

## 동적 환경에서의 효과적인 움직임은 객체 추출

### An effective background subtraction in dynamic scene

한재혁, Jae-Hyek Han\*, 김용진, Yong-Jin Kim\*\*, 유세운, Sae-Woon Ryu\*\*,  
이상화, Sang Hwa Lee\*\*, 박종일, Jong-Il Park\*\*\*

**요약** 컴퓨터 비전 분야에서 전경을 추출하기 위한 영역 분할(segmentation) 방법에 대한 연구가 활발히 진행되어 왔다. 특히, 전경이 배제된 배경 영상과 현재 프레임의 차이를 이용하여 전경을 추출하는 배경 차분(background subtraction) 방법은 요구하는 계산량에 비해 우수한 품질의 전경 추출이 가능하므로 실시간 처리가 필요한 비전 시스템에 다양하게 응용되고 있다. 그러나 배경 차분 방법만을 이용하여서는 배경이 동적으로 변하는 환경에서 정확한 전경을 추출해 내지 못하는 단점이 있다.

본 논문에서는 정적인 배경과 동적인 배경이 공존하는 환경에서 영역 분할을 효과적으로 수행하는 방법을 제안한다. 제안된 방법은 정적인 배경 영역에 대해서는 기존의 배경 차분 방법을 이용하여 전경을 추출하고, 동적인 배경 영역에 대해서는 깊이 정보를 이용하여 전경을 추출하는 하이브리드 방식을 사용한다. 정적인 배경에 동적인 영상을 프로젝터로 투영하는 환경에서 제안된 방법의 효율성을 검증하였다.

**Abstract** Foreground segmentation methods have steadily been researched in the field of computer vision. Especially, background subtraction which extracts a foreground image from the difference between the current frame and a reference image, called as "background image" have been widely used for a variety of real-time applications because of low computation and high-quality. However, if the background scene was dynamically changed, the background subtraction causes lots of errors.

In this paper, we propose an efficient background subtraction method in dynamic environment with both static and dynamic scene. The proposed method is a hybrid method that uses the conventional background subtraction for static scene and depth information for dynamic scene. Its validity and efficiency are verified by demonstration in dynamic environment, where a video projector projects various images in the background.

**핵심어:** *Background Subtraction, Image Segmentation, Computer Vision*

본 논문은 2008 년 한국전자통신연구원의 디지털 교과서 및 u-러닝 활성화를 위한 요소 기술 개발 과제의 지원으로 수행되었습니다

\*주저자 : 한양대학교 전자컴퓨터통신공학과 e-mail: [turbostar@mr.hanyang.ac.kr](mailto:turbostar@mr.hanyang.ac.kr)

\*\*공동저자 : 한양대학교 전자컴퓨터통신공학과 e-mail: [ladinjin@mr.hanyang.ac.kr](mailto:ladinjin@mr.hanyang.ac.kr)

\*\*공동저자 : 한양대학교 전자컴퓨터통신공학과 e-mail: [ryuseun@mr.hanyang.ac.kr](mailto:ryuseun@mr.hanyang.ac.kr)

\*\*공동저자 : 서울대학교 전기공학부 e-mail: [ish@ipl.snu.ac.kr](mailto:ish@ipl.snu.ac.kr)

\*\*\*교신저자 : 한양대학교 전자컴퓨터통신공학과 e-mail: [jipark@hanyang.ac.kr](mailto:jipark@hanyang.ac.kr)

