

# 개선된 적응적 가우시안 혼합 모델을 이용한 객체 검출

## Advanced Gaussian Mixture Learning for Complex Environment

박대용, 김재민, 조성원, 김준범  
홍익대학교 전자전기공학부

Dae-Yong Park, Jae-Min Kim, Seong-Won Cho

Dept. of Electronic and Electric Engineering Hong Ik National University

E-mail : tkc-thmk@hanmail.net

### 요 약

Background Subtraction은 움직이는 물체 검출에 가장 많이 사용되는 방법 중 하나이다. 배경이 복잡하고 변화가 심한 경우, 배경을 실시간으로 얼마나 정확하게 학습하는가가 물체 검출의 정확도를 결정한다. Gaussian Mixture Model은 이러한 배경의 모델링에 가장 많이 쓰이는 방법이다. Gaussian Mixture Model은 확률적 학습 방법을 사용하는데, 이러한 방법은 물체가 자주 지나다니거나 물체가 멈춰있는 경우, 배경을 정확하게 모델링하지 못한다.

본 논문에서는 밝기 값에 대한 확률적 모델링과 밝기 값의 변화에 따른 처리를 결합하여 혼잡한 환경에서 배경을 정확하게 모델링할 수 있는 학습 방법을 제안한다.

### 1. 서론

최근 일어난 여러 테러 사건들로 국내외적으로 사회 안전망 구축에 많은 관심이 고조되고 있다. 이에 따라 보안·감시시장이 급속하게 성장되고 있으며, 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 시스템에 있어서 배경과 물체를 정확하고 효과적으로 분리해 내는 것은 기본적인 매우 중요한 일이다. 물체를 정확히 분리해내지 못한다면 그 다음 단계인 물체의 분류와 행동 분석은 매우 힘든 작업이 된다.

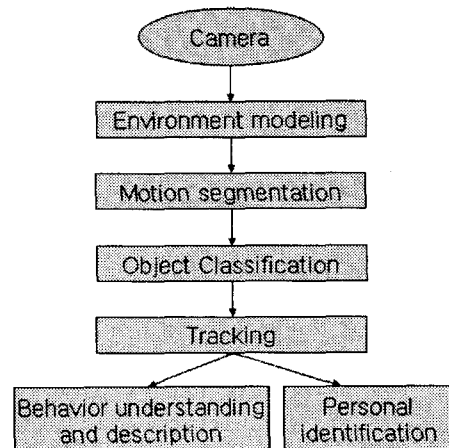


그림 1. 영상 감시 시스템의 일반적인 구조

#### 1.1 가우시안 혼합 모델

입력되는 영상은 시간에 따라 영상의 잡음, 조명이나 날씨의 변화, 물체들의 이동에 따라 변화한다. 이러한 다양한 변화 속에서 배경만을 모델

링하기 위해 자주 사용하는 방법이 확률적인 모델을 이용한 학습이다. 가우시안 혼합 모델은 이러한 확률적 모델을 이용한 학습에 가장 많이 사용되는 방법으로 영상의 각 픽셀들을 가우시안 혼합 확률분포를 이용해 모델링한다.[4,5,6] 각 가우시안 혼합 분포를 이루는 가우시안들은 online approximated EM algorithm에 따라 입력

\*본 논문의 연구는 홍익대학교 RA지원으로 이루어짐

되는 색 정보를 학습해나감으로 배경의 변화에 적응한다. 또한 기존의 가우시안을 새로운 가우시안으로 교체함으로 물체의 이동에 따른 변화에 적응한다.

가우시안 혼합 모델은 잡음에 강하며 나무 가지의 흔들림이나 파도 물결 등의 움직임을 나타내는 배경도 모델링할 수 있는 장점을 지닌다.

