

## SuBSENSE 알고리즘 구조 분석

이상하, \*유지상  
 광운대학교, \*광운대학교  
 kcv456@kw.ac.kr, \*jsyoo@kw.ac.kr

### SuBSENSE algorithm structure analysis

SangHa Lee \*JiSang Yoo  
 Kwangwoon University \*Kwangwoon University

#### 요 약

최근 카메라로부터 획득한 영상을 이용하여 지능적으로 상황을 인지하고 판단하고 결정하는 알고리즘의 연구개발이 활발하게 이루어지고 있다. 예를 들어 영상 내의 객체를 검출하는 알고리즘, 영상 내 화재와 같은 위험 상황을 알려주는 알고리즘 등이 있다. 본 논문에서는 SuBSENSE 라고 하는 영상 내 객체를 검출하는 알고리즘에 대해서 분석하고자 한다. SuBSENSE 는 background subtraction 기반으로 동작하는 객체 알고리즘으로서 다양한 상황에도 강건하게 객체를 추출하기위한 몇 가지 과정들이 존재한다. 본 논문에서는 SuBSENSE 알고리즘 구조 분석 및 해당 구조에서 동작하는 파라미터들의 역할에 대해 살펴보고자 한다.

#### 1. 서론

최근 카메라로부터 획득한 영상을 이용하여 지능적으로 상황을 인지하고 판단하고 결정하는 알고리즘의 연구 개발이 활발하게 이루어지고 있다. 예를 들면 영상 내의 객체를 검출하는 object detection, 출입구를 기준으로 드나드는 사람의 수를 세는 people counting, 영상 내에 화재와 같은 위험한 상황을 알려주는 알고리즘 등이 있다. 본 논문은 영상 내의 객체를 검출하는 object detection 알고리즘 SuBSENSE: A Universal Change Detection Method With Local Adaptive Sensitivity [1]에 대해서 분석하고자 한다. SuBSENSE 는 background subtraction 기반으로 하여 객체를 검출하는 알고리즘들 [4], [5], [6] 중 하나이다. SuBSENSE 는 다양한 상황에 강건하게 객체를 검출하기 위한 파라미터들이 존재하고 이를 계산하기 위한 많은 과정들이 요구된다. 본 논문에서는 SuBSENSE 알고리즘 분석 및 파라미터들의 역할에 대해 살펴보고자 한다.

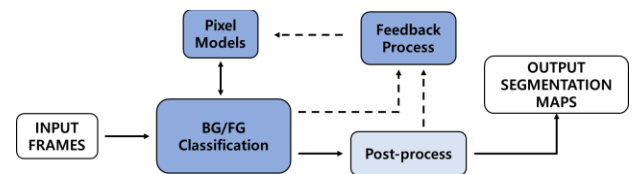


그림 1. SuBSENSE 알고리즘 흐름도

SuBSENSE 는 크게 Pixel Models, BG/FG Classification, Post-process, Feedback Process 으로 총 4 가지 과정으로 이루어져 있다. Pixel models 이란 background subtraction 을 할 때 필요한 background sample 50 개를 보관하고 업데이트 하는 모듈이다. Background sample 은 입력 영상과 해상도가 같은 이미지의 형태로 이루어져 있으며 각 픽셀은 RGB color 성분과 LBSP(Local Binary Similarity Pattern) [2]이라는 특징 성분을 포함하고 있다. LBSP 란 LBP(Local Binary Pattern) [3]와 유사한 특징으로 background subtraction 을 진행할 때 LBP 보다 더 좋은 성능을 보여 많이 사용되는 특징이다. 그림 2 는 LBP 와 LBSP 의 마스크를 나타낸다.

#### 2. SuBSENSE 분석

SuBSENSE 알고리즘은 비디오 내에서 background subtraction 을 기반으로 움직이는 객체를 추출하는 알고리즘이다. SuBSENSE 는 다양한 상황에서도 강건하게 객체를 추출하기 위해 여러 가지 파라미터를 사용한다. 그림 1 은 SuBSENSE 알고리즘의 흐름도 이다.

