

# 혼잡한 환경에서 적응적 가우시안 혼합 모델을 이용한 배경의 학습 및 객체 검출

|         |
|---------|
| 論 文     |
| 55D-2-2 |

## Adaptive Gaussian Mixture Learning for High Traffic Region

朴大龍<sup>†</sup> · 金裁敏<sup>\*</sup> · 趙成元<sup>\*\*</sup>  
(Dae-Yong Park · Jae-Min Kim · Seong-Won Cho)

**Abstract** - For the detection of moving objects, background subtraction methods are widely used. An adaptive Gaussian mixture model combined with probabilistic learning is one of the most popular methods for the real-time update of the complex and dynamic background. However, probabilistic learning approach does not work well in high traffic regions. In this paper, we propose a reliable learning method of complex and dynamic backgrounds in high traffic regions.

**Key Words** : GMM, Background Modeling, Motion Detection, Background Subtraction, High Traffic Region

### 1. 서 론

최근 일어난 여러 테러 사건들로 국내외적으로 사회 안전망 구축에 많은 관심이 고조되고 있어, 보안·감시시장이 급속하게 성장되고 있다. 이에 따라 영상을 효과적으로 처리하여 물체를 효율적으로 감시하는 영상 감시 시스템에 대한 연구가 활기를 띠고 있다[1,2,3]. 그림 1은 이러한 영상 감시 시스템의 일반적인 구조를 보여준다.

영상 감시 시스템뿐만 아니라 실시간으로 비디오 영상을 이용하는 시스템에 있어서 배경과 물체를 정확하고 효과적으로 분리해 내는 것은 기본적으로 매우 중요한 일이다. 물체를 정확히 분리해내지 못한다면 그 다음 단계인 물체의 분류와 행동 분석은 매우 힘든 작업이 된다.

#### 1.1 적응적 배경 차분을 이용한 물체 검출(adaptive background subtraction)

물체를 분리해내기 위해서는 먼저 배경을 정확하게 모델링하는 것이 필요하다. 영상의 각 화소에 대해 배경일 경우 어떠한 값을 갖게 되는지 기억한 후, 새롭게 들어오는 입력영상의 각 화소의 색 정보를 기억된 배경에 대한 정보와 비교함으로써 입력 영상에 나타난 물체의 영역을 찾아내게 되는 것이다. 이 때 배경은 시간이 지남에 따라 변화하게 된다. 예를 들어 구름이 지나가면 햇빛의 밝기가 변하고 그에 따라 배경의 값들도 변한다. 또한 어떤 경우에는 배경에 주차된 차가 이동하기도 한다. 따라서 배경 모델은 실제 배경을 정확히

나타내도록 시간에 따라 지속적으로 학습되어야 한다. 그렇지 않다면 나타난 물체를 정확히 검출해낼 수 없다. 이렇게 시간에 따라 변화하는 배경을 모델링하여 배경과 배경에 등장한 물체를 구분해 내는 방법을 적응적 배경 차분이라 한다. 이 방법은 장면 차분(frame differencing)[4]이나 물체가 없는 배경 상태를 설정하여 이미 학습된 배경과의 차를 구하는 배경 차분(background subtraction)과 달리 물체의 움직임이 없거나 배경이 변화하는 경우에도 물체의 전체 영역을 찾아낼 수 있다.

#### 1.2 적응적 가우시안 혼합 모델 (adaptive gaussian mixtures)

기본적으로 배경의 각 화소들은 하나의 색정보를 갖고 있으며 영상 취득 장치의 잡음에 따라 약간의 변화를 보인다. 이 값은 조명의 변화에 따라 점차 변화하기도 하고, 환경의 변화에 따라 전혀 새로운 값으로 변화하기도 한다. 이러한 배경의 값은 적응적 가우시안 확률 분포로 효과적으로 모델링할 수 있으며, 가우시안의 평균과 분산은 매 장면마다 새로 입력되는 값을 학습함으로써 환경의 변화에 적응해간다. 적응적 배경 차분의 방법 중의 하나인 적응적 가우시안 혼합 모델은 영상의 각 화소의 색 정보들을 단지 하나의 가우시안 분포가 아닌, 여러 개의 가우시안의 혼합 분포를 이용해 확률적으로 모델링 한다[5]. 이 방법의 장점은 배경이 조명이나 날씨에 따라 변하는 경우나, 배경에 새로운 물체가 첨가되거나 기존 물체가 사라지는 경우에도 이러한 배경의 변화를 효과적으로 학습한다는 것이다. 또한 이 방법은 단순히 각 화소의 값을 평균과 문턱값으로 모델링하는 방법과 비교해 볼 때, 확률적 분포를 사용하기 때문에 베이시안 결정이론 등의 확률적 접근이 용이하다. 어떤 경우에는 배경이 바람에 흔들리는 나뭇가지나 물이 솟아나는 분수를 포함하기도 한다. 이러한 경우 나뭇가지와 분수 물의 움직임 또한 물체로 인식되어 물체를 검출해내는 것을 방해한다. 이러한 경우에도 적응

<sup>†</sup> 교신저자, 正會員 : 弘益大學 電氣情報制御工學科 碩士課程  
E-mail : tkc-tmhk@hanmail.net

<sup>\*</sup> 正會員 : 弘益大學 電氣情報制御工學科 副教授 · 工博

<sup>\*\*</sup> 正會員 : 弘益大學 電氣情報制御工學科 教授 · 工博

接受日字 : 2005年 9月 26日

最終完了 : 2005年 12月 17日

적 가우시안 혼합 모델은 나뭇가지의 흔들림 등의 자주 일어나는 변화를 확률적 분포를 이용해 배경으로 모델링할 수 있고, 따라서 이러한 배경에서도 나타난 물체의 영역을 찾아낼 수 있다. 이러한 이유 때문에 가우시안 혼합 모델은 많은 실시간 영상 감시 시스템에 사용되어 왔다[5,6,7]. 그림 2에는 한 화소에 대하여 바람에 따라 보였다 사라지는 서로 다른 세가지 색정보가 가우시안 혼합 모델로 학습되었음을 확인할 수 있다.

### 1.3 적응적 가우시안 혼합 모델의 학습 방법과 혼잡한 환경에서 나타나는 문제점

적응적 가우시안 혼합 모델은 영상의 각 화소의 색 정보를 K개의 가우시안을 혼합한 분포로 모델링한다. Stauffer등은 이 모델들을 online approximated EM algorithm에 따라 학습하였다[5]. 가우시안 혼합 모델의 학습(gaussian mixtures learning)의 기본적인 개념은 다음의 recursive filter식을 통해 알 수 있다.

$$\alpha(t) = (1 - \gamma(t)) \cdot \alpha(t-1) + \gamma(t) \cdot \nabla(K(t); \alpha(t-1)) \quad (1)$$

t에서의 모델  $\alpha(t)$ 는  $\alpha(t-1)$ 과 예측된 변화량  $\nabla(K(t); \alpha(t-1))$ 의  $\gamma(t)$ 에 따른 비율적인 합을 계산하여 변화하는 입력  $K(t)$ 에 적응하게 된다.  $K(t)$ 는 시간 t에 특정 영상 좌표로부터 입력된 값이다. 각 가우시안 분포들의 평균과 분산, 혼합 비율은 위 식에 따라 학습되며 이를 통해 배경의 색 정보가 확률적으로 모델링된다. K개의 가우시안들은 시간에 따라 입력되는 색 정보가 자신의 확률 분포에 얼마나 해당되는 지에 따라 학습되며 이러한 과정을 거쳐 각 가우시안들은 점차 배경의 색 정보가 가지는 확률적 분포를 정확히 나타내게 되고 또한 배경의 변화를 적응해나간다. 이때 배경 모델의 수렴 속도와 안정도는  $\gamma(t)$ 에 따라 결정된다.

