ⅰ. 서 론

제스처 언어는 원시시대부터 수렵을 할 때, 혹은 언어가 통하지 않는 부족간의 의사소통을 위해서 사용되었다. 오늘날에도 음성을 이용한 의사소통이 불가능하거나 효율적이지 못한 경우, 동작을 통한 의사전달이 효과적인 대안으로 고려될 수 있다. 예를 들면, 청각 장애자를 위한 수화나, 지휘자와 연주자 사이에서의 동작을 통한 의사소통 등이 제스처 언어의 예가 될 수 있을 것이다. 이중, 손을 이용한 제스처가 가장 표현력이 좋아서 인간-컴퓨터 인터페이스로의 응용도 활발하다[3][6][11].

본 논문에서는 지휘자의 손동작으로부터 지휘하는 박자를 자동적으로 인식하는 방법을 제안하고 있다. 제안된 방법에서는 하나의 지휘 패턴을 양자화하여 기호열로 표현하고, 그 기호열이 미리 학습된 모델의 기호열과 얼마나 유사한지를 판단하는 방법을 사용한다. 변형을 포함하는 기호열의 정합을 위해서는 DTW (Dynamic Time Warping)을 이용하는 방법[4][8][12], 인공신경망(Artificial Neural Network)이론을 이용한 방법[6][10], 그리고 은닉 마르코프 모델(Hidden Markov Model)[1][2][3][5][9][13][14][15]을 이용한 방법 등이 사용된다.

DTW는 동적 프로그래밍을 기반으로 주어진 두 이벤트 사이의 패턴이 가질 수 있는 전역 유사도(global similarity)를 측정하는데 쓰인다. DTW는 미리 정해진 패턴을 바탕으로 하여 각 시간대별로 비슷한 패턴을 갖는 여러 개의 후보를 가질 수 있으며, 훈련 데이터가 적은 환경에서도 사용이 가능하다. 그러나 정해진 패턴과의 정합과정에서 유사도를 계산하여야하기 때문에 공간적인 유사도를 처리하기 위한 추가적인 참조패턴이 필요하며, 이를 정합하기 위한 시간이 필요하므로 분석하는 데 많은 시간이 소요된다. 또한 새로운 패턴을 표현하기 위해서는 새로운 참조패턴이 필요하므로, 이러한 새로운 패턴에 대한 분석 방법을 갖지 못하다는 단점이 있다.

인공신경망 알고리즘은 생물학적 신경망을 수리적으로 모델링하고 그들의 학습기능을 컴퓨터 알고리즘으로 시뮬레이션 한다. 신경망은 뉴런을 모델링한 유닛(unit)들과 그 유닛 사이의 가중치 연결(weighted-connection)들로 이루어지며, 각 신경망 모델에 따라 다양한 구조와 각기 독특한 학습 규칙을 갖는다. 이러한 신경망은 불완전하거나 사전에 알 수 없던 입력을 표현하는 경우에 합리적인 반응생성이 가능하고, 새로운 환경에서 즉각적으로 프로그램을 갱신하고 유지하는데 사용 될 수 있는 장점이 있지만, 주로 정적인 패턴을 인식하는데 사용되고 있다.

HMM은 1980년대 중반 경에 그 구체적인 방법과 이론이 Bell연구소 연구자들에 의해 발표되었으며, 시계열 패턴 인식에 널리 활용되고 있다.

본 논문에서는 은닉 마르코프 모델을 이용하여 지휘영상에서의 박자인식 방법을 연구하였으며, 2장에서는 박자인식 시스템의 구성에 대한 설명을 다루었으며, 3장에서는 HMM을 이용한 구체적인 모델을 제시하였고, 4장에서는 실험결과를, 마지막 5장에서는 결론을 논하였다.

