

지능형 영상 감시 시스템에서의 은닉 마르코프 모델을 이용한 특이 행동 인식 알고리즘

정창욱* 강동중

부산대학교 기계공학과

jcwook*@pusan.ac.kr, djikang@pusan.ac.kr

Specific human behaviors recognition algorithm using Hidden Markov Models in an intelligent surveillance system

Changwook Jung* Dongjung Kang

Dept. of Mechanical Engineering, Pusan National University

요 약

본 논문은 Hidden Markov Model을 사용하여 사람의 특정한 행동을 인식하여 사용자에게 알려주는 지능형 영상 감시 시스템을 제안한다. 본 방법에는 카메라를 통해 입력된 영상에서 사람 영역을 찾은 후 발영역만을 추출하여 특징이 되는 관측열을 생성한다. 특징 영역은 입력 영상의 각 프레임을 16개의 영역으로 나누어 발바닥이 위치한 곳의 코드를 읽어 사용하고, 인식하고자하는 패턴 행동들에 대해서는 각각의 관측열을 구하고 HMM의 Baum-Welch 알고리즘을 사용하여 학습한다. 인식에는 전향 알고리즘을 사용하여 입력된 행동과 학습된 행동을 확률적으로 비교함으로써 영상 내의 행동이 어떤 패턴 행동인지를 결정하여 출력하도록 한다. 제시된 방법은 복도에서 사람의 특정 행동을 인식하는데 성공적으로 적용될 수 있음을 실험을 통해 확인 하였다.

1. 서 론

영상 감시 시스템은 교통량 감시, 출입자 확인, 그리고 공공장소에서의 범죄 예방 등 많은 분야에서 인력을 보조하거나 대체하는데 유용하게 사용될 수 있다. 특히 전자기술의 급속한 발전에 따라 산업계의 공장 자동화나 빌딩 자동화 시스템에서의 영상 감시 시스템의 사용은 이미 일반화되었고, 그 외 일반 사회생활의 여러 분야, 아파트 관리 시스템이나 현금 입출기, 슈퍼마켓, 편의점, 교차로 등의 무인 경비 및 관리 시스템에서의 영상 감시 시스템의 사용도 빠른 속도로 확산되어 가고 있다. 따라서 각 상황에 맞는 지능형 영상 감시 시스템 개발의 필요성은 점차 커지고 있다.

CMU의 Video Surveillance and Monitoring(VSAM) 프로젝트[1], MIT AI Lab의 Forest of Sensors 프로젝트[2]가 지능형 영상 감시 시스템의 좋은 예라고 볼 수 있다. 지능형 영상 감시 시스템의 장점으로는 기존의 영상 감시 시스템에서는 도난 사건이 일어나면 상황이 종료된 후 저장된 기록을 통해서 증거를 획득하게 되지만 지능형 영상 감시 시스템에서는 상황 발생 즉시 도난 사건을 인지하고 처리 하게 되는 것이다. 또한 기존의 감시 시스템에서는 사람이 항상 모니터를 주시하고 상주해야 되나 지능형 영상 감시 시스템에서는 특이 상황 발생 즉시 사용자에게 알려주기에 그럴 필요가 없다.

이러한 지능형 영상 감시 시스템에 효과적으로 적용될 수 있는 신뢰성이 높은 알고리즘을 제안하고자 한다. 기

존의 영상 감시 시스템 연구에서는 특이 영역을 고속 추적하여 다양한 사물에 대해서 위치를 추적하는 시스템 등이 제안되어 있다[3]. 이러한 방법은 어떤 특정한 행동을 인식하고 알려주는 지능형 영상 감시 시스템에 적용하기에는 한계가 있다. 따라서 본 논문에서는 Hidden Markov Model (HMM)[4] 알고리즘을 사용하여 사람의 특정 행동을 인식하는 시스템을 제안한다. HMM은 최근 몇 년 동안 컴퓨터 비전에서 움직임 검출 알고리즘으로 많이 사용되었으며, 특히 핸드 제스처 인식[5], 얼굴 표정 인식[6], 테니스 스트로크 인식[7] 등이 대표적인 예이다. 적용사례들의 공통적 특징은 시계열 패턴, 즉 시간의 흐름에 따라 움직임의 변화가 일어나는 것을 특징으로 사용한다는 것인데 이는 HMM이 시계열 패턴 인식에 잘 적용된다는 것을 잘 보여주는 예이다.

