

혼잡한 환경에서 적응적 가우시안 혼합 모델을 이용한 계층적 객체 검출

Layered Object Detection using Adaptive Gaussian Mixture Model in the Complex and Dynamic Environment

이진형*, 조성원*, 김재민*, 정선태**

Jinhyung Lee, Seongwon Cho, Jaemin Kim and Sun-Tae Chung

* 서울시 마포구 홍익대학교 전기정보제어공학과

E-mail: pantera34@naver.com

** 서울시 동작구 숭실대학교 정보통신전자공학부

요 약

움직이는 객체를 검출하기 위해서 정확한 배경을 사용하기 위해 널리 사용되는 방법으로는 가우시안 혼합 모델이다. 가우시안 혼합 모델은 확률적 학습 방법을 사용하는데, 이 방법은 움직이는 배경일 경우와 이동하던 물체가 정지하는 경우 배경을 정확히 모델링하지 못한다.

본 논문에서는 확률적 모델링을 통해 혼잡한 배경을 모델링하고 객체의 계층적 처리를 통해 보다 정확한 배경으로 갱신할 수 있는 학습 방법을 제안한다.

키워드 : 객체 탐색, 계층, 혼잡한 환경 모델링, 가우시안 혼합 모델

Abstract

For the detection of moving objects, background subtraction methods are widely used. In case the background has variation, we need to update the background in real-time for the reliable detection of foreground objects. Gaussian mixture model (GMM) combined with probabilistic learning is one of the most popular methods for the real-time update of the background. However, it does not work well in the complex and dynamic backgrounds with high traffic regions. In this paper, we propose a new method for modelling and updating more reliably the complex and dynamic backgrounds based on the probabilistic learning and the layered processing.

Key Words : Object Detection, Layer, Complex Environment Modeling, Gaussian Mixture Model

1. 서 론

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. 움직이는 객체를 정확히 분리하기 위해서는 먼저 배경을 정확하게 모델링하는 것이 필요하다. 따라서 배경 모델은 실제 배경을 정확히 나타내도록 시간에 따라 지속적으로 학습되어야 한다.

본 논문에서는 배경 모델을 모델링하는 방법 중에 널리 사용되는 적응적 가우시안 혼합 모델을 사용하여 움직임이 있는 배경을 학습하고, 찾아진 객체 중에 이동이 멈춘 객체를 배경에 포함 시키는 과정에서 계층적인 방법으로 배경에

추가하는 방법을 제안한다.

2. 관련이론

2.1. 적응적 가우시안 혼합 모델

기본적으로 배경 화소는 하나의 화소 값을 가지며 조명의 변화나 영상 취득 장치의 잡음에 따라 약간의 변화를 보인다. 이러한 배경의 값은 적응적 가우시안 확률 분포로 효과적으로 모델링할 수 있으며[1], 가우시안의 평균과 분산은 매 장면마다 새로 입력되는 값을 학습함으로 환경의 변화에 적응해간다. 적응적 가우시안 혼합 모델은 영상의 각 화소 값을 K개의 가우시안을 혼합한 분포로 모델링한다. 가우시안 혼합 모델의 학습의 기본적인 개념은 식(1)과 같이 recursive filter를 통해 알 수 있다.

$$\theta(t) = (1 - \eta(t)) \cdot \theta(t-1) + \eta(t) \cdot \nabla(I(t); \theta(t-1)) \quad (1)$$

t에서의 모델 $\theta(t)$ 는 $\theta(t-1)$ 과 예측된 변화량 $\nabla(I(t); \theta(t-1))$ 의 $\eta(t)$ 에 따른 비율적인 합을 계산하여 변화하는 입력 $I(t)$ 에 적응하게 된다. $I(t)$ 는 시간 t에 특정 영상

접수일자 : 2007년 10월 30일

완료일자 : 2008년 5월 30일

이 논문은 2006년 정부(교육인적자원부)의 재원으로 학술진흥재단의 지원을 받아 수행된 연구임(KRF-2006-311-D00821)

좌표로부터 입력된 값이다. 각 가우시안 분포들의 평균과 분산, 혼합 비율은 위 식에 따라 학습되며 이를 통해 배경의 색 정보가 확률적으로 모델링된다.

