

# 고정된 비디오 카메라로부터 효율적인 배경영상 생성에 관한 연구

이동열<sup>0</sup>, 신옥선, 이창훈  
건국대학교 컴퓨터공학과

e-mail : ([wildeye](mailto:wildeye@konkuk.ac.kr), [wsshin](mailto:wsshin@konkuk.ac.kr), [chlee](mailto:chlee@konkuk.ac.kr))@konkuk.ac.kr

## The Study On Efficiently Estimating A Background Image From A Stationary Video Camera

Dong-Yeol Lee, Wook-Sun Shin, Chang-Hun Lee  
Dept. of Computer Science, KonKuk University

### 요 약

감시, 인지, 보안 시스템으로부터 얻은 비디오 영상에서 원하는 객체를 탐지해 내는 것은 매우 중요하다. 객체 추출 방법은 여러 가지가 있지만 가장 많이 쓰이는 방법이 배경을 이용하는 방법이다. 이때 실외 환경에 설치된 카메라의 경우 날씨, 시간에 따른 태양의 밝기등과 영상 내의 객체의 변화 량에 따라서 효율적으로 적응할 수 배경 추출 알고리즘이 필요하다. 본 논문에서는 빠르고 정확하게 배경을 얻기 위한 기본적인 방법인 평균값과 최빈값을 이용한 방법을 혼합하여 영상의 변화 량에 따른 빠르고 정확한 배경을 추출하는 알고리즘을 제안하고자 한다.

### 1. 서론

교통 정보를 알기 위해 도로상에 설치된 카메라나 침입탐지를 위한 CCTV 등 요즘에는 여러 많은 카메라들이 설치 되어 있다. 하지만 실제 이러한 카메라에서 자동으로 정보를 얻기 위한 시스템은 많이 미흡한 실정이다. 예를 들어 교통현황이나 정보 등을 파악하는 시스템, 지하철, 상가, 무인 시설물 등에 설치된 보안시스템 등이 있다. 본 논문에서는 고정된 비디오 카메라에서 정보를 얻기 위한 가장 기본적인 방법인 배경 추출에 관해 논의 하고자 한다.

배경이미지를 생성하는 것은 픽셀의 누적 데이터를 이용하면 얻을 수 있다. 하지만 만들어진 배경이 모든 영상에서 고정되게 사용될 수는 없고, 상황에 따라 빠르고 정확한 배경이 생성되는 것이 현재 객체를 추출해 내는데 훨씬 유리 하다. 그 이유는 첫째, 카메라를 통해 들어온 정보는 실제 많은 노이즈를 가지고 있다. CCD 센서의 출력을 증폭하는 Nonlinear 요소를

포함하는 회로로 인하거나, 아날로그 신호를 디지털 신호로 변형될 때 생기는 quantization error 등의 노이즈를 포함한다. 둘째, 실외에 설치된 카메라의 경우 날씨에 따라서 밝기의 변화가 심하다. 구름이나 그늘 등의 여러 가지 요소로 배경의 변화가 생긴다. 셋째, 카메라의 진동 등으로 생기는 화면의 변화이다. 바람 또는 인위적으로 카메라의 위치 변화는 실제 배경영상 추출에 많은 영향을 끼친다. 마지막으로 인위적인 물체에 인한 배경의 변화이다. 정차된 차량이나 물건의 변화, 사람의 움직임 등 인위적으로 조작된 배경의 변화로 인해 배경을 추출하는데 많은 문제점이 발생할 수 있다.[3][4]

본 논문에서는 위와 같은 문제점을 고려하여 기본적인 방법인 평균값과 최빈값을 혼합한 배경추출 방법을 사용하여 빠르게 정확하게 초기배경을 추출하고 환경에 따른 적응적인 배경 갱신을 위한 방법에 대해 제안 하고자 한다. 2 장에서는 기본적인 방법인 평균값과 최빈값을 이용한 방법에 대해 알아보고 3 장에서는 제안하고자 하는 방법에 대해 알아보고, 4 장에서는 이러한

<sup>0</sup> 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성지원사업의 연구결과로 수행되었음

방법을 사용하여 생성된 배경에 대해 MSE 값을 통해 배경의 정확도를 측정하여 결과를 나타내었다. 마지막 5 장에서는 결론과 향후 연구방향에 대해 논해 보고자 한다.