### 1.2 확률적인 학습 방법의 한계

확률적 모델을 이용한 학습의 기본적인 가정은 배경의 색정보를 모델링한 가우시안이 물체의 색정보를 모델링한 가우시안 보다 더 높은 확률을 가진다는 것이다. 따라서 물체의 색정보의 확률이 높아질수록 물체의 색정보를 학습한 가우시안과 배경의 색정보를 학습한 가우시안 사이에 구분이 모호하게 되고 결과적으로 배경 모델링의 정확도가 저하된다. 예를 들면 이동하는 물체가 정지하는 경우, 일정 시간이 지나면 기존 배경을 나타내는 확률보다 물체를 나타내는 확률이 높아진다. 물체의 색정보를 배경의 색정보로 학습하게 되는 것이다. 또한 물체가 자주 이동하는 경우에도 비록 물체가 멈춘 것은 아니지만 물체의 색정보의 확률이 높아짐으로 물체 검출의 정확도가 저하된다.

### 1.3 관련 연구들

가우시안의 학습 비율을 적절하게 조절하는 방법들은 복잡한 환경에서 물체 검출의 정확도를 높여준다. 최근에 D. Lee는 한 픽셀을 모델링하는 각 가우시안들의 학습 비율을 개별적으로 조절하는 방법을 제안하였다[7]. 그는 학습비율을 조절을 통해 가우시안 혼합 모델의 정확도를 높여온 기존의 여러 방법들을 정리하고[4,8,9], 배경모델의 정확도와 안정도를 크게 높인 학습 방법을 제안하였다. 그 결과 물체가 자주 다니지 않는 환경에서는 물체 검출의 정확도가 향상되었다. 그러나 물체가 어느 정도 이상 빈번하게 지나다니거나, 이동하는 물체가 멈추는 경우에는 배경을 정확하게 모델링하지 못한다.

확률적인 방법의 한계를 해결하는 방법들 중 가장 직관적으로 생각할 수 있는 해결책은 스테레오 비전의 사용이다[10]. M. Harville등은 스테레오 비디오를 통해 얻은 각 픽셀의 깊이(depth)에 따라 현재 입력이 배경을 나타내는 가우시안에 해당되는 값인지 물체를 나타내는 가우시안에 해당되는 값인지를 구분하였고, 이에 따라 배경 모델을 학습하였다[5]. 그러나 현재의 스테레오 비전의 기술로는 얻어낸 깊이의 정확도

가 낮고 연산량이 많다. 또한 근본적으로 이 방법은 카메라를 두 대 이상 사용하여야 한다.

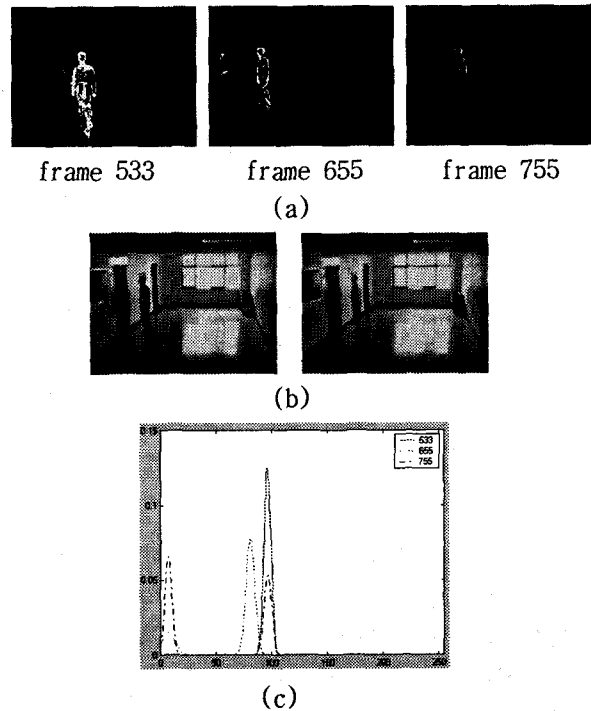


그림 2. (a) 기존 방법으로 학습한 배경에 따라 검출한 정지하는 물체 검출 결과 영상 (b) 프레임 755번에서 입력 영상과 학습된 배경 영상 (c) 십자로 표시된 화소에서 학습된 배경의 가우시안 혼합 모델의 pdf 변화