이때 배경 모델의 수렴 속도와 안정도는 $\eta(t)$ 에 따라 결정된다[1,2,3]. 만약 현재 학습된 분포에 해당되지 않은 입력 값이 들어오면 새로운 가우시안 분포를 만들고 각 화소에서 정해진 가우시안 수의 한계치인 K 가 넘는 경우에만 이미 존재했던 가우시안 중에서 혼합 비율이 가장 낮은 것을 제거한다. 이런 과정을 통해 혼합 비율이 높은 K 개의 가우시안이 혼합 모델을 만들고 학습되어간다. 이러한 가우시안 혼합 모델이 현재의 배경 변화를 학습하여 적응해 적응하는 학습 속도 $\eta(t)$ 조절이 중요하다. 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. 최근 보고된 안정도를 유지하면서 수렴 속도를 향상시키는 방법으로 각 가우시안에 동일하게 적용했던 고정된 학습 비율을 각 가우시안에 맞게 적절하게 조절해 나가는 방법이 있다[4].

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

식(2)를 보면 어떠한 시간 t 에서 화소 (x,y) 의 학습된 모델은 K 개의 가우시안들의 w_k 에 따른 합으로 표현된다. 입력되는 화소 값 정보들을 K 개의 가우시안으로 모델링하는데, 각각의 가우시안은 배경의 화소 값 분포를 학습한 가우시안일 수도 있고 물체의 화소 값 분포를 학습한 가우시안일 수도 있다. 이 때 물체의 화소 값 정보를 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 σ_k 값을 가지게 된다. 이와 반대로 배경을 학습한 가우시안은 상대적으로 큰 w_k 과 작은 σ_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 입력되는 화소 값의 입력이 많아져 w_k 가 높아지고, 자주 나타나는 물체들이 비슷한 값을 가지면 가우시안의 σ_k 가 낮아지는 문제가 있다.

2.2. Temporary moving region mask

위에서 살펴본 것과 같이 단순히 학습 비율을 조절하는 것으로 객체를 찾기에는 어려움이 있다. 입력되는 화소는 식(3)과 같이 나타낼 수 있다.

$$I(t) = I(t-1) + I_{noise}(t) + I_{environmental\ change}(t) + I_{object}(t) \quad (3)$$

영상 획득 장치에 의한 잡음이나 배경 환경의 변화는 가우시안 혼합 모델에 효과적이고 정확하게 반영되어야 한다. $I_{object}(t)$ 는 배경 모델의 학습에 영향을 주어서는 안 된다. 이를 해결하기 위해 temporary moving region mask를 사용한다. 물체의 움직임이 연속적이라는 사실을 이용하여, 장면 차분을 통해 얻은 움직이는 물체의 일부 영역을 시간에 따라 효과적으로 간직한다면 배경의 값을 알지 못하는 상황에서도 움직이는 물체의 대부분의 영역을 찾아낼 수 있다.

$$T = \max\{I(t) - I(t-j)|, \forall j \in [1, M]\} \quad (4)$$

T 는 각 화소의 입력 값의 차의 최대값을 N 장면 동안 기억함으로써 현재까지 N 개의 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 T 값만으로 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.

$$S = \frac{M \sum_{j=1}^M I(t+j)^2 - (\sum_{j=1}^M I(t+j))^2}{M(M-1)} \quad (5)$$

S 는 각 화소에서 현재 이후의 M 개의 장면 동안 입력 값의 분산을 계산하여 문턱 값 이하가 되는 경우, 그 화소에서 움직임이 없는 것으로 처리한다. 현실적으로 미래의 입력을 안다는 것은 사실 불가능하다. 이와 같은 S 값을 이용하기 위해서 입력되는 모든 값이 M 장면 지연되어 사용된다. 지연되는 시간은 수 초 이내이며 자동차와 같이 빠른 움직임이 아닌 경우 관찰자는 이것을 거의 느껴지지 못한다.

이와 같이 T 와 S 를 이용하면 배경이 없는 상태에서도 움직임이 있는 영역을 얻어낼 수 있다[4]. 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 M 으로 나타내었다. 이 영역에서 들어오는 입력을 $I_{object}(t)$ 로 보고 이러한 입력에 대하여 가우시안 혼합 모델의 학습에 제한을 두었다.

$$M(x,y) = \begin{cases} 1 & T(x,y) > Threshold_T, S(x,y) > Threshold_S \\ 0 & otherwise \end{cases} \quad (6)$$

Temporary moving region mask는 $I_{environmental\ change}(t)$ 가 들어오는 영역을 제외해야 한다. 조명과 날씨에 따른 색 정보의 변화는 물체의 이동에 따른 변화에 비해 상대적으로 작은 변화 값을 가지기 때문에 이러한 문제는 적절한 $Threshold_T$ 의 선택을 통해 해결된다.

다음의 방법은 움직이는 영역을 찾아내어 가우시안 혼합 모델의 학습을 효과적으로 막아주나 배경에서 항상 움직임이 있는 영역은 움직이는 객체로 판단하여 학습을 제한한다.

