

색 공간 정보를 이용한 지지벡터 영역 모사 기반의 장면 이해

김수완\*, 장형진\*\*, 강우성\*\*\*, 최진영\*\*\*\*  
 서울대학교 전기·컴퓨터 공학부 인지지는 연구실

SVDD based Scene Understanding using Color Space Information

Soowan Kim\*, Hyung Jin Chang\*\*, Woosung Kang\*\*\*, Jin Young Choi\*\*\*\*  
 Perception and Intelligence Lab, Seoul National University

**Abstract** - 기존 영상감시 시스템의 물체 탐지 알고리즘은 주로 배경 모델링 기법을 기반으로 하고 있다. 이 기법은 차영상 기법보다는 성능이 뛰어나기는 하지만 여전히 정지 카메라에서만 활용이 가능하고, 주변 환경에 따라 알고리즘 상의 많은 임계값을 현재 상황에 맞춰 일일이 조절해 주어야 한다는 한계점이 있다. 따라서 이 논문에서는 배경모델링 기법을 사용하지 않고 입력되는 영상의 Color 정보를 이용하여 영상 내에 있는 여러 대상을 직접 판단하여 관심 있는 물체를 탐지하는 방법을 제안하고자 한다. 제안된 알고리즘은 먼저 현재의 영상을 하나의 물체로 추정되는 영역이 하나의 영역으로 구분되어지게 간단하게 분할해낸다. 그리고 나누어진 영역마다 대표 Color 값을 계산하여 미리 학습된 데이터를 기준으로 Support Vector Domain Description (SVDD) 알고리즘을 사용하여 구별해내고 그 결과를 바탕으로 영역이 무엇인지를 판별해낸다. 이 방법은 정지되어 있는 카메라뿐만 아니라 움직이는 카메라 상에서도 사용되어질 수 있으며 알고리즘 상에서 사용되는 임계값의 종류가 적기 때문에 많은 상황에서 일반적으로 쓰일 수 있다.

1. 서 론

영상 감시 시스템은 영상처리 기술과 다른 센서보다 믿을 수 있는 시각 정보를 이용하기 때문에 보안 감시, 화재 감시, 시스템 감시 등에 효율적으로 이용될 수 있다. 영상 감시 시스템을 이용하지 않은 채 감시자가 24시간 감시를 하고 모든 상황에서 정확하게 대상 물체를 탐지하고 추적 해낸다는 것은 불가능하기 때문에 강한 영상 감시 시스템이 필요하다.

영상 감시 분야는 크게 물체 탐지, 물체 추적, 물체 인식 이 세 가지 분야로 나뉘어진다. 물체 탐지는 관심 대상 물체를 현재 영상에서 찾아내는 방법이고, 물체 추적은 관심 대상 물체의 움직임을 포착하여 이동 경로를 파악하고 현재 위치를 찾아내는 방법이다. 그리고 물체 인식은 탐지한 물체가 무엇인지를 판단하는 방법이다. 본 논문에서 다루는 scene understanding 기법은 이 중 물체 탐지와 물체 인식을 하는 알고리즘으로 이다. 기존의 물체 탐지 알고리즘으로는 차영상 기법, 배경 모델링 기법, 광흐름 기법 등이 있는데 Stuffer와 Grimson의 가우시안 혼합 모델링을 사용하는 배경 모델링 기법[1]이 가장 널리 사용된다. 이 알고리즘은 감시 시스템의 관심 대상 물체를 탐지해내고 흔들리는 나무나 깃발, 일부 파도 등의 여러 노이즈를 제거 하는 데에 강한 성능을 보여준다. 하지만 배경 모델링 방법은 모든 픽셀의 과거 값을 배경 히스토리로 저장하고 그 히스토리를 기준으로 물체를 탐지하는 방법이기 때문에 영상 전체가 흔들리거나 움직이는 카메라 상에서는 정확한 배경을 만들 수 없고 그 결과 물체 탐지를 할 수 없게 된다. 정지카메라 상에서의 감시시스템 또한 실제 환경에서는 바람이나 비 등 여러 환경 요인들이 있기 때문에 배경 모델링 기법은 한계가 있다고 할 수 있다.

이런 배경 모델링 기법의 문제점은 scene understanding 기법을 통해 해결 될 수 있다. 이는 픽셀 단위로 배경을 만드는 것이 아니라 특징 값을 학습하여 현재 영상을 판단하는 방법으로 A.Hoogs 에 의해 제안되었다. [2] A.Hoogs는 논문에서 edge를 기반으로 scene segmentation을 하고 여러 종류의 특징 값들을 추출하여 Multiclass-Adaboost [3] 을 사용하여 scene understanding 하는 방법을 소개하였다.

