

SVM을 이용한 스테레오 비전 기반의 사람 탐지

* 정상준 (고려대학교 메카트로닉스학과)
 ** 송재복 (고려대학교 기계공학과)

Stereo Vision based Human Detection using SVM

Sang-Jun Jung (TEL: 02-929-8501, E-mail: hound98@korea.ac.kr)
 Jae-Bok Song (TEL: 02-3290-3363, E-mail: jbsong@korea.ac.kr)

Abstract - A robot needs a human detection algorithm for interaction with a human. This paper proposes a method that finds people using a SVM (support vector machine) classifier and a stereo camera. Feature vectors of SVM are extracted by HoG (histogram of gradient) within images. After training extracted vectors from the clustered images, the SVM algorithm creates a classifier for human detection. Each candidate for a human in the image is generated by clustering of depth information from a stereo camera and the candidate is evaluated by the classifier. When compared with the existing method of creating candidates for a human, clustering reduces computational time. The experimental results demonstrate that the proposed approach can be executed in real time.

1. 서 론

과거의 로봇은 공장에서 사람의 일을 대신하여 주는 산업용 위주인 반면에, 요즘에는 가정환경에 적용되어 사람에게 각종 서비스를 제공하는 추세이다. 이런 가정환경에서는 사람과 로봇의 상호작용을 위해 로봇의 사람 탐지 기술이 필수적이다. 사람을 탐지하기 위해서 사람의 외형 정보를 제공할 수 있는 비전센서가 주로 사용되고 있다.

비전 영상에서 사람을 탐지하기 위해 이전 영상과 현재 영상의 차이를 이용하여 움직이는 물체를 탐지하여 사람으로 간주하는 기법이 제시되었다 [1]. 그러나 이 방법에서는 사람이 아닌 움직일 수 있는 물체도 사람으로 인식하는 문제가 있다. 사람의 외형을 실루엣으로 변환하여 템플릿 정합의 유사도를 통해 사람을 탐지하는 기법이 제시되었으나, 사람의 다양한 외형을 탐지하지 못하는 단점이 있다 [2]. 이러한 단점을 보완하기 위해 다양한 사람의 이미지를 특징벡터로 사용하여 학습 알고리즘을 적용시키는 연구가 활발히 진행되고 있다. Dalal & Triggs는 사람이 존재하는 이미지들과 사람이 없는 이미지들을 기울기 히스토그램 기법(histogram of gradient)을 사용하여 특징벡터를 생성하였다[3]. 이 특징벡터들을 SVM기법으로 학습시켜 사람 탐지 분류기(classifier)를 생성하였는데, 이 분류기는 사람 탐지에 우수한 성능을 보여주었다. 하지만 한 영상(320 x 240의 해상도 기준)에서 사람을 탐지하기 위해서는 1000번 이상 조밀하게 검색하기 때문에 시간이 오래 걸려 실시간성이 중요한 로봇의 수행에 적합하지 않다.

본 논문에서는 검색시간을 줄이기 위하여 스테레오 카메라의 거리정보를 군집하는 기법을 적용하였다. 입력영상을 조밀하게 검사하지 않고 군집된 영역에서만 분류기로 평가하기 때문에 검색시간을 줄일 수 있다. 이를 통해 로봇이 사람 탐지 외의 다른 서비스에 자원을 활용할 수 있다.

본 논문은 다음과 같이 구성된다. 2.1절에서는 SVM 학습에 필요한 벡터를 추출하는 알고리즘인 R-HOG를 소개하고, 2.2절에서는 SVM 재학습을 통해 오류에 강인한 분류기의 생성과정을 설명한다. 2.3절에서는 생성된 분류기를 거리정보 기반의 사람 탐지 알고리즘에 적용하는 과정을 다룬다. 2.4절에서는 본 논문에서 제안하는 사람탐지 기법과 기존의 사람탐지 알고리즘을 실제 실험을 통하여 비교하고, 3절에서는 결론을 도출한다.