본 논문에서는 1장의 서론에 이어 2장에서는 지능형 영상 감시 시스템에 대한 개괄적인 설명을 하고 3장에서는 시스템의 실험 결과를 기술한다. 마지막으로 4장에서는 본 논문의 결론 및 향후 과제에 대해 기술한다.

2. 지능형 영상 감시 시스템

2.1 시스템 구성

전체적인 시스템 구성은 그림 1과 같다. PC캠으로부터 받은 영상을 영역 추출기를 통해 원하는 사람 영역만을 추출해 낸 후, 다시 발자국 영역을 추출한다. 각 Frame

마다 발자국 영역을 특징 데이터로 하는 관측열을 생성하고, 미리 학습된 HMM 엔진을 이용하여 PC캠을 통해 입력된 사람의 행동을 알아낸 후 화면에 결과를 출력하는 순으로 시스템을 구성하였다.

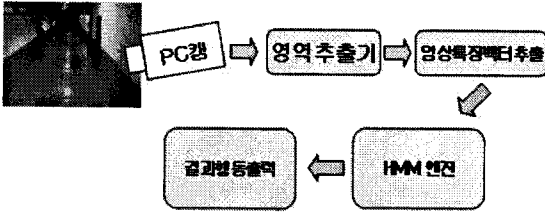


그림 1 지능형 영상 감시 시스템 구성도

$Q=(q_1, q_2, \dots, q_N)$, M 개의 관측 가능한 기호 $V=(v_1, v_2, \dots, v_M)$ 을 가진다고 가정한다. 그리고 시간 t 에서의 상태를 S_t 라고 표시하고 HMM의 파라미터를 $\lambda=(A, B, \pi)$ 로 표현한다.

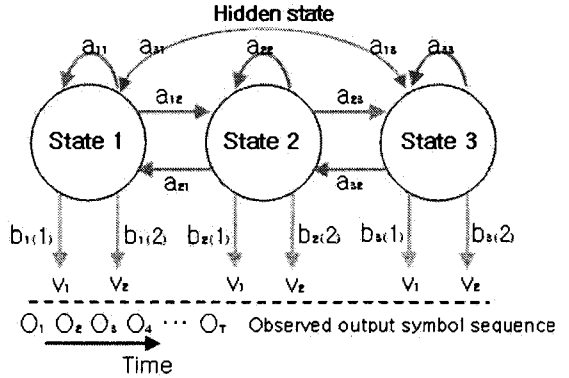


그림 2 이산적 데이터의 HMM 처리과정

2.2 복도에서의 사람 행동인식

가. Hidden Markov Model

HMM은 관찰이 불가능한 미지(hidden)의 확률적 과정(stochastic process)을 관찰이 가능한 기호(symbol)를 발생시키는 다른 확률적 과정을 통하여 모형화(modeling)하는 이중의 확률론적 과정이다. 예를 들어 우리가 쉽게 관찰 할 수 있는 음파나 공기의 압력, 또는 발음기관의 위치 등을 이용하여 두뇌의 활동을 모형화하는 것을 생각해 볼 수 있다. 이러한 HMM은 <표 1>과 같이 2개의 상태집합과 3개의 확률 집합으로 구성되는 5개의 요소를 갖는다.

표 1. HMM의 5가지 요소

1. 은닉상태집합(hidden state set) : 마르코프 프로세스에 의해서 설명되는 상태들의 집합
2. 관찰가능 상태집합(observable state set): 의형적으로 눈에 보이는 전이 상태들의 집합
3. π 벡터: 특정 은닉 상태가 시간 $t=1$ 일 때 모델의 확률
4. 상태전이 행렬 : 이전의 은닉상태에서 현재의 은닉상태로의 전이 확률을 나타내는 것으로 모델 내부의 은닉상태들 간의 전이 확률을 나타내는 행렬
5. 관찰확률 행렬 : 특정 은닉상태에서의 관찰 가능한 각각의 상태들에 대한 확률을 나타내는 행렬

이와 같이, HMM은 관찰 가능한 상태들과 은닉상태들 간의 확률적 관계를 이용하여 계산되어지는 알고리즘이다.