3. 혼잡한 환경에서 가우시안 혼합모델을 이용한 계층적 객체검출

3.1 혼잡한 배경 검출 및 학습

반복된 움직임이 있는 배경은 Temporary moving region mask에서 제외해줘야 한다. 제외하기 위해서는 dynamic background rate r 을 계산한다.

$$r_{x,y,t} = (1-\lambda)r_{x,y,t-1} + \lambda f_{threshold_T}(|I_{x,y,t} - I_{x,y,t-1}|) \quad (7)$$

$$f_{threshold_T}(I) = \begin{cases} I < threshold_T & 0 \\ I > threshold_T & 255 \end{cases} \quad (8)$$

객체가 움직이는 영역과 반복적인 움직임이 있는 배경 모두 픽셀 값 변화의 크기와 분산이 크다. 그러나 객체가 많이 다니는 경우, 객체 간에 안정된 구간이 잠깐씩 나타나는 반면 반복된 움직임이 있는 배경은 픽셀 값이 안정된 구간이 거의 없는 것을 확인할 수 있다. 따라서 시간당 변화가 $Threshold_T$ 이상인 유효 움직임이 얼마나 자주 나타나는지를 학습하면 이 영역이 반복적인 움직임을 나타내는 영역인지 객체가 많이 다니는 영역인지 알 수 있다. dynamic background rate의 학습비율 λ 은 장면속도를 고려해서 정해야 하면 다른 학습의 학습비율에 비해 상대적으로 낮다. 움직임이 자주 나타나는 영역에서 r 값은 천천히 상승하며 일정 시간이 지나면 학습에 따라 반복적인 움직임이 있는 영역만 찾아낸다. 이러한 학습 방법은 나뭇잎의 흔들림과 같이 시간에 따라 움직임이 잠시 멈추는 특성을 가지는 영역의 경우에도 반복적인 움직임을 나타내는 영역을 효과적으로 유지할 수 있다. 이렇게 해서 학습된 $r(x,y)$ 의 값이 $Threshold_r$ 이상인 경우에는 이 영역을 dynamic background로 설정하고 temporary moving region mask에서 제외하여 가우시안 혼합 모델로 학습되도록 한다. 이를 위해 식(6)은 식 (9)와같이 수정된다.

$$M(x,y) = \begin{cases} 1 & T(x,y) > Threshold_T, S(x,y) > Threshold_S \\ & r(x,y) < Threshold_r \\ 0 & otherwise \end{cases} \quad (9)$$

3.2 계층적 객체 검출

객체를 배경과 구분하여 계층적으로 검출하기 위해서는 입력 영상의 화소 분석과 객체로 판단되는 지역 분석이 이루어진다. 기본적으로 화소 분석은 temporary moving region mask를 이용하여 $I_{object}(t)$ 를 배경 학습에 영향을 주지 않도록 하는 것과 동일하다. 하지만 temporary moving region mask를 이용한 적응적 가우시안 혼합 모델은 움직이던 객체가 멈추는 경우에 단순히 배경으로 학습해버리게 되므로 추적하던 객체의 정보는 사용할 수 없게 된다. 이러한 단점을 극복하기 위해 화소들의 T, S값을 이용하여 temporary moving region mask가 1인 영역을 하나의 덩어리로 만들어 객체의 지역 정보로 사용하여 계층에서 사용될 지역 분석을 행한다.

그림2에는 화소를 분석하여 덩어리로 만들어져 객체로 인식하여 객체의 지역 정보를 이용하여 멈춰있거나 이동하는지 판별한다. 화소가 배경의 가우시안일 때 T값의 변화가 큰 입력이 들어오게 되면 일단 객체로 판단한다. T, S가 일정한 값을 가져 temporary moving region mask가 $I_{object}(t)$ 로 보이는 화소를 덩어리로 묶어 객체 1의 지역 정보를 알 수 있게 된다. 시간이 지나서 객체가 멈추게 된다면 O1 즉 객체 1의 화소 값으로 정의하고 객체 1의 화소 값 가우시안을 함께 가진다. 덩어리로 묶인 객체도 멈춰있는 객체로 판단하여 배경에 계층적으로 더해져서 실제 배경으로 사용된다. 같은 방식으로 새로운 객체가 입력영상에 들어와서 이동하다가 멈추게 되어도 새로운 계층으로 배경 영상에 더해진다[5].

계층적으로 객체를 검출한 알고리즘은 그림 3과 같이 객체의 지역 정보 O에서 지역 내의 화소의 T와 S값을 토대로 멈춰 있다는 판단을 하면 배경에 객체를 계층적으로 추가한다.