만약 현재 학습된 분포에 해당되지 않은 입력 값이 들어 오면 새로운 가우시안 분포를 만들고 각 화소에서 정해진 가우시안 수의 한계치인 K가 넘는 경우에만 이미 존재하던 가우시안 중에서 혼합 비율이 가장 낮은 것을 제거한다. 이런 과정을 통해 혼합 비율이 높은 K개의 가우시안이 혼합 모델을 만들고 학습되어간다.

이러한 가우시안 혼합 모델이 현재의 배경 변화를 적응해나가는 학습 속도  $\gamma(t)$  조절에 관련된 많은 연구가 있었다[5,8,9]. 학습의 속도가 너무 느린 경우, 물체 검출 시스템은 실제 배경이 변화할 때 배경에 대한 모델이 이를 반영하지 못하므로 배경의 변화를 물체로 인식하게 된다. 또한 학습의 속도가 너무 빠르면 천천히 이동하는 물체들을 배경으로 학습하게 되기 때문에 잘못된 배경 모델을 가지게 되고, 그 결과 실제 배경의 값이 입력되는 경우에도 물체로 인식하게 된다. 따라서 학습 영역이 조명이나 새롭게 위치한 물체로 인하여 기존의 배경과는 다른 새로운 배경이 되는 경우는 수렴 속도를 높여 빠르게 새로운 배경의 값을 학습하고, 새로운 배경이 아니라 물체가 지나가는 경우에는 입력의 영향을 최소화함으로써 과거에 학습한 배경 정보를 안정적으로 유지하는 것이 필요하다. 그러나 그동안 이에 대한 명확한 해결책이 제시되지 않았다. 이러한 문제를 해결하기 위해 최근 안정도

를 유지하며 수렴 속도를 향상시키는 학습 방법이 보고되었다.[10] Lee는 이 논문에서 그 동안의 여러 방법들을 분석하고 각 가우시안에 동일하게 적용했던 고정된 학습 비율을 각 가우시안에 맞게 적절하게 조절해 나가는 방법을 제시하였다. 그 결과 안정도를 유지하면서 수렴 속도를 성공적으로 높였고 이에 따라 물체를 구분해내는 정확도도 향상되었다. 그러나 물체가 자주 나타나는 상황에서는 이러한 방법도 한계를 드러낸다.

$$\alpha(x, y) = \sum_{k=1}^K w_k \cdot g(x, y; \mu_k, \sigma_k) \quad (2)$$

이러한 배경 학습방법은 배경과 물체를 구분하지 않은 상태로 학습을 시작하지만 일정 시간 동안 학습을 하면 물체의 움직임에 영향을 받지 않고 배경을 효과적으로 모델링하고 그 변화를 학습해 나갈 수 있다. 이는 확률적인 학습 과정을 통해 현재 입력이 물체에서 입력되는 색정보인지 또는 배경에서 입력되는 색정보인지를 구분하기 때문이다. 식(2)를 보면 어떠한 시간 t에서 화소(x,y)의 학습된 모델은 K개의 가우시안들의  $w_k$ 에 따른 합으로 표현된다. 입력되는 색정보들을 K개의 가우시안으로 모델링하는데, 각각의 가우시안은 배경의 색분포를 학습한 가우시안일 수도 있고 물체의 색분포를 학습한 가우시안일 수도 있다. 이 때 물체의 색정보를 학습한 가우시안은 배경의 경우에 비해 물체의 색정보가 입력되는 횟수가 상대적으로 작기 때문에 작은  $w_k$ 값을 가지고, 또한 서로 다른 물체의 색 정보는 대부분 그 값의 차이가 커서 하나의 가우시안을 이루지 못하거나, 이룬다 해도 큰  $\sigma_k$  값을 가지게 된다. 이와 반대로 배경을 학습한 가우시안은 상대적으로 큰  $w_k$ 과 작은  $\sigma_k$ 을 가지게 된다. 이를 이용해 배경을 학습한 가우시안을 유지하며 그 변화를 학습함으로써 배경 모델을 구성한다. 대부분의 경우 이러한 방법으로 배경을 모델링하고 물체를 검출해낼 수 있다. 하지만 통행량이 많은 영역에서는 상대적으로 물체의 색정보가 입력되는 양이 높아지게 되고 따라서 물체의 색정보를 학습한 가우시안의  $w_k$ 가 높아지게 된다. 또한 어떤 경우에는 자주 나타나는 물체들이 서로 비슷한 색을 가짐으로 물체들의 색 분포를 학습한 가우시안의  $\sigma_k$ 가 낮아질 수 있다. 그 결과 물체의 색 정보를 학습한 가우시안과 배경의 색정보를 학습한 가우시안의 구분이 모호해지고, 물체의 색정보가 배경 모델에 영향을 끼치게 된다. 이는 물체 검출의 정확도를 저하시킨다. 실제 감시가 필요한 공항이나 지하철, 도로 등에서 사람과 자동차 등의 물체가 자주 다닌다는 점을 고려할 때, 이러한 문제점은 가볍게 볼 수 없다. 따라서 이러한 환경에서 배경의 색 정보만을 학습하면서 가우시안 혼합 모델이 배경을 정확히 나타내도록 유지하고, 또한 배경의 시간에 따른 변화를 신속하고 안정적으로 학습해나가는 방법이 필요하다.

### 1.4 혼잡한 환경에서 가우시안 혼합 모델을 이용한 물체 검출에 대한 기존 연구

이러한 문제에 대한 명확한 해결은 제시되지 않았지만 이를 개선한 몇몇 연구 결과들이 있다. 물체와 배경의 차이를

생각해볼 때 이 문제를 해결하는 가장 간단한 방법은 스테레오 비디오를 이용해 각 화소의 깊이를 구하는 것이다[11]. 배경은 당연히 물체보다 뒤에 위치하게 되기 때문에 동일한 화소에서 얻어진 서로 다른 색 정보 중 어느 하나가 깊이값이 크다면 이를 배경의 색 정보로 인식하면 된다. Harville등은 [6]에서 이러한 원리를 이용하여 스테레오 비디오를 통해 얻은 각 화소의 깊이에 따라 현재 입력이 배경을 나타내는 가우시안에 해당되는 값인지 물체를 나타내는 가우시안에 해당되는 값인지를 구분하여 배경 모델을 학습하였다. 그러나 일반적인 스테레오 시스템이 동일 화소에 대하여 빛의 각도에 따른 색 정보의 변화와 가림(occlusion)현상으로 각 화소의 깊이값을 얻는데 어려움을 겪듯이 이 방법도 각 화소의 깊이값에 대한 신뢰도가 문제가 되었다. 그래서 이들은 깊이값에 대한 신뢰도를 계산하였고 그 값이 문턱값 이상인 경우만을 사용함으로써 물체 검출의 정확도를 개선하였다. 그러나 깊이값과 그 신뢰도를 구하는데 있어 많은 연산량을 요구한다. 그리고 이 방법은 근본적으로 카메라를 두 대 이상 사용하여야 한다.

이들은 이러한 깊이 값을 비교하는 방법 외에도 activity-based learning modulation을 제안하였다. 식(3)은 영상의 화소(x,y)에서 시간 t에 따라 activity level 값 A가 학습되는 방법을 보여준다. Y는 영상의 입력을 나타낸다.

$$A_{x,y,t} = (1-\lambda)A_{x,y,t-1} + \lambda|Y_{x,y,t} - Y_{x,y,t-1}| \quad (3)$$

Activity-based learning modulation은 activity level 값 A가 미리 설정한 상한 값 H보다 높아지면, 해당 화소에 해당하는 가우시안들의 새로운 입력 Y를 학습하는 비율들을 일괄적으로 감소시킨다. 복잡한 환경에서 이를 이용하면 자주 나타나는 물체의 색정보가 배경 모델에 미치는 영향을 줄임으로, 이미 학습한 배경에 대한 올바른 정보를 오래 유지할 수 있다. 이 방법은 배경에 따른 색 정보와 물체에 따른 색 정보를 구분하지 않고 단지 물체가 자주 나타나는 영역에 대하여 일괄적으로 학습 비율을 낮춤으로 학습된 배경 모델의 안정도를 높인 것이다. 그러나 그 결과 activity level값이 문턱값 H 이상으로 증가한 영역들은 환경의 변화에 적응하는 속도가 느려진다. 또한 물체가 자주 지나가는 환경이 지속된다면 물체를 낮은 학습 비율로 계속 학습해 나갈 것이고 이러한 학습의 결과 배경 모델은 잘못된 분포를 가지게 된다. 더욱이 일괄적으로 낮아진 학습 비율은 잘못된 학습의 결과를 오랜 시간 동안 유지시키고 정확도 저하에 지속적인 영향을 준다.