그림3. 기존 방법에 [4]의 알고리즘으로 학습된 배경

## 2. Object Region Mask

본 절에서는 위에서 논의한 확률적인 방법의 한계를 넘기 위해 각 픽셀들의 변화값을 이용하는 방법을 제시한다. 픽셀의 변화를 이용하는 방법은 대표적으로 frame differencing을 들 수 있다[11]. 이 방법은 배경 모델이 필요 없고 환경의 변화에 강인한 장점을 가지나 물체의 전체 영역을 얻어 낼 수 없는 한계를 지닌다. 따라서 얻어낸 물체의 일부 영역에서 전체 물체 영역을 얻

어내는 영상처리기법을 필요로 한다.

시간  $t$ 에서 배경 모델링을 위해 입력되는 영상에서 각 픽셀들의 변화는 다음과 같이 나타낼 수 있다.

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

정확한 배경의 모델링을 위해서는 이러한 픽셀 값의 여러 변화 중 물체로 인한 변화를 구별하여, 이러한 입력이 배경 모델에 영향을 주지 못하도록 하여야 한다.

본 절에서는 frame differencing과 가우시안 혼합 모델을 결합하여 대부분의 물체의 영역을 찾아내는 방법을 제시한다.

### 2.1 Moving Region

먼저 움직이는 물체로 인한 입력을 구분해야 한다. Collins등은 frame differencing방법을 확장하여 각 픽셀의 최대 변화값과 시간에 따른 안정도를 시간에 따라 유지하고 비교하여, 움직이는 영역들을 얻어내는 방법을 제안하였다[12]. 비록 이동 중에 정지하는 물체 검출에 대한 정확도는 부족하지만 이와 같은 접근 방법을 통해 배경의 색정보를 학습한 가우시안과 이동 중인 물체의 색정보를 학습한 가우시안을 구분할 수 있다.

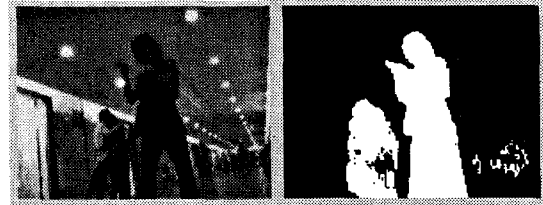
현재부터 입력되는 영상 (M+N)개를 매 프레임마다 저장한다. 저장된 영상 중 M번째 들어온 것을 현재 영상이라 가정하고 먼저 입력된 M개에서 S를 구하고 나중에 입력된 N개에서 T값을 구한다. 이 두 값의 비교를 통해 배경 모델이 없는 상태에서도 움직임이 있고 안정되지 않은 영역들을 찾아낼 수 있다. 조명에 따른 입력의 변화는 서서히 이루어지기 때문에 큰 값을 가지지 않는다. 따라서  $threshold_T$  이상인 변화들만을 이용하면 조명에 따른 변화 영역을 제외한 움직이는 물체의 대부분의 영역을 찾을 수 있다. 하지만 물체의 색상이 일정하여 S값이  $threshold_S$  이하가 되는 경우에는 물체에 의한 입력임을 구별하지 못한다.

이렇게 찾은 영역들을 moving region 으로 설정하고 이 영역에 대해서 가우시안 혼합 모델의 학습을 제한하였다.

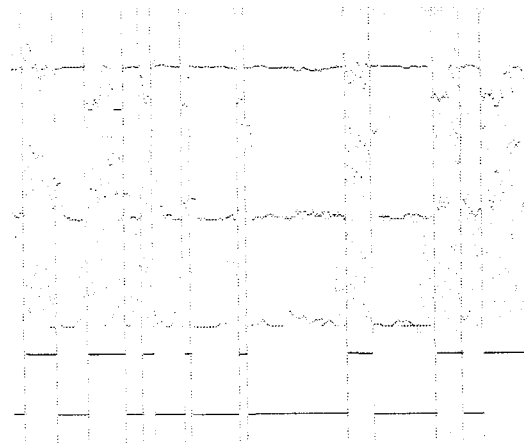
$$T = \max\{|I(t) - I(t-j)|, \forall j \in [1, M]\} \quad (2)$$

$$S = \frac{M \sum_{j=1}^M I(t+j)^2 - (\sum_{j=1}^M I(t+j))^2}{M(M-1)} \quad (3)$$

$$M(x,y) = \begin{cases} 1 & T(x,y) > threshold_T \\ & S(x,y) > threshold_S \\ 0 & otherwise \end{cases} \quad (4)$$



