

## 연속적인 배경 모델 학습을 이용한 코드북 기반의 전경 추출 알고리즘

정재영\*

### 요약

이동 물체의 검출은 비디오 감시, 보행자의 행동 분석과 같은 컴퓨터 시각 분야에서 매우 중요한 전처리 작업이다. 이는 실제 외부 환경을 대상으로 할 때, 영상 시퀀스에 존재하는 배경의 불규칙한 움직임, 조명 변화, 그림자, 배경 물체의 위상 변화 및 잡음 등으로 인하여 매우 어려운 작업이다. 본 논문에서는 코드북 기반의 전경 검출 알고리즘을 제안한다. 코드북은 입력 영상으로부터 얻어지는 배경 화소에 대한 정보 데이터베이스이다. 먼저, 첫 번째 프레임을 배경 영상으로 가정하고 이를 입력 영상과 비교하여 차 영상을 구한다. 구해진 차 영상에는 순수한 이동 물체뿐만 아니라, 잡음까지 포함된다. 둘째로, 전경으로 검출된 화소의 색상과 밝기 값을 가지고 코드북을 조사하여 존재하는 경우 잘못 추출된 전경 화소로 판단하고 전경에서 제거한다. 마지막으로, 다음번 입력되는 프레임을 반복 처리하기 위하여 배경 영상을 새롭게 갱신하는데, 배경 화소로 검출된 화소의 경우에는 현재의 입력 영상으로부터 추정되며, 전경 화소로 검출된 경우에는 이전 배경 영상의 화소 값을 복사하여 사용한다. 제안한 알고리즘을 PETS2009 데이터에 적용한 결과를 GMM 알고리즘과 표준 코드북 알고리즘의 결과와 비교하여 보인다.

키워드 : 배경 모델 학습, 배경 제거, 전경 추출, 코드북

## Codebook-Based Foreground Extraction Algorithm with Continuous Learning of Background

Jae-young Jung\*

### Abstract

Detection of moving objects is a fundamental task in most of the computer vision applications, such as video surveillance, activity recognition and human motion analysis. This is a difficult task due to many challenges in realistic scenarios which include irregular motion in background, illumination changes, objects cast shadows, changes in scene geometry and noise, etc. In this paper, we propose an foreground extraction algorithm based on codebook, a database of information about background pixel obtained from input image sequence. Initially, we suppose a first frame as a background image and calculate difference between next input image and it to detect moving objects. The resulting difference image may contain noises as well as pure moving objects. Second, we investigate a codebook with color and brightness of a foreground pixel in the difference image. If it is matched, it is decided as a fault detected pixel and deleted from foreground. Finally, a background image is updated to process next input frame iteratively. Some pixels are estimated by input image if they are detected as background pixels. The others are duplicated from the previous background image. We apply our algorithm to PETS2009 data and compare the results with those of GMM and standard codebook algorithms.

Keywords : Background model Learning, Background elimination, Foreground extraction, Codebook

※ 교신저자(Corresponding Author): Jae. Y. Jung  
접수일: 2014년 06월 02일, 수정일: 2014년 07월 23일  
완료일: 2014년 07월 27일

\* 동양대학교 컴퓨터정보전학과

Tel: +82-54-630-1055, Fax: +82-54-630-1141

email: [jjung@dyu.ac.kr](mailto:jjung@dyu.ac.kr)

■ 본 연구는 동양대학교의 2013학년도 교내연구비

### 1. 서론

최근 CCTV의 보급과 더불어 이동 객체의 추적을 통한 출입 감시, 인간의 행동 분석, 위험 상황 분석을 위한 연구가 활발하다. 객체 검출은 움직이는 전경 객체와 배경을 분리하는 기술로서 객체 추적의 성능을 크게 좌우한다. 실외 환경에서는 전-배경 분리를 어렵게 하는 많은 요인들이 존재하는데, 조명 변화, 그림자, 잡음, 바람 등의 영향으로 배경의 흔들림, 주차장 등에서 나타날 수 있는 배경의 위상 변화 등이 있다.