2. 관련 연구

영상에서 배경을 구하는 가장 기본적인 방법은 각 픽셀마다의 누적 데이터를 이용하여 배경영상을 구성하는 것이다. 이 방법은 각 픽셀마다의 누적 데이터의 평균을 이용하는 방법과 각 픽셀의 누적값들의 빈도수를 계산하여 최빈값을 이용하는 방법 두 가지로 나뉘게 된다. [1][2]

먼저 최빈값을 이용한 방법을 알아보면, 픽셀의 위치를 (x,y) 라고 했을 경우 아래의 ①식으로 배경영상을 얻을 수 있다.

$$B_{N-1}(x,y) = \text{mode}\{I_k(x,y) | k = 0, \dots, N-1\} \dots \text{①}$$

$B_{N-1}(x,y)$  는 얻어진 배경픽셀을 의미하고,  $I_k(x,y)$  가 N 개의 연속된 이미지에서 k 번째 영상값을 사용하는 것을 의미한다. 그리고

$$\text{mode}\{I_k(x,y)\} = \max\{\rho_{x,y}(r) | 0 \leq r \leq (L-1)\}$$

$\rho_{x,y}(r)$  는 (x,y)에서 가질 수 있는 픽셀 값의 개수를 의미하고 L 은 픽셀의 밝기 값의 범위(0~255)를 나타낸다.

다음으로 평균값을 이용한 방법을 알아보면 위의 방법과 같이 픽셀의 위치를 (x,y) 라고 했을 경우, 아래의 ②식으로 배경영상을 얻을 수 있다.

$$B_{avg,n}(x,y) = \alpha * B_{avg,n-1}(x,y) + (1-\alpha)B_n(x,y) \dots \text{②}$$

$B_{avg,n}(x,y)$  는 얻어진 배경픽셀을 의미하고,  $\alpha$  값은 최근 페이지를 얼마나 많은 비율로 배경생성에 영향을 줄 것인가를 결정하는 인자이다.

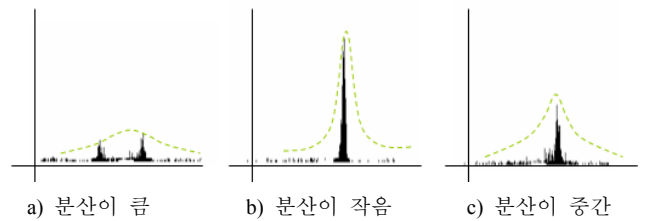
위의 두 방법은 서로 다른 특성을 가지고 있다. 속도면에서 본다면, 평균을 이용하는 방법의 경우는 최빈을 이용하는 경우보다 속도가 훨씬 빠르다. 메모리 면에서도 평균을 이용하는 방법이 훨씬 우위에 있다. 최빈을 이용하는 방법의 경우는 누적 데이터에 대한 자료를 모두 보관하고 있어야 하기 때문 메모리도 많이 차지하고 속도도 느리다. 하지만 배경의 질면에서 따졌을 경우, 평균값을 이용하는 경우 배경영상과 각 픽셀값을 비교해 보면 Threshold 값이 커야지만 잡음을 줄일 수가 있게 된다. 즉 배경빼기 방법을 사용하여 객체를 추출 시에 잡음 발생빈도가 상당히 높다. MSE(mean square error)값을 비교해 보아도 차이가 있음을 알 수 있다. 반대로 최빈값을 이용하는 경우 실제 배경영상에 가장 근접한 배경영상을 생성해 낼 수 있다. 즉 MSE 값이 비교적 낮게 측정 된다. 다른 측면에서 볼 때, 평균값을 이용하는 경우 영상의 변화가

심한 경우에도 크게 배경이 변화하지 않고, 실제로 크게 차이가 나지 않는다. 하지만 최빈값을 이용하는 경우에서 예를 들어, 차량이 정체가거나 잠시 정차하는 경우, 혹은 사람이 한자리에 머무는 경우 등의 상황에서는 좋지 않은 결과를 나타낼 수 있다.

위와 같이 두 가지의 방법은 각기 다른 특징을 가지고 있다. 이러한 두 가지 배경 생성 방법의 장점을 혼합한다면 훨씬 정확하고 상황에 따른 최적의 배경을 생성할 수 있는 방법을 생각해 볼 필요가 있다.

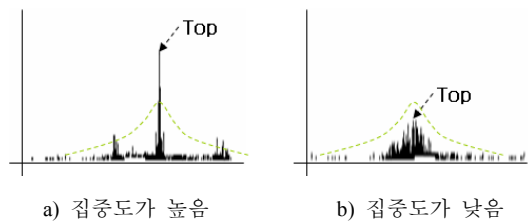
3. 제안된 배경 생성 방법

배경 생성시 하나의 픽셀 P(x, y)의 누적 값 분포를 살펴보면 몇 가지 특성을 알 수 있다. 아래의 그래프는 짧은 시간에 모아진 누적 데이터의 분포를 그래프로 나타낸 것이다.