2. 사람 탐지 알고리즘

2.1 벡터 추출 알고리즘

벡터 추출 알고리즘으로 R-HOG, R2-HOG, C-HOG 등 다양한 방법이 제시되었는데 [3], 본 논문에서는 기본적인 방법인 R-HOG를 사용하였다. 먼저, 영상 내에서 가장자리를 제외한 모든 화소의 기울기와 기울기의 크기를 구한다. 만약, 해상도 66 x 130의 영상이라면 가장자리 화소를 제외한 64x128 영상에서 각 화소가 갖는 값에 대한 기울기를 식 (1)을 통하여 계산한다.

$$\text{Gradient}(x, y) = \frac{I(x, y-1) - I(x, y+1)}{I(x+1, y) - I(x-1, y)} = \frac{dy}{dx} \quad (1)$$

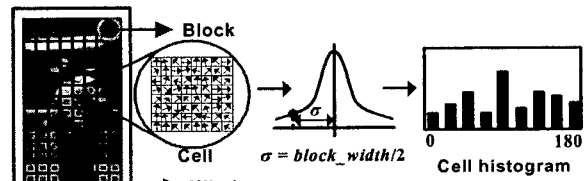
여기서 I 는 입력영상을 의미하고, x 와 y 는 각각 영상의 가로좌표와 세로좌표를 나타낸다. 셀(cell)은 히스토그램을 작성하기 위한 화소단위의 영역이다. 셀 히스토그램을 얻기 위해 다음과 같은 절차가 필요하다.

1. 화소의 기울기에 삼각 선형보간법(trilinear interpolation)을 적용시켜 히스토그램의 가로축에 해당하는 기울기의 각도와 가중치를 얻는다.
2. 블록은 2차원 셀의 집합으로 영상에서 특징벡터를 추출하는 화소단위이다. 블록의 중심과 화소까지의 거리(d), 기울기 크기(m), 표준편차($\sigma = 0.5 \times$ 블록의 길이)와 식 (2)을 통해 가우시안 가중치를 계산한다.

$$W(m, d, \sigma) = \frac{m}{2\pi\sigma^2} e^{-m/2\sigma^2} \quad (2)$$

3. 단계 1을 통해 얻은 기울기의 각도에 단계 2로부터 얻은 가우시안 가중치를 셀 히스토그램에 누적시킨다.
4. 셀의 모든 화소에 단계 1-3을 적용한다.

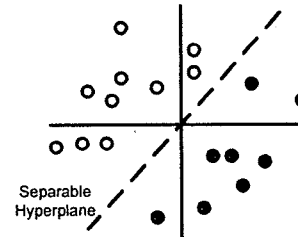
<그림 1>은 단계 1-4의 과정을 통해 셀의 기울기 정보가 히스토그램으로 축적되는 과정을 표현한 것이다. 블록 안에 있는 모든 셀의 히스토그램 작성이 완료되면 블록 정규화(block normalization)한다. 정규화 과정까지 거친 히스토그램은 블록에 해당하는 영상 영역에서 추출된 최종적인 특징 벡터이다. 이와 같은 방법을 영상의 모든 블록에 적용하여 영상내의 모든 특징 벡터를 추출한다. 그리고 영상에서 특징벡터를 추출한 영역을 윈도우(window)라고 정의한다.



<그림 1> 셀 히스토그램

2.2 오류에 강인한 SVM 분류기 생성

SVM은 주어진 두 클래스에서 최적의 분류 경계를 찾는 알고리즘이다. <그림 2>는 2차원 특징공간에서 흰 공과 붉은 공을 구분 할 수 있는 분리 경계(separable hyperplane)를 보여준다.



<그림 2> 분리 경계

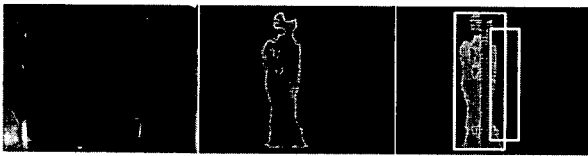
본 논문에서는 사람이 존재하는 영상과 사람이 존재하지 않은 영상에서 추출한 벡터를 SVM 기법으로 학습하여 최적의 사람탐지 판단모델을 찾는다. 학습 데이터에서 사람이 존재하는 영상의 경우 영상 내에서 사람이 있는 위치와 크기가 일관성이 있어야 한다. 그렇지 않으면 잘못된 특징벡터를 추출하게 되거나 특징벡터의 크기가 달라질 수 있기 때문이다. 반면에, 사람이 없는 영상에서는 벡터 추출 위치와 개수가 상관 없기 때문에 다수의 윈도우에서 벡터 추출이 가능하다.