전-배경 분리를 위한 연구는 크게 픽셀 기반의 방법과 영역 기반의 방법으로 분류할 수 있는데, 픽셀 기반의 배경 모델링 기법에는 커널(kernel)에 기초한 방법[1-2], GMM(Gaussian Mixture Model)에 기초한 방법[3-6], 코드북에 기초한 방법[7-11] 등이 있다. 차 영상을 이용한 방법은 계산이 간단하다는 장점은 있으나 다양한 환경에서 정확한 배경 모델링이 어려우며, GMM 방법은 매개변수로 설정하는 배경 학습율에 따라 점진적인 조명 변화에는 잘 적응하지만 배경이 빠르게 변화하거나 불규칙하게 배경이 흔들리는 환경에는 만족스러운 결과를 보이지 못하고 있다. 코드북 방법은 배경 모델링의 정확도와 처리 시간 측면에서 강점을 보이는 최신의 기술로서 최근 10여 년간 활발하게 연구되고 있다. Kim[7]은 훈련 영상 시퀀스로부터 동일 위치의 화소 값들을 양자화하여 배경을 모델링하는 코드북 기반의 방법을 처음으로 제안하였다. 그러나 그의 방식에서는 급격한 조명의 변화, 매개변수 설정의 어려움 등의 문제를 보였으며, 이를 개선하기 위한 많은 방법[8-9]들이 제시되었다. Guo[10], Quan[11] 등은 블록기반 방식과 픽셀기반 방식을 혼합한 계층적인 코드북 방식을 제안하였다.

2장에서는 표준적인 코드북 기반의 전경 추출 알고리즘을 설명하고, 3장에서는 제안한 알고리즘을 설명한다. 4장에서는 PETS2009에 공개된 몇 개의 벤치마크 데이터에 대하여 적용한 결과를 기존의 GMM, 표준 코드북 알고리즘과 비교하여 보이고, 5장에서는 결론을 맺는다.

지원에 의해 수행되었음

### 2. 표준 코드북 모델[7]

코드북 알고리즘은 양자화 기법을 적용하여 훈련 영상 시퀀스로부터 배경을 모델링하여 화소 단위로 코드북  $C = \{c_1, c_2, \dots, c_L\}$ 을 생성한다. 이 때,  $L$ 은 코드북을 구성하는 코드워드의 개수로서 화소 값의 변화 특성에 따라 화소마다 그 크기가 다를 수 있다. 코드워드  $c_i$ 는 식(1)과 같이 RGB 벡터  $v_i$ 와 6차원의 속성벡터  $aux_i$ 로 구성되는데, 영상 시퀀스로부터 해당 화소 위치에서의 평균 색상, 그 코드워드가 포용하는 밝기 범위, 출현 빈도, 출현하지 못한 최장 시간, 최초 출현 시점, 최근 출현 시점을 의미한다.

$$\begin{cases} v_i = (\overline{R}_i, \overline{G}_i, \overline{B}_i) \\ aux_i = \langle \tilde{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i \rangle \end{cases} \quad (1)$$

생성된 초기 코드북은 전경 및 배경을 모델링한 결과가 함께 나타나는데,  $\lambda$  값을 기준으로 오랫동안 사용되지 않은 코드워드는 전경의 코드워드로 판단하고 코드북에서 삭제한다. 코드북 생성의 구체적인 알고리즘은 <표 1>과 같다.

<표 1> 코드북 생성 알고리즘

```

//N: the number of training images
I. L=0, C=∅ (empty set)
II. for t=1 to N do
(i)  $x_t = (R, G, B)$ ,  $I = \sqrt{R^2 + G^2 + B^2}$ 
(ii) Find the  $c_i$  in C matching to  $x_t$  based on the following two conditions.
(a)  $Color\_Distortion(x_t, v_i \text{ of } c_i)$ ,  $d_{x_t, v_i} \leq \epsilon$ 
(b)  $Brightness(I, \langle \tilde{I}_i, \hat{I}_i \rangle \text{ of } c_i)$ ,  $b_{x_t, c_i} = \mathbf{true}$ 
(iii) If there is no match, then  $L=L+1$ ,
Create a new codeword  $c_L$  by setting
 $v_L = (R, G, B)$ ,  $aux_L = \langle I, I, t-1, t, t \rangle$ 
(iv) Otherwise, update the matched codeword  $c_i$ .
 $v_i = \left( \frac{f_i \overline{R}_i + R}{f_i + 1}, \frac{f_i \overline{G}_i + G}{f_i + 1}, \frac{f_i \overline{B}_i + B}{f_i + 1} \right)$ ,
 $aux_i = \langle \min\{I, \tilde{I}_i\}, \max\{I, \hat{I}_i\}, f_m + 1, \max\{\lambda_i, t - q_i\}, p_i, t \rangle$ 
end for
III. Refine C based on maximum negative run-length,  $\lambda$ 
 $C = \{c_i | c_i \in C \wedge \lambda_i \leq Thr\}$ 
    
```

