

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

정창욱\*, 강동중\*\*

## 요 약

본 논문은 은닉 마르코프 모델을 사용하여 사람의 특정한 행동을 인식하여 사용자에게 알려주는 지능형 영상 감시 시스템을 제안한다. 본 방법에는 카메라를 통해 입력된 영상에서 사람 영역을 찾은 후 발 영역만을 추출하여 특징이 되는 관측열을 생성한다. 특징 영역은 입력 영상의 각 프레임을 16개의 영역으로 나누어 발바닥이 위치한 곳의 코드를 읽어 사용하고, 인식하고자 하는 패턴 행동들에 대해서는 각각의 관측열을 구하고 HMM의 Baum-Welch 알고리즘을 사용하여 학습한다. 인식에는 전향 알고리즘을 사용하여 입력된 행동과 학습된 행동을 확률적으로 비교하여 영상 내의 행동이 어떤 패턴 행동인지를 결정하여 출력한다. 제시된 방법은 복도에서 사람의 특정 행동을 인식하는데 성공적으로 적용될 수 있음을 실험을 통해 확인했다.

## A Recognition Algorithm of Suspicious Human Behaviors using Hidden Markov Models in an Intelligent Surveillance System

Chang-Wook Jung\*, Dong-Joong Kang\*\*

### ABSTRACT

This paper proposes an intelligent surveillance system to recognize suspicious patterns of the human behavior by using the Hidden Markov Model. First, the method finds foot area of the human by motion detection algorithm from image sequence of the surveillance camera. Then, these foot locus form observation series of features to learn the HMM. The feature that is position of the human foot is changed to each code that corresponds to a specific label among 16 local partitions of image region. Therefore, specific moving patterns formed by the foot locus are the series of the label numbers. The Baum-Welch algorithm of the HMM learns each suspicious and specific pattern to classify the human behaviors. To recognize the inputted human behavior pattern in a test image, the probabilistic comparison between the learned pattern of the HMM and foot series to be tested decides the categorization of the test pattern. The experimental results show that the method can be applied to detect a suspicious person prowling in corridor.

**Key words:** Suspicious pattern recognition(특이행동 감지), Human behavior classification(인간행동 분류), HMM(HMM), Intelligent surveillance system(지능형 영상 감시 시스템)

※ 교신저자(Corresponding Author) : 강동중, 주소 : 부산시 금정구 장전동 산30번지(609-735), 전화 : 051)510-2163, FAX : 051)514-1118, E-mail : dj kang@pusan.ac.kr  
접수일 : 2008년 5월 19일, 완료일 : 2008년 9월 24일  
\* 부산대학교, 지능기계공학 석사과정  
(E-mail : jcwook@pusan.ac.kr)

\*\* 부산대학교 기계공학부 조교수  
※ 본 연구는 교육과학기술부와 한국산업기술재단의 지역 혁신인력양성사업으로 수행된 연구결과입니다. 본 과제(결과물)는 교육인적자원부·산업자원부·노동부의 출연금으로 수행한 산학협력중심대학육성사업의 연구결과입니다.

# 1. 서 론

## 1.1 영상 감시 시스템

비디오 감시 시스템이 보편화 되면서 절도나 테러 등을 사전에 예방할 수 있는 강력한 수단이 되기는 했지만 실제로 보안 산업은 예상치 못한 도전에 당면하고 있다. 이는 방대한 양의 비디오 데이터에 대한 문제이다. 비디오 데이터의 홍수에서 벗어나기 위해 정부 기관과 대기업들은 고성능 디지털 신호처리 프로세서로 작동하는 지능형 소프트웨어에 관심을 돌려 솔루션을 찾고 있다.

비디오 감시 시스템은 지난 수년간 사회 안전 시스템(인프라)을 구축하려는 정부와 이에 보조를 맞추는 기업에서 높은 중요도를 차지해왔다. 아날로그에서 디지털 시스템으로 이동하는 추세는 원활히 진행되고 있으나 압축 비디오의 인력에 의한 감시 및 내용분석이라는 광범위하게 수용되는 패러다임은 아직도 과거 지향적인 전략이다.

지능형 감시의 구현은 보안 시스템에서 패러다임의 변화를 의미한다. 이러한 솔루션은 증가된 효율로 인해 디지털 비디오가 아날로그를 대체하는 추세를 가속화시킬 것이다. 제어는 중앙 집중 방식의 모델에서 분산형 모델로 발전하게 될 것이며, 보안 직원은 통제실에서 나와 현장에 보다 가까이 다가갈 수 있게 될 것이다. 초고속 인터넷 망의 보급을 통한 스트리밍 비디오의 대역폭 요건과 비교해 볼 때, 비디오 영상을 검출 초기 단계에서 소프트웨어가 인간을 대체할 경우 고도로 지능화된 선택을 하는 데에는 그다지 많은 이미지 정보가 필요하지 않다. 콘텐츠 분석 엔진과 추론 엔진을 포함한 지능형 소프트웨어를 사용하여 비디오에서 발생된 이벤트를 분석하는 것은 감시 시스템의 모든 수준에서 장점을 갖는다. 뿐만 아니라 카메라나 카메라모듈 내부에 이러한 지능을 가능한 한 많이 통합할 경우 프로세싱 부하를 더욱 줄이고 정보의 품질을 높이며 보다 신속하게 다양한 상황에 대처할 수 있다.