학습 이미지에서 추출된 특징벡터들을 SVM으로 학습시킨 후, 오류에 강

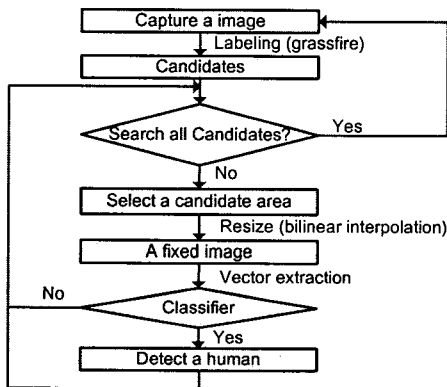
인한 분류기 생성을 위하여 재학습 과정을 거친다. 이 과정을 위해, 처음 생성된 분류기를 사람이 없는 영상에서 일정한 화소 단위로 이동하면서 윈도우 영역에 사람이 존재하는지를 평가한다. 분류기가 사람이 존재한다고 판단하는 경우, 이는 사람이 존재하지 않는 영상에서 사람으로 판단한 것이므로 오류가 발생한 것이다. 이 윈도우에 있는 벡터를 기존의 특징벡터 데이터에 추가한다. 이 과정을 사람이 없는 모든 학습 영상에 적용하면 기존의 특징벡터 데이터에 오류가 발생된 벡터들이 추가하게 된다. 갱신된 특징벡터 데이터를 다시 학습하게 되면 사람이 아닌 영역을 사람으로 탐지하는 오류(false positive rate)에 강한 분류기를 만들 수 있다.

2.3 거리정보 기반의 사람탐지 알고리즘

스테레오 카메라는 두 카메라의 간격과 사물의 각도를 통해 사물의 거리를 얻기 때문에 각 화소의 거리정보 값을 얻을 수 있다. 영상에서 비슷한 거리값을 갖는 화소들을 레이블링(labeling)하면 비슷한 거리에 있는 사물을 얻을 수 있다. 거리정보를 군집하기 위해 grassfire 알고리즘[5]을 적용하여 영상 내에서 사람과 배경을 구분하였다. <그림 3>은 레이블링을 적용하여 사람 후보자들을 추출한 결과이다. (a)의 원본영상에 대응하는 거리정보를 기반으로 (b)의 영상을 만들고, 레이블링을 적용하여 (c)의 영상과 같이 사물을 군집한다. 군집된 사물들은 크기가 일정하지 않기 때문에 학습 데이터와 일치하기 위해 양선형 보간법(bilinear interpolation)을 사용한다. 크기가 일치된 후보자들을 R-HOG 알고리즘을 통해 특징벡터를 추출한다. <그림 4>는 위 과정을 도식화한 것이다.



<그림 3> (a) 원본영상, (b) 거리정보 영상, (c) 군집 영상


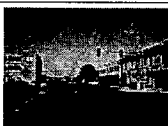


<그림 4> 스테레오 카메라의 거리 정보를 이용한 사람 탐지 흐름도

2.3 실험

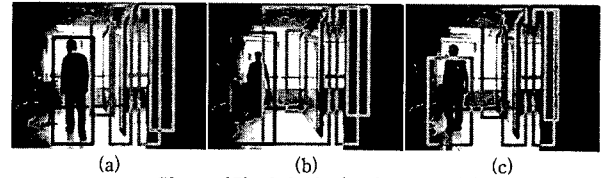
본 논문의 실험을 위해 Pentium 4-M 2.2GHz 컴퓨터와 Videre design사의 스테레오 카메라인 STH-MDCS-C이 장착된 로봇을 사용하였다. SVM-Light[5]를 사용하였다. 본 실험에서의 셀과 블록, 블록 이동 크기(stride size)는 8x8 화소, 2x2 셀, 8화소로 설정하였다. 히스토그램 가로축 크기는 9이고, 기울기 각도의 범위는 0°-180°로 제한하였다. 초과하는 기울기에 대해서는 무시하였다. 히스토그램의 벡터 크기를 0.2를 넘지 않도록 제한하는 L2-Norm 정규화 기법을 적용시켰다. 이 경우 하나의 윈도우에서 105 블록, 3780 차원의 특징벡터를 추출 할 수 있다.