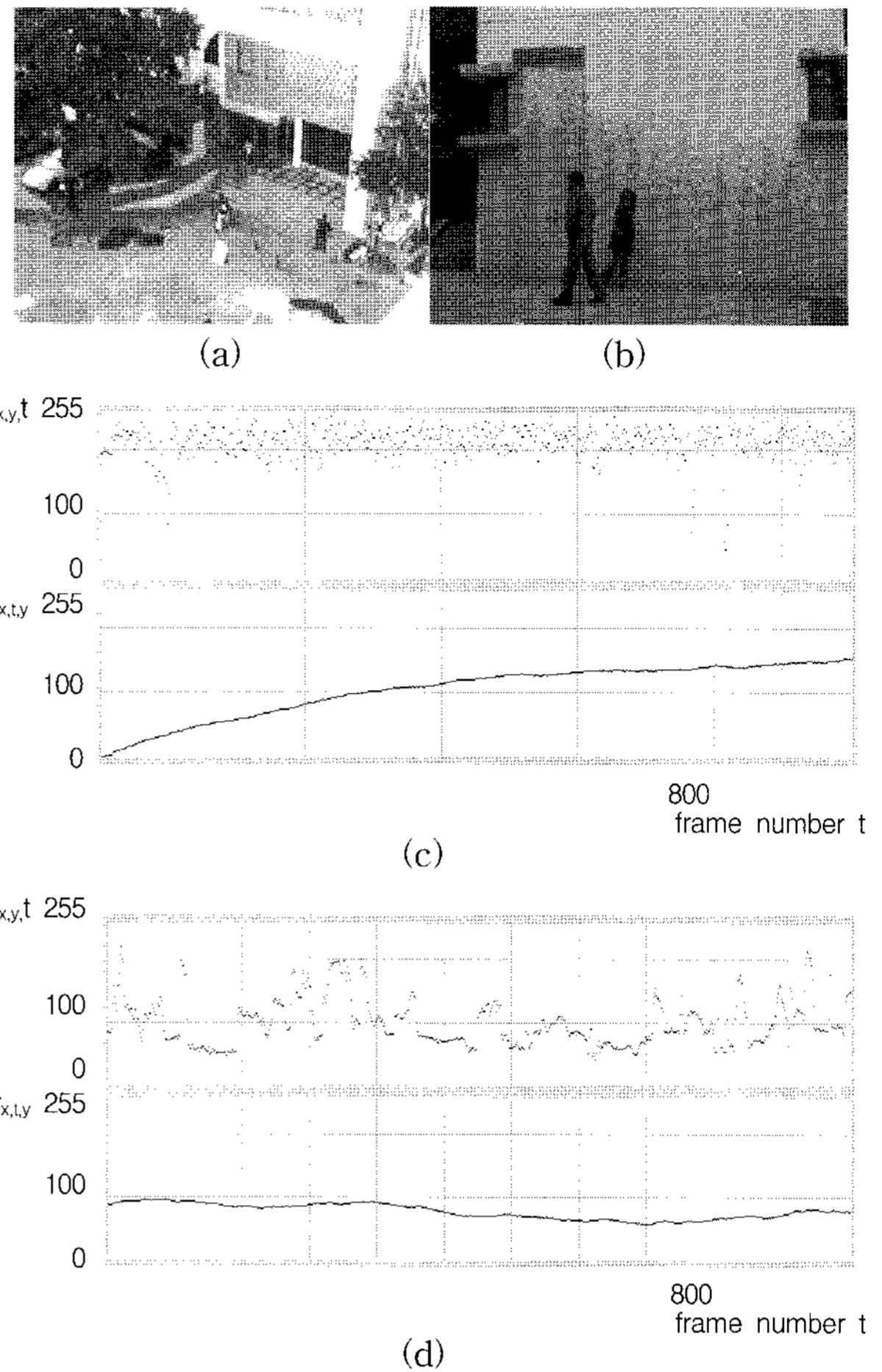


그림 1. 흔들리는 나뭇가지(a)(c)와 객체가 자주 다니는 공간(b)(d)에서 영상 입력값 I(t)(위)와 dynamic background rate r(아래)의 비교
Fig. 1. Comparison of image input value and dynamic background rate