그림 2는 이산적인 시계열 데이터들의 HMM 처리 과정을 나타내었다. 다른 상태로의 변환 확률은 이전의 상태들과는 무관하며, 단지 바로 전 단계의 상태에 의해서만 영향을 받는 것을 볼 수 있다. HMM이 N 개 상태

$$A \equiv \{a_{ij} | a_{ij} = P(S_{t+1} = q_j | S_t = q_i)\} \quad (1)$$

$$B \equiv \{b_i(k) | b_i(k) = P(v_k | S_t = q_i)\} \quad (2)$$

$$\pi \equiv \{\pi_i | \pi_i = P(S_1 = q_i)\} \quad (3)$$

식 (1)은 $N \times N$ 상태변환 확률 행렬을 나타내고, 식 (2)는 $M \times N$ 상태들에서의 관측 기호의 확률 행렬을 나타내며, 식 (3)은 초기상태 확률 벡터를 나타낸다.

HMM $\lambda=(A, B, \pi)$ 에 대해서, 관찰되어진 관측열 $O=(O_1, O_2, \dots, O_T)$ 의 확률은 베이즈의 법칙 $P(O|\lambda)$ 로 계산되어진다. 이 확률값은 forward 알고리즘[4]을 사용하여 추정되어진다.

$$P(O|\lambda) = \sum_{i=1}^N \alpha_i(i) |_{t=T} \quad (4)$$

forward 변수 $\alpha_i(i)$ 는 아래와 같이 정의되고, 이는 시간 t 에서 상태가 i 에 있을 때 부분상태 열에 대한 관찰 확률을 의미한다.

$$\alpha_i(i) \equiv P(O_1, O_2, \dots, O_t, S_t = q_i | \lambda) \quad (5)$$

HMM의 학습은 학습데이터로부터 $P(O|\lambda)$ 를 최대화하는 최적의 λ 를 구하는 과정으로 Baum-welch 알고리즘[4]이 사용된다.

backward 변수 $\beta_i(i)$ 는 다음과 같이 정의되고 이는 앞서 설명한 forward 변수와 같은 관찰 확률을 의미한다.

$$\beta_i(i) \equiv P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = S_i, \lambda) \quad (6)$$

또한 γ, ξ 도 아래와 같이 정의하고 이는 Baum-Welch 알고리즘에 사용되는 변수로, γ 는 시간 t 에서 주어진 관찰열을 가지고 상태 i 에 있을 확률을 의미하고 ξ 는 시간 t 에서 상태 i 에 있고, 시간 $t+1$ 에서 상태 j 에 있을 확률을 의미한다.

$$\gamma_i(i) = P(S_t = q_i | O_1, \dots, O_T, \lambda) = \frac{\alpha_i(i)\beta_i(i)}{P(O|\lambda)} \quad (7)$$

$$\xi_i(i, j) = P(S_t = q_i, S_{t+1} = q_j | O_1, \dots, O_T, \lambda) = \frac{\alpha_i(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} \quad (8)$$

이전에 주어진 파라미터 λ 는 아래의 식(9)에 의해 개선된 λ' 로 추정되어 진다.

$$a'_{ij} \equiv \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, b'_i(k) \equiv \frac{\sum_{t \in \{t | Q_t = v_k\}} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (9)$$

$$\pi'_i \equiv \gamma_1(i)$$

학습은 $\lambda' = \lambda$ 가 될 때 즉 $P(O|\lambda)$ 의 극대화값이 수렴할 때까지 반복한다.

HMM알고리즘은 각 클래스의 관찰열을 가장 잘 표현하는 파라미터 λ 로 학습되어지고 미지의 관찰열이 들어왔을 때 학습되어진 λ 와 관찰열을 이용해서 확률적으로 미지의 관찰열을 알아내는 것이다.

나. 복도에서의 사람 행동 인식

지능형 영상 감지 시스템에 사용한 복도에서의 사람 행동은 5가지(A방으로 들어가는 행동, B방으로 들어가는 행동, A방에서 나오는 행동, B방에서 나오는 행동, 서성거리는 행동)이며 그림 3과 같다.