[그림 1] 누적 데이터의 분포

가로축은 픽셀의 밝기 값의 범위(0~255)이고, 세로축은 누적된 값의 개수이다. 그래프는 각기 다른 분포를 가지며 이들간의 차이는 분산 값의 차이로 구분할 수 있다. [그림 1]의 a)는 가장 분산이 큰 경우로 다양한 값의 분포하고 집중도가 가장 낮다. 이는 동영상의 하나의 픽셀에서 픽셀의 밝기 값의 변화가 크다는 것을 의미한다. 즉 배경 픽셀과 그 외 픽셀이 골고루 분포하고 있는 경우이다. b)는 픽셀의 분포가 한두 밝기 값에 집중되어 있어서 가장 작은 분산 값을 보인다. 이는 픽셀의 값이 시간에 따라 변화가 적다는 것을 의미하는 것으로 배경이 배경 픽셀일 가능성이 가장 높다. a)의 경우 최빈값이 배경 픽셀일 것인 확률이 낮기 때문에 누적 픽셀의 평균값을 이용하면 좋고 b)의 경우는 최빈값이 배경일 확률이 높기 때문에 배경픽셀로 사용한다. 따라서 분산값은 평균값 또는 최빈값 중 어떤 값을 사용하는지 판단하는 가장 중요한 요소이다.

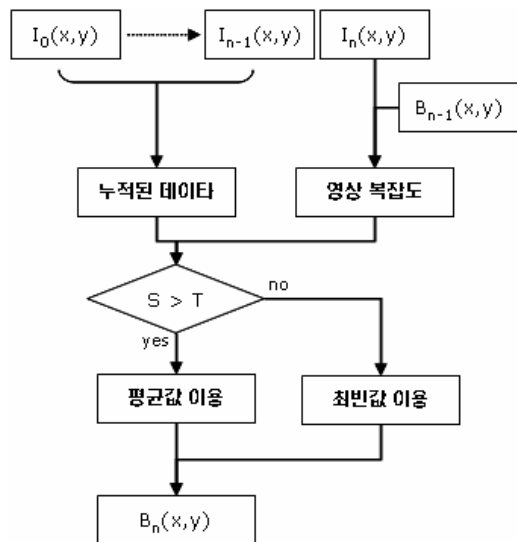


[그림 2] 분포의 형태

하지만 픽셀에 대한 누적 값은 확률적 분포 모델을 정확하게 따르고 있지 않는 경우가 있다. 위의 [그림

2]는 비슷한 분산값을 가지지만 분포형태는 다르다는 것을 알 수 있다. 이러한 이유로 분산값 만으로 최빈/평균값의 사용을 판단할 수 없으므로 가장 높은 빈도수인 top 의 값을 사용하여 보정되어야만 한다. 분산이 작더라도 top 의 값이 작다면 실제로 일정 지역에 값이 몰려 있을 수 있다. 이러한 경우에도 어떠한 값이 더 정확한 값인지 알기란 쉽지 않다. 이러한 두 개의 요소(분산값과 top)를 이용하여 다음의 과정을 거쳐 정확하게 배경을 생성하였다.

동영상 내의 각 프레임을 사용하여 각 픽셀에 대한 누적 밝기값을 구한다. 누적된 데이터는 현재프레임 이전 프레임들의 데이터로 구하고, 영상의 복잡도는 이전 프레임의 데이터로 생성된 배경과 현재 프레임을 배경빼기를 한 후 threshold 값을 통해 필터링을 하면 객체가 남는다. 그러한 객체가 가지는 픽셀 수를 세어 구하게 된다.



[그림 3] 배경 생성 과정

기존의 누적된 자료를 통해서 분산과  $w$  를 구하고,  $S$  와  $L$  값은 아래의 식으로 계산된다.

$$S = \left[ \frac{1}{n-1} \sum_{i=0}^{n-1} (x_i - m)^2 \right] / top \quad \dots \textcircled{3}$$

$$T_{IC} = (IC / \alpha) + \beta \quad \dots \textcircled{4}$$

④번 식의  $T$  값은 평균값과 최빈값 중 어떤 값을 선택할 지에 대한 Threshold 가 된다. IC(image complexity)의 값은 복잡도를 뜻한다. 영상의 복잡도에 따라 Threshold 를 구하게 되는 이유는 복잡도가 작은 영상은 대체적으로 최빈값을 사용하는 것이 우수하고, 복잡도가 큰 영상은 객체가 많기 때문에 평균값을 쓰는 것이 대체로 유리한 상황이 많다. 그래서 적절하게 영상의 복잡도를 threshold 값으로 적용함으로써 배경 생성에 정확성을 기하였다.

