한국컴퓨터정보학회 동계학술대회 논문집 제20권 제1호 (2012. 1)

HOG와 SVM을 사용한 지하철 실내 픽토그램 인식

김설호°, 최형일**, 김계영*
°*숭실대학교 컴퓨터학과
**숭실대학교 미디어학과
e-mail: sulho@ssu.ac.kr, hic@ssu.ac.kr, gykim11@ssu.ac.kr

Pictogram Sign Recognition in Subway using HOG and SVM

Sul-Ho Kim°, Hyung-Il Choi**, Gye-Young Kim*
°*Dept. of Computing, Soong-Sil University
**Dept. of Media, Soong-Sil University

● 요 약 ●

지하철 실내의 표지판은 다양한 정보들을 포함하고 있다. 그 중에서 시설물의 형상을 단순화하여 나타낸 픽토그램 사인은 국제적으로 표준화 되어 있어 검출이 용이하다고 볼 수 있다. 일반적으로 객체를 인식하는 방법은 후보영역을 설정하는 검출단계와 후보영역을 인식하는 단계로 나뉘어 진다. 본 논문에서는 후보영역 검출단계에서 단일 값을 가지는 이진화로 픽토그램 영역이 분리가 잘 안되는 문제점을 해결하기 위하여 적응적 이진화를 사용하였고 인식을 위한 특징추출로 HOG서술자를 사용하고 학습과 인식에는 SVM을 사용하였다. 실험 결과를 통하여 HOG서술자로 픽토그램 사인을 인식하는 것이 적합한 것인지 확인한다.

키워드: 사인인식(sign recognition), 적응적 이진화(adaptive thresholding), 호그(hog)

I. 서론

객체 인식의 한 분류인 표지판 인식은 일반적으로 후보영역 검출과정과 인식과정을 거친다. 자연영상은 빛과 눈, 비등 다양한 환경변화들이 존재하기 때문에 다양한 문제점이 발생하게 되고 이에 대한 처리를 위해서는 많은 처리 시간이 소요된다. 반면에 지하철 실내의 자연영상은 빛의 변화등의 자연현상에 대하여 비교적 제한된 환경을 가지고 있다. 픽토그램은 그림(Picture)와 전보(Telegram)의 합성어로 사물이나 시설을 상징적으로 나타내며 의미를 빠르게 전달하는 상징문자이다. 지하철 실내의 표지판에는 이러한 픽토그램이 다수 사용되고 있으며 국제표준화 되어있기에 검출이 용이하다. 본 논문에서는 지하철 실내에서의 픽토그램을 인식하는 과정에 관하여 서술한다. 앞서 말한 과정과 동일하게 후보영역 검출과정과 인식과정으로 나뉘어지며 HOG서술자로 특징을 추출하여 SVM을 통해 학습과 인식을 한다.

그림 1-1. 픽토그램 사인의 예

II. 관련 연구

기존의 방법들은 주로 도로 표지판 인식에 많이 사용되고 있으며 표지판의 색상을 이용한 방법과 모양을 이용한 방법, 그리고 이 방법들을 병합한 방법이 주로 사용되고 있다. 이러한 방법들을 통하여 얻어진 후보영역으로 특징벡터를 추출하여 학습하는 과정을 가지거나 템플릿매칭등을 수행하여 최종적인 인식단계에 이르게 된다. 픽토그램은 주로 검정색과 흰색으로 구성되어있으며 일부 빨간색, 노란색, 파란색등이 사용되기도 한다. 대부분의 경우 배경이 검정색, 내부가 흰색으로 이루어져 대비를 통해 정보를 전달하기 때문에 컬러정보를 이용하는 것보다 그레이 이미지를 사용하는 것이 연산속도나 일반화적인 면에서 더 유리하다. 따라서 입력영상을 그레이 이미지로 바꾸어서 후보영역 검출 과정을 진행한다. 또한 픽토그램 사인이 영상에서 차지하는 비중은 작은 편이다. 이처럼 작은 영역을 차지하는 픽토그램에서 SIFT나 SURF를 사용하여 특징을 추출할 경우 특징점이 추출되지 않거나 추출이 되더라도 매칭을 시도하기 어려웠다. 따라서 픽토그램이 가지는 대비와 모양을 활용할 수 있는 방법이 필요하다. HOG서술자는 주로 보행자 인식에 많이 사용되는 방법으로 배경과 객체간의 대비에 강건하고 다소의 기울어짐과 이동등의 노이즈에 대해 강건한 특징으로 알려져 있다. HOG서술자는 에지의 방향성분과 크기의 히스토그램을 이용한 방법이다. 따라서 픽토그램의 인식에 적합할 것으로 예상할 수 있다. 따라서 본 논문에서는 보행시의 카메라 흔들림과 회전에도 강건한 모습을 보여줄 수 있는 HOG서술자를 사용

하여 특징벡터를 추출하고 일반화성능이 뛰어나기 때문에 다양한 환경에서 우수한 성능을 보여줄 수 있는 SVM을 사용하여 학습과정을 통해 분류기를 만들고 인식하는 과정을 가진다.