## 1. 서론

컴퓨터 비전 분야에서 전경 영역을 추출하기 위한 영역 분할(segmentation) 기법들은 꾸준히 개발되어 왔다. 이중에서도 비교적 계산량이 적으면서 고품질의 전경 영역을 추출할 수 있는 기술로서 배경 차분(background subtraction)방법이 많이 이용되고 있다. 배경 차분 방법은 배경을 이미 가지고 있는 영상 성분으로 가정하고, 전경 영역을 배경과의 차이를 비교하여 분리하는 방법으로 지금까지 다양한 방법들이 제안되어왔다. Horparasert 등은 배경 모델링 시 통계학적인 방법을 이용하였다[1]. 이 방법은 시간 복잡도가 낮은 장점이 있으며, 영역 분할 뿐 아니라 그림자 영역을 분리하는 데에도 우수한 결과를 보여준다. Stauffer 와 Grimson 은 적응적 배경 혼합 모델을 제안하였다[2]. 이 방법은 다중 가우시안 모델을 사용하여 다양한 장면 변화에도 전경 영역을 추출할 수 있다는 장점이 있으나 높은 시간 복잡도로 인하여 고해상도 영상에 대해서는 실시간 처리가 어렵다. 이 밖에도 다양한 배경 차분 방법들이 제안되었다[3-6].

그러나 일반적으로 동적인 배경을 포함하는 다양한 환경에서 배경 차분 방법을 영상에 적용하기에는 많은 문제점이 있다. 왜냐하면 기존의 배경 차분 방법에서는 배경 영상을 고정된 정보로 가정하기 때문이다. 예를 들어 프로젝터의 영상이 투영되고 있는 실내 환경에서 배경 차분 방법을 사용하여 전경 영역을 추출하는 경우 배경 이 시간이 지남에 따라서 수시로 변하게 되어 올바른 전경 영역을 추출해내기가 어렵다. 따라서 본 논문에서는 정적인 영역과 동적인 영역이 혼합되어 있는 다양한 환경에서도 효율적으로 전경 영역을 추출해 낼 수 방법을 제안한다.

본 논문에서 제안하는 방법은 정적인 배경 영역에 대해서는 Horparasert 등이 제안한 방법[1]을 시스템에 맞게 최적화한 후 적용하여 전경 영역을 추출하고, 동적인 배경 영역에 대해서는 스테레오 카메라를 사용한 깊이 정보를 이용하여 전경을 추출하는 혼합형 배경 차분 방식을 사용한다. 특히 스테레오 카메라의 기존 영상에서 정적인 영상 영역은 배경 차분 방법을 사용하고, 동적인 영역에서는 스테레오 카메라의 참조영상을 이용하여 깊이정보 계산에 활용하는 하이브리드 방식을 사용하여 고속화 및 최적화 한다(그림 1 참조). 따라서 그림 2 와 같이 동적인 영역을 포함하는 환경에 대해서도 올바른 전경 영역 추출 결과를 얻을 수 있다.

## 2. 전경 영상과 배경 영상의 분리

그림 1 은 본 논문에서 제안하는 전경 영역 추출 과정의 전체적인 흐름을 나타낸다. 깊이 정보와 색상 정보를 동시에 획득할 수 있는 스테레오 카메라를 이용하여 프로젝터 영상이 화면이 투영되는 동적인 영역을 포함하여 영상을 획득한다. 획득한 영상을 본 논문에서 제안하는 방법을 사용하여 정적인 배경

영역에서는 배경 차분 방법을 사용하여 전경 영역을 추출하고 동적인 배경 영역에서는 깊이 정보를 이용하여 전경을 효율적으로 추출한다. 깊이 정보와 배경 차분 방법을 이용하여 추출한 전경 영상을 후처리 과정을 통하여 잡음 또는 구멍(hole)을 제거한다. 마지막으로 얻어진 두 개의 전경 영상을 적절히 혼합하여 동적인 영상을 포함하는 배경에서도 효율적으로 전경 영상을 추출한다.

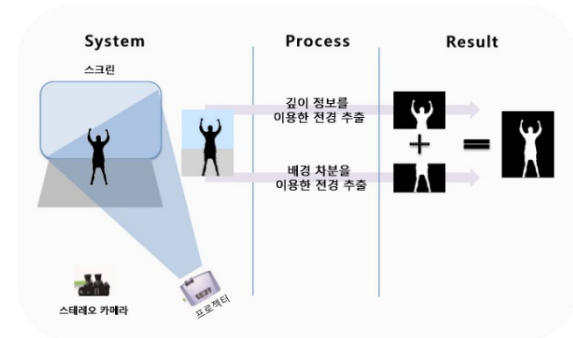


그림 1. 제안한 방법의 흐름도.