(a)



(b)

그림4. (a) 입력 영상과 얻어낸 moving region (b) 표시된 점에서 프레임1038에서 프레임1868까지의 값의 변화. 위로부터 아래로 입력 영상 값 I(t), T, S, M

### 2.2 Stationary Region

움직이던 물체가 정지하면 정지한 영역이 배경인지 물체인지 T와 S값만으로는 알 수 없다. 움직이던 물체의 영역 M인 픽셀이 안정화되어 S값이  $threshold_S$  이하가 되는 픽셀들을 정지 물체 후보 영역 C로 설정한다. 영역 C의 안정화 된 값들이 배경인지 멈춘 물체인지 구별함으로 이 영역의 입력을 학습할지 여부를 결정한다. 이 때 배경으로부터 들어오는 입력을 정지한 물체로부터 들어오는 입력으로 결정하면 학습한 배경의 안정도가 크게 손상된다. 배경의 특성에 따라 큰 분산을 가지는 픽셀도 있기 때문에 모든 픽셀에 대하여 동일한 문턱값을 적용할 경우 정확도가 저하된다. 이미 학습해둔 가우시안 모델의 표준편차를 이용하면 정지한 물체의 영역을 안정적으로 찾아낼 수 있다.

$$M(x,y) = \begin{cases} 1 & T(x,y) > threshold_T, \\ & S(x,y) > threshold_S \\ 2 & (x,y) \in C, \\ & P(x,y) < P(mean - 2.5 * \sigma) \\ 0 & otherwise \end{cases} \quad (5)$$

**2.3 Dynamic Background Region**

움직임이 있는 영역 중에서도 배경인 영역이 있다. 바람에 따른 나뭇가지의 영역이나 물이 솟아나는 분수, 파도가 치는 바다 등은 픽셀 값에 큰 변화가 있다. 이러한 영역을 물체로 인식할 경우 물체 검출의 정확도는 크게 떨어진다.

이러한 영역을 모델링할 수 있는 것이 가우시안 혼합 모델의 장점이다. 위와 같이 moving region을 설정해 줄 경우, 이러한 영역은 배경으로 모델링하지 못한다. 따라서 이러한 움직임이 있는 배경을 찾아 moving region에서 제외시켜주는 작업이 필요하다. 이는 dynamic background rate r의 계산을 통해 이루어진다. r이 문턱값 이상인 화소의 영역을 dynamic background region으로 설정하고

$$r_{x,y,t} = (1 - \lambda)r_{x,y,t} + \lambda f(|I_{x,y,t} - I_{x,y,t-1}|) \quad (6)$$

$$f(I) = \begin{cases} I < threshold_m & 0 \\ I > threshold_m & 255 \end{cases} \quad (7)$$

따라서 결과적으로 moving region은 다음과 같이 정리된다.

$$M(x,y) = \begin{cases} 3 & T(x,y) > threshold_T, \\ & S(x,y) > threshold_S \\ & r(x,y) < thrshhold_r \\ 2 & (x,y) \in C, \\ & P(x,y) < P(mean - 3 * variance) \\ & r(x,y) < thrshhold_r \\ 1 & r(x,y) > thrshhold_r \\ 0 & otherwise \end{cases} \quad (8)$$