그림 1은 시대적 요구에 따른 감시 시스템의 변화를 보여주고 있다.

얼마 전 국보 1호인 승례문이 방화로 전소되는 사건이 있었다. 수백 년 세월을 지내온 자랑스러운 민족의 유산이 소홀한 감시와 관리로 인해 한 순간에 사라져 버리게 된 것이다. 또한 최근에는 무고한 시

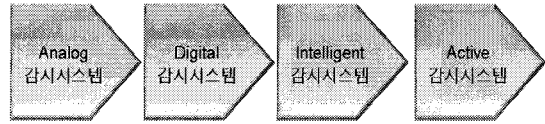


그림 1. 시대적 요구에 따른 감시 시스템의 변화

민을 대상으로 자행되는 지하철 테러와 수십 년을 지켜온 산림을 한 순간에 잿더미로 만드는 산림 방화 등과 같은 사건이 빈번하게 발생하고 있어 이를 관리/감시하기 위해 매년 많은 비용과 인력이 투입되고 있는 상황이다.

그럼에도 불구하고 최근에 있었던 익산 귀금속 센터 보석 도난 사건(2005년 말에 있었던 익산 귀금속 센터 보석 도난 사건은 국내 최대 규모의 익산 귀금속 센터에서 67억 원어치의 보석이 도난당한 절도 사건으로, 무인경비 보안 업체 및 CCTV 감시 카메라와 같은 첨단 보안 시스템을 구축했음에도 불구하고 실시간 대처 미흡으로 막대한 손실을 초래 하였다. (2005년 5월11일, 조선일보))에서 보는 바와 같이 무인 경비 시스템이나 CCTV 감시 카메라와 같은 기존의 보안 시스템은 사람의 집중과 판단에 많이 의존하기 때문에 관리 소홀로 인한 취약점에 어쩔 수 없이 노출되어 있다.

지능형 감시 시스템(Intelligent Surveillance System)은 감시 카메라의 영상 정보를 디지털화 하여 시스템이 스스로 사물의 위치 및 패턴을 분석하고 위험 상황이 발생할 경우 경보 및 출입 차단 시스템을 작동하거나 보안 담당자에게 즉각 현황을 알려 사람의 관리로 인해 발생하는 취약점을 대체/보완할 수 있게 해준다.

아직까지 지능형 감시 시스템의 구축에는 프라이버시 침해에 따른 법적인 문제와 고해상도 카메라, 광대역 전송매체와 같은 기술적인 이슈, 발생 가능한 위반 사항을 미리 고려해서 구축해야 하는 관리 시나리오의 구성과 같은 운영상의 이슈가 존재하고 있지만, U-City 구축과 같은 차세대 도시 관리 시스템에 대한 요구, 보안 인건비 상승에 따른 기업의 운영비용 증가, 감시카메라 영상의 법적 증거 채택이 가능하도록 하는 법령 개정의 추진, 보안 업계와 고객의 인식 제고라는 호재로 인해 향후 지능형 감시 시스템 구축에 대한 요구는 더욱 높아질 것으로 전망된다.

### 1.2 지능형 영상 감시 시스템의 필요성

영상 감시 시스템은 교통량 감시, 출입자 확인, 그리고 공공장소에서의 범죄 예방 등 많은 부분에서 인력을 보조하거나 대체하는데 유용하게 사용될 수 있다. 특히 전자 기술의 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 인출기, 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 처리하게 되는 것이다. 또한 기존의 감시 시스템에서는 사람이 항상 모니터를 주시하고 상주해야 하나 지능형 영상 감시 시스템에서는 특이 상황 발생 즉시 담당자에게 알려준다.