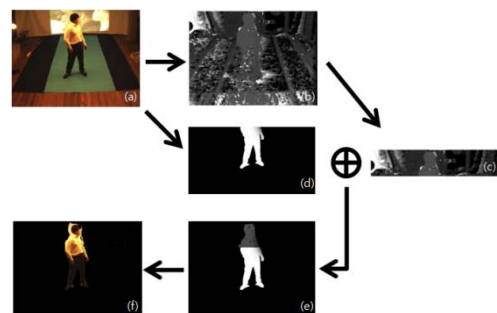


그림 2. (a) 원본 영상, (b) 깊이 정보 영상, (c) 깊이 정보를 이용한 추출 된 영상, (d) 배경 차분 방법을 사용하여 추출 된 영상, (e) (c),(d)영상을 조합한 영상, (f) 전경 영상 분리 결과.

그림 2 는 본 논문에서 제안한 방법을 사용하여 얻어진 전경 영역을 분리한 결과 영상이다. 원본 영상에 프로젝터에 의해 투영된 동적인 영상이 포함되어 있음에도 불구하고 효율적으로 전경 영역을 추출한 것을 그림 2-(f)를 통해 확인할 수 있다.

### 2.1 정적인 영상에서의 배경 영상 분리

정적인 영상에서의 배경 분리는 앞에서 설명한 것처럼 이미 가지고 있는 배경의 화소 정보를 알고 있을 때 가능한 방법이다. 일반적으로 배경은 영상에서 전경 영역을 제외한 나머지 영역을 배경으로 간주하며 초기 몇 초 동안의 수 프레임(frame)을 전경 영역이 없는 상태로 배경을 등록하여 이후에 들어 오는 영상이 배경과 얼마나 차이가 나는지를 분석하여 배경으로부터 전경 영역을 분리 추출할 수 있다.

### 2.1.1 배경 영역 분리

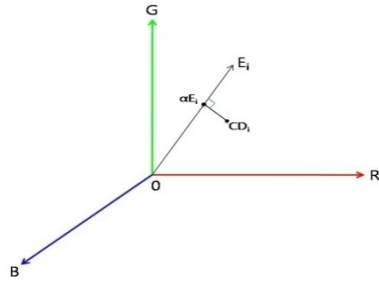


그림 3. 두 픽셀간의 밝기 왜곡과 색차 왜곡.

이 방법은 RGB 공간에서 두 화소간의 차이 밝기 왜곡(brightness distortion)과 색차 왜곡(chromatic distortion)을 계산하여 배경 영역과 전경 영역을 분리한다(그림 3). 그림 3의  $E_i$ 는 기대치 값(expected value) 즉 배경의 화소 값을 나타낸다. 본 논문에서 사용하는 배경의 화소 값의 등록은 초기 10 프레임 정도의 화면을 캡처한 후 각 화소 별로 평균값을 계산하여 기대치 값으로 취한다.  $I_i$ 는 현재 프레임에서 관찰된 화소 값을 의미하며  $\alpha_i$ 와  $CD_i$ 는 각각 밝기 왜곡과 색차 왜곡을 나타낸다. 두 값의 계산식은 다음과 같다.

$$E_i = [\mu R(i), \mu G(i), \mu B(i)] \quad (1)$$

$$\alpha_i = \min \left[ \left( \frac{I_R(i) - \alpha_i \mu R(i)}{\sigma R(i)} \right)^2 + \left( \frac{I_G(i) - \alpha_i \mu G(i)}{\sigma G(i)} \right)^2 + \left( \frac{I_B(i) - \alpha_i \mu B(i)}{\sigma B(i)} \right)^2 \right] \quad (2)$$

$$CD_i = \sqrt{\left( \frac{I_R(i) - \alpha_i \mu R(i)}{\sigma R(i)} \right)^2 + \left( \frac{I_G(i) - \alpha_i \mu G(i)}{\sigma G(i)} \right)^2 + \left( \frac{I_B(i) - \alpha_i \mu B(i)}{\sigma B(i)} \right)^2} \quad (3)$$