Ⅱ. 시스템 구성

전체적인 시스템 구성은 (그림 1)과 같다. USB PC 카메라로부터 받은 영상을 영역 추출기를 통해 원하는 손 영역만을 추출해 낸 후, 손 추적기를 통해 지휘자의 손을 추적해 간다. 지휘자가 한 박자를 완료했을 때 Frame 결정기를 이용하여 HMM에 필요한 관측열을 생성하고, 미리 학습된 HMM 엔진을 이용하여 지휘자의 박자를 알아낸 후 화면에 그 결과를 출력하는 순으로 시스템을 구성하였다.

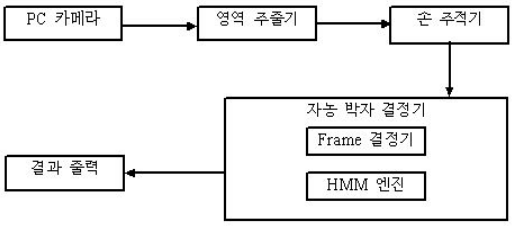

그림 1. 시스템 구성도
Fig. 1. System configuration

영역 추출기에서는 먼저 USB PC 카메라로부터 받은 영상을 YIQ 칼라채널(Color Channel)을 이용하여 영상을 분리한다. 이는 RGB 칼라채널을 이용하는 것보다 YIQ 채

널의 I 값이 조명의 변화에 덜 민감하며, 동양인의 피부색을 잘 표현해 주기 때문이다. 분리된 채널의 I 값을 이진화 하여 영상을 단순화시킨다. RGB 와 YIQ 의 관계는 다음과 같다.

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.28 & -0.32 \\ 0.21 & -0.52 & 0.31 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \cdots\cdots\cdots\cdots (1)$$

손 추적기는 영역 추출기를 통해서 나온 결과영상에서 원하는 대상(손)만을 찾고 그 움직임을 추적하는 역할을 한다. 손 영역만을 찾기 위해서는 영역 추출기를 통해 나온 영상에서의 연결요소(connected com-ponent)에 레이블을 부여한다. 레이블링 된 영상에서 얼굴과 손 그리고 기타(배경, 잡음 등)영역의 면적 이용하여 원하는 손 영역을 찾아내고 중심점을 구한다.

손 추적기를 통해 찾은 손의 특징 점을 프레임 결정기를 통해 손의 이동위치를 추적해 가고 박자가 끝났을 때 지금까지의 손 위치 점의 최대, 최소 포인트를 이용하여 정규화된 프레임으로 만들어주며 각 각 손의 위치들이 어느 프레임요소에 속하는지를 표시하여 sequence를 생성한다. 본 논문에서는 프레임의 크기를 60*70 으로 하였다. (그림 2)는 지휘자가 2/4박자를 지휘했을 때 프레임 생성의 예이다.

예에서는 2/4박자를 지휘하기 위해 18프레임의 영상을 사용하였다. 정규화 된 프레임을 생성하는 과정은 다음과 같다.

- 손 추적기를 이용한 손의 위치 저장
- 저장된 위치에서 최소 점(x, y)와 최대 점 (x, y)를 구한다.
- 프레임의 크기인 60*70에 맞게 사각형을 형성한다.

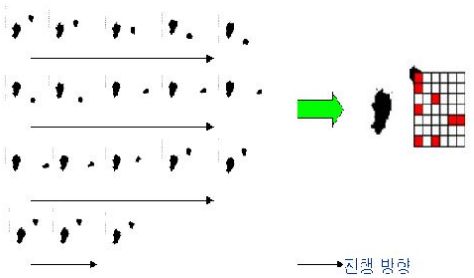

그림 2. 2/4박자 프레임 생성의 예
Fig 2. Frame generation for 2/4beat

III. HMM과 이를 이용한 지휘패턴인식

1. HMM (Hidden Markov Model)

인식할 패턴이 시간에 따라서 형성되는 시계열 패턴의 경우, 패턴의 생성과정을 잘 모델링 할 수 있다면 패턴인식을 효율적으로 수행할 수 있는 경우가 있다. 은닉 마르코프 모델이란 마르코프 연쇄에 기초한 시계열 패턴의 생성과정에 대한 확률적인 모델이다. HMM의 매개변수들을 정의하면 다음과 같다.