<Table 1> Codebook construction algorithm

전경 검출 과정에서는 입력 영상의 각 화소 단위로 코드북을 조사하여 매칭이 이루어지면 배경 화소로 판단하고 해당 코드워드의 속성 값

을 수정하며, 매칭이 이루어지지 않으면 전경 화소로 판단한다. 매칭 여부는 식 (2), (3)과 같이 색상 및 밝기 차이를 기준으로 한다.

$$d_{x,v_i} = \sqrt{(R^2 + G^2 + B^2) - \frac{(RR_i + GG_i + BB_i)^2}{R_i^2 + G_i^2 + B_i^2}} \quad (2)$$

$$b_{x,c_i} = \begin{cases} true & \text{if } \alpha \tilde{I}_i \leq \|x\| \leq \min\{\beta \tilde{I}_i, \tilde{I}_i/\alpha\} \\ false & \text{otherwise} \end{cases} \quad (3)$$

### 3. 제안 알고리즘

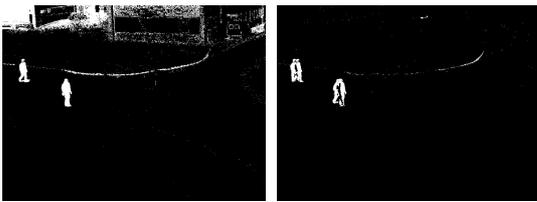
#### 3.1 배경 모델링

차 영상을 이용한 배경 모델링에서 두 영상의 시간 간격이 클 경우에는 조명의 변화에 많은 영향을 받으며, 시간 간격이 작을 경우에는 전경 물체의 영향을 받아 정확한 배경을 모델링하기 어렵다. (그림 1d)는 lc-al를 계산한 결과로 많은 시간 차이로 인하여 건물 벽면에 많은 변화가 있는 것을 알 수 있으며, (그림 1e)는 lc-bl를 계산한 결과로 두 프레임에 존재하는 각각의 물체가 결과로 나타나게 되어 현재 영상에서의 전경 물체만을 추출하는 것이 어렵다.

(그림 1) 차 영상



(a) 1번째 (b) t-1 번째 (c) t번째



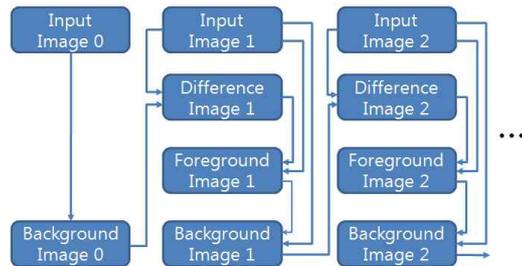
(d) 차영상(lc-al) (e) 차영상(lc-bl)

(Figure 1) Difference images

본 논문에서는 차 영상에 의한 배경 모델을 1 단계 배경 모델로, 코드북에 의한 배경 모델을 2 단계 배경 모델로 계층화하여 1단계 모델에 의해서는 개략적인 전경 물체를 추출하고, 2단계에서는 순수한 전경 물체만을 검출하도록 하였다.

1단계의 인접 영상 간 차 영상에 의한 배경 모델링 과정에서 (그림 1e)의 문제를 해결하기 위하여 (그림 2)와 같이 매 프레임마다 자동적으로 배경 영상을 습득하도록 하였다.