이에 반하여 Lee는 한 화소를 모델링하는 각 가우시안들의 학습 속도를 개별적으로 조절하는 방법을 제시하였다. 이 방법은 기존 방법[5]에 비해 실제 영상의 확률 분포에 수렴하는 속도를 크게 높였고 안정도 또한 성공적으로 유지하였다. 그의 논문은 복잡한 상황에서의 배경 학습을 고려하지는 않았지만 그가 제안한 알고리즘은 혼잡한 배경에서도 기존 방법보다 우수한 성능을 보인다. 기존 접근 방법들은 동일 화소들의 가우시안들에 대하여 동일한 학습 비율을 적용하였고 이것들의 학습 비율을 동일하게 낮춤으로 학습된 배경에 대한 안정도를 높이려 했다. 그러나 Lee는 동일할 화소를 나타내는 각 가우시안의 학습 비율을 개별적으로 조절하였다.

$$q_k = \frac{p_k}{\sum_{j=1}^K p_j} \quad (4)$$

$$c_k = c_{k-1} + q_k \quad (5)$$

$$\eta_k = q_k \left( \frac{1-a}{c_k} + a \right) \quad (6)$$

이는 각 가우시안이 현재 입력을 얼마나 잘 표현하느냐에 따른 상대적인 확률값  $q_k$ 와 그동안 가우시안이 입력의 분포를 얼마나 잘 모델링해왔는가를 나타내는 상대적인 양  $c_k$ 를 사용해 각 가우시안의 학습률  $\eta_k$ 를 변화시킴으로 행해진다. K는 각 화소가 가질 수 있는 최대 가우시안 수를 나타내며 k는 각각의 가우시안을 표시한다.  $p_k$ 는 각 가우시안 분포에 따라 구한 현재 입력 색정보의 확률이며,  $a$ 는 고정된 학습 비율이다.

$c_k$ 는 새롭게 생성된 가우시안 분포에 해당하는 입력이 많이 들어올수록 그 가우시안의 매개 변수(평균값,분산)의 학습 비율을 낮춘다. 따라서 많은 입력 값들로부터 지지를 받는 가우시안의 경우에는 점차 안정되게 되지만, 그렇지 못한 가우시안의 경우 그 가우시안 분포의 매개 변수들은 안정되지 못하고 높은 학습 비율에 따라 평균값의 이동과 큰 분산을 가지게 된다. 이는 새롭게 나타난 물체에 대해 빠르게 가우시안 모델을 만들도록 해주지만 만든 가우시안 분포에 해당하는 입력이 적은 경우에는 빨리 사라지도록 해준다. 결국 배경을 나타내는 가우시안은 상대적으로 천천히 학습하며 그 분포를 유지하지만 물체를 나타내는 가우시안은 빠르게 생겨나고 사라짐을 반복하게 된다. 따라서 비록 혼잡한 환경이라 할지라도 나타나는 물체들의 색 분포의 차가 크고 이러한 물체들의 지나가는 속도가 빠르다면 Lee의 알고리즘은 상대적으로 배경을 정확하게 모델링할 수 있다. 그러나 물체들의 색, 분포의 차가 크지 않고 이러한 물체들이 비교적 천천히 이동하는 경우에는 잘못 학습된 가우시안들의 혼합비율이 높아지게 되고 이는 물체 검출의 정확도를 떨어트린다.

### 1.5 복잡한 환경에서 학습 비율을 조정하는 학습 방법의 한계

위에서 살펴본 것과 같이 단순히 학습 비율을 조절해서는 좋은 결과를 얻을 수 없다. 이러한 학습 방법들의 기본 형태는 식 (1)과 같다. 식(1)에서  $I(t)$ 의 변화는 다음과 같이 나타낼 수 있다.

$$I(t) = I(t-1) + I_{noise}(t) + I_{environmental\ change}(t) + I_{object}(t) \quad (7)$$

영상 획득 장치에 의한 잡음이나 배경 환경의 변화 등은 가우시안 혼합 분포의 학습에 의해 배경 모델의 확률 분포에 반영된다. 이렇게 얻어진 가우시안 확률 분포는 날씨나 조명의 변화를 적용하며 변화하는 배경을 효과적이고 정확하게 모델링한다. 이 때 물체를 나타내는 색정보  $I_{object}(t)$ 는 배경 모델의 학습에 영향을 주어서는 안 된다. 기존 학습 방법에

서는 winner-take-all[5]이나 soft-partition[10] 을 이용하여 혼합비율이 높은 가우시안에 해당되는 입력  $I(t)$ 를  $I_{environmentalchange}(t)$ 라 가정하고 이에 대한 혼합 비율을 높이는 방법을 취했다. 혼합 비율이 낮은 가우시안 분포는 K개의 가우시안들 중 하나로 남아있을 뿐, 혼합 비율이 점점 낮아지거나 새로운 분포로 대체되어 사라지게 된다. 그러나 이것은 기본적으로 배경의 색 정보에 대한 입력이 물체의 색 정보에 대한 입력에 비해 상대적으로 많이 크며 서로 다른 물체의 색정보를 나타내는 가우시안들이 서로 다른 가우시안 분포로 설정되고 학습된다는 가정에 따른 것이다. 배경 모델의 확률적 분포에 해당되지 않는 색 정보를 가진 물체가 가끔씩 지나간다면 이 때 입력되는  $I(t)$ 는 이미 학습된 배경 모델에 영향을 주지 않는다. 이러한 입력만이 들어온다면 배경 모델은 학습된 분포를 잘 유지함으로써 현재 배경을 정확히 모델링할 수 있으며 나타난 물체를 정확히 검출할 수 있게 된다. 그러나 배경 모델의 색정보 분포에 해당하는 물체의 거듭된 등장은  $I_{object}(t)$ 가 배경에 대한 확률분포를 변화시키도록 하고 배경 모델은 그 정확도를 잃게 된다. 배경과 물체가 동일한 가우시안으로 학습되는 경우 평균의 이동은 비록 적지만 분산은 크게 증가하며, 물체의 분포를 학습한 가우시안의 출현은 실제 배경을 나타내는 가우시안의 혼합 비율을 감소시킴으로써 결국 검출 정확도의 저하를 가져오게 된다. 색분포가 비슷한 물체가 자주 나타나는 환경에는 이러한 영향이 심각하게 나타나게 된다. 특히 공장이나 군대에서는 옷의 색이 동일하기 때문에 이를 분리해주지 않는다면 물체를 올바르게 검출해 내지 못한다. 이를 방지하기 위해서는 단순히 학습 속도를 조절해 주는 방법으로는 한계가 있다. 본 논문은 조명과 날씨에 따른 색 정보의 변화와 물체의 등장에 따른 색 정보의 변화를 효과적으로 분리하여 학습하는 방법을 제안한다.