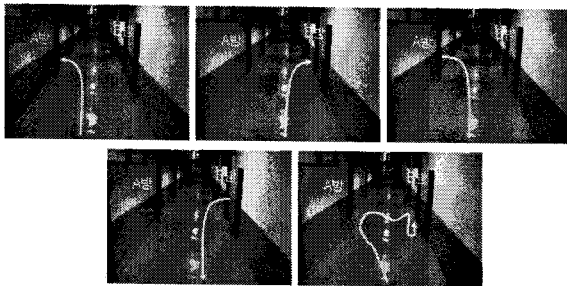


그림 3 복도에서의 사람 행동 종류

HMM의 학습은 Baum-welch 알고리즘 [5]을 사용하였고 HMM의 인식에는 forward 알고리즘을 사용하였다. 인식 실험에 사용한 학습데이터의 관찰열은 PC캠을 통해 획득한 영상을 이용하여 데이터를 얻었다.

실험영상에서 학습에 필요한 관찰열(Observation

Sequence)을 구하는 방법은 각 프레임에서 사람 영역을 찾은[8] 후 발 영역을 프레임에 체크함으로 관찰열을 생성하였다. 그림 4는 관찰열을 생성하는 것을 보여준다. 움직임을 추적하는 것은 확률적 운동 영역 감지 기술을 이용한다[9].

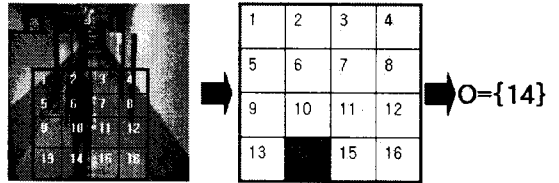


그림 4 관찰데이터 생성 방법

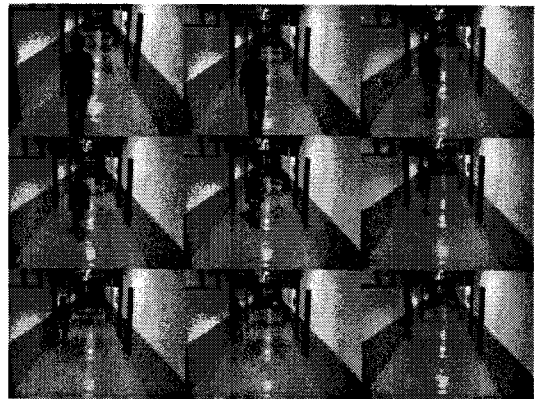


그림 5 A방으로 들어가는 영상

그림 5는 PC캠으로 촬영된 영상으로 발바닥 영역은 그림 6과 같이 생성된다.

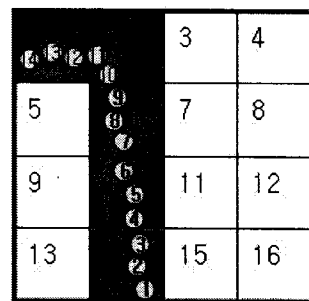


그림 6 영상에서의 발바닥 영역 표시

위 프레임으로부터 관찰열 O를 아래와 같이 얻을 수 있다.

$$O = \{14, 14, 14, 10, 10, 10, 6, 6, 6, 2, 2, 1, 1, 1\} \quad (10)$$

생성한 관찰열은 HMM을 이용하여 여러 행동 중 가장

확률적으로 높은 행동을 그 결과로 출력한다.

3. 실험결과

실험환경은 다음과 같다. Window XP를 운영체제로 하는 Intel Pentium-4 1.8G processor의 DeskTop 컴퓨터를 사용하였다. 초당 20 프레임의 속도로 320x240 해상도를 갖는 PC 캠을 입력 장치로 사용하였으며 Matlab을 이용하여 프로그램을 작성하였다.

실험에는 각 행동, 즉 A,B방으로 각각 들어가고 나오는 행동, 서성이는 행동에 대해서 20개씩의 관측열 구해서 Baum-Welch알고리즘으로 학습하여 5개의 HMM $\lambda = (A, B, \pi)$ 파라미터를 생성하였고 PC캠으로부터 입력된 복도 영상과 구해진 λ 들과의 확률 계산을 이용해 특정 행동을 인식하고자 하였다. HMM 상태 수 $N=3$ 으로 하였으며, 관측 가능한 기호 $M=16$ 으로 하였다. 그림 7은 HMM의 학습에 의한 우도값(Likelihood)의 그래프를 나타낸 것이다. 이 결과에서 보듯이 Baum-welch알고리즘은 반복이 되면서 우도가 지역적 최적화가 되는 값으로 수렴하는 것을 볼 수 있다.

