

# 모델 학습 방법을 이용한 사람 탐지 시스템의 구현

서경민<sup>o</sup> 이철우  
전남대학교 컴퓨터공학과  
{seokm, leecw}@image.chonnam.ac.kr

## Implementation of person tracking system using learning model method

Kyung-Min Seo<sup>o</sup> Chil-Woo Lee  
Computer Department, Chonnam National University

### 요약

본 논문은 입력 영상들로부터 추적이나 탐지의 대상이 되는 모델을 학습에 의해 생성하는 방법에 대해 기술한다. 일반적으로 탐지나 검출 시스템을 구성할 경우, 사용되는 대상 모델은 초기에 인위적으로 주어지게 된다. 이 경우 시스템이 동작하는 주위의 환경이 변하게 되면, 그에 맞게 새로운 대상 모델이 다시 주어져야 하는 단점이 있다. 또한 탐지 시스템 개발에 있어서 일반적인 문제점은 탐지 대상이 가려지거나 겹칠 경우 인식 성공률이 크게 떨어진다는 것이다. 본 논문에서는 사람 탐지 시스템의 일반적인 문제점들에 대응하고 탐지의 성능을 높이기 위하여 최소한의 제약 조건만이 미리 주어지고 실제 탐지 대상의 모델은 입력 영상으로부터 학습을 통해 구성 요소별로 생성하는 방법에 대해 기술한다.

### 1. 서론

본 논문에서는 탐지의 대상이 되는 사람을 입력 영상으로부터 학습하여 모델링한 후, 새로운 영상들로부터 탐지하는 시스템에 대하여 기술한다. 개발 시스템의 보다 구체적인 범위는 고정된 카메라를 이용하여 실내에서 다수의 사람을 동시에 탐지하며 그 결과물로 탐지 대상의 위치를 출력하는 것이다. 사람 탐지 시스템 개발에 있어서 해결하여야 할 일반적인 문제점들은 탐지 대상인 사람이 여러 형태로 변형이 가능하고, 각 부분들끼리 겹칠 수도 있으며 다양한 물체에 의해 탐지 대상이 가려질 수 있다는 것이다[1][2]. 기존의 사람 탐지 시스템으로는 크게 외형 기반의 방법[1], 시공간 정보 기반의 방법[3], 블랍(blob)의 통계적 특징을 이용한 방법[4] 등이 있다. 기존 방법은 모델을 생성할 때와 모델을 탐지할 때의 환경이 다르면 탐지의 성능이 떨어지게 되는 단점이 있다. 본 논문에서는 사람 탐지 시스템의 일반적인 문제점들에 대응하고, 탐지의 성능을 높이기 위하여 탐지가 이루어지는 시점에 탐지 대상물 구성 요소별로 학습하는 방법을 제안한다. 제안된 방법은 탐지 대상이 되는 물체에 관하여 환경 불변적인 기본 모델만을 가지고 있으며, 실질적인 모델은 탐지 시스템 동작 중에 학습에 의해 생성하고 갱신된다는 점에서 기존 방법과 차별화 된다. 제안된 방법을 이용한 전체 시스템의 구성은 1) 전경과 배경의 분리, 2) 사람 모델의 학습, 3) 사람 탐지, 4) 모델 갱신의 4단계로 이루어지며, 각각의 내용은 2장에서 자세히 설명된다.

### 2. 본론

본 논문에서 제안한 방법을 이용한 사람 탐지 시스템의 전체 구성은 그림 1과 같다. 전경과 배경의 분리 단계는 고정된 카메라로부터 입력되는 영상들을 미리 생성된 배경 모델과 비교하여 사람일 가능성이 상대적으로 높은 전경 부분만을 분리하는 단계이다. 사람 모델의 생성 단계는 전경 영역 중 탐지의 대상인 사람의 모델을 미리 주어진 기본 모델과 여러 제약 조건들을 이용하여 확률적으로 생성하는 단계이다. 생성되는 모델은 사람의 머리, 팔, 다리로 각각 구성이 된다. 사람 탐지 단계는 전경 내에서 탐지의 대상인 사람을 각 부분별로 먼저 탐

색하고, 각 부분별 탐색 결과를 기하학적인 위치제약조건을 이용하여 완전한 사람으로 탐지하는 단계이다. 모델 갱신 단계는 새로 생성된 모델과, 탐지의 결과를 이용하여 이전에 생성된 배경, 사람 모델에 새로운 정보를 반영하여 주는 단계이다.