본 논문에서는 명도를 이용한 segmentation을 하고 그 결과 영역에서 색 공간 정보를 추출 하여 SVDD classification을 통해 현재 영상을 분류하는 방법을 제안하였다. 여러 상황의 다른 동영상들로 이 알고리즘을 테스트 하고 그 결과를 실제 값과 비교하여 그 정확도를 평가하도록 한다.

2. 배경 이론

2.1. SVDD

SVDD는 hyperplane을 가정하여 패턴을 분류할 수 있는 판단 경계면(decision boundary)을 학습하는 방법과 유사하지만, 특징공간에서의 클래스의 경계를 hyperplane이 아닌, a를 중심으로 하고, R을 반지름으로 하여 가능한 모든 학습 영역을 포함하는 hypersphere를 가정하여 학습한다는 차이점을 가지고 있다. [4]

SVM과 유사하게 hypersphere에 의해 결정되는 목적 함수를 정의하면 다음과 같다.

$$F(R, a) = R^2 \tag{1}$$

제약조건:  $\|x_i - a\|^2 \leq R^2, \quad \forall i$

여기서 hypersphere의 외곽 경계면 근처에 있는 학습 데이터에 대한 학습을 고려하기 위해, slack 변수  $\xi_i \geq 0$ 을 도입하여 위의 식을 다시 표현하면,

$$F(R, a) = R^2 + C \sum_i \xi_i \tag{2}$$

제약조건:  $\|x_i - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i$   
 와 같이 된다. 위의 식에서 모수 C는 구면체의 부피와 오차사이의 절충 값을 조절한다.

목적함수와 제약조건을 라그랑지 승수법을 사용하여 표현하면,

$$L(R, a, \alpha_i, \gamma_i, \xi_i) = R^2 + C \sum_i \xi_i - \sum_i \alpha_i \{R^2 + \xi_i - (\|x_i\|^2 - 2a \cdot x_i + \|a\|^2)\} - \sum_i \gamma_i \xi_i \tag{3}$$

와 같다.

여기서 라그랑지 승수들은  $\alpha_i \geq 0, \gamma_i \geq 0$ 가 되며, L은 R, a,  $\xi_i$ 에 관해서는 최소화 되어야하고,  $\alpha_i, \gamma_i$ 에 관해서는 최대화되어야 한다. L을 R, a,  $\xi_i$  각각에 대해서 편미분한 결과를 0으로 놓은 후, 이를 QP을 이용하여 학습한다. 라그랑지 승수  $\alpha_i$ 가 학습 데이터  $x_i$ 에 대해  $0 < \alpha_i < C$ 를 만족 하는 데이터  $x_i$ 들을 support vector (SV)라 하고,  $\alpha_i = C$ 인 데이터  $x_i$ 들을 바깥 점(outlier)이라고 한다.

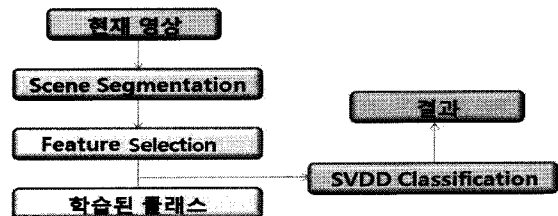
새로운 데이터 z에 대한 클래스를 판단하기 위해서는 sphere의 중심으로부터의 거리를 계산해야 한다.

$$\|z - a\|^2 = K(z, z) - 2 \sum_i \alpha_i K(z, x_i) + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \leq R^2 \tag{4}$$

여기서 K는 non-separable 문제를 해결하기 위하여 도입된 커널 함수를 나타낸다. 계산 결과 거리가 반지름 R보다 작거나 같으면 데이터 z는 같은 클래스에 속하는 것으로 판단된다.

3. Scene understanding

본 논문에서 다루고 있는 색 공간 정보를 이용한 scene understanding 기법은 크게 세 가지 단계로 이루어진다. 현재 입력되는 영상을 여러 개의 영역으로 분할하는 scene segmentation 단계, 각 영역에서 특징 정보를 추출하는 단계, 그리고 그 정보와 미리 학습된 정보를 통해 영역을 판단하는 단계이다. 전체 알고리즘의 흐름은 <그림 1> 과 같다.