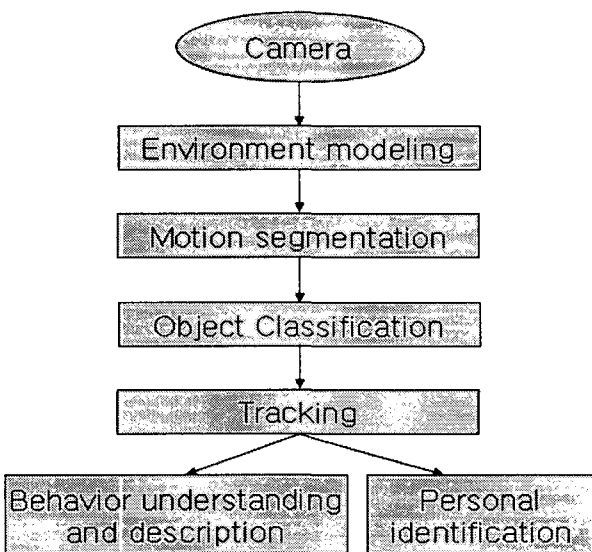


그림 1 영상 감시 시스템의 일반적인 구조  
 Fig. 1 General framework of visual surveillance systems

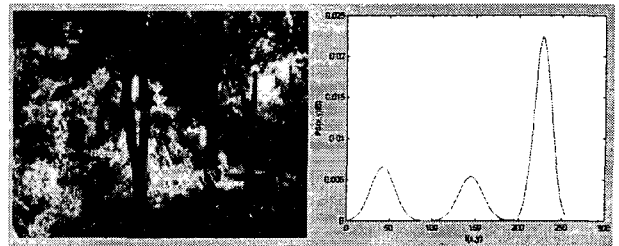


그림 2 반복되어 흔들리는 나뭇잎 영역의 입력 영상(좌), 십자로 표시된 화소에서 가우시안 혼합 모델을 사용해 학습한 배경 모델의 pdf(우).

Fig. 2 Example of dynamic background region and its modeled pdf.

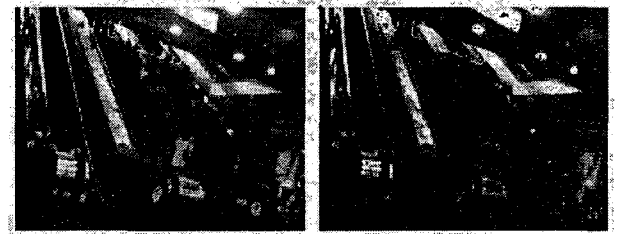


그림 3 기존 방법에 [5]의 알고리즘에 따라 학습된 보행자가 많은 지하철 에스컬레이터에서의 배경

Fig. 3 Example of high traffic region and its modeled background using the algorithm of [5].

## 2. 본 론

위에서 살펴보았듯이 이상적인 것은  $I_{object}(t)$ 을 분리하여 그에 대한 학습을 막는 것이다. 이론적으로 스테레오 카메라의 경우 이를 구분하는 것은 어려운 문제가 아니다. 그러나 카메라 한대를 통해 들어오는 2차원 영상에서는 이를 구분하는 것이 간단하지 않다. 근본적으로 적응적 배경 모델을 만드는 이유가  $I_{object}(t)$ 를 찾기 위한 것이기 때문에 배경 모델을 학습하는데 있어  $I_{object}(t)$ 을 분리하여 그에 대한 학습을 막는다는 결론은 모순이다. 따라서 이러한 문제를 해결하기 위해서는 배경 모델을 사용하지 않고  $I_{object}(t)$ 를 구분해내는 접근 방법이 필요하다.

### 2.1 Temporary Moving Region Mask

장면 차분은 배경의 정보가 없는 상황에서도 움직이는 물체를 검출할 수 있다[4]. 하지만 장면(frame) 간의 변화가 문턱값 이상인 영역만을 찾기 때문에 움직이고 있는 물체의 영역 중에서도 색 정보의 변화를 나타내는 일부 밖에는 검출할 수 없다는 한계를 지닌다. 그러나 물체의 움직임이 연속적이란 사실을 이용, 장면 차분을 통해 얻은 움직이는 물체의 일부 영역을 시간에 따라 효과적으로 간직한다면 배경의 값을 알지 못하는 상황에서도 움직이는 물체의 대부분의 영역을 찾아낼 수 있다.

$$T = \max\{I(t) - I(t-j), \forall j \in [1, M]\} \quad (8)$$

T는 각 화소의 입력 값의 차의 최대값을 N 장면 동안 기억함으로 현재까지 N개의 장면 동안의 최대 변화 값을 최대값을 나타낸 장면으로부터 N 장면 동안 간직한다. 이는 장면 차분을 통해 얻어진 변화가 있는 영역을 N 장면 동안 유지하게 해준다. 이를 통해 현재 시간에 움직인 물체의 대부분의 영역을 얻게 된다. 그러나 T값만으로 얻어진 영역은 단지 움직임이 있었던 영역들을 N 장면동안 간직한 결과이기 때문에 움직임이 끝난 영역들을 많이 포함하고 있다. 움직임이 있는 영역만을 얻어내기 위해서는 화소의 분산을 검사하여 움직임이 끝난 안정된 화소들을 제외해야 한다.

$$S = \frac{M \sum_{j=1}^M I(t+j)^2 - (\sum_{j=1}^M I(t+j))^2}{MM-1} \quad (9)$$

S는 각 화소에서 현재 이후의 M개의 장면 동안 입력값의 분산을 계산하여 문턱값 이하가 되는 경우, 그 화소에서 움직임이 없는 것으로 처리한다. 현실적으로 미래의 입력을 안다는 것은 사실 불가능하다. 이와 같은 S값을 이용하기 위해서 입력되는 모든 값이 M 장면 지연되어 사용된다. 지연되는 시간은 수 초 이내이며 자동차와 같이 빠른 움직임이 아닌 경우 관찰자는 이것을 거의 느끼지 못한다.

이와 같이 T와 S를 이용하면 배경이 없는 상태에서도 움직임이 있는 영역을 얻어낼 수 있다[12]. 이것은 물체 검출에 쓰기에는 정확도가 낮지만, 배경이 없는 상태에서 이동하는 물체의 영역만을 찾아내는 데는 매우 효과적인 방법이다. 이를 이용하여 물체의 이동에 따른 색 정보의 변화를 구별해낼 수 있다. 이러한 변화를 나타내는 영역을 temporary moving region mask로 설정하고 이를 M으로 나타내었다. 이 영역에서 들어오는 입력을  $I_{object}(t)$ 로 보고 이러한 입력에 대하여 가우시안 혼합 모델의 학습에 제한을 두었다.

$$M(x, y) = \begin{cases} 1 & T(x, y) > Threshold_T, S(x, y) > Threshold_S \\ 0 & otherwise \end{cases} \quad (10)$$

Temporary moving region mask는  $I_{environmentalchange}(t)$ 가 들어오는 영역을 제외해야한다. 조명과 날씨에 따른 색 정보의 변화는 물체의 이동에 따른 변화에 비해 상대적으로 작은 변화값을 가지기 때문에 이러한 문제는 적절한  $Threshold_T$ 의 선택을 통해 해결된다. 그림 5는 temporary moving region mask가 움직이는 물체에 대한 색 정보의 변화 구간을 찾아내는 것을 보여준다. 결과값 M을 보면 S와 T값을 이용하여 물체의 이동에 따른 입력 영상의 구간을 찾아낸 것을 확인할 수 있다.