(그림 2) 연속적인 배경 모델 갱신



(Figure 2) Continuous updating of background model

우선, 입력 시퀀스의 첫 번째 영상을 이동 물체가 없는 순수한 배경 영상으로 가정하고, 이후 입력되는 영상에 대하여 다음의 과정을 반복한다. 첫째, 입력 영상  $I_t$ 와 이전의 배경 영상  $B_{t-1}$ 과의 차 영상을 구하고, 식 (4)와 같이 초기에 설정한 임계값을 기준으로 이를 이진화한다.

$$D_t = \begin{cases} f, & \text{if } |I_t - B_{t-1}| \geq \epsilon \\ b, & \text{otherwise} \end{cases} \quad (4)$$

둘째, 이진화된 차 영상  $D_t$ 에서  $f$ 로 표현되는 전경 화소는 이동 객체에 의한 실제의 전경뿐만 아니라, 조명 변화, 배경 변화, 잡음 등의 영향으로 잘못 검출된 전경이 함께 나타나게 되는데,  $D_t$ 로부터 배경 및 잡음을 제거(3.2절에서 설명)하고 전경 영역을 구성하는 화소들만 추출한다.

셋째, 식 (5)와 같이 현재의 입력 영상  $I_t$ 와 추출된 전경 영상  $F_t$ 을 이용하여 배경 영상을 자동적으로 갱신한다. 현재의 입력 영상에서 전경 화소로 검출된 경우 이전의 배경 화소를 그대로 사용하고, 배경 화소로 검출된 경우에는 입력 영상과 이전 배경 영상을 기반으로 현 시점의 배경 영상을 추정한다.

$$B_t(x,y) = \begin{cases} B_{t-1}(x,y), & \text{if } F_t(x,y) = f \\ \alpha I_t(x,y) + (1-\alpha)B_{t-1}(x,y), & \text{otherwise} \end{cases} \quad (5)$$

### 3.2 전경 추출

일반적으로 차 연산 과정에서 참조 영상으로 사용되는 이전 배경 영상  $B_{t-1}$ 과 현재 입력 영상  $I_t$ 를 비교할 때,  $I_t$  영상 내의 전체 화소들은 다음과 같이 분류된다.

$$I_t = \{B'_{t-1}, B^N_{t-1}, N_t, O_t\} \quad (6)$$

- $B'_{t-1}$ : 이전 배경  $B_{t-1}$ 의 값이 그대로 나타나거나, 조명 변화에 의하여 그 값이 미세하게 변하는 경우의 화소 집합
- $B^N_{t-1}$ : 바람 등의 영향으로 배경이 흔들려서 주변 화소의 값이 나타나는 경우의 화소 집합
- $N_t$ : 백색 잡음 등의 영향으로 배경 영상과 전혀 다른 값이 부분적으로 나타나는 경우의 화소 집합
- $O_t$ : 이동 물체의 영향으로 이전 배경 영상과 전혀 다른 값이 나타나는 경우의 화소 집합

이 중에서  $O_t$ 만이 실제적인 전경 화소 그룹이므로, 최종적인 객체 검출 결과에서 추출되어야 한다. 본 논문에서는  $B'_{t-1}$ 의 영향을 배제하기 위하여 식 (5)와 같이 계속적으로 배경을 학습해 나가도록 하였고,  $B^N_{t-1}$ 의 영향을 배제하기 위하여 코드북 검색 시에 현재 화소 위치에서의 코드북 내에 일치하는 코드워드가 없을 경우 주변 화소의 코드북을 함께 조사하도록 하였다.  $N_t$ 를 제거하기 위하여 차 영상에서 고립점의 형태로 존재하는 점들은 삭제하도록 하였다.

구체적인 전경 추출 알고리즘은 <표 2>와 같다. 처음 입력되는 영상의 화소 값들을 그대로 사용하여 초기 코드북을 생성하고, 첫 번째 배경 영상으로 사용한다. 이를 이후에 입력되는 영상과 비교하여 차 영상을 구한다. 차 영상에서 인접 화소와 연결되어 있지 않은 고립점은 제거되며, 배경으로 검출된 검은 화소와 함께 전경으로 검출된 흰 화소 중에서 주변 코드북에서도 일치하는 코드워드가 없는 경우 최종적인 배경으로 판단한다. 마지막으로 이러한 과정을 매 프레임마다 반복 적용하기 위하여 식 (5)에 의해 배경 영상을 갱신한다.