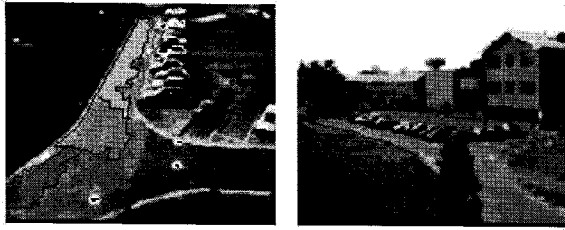


<그림 1> Scene Understanding

3.1 Scene Segmentation

Scene Understanding은 영상의 각 부분을 미리 학습된 training set 중의 하나로 판단하는 방법이고 이를 영상감시에 사용하기 위해서는 영상의 모든 부분에 대한 판단 결과가 필요하다. 하지만 보통 영상 감시에서 사용하는 영상은 그 픽셀 수가 많기 때문에 픽셀 단위로 판단을 내리는 것은 보다 정확할 수 있지만 알고리즘 수행 시간이 너무 많이 든다는 단점이 있다. 또한 픽셀 단위로 클래스 판단을 할 경우에는 판단에 사용할 수 있는 특징의 종류가 그 픽셀의 색 정보 등 극히 단순한 정보밖에 없다는 단점이 있다고 할 수 있다. 이런 여러 단점을 보완하기 위해 하나의 물체와 다른 물

체는 그 경계에서 명도가 바뀐다는 가정을 통해 영상에 대해 Scene Segmentation을 하였다.



<그림 2> Intensity Scene Segmentation result

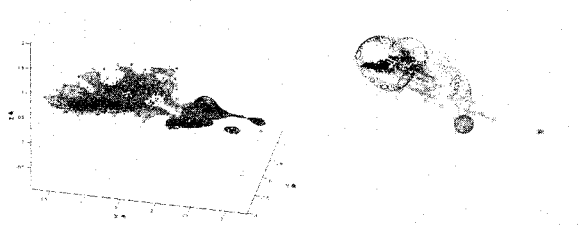
<그림 2>에서 볼 수 있듯이 segmentation 결과 명도가 비슷한 근처 픽셀들이 하나의 영역으로 결합되었음을 확인할 수 있다. 위 그림에서 하나의 색으로 표현된 부분을 일부 빨간색으로 표시 하였으며 하나의 영역으로 분할된 영역이다. 이를 SVDD 학습의 데이터로 사용하였다. 하나의 물체를 정확하게 segmentation 하는 데에는 복잡한 알고리즘이 필요하고 연산량이 매우 크기 때문에 명도만을 이용하여 각 영역을 분할하였다. 위 Segmentation 알고리즘은 명도 차이만을 이용하였기 때문에 속도가 매우 빠르고, 전체 알고리즘 속도에 크게 영향을 주지 않음을 알 수 있었다. 또한 이 방법을 사용하여 학습함에 있어서의 test set의 크기를 현저하게 줄일 수 있었으며, 이는 전체 알고리즘의 속도 개선에 매우 큰 역할을 한다.

### 3.2 Feature selection

위의 segmentation 된 결과 내에서 의미 있는 색 공간 정보를 추출하는 단계이다. 본 논문에서는 특징 값을 추출하기 위해 segmentation 된 모든 영역들에서 영역 안의 픽셀들을 LUV 색 공간으로 변환하였는데 실험결과 RGB 색 공간에서보다 성능이 뛰어남을 알 수 있었다. 색 공간 변환 후 각 영역 내 픽셀 값들의 평균을 각 영역의 대표 값으로 정하였다. 이 추출된 정보(특징 값)는 다음 classification 단계에서 사용하였고 대표 값의 분류 결과에 따라 해당 영역이 모두 하나의 동일한 결과를 갖게 된다. 색 공간 정보, texture 정보, 기하학적 정보, 주변 픽셀과의 관계 정보등 사용할 수 있는 Feature의 종류가 매우 많기 때문에 상황에 따라 적절한 Feature를 고르는 것이 알고리즘의 결과를 향상 시킬 수 있는 데에 중요한 부분이라고 할 수 있다. 본 논문에서는 가장 간단한 색 공간 정보만을 이용하여 scene understanding을 하였는데 색 공간 정보만 사용 하는 것은 계산은 쉽지만 그 판단 기준이 다른 여러 특징들을 사용할 때에 비해 떨어질 수 있기 때문에 여러 종류의 특징들을 learning에 고려함으로써 알고리즘의 성능이 개선될 여지가 있다. [5][6]