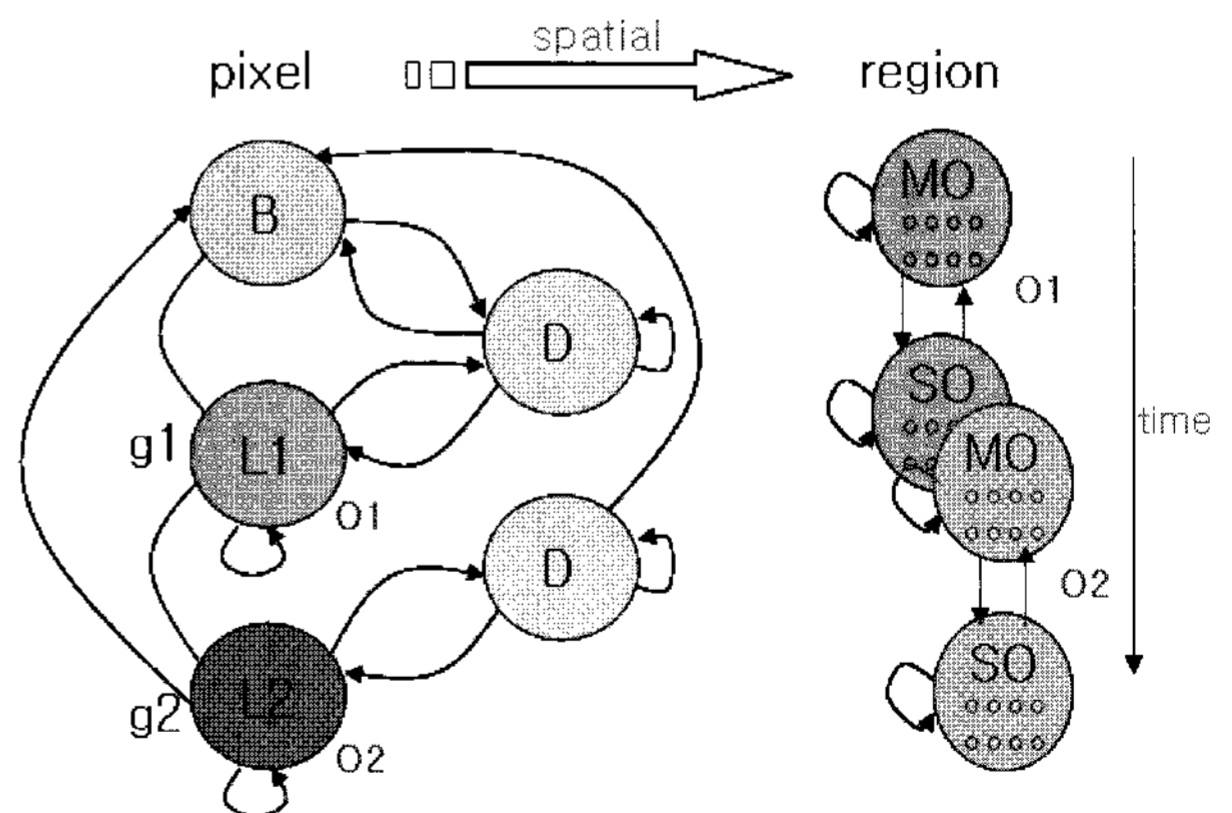


그림 2. 계층적인 객체 검출을 위한 화소 분석 및 지역분석의 개념도
Fig. 2. Diagram for pixel analysis and region analysis

```

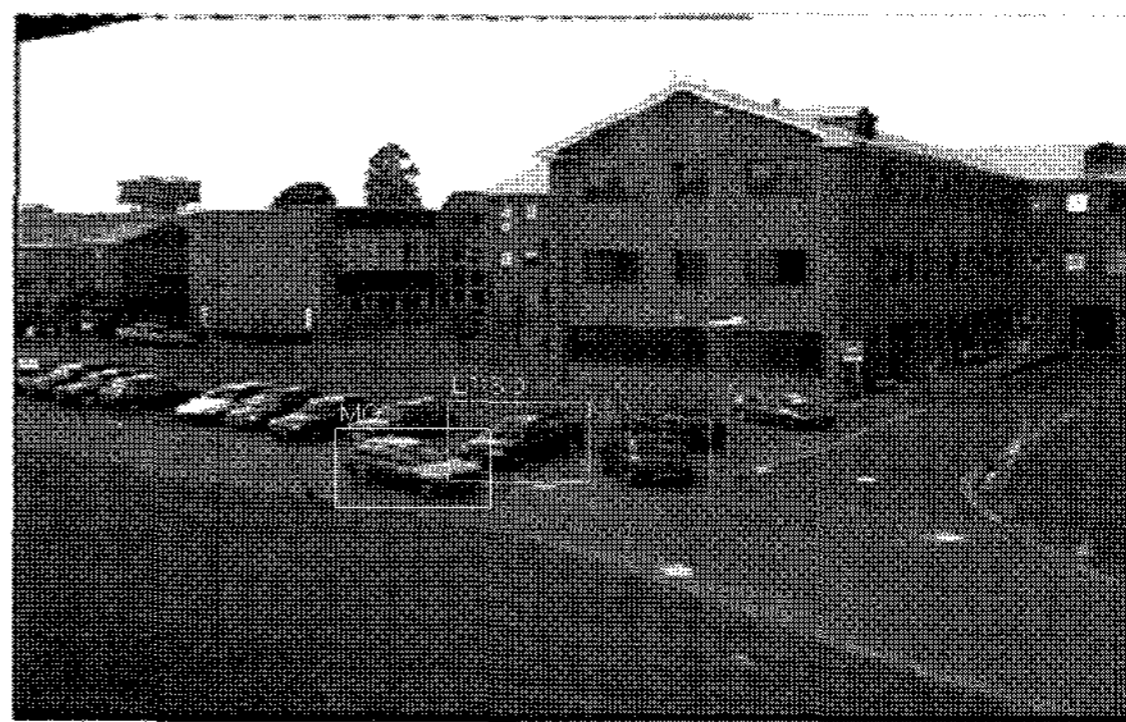
if (O = 움직임){ //T>ThresholdT;O내부의 모든 화소
    O -> 움직이는 객체
}
elseif(O = 멈춤){//S<ThresholdS;O내부의 모든 화소
    //O = O - (L(0)+L(1)+ ... +L(j))미리 특정 계층에
    포함된 화소값은 제거
    //어떤 계층에도 포함되지 않았으면 새로운 계층을
    생성
    if(O != 0){
        새로운 계층 L(j+1) = O 생성
        O -> 멈춰있는 객체
    }
}
else{//O내부의 화소가 T>ThresholdT인 것을 포함하고,
    S<ThresholdS인 것을 함께 포함하는 경우 각
    부분을 공간적으로 덩어리로 묶어 SO를 생성
    if(SO = 움직임){
        SO -> 움직이는 객체
    }
    if(SO = 멈춤){
        새로운 계층 L(j+1) = SO 생성
        SO -> 움직이는 객체
    }
    if(SO = 움직임 + 멈춤){
        SO -> 움직이는 객체
    }
}
}