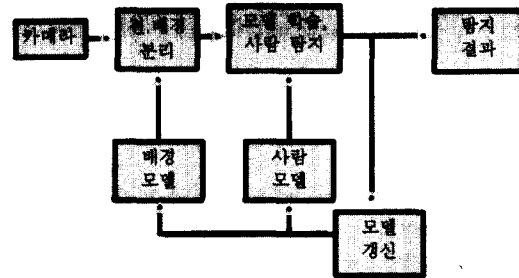


그림 1. 전체 구성

#### 2.1 전경과 배경의 분리

입력되는 영상은 전경과 배경으로 분리된 후 사람일 가능성이 높은 전경 영역에 대해서만 처리가 이루어지도록 하였다. 영상을 전경과 배경으로 분리하기 위해서는 전경이 없는 영상만으로 배경 모델을 생성한 후 새로운 영상과 배경 모델과의 차이를 전경으로 판단할 수 있다. 전경과 배경을 실제 영상을 이용하여 분리할 때는 배경 모델의 생성 방법, 배경 모델의 갱신, 배경 모델과 새로운 영상의 차이를 계산하는 방법 등을 고려하여야 한다. 본 논문에서는 영상 입력의 초기에는 전경이 포함되어 있지 않다고 가정하고, 입력 초기 영상들의 평균 영상으로 배경 모델을 생성한 후, 처리 도중 전경이 검출되지 않는 입력 영상을 추가로 더해 주는 방식을 이용하였다. 배경 모델과 새로운 영상의 차이는 영상의 밝기에 상관없는 전경과 배경의 분리를 위해 식 (1)과 같은 Yuji[5]의 함수를 이용하여 계산하였다.

$$f(a,b) = 1 - \frac{2\sqrt{(a+1)(b+1)}}{(a+1)+(b+1)} \cdot \frac{2\sqrt{(256-a)(256-b)}}{(256-a)+(256-b)}$$

$$(0 \leq f(a,b) < 1)$$

$$(0 \leq a(x,y), b(x,y) \leq 255)$$
(1)

여기서,  $f(a,b)$ 는 영상  $a, b$ 간의 차이,  
 $a(x,y), b(x,y)$ 는 각 영상의 밝기,  
 $(x,y)$ 는 영상 내 좌표.

### 2.2 후보 영역 추출

입력 영상으로부터 추출된 전경 영역은 사람에 의해 만들어진 전경 영역 외에도 노이즈, 사람이 아닌 물체의 이동 등을 포함하고 있다. 먼저 노이즈를 제거하기 위해 기본적인 morphological 연산인 erosion과 dilation을 전경 영역에 몇 차례 수행하여 주었다. 노이즈가 제거된 전경 영역은 크기 정규화 후, 그림 2처럼 미리 주어진 사람 모델과의 template matching을 통하여 사람의 후보 영역만을 추출하였다. 그림 2의 왼쪽 그림은 사람의 일반적인 정면 모습이고, 오른쪽 그림은 옆에서 본 사람의 걷는 모습이다. 전경 영역들 중 시공간에서 근접한 영역들은 동일한 사람에 의한 전경 영역이라 보고, 모델 생성 과정에서 보다 다양한 후보 영역들을 이용하도록 하기 위하여 후보 영역에서 제거하였다. 또한 전경 영역과 사람 모델과의 크기를 일치시키는 정규화 과정에서 사용되는 비례 계수(scale factor)값들의 평균과 분산을 함께 구하여 모델 탐지 단계에서 탐지 대상의 크기를 역으로 제한하는데 이용하도록 하였다.



그림 2. 사람 모델

### 2.3 모델 학습

전 단계에서 얻어진 사람의 후보 영역들은 그림 3과 같이 머리, 팔, 다리로 나누어진 후 각각의 부분별로 모델을 생성하게 된다. 모델을 생성하기 위해 사용될 특징은 색상 정보, 에지 정보, texture등 여러 가지가 사용될 수 있지만 사람의 경우 입고 있는 옷이 다양한 색상을 가질 수 있으므로 본 논문에서는 색상과 관계없는 에지 정보를 모델의 특징으로 사용하였다.