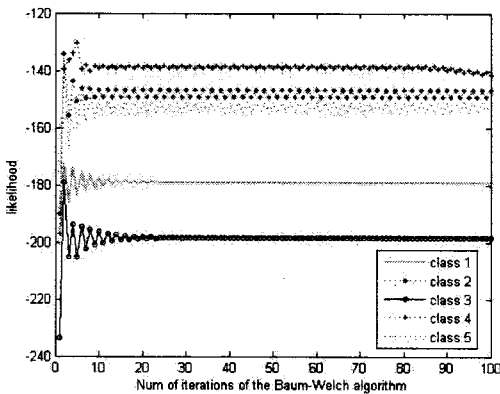


그림 7 HMM의 학습에 의한 우도

표 2 각 행동에 대한 로그-우도의 결과

	A	B	C	D	E
A방Enter	-20.13	$-\infty$	-77.78	$-\infty$	$-\infty$
B방Enter	$-\infty$	-25.43	$-\infty$	-67.34	$-\infty$
A방Exit	$-\infty$	$-\infty$	-24.31	$-\infty$	$-\infty$
B방Exit	-545.90	$-\infty$	$-\infty$	-16.74	$-\infty$
서성임	$-\infty$	$-\infty$	$-\infty$	$-\infty$	-26.42

<표 2>는 PC캠으로부터 입력된 영상 관측열에 대해서 각 패턴에 대한 우도값의 결과를 나타낸 것이다. 입력패턴 A,B,C,D,E는 각각 A방으로 들어가는 행동, B방으로 들어가는 행동, A방에서 나오는 행동, B방에서 나

오는 행동, 서성이는 행동을 나타낸다. 결과에서 알 수 있듯이 각각의 해당 행동이 다른 패턴 행동에 비해서 가장 큰 수치를 보임을 알 수 있다. 이 결과를 이용해서 현재의 입력 영상이 어떤 행동인지를 출력하게 되는 것이다.

4. 결론

본 논문에서는 PC 캠으로부터 입력된 영상을 이용하여 복도에서의 사람행동을 인식하는 방법을 제안하였다. 제안된 시스템은 영상 이미지에서 시계열 패턴인 관측열을 생성하고, HMM의 Baum-Welch알고리즘으로 학습하고 forward 알고리즘을 사용하여 인식하는 시스템을 구성하였다. 입력 행동과 미리 학습된 행동들의 비교를 통해 확률적으로 가장 높은 행동을 결과로 인식하는 것을 보였으며 성공적인 결과를 실험을 통해서 확인하였다.

본 논문에서는 지능형 영상 감시에 초점을 둔 기초연구의 목적으로 5가지의 단순한 행동에 대해서만 학습하고 인식하였다. 따라서 추후에는 다양한 행동, 예를 들면 사람이 넘어지는 행동, 폭력적인 행동 등에 대해서 실험하고 연구할 계획이다.

참고문헌

- [1] R.T.Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson, "A System for Video Surveillance and Monitoring," Technical Report CMU-RI-TR-00-12, Carnegie Mellon University, 2000
- [2] C. Stauffer and W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no. 8, pp. 747-757, August 2000.
- [3] M. Isard and A.Blake, "Condensation-conditional density propagation for visual tracking," Int. J. Computer Vision, vol 29, no.1, pp.5-28, 1998.
- [4] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol.77, no2, pp.257-286, 1989.
- [5] T.E. Starner and A. Pentland, "Visual Recognition of American Sign Language Using Hidden Markov Models," Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, pp. 189-194, 1995.
- [6] N. Oliver, A. Pentland, and F.Berard, "LAFTER: A Real-time Face and Lips Tracker with Facial Expression Recognition," Pattern Recognition, vol. 22, pp. 1369-1382, 2000.
- [7] J.Vamato, J. Ohya, and K.Ishii, "Recognizing Human Action in Time Sequential Images Using Hidden Markov Model," 1992 IEEE Computer

- Society Conference on Computer Vision and Pattern Recognition, PP. 379-385, 1992.
- [8] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," Proceedings of International Conference on Pattern Recognition, pp. 627-630, 2000
- [9] Ahmed Elgammal, David Harwood, Larry Davis "Non-parametric Model for Background Subtraction," 6th European Conference on Computer Vision. Dublin, Ireland, Vol.1843, pp. 751-767, 2000.