본 논문에서는 지능형 영상 감시 시스템에 효과적으로 적용될 수 있는 신뢰성이 높은 알고리즘을 제안하고자 한다. 기존의 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 은닉 마르코프 모델 (Hidden Markov Model : HMM)[4] 알고리즘을 사용하여 사람의 특정 행동을 인식하는 시스템을 제안한다. HMM은 최근 몇 년 동안 컴퓨터 비전에서 움직임 검출 알고리즘으로 많이 사용되었다. 특히 J. Yamato 등은 테니스 선수의 자세를 인식하기 위해 이산 HMM을 사용하였으며[5], A. Kale 등은 이진화된 이미지에서 실루엣의 폭을 이용해 특징 벡터를 생성하고 나서, 생성된 벡터를 이용해 k-평균 분할 (k-means clustering)을 수행하여 5개의 기본자세를

선택하였다. 이들 5개의 기본자세와 입력 특징 벡터와의 유클리드 거리를 값으로 하는 5차원 벡터를 생성하고, HMM을 인식기로 하여 89.09%의 성능을 얻었다[6,7]. 한편, J.J Little 등은 광유(optical flow)에 의한 방식을 사용하여 개개인의 걸음걸이에 대한 움직임 및 주파수와 위상(phase)의 특징을 찾고, 이를 인식에 이용하였다[8]. R. Collins 등에 의한 실험에서는 형틀 정합을 이용하여 관찰하는 위치에 의존적인 방식을 소개하였으며[9], C. Ben Abdelkader 등은 추출된 이미지간의 상관관계를 이용하여 자기유사도(self-similarity)를 계산하고, PCA와 k-NN 방법을 이용하여 벡터의 차원을 낮추어 인식하는 방법을 제안하였다[10]. 이후로도 핸드 제스처 인식[11], 얼굴 표정 인식[12] 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 보여주는 예이다.

본 논문에서는 1장의 서론에 이어 2장에서는 지능형 영상 감시 시스템의 행동 패턴분류 알고리즘에 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 2와 같다. PC 캠으로부터 받은 영상을 움직임영역 추출기를 통해 원하는 사람 영역만을 추출해 낸 후, 다시 발자국 영역을 추출한다. 각 프레임마다 발자국 영역을 특징 데이터로 하는 관측열을 생성하고, 미리 학습된 HMM 엔진을 이용하여 PC 캠을 통해 입력된 사람의 행동을 알아낸 후 화면에 결과를 출력하는 순으로 시스템을 구성하였다.

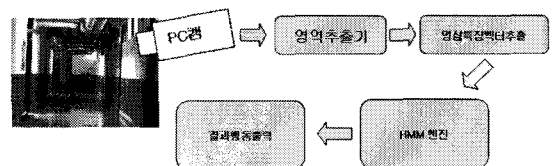


그림 2. 시스템 구성도

2.2 복도에서의 사람 행동인식

2.2.1 은닉 마르코프 모델

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 관찰할 수 있는 음파나 공기의 압력, 또는 발음기관의 위치 등을 이용하여 두뇌의 활동을 모형화하는 것을 생각해 볼 수 있다. HMM은 표 1과 같이 2개의 상태 집합과 3개의 확률 집합으로 구성되는 5개의 요소를 갖는다.

이와 같이, HMM은 관찰 가능한 상태들과 은닉 상태들 간의 확률적 관계를 이용하여 계산되는 알고리즘이다.

그림 3은 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

표 1. HMM의 5가지 요소

1. 은닉 상태 집합(hidden state set) : 마르코프 프로세스에 의해서 설명되는 상태들의 집합
2. 관찰가능 상태 집합(observable state set): 외형적으로 눈에 보이는 전이 상태들의 집합
3.  $\pi$  벡터: 특정 은닉 상태가 시간  $t=1$ 일 때 모델의 확률
4. 상태전이 행렬 : 이전의 은닉 상태에서 현재의 은닉 상태로의 전이 확률을 나타내는 것으로 모델 내부의 은닉 상태들 간의 전이 확률을 나타내는 행렬
5. 관찰확률 행렬 : 특정 은닉 상태에서의 관찰 가능한 한 각각의 상태들에 대한 확률을 나타내는 행렬

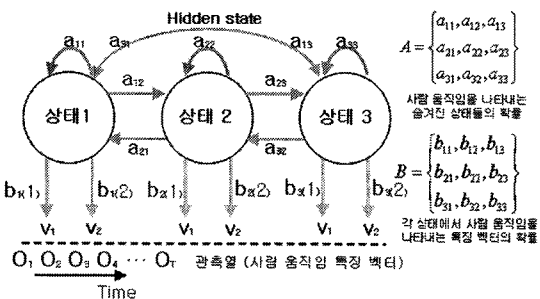


그림 3. 이산적 데이터의 HMM 처리과정

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

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

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

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

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 초기 상태 확률 벡터를 나타낸다.

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

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

전향변수  $\alpha_i(i)$ 는 아래와 같이 정의되고, 이는 시간  $t$ 에서 상태가  $i$ 에 있을 때 부분 상태 열에 대한 관측 확률을 의미한다.