## 2.2 Dynamic background detection

위의 방법은 움직이는 영역을 찾아내어 그에 대한 가우시안 혼합 모델의 학습을 효과적으로 막아준다. 그러나 문제는 배경 중에서도 항상 움직이는 부분이 있다는 것이다. 서론에서 가우시안 혼합 모델의 장점으로 제시했던 반복적인 움직

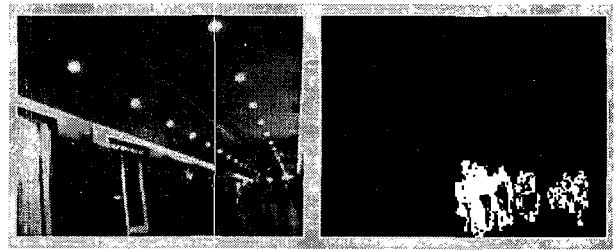


그림 4 1124장면에서 설정된 temporary moving region mask 영역

Fig. 4 Example of temporary moving region mask. (frame number: 1124)

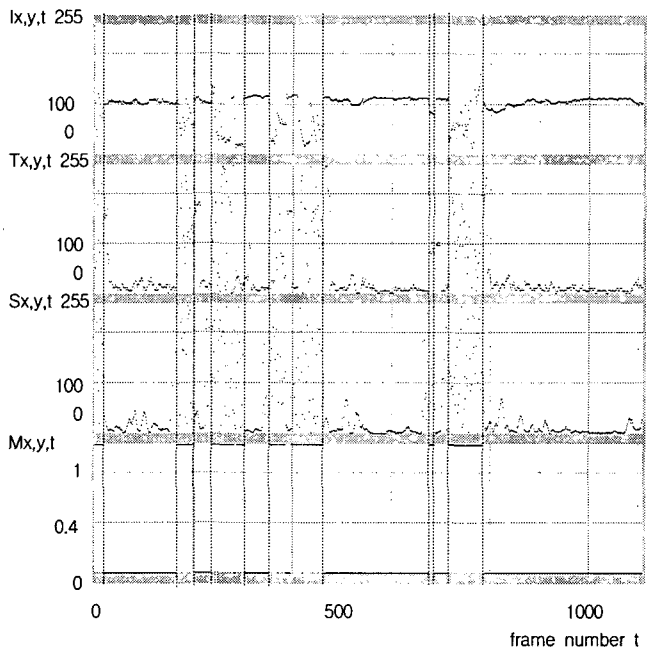


그림 5 그림 4에서 십자로 표시된 화소(x,y)에서 입력 영상 I에 따라 구한 T, S, M 값의 변화.

Fig. 5 The changes of I,T,S,M according to the pixel of Fig. 4.

임이 있는 배경(dynamic background)에서는 temporary moving region mask의 문제점이 드러난다. 나뭇가지의 흔들림이나 분수의 움직임 등의 경우 일부 영역은 그 색 정보의 변화의 크기와 분산이  $Threshold_T$ 와  $Threshold_S$ 보다 크다. 이런 경우 temporary moving region mask는 이를 물체의 움직임으로 인식하고 이에 대한 학습을 제한하게 된다. 따라서 이 영역은 배경에 대한 학습이 되지 않고 그 결과 항상 물체가 있는 것으로 나타나게 된다. 반복적인 움직임이 있는 배경을 학습할 수 있는 가우시안 혼합 모델의 장점을 잃어버리게 되는 것이다. 따라서 반복적인 움직임이 있는 배경을 찾아 Temporary moving region mask에서 제외해주는 것이 필요하다. 이는 dynamic background rate r의 계산을 통해 이루어진다.

$$r_{x,y,t} = (1-\lambda)r_{x,y,t-1} + \lambda f_{threshold_T}(I_{x,y,t} - I_{x,y,t-1}) \quad (11)$$

$$f_{threshold_r}(I) = \begin{cases} I < threshold_r & 0 \\ I > threshold_r & 255 \end{cases} \quad (12)$$

이 때 물체가 자주 지나다니는 영역과 반복적인 움직임이 있는 배경에 대한 구분이 필요하다. 이 두 영역을 비교해 보면 두 영역 모두 색 정보 변화의 크기와 분산이 크다. 그러나 물체가 자주 지나다니는 배경의 경우, 지나가는 물체가 많을 지라도 물체와 물체 사이와 물체의 색 분포에 따라 안정된 구간이 잠깐씩 나타난다. 그렇기 때문에 물체가 자주 지나다니는 영역의 한 화소에서 색 정보의 장면 당 변화가 일정 값  $threshold_r$  이상 나타난 횟수와 나타나지 않는 횟수를 비교해 보면 장면당 변화값이 안정된 구간이 상당부분을 차지함을 알 수 있다. 그러나 반복되는 움직임이 있는 배경의 영역에서 이 둘의 비율을 비교해보면 색 정보가 안정되는 구간이 거의 없는 것을 확인할 수 있다. 따라서 시간당 변화가  $threshold_r$  이상인 유효 움직임이 얼마나 자주 나타나는지를 학습하면 이 영역이 반복적인 움직임을 나타내는 영역인지 또는 물체가 자주 지나가는 영역인지 확인할 수 있다. dynamic background rate의 학습 비율  $\lambda$ 은 장면 속도를 고려해서 정해져야 하며 다른 학습의 학습 비율에 비해 상대적으로 낮다. 움직임이 자주 나타나는 영역에서  $r$ 값은 천천히 상승하며 일정 시간이 지나면 학습에 따라 반복적인 움직임이 있는 영역만을 찾아낸다. 이러한 학습 방법은 나뭇잎의 흔들림과 같이 시간에 따라 움직임이 잠시 멈추는 특성을 가지는 영역의 경우에도 반복적인 움직임을 나타내는 영역을 효과적으로 유지할 수 있다.

Harville등은 [6]에서 이와 비슷한 움직임에 대한 학습을 소개하였다. 그러나 이들은 움직임에 대한 학습을 혼합한 환경을 찾아 이 영역에 대해 해당하는 모든 가우시안 분포의 학습 비율을 낮추는데 사용하였다. 이러한 움직임에 대한 학습은 혼합한 영역의 배경을 정확하게 모델링하는 데에는 기여하는 바가 적지만 반복적인 움직임이 있는 배경을 찾는 경우에는 매우 유용하다. 본 논문에서는 이 방법을 약간 보완하여 반복적인 움직임이 있는 배경을 찾아내었다. Dynamic background rate 식(11)을 activity-based learning modulation 식(3)과 비교해보면  $r$ 은 반복적인 움직임이 있는 영역을 찾기 위해 장면 단위의 색 정보의 차가 얼마나 보다는 그것이 실제 물체의 움직임에 따른 것이냐와 그것이 얼마나 반복되었느냐를 시간에 따라 학습한다. 이를 위해  $threshold_r$ 를 사용하는데 이는 temporary moving region mask에 해당하지 않는 움직임의 경우는 고려할 필요가 없기 때문이다. 이렇게 해서 학습된  $r(x,y)$ 의 값이 문턱값  $threshold_r$  이상인 경우에는 이 영역을 dynamic background로 설정하고 temporary moving region mask에서 제외하여 가우시안 혼합 모델로 학습되도록 한다.

따라서 위에서 제시한 식(10)은 가우시안 혼합모델이 반복적인 움직임을 나타내는 배경을 모델링하도록 다음과 같이 수정된다.

$$M(x,y) = \begin{cases} 1 & T(x,y) > Threshold_T, S(x,y) > Threshold_S, \\ & r(x,y) < Threshold_r, \\ 0 & otherwise \end{cases} \quad (13)$$

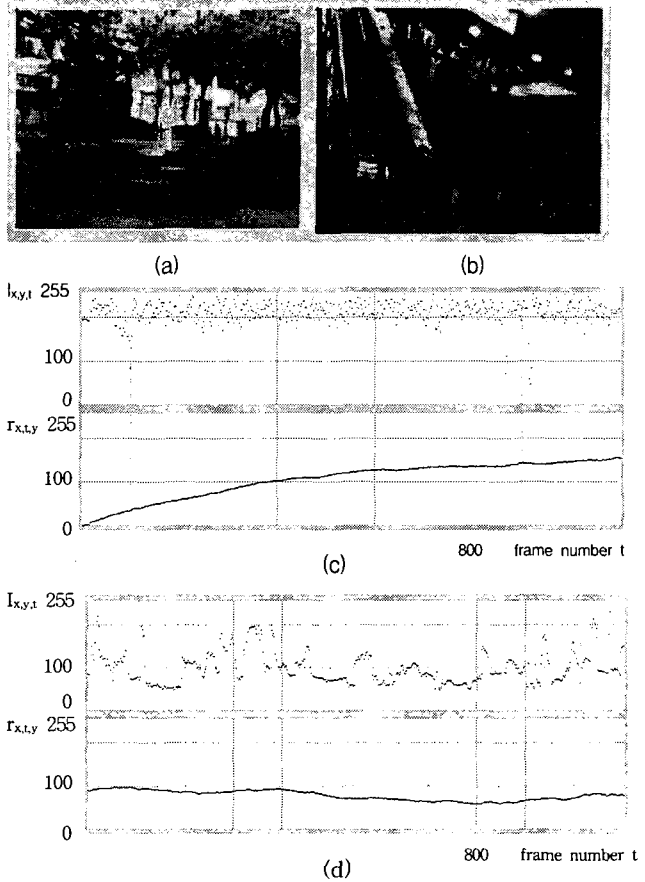


그림 6 분수(a)(c)와 혼합한 지하철 에스컬레이터(b)(d)에서 영상 입력값  $I(t)$ (위)와 dynamic background rate  $r$ (아래)의 비교

Fig. 6 The Comparison of dynamic background rate between dynamic region and high traffic region.