```

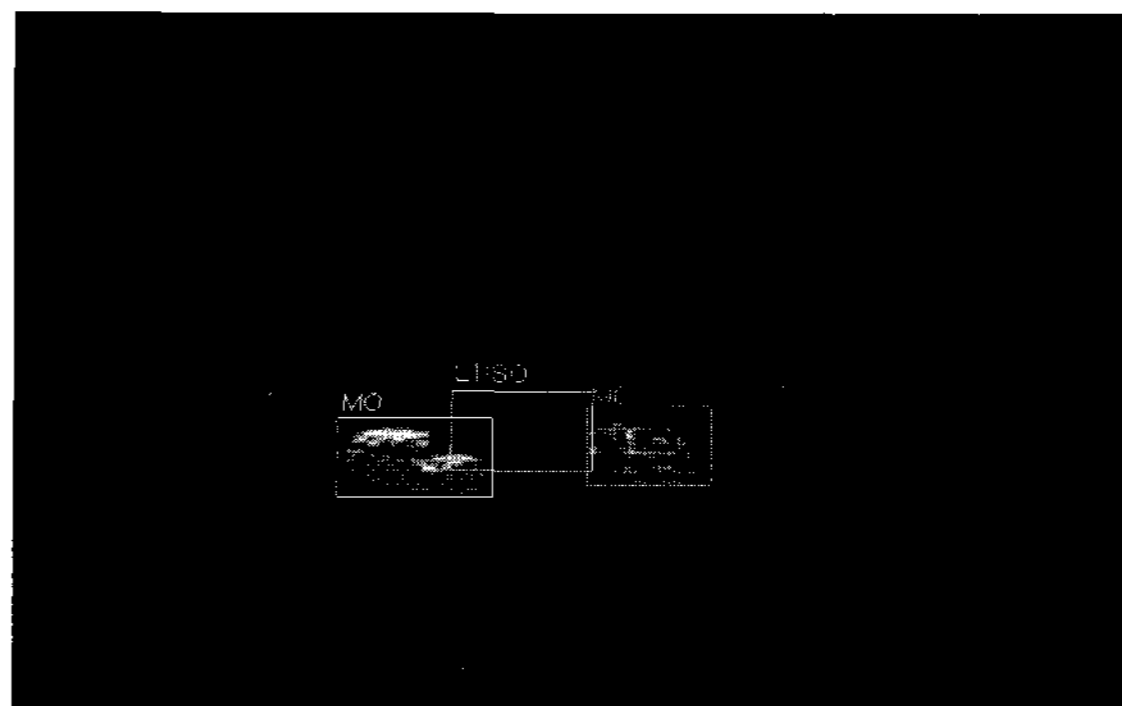
그림 3. 객체의 계층적 정렬 및 객체의 움직임 상태 분석 알고리즘

Fig. 3. Object moving analysis algorithm

그림 4는 계층적 객체 검출을 사용하여 객체를 배경에 새로운 계층으로 추가하여 실험한 결과를 나타낸다. MO는 움직임이 있는 객체를 나타낸 것이고, SO는 움직이던 객체가 멈추어 배경에 계층 L1이 추가된 것을 나타낸다. 그림 5는 본 논문에서 제안한 알고리즘인 혼잡한 환경에서 가우시안 혼합모델을 이용한 계층적 객체검출을 사용하여 객체를 배경에 새로운 계층으로 추가하여 실험한 결과를 나타낸다. 그림 5(b)를 보면 분수는 움직임이 있는 배경으로 분류하여 회색으로 표시되었고, 움직임 검출이 되지 않았음에도 객체로 표시되어 있는 부분은 움직이다 멈춘 객체로 계층 L1로 포함되었다.