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

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

후향(backward)변수  $\beta_i(i)$ 는 다음과 같이 정의되고 이는 앞서 설명한 전향변수와 같은 관측 확률을 의미한다.

$$\beta_i(i) \equiv P(O_{t+1}, O_{t+2}, \dots, O_T | S_t = q_i, \lambda) \tag{6}$$

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

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

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

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

$$a'_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{i=1}^{T-1} \gamma_t(i)}, \quad b'_i(k) = \frac{\sum_{t \in \{d|Q_t=v_k\}} \gamma_t(i)}{\sum_{i=1}^{T-1} \gamma_t(i)}, \quad \pi'_i = \gamma_1(i) \quad (9)$$

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

HMM 알고리즘은 각 클래스의 관측열을 가장 잘 표현하는 파라미터  $\lambda$ 로 학습 되고 미지의 관측열이 들어왔을 때 학습 된  $\lambda$ 와 관측열을 이용해서 확률적으로 미지의 관측열을 알아내는 것이다.

### 2.2.2 사람 운동 영역의 감지

사람의 특이 행동 인식을 위한 첫 번째 단계는 각 영상 프레임에서 사람을 찾는 것이다. 복도의 사람이 없는 배경 화면에서 전경 화면을 빼기 연산함으로써 사람 영역을 찾을 수 있다. 배경 모델  $B_n(x,y)$ 은 복도에 사람이 전혀 없는 연속된 비디오 프레임의 5프레임의 평균으로 초기화한다. 임계값  $T_n(x,y)$ 는 모든 픽셀 값을 50으로 초기화한다. 그리고 전경 모델을  $I_n(x,y)$ 로 둔다. 식 10은 전경 모델  $I_n(x,y)$ 에서 배경 모델  $B_n(x,y)$ 을 뺀 값이 임계값 보다 큰지를 나타낸다. 이 연산을 통해서 각 프레임에서 사람 영역을 찾아낼 수 있다[13].

$$|I_n(x,y) - B_n(x,y)| > T_n(x,y) \quad (10)$$

배경 모델과 임계값은 시간의 평균을 사용한 모든 프레임에 의해 업데이트되어 진다. 식 11, 12와 같다[13]. 그림 4는 운동 영역 감지 결과를 나타낸다.

$$B_{n+1}(x,y) = \begin{cases} B_n(x,y) & \text{if } (x,y) \text{ foreground} \\ \alpha B_n(x,y) + (1-\alpha)I_n(x,y) & \text{otherwise} \end{cases} \quad (11)$$

$$T_{n+1}(x,y) = \begin{cases} T_n(x,y) & \text{if } (x,y) \text{ foreground,} \\ \alpha T_n(x,y) + 2(1-\alpha)|I_n(x,y) - B_n(x,y)| & \text{otherwise} \end{cases} \quad (12)$$

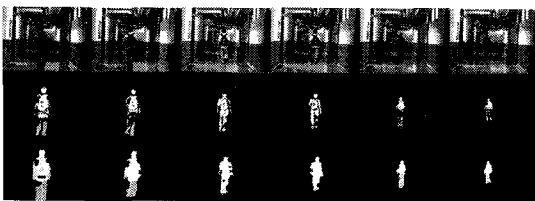


그림 4. 운동 영역의 감지

### 2.2.3 복도에서의 사람 행동 인식

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

HMM의 학습은 Baum-welch 알고리즘[5]을 사용하였고 HMM의 인식에는 전향 알고리즘을 사용하였다. 인식 실험에 사용한 학습데이터의 관측열은 PC 캠을 통해 획득한 영상을 이용하여 데이터를 얻었다.

실험 영상에서 학습에 필요한 관측열(Observation Sequence)을 구하는 방법은 각 프레임에서 사람 영역을 찾은[13] 후 발 영역이 위치하는 코드북을 체크하여 관측열을 생성하였다. 그림 6은 관측열을 생성하는 것을 보여준다. 그림 6에서 복도 영상에 씌운 영역 표시 마스크는 발이 위치할 수 있는 복도 바닥에 대해서 16 분할했으며 이해를 돕고자 영역을 중첩시켜 보여주고 있다. 카메라가 고정되어 있어 16개 영역은 일정하며, 이는 1 영역인 X, Y 좌표 (91, 158), (114, 158), (75, 178), (106, 178)의 네 점으로 이루어진 사각형으로부터 16 영역인 (191, 219), (243, 219), (203, 239), (270, 239)의 네 점으로 이루어진 사각형까지로 구성된다. 발 영역은 움직임 영역 추출을 통해 운동 영역인 사람 영상을 찾은 후 네 번의 모폴로지 팽창 연산과 네 번의 모폴로지 침식 연산의 과정을 거쳐