### 2.3 Temporary moving region mask를 이용한 가우시안 혼합 모델의 학습

학습 초기에는 어느 것이 배경인지 물체인지에 대한 확실한 정보가 없다. 일단 색 정보의 변화가 적은 화소들을 배경이라 가정한다. 따라서 학습은 temporary moving region mask가 0인 영역에 대해서만 행해진다. 물체가 이동하면 그 이동이 나타나는 영역의 화소에서는 이동하는 동안 temporary moving region mask의 값이 1이 된다. 그 동안 이 화소에서 입력되는 값에 대한 가우시안 혼합 모델의 학습은 잠시 멈춰진다. 물체의 이동이 끝나고 temporary moving region mask의 해당 값이 0이 되면 학습은 다시 진행된다.

날씨나 조명의 변화 등은  $threshold_T$  이하이기 때문에 이러한 변화에 대하여는 temporary moving region mask의 해당 값이 0으로 유지된다. 따라서 가우시안 혼합 모델의 학습이 진행되어 변화에 적용하게 된다. 배경을 이루고 있던 기존 물체가 사라지거나 배경에 새로운 물체가 첨가되는 경우에는 입력 영상 값의 변화가  $threshold_T$ 와  $threshold_S$  보다 크다. 이러한 변화에 대한 학습은 물체가 사라지거나 나타나는 동안 temporary moving region mask에 의해 잠시 제한되지만 움직임이 끝나면 학습은 곧 진행되어 이러한 변화에 적용

하게 된다. 이는 움직임이 끝나고 수 초 안에 완료된다. 반복적인 움직임을 나타내는 배경의 경우에는 dynamic background rate를 사용하여  $threshold_r$  이상인 영역에 대하여는 기존에 구한 temporary moving region mask에 상관없이 가우시안 혼합 모델로 학습되도록 한다.

그림7 에는 입력 영상값 중 물체의 이동에 따른 입력만을 분류하여 학습에서 제외시킴으로써, 기존의 가우시안 학습 방법의 장점을 유지하며 배경 모델의 정확도를 높이는 학습 과정의 흐름이 도시되어 있다. 이러한 처리는 기존의 확률적 방법을 이용하여 작은  $w_k$  과 큰  $\sigma_k$  를 갖는 가우시안의 영향을 줄인 방법과 함께 정확한 배경 모델을 구성하도록 한다. 정확한 배경 모델을 배경 모델의 정확도의 향상은 물체 검출의 정확도의 향상과 직결된다.

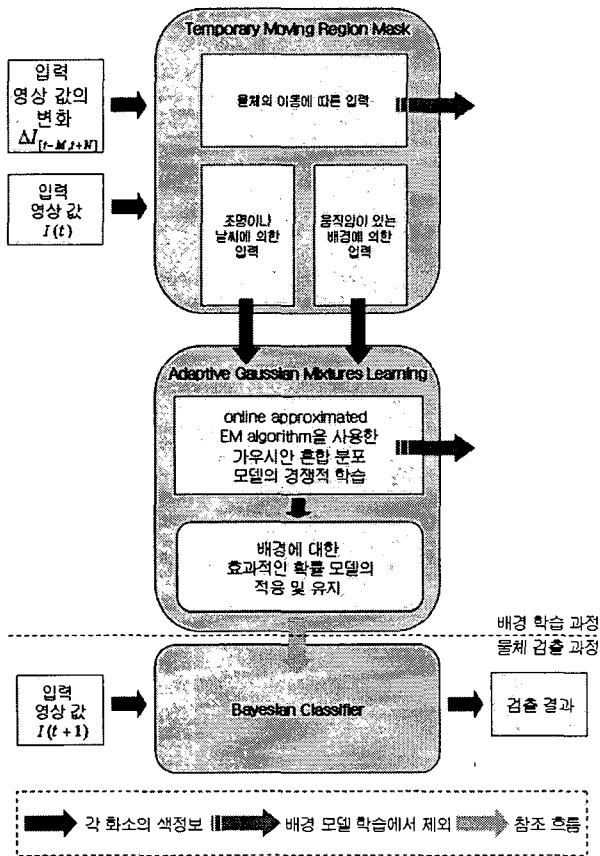


그림 7 제안하는 학습 방법의 흐름도.  
Fig. 7 Architecture of proposed method

2.4 실험 결과

실험은 펜티엄 2.4GHz Pentium 4 PC에서 행해졌다. 영상은 USB 카메라를 통해 얻어진 320\*240의 RGB 입력을 YUV로 변환하여 사용하였고 Lee가 제안한 알고리즘[10]을 사용하여 가우시안 혼합 모델의 학습 및 물체 검출을 하였다. 영상은 백화점 통로, 자동차 도로, 지하철 에스컬레이터와 및 학교 주위의 여러 환경에서 획득하여 제안한 방법이 실내와 실외, 반복적인 움직임이 있는 배경, 날씨의 변화, 혼잡도의

변화 등의 여러 요인들에 대하여 얼마나 효과적인지 검토하였다.

백화점 통로에서 획득한 영상에 따른 결과를 보면 제안된 방법이 혼잡한 환경에서 배경 모델링의 정확도를 향상시키는 것을 확인할 수 있다. 각 결과들은 그림 9에 나타나있다. 표1을 보면 물체의 색 정보 중 자주 나타나는 색 정보를 배경의 색정보와 동일한 가우시안 분포로 인식함으로 평균과 분산이 잘못 학습된 것을 확인할 수 있다. 상대적 학습량(C)[10]은 해당 가우시안이 입력에 대하여 학습한 상대적인 양을 나타낸다. 이를 비교해보면 기존의 학습 방법은 학습된 가우시안 분포가 배경이 아닌 잘못된 정보에 대하여 상대적인 양 800 이상의 학습을 하였다. 제안한 방법은 자주 나타나지 않는 물체의 색 정보를 상대적인 양 600이상을 학습과정에서 제외시켰다. 학습된 배경을 통해서도 이것을 확인할 수 있다. 육안으로는 비슷하게 보일지 모르지만 기존의 방법으로 학습한 배경 영상은 사람이 많이 지나다니는 영역의 밝기 값이 평균 10정도 낮다. 이것은 배경보다 밝기 값이 작은 사람들의 옷 영역이 학습되었기 때문이다. 따라서 밝기 값이 낮은 물체에 대한 인식률이 저하되게 된다. 이를 베이스안 결정 이론을 이용한 검출 영상을 통해 확인할 수 있는데, 비교해보면 제안된 방법은 여자의 다리 부분을 정확히 검출해 내는 것을 확인할 수 있다. 그림 4,5의 동일한 입력 영상에 따른 결과를 보면 표시된 화소에서 8명의 지나가는 물체 영역의 학습이 temporary moving region mask를 통해 효과적으로 제한된 것을 알 수 있다.

|            |                       |
|------------|-----------------------|
| 입력 영상      | 학습한 배경 영상             |
| 물체를 검출한 영상 | temporary region mask |

그림 8 결과 영상의 구성  
Fig. 8 The organization of images

그림11은 비오는 날 육교위에서 획득한 도로에서 자동차의 움직임에 대한 실험 결과이다. 지나가는 차의 양이 비교적 많고 속도 또한 빠르기에 검출영상에서는 큰 차이를 보이지 않는다. 하지만 표2의 각 가우시안의 매개변수 값을 보면 기존의 방법이 물체의 색정보의 영향을 받은 것을 확인할 수 있다. 차의 속력이 느리고 횡수가 높아진다면 이러한 결과는 더욱 뚜렷해진다. 이 결과를 통해 비가 오는 환경의 변화에도 제안한 방법이 배경 모델의 정확도를 높임을 확인할 수 있다.