<표 2> 전경 추출 알고리즘

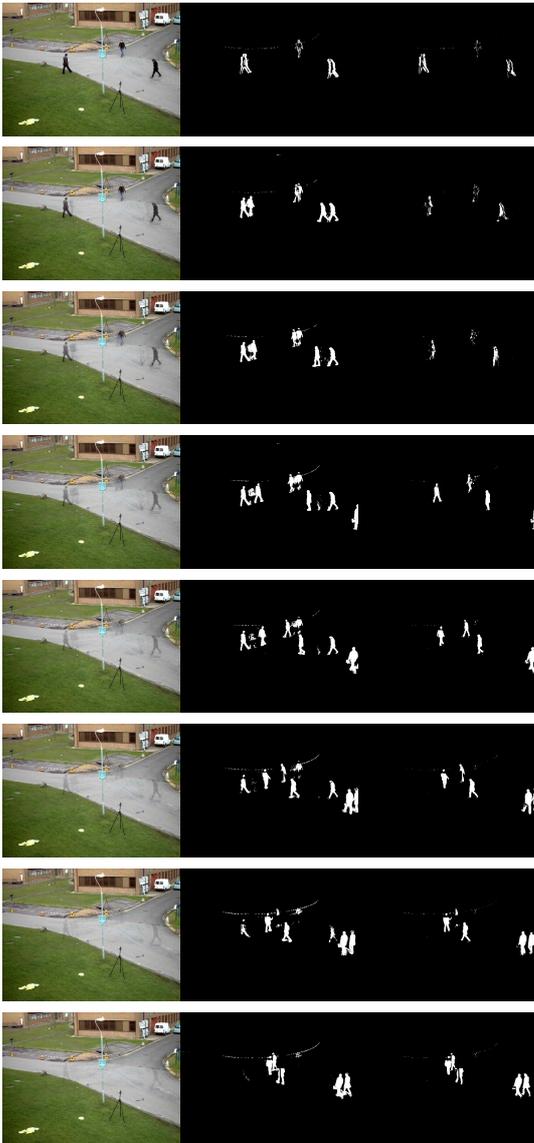
<p>Notation <math>C^N</math>: an union of several neighboring codebooks</p> <p>I. Generate an initial codebook consisting of one codeword for each pixel using the first input image, <math>I_0</math>.</p> <p>II. <math>I_0</math> is used as <math>D_0</math>, background image.</p> <p>III. for each input image, <math>I_t</math></p> <p>(i) Calculate <math>D_t</math> by formular (4)</p> <p>(ii) Eliminate isolated points in the <math>D_t</math></p> <p>(iii) If <math>D_t(x, y) = b</math>,</p> <p>1) determine as a background pixel,</p> <p>2) if matched = <b>true</b> in <math>C^N</math>, update the matched codeword. else add a new codeword in C</p> <p>(iv) If <math>D_t(x, y) = f</math> and matched = <b>true</b> in <math>C^N</math>,</p> <p>1) determine as a background pixel, 2) update the matched codeword.</p> <p>(v) If <math>D_t(x, y) = f</math> and matched = <b>false</b> in <math>C^N</math>,</p> <p>1) determine as a foreground pixel, 2) add a new codeword in C</p> <p>(vi) Update background image using formular (5)</p> <p>end for</p>
---