**3. Object Region Mask를 이용한 학습**

학습은 moving region mask의 값이 2 이상인 영역에 대하여 행해진다. 물체가 등장하고 움직이고 있으면 그 영역의 moving region은 3으로 설정되고 가우시안 혼합 모델의 학습이 제한된다. 움직이는 물체가 멈추는 경우나 움직이는 물체의 영역 내부가 비슷한 색일 경우에는 2로 설정되어 학습이 제한된다. 움직임이 끝나고 배경 모델에 해당되는 입력이 들어오면 0으로 설정되어 학습이 계속된다. 1인 경우에는 다른 변화에

상관없이 학습이 진행된다. dynamic background region의 경우에는 물체의 움직임이 적은 상태에서 충분히 학습시키는 것이 필요하다. 물체가 많이 다니는 환경에서는 반복적인 움직임을 나타내는 배경 영역의 구분이 모호해지기 때문이다.

그림5에는 입력값을 분류하여 학습시킴으로 기존의 가우시안 학습 방법의 장점을 유지하며 배경 모델의 정확도를 높이는 학습 과정의 흐름이 도시되어 있다. 기존의 방법에서는 배경 모델의 형성을 위해 각 화소의 입력 값들이 구분 없이 학습되었지만, 제안하는 방법에서는 이미 형성된 배경을 참조하는 object region mask를 통해 배경 모델에 해당하는 입력 값들을 1차적으로 구분해 낸다. 이렇게 얻어진 값들만을 가우시안 혼합 모델로 학습하고 그 중에서도 높은 확률을 가지는 가우시안 분포를 배경 모델로 유지한다. 배경 모델의 정확도의 향상은 물체 검출의 정확도의 향상과 직결된다.

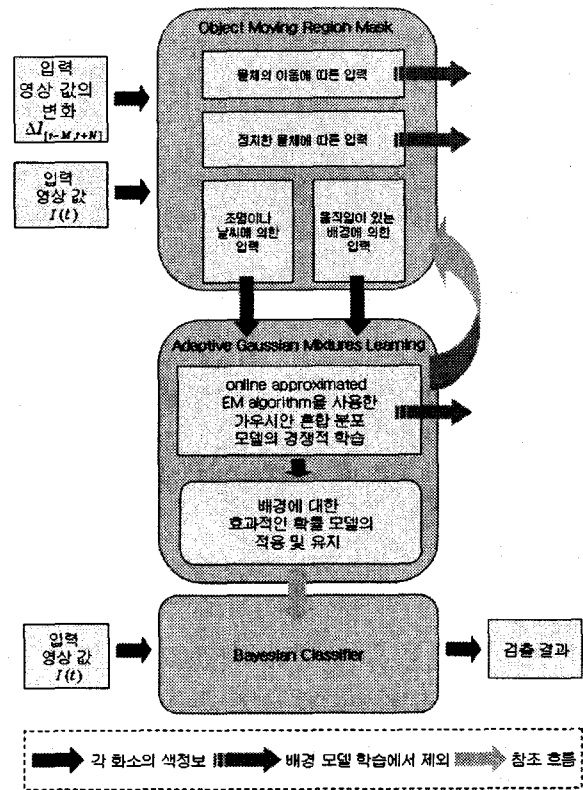


그림5. 전체 시스템 흐름도

**4. 실험 결과**

실험은 펜티엄 2.4GHz pentium 4 PC를 이용했다. 영상은 USB 카메라를 통해 얻어진 320\*240의 RGB 입력을 YUV로 변환하여 사용하였고 D. Lee가 제안한 알고리즘[4]을 사용하여 가우시안 혼합 모델의 학습 및 물체 검출을 하였다. 영상은 백화점 통로, 자동차 도로, 지하철

카메라로부터 들어오는 입력 영상	가우시안 혼합 모델로 학습한 배경 영상
물체 검출 영상	object region mask 영상

그림 6. 결과 영상의 구성

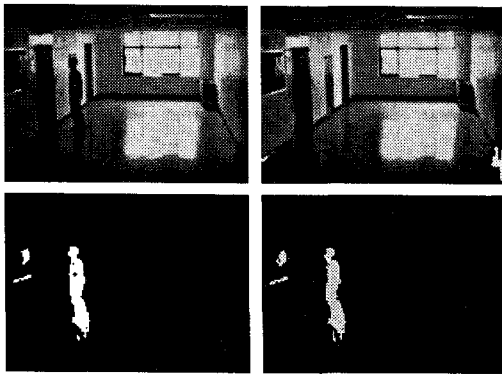


그림 7. 프레임 755에서 설정된 stationary region 영역. 좌측 위부터 시계방향으로 입력영상, 학습된 배경영상, stationary region, 검출된 물체영역