식(2), (3)의  $\sigma$ 는 등록된 배경 영상의 화소값들 간의 표준 편차를 의미하고 식(1),(2),(3)의  $\mu$ 는 배경 영상의 화소값들의 평균을 나타낸다. 이와 같이 계산된 밝기 왜곡과 색차 왜곡의 정보를 이용하여 표 1)의 조건에 의해 각 영역을 판별할 수 있다.

표 1: 밝기 왜곡 정보와 색차 왜곡 정보를 이용한 각 영역 판별 방법

영역	판별 방법
원 배경	대응되는 배경 이미지와 새로운 이미지의 밝기와 색차가 비슷할 때
그림자 및 어두운 배경	색차는 비슷 하나 대응되는 배경 이미지의 화소의 밝기가 낮을 때
전경	대응되는 새 이미지의 화소의 색차가 배경 이미지의 기대치와 다를 때

### 2.1.2 잡음(Noise) 제거

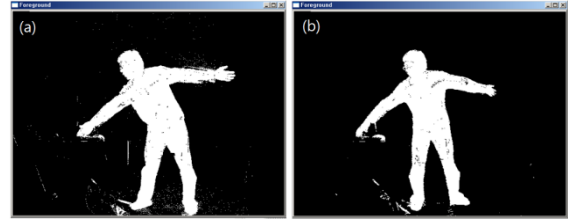


그림 4. (a) 잡음 제거 전의 전경 영상 (b) 잡음 제거 후의 전경 영상

2.1.1 절의 과정을 통해 얻어진 이진(binary)전경 영역 그림 4-(a)는 많은 잡음(salt and pepper noise)을 포함하고 있다. 이러한 문제를 해결하기 위해서 다음 세 단계의 후처리 과정을 수행한다. 첫 번째 후처리 과정은 salt and pepper 잡음을 제거 하기 위해 중앙 필터(median filter)를 사용한다.

본 논문에 사용된 중앙 필터는 필터와 필터 아래에 놓여진 원본 이미지 영역간의 대응되는 화소에 AND 연산을 취하여 나타나는 결과가 임계값(threshold) 이하이면 잡음으로 판단 하여 제거하는 방법을 사용한다 식(4).

$$result = f(x + i, y + j) \& filter(i, j)$$

$$if result = 255 \text{ Then increase counter}$$

$$if counter < T \text{ Then } f(x, y) = 0$$

where T is threshold value ,

$$f(x, y) \text{ is binary image and filter} \quad (4)$$

이러한 후처리 과정을 거치고 나면 그림 4-(b)와 같이 많은 잡음(salt and pepper noise)들이 제거 됨을 볼 수 있다. 그러나 그림 4-(b)에서처럼 객체가 아닌 부분이 객체로 추출되는 오류들이 여전히 남아있는 것을 확인 할 수 있다. 이러한 오류들은 두 번째 후처리 과정을 통해서 제거한다.

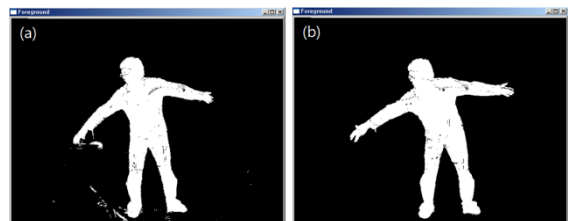


그림 5. (a) 라벨링을 수행하기 전의 전경 영역 (b) 라벨링을 수행한 후의 전경 영역

두 번째 후처리 과정은 그림 5와 같이 실루엣 영상에서 가장 큰 영역을 차지하는 부분만을 전경으로 결정하는 라벨링(labeling)을 수행하여 임계값(threshold)보다 작게 분류된 영역은 잡음으로 가정하여 제거하고 원하는 전경

영역만 추출한다. 라벨링은 표 2 와 같은 방법을 사용하여 수행한다.