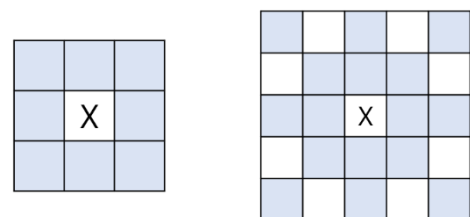


그림 2. LBP 마스크(왼), LBSP 마스크(오)

LBP 는 중심 픽셀  $X$  주변의 8 개의 픽셀을 이용한다. 반면에 LBSP 는 중심 픽셀  $X$  주변의 8 개의 픽셀과 추가로 최외각 부분 8 개의 픽셀, 총 16 개의 픽셀을 이용한다. 마스크 값을 인코딩하는 방법은 LBP 와 유사하다.

$$d(i_p, i_x) = \begin{cases} 1 & \text{if } |i_p - i_x| \leq T_r \cdot i_x \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$LBSP(x) = \sum_{p=0}^{15} d(i_p, i_x) \times 2^p \quad (2)$$

수식 (1)의  $i_p$ 는  $p$  위치의 밝기 값,  $i_x$ 는 중심 픽셀의 밝기 값  $T_r$ 은 noise-free parameter 로서 약 0.3 으로 설정되어 있다.  $d(i_p, i_x)$ 는 이진화 된  $p$  위치의 값이다. 수식 (2)는 이진화 된 마스크 값을 16 bit 숫자로 인코딩하는 과정을 나타낸다. Background sample 은 이렇게 인코딩 된 LBSP 와 RGB color 성분을 가지고 있다. SuBSENSE 의 background sample 들은 시간이 지나갈 때마다 background update parameter  $T^t(x)$ 를 이용하여 배경 성분 업데이트를 한다. 구체적으로 말하자면  $1/T^t(x)$  확률로 현재 입력 영상의 color/LBSP 성분을 50 개의 background sample 중 랜덤하게 업데이트를 진행한다. 랜덤하게 업데이트를 진행함으로써 long term 과 short term 성분을 동시에 가지고 있을 수 있는 효과를 얻게 된다. background update parameter  $T^t(x)$ 는 Feedback Process 모듈에서 매 프레임마다 계산된다.

BG/FG Classification 은 Pixel models 안에 있는 background sample 50 개와 현재 프레임과 color/LBSP distance 를 계산하여 해당 픽셀이 객체 (foreground)인지 배경 (background)인지를 결정하는 모듈이다. 이러한 과정은 수식 (3)과 같이 표현된다.

$$S^t(x) = \begin{cases} 1 & \text{if } N\{dist(I^t(x), B_n^t(x)) < R^{t-1}(x), \forall n\} < 2 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$



그림 3. 입력 영상(왼), 출력 영상(오)

수식 (3)의  $t$ 는 frame index,  $x$ 는 pixel index,  $S^t(x)$ 는 이진 마스크로 표현된 결과 영상(output segmentation map)이다.  $dist(I^t(x), B_n^t(x))$ 는 현재 프레임  $I^t(x)$ 와 배경 샘플  $B_n^t(x)$  사이의 color/LBSP distance 를 반환한다. Color/LBSP distance 에서 color distance 는 L1 distance 를 사용하고 LBSP distance 는 Hamming distance 를 사용한다. 반환된 color/LBSP distance 가 distance threshold  $R^{t-1}(x)$ 보다 작은 개수가 2 보다 작으면 객체로 판단한다.  $R^{t-1}(x)$ 은 주로 움직이는 배경에서 발생하는 오 탐지(false positive)를 막기위한 파라미터로 움직이는 배경에서 높은 값을 가진다.

Feedback Process 는 다양한 상황에서도 강건하게 객체와 배경을 구분하고 상황에 맞게 배경 성분을 업데이트 하기 위해

$R^t(x)$ 와  $T^t(x)$  파라미터를 계산하는 모듈이다. 수식 (4), (5)는 파라미터를 계산하는데 사용되는 파라미터에 관련된 수식이다.

$$D_{min}^t(x) = D_{min}^{t-1}(x) \cdot (1 - \alpha) + d_t(x) \cdot \alpha \quad (4)$$

$$v^t(x) = \begin{cases} v^{t-1}(x) + v_{incr} & \text{if } X^t(x) = 1 \\ v^{t-1}(x) - v_{decr} & \text{otherwise} \end{cases} \quad (5)$$