<Table 2> Proposed foreground extraction algorithm

### 4. 실험결과

제안한 알고리즘을 Pets2009[12]에 공개된 두 가지 Benchmark Dataset에 적용 실험하였다. (그림 3)은 초반부에 3명의 보행자가 중앙을 향하여 이동하는 시나리오에서 시간의 흐름에 따라 자동적으로 배경 영상이 추정되고 있는 과정을 보인 것이다. 1열을 살펴보면, 시작 단계에서는 3명의 보행자를 포함하고 있는 입력 영상을 배경 영상으로 사용하지만, 점차적으로 순수 배경 영상만을 포함하도록 변화하고 있는 것을 볼 수가 있다. 2열은 1열에서 보인 배경 영상과 현재 입력 영상과의 차 영상을 보인 것으로, 23번째 프레임부터 순수한 배경 영상이 나타나고 있다. 그림에서 2열의 마지막 행에 있는 그림을 확대하여 살펴보면 미세하게 보행자의 흔적이 남아있긴 하지만, 입력 영상과의 밝기 차가 임계값 이하여서 차 영상을 구하는 과정에서는 무시된다. 3열은 2열의 차 영상을 바탕으로 전경 물체만을 검출한 결과로, 3행(입력 시퀀스의 8번째 프레임)까지는 최초의 사람 위치와 현재 프레임에서의 사람의 위치가 겹치는 부분이 존재하여 전경 물체가 뚜렷하게 검출되지는 않지만, 이후 프레임부터는 정확하게 전경을 검출하고 있다.

(그림 3) 자동적인 배경 추정 과정



(Figure 3) Background estimation process  
 (Row: 2, 5, 8, 11, 14, 17, 20, 23th frame,  
 1st column: Automatically estimated  
 background, 2nd column: Difference image  
 3rd column: Foreground detection results)

(그림4~5)는 제안한 방법의 객체 검출 결과를 GMM 방법과 표준 코드북 방법의 결과와 비교하여 보인 것이다. Dataset 1은 다중의 보행자가 캠퍼스를 활보하는 영상으로 첫 번째 영상부터 보행자가 존재하는 영상이고, Dataset 2는 소수

의 보행자뿐만 아니라 자동차가 빠르게 이동하는 영상으로 시간의 흐름에 따라 조명의 변화가 존재하는 경우이다. 또한, 두 Dataset 모두에 잔디 보호를 위한 진입 차단용 바줄이 바람에 심하게 흔들리고 있다.

표준 코드북 방법의 실험에서 사용한 매개변수는  $\epsilon, \alpha, \beta$ 의 값은 각각 10, 0.5, 1.5를 사용하였고, 훈련 프레임 수는 두 가지 Dataset에 대하여 각각 100, 50 프레임을 사용하였다. 제안한 알고리즘의 경우에도 동일한  $\epsilon, \alpha, \beta$  값을 사용하였으며, 차영상의 이진화를 위한 임계값으로는 밝기 값 25를 사용하였다. GMM 방식의 경우, 바람에 흔들리는 배경의 위치 변화에 의한 영향이나 조명의 변화에 의한 영향을 많이 받고 있으며, 표준 코드북 방식의 경우 배경 화소가 다양한 변화 패턴을 보일 경우 적절하게 처리하지 못하고 있음을 볼 수 있다. 제안한 방식의 결과에서는 조명에 의한 배경의 변화나 다양한 변화 패턴을 보이는 화소에 대해서도 우수한 결과를 보이고 있다. 진입차단용 바줄의 경우 세 가지 방식 모두에서 충분하게 제거되지 못하고 있다. 본 알고리즘의 경우 중심 화소를 기준으로 5X5 블록 안에서 이웃하는 코드북을 함께 조사하여 배경 여부를 판단하고 제거하였으나, 바줄이 임의의 방향으로 크게 흔들림으로써 충분한 제거 효과를 얻지 못한 것으로 판단된다.

### 5. 결론

본 논문에서는 전경 객체를 검출하기 위하여 차영상에 의한 배경 모델과 코드북에 의한 배경 모델을 혼합한 배경 모델링 방법을 제안하였다. 효과적인 차영상 모델링을 위하여 매 프레임마다 배경 영상을 자동적으로 추정하였고, 차 연산을 통하여 개략적인 전경 물체를 빠르게 검출한 후, 코드북 검색을 통하여 정확한 전경 물체만을 검출하도록 하였다. PETS2009에 공개된 몇 가지 영상에 적용하여 배경 변화에 민감한 GMM 기반 방법과 잡음에 민감한 코드북 기반 방법에 비하여 보다 우수한 결과를 보이고 있음을 확인하였다. 차 연산의 결과에서 배경과 유사한 색상의 전경 영역에 대해서는 배경으로 잘못 판단되는 문제를 여전히 안고 있으며, 향후에 지속적인

로 보완할 필요가 있다.