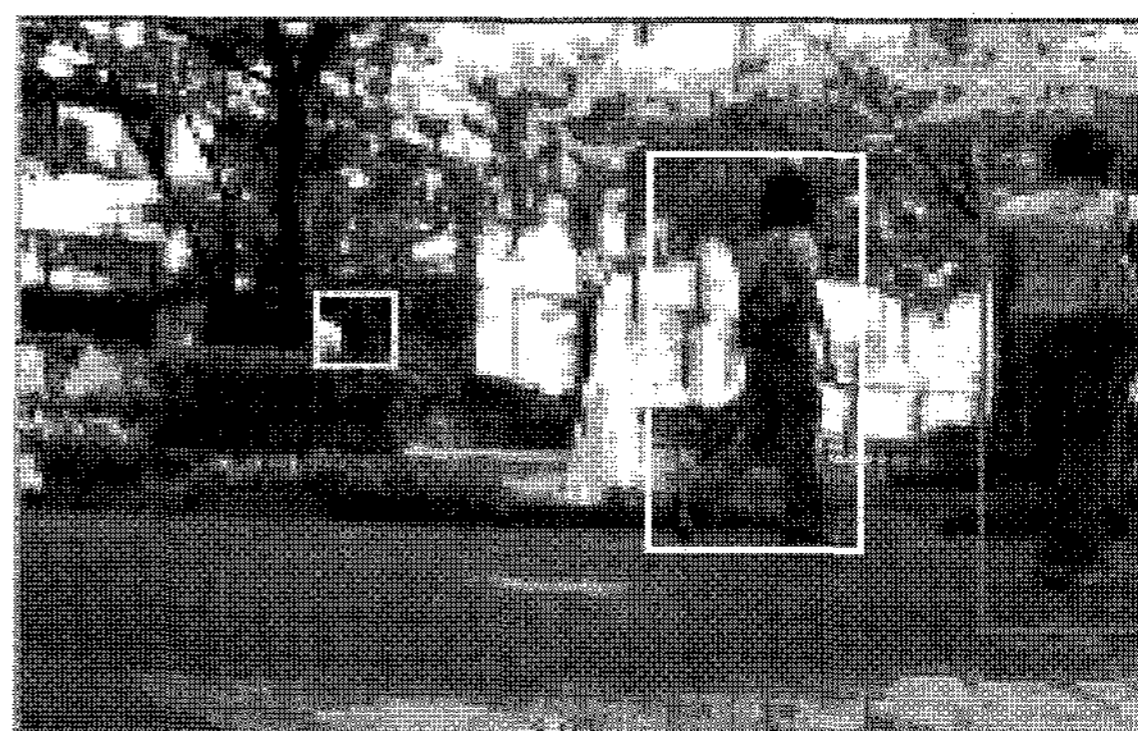
a) 입력 영상에서 결과 표시



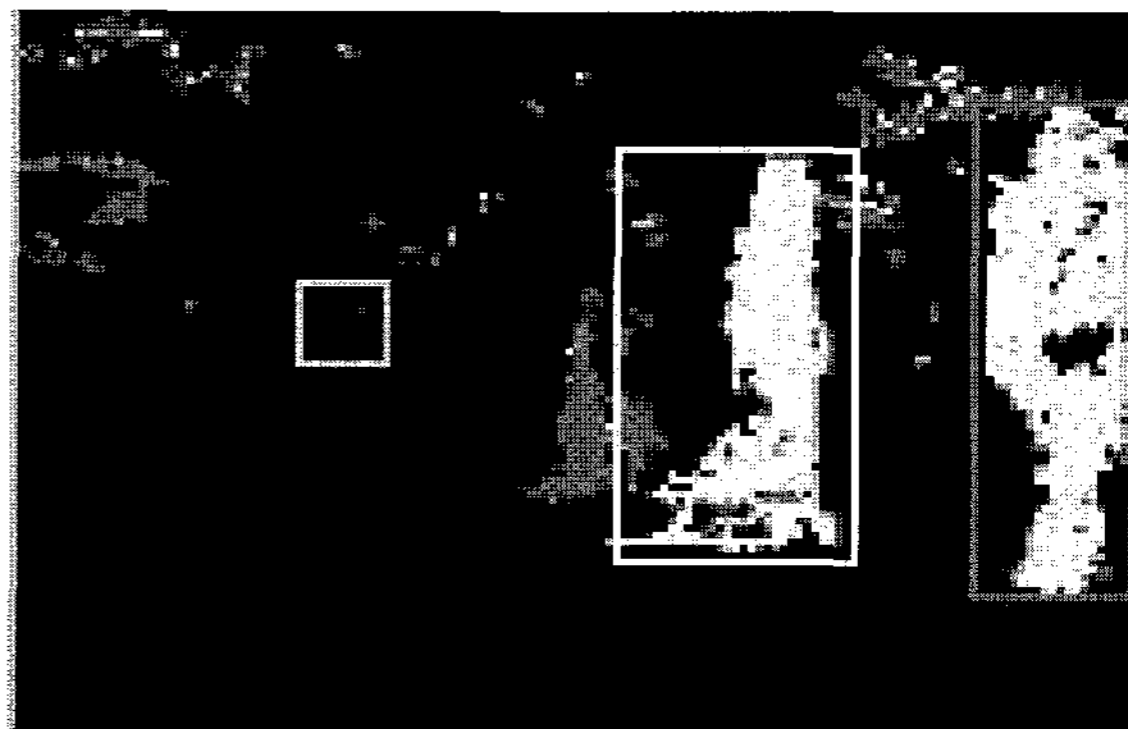
b) 움직임 검출영상에서 결과 표시

그림 4. 계층적 객체 검출

Fig. 4. Layered object detection



a)입력 영상에서 결과 표시



b)움직임 검출영상에서 결과 표시

그림 5. 혼잡한 환경에서 계층적 객체 검출

Fig. 5. Layered object detection in the complex and dynamic environment

4. 결과 고찰

본 논문에서는 기존 방법으로 움직이던 객체가 멈추게 되는 경우 단순한 배경처리로 객체의 정보를 사용하지 못하는 문제를 지적하였고, 움직임이 있는 배경을 모델링하지 못하는 문제를 지적하여 이를 개선하기 위해 배경 모델이 없는 상태에서 움직이는 물체만의 영역을 얻어내고 이러한 영역에 대한 학습을 제한하였다. 또한 움직이던 물체가 멈추는 경우 계층적으로 객체를 배경에 추가하여 객체의 정보를 사용할 수 있게 하였다. 또한 dynamic background rate로 움직이는 배경을 모델링하여 객체로 검출되지 않게 하였다. 그 결과 배경 모델의 정확도 향상을 통해 물체 검출의 정확도가 높아짐을 실험 결과를 통해 보였다. 이러한 학습 방법은 기존의 시스템에 추가하기에 용이하다. 또한 각 화소의 변화에 대한 객체의 지역 정보의 처리이므로 구현의 측면에서도 장점을 가진다.

저 자 소개



이진형(李辰炯)

2006년 : 홍익대 전자전기공학부 졸업

2008년 : 홍익대 전기정보제어공학과 대학원 졸업

Phone : 02 - 320 - 1634

Fax : 02 - 320 - 1110

E-mail : pentera34@naver.com

참 고 문 헌

- [1] 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.
- [2] 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*, September 2001.
- [3] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach", *Proc. 13th Conf. Uncertainty in Artificial Intelligence*, August 1997.
- [4] 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.
- [5] 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.



조성원(趙成元)

1982년 : 서울대 전기공학과 졸업.

1992년 : Purdue Univ. Electrical and Computer Engineering 졸업(공박)

현재 : 홍익대 전자전기공학부 교수

Phone : 02 - 320 - 1493

Fax : 02 - 320 - 1110

E-mail : swcho@wow.hongik.ac.kr



김재민(金載敏)

1984년 : 서울대 전기공학과 졸업.

1994년 : R. P. I. 전기공학과 졸업(공박).

현재 : 홍익대 전자전기공학부 부교수

Phone : 02 - 320 - 1634

Fax : 02 - 320 - 1110

E-mail : jaemin@wow.hongik.ac.kr



정선태(鄭善太)

1983년 2월 : 서울대학교 전자공학과 학사

1990년 12월 : 미국 미시간대학교(앤아버) 전자 및 컴퓨터 박사

1991년~현재 : 숭실대학교 정보통신전자공학부 교수