수식 (4)의  $d^t(x)$ 는 입력 영상과 배경 샘플 사이에 계산했던 모든 color/LBSP 거리를 정규화 한 것의 최소 값,  $\alpha$ 는 learning rate 이다.  $d^t(x)$ 의 값이 작으면 배경, 크면 움직이는 배경 또는 객체일 확률이 높다.  $D_{min}^t(x)$ 는 과거의 값을 feedback 을 받아 신뢰할 수 있는 객체의 위치에 대한 정보를 담고있다.  $v^t(x)$ 는 나뭇잎이나 분수대 같은 움직이는 배경에 대해 수치적으로 표현한 파라미터이다. 수식 (5)의  $v_{incr}, v_{decr}$ 는 각각 1, 0.1 의 값을 가지는 상수이다.  $X^t(x)$ 는 이전 결과  $S^{t-1}(x)$ 와 현재 결과  $S^t(x)$ 에 XOR 연산한 값으로 서로 결과가 다른 경우 1 같은 경우 0 이라는 값을 가진다. SuBSENSE 알고리즘은 움직이는 배경은 이전과 현재 결과가 다른 경우가 주기적으로 발생할 것이라는 가정을 가지고있다. 다시 말해  $v^t(x)$ 는 움직이는 배경 영역에서는 큰 값을 가지고 움직이지 않는 배경 영역에서는 작은 값을 가진다.

수식 (6), (7)은  $R^t(x)$ 와  $T^t(x)$  파라미터를 계산하는 수식이다.

$$R^t(x) = \begin{cases} R^{t-1}(x) + v^t(x) & \text{if } R^{t-1}(x) < (1 + D_{min}^t(x) \cdot 2)^2 \\ R^{t-1}(x) - \frac{1}{v^t(x)} & \text{otherwise} \end{cases} \quad (6)$$

$$T^t(x) = \begin{cases} T^{t-1}(x) + \frac{1}{v^t(x) \cdot D_{min}^t(x)} & \text{if } S^t(x) = 1 \\ T^{t-1}(x) - \frac{v^t(x)}{D_{min}^t(x)} & \text{otherwise} \end{cases} \quad (7)$$

$R^t(x)$ 은  $v^t(x)$ 에 의해서 업데이트 되는 파라미터로서 움직이는 배경에 대해서 높은 수치를 가진다.  $R^{t-1}(x) < (1 + D_{min}^t(x) \cdot 2)^2$  조건은  $R^t(x)$ 의 값을 증가시킬 때 지수적인 관계를 가지게 하기 위함이다. 이는 심하게 흔들리는 나무의 경우 높은  $R^t(x)$  값을 반환하여 오 탐지를 줄일 수 있다.

$T^t(x)$ 는 객체가 탐지된 영역에서는 큰 값을 가지고 배경이 탐지된 영역은 작은 값을 가지도록 진행된다.  $T^t(x)$  값이 작아질 때 움직이는 배경에선 보통 배경에 비해 더 많이 작아진다. 이 역시 움직이는 배경의 color/LBSP 성분을 추가하여 움직이는 배경을 객체로 탐지하는 오 탐지를 줄이는 효과를 준다.

Post-process 모듈은 BG/FG Classification 모듈에서 계산했던 결과 영상  $S^t(x)$ 에 morphology, median filter 연산을 진행하여 noise 를 제거하여 조금 더 깔끔한 영상을 얻는 과정을 수행한다.

### 3. 결론

본 논문에서는 SuBSENSE 알고리즘이 다양한 상황에도 강건하게 객체를 추출할 수 있는 방법에 대해 알아보았다. 입력 영상에 대해 객체와 배경으로 classification 을 하면서 상황을 분석하고 다음 classification 에 사용될 파라미터를 조절하여 객체를 추출하여 여러 가지 상황에서도 강건하게 추출되고 있다. 하지만 여러 가지 상황을 완벽하게 인지하는 것은 아니기 때문에 상황에 따른 문제점들을 해결하기 위한 연구가 활발히

이루어 지고있다. 이러한 문제들이 해결이 된다면 여러 응용 분야에 대한 확장을 기대해볼 수 있다.

#### ACKNOWLEDGEMENT

이 논문은 기업(한화테크윈)의 지원을 받아 수행된 연구임.

#### 4. 참고문헌

- [1] St-Charles, Pierre-Luc, Guillaume-Alexandre Bilodeau, and Robert Bergevin. "Subsense: A universal change detection method with local adaptive sensitivity." *IEEE Transactions on Image Processing* 24.1 (2015): 359-373.
- [2] G.-A. Bilodeau, J.-P. Jodoin, and N. Saunier, "Change detection in feature space using local binary similarity patterns," in *Proc. Int. Conf. Comput. Robot Vis.*, May 2013, pp. 106-112.
- [3] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657-662, Apr. 2006.
- [4] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709-1724, Jun. 2011.
- [5] M. Van Droogenbroeck and O. Paquot, "Background subtraction: Experiments and improvements for ViBe," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 32-37.
- [6] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 38-43.