$\alpha$  와  $\beta$  값은 실제 영상에 따라 최적값이 약간씩 변

화가 있지만, 실험을 통해서 여러 가지 영상의 종류와 영상의 복잡도에 따라 적절한 값을 찾아 낼 수 있었다.

## 4. 실험 결과

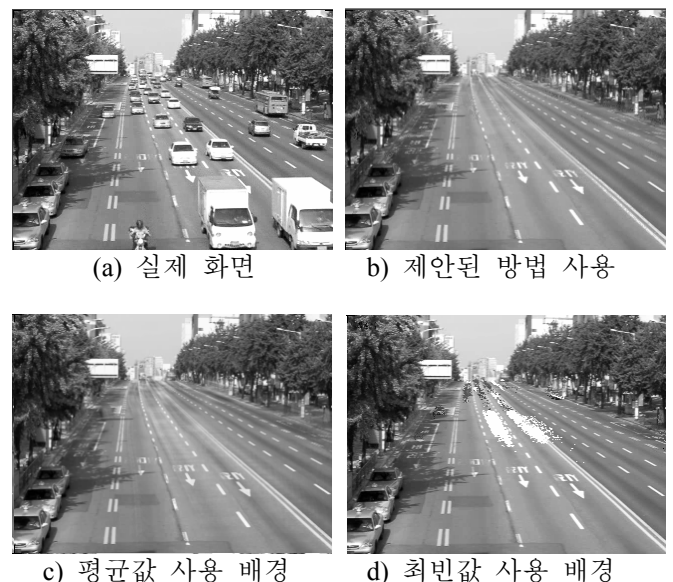
### 4.1 실험 환경

제안된 배경생성에 대한 실험은 차량과 사람이 다니는 장소에서 직접 캠코더로 촬영한 자료를 바탕으로 실험을 하였다. 틀은 Visual Studio .NET 을 사용하였고, 사용 언어는 MFC 를 사용하여 제작하였다. 동영상은  $320 \times 240$  의 RGB 24bit, 15frame/sec 컬러영상을 사용하였다.

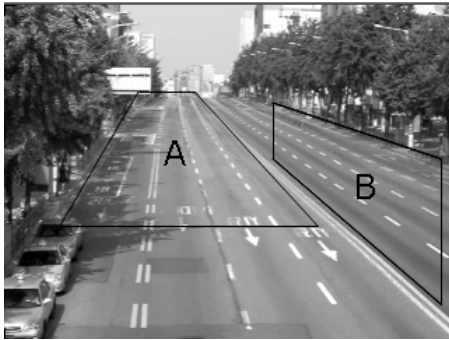
### 4.2 실험 결과

본 논문에서 제안한 방법에 대한 결과를 측정하기 위해 평균값을 사용한 배경과 최빈값을 사용한 배경 그리고 제안된 방법을 사용한 배경 등 3 가지의 배경을 실제 영상에서 선택한 배경과 차이를 통해서 결과를 분석하였다. 차이를 구하는 방법은 MSE(mean square error)를 사용하였다. MSE 값은 두 영상의 차이를 수치로 나타내었으므로 수치가 낮을수록 영상의 차이가 작다.

여러 가지 상황을 고려한 6 가지의 영상을 촬영하여 실험 데이터로 사용하였다. 영상은 차량이 다니는 도로에서 촬영한 3 가지의 영상과 사람이 다니는 길에서 촬영한 3 가지의 영상을 각기 복잡도가 다르게 영상을 추출하여 영상을 만들었다. 제안된 방법에서 사용된  $\alpha$  와  $\beta$  값은 여러 가지 조건에서 테스트한 결과  $\alpha = 200$  ,  $\beta = 80$  에서 최적의 배경을 생성하였다.