그림 13에서 dynamic background rate를 사용하여 얻어진 반복적인 움직임을 나타내는 배경 영역이 temporary moving region mask의 영역에서 제외된 것을 확인할 수 있다. 그 결과 분수와 나뭇가지의 흔들림이 가우시안 혼합모델의 학습을 통해 배경으로 모델링되었다. 물체의 움직임을 검출한 영상을 보면 분수와 나뭇가지의 영역이 물체의 움직임으로 검출

되지 않았음을 보여준다.

### 3. 결 론

가우시안 혼합 모델은 다양한 배경에 대하여 적응적 배경을 만들고 환경 변화를 학습해나가는 데 매우 효과적인 방법이다. 본 논문에서는 기존 방법으로 물체가 자주 지나다니는 배경을 학습하는데 있어 나타나는 문제를 분석하였고, 단순히 각 가우시안의 학습 비율을 조정함으로써 배경과 물체를 구분하여 배경을 모델링하는 방법의 한계를 지적하였다. 이를 개선하기 위해 배경 모델이 없는 상태에서 조명의 변화를 제외한 움직이는 물체만의 영역을 얻어내고 이러한 영역에 대한 학습을 제한하였다. 또한 반복적인 움직임이 나타나는 영역을 찾아내어 기존의 방법의 이러한 영역에 대해 학습할 수 있는 장점을 유지하였다. 그 결과 기존의 장점들을 모두 유지하면서 혼잡한 환경에서 모델의 정확도를 크게 높였다. 또한 배경 모델의 정확도 향상을 통해 물체 검출의 정확도가 높아짐을 실험 결과를 통해 보였다. 이러한 성능 향상은 가우시안 혼합 모델뿐만이 아닌 online approximated EM algorithm의 학습 비율을 조정함으로써 배경과 물체에 대한 색 정보의 변화를 구분하는 다른 적응적 배경 모델 방법에서도 동일한 것으로 기대된다. 이러한 학습 방법은 기존의 시스템에 추가하기에 용이하다. 또한 각 화소의 변화에 대한 처리이므로 구현의 측면에서도 장점을 가진다. 이 연구의 결과가 가우시안 혼합 모델을 사용하는 많은 비디오 시스템의 정확도 향상에 기여할 것으로 기대된다.

#### 감사의 글

이 연구는 2003년 홍익대학교 연구 기금의 지원으로 수행되었습니다.

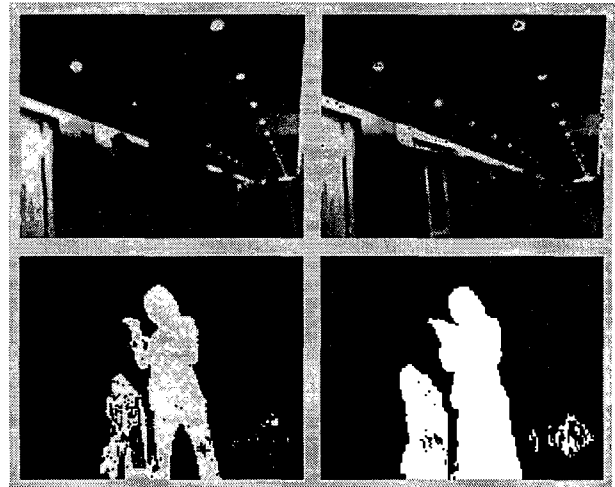
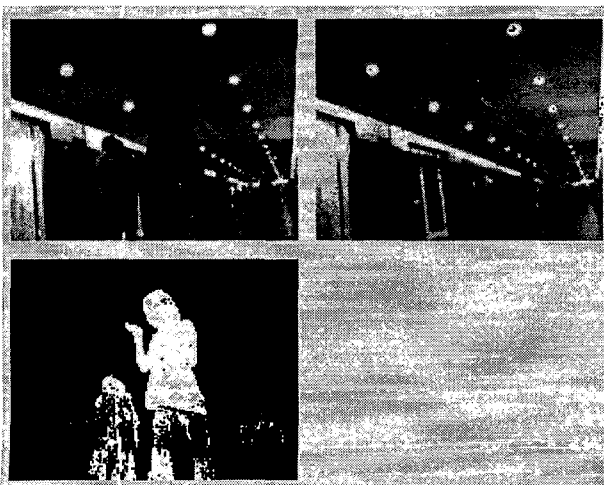


그림 9 백화점 통로 영상의 1863번째 장면에서 배경 학습과 물체 추출 결과 비교(아래측이 제안한 방법)

Fig. 9 Comparison of results with algorithm [10] on department store video (frame number: 1863).

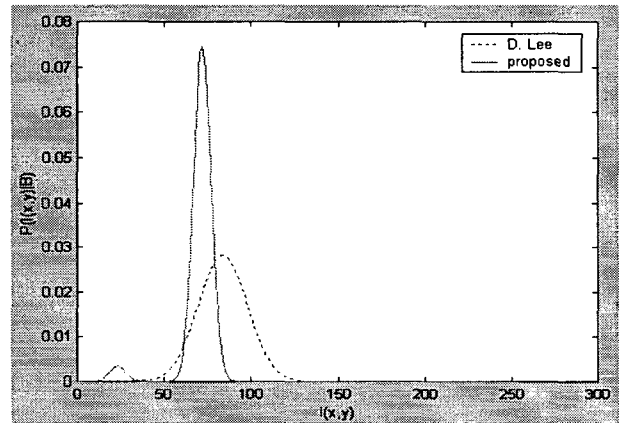


그림 10 그림 9영상의 표시된 점에서 학습에 따라 가우시안 혼합 분포로 추정된 pdf 비교

Fig. 10 Comparison of modeled pdf at the pixel marked as cross of Fig. 9.

표 1 백화점 통로에서 가우시안 혼합 모델의 각 매개변수 값의 비교

Table 1 Comparison of parameters at the pixel marked as cross of Fig. 9.

| 가우시안 번호 (K=3) | 기존 방법[10] |    |    | 제안하는 방법 |        |    |
|---------------|-----------|----|----|---------|--------|----|
|               | Y1        | Y2 | Y3 | Y1      | Y2     | Y3 |
| 평균            | 83.77     | -  | -  | 72.05   | 23.03  | -  |
| 분산            | 200       | -  | -  | 25.75   | 21.23  | -  |
| 상대적 학습률       | 1849      | -  | -  | 1036.44 | 181.56 | -  |
| 구성 비율         | 1         | -  | -  | 0.95    | 0.048  | -  |