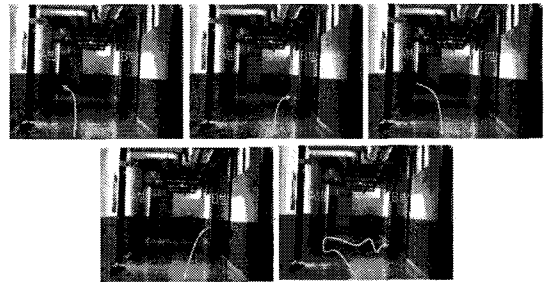


그림 5. 복도에서의 사람 행동 종류

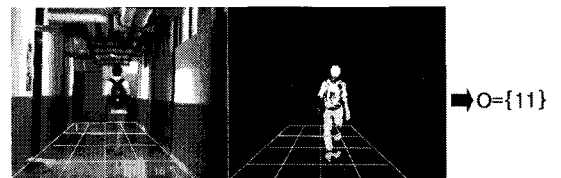


그림 6. 관측데이터 생성 방법

사람 영역이 한 덩어리가 되게 하였다. 그 후 이진화된 사람 영역의 상, 하, 좌, 우 각각의 최 외각 픽셀을 경계로 하는 MBR(Minimum Boundary Rectangle)을 추출하고 사각 박스 하단의 중심 픽셀 좌표를 사용하여, 16개 영역으로 나뉜 코드북에서 좌표가 위치하는 영역을 찾아 특징 벡터로 사용하였다.

움직임 영역의 추출은 확률적 운동 영역 감지 기술을 이용한다[14]. 사람 운동 영역이 MBR은 조명조건, 물체의 색과 밝기특성 등의 요인으로 한 사람이 여러 개의 MBR 영역으로 분할되어 나타날 수 있으며 이 경우 추출된 MBR중 가장 하단에 위치한 MBR을 발 영역으로 취급하였다.

그림 7은 PC 캠으로 촬영된 A방으로 들어가는 영상이다. 발 영역 표시를 위해 영상은 0.5초 단위로 추출하여 사용하였으며 사람이 복도에서 멈추지 않고 계속 걷는다고 가정했다. 관측열의 시작은 사람의 움직임이 관측되고 발 영역까지 영상에 나오는 평균 시간을 측정하여 움직임 관측 후 3초 후에 시작했다. 관측열의 끝은 사람이 복도에서 없어질 때 또는 임의로 정하였다.

위 프레임으로부터 관측열 O를 아래와 같이 얻을 수 있다.

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

각 행동, 즉 A, B방으로 각각 들어가고 나오는 행동, 서성이는 행동에 대해서 20개씩의 관측열을 구해서 Baum-Welch 알고리즘으로 학습하여 5개의 HMM  $\lambda=(A,B,\pi)$ 변수를 생성한다. HMM의 학습은 학습데이터(각 행동 영상에서 구한 관측열 O)로부터

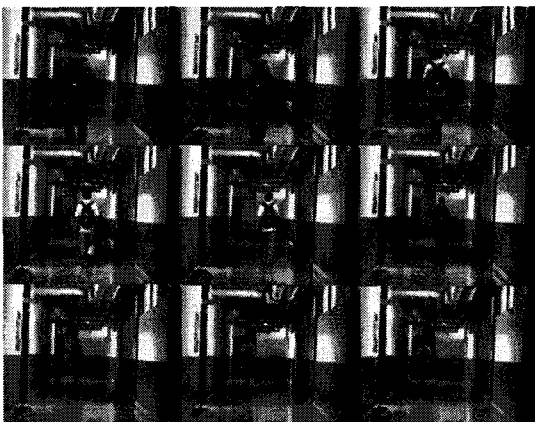


그림 7. A방으로 들어가는 영상

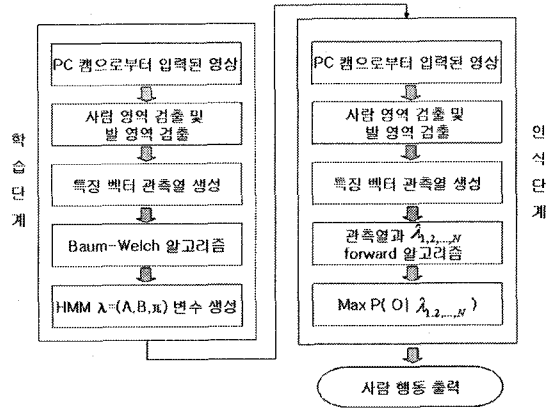


그림 8. 사람 행동 인식 단계도

식 (4)를 사용하여  $P(O|\lambda)$ 를 최대화하는 최적의  $\lambda$ 를 구하는 과정이다. 이는 다시 식 (9)에 의해서  $\lambda' = \lambda$ 가 될 때 즉  $P(O|\lambda)$ 의 극대화 값이 수렴할 때까지 반복해서 구해진다.