N : 은닉상태들의 수
M : 관측 기호 집합의 원소의 개수
T : 관측열의 길이
$O = \{ O_1, O_2, \cdots, O_T \}$: 관측열
$Q = \{ q_1, q_2, \cdots, q_N \}$: 모델 내의 상태 집합
$V = \{ v_1, v_2, \cdots, v_M \}$:
관측 가능한 기호들의 이산집합
$A = \{ a_{ij} \}$: 상태 전이 확률분포
$a = P(q_{t+1} = j \mid q_t = i)$ $\quad 1 \le i, j \le N$
$B = \{ b_j(k) \}$: 상태 j에서의 관측 기호 확률 분포
$B(k) = P(o_t = k \mid q_t = j)$ $\quad 1 \le k \le M$
$\pi = \{ \pi_i \}$: 초기상태 분포
$\pi = P(q_i \ at \ t = 1)$: 상태 I의 초기 확률 값

여기서 A, B, π 의 원소들은 확률변수이므로 다음과 같은 속성을 만족한다.

$$\sum_i a_{ij} = 1, \ a_{ij} \ge 0, \text{ for all i} \cdots\cdots\cdots\cdots (2)$$

$$\sum_k b_j(k) = 1, \ b_j(k) \ge 0, \text{ for all j} \cdots\cdots\cdots\cdots (3)$$

$$\sum_i \pi_i = 1, \ \pi_i \ge 0 \cdots\cdots\cdots\cdots (4)$$

HMM은 $\lambda=(A, B, \pi)$로 표기한다. 모델 λ를 설정하기 위해서는 N과 M을 선택해야하며, 세 가지 확률밀도 A, B, π를 정의해야 한다. HMM은 이러한 매개변수들을 사용해서 주어진 관측 열들을 표현한다. HMM은 외부에서 상태전이 과정을 관찰할 수 없다. 그러나 HMM 자체의 확률 매개변수를 사용하여 마르코프 과정의 확률 함수로 모델링할 수 있다.

HMM은 다음과 같은 세 가지 문제의 해결에 의해서 최적의 $\lambda=(A, B, \pi)$를 구성하고, 입력 열에 가장 적합한 모델을 찾을 수 있다.

- 첫 번째 문제는 관측열 $O=\{O_1, O_2, \cdots, O_T\}$ 와 모델 $\lambda=(A, B, \pi)$가 주어졌을 때 관측열의 확률 $P(O|\lambda)$를 어떻게 효율적으로 계산할 것인가의 계산문제로서 전향변수와 후향변수를 이용한 계산방식이 사용된다[5].
- 두 번째 문제는 관측열 $O=\{O_1, O_2, \cdots, O_T\}$ 와 모델 $\lambda=(A, B, \pi)$가 주어졌을 때 최적의 상태 열 $q=\{q_1, q_2, \cdots, q_N\}$ 를 어떻게 선택할 수 있는가의 은닉 상태 열을 찾는 문제로서 Viterbi 알고리즘이 사용된다[7].
- 세 번째 문제는 $P(O|\lambda)$를 극대화시키는 모델 $\lambda=(A, B, \pi)$의 매개변수를 어떻게 조정할 수 있는가의 학습문제로서 Baum-Welch 알고리즘이 사용된다[5].

2. HMM을 이용한 지휘패턴인식

자동 박자 인식에서 사용한 지휘법은 4가지(2박자, 3박자, 4박자, 6박자)를 실험하였으며 (그림 3)과 같다.

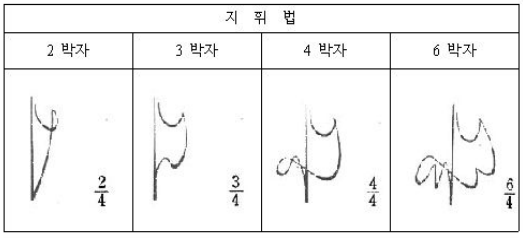

그림 3. 지휘법의 종류
Fig 3. Types of conducts for beats

같은 박자의 악곡이라도 곡의 빠르기에 따라 박자짓는 모양이 달라질 수 있으나 실험에서는 이 사항을 고려하지 않고 실험을 하였다.