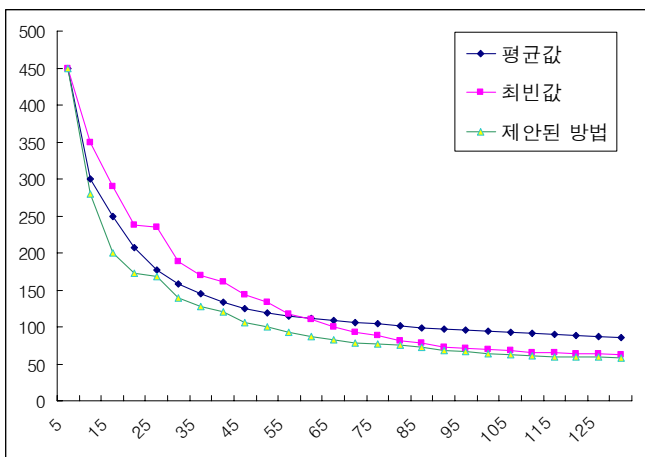
[그림 4] 테스트에 사용된 영상



[그림 5] 제안된 방법으로 만든 배경 영상

위의 [그림 4]은 약 50 프레임 정도에서 차량을 촬영한 영상중의 하나이다. a)그림은 실제 영상이고 b) c) d)는 배경을 추출한 화면이다. [그림 5]을 분석해 보면 A, B 부분은 양쪽의 복잡도가 다르게 나타난다. A는 차량이 많이 지나가고 B 부분은 차량이 상대적으로 적게 다닌다. 이러한 부분이 실제 배경 생성에 그대로 영향을 미친 것을 볼 수 있다. A는 분산이 높고 top이 낮은 형태가 취해져서 평균값의 배경을 따라 가고, B는 분산이 낮고 top이 높은 형태를 취해져서 최빈값이 사용 되어졌다. 이렇듯 제안된 방법을 사용하면 여러 가지 상황에 따른 더 좋은 배경을 생성하게 되는 것을 알 수 있다. MSE 값으로 평가해 보았을 때에도 평균이나 최빈을 사용한 배경보다 더 좋은 결과를 나타내고 있다.

위와 같은 방법으로 촬영된 각각의 영상을 5 프레임 단위로 MSE 값을 측정하여 기록하였고, 각각 영상에서 추출한 MSE의 평균값으로 종합하여 아래의 그래프로 나타내었다.



[그림 6] 프레임에 따른 MSE의 변화량

위의 [그림 6]에서 가로축은 프레임이고, 세로축은 MSE 값이다. 처음 5 프레임까지는 그다지 차이가 없지만 그 후 60 프레임까지 평균이 우세를 보이다 그 뒤로는 최빈값이 우세를 보이는 경향을 보인다. 하지만 제안된 방법은 두 가지의 장점을 살려 더 낮은 MSE 값을 보인다. 여러 영상에 따라 평균을 사용한 방법이 좋은 경우도 있고, 최빈값을 사용한 방법이 좋은 경

우도 있다. 하지만 위의 방법을 사용하면 대부분의 경우에 더 좋은 결과를 나타내었다.

### 5. 결론 및 향후 연구 방향

본 논문에서는 고정된 비디오 카메라에서 얻어진 동영상으로부터 배경영상을 효율적으로 얻어오는 방법에 대해서 논의하였다. 제안된 방법은 평균값을 이용할 때의 빠른 배경생성과 최빈값을 이용할 때의 정확성의 장점을 모두 가진다. 효율적인 배경영상의 생성은 자동화된 교통, 보안, 인지 시스템에서의 객체 추출을 위한 배경 빼기와 같은 다른 응용의 결과에 많은 영향을 주기 때문에 중요하다고 할 수 있다.

향후 연구 방향으로는 극심한 움직임이 발생하거나 장시간의 움직임의 정체가 발생하는 동영상에서의 원활한 배경추출을 위한 개선된 방법의 제안이 필요하다.

### 참고문헌

- [1] Lai, A.H.S.; Yung, N.H.C. "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence" IEEE Circuits and Systems, Volume 4
- [2] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder:Real-time Tracking of the Human Body," IEEE Trans. vol. 19, no. 7.
- [3] 권영탁, "영상처리를 이용한 공간 교통정보 수집", 명지대 박사 학위 논문
- [4] 도명환, "재귀적으로 modeling 한 두 가지 배경화면을 이용하는 영상 감시 시스템 설계", 연세대 석사 학위 논문
- [5] Ismail Haritaoglu, David Harwood, Larry S. Davis, "A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance" IEEE Pattern Recognition Volume 4, 3-7 Sept. 2000 Page(s):179 - 183 vol.4
- [6] C. Stauffer, W.E.L. Grimson, "Adaptive background mixture modelsfor real-time tracking", Proc. of CVPR 1999, pp. 246-252.
- [7] M. Seki, T. Wada, H. Fujiwara, K. Sumi, "Background detection based on the cooccurrence of image variations", Proc. of CVPR 2003, vol. 2, pp. 65-72.