표 2: 라벨링 수행 방법

Pass	Method
	영상의 모든 영역을 탐색하여 다음 조건을 검사 한다
First Pass	1) 모든 이웃이 배경 화소이면 새로운 레이블을 할당한다 2) 이웃 중에 다른 라벨이 하나라도 포함되어 있으면 해당되는 레이블을 할당한다
Second Pass	같은 영역이지만 First Pass 에서 다른 라벨로 판단 된 것을 새로운 하나의 라벨로 묶어 준다.

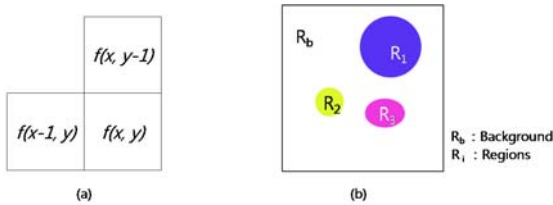


그림 6. (a) 이웃 화소 판별 방법. (b) 라벨링 후의 전경 영상들.

표 2)의 방법으로 라벨링을 수행 하고 나면 그림 6-(b)와 같이 배경 영상과 여러 전경 영상 영역들로 분리된다.

$$\begin{aligned} & \text{if } \max(R_i) \text{ then, assign to foreground} \\ & \text{else eliminate regions} \end{aligned} \quad (5)$$

그림 6 의 결과에서 식(7)의 조건을 만족하는 영역을 찾아내면 그림 5-(b)의 결과에서 살펴 볼 수 있듯이 가장 큰 물체만 추출되어 전경 영상에 나타난다. 하지만 추출된 전경 영상의 물체 내부에 작은 구멍(hole)이 아직 남아있는 것을 확인할 수 있다. 이러한 문제를 해결 하기 위해 마지막 후처리 과정을 거친다.

마지막 후처리 과정은 분류된 전경 영역의 내부에 생긴 구멍을 제거하기 위하여 팽창(dilation) 필터를 사용한다. 팽창 필터를 사용하고 나면 그림 7 과 같이 분류된 전경 영역의 내부에 있는 구멍이 거의 사라지는 것을 확인할 수 있다.

## 2.2 동적인 영상에서의 배경 영상 분리

본 논문에서는 스테레오 카메라를 사용하여 깊이 정보 영상을 획득한다. 스테레오 카메라를 사용하여 깊이 정보를 정교하게 도출하기 위해서는 촬영하는 영상의 스테레오 영상간 대응되는 특징점을 잘 추출할 수 있어야 한다. 예를 들어 단일색을 가지는 영상은 대응되는 점을 찾기 어렵기 때문에 보다 의미있는 색의 조합으로 이루어진 영상이 우수한 깊이 정보를 도출해 낼 수 있다

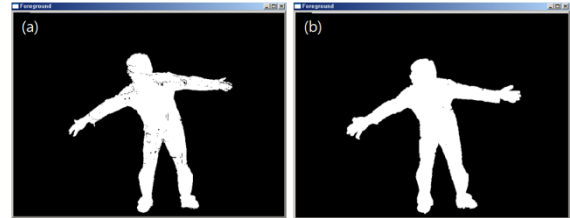


그림 7. (a) 팽창 필터를 적용하기 전의 전경 영역 (b) 팽창 필터를 적용하고 난 후의 전경 영역

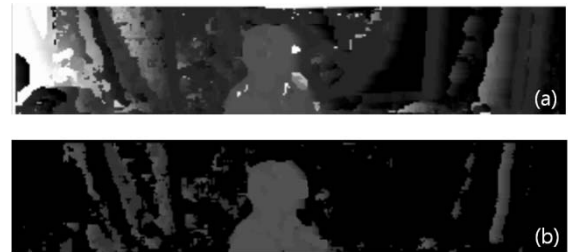


그림 8. (a)스테레오 카메라로부터 얻어지는 초기 상태의 깊이 정보 (b) 임계값을 기준으로 분류한 영상 정보.