III. 본 론

앞서 그레이이미지로 변환한 영상은 이진화 과정을 통해 픽토그램에 해당하는 전경과 배경에 해당하는 영역으로 나누어 주어야 한다. 이진화 방법에는 여러 가지 방법이 사용되고 있다.

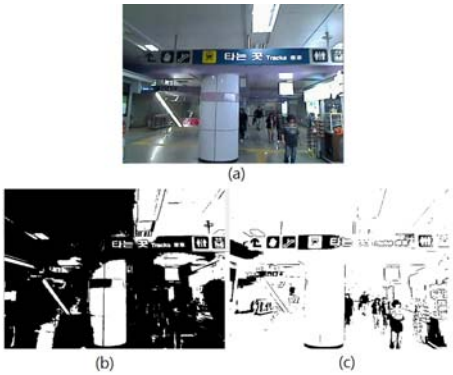

그림 3-1. 이진화 결과영상.
(a)입력영상. (b)Otsu이진화. (c)임계값 50

[그림 3-1]은 이진화 방법에 따른 픽토그램 영역의 분리 상황을 보여주고 있다. 영상에서 픽토그램은 작은 부분만을 차지하기 때문에 영상전체에 대하여 수행하는 Otsu이진화 같은 방법은 픽토그램 영역에서 손실을 가져올 가능성이 높다. 또한 여러 영상에서의 실험을 통해 얻어낸 임계값 50을 사용할 경우 비교적 픽토그램 영역을 잘 추출하지만 천장에 조명이 존재하여 표지판 영역에 영향을 미칠 경우 픽토그램 영역에 조명에 따른 그라데이션 효과가 나타나기 때문에 단일 임계값으로는 좋은 결과를 얻어낼 수가 없다. 따라서 픽셀마다 다른 임계값을 적용시키는 방법이 필요하다.

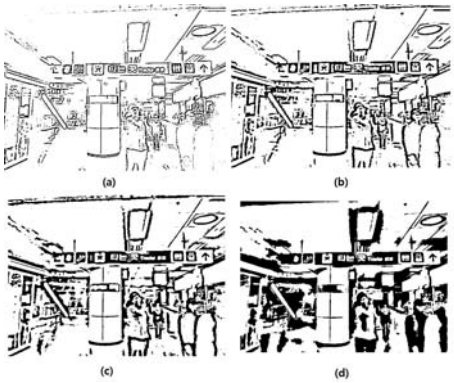

그림 3-2. 이웃픽셀 범위에 따른 이진화 결과
(a)이웃픽셀범위 7. (b)이웃픽셀범위 14.
(c)이웃픽셀범위 21. (d)이웃픽셀범위 49.

적응적 이진화는 기준 픽셀을 중심으로 주변 이웃범위의 픽셀 값을 이용하여 기준 픽셀의 임계값을 결정하는 방법이다. 적응적 이진화의 임계값을 구하는 방법으로는 이웃 범위의 픽셀 값의 평균, 중간 값, 최소최대평균등의 방법등 많은 방법이 있다. 본 논문에서는 3가지 방법 중 비교적 좋은 결과를 보여주는 평균 값을 임계값으로 설정하였다. 실험에서는 로지텍 C905 200만화소 웹캠으로 640 * 480으로 영상을 촬영하였는데 이 때 픽토그램의 사이즈는 10 * 10픽셀에서 50* 50픽셀 정도의 크기이며 25 * 25 ~ 45 * 45범위가 특징 추출에 가장 적합하다고 판단하였다. 그 이유는 너무 작으면 추출된 특징의 비교가 어렵고 너무 크게 되면 픽토그램의 내부 그림까지 객체단위로 분할되기 때문이다. 이와 같은 사이즈를 가지는 픽토그램 후보영역 추출을 위해 적응적 이진화를 수행하게 된다. 이 때, [그림 3-2]와 같이 주변이웃범위에 따라 이진화되는 영상의 결과가 달라지는데 이웃범위가 작으면 픽토그램 영역의 안 쪽 부분에 흰색 구멍이 뚫리게되고 범위가 너무 크면 픽토그램이 배경영역과 달라붙는 현상이 나타난다. 또한 이웃범위가 커질수록 계산량이 증가하게 되기 때문에 적합한 값을 찾는 것이 중요하다. 여러 값을 대입하며 실험한 결과 픽토그램의 최대사이즈의 절반정도가 적합함을 알 수 있었고 뒤에 이어질 실험에서는 25로 설정하였다. 이러한 과정으로 얻어진 이진화영상은 레이블링 과정을 거치게 된다. 레이블링에는 그래스파이어 알고리즘을 적용하였다.