그림 8. 우측 아래 영상의 짙은 회색 영역이 dynamic background region이다.(얇은 회색은 stationary region이다) 좌측 아래 영상은 분수의 움직임이 물체의 움직임으로 검출되지 않았음을 보여준다.

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.

그림9,10과 표1를 통해 백화점 통로 영상의 실험 결과를 살펴보면 기존 방법은 물체에 대한 정보를 학습함으로써 물체 검출의 정확도가 저하됐음을 확인할 수 있다. 입력에 대한 학습량을 나타내는 c값을 보면 기존 방법의 경우 900프레임 동안 899를 학습했지만 제안하는 방법의 경우는

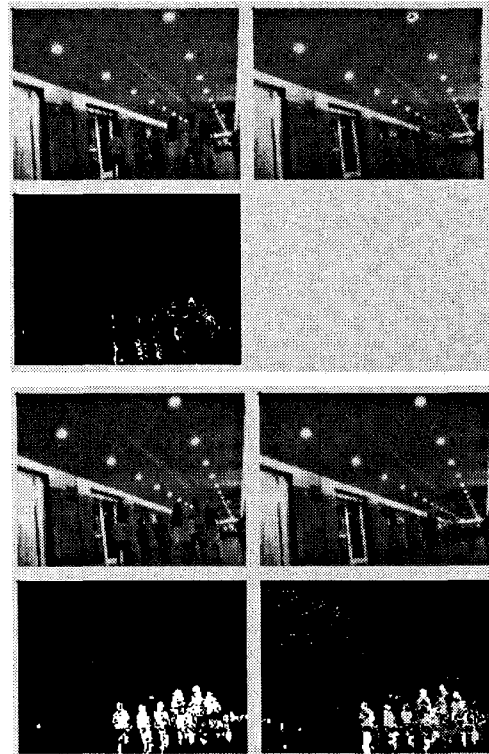


그림 9. 프레임 900에서 기존 검출 영상(위)과 개선 후 검출 영상(아래) 좌측 아래 영상들을 비교해 보면 물체 검출의 정확도가 크게 향상되었음을 확인할 수 있다.

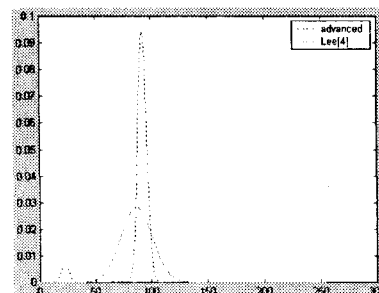


그림 10. 그림8의 표시된 점에서 학습된 배경 가우시안 모델의 pdf 비교

417만 학습하여 물체를 나타내는 색정보의 입력이 배경 모델에 영향을 주는 것을 차단했다.

그림11에서는 흔들리는 나뭇가지를 배경으로 멈춘 사람의 영역에 대하여 실험하였다. 물체 검출 결과를 보면 제안한 방법은 움직임이 많은 배경에 대하여 정지한 물체의 영역의 대부분을 검출

표1. 그림9의 표시된 점에서 각 가우시안의 매개변수 비교

	기존 방법			제안하는 방법		
	Y1	Y2	Y3	Y1	Y2	Y3
Mean	86.51	-	-	92.09	22.06	72
Variance	200	-	-	15.68	11.68	40
C	899	-	-	364	52	1
Weight	1	-	-	0.936	0.058	0.004