### 2.2.1 배경 영상 분리

그림 8 은 스테레오 카메라로부터 얻어지는 초기 상태의 깊이 정보 영상을 임계값(threshold)을 기준으로 분류하여 객체와 배경을 분리해 낼수 있다. 그러나 임계값 만 가지고 분류했을 경우 그림 8-(b)와 같이 객체가 아닌 부분을 객체로 추출하는 오류가 발생하게 된다. 따라서 올바른 객체 추출 결과를 도출해 내기 위해서는 후처리 과정을 통해 잡음 제거를 해 주어야 한다.

### 2.2.2 잡음(Noise) 제거

2.2.1 절의 과정을 통해 얻어진 이진 전경 영역은 깊이 정보의 부정확성 문제때문에 원하지 않는 영역도 전경 영역으로 추출하게 된다. 이러한 문제점을 해결하기 위해서는 그림 9 와 같이 우리가 원하는 객체가 가장 크다고 가정을 하고, 2.1.2 절의 라벨링(labeling)을 방법을 사용하여 가장 많은 영역을 차지하는 객체를 추출하면 된다.

그러나 그림 9-(b)에서 보듯이 분류된 전경 영역은 원본 컬러(color) 영상의 실제 물체보다 비교적 크게(fatting)

추출 되는 경향이 있다. 이러한 문제를 해결하기 위해 마지막으로 침식(erosion) 필터를 사용하여 추출된 전경 영역의 물체를 원본 컬러 영상의 물체와 비슷하게 추출할 수 있다. 침식 필터는 원본 컬러 영상에서 전경 성분의 얼굴의 살색과 머리카락 색상은 변화하는 성분이 아니라는 가정을 하고, 침식의 범위를 살색과 머리카락 색상이 검출되는 영역까지로 지정하면 비교적 실제 전경 영역에 가깝게 추출이 가능하다.



그림 9. (a) 임계값을 기준으로 분류한 영상 정보. (b) 라벨링 방법을 사용하여 추출해낸 영상

### 3. 실험 결과 및 분석

그림 10 은 본 논문에서 사용된 실험 환경을 보여준다. 카펫이 깔린 부분(가로 3m, 세로 3m)을 활동 영역으로 가정하면 움직이는 객체의 윗 부분은 프로젝터가 투영되는 스크린과 겹치게 된다. 따라서 스크린이 객체와 겹치지 않는 부분은 배경 차분 기술을 사용 하고 스크린이 겹치는 부분은 깊이 정보를 이용한 전경 영역을 추출 방법을 사용하여 실험을 진행하였다. 실험을 진행하기 위하여 깊이 정보 추출이 가능한 Point Grey 사의 스테레오 카메라 Bumblebee2 를 사용하고 640X480 크기의 영상을 이용하였다. 또한 본 실험에 사용된 PC 사양은 표 3 과 같다.

표 3: 실험에 사용된 PC 사양

장치	사양
CPU	Intel® Core™2 Quad CPU Q9400 @ 2.66GHz
RAM	2GB RAM
VGA	Nvidia Geforce 9600 GT
CAMERA	PointGrey Bumblebee 2

그림 11 는 실험환경 내에서 제안한 방법을 적용하여 얻어낸 결과를 나타낸다. 본 논문에서 제안하는 하이브리드 배경 차분 기술을 사용한 결과 비교적 정확하게 전경 영역이 추출되는 것을 확인할 수 있었다. 또한, 초당 10fps~15fps 정도의 속도로 객체를 추출해 낼 수 있었다. 배경 차분 방법만을 사용 하면 좀더 빠른

속도로 객체를 추출해 낼 수 있지만 깊이 정보를 획득하는데 많은 시간이 소요되기 때문에 속도 저하가 발생하였다.