<표 1> INRIA person data set

	사람이 있는 영상 (positive image)	사람이 없는 영상 (negative image)
해상도	 96x128(fixed size)	 No fixed images
파일 개수	2,416 images	1,261 images

<표 1>에서 INRIA에서 제공하는 사람 데이터 집합(<http://lear.inrialpes.fr/data/human>)의 영상을 보여주고 있다. 이 영상들로부터 학습 벡터를 추출한다. 사람이 있는 영상(positive images)의 경우 크기가 일정한 96x160이지만, 가장자리 16픽셀을 제외한 64x128 영역만 사람이 있는 영역으로 학습을 시킨다. 사람이 없는 영상(negative images)은 한 영상마다 10 윈도우를 추출하였고, 오차를 줄이기 위한 재학습에서 윈도우 이동 단위를 16 화소 설정하였다.

추출된 특징벡터를 SVM-Light로 학습하여 생성된 분류기를 사람 탐지 알고리즘에 적용시켰다. <그림 4>는 사람 탐지에 성공한 경우와 실패한 경우를 보여준다. 붉은 사각형은 사람이라고 판단한 영역이고 초록색 사각형은 후보자들이다. <그림 4> (a)와 (b)는 올바른 거리 정보로 인해 군집이 잘되어 탐지를 성공한 반면에, <그림 4>의 (c)는 노이즈와 같은 이유로 잘못된 군집으로 탐지를 실패한 경우이다.



<그림 4> 사람 탐지 성공(a,b) 및 실패(c)

기존의 알고리즘은 리눅스 버전으로 웹 사이트(<http://pascal.inrialpes.fr/soft/olt/>)에서 제공되고 있다. 본 논문과 성능 비교를 위해 실행 파일을 사용하여 <표 1>의 INRIA 영상 데이터를 학습 시켰다. 영상의 모든 영역을 조밀하게 탐지하는 방법보다 거리정보를 기반으로 후보자 영역만 분류기로 평가하는 것이 빠르며, <표 2>에 본 논문에서 제시한 알고리즘과 기존의 알고리즘의 성능을 비교한 결과가 나타나있다.

<표 2> 거리 정보 기반 알고리즘과 기존의 알고리즘 비교

	거리정보 기반 알고리즘	기존의 알고리즘
이미지 당 윈도우 (해상도 320 x 240)	0-10 window	3412 윈도우
소요 시간 (해상도 320 x 240)	20-230 ms	2-3 sec

3. 결 론

본 논문에서는 스테레오 카메라의 거리정보를 이용하여 영상에서 사람의 위치를 실시간으로 탐지하는 방법을 제시하였다. 기존의 사람탐지 알고리즘은 높은 성공률을 보이지만, 시간이 오래 걸리기 때문에 로봇에 적합하지 않았다. 본 논문에서는 스테레오 카메라의 거리정보를 사용하여 검색영역을 사람 후보자 영역으로 한정시킴으로써 사람을 탐지하는 데 소요되는 계산 시간을 감소시켰다. 따라서 본 논문에서 제안한 사람 탐지 알고리즘은 로봇에 실시간으로 적용할 수 있으므로 실용적이다.

[후 기]

이 논문은 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었음.

[참 고 문 헌]

- [1] 임종석,곽원욱,이동카메라 영상에서 히스토그램과 컬러 정보를 이용한 다수 보행자 검출 및 추적, 전자공학회논문지 제 41권 SP편 제 5호,2004. 9
- [2] M. Dimitrijevic, V. Lepetit and P. Fua, Human Body Pose Detection Using Bayesian Spatio-Temporal Templates, Computer Vision and Image Understanding, Vol. 104, Nr. 2, pp. 127 - 139, December 2006.
- [3] N.Dalal, B.Triggs, Histograms of Oriented Gradients for Human Detection, in: Proceedings of the Conference on Computer Vision and Pattern Recognition, June 2005, vol. 2, p. 886-893
- [4] R. M. Haralick, L. G. Shapiro, Computer and Robot Vision, Addison-Wesley, 1992.
- [5] Thorsten Joachims, Learning to Classify Text Using Support Vector Machines. Dissertation, Kluwer, 2002.