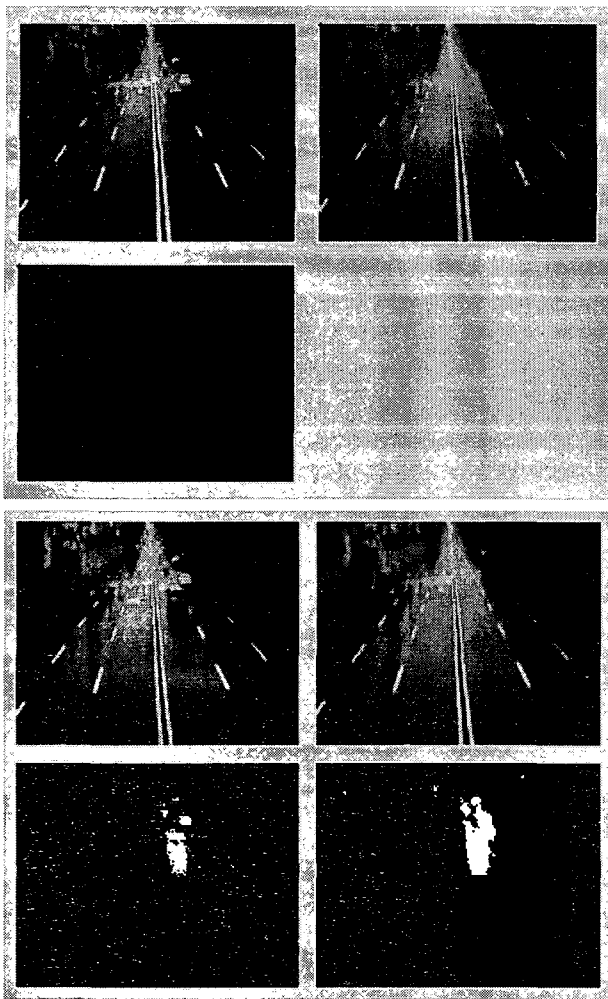


그림 11 비오는 날 도로 영상의 773번째 장면에서 비교(아래 쪽이 제안하는 방법)

Fig. 11 Comparison of results with algorithm [10] on rainy street video (frame number: 773).

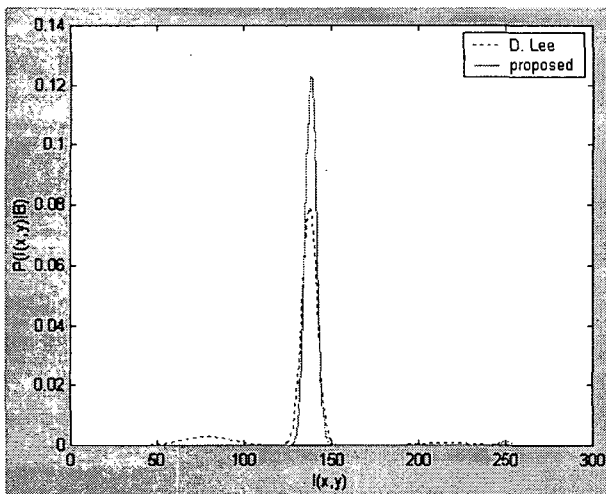


그림 12 비오는 날 도로 영상의 표시된 점에서 학습에 따라 가우시안 혼합 확률 분포로 추정된 pdf 비교

Fig. 12 Comparison of modeled pdf at the pixel marked as cross of Fig. 11.

표 2 도로 영상에서 가우시안 혼합 모델의 각 매개변수 값의 비교

Table 2 Comparison of parameters at the pixel marked as cross of Fig. 11.

|               | 기존 방법[10] |        |       | 제안하는 방법 |        |    |
|---------------|-----------|--------|-------|---------|--------|----|
|               | Y1        | Y2     | Y3    | Y1      | Y2     | Y3 |
| 가우시안 번호 (K=3) |           |        |       |         |        |    |
| 평균            | 137.74    | 215.99 | 78.97 | 138.45  | 249.99 | -  |
| 분산            | 18.60     | 200    | 200   | 10.13   | 15.60  | -  |
| 상대적 학습량       | 643.25    | 10.00  | 28.76 | 630     | 4      | -  |
| 구성 비율         | 0.86      | 0.04   | 0.10  | 0.99    | 0.012  | -  |



그림 13 Dynamic region의 학습 이전(장면 번호: 180)과 학습 이후(장면 번호: 957)의 temporary moving region mask의 비교. 짙은 회색 영역이 dynamic region이다.

Fig. 13 Example of learning dynamic region filled dark gray.(frame number: 180, frame number: 957)

참 고 문 헌

[1] A. Ghosh, S. Devadas, K. Keutzer and J. White, "Estimation of Average Switching Activity in Combinational and Sequential Circuits," ACM/IEE Design Automation Conf., pp. 253-259, 1992.

[1] Gian Luca Foresti, Christian Micheloni, Lauro Snidaro, Paolo Remagnino, and Tim Ellis "Active Video-Based Surveillance System", IEEE Signal Processing Magazine, pp. 25-37, March 2005

[2] Arun Hampapur, Lisa Brown, Jonathan Connell, Ahmet Ekin, Norman Haas, Max Lu, Hans Merkl, Sharath Pankanti, Andrew Senior, Chiao-Fe Shu, and Ying Li Tian "Smart Video Surveillance", IEEE Signal Processing Magazine, pp. 38-51, March 2005

[3] Trista P. Chen, Horst Haussecker, Alexander Bovyryn "Computer Vision Workload Analysis Case Study of Video Surveillance Systems", Intel Technology Journal Vol 9, Issue 2, May 19 2005

[4] C. Anderson, Peter Burt, and G. van der Wal. "Change detection and tracking using pyramid transformation techniques". In Proceedings of SPIE - Intelligent Robots and Computer Vision, volume 579, pp. 72-78, 1985.

[5] C. Stauffer and W.E.L. Grimson "Adaptive Background Mixture Models for Real-Time Tracking", Proc. Conf. Computer Vision and Pattern Recognition, vol 2, pp. 246-252, June 1999

[6] M. Harville, G. Gordon, and J. Woodfill, "Foreground Segmentation Using Adaptive Mixture Models in Color and Depth," Proc. ICCV Workshop Detection and Recognition of Events in Video, July 2001.

[7] S.J. McKenna, Y. Raja, and S. Gong, "Object Tracking Using Adaptive Color Mixture Models," Proc. Asian Conf. Computer Vision, vol. 1, pp. 615-622, Jan. 1998.

[8] P. Kaew, Tra Kul Pong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection," Proc. European Workshop Advanced Video Based Surveillance Systems, Sept. 2001

[9] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," Proc. 13th Conf. Uncertainty in Artificial Intelligence, Aug. 1997.

[10] Dar-Shyang Lee "Effective Gaussian Mixture Learning for Video Background Subtraction" IEEE Trans. Pattern Analysis and Machine Intelligence Vol. 27, pp. 827-832, May. 2005.

[11] C. Eveland, K. Konolige, R. Bolles. "Background Modeling for Segmentation of Video-rate Stereo Sequences". In CVPR'98, pp. 266-271, June 1998.

[12] Robert T. Collins, Alan J. Lipton, Takeo Kanade,

Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt and Lambert Wixson "A System for Video Surveillance and Monitoring" Carnegie Mellon University 2001.

저 자 소 개



박 대 용(朴大龍)

2003 홍익대 전자전기공학부 졸업  
 2003~ 홍익대 전기제어정보대학원  
 Phone: 02 - 320 - 1493  
 Fax: 02 - 320 - 1110  
 E-mail: tkc-tmhk@hanmail.net



김 재 민(金載敏)

1984년 서울대 전기공학과 졸업.  
 1994년 R. P. I. 전기공학과 졸업(공학박)  
 현재 홍익대 전자전기공학부 부교수  
 Phone: 02 - 320 - 1634  
 Fax: 02 - 320 - 1110  
 E-mail: jaemin@wow.hongik.ac.kr



조 성 원(趙成元)

1982년 서울대 전기공학과 졸업.  
 1992년 Purdue Univ. 전기공학과 졸업(공학박)  
 현재 홍익대 전자전기공학부 교수  
 Phone: 02 - 320 - 1493  
 Fax: 02 - 320 - 1110  
 E-mail : swcho@wow.hongik.ac.kr