PC 캠으로부터 입력된 복도 영상과 미리 학습에 의해 구해진  $\lambda$ 들과의 확률 계산을 이용해 여러 행동 중 가장 확률적으로 높은 행동을 그 결과로 출력한다. 그림 8은 복도에서 사람 행동의 인식을 위한 알고리즘의 전체 흐름도를 나타낸다. 전체 시스템은 학습 단계와 인식단계로 나누어지고 학습단계에서는 운동 영역(발 영역)의 검출과 Baum-Welch 알고리즘을 이용한 HMM 모델의 학습을 표현한다. 인식단계에서는 카메라에서 입력된 사람의 발 영역을 감지하여 이의 관측열과 전향 알고리즘을 이용하여 학습된 행동패턴을 분류하는 것을 보여준다.

### 3. 실험결과

#### 3.1 학습과 인식

실험환경은 다음과 같다. Window-XP를 운영체제로 하는 Intel Pentium-4 1.8G 속도의 데스크톱 컴퓨터를 사용하였다. 초당 20프레임의 속도로 320×240 해상도를 갖는 PC 캠을 입력 장치로 사용하였으며 Matlab을 이용하여 프로그램을 작성하였다.

실험에는 각 행동, 즉 A, B방으로 각각 들어가고 나오는 행동, 서성이는 행동에 대한 5개의 HMM  $\lambda=(A,B,\pi)$ 변수와 PC 캠으로부터 입력된 복도 영상에 대한 관측열과의 확률 계산을 이용하여 특정 행동을 인식하였다.

HMM의 관측 가능한 기호 M을 (3×3), (4×4), (5×5)으로 가변해 가면서 실험을 하였으며, 상태 수 N을 2~5로 가변해 가면서 실험을 하였다.

HMM은 상태의 주요한 변화를 ‘마르코프 과정’으로 모델링한다. 상태 변화 시퀀스는 겉으로 드러나지 않으며(hidden), 직접 관측이 가능한 데이터를 통해서 간접적으로 알 수 있기 때문에 상태 수 N을 가변해 가면서 가장 확률이 높을 때의 상태 수를 구하게 되는 것이다. 표 2는 M이 9일 때 상태 수에 따른 인식률을 보여주고 표 3은 M이 16일 때, 표 4는 M이 25일 때 각각의 인식률을 보여준다. 그림 9, 10, 11은 각각의 실험에서 인식기의 후보 개수를 고려하였을 때의 결과를 나타낸 것이다. 이 그래프의 세로축은 인식률을, 가로축은 분류해야 할 행동 패턴의 수를 나타낸다.

이들 실험에서 HMM의 관측 가능한 기호 M을 16로 하고 상태 수 N을 3으로 했을 때 가장 인식률이 높았다.

그림 12는 상태수 N=3, 관측 가능한 기호 M=16으로 했을 때 HMM의 학습에 의한 우도값(Likelihood)의 그래프를 나타낸 것이다. 앞 절에 설명했듯이 HMM의 학습 알고리즘, Baum-welch 알고리즘은  $\lambda' = \lambda$  가 될 때 즉  $P(O|\lambda)$ 의 극대화 값이 수렴할 때까지 반복한다. 결과에서 보듯이 Baum-welch 알고리즘은 반복이 되면서 우도가 지역적 최적화가 되는

표 2. HMM 인식률 결과(M=9)

HMM의 관측 가능한 M=(3×3)				
상태수(N)	2	3	4	5
인식률(%)	32.82	64.87	67.53	58.25

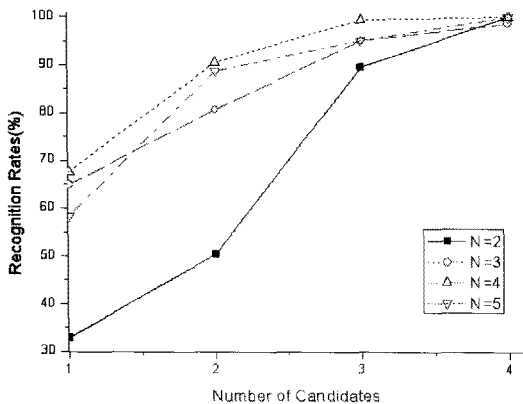


그림 9. 후보 수를 고려한 HMM 인식률 변화(M=9)

표 3. HMM 인식률 결과(M=16)

HMM의 관측 가능한 M=(4×4)				
상태수(N)	2	3	4	5
인식률(%)	56.78	89.13	68.54	60.83

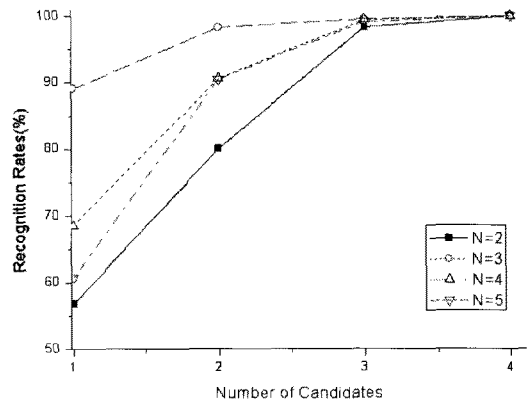


그림 10. 후보 수를 고려한 HMM 인식률 변화(M=16)