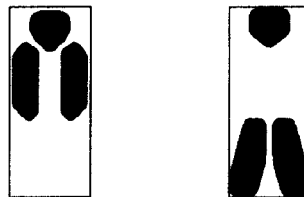


그림 3. 부분 모델

$n$ 개의 입력 영상으로부터 검출된 각각의 부분별 후보 영역 내의 위치  $(x, y)$ 의 에지 정보들을  $i_n(x, y)$ 라 하면 얻고자 하는 탐지 대상의 모델은  $i_n(x, y)$ 의 대표치로서 나타낼 수 있다. 사람의 경우 앞모습과 뒷모습, 왼쪽에서 본 모습과 오른쪽에서 본 모습이 각각 다르므로  $i_n(x, y)$ 는 k-means 알고리즘[6]을 이용하여 2개의 집단으로 군집화(clustering)되도록 하였다. 군집화 후 각 집단의 분산을 계산하여 분산이 미리 주어진 임계치 이상으로 클 경우의 모델 정보는 의미가 없는 것으로 판단하여, 탐지 단계에서 모델 매칭을 계산하는 과정에 영향을 주지 않도록 하였다. k-means 알고리즘을 이용한 모델 생성의 과정은 다음과 같다.

- (1)  $i_n(x, y)$ 의 값 중 발생 빈도수가 가장 높은 2개의 값을 k-means 알고리즘의 초기 값으로 준다.
- (2) 남아있는  $i_n(x, y)$ 값들은 초기 값과의 거리를 계산하여 가까운 쪽의 군집으로 포함되도록 한다.
- (3) 군집의 중심을 다시 계산한다.
- (4) 모든 군집에 대해 변화가 없으면 계산을 종료하고, 그렇지 않으면 (2)단계로 간다.

### 2.4 모델 탐지

모델 탐지는 전 단계에서 구하여진 사람 모델을 새로운 입력 영상의 전경 영역에서 찾는 단계이다. 입력 영상 내에서 사람은 다양한 크기로 존재할 수 있으므로, 탐색에 사용될 사람 모델은 '2.2 후보 영역 추출' 단계에서 구한 비례 계수(scale factor)와 다음 식 (2)를 이용하여  $n$ 개의 다양한 크기로 변형된 후 탐지에 이용된다.

$$\text{탐지모델}_n = \text{사람모델} \times (\mu_{\sigma} \pm \alpha_n)$$

$$\alpha_n \leq c \times \sigma_{\sigma} \quad (c = \text{상수})$$
(2)

여기서,  $\mu_{\sigma}$ 는 scale factor의 평균값,  
 $\sigma_{\sigma}$ 는 scale factor의 분산.

입력 영상 내의 전경 영역은 식 (3)의 상관값(correlation)계산을 이용 사람 모델의 각 부분과의 상관도를 계산하여 임계치 이상의 상관도를 가지면 대응되는 머리, 팔, 다리 등으로 판정이 나게 된다. 현재 처리 중인 block을  $b_1, b_2$ 라 하면, block  $b_1$ 과  $b_2$ 간의 상관값은 다음 식에 의하여 계산된다.

$$\rho_{b_1, b_2} = \frac{\sigma_{b_1, b_2}}{\sigma_{b_1} \sigma_{b_2}}$$

$$= E \left[ \left( \frac{b_1 - \mu_{b_1}}{\sigma_{b_1}} \right) \left( \frac{b_2 - \mu_{b_2}}{\sigma_{b_2}} \right) \right]$$
(3)

여기서,  $\rho_{b_1, b_2}$ 는  $b_1$ 과  $b_2$ 간의 상관값,  
 $\mu_{b_n}$ 는 block  $b_n$ 의 평균값,  
 $\sigma_{b_n}$ 는 block  $b_n$ 내의 분산.

찾아진 머리, 팔, 다리의 후보 영역은 신체 각 부분의 기하학적인 위치 제약을 적용하여 실제로 가능한 경우들만을 최종적인 사람으로 평가하였다.

### 2.5. 모델 갱신

본 논문에서 제안된 사람 탐지 시스템에 사용되는 배경 모델

과 사람 모델을 각각 일정한 주기로 갱신이 되어 항상 최신의 정보를 반영하도록 하였다. 입력 영상은 전경과 배경으로 구분된 후 표1.과 같은 조건으로 배경 모델을 갱신하도록 하였다.

표 1. 배경 갱신 조건

배경 \ 입력	전경		배경
	이동중	고정	
갱신	×	○	○

전경과 배경을 분리하여 배경으로 분리된 영역은 모델 갱신 주기마다 갱신을 하여 주고, 전경으로 분리된 영역은 각 영역의 중심을 전 프레임과 비교하여 영역의 움직임을 분석한 후, 갱신 주기동안 움직이지 않은 고정 영역만을 갱신하도록 함으로써, 사람을 포함하고 있는 전경 영역은 계속 전경으로 유지하고, 사물의 움직임에 의해 잘못 인식된 전경 영역은 배경으로 갱신하도록 하였다.