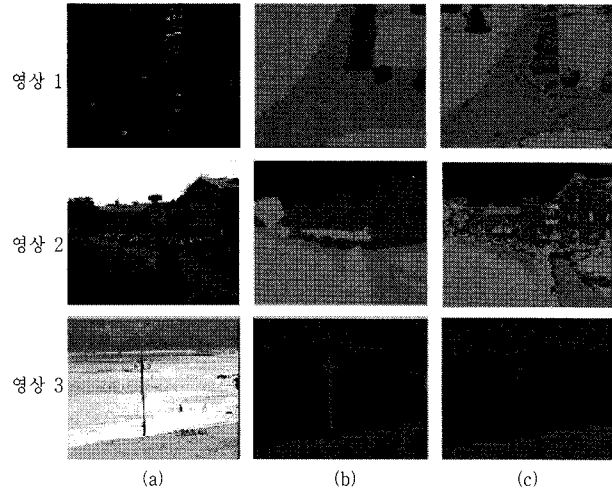
### 3.3 SVDD classification

3.1과 3.2에서 설명된 Scene Segmentation을 통해 얻어진 영역에서 추출된 특징들을 2.1에 설명된 SVDD를 통해 classification을 하였다. 학습을 하기 위해 영상의 첫 프레임들 직접 나누어 색 공간 정보에 따른 실제 라벨을 만들어 주고 학습과정을 통해 support vector를 구하여 다음 프레임부터 SVDD classification을 하기 시작한다. 이때 segmentation된 영역을 test하며 그 결과 값을 영역의 클래스 값으로 지정해 준다.



<그림 3> SVDD 학습

<그림 3>는 테스트 영상 중 두 개의 SVDD fh 묘사된 모습으로 각 set의 support vector에 의해 training set이 하나의 라벨로 묶여진 것을 볼 수 있다. 이 과정을 통해 현재 영상의 각 영역이 여러 training set 중에 어떤 set에 속하는지를 판단하여 scene understanding을 한다.



<그림 4> SVDD classification 결과

<그림 4>에서 (a) 열은 영상 감시 시스템의 현재 영상, (b) 열은 현재 영상의 실제 값, (Ground truth) 그리고 (c) 열은 본 논문에서 제시한 알고리즘을 통해 분류된 결과이다. 영상 2의 경우 나무의 실제 색 정보가 여러 종류이기 때문에 알고리즘 테스트 시 두 training set을 나누라고 할당하였다. 이 영상에서는 앞의 나무가 흔들리는 것과 상관없이 정확하게 판단을 내리는 것을 볼 수 있다. 영상 3의 경우 계속해서 파도가 치지만 역시 정확하게 파도 부분을 분류함을 볼 수 있다.

<표 1> 실제 값과 알고리즘 결과 비교

영상	클래스 수	정확도
1	3	82%
2	6	71%
3	3	87%

## 3. 결 론

본 논문에서는 물체 탐지에 있어서 널리 사용되고 있는 배경 모델링을 이용하는 방법들의 한계점인 움직임은 영상 내에서의 탐지와 물체 인식을 위해 SVDD를 이용하여 보다 정확한 탐지와 인식을 하는 scene understanding에 관한 알고리즘을 제안하였다. 위의 <그림 4>와 <표 1>에서 볼 수 있듯이 여러 영상에서 테스트 해 본 결과 실제 값과 비슷한 결과를 낼 수 있었다. 제안된 색 공간 정보를 이용한 scene understanding 알고리즘의 세 단계인 scene segmentation, feature selection, SVDD classification은 그 자체로도 컴퓨터 비전 분야에서 많이 연구되는 분야로서, 추후 이 세 분야에 관한 연구를 통해 보다 강인한 segmentation, 여러 종류의 feature, 보다 빠르고 정확한 SVDD classification을 통해 전체 알고리즘의 성능 향상을 도모할 계획이다. 또한 각 알고리즘들을 조합하는 방법이나 다른 데이터들을 이용하여 연구를 진행하여 특정 상황에 맞는 scene understanding 알고리즘을 개발하고 이를 실제 활용해 볼 계획이다.

## [참 고 문 헌]

- [1] C. Stauffer and W.E.L. Grimson, "Adaptive Background mixture Models for Real-time Tracking", CVPR99, June, 1999
- [2] J. Kaufhold, R. Collins, A. Hoogs, and P. Rondot. Recognition and segmentation of scene content using region-based classification. In IEEE International Conference on Pattern Recognition, Hong Kong, 20 - 24 Aug. 2006
- [3] Robert E. Schapire, "A Short Introduction to Boosting", Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999.
- [4] David M.J. Tax, "Support Vector Data Description", Machine Learning, vol. 54, pp. 45-66, 2004.
- [5] J.Kaufhold, A.Hoogs, Learning to Segment Images Using Region-Based Perceptual Features, IEEE Computer Vision and Pattern Recognition 2004
- [6] C.J. Setchell, N.W. Campbell, Using Colour Gabor Texture Features for scene understanding.