HMM의 학습은 Baum-Welch 학습방법을 사용하였다. HMM의 인식에는 Forward Algorithm을 사용하였으며 인식 실험에 사용한 학습데이터의 입력 열은 USB PC 카메라를 통해 획득한 영상을 이용하여 데이터를 얻었다.

실험영상에서 학습에 필요한 관측열(Observation Sequence)을 구하는 방법은 2장에서 프레임 결정기에 관해서 언급한대로 각 프레임의 손의 위치를 해당 프레임 에 체크함으로 입력 열을 생성하였다. (그림 4)는 3절의 (그림 2)에서 결과 영상만을 가져온 것이다.

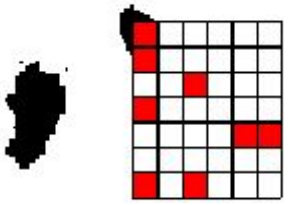

그림 4. 생성된 관측열
Fig 4. Generated observation sequence

위의 예에서 첫 프레임의 값을 0 이라 했을 때 관측 열 O 를 다음과 같이 얻을 수 있다

O = {0, 5, 15 ,30, 30, 30, 31, 23, 24, 24, 24,
24, 23, 12, 0, 0, 0}

생성한 관측 열을 HMM 엔진을 이용하여 여러 모델 중 가장 확률이 높은 모델을 그 결과로 출력한다. 본 논문에서는 (그림 5)와 같은 우향(left-to-right)구조로 구성하였다.

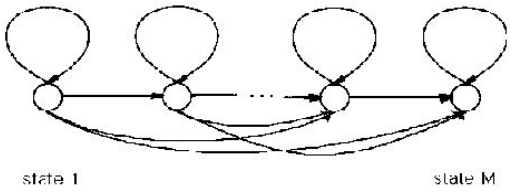

그림 5. M개의 상태를 갖는 left-to-right 모델
Fig 5. Left-to-right model with M states

또한 HMM 매개변수들 중 M(관측 기호 집합의 원소의 개수)을 42로, N(은닉 상태들의 수)의 크기를 12~18 사이로 하였으며 δ값을 0.024로 하여 학습하였다.

Ⅳ. 실험결과

실험은 Windows 98을 운영체계로 하는 Pentium III CPU를 사용하였으며, 초당 20 프레임의 속도로 160*120 해상도를 갖는 USB PC 카메라를 입력 장치로 사용하고 Visual C++를 이용하여 프로그램을 작성하였다.

〈표 1〉은 지휘자가 각각 2박자, 3박자, 4박자, 6박자를 순서대로 지휘하였을 경우 Forward Algorithm을 통해 나온 결과치를 보여주고 있다.

표 1. 결과치 비교표
Table 1. Comparison table for the results

입력 / 모델	A	B	C	D
모델 1 (2박자)	9.387	8.34	1.48	1.92
모델 2 (3박자)	1.73	7.13	2.38	5.07
모델 3 (4박자)	4.17	3.07	8.83	2.05
모델 4 (6박자)	1.71	3.27	5.58	7.34

입력패턴 A, B, C, D는 각각 2박자, 3박자, 4박자, 6박자의 지휘패턴을 나타낸다. 〈표 1〉의 결과에서 알 수 있듯이 결과의 각각의 해당 박자가 다른 모델에 비해서 가장 큰 수치를 보임을 알 수 있다. 이 수치는 여러 가지 변수에 의해서 달라질 수 있다. 실험결과 화면은 (그림 6)과 같다.