### 3. 실험

#### 3. 실험 환경

제안된 방법은 Pentium III 500Mhz PC와 Visual C++ 6.0을 사용하여 구현되었다. 입력은 연구실 내부를 다양한 조건에서 촬영한 영상을 이용하였으며, 영상은 320x240 크기의 8bit 그레이 영상으로 통일하였다.

#### 3.1 실험 결과

그림 4.는 입력 영상으로부터 전경만을 분리한 후, 사람의 후보 영역을 추출하여 그 윤곽선 정보를 구한 그림이다. 상단 왼쪽 영상이 입력 영상이며, 하단 왼쪽 영상은 후보 영역의 위치와 범위를 나타낸다. 중간과 오른쪽 영상은 후보 영역 내의 영상과 윤곽선 정보를 각각 나타낸다. 전경과 배경의 차이를 결정하는 임계치와 erosion과 dilation의 적용 횟수등 기본적인 설정과 임계값은 반복되는 실험 과정에서 경험적으로 적용되었다. 다음 그림 5는 전 단계에서 구한 후보 영역들을 이용하여 사람 모델을 생성한 결과이다. 각각 머리, 팔, 다리로 구분이 되며, 팔과 다리의 중간에 보이는 수평 윤곽선은 입력 영상 내에 등장하는 사람들의 반팔 옷과 반바지의 경계면에 의해 생긴 것들이다. 그림 6은 입력 영상으로부터 구한 사람의 부분별 모델을 이용하여 사람을 탐지하는 영상이다. 모델 생성에 사용된 영상과 동일한 조건에서 촬영된 다른 영상을 이용하여 사람을 탐지하였으며, 각각 입력 영상, 모델 영상, 윤곽선 영상에서의 탐지 결과, 입력 영상에서의 탐지 결과를 나타낸다.

### 4. 결론

본 논문에서 제안한 방법은 탐지의 대상이 되는 모델을 시스템이 동작하는 환경에 맞추어 학습함으로써 환경의 변화에 덜 민감한 탐지 시스템 개발에 이용할 수 있다. 또한 탐지의 대상인 사람을 각 부분별로 학습함으로써 다른 물체에 의해 사람이 부분적으로 가려지는 경우에도 탐지를 계속할 수 있다. 그러나 학습이 이루어지는 영상 입력 초기에는 탐지의 정확도가 아주 낮게 떨어지며, 모델 학습을 주기적으로 반복해 줄 경우, 계산량이 많아져 실시간 시스템에 적용할 수 없는 부분들은 개선되어야 할 부분이라 보여진다. 차후에는 학습 과정을 개선하여 실시간 감시 시스템에 적용할 수 있도록 하고, 시계열 정보를 함께 이용한 모델 탐지 시스템을 개발할 예정이다.



그림 4. 전경 분리와 후보 영역 추출

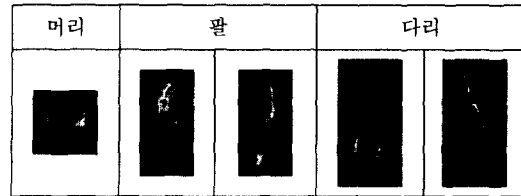


그림 5. 생성된 모델



그림 6. 사람 탐지 결과

### 【 참고문헌 】

[1] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-Based Object Detection in Images by Components", PAMI, Vol. 23, NO. 4, pp. 349-361, APRIL 2001.  
 [2] D. M. Gavrila, "The Visual Analysis of Human Movement: A Survey", Computer Vision and Image Understanding, Vol. 73, No. 1, January, pp. 82-98, 1999.  
 [3] S. Niyogi and E. Adelson, "Analyzing and recognizing walking figures in XYT", CVPR, 1994, pp. 469-474  
 [4] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body", PAMI, Vol. 19, No. 7, pp. 780-785, 1997.  
 [5] Yuji KUNO, Takahiro WATANABE, "Automated Detection of Human for Visual Surveillance System", ICPR'96, 865-869, 1996  
 [6] Earl Gose, Richard Johnsonbaugh, and Steve Jost, "Pattern Recognition and Image Analysis", Prentice Hall, pp.213-215, 1996.