표 4. HMM 인식률 결과(M=25)

HMM의 관측 가능한 M=(5×5)				
상태수(N)	2	3	4	5
인식률(%)	40.65	69.34	62.61	53.91

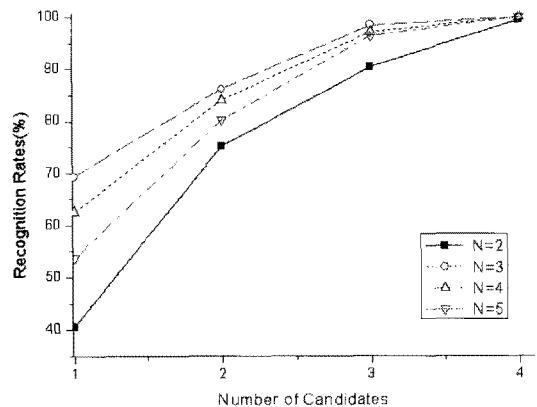


그림 11. 후보 수를 고려한 HMM 인식률 변화(M=25)

값으로 수렴하는 것을 볼 수 있다.

표 5는 PC 캠으로부터 입력된 영상 관측열에 대해서 각 행동에 대한 우도값의 결과를 나타낸 것이다. 표의 가로축 입력 행동 A, B, C, D, E는 각각 A방으로 들어가는 행동, B방으로 들어가는 행동, A방에서 나오는 행동, B방에서 나오는 행동, 서정이는

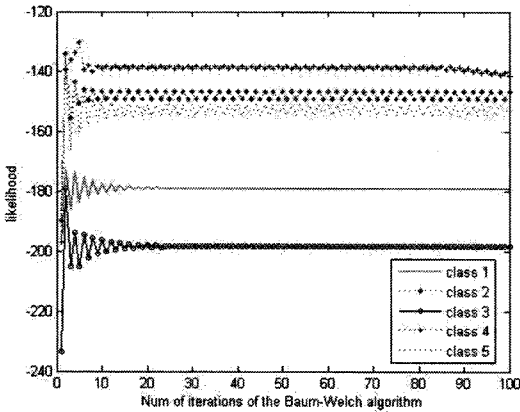


그림 12. HMM의 학습에 의한 우도

표 5. 각 행동에 대한 로그-우도의 결과

	A	B	C	D	E
A방Enter	-18.03	-486.15	-64.75	-134.94	-387.64
B방Enter	-164.27	-26.84	-157.27	-67.34	-351.27
A방Exit	-259.72	-315.78	-22.53	-351.54	-413.24
B방Exit	-419.25	-465.24	-297.64	-23.79	-369.45
서성입	-502.14	-432.16	-413.65	-511.47	-29.15

행동에 대한 관측열을 나타낸다. 그리고 세로축은 각 행동에 대해서 20개의 관측열을 HMM 학습의 결과로 구해진  $\lambda$  변수를 나타낸다. 각 행동의 관측열과 미리 학습에 의해서 구해진  $\lambda$  변수를 식 (5)에 대입하여 우도를 계산한다. 결과에서 알 수 있듯이 각각의 해당 행동이 다른 행동에 비해서 우도(확률값)가 가장 큰 수치를 보임을 알 수 있다. 이 결과를 이용해서 현재의 입력 영상이 어떤 행동인지를 출력하게 되는 것이다.

### 3.2 성능 비교

본 논문에서 제시한 방법의 적합성을 알아보기 위해 M. Brand 등[15]의 방법을 비교해 보았다. M. Brand 등의 연구에서는 사무실에서의 행동 중 사무실로 들어오고 나가고 하는 단순한 행동에 대한 인식을 위해 HMM을 사용하였다. 본 논문에서는 다양한 위치에서 각각 들어가고 나가고 하는 행동에 대해서 이의 인식방법을 HMM의 시변 관측열 인식특징을 도입하여 학습하고 분류해 내는 방법을 제시하였고, 이의 적용 예로써 복도에서 사무실로 입출과 특이한

행동을 하는 사람을 인식하는 것을 보였다. 특히, 지나가 버리지 않고 복도에서 서성이며 특이한 행동을 하는 사람을 HMM이 인식할 수 있다는 것을 보였다. 이는 복도에서 수상한 행동을 하는 사람을 발견하고 알려 범죄에 즉흥적으로 대응할 수 있음을 나타낸다.