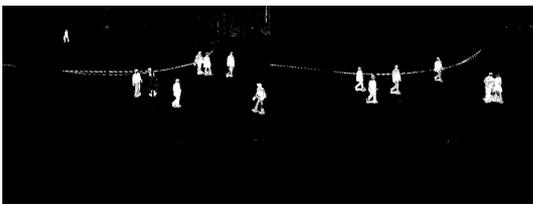
(그림 4) Dataset 1: 다중 보행자 영상



(a) frame #1



(b) frames #117, #328



(c) Results of GMM



(d) Kim's results



(e) Our results

(Figure 4) Dataset 1: multiple pedestrians

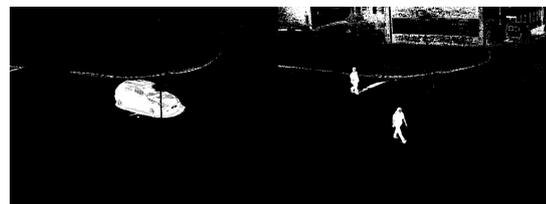
(그림 5) Dataset 2: 조명 변화 영상



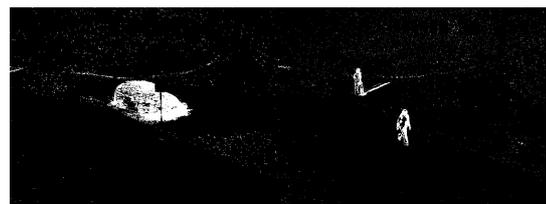
(a) frame #1



(b) frames #70, #159



(c) Results of GMM



(d) Kim's results



(e) Our results

(Figure 5) Dataset 2: illumination change

References

[1] Mittal A, Paragios N. "Motion-based background subtraction using adaptive kernel density estimation." Proceedings of IEEE Conference on CVPR, Vol.2, pp.302-309, 2004.

[2] Sheikh Y, Shah M. "Bayesian modeling of dynamic scenes for objects detection." IEEE Trans. on PAMI. Vol.27, pp.1778-1792, 2005.

[3] Stauffer C, Grimson WEL. "Adaptive background mixture models for real-time tracking." Proceedings of IEEE Conference on CVPR, pp. 246-252. 1999.

[4] Dalley G, Migdal J, Grimson WEL. "Background subtraction for temporally irregular dynamic textures." IEEE Workshop on Applications of Computer Vision, pp.1-7. 2008.

[5] X. Zhang, F. Liu, Z. Li. "An improved foreground object detection method based on Gaussian mixture models," IEEE International Conference on Multimedia Communications, pp. 90-93, 2010.

[6] J. Kim, K. Lee, T. Hwang, "Illumination influence minimization method for efficient object extraction," Journal of Digital Contents Society, Vol.14, No.1, pp. 117-124, 2013.

[7] K. Kim, T.H.Chalidabhongse, D. Harwood, L. Davis, "Real-time foreground - background segmentation using codebook model." Real-Time Imaging Vol.11, pp.172-185, 2005.

[8] M. Shah, J. Deng, B. Woodford, "Enhanced Codebook Model for Real-Time Background Subtraction, Neural Information Processing," International Conference on Neural Information Processing, pp. 449 - 458, 2011.

[9] M. Wu, X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction." International Journal of Electronics and Communications, Vol.64, No.8, pp.739-747, 2010.

[10] J.M. Guo, Y.F. Liu, C.H. Hsia, "Hierarchical method for foreground detection using codebook model," IEEE Transactions on Circuits and Systems for Video Technology, Vol.21, No.6, pp.804-815, 2011.

[11] S. Quan, T. Zhixing, H. Songchen, "Hierarchical codebook for background subtraction in MRF," Infrared Physics & Technology Vol.61, pp.259-264, 2013.

[12] <http://pets2009.net>



정재영

1989년 : 성균관대학교 (공학사)  
 1993년 : 성균관대학교 (공학석사)  
 1997년 : 성균관대학교 (공학박사)

1997년~현재 : 동양대학교 컴퓨터정보전학과 교수  
 관심분야 : 인공지능(Artificial Intelligence), 패턴인식 (Pattern Recognition), 움직임 추적(Motion Tracking) 등