그림 6. 4박자 테스트화면
Fig 6. Screen of test for 4 beat conduct

Ⅴ. 결론

본 논문에서는 PC 카메라로부터 입력된 영상에서 손동작만을 인식한 후 HMM을 이용하여 지휘자의 박자를 인식하는 시스템을 구현하였다. 지휘영상에서 칼라변환과 이진화를 통해 영상을 단순화시키고 레이블링을 통해 손의 위치를 식별하도록 하였으며, 프레임 결정기를 통하여 정형화된 크기의 영역을 결정하도록 함으로써 지휘자의 위치에 따른 변이를 흡수하도록 함으로써 신뢰도를 높였다. 또한 HMM을 이용해서 지휘패턴을 분류함으로 써 매 지휘마다 발생할 수 있는 적절한 범위 내의 위치변이나 시간 차이를 극복하도록 하였다.

본 논문에서는 박자 인식에 초점을 둔 연구를 목적으로 하였기 때문에 순수하게 지휘 박자를 인식하는데 중점을 두었다. 따라서 실행과 동시에 지휘가 시작하는 것으로 약속하였으며, 영상 내에서 손의 위치를 찾지 못 하였을 경우를 박자의 끝으로 정의하였는데, 이러한 제약조건은 박자의 시작과 끝의 패턴을 연구하면 해결할 수 있을 것으로 기대된다.

참고문헌

[1] 이성환, 문자인식 이론과 실제 제 1 권, 홍릉과학출판사, 14장, 1997

[2] 이성환, 패턴인식의 원리 II권, 홍릉과학출판사, 1997

[3] 이현규, 김진형, PowerGesture: 제스처 Spotting기법을 이용한 발표 지원 시스템, HCI 97 학술대회 발표논문집, 1997

[4] Bemdt, D. & Clifford, J. "Using dynamic time warping to find patterns in time series", AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94), Seattle, Washington, 1994

[5] Carlos Morimoto, Yaser Yacoob, Larry Davis, "Recognition of Head Gestures Using Hidden Markov Models," International Conference on Pattern Recognition, Austria, pp461-465, 1996.
[6] C.Maggioni, "A Novel Gesture Input and Device for Virtual Reality", Proc, of IEEE Reality Annual International Symposium, 1993
[7] G.D.Forney Jr., "The Viterbi Algorithm," Proc. IEEE, vol 61, no. 3. 1973
[8] K.Takahashi, S.Seki, and R.Oka, "Spotting Recognition of Human Gesture from Motion Images", Technical Report IE92-134, The Institute of Electronic, Information and Communication Engineers(Japan), 1992
[9] L. R. Rabiner and B. H. Jang, "An introduction to Hidden Markov models," IEEE ASSP Mag, 1986
[10] R,Kjeldsen and J.Kender, "Visual Hand Gesture Recognition for Window System Controls", Proc. Int. Workshop on Automatic Face and Gesture Recognition (IWAFGR), 1995
[11] S.Fels and G.E.Hint on, "Glove-talk: A Neural Network Interface between a Dataglove and a Speech Synthesizer", IEEE transactions on Neural Newworks, Vol. 4, 1993
[12] S.Seki, K.Takahashi and R.Oka, "Gesture Recognition from Motion Images by Spot ting Algorithm", Proc. of Asia Conference on Computer Vision(ACCV), 1993
[13] T.Starner and A.Pentland, Real-Time American Language Recognition from Video Using Hidden Markov Models, Technical Report TR-375, Media Lab, MIT, 1995
[14] X.D.Huang, Y.Ariki, and M.A.Jack, Hidden Markov Models for Speech Recognition, Edinburgh Univ. Press, 1990
[15] Yamron, J., Carp, I., Gillick, L., Lowe, S., & van Mulbergt, P. A Hidden Markov model approach to text segmentation and event tracking. In Proceedings of the IEEE ICASSP Seattle, Washington, 1998

저 자 소 개

문 형 득

1997년 충북대 수학과 학사
2002년 단국대학교 대학원 전산통계학과 석사
현 재 (주)프라임 모바일

구 자 영

1977년 서울대학교 전자공학과 학사
1980년 한국과학기술원 전기 및 전자공학부 석사
1986년 한국과학기술원 전기 및 전자공학부 박사
1986년 ~ 현재 단국대학교 정보 컴퓨터학부 교수