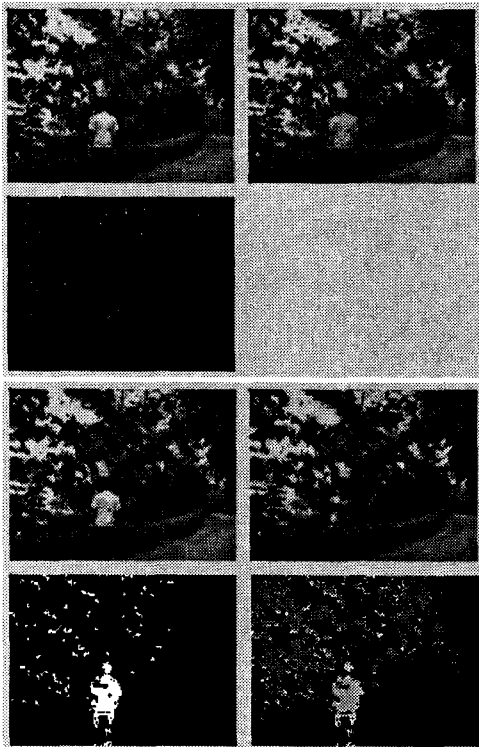


그림11. 프레임 5678에서 결과 영상(위)과 개선 후 결과 영상(아래)

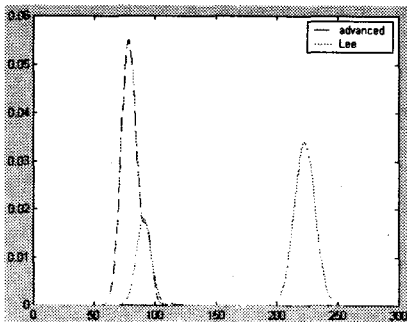


그림12. 그림 10에서 표시된 점에서 학습된 배경 가우시안 모델의 pdf 비교

한 것을 확인할 수 있다. 그러나 나뭇가지의 흔들림에 대한 처리는 기존의 방법이 약간의 우위를 나타낸다.

### 5. 결론 및 향후 과제

가우시안 혼합 모델은 복잡한 배경에 대하여 적응적 배경을 만들고 환경 변화를 학습해나가는 데 매우 효과적인 방법이다. 본 논문에서는 기존 방법으로 물체가 자주 지나다니거나 정지하는 배경을 학습하는데 있어 나타나는 문제를 분석하였고, 확률적으로 배경 모델을 학습하는 방법의 한계를 지적하였다. 이를 개선하기 위해 배경 모델이 없는 상태에서 조명의 변화를 제외한 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 대부분 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 방법에도 응용할 수 있다. 이러한 학습 방법은 기존의 시스템에 추가하는 면에서나 구현의 측면에도 장점을 가진다.

그러나 기존의 방법에 비해 계산량이 많은 단점을 지니며, 물체와 배경의 색이 비슷한 경우에는 여전히 검출의 정확도가 낮다. 하드웨어 속도의 향상과 최적화를 통해 계산량의 문제는 개선되겠지만 비슷한 배경과 물체를 구분하는 문제는 계속적인 연구가 필요하다.

### 6. 참고문헌

- [1] Gian Luca Foresti, Christian Micheloni, Lauro Snidaro, Paolo Remagnino, and Tim Ellis "Active Video-Based Surveillance System" IEEE Signal Processing Magazine, pages 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, pages 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.Stauffer and W.E.L. Grimson "Adaptive Background Mixture Models for Real-Time Tracking" Proc. Conf. Computer Vision and Pattern Recognition, vol 2, pages 246-252, June 1999
- [5] 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.
- [6] 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.
- [7] Dar-Shyang Lee "Effective Gaussian Mixture Learning for Video Background Subtraction" IEEE Trans. Pattern Analysis and Machine Intelligence Vol. 27, pages 827-832, May. 2005.
- [8] P. KaewTraKulPong 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] C. Eveland, K. Konolige, R. Bolles. "Background Modeling for Segmentation of Video-rate Stereo Sequences". In CVPR'98, pp.266-271, June 1998.
- [11] 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, pages 72-78, 1985.
- [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 2000