그림 10. 실험 환경

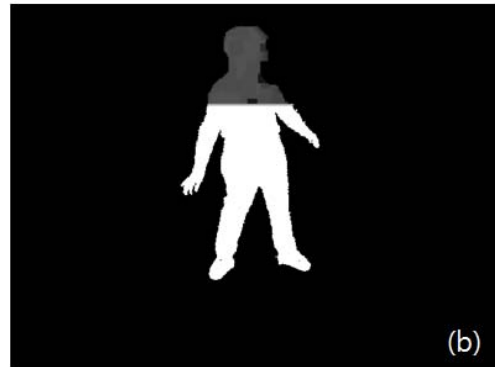


그림 11. (a) 원본 영상 (b) 배경 차분 기술과 깊이 정보를 이용하여 추출 한 전경 영역 (c) 전경 영역을 이용하여 추출 한 객체 이미지

#### 4. 결론

본 논문에서는 동적인 배경을 포함하는 영상에서 효율적으로 전경 영역을 추출 하는 방법을 제안하고 실험을 통해 기존의 배경 차분 방법에 비해 동적인 영상에서 추출된 전경 영역의 품질의 우수성을 확인 하였다. 또한 실험을 통해 실험 환경이 깊이 정보를 추출 하는데 많은 영향을 끼치는 것을 확인할 수 있었다.

기존의 배경 차분 기술은 동적인 환경에서 추출이 이론적으로 불가능하였다. 그러나 본 논문에서 제안하는 방법을 통해서 배경영상이 동적인 다양한 환경에서도 비교적 정확한 전경 영역을 추출할 수 있었다. 향후 본 논문에서 제안한 방법을 고해상도 영상에 적용하기 위해서 GPU(Graphics Processing Unit)를 이용한 병렬 처리 과정을 추가하는 연구를 수행 중에 있다.

#### 참고문헌

[1] Thanart Horprasert, David Harwood, Larrs S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," Proc. Frame Rate Workshop at ICCV' 99, pp. 1-19, 1999.

[2] Chris Stauffer, W.E.L Grimson, "Adaptive background mixture models for real-time tracking," IEEE International Conf. Computer Vision and pattern Recognition, pp 246- 252, 1999.

[3] P. KaewTraKulPong, R. Bowden, "An Improved adaptive background mixture model for real-time tracking with shadow detection," Proc. the 2<sup>nd</sup> European Workshop on Advanced Video-Based Surveillance Systems, 2001.

[4] K. Toyama, J. Krumm, B.Brumitt, B Meyers, "Wallflower Principles and practice of background

maintenance," Proc. the 7<sup>th</sup> IEEE international conf, on Coumputer Vision, pp 252-261, 1999.

[5] P. Kumar, S. Ranganath, W. Huang, "Queue based fase background modeling and fast hysteresis thresholding for better foreground segmentation," Proc. The 2003 joint Conf of the Fourth ICICS and PCM, pp 2A 2.5, 2003.

[6] Bo Li, Zhen Tang, Baozong Yuan, Zhenjiang Miao, "Segmentation of Moving Foreground Objects using Codebook and Local binary Patterns," Proc. Image and Signal Processing Vol.4, pp 239- 243, 2008.

[7] O. Javed, K. Shafique, and M. Shah. "A hierarchical approach to robust background subtraction using color and gradient information," Proc. Motion and Video Computing, pp 22-27, 2002.

[8] E. Salvador, P. Green, and T. Ebrahimi, "Shadow identification and classification using invariant color models," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pp 1545-1548, 2001.

[9] Sen and Chandrika Kamath. "Robust techniques for background subtraction in urban traffic video," Proc. SPIE, Visual Communications and Image Processing 2004, volume 5308, pp 881-892, 2004.

[10] M. M. Trivedi A. Prati, I. Mikic and R. Cucchiara. "Detecting moving shadows: Algorithms and evaluation," IEEE Trans. Pattern Analysis and Machine Intelligence, 25(7), pp918-923, 2003.

[11] C. S. Cheung and C. Kamath. "Robust background subtraction with foreground validation for urban traffic video," Proc EURASIP Journal on Applied Signal Processing, pp 1-11, 2005.