그림 3-3. 이진화된 영상(좌)와 레이블링된 영상(우)

레이블링 된 영상은 수 많은 객체영역을 가지게 된다. 이렇게 수 많은 객체 영역에 대하여 HOG특징을 모두 추출하게 되면 계산상에 있어서 매우 불리하다. 따라서 앞서말한 25 * 25 ~ 45 * 45에 해당하는 범위로 객체를 한 번 필터링해준다. 픽토그램이 정면이 아닌 측면에서 나타날 경우 픽토그램이 세로로 찌그러지게 된다. 따라서 가로2 : 세로1의 비율을 객체로 두 번 필터링해준다. 이 과정까지 거치게 되면 한 프레임에 수백 개씩 검출되는 객체가 10개 이하의 객체로 줄어들게 되고 이에 대해서만 HOG특징을 추출하면 된다. HOG는 주로 보행자 인식에 사용되는 방법으로 Navneet Dalal 과 Bill Triggs[1]에 의해 제안된 방법이다. HOG는 방향별 이미지의 강도를 누적하여 계산하며 본 논문에서는 40 * 40픽셀로 후보영역들을 정규화 하여서 실험과정에서 이용하였다.

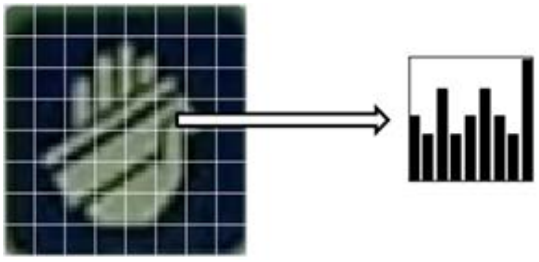

그림 3-4. HOG 특징 추출

40 * 40픽셀로 정규화 하였을 때 [그림 3-4]와 같이 후보영역을 8 * 8의 셀로 나눌 수 있다. 이 때 한 셀은 5픽셀에 해당한다. 이제 셀 내의 경사 값을 구하여 경사의 방향별로 누적하는 과정을 거친다. 경사 값을 구하기 위하여 두 픽셀간의 차이의 1차미분을 근사화한 $(-1\ 0\ 1)$과 $(-1\ 0\ 1)^T$ 필터를 사용하고 이로부터 간단하게 값을 얻을 수 있다. 이는 아래의 식(1)과 같다.

$$\begin{cases} f_x(x,y) = I(x+1,y) - I(x-1,y) \\ f_y(x,y) = I(x,y+1) - I(x,y-1) \end{cases} \quad 식(1)$$

식(1)에서 얻은 f_x와 f_y를 이용하여 경사의 크기와 경사의 방향을 계산할 수 있으며 그 식은 아래의 식(2)와 (3)에 나타난다.

$$magnitude(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad 식(2)$$

$$Orientation(x,y) = \tan^{-1}(f_y(x,y)/f_x(x,y)) \quad 식(3)$$

구현 시에는 atan2함수를 통하여 쉽게 계산할 수 있다. 얻어진 방향 값을 빈으로 놓고 크기 값을 누적하여 그림4의 오른쪽 그림처럼 셀 단위로 히스토그램을 작성할 수 있다. 그 후, 이러한 셀 단위의 히스토그램을 블록단위로 정규화하는 과정을 거치게 된다.

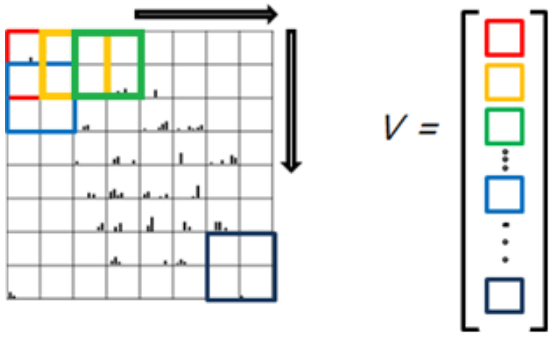

그림 3-5. 정규화 과정

[그림 3-5]는 블록단위로 특징벡터를 생성하는 모습을 보여준다. 좌상단에서 시작하여 x축 방향으로 이동한 후 y축방향으로 이동하면서 벡터열을 만들게된다. 실험 과정에서는 한 개의 블록을 가로 3개 세로 3개의 셀로 구성하였다. 블록 정규화 과정을 거치게 되면 부분적인 조명의 차이와 약간의 이동등 잡음의 영향을 줄일 수 있게 된다. 정규화 방법으로는 L2-norm을 사용하였다. 공식은 식(4)에 표현되며 ε값으로는 1.0을 주었다.