M. Brand 등의 연구에서 특징 벡터 추출 방법은 전경(foreground) 픽셀에서 2D 가우시안의 고유벡터와 중심 값으로부터 계산되어진 8개의 변수를 사용하였다. 본 논문에서는 사람영역을 찾은 후 상, 하, 좌, 우 각각의 최 외각 픽셀을 경계로 하는 MBR을 추출하고 사각 박스 하단의 중심 픽셀 좌표를 사용하여, 16개 영역으로 나뉜진 코드북에서 좌표가 위치하는 영역을 찾아 특징 벡터로 사용하였다. 본 논문에서는 상태수  $N=3$ , 관측 가능한 기호  $M=16$  으로 했을 때 최고 89%의 인식률을 보였다.

## 4. 결 론

본 논문에서는 PC 캠으로부터 입력된 영상을 이용하여 복도에서의 사람 행동을 인식하는 방법을 제안하였다. 제안된 시스템은 영상 이미지에서 먼저 사람 영역을 찾고 그다음 발 영역을 찾아 특징 벡터인 관측열을 생성하고, HMM의 Baum-Welch 알고리즘으로 학습하여 각 행동에 대한 HMM의  $\lambda$  변수를 구하였다. 그 후 현재 입력되는 영상의 관측열을 구하여 각 행동에 대한  $\lambda$  변수와 전향 알고리즘을 사용하여 확률 계산을 하였으며 확률이 가장 높은 행동을 결과 행동으로 인식하도록 하였다. 실험을 통해서 HMM의 상태수  $N=3$ , 관측 가능한 기호  $M=16$ 일 때 인식률이 가장 높았고 이 설정을 알고리즘에 적용하여 현재 입력된 행동이 학습된 행동과 같은 행동에서 확률이 가장 높은 결과를 얻었다.

제안하는 알고리즘은 복도에서 수상한 행동을 하는 사람을 인식하여 사용자에게 알려 범죄 예방에 효과적으로 사용될 수 있음을 보였다.

본 논문에서는 지능형 영상 감시에 초점을 둔 기초 연구의 목적으로 5가지의 단순한 행동에 대해서만 학습하고 인식하였다. 따라서 추후에는 영상에서 사람의 팔, 다리, 머리, 몸통 부분을 구분하여 인식해 사람의 복잡하고 다양한 행동, 예를 들면 사람이 넘어지는 행동, 폭력적인 행동 등에 대해서 실험하고 연구할 계획이다.



참 고 문 헌

[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 Mellin 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, Aug. 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, No.2, pp. 257-286, 1989.

[5] J.Yamato, 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.

[6] A. Kale, A. N. Rajagopalan, A. Sundaresan, N. Countoor, A. N. RoyChowdhury, V. Krueger, and R. Chellappa, "Identification of humans using gait," *IEEE Transactions on Image Processing*, Sep. 2004.

[7] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger. "Gait-based recognition of humans using continuous HMMs," *Proceedings of the IEEE Conference in Automatic Face and Gesture Recognition*, pp. 321-326, 2002.

[8] J.J. Little and J.E. Boyd. "Recognizing people by their gait: the shape of motion," *Videre*, Vol.1, No.2, pp. 1-32, 1998.

[9] R. Collins, R. Gross, and J.Shi. "Silhouette-based human identification from body shape and gait," *IEEE Conf Automatic Face and Gesture Recognition*, pp. 351-356, 2002.

[10] C. Ben Abdelhader, R. Cutler, and L. Davis. "Motion-based recognition of people in eigengait space," *IEEE Conf Automatic Face and Gesture Recognition*, pp. 254-259, 2002.

[11] 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.

[12] 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.

[13] 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.

[14] A. Elgammal, D. Harwood, and L. Davis "Non-parametric Model for Background Subtraction," *6th European Conference on Computer Vision. Dublin, Ireland*, Vol.1843, pp. 751-767, 2000.

[15] M. Brand and V. Kettner, "Discovery and Segmentation of Activities in Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, No.8, pp. 844-851, Aug. 2000.



정 창 욱

2007년 부경대학교 제어자동화 공학과 학사  
 2007년~현재 부산대학교 지능기계공학 석사과정  
 관심분야 : 컴퓨터비전, 인공지능, 패턴인식



강 동 중

1988년 부산대학교 정밀공학과  
학사

1990년 KAIST 기계공학과 석사

1999년 KAIST 자동화및설계  
공학과 박사

1990년~1992년 현대전자 산업전  
자연연구소 주임연구원

1997년~1999년 삼성종합기술원 신호처리연구실 선임  
연구원

2000년~2005년 동명대학교 메카트로닉스공학과 조교수

2006년~현재 부산대학교 기계공학부 조교수

관심분야 : 머신비전, 패턴인식, 영상검사, 이동로봇