$$L2-norm = \frac{V}{\sqrt{\| V \|^2 + \varepsilon}} \quad 식(4)$$

블록 정규화의 슈도코드는 다음과 같다.

```
norm initialize;
for block dimension
norm += vec[i] * vec[i];
for block dimension
vec[i] /= sqrt(norm + 1.0);
```

이 과정을 거치게 되면 40 * 40픽셀의 후보영역에서 총 2916차원의 특징벡터를 얻을 수 있다. 정리해 보면 한 개의 블록의 차원은 3 * 3 * 9로 81차원이며 전체 이미지 특징의 벡터차원은 81 * 6 * 6으로 2916차원의 특징벡터가 생성된다. 이는 식(5)에 나타난다.

$$\begin{aligned} 블록차원 &= 블록가로셀 \times 블록세로셀 \times 방향수 \\ 전체차원 &= 블록차원 \times 가로블록수 \times 세로블록수 \end{aligned} \quad 식(5)$$

이렇게 추출한 특징벡터의 학습과 인식에는 SVM을 사용한다. SVM은 신경망, 은닉 마코프 모델등과 같이 가장 많이 사용되는 분류기로 Vapnik이 제안한 방법이다. 신경망과 같은 방법들은 오류율을 최소화하기 위한 목적을 가지고 있지만 SVM의 경우는 오류율의 최소화가 목적이 아닌 여백의 최소화에 초점을 맞추고 있다. 각각의 서포트 벡터간의 마진이 최소화 되었을 때 새로운 특징들을 인식해도 정확하게 분류를 할 가능성이 높다. 즉 일반화 능력이 뛰어나며 분류기의 품질도 뛰어나다고 볼 수 있다.[3] 본 논문에서는 이하 SVM의 quadratic optimization 문제등에 대해서는 설명하지 않는다. SVM을 통하여 학습을 진행하고 만들어진 분류기를 사용하여 인식을 하게된다.

Ⅳ. 실험결과

실험에 사용된 PC는 Intel E6600, 4G RAM, Intel SSD로 구성되어 있으며 Visual Studio 2010의 MFC와 OpenCV 1.1 pre를 사용하여 프로그램을만들었다. 학습과 인식에 사용하는 SVM은 오픈소스인 T. joachims[2]의 SVM-Light를 사용하였다. 동영상 본론에 말한 장비와 해상도로 대전광역시 지하철 역사 여러 곳에서 촬영하였고 총45개, 20분분량이다. 이 중 27개를 훈련 집합으로, 나머지 18개를 테스트집합으로 설정하였다. 분류기로 분류하게 될 픽토그램의 종류는 아래의 [표 4-1]과 같다. 또한 적응적 이진화가 적용된 영상에서 HOG특징을 추출하였다.

표 4-1. 픽토그램의 종류

픽토그램	형태			
매표소		-	-	-
휠체어		-	-	-
화장실				
엘리베이터			-	-
에스컬레이터		-	-	-
전화기		-	-	-

테스트집합에서 추출된 배경영역의 개수는 24,817개이며 픽토그램 후보영역의 개수는 1,621개이며 픽토그램별 개수는 [표 4-2]와 같다.

표 4-2. 테스트 집합의 전경영역 개수

	매표소	휠체어	화장실	엘리베이터	에스컬레이터	전화기
테스트집합	276	128	371	621	125	100

테스트 집합의 인식결과는 [표 4-3]과 같다. 각 픽토그램의 첫 글자만 따서 표시하였으며 가로방향이 분류기를 뜻하며 세로방향은 픽토그램 후보영역을 뜻한다.

표 4-3. 테스트 집합 인식결과(단위: %)

	화	엘	에	전	매	휠
화	97.02	22.49	0	0	0	0
엘	32.79	99.52	0	0	0	0
에	0	0	97.56	0	0	0
전	40.82	30.61	0	61.22	0	0
매	9.49	29.56	0	0	91.24	0
휠	0	57.75	0	0	0	97.52
배	99.49	97.95	99.99	100	99.98	100

예를 들어 엘리베이터 분류기로 엘리베이터 화장실 픽토그램을 분류한 결과 화장실의 22.49%를 엘리베이터로 잘못 분류한 것이다. 그리고 화장실 분류기로 전화기 픽토그램을 분류한 결과 40.82%의 전화기 픽토그램이 화장실이라고 잘못 분류된 것이다. 각 분류기를 이용하여 자신의 후보영역을 분류하였을 때는 표의 회색 영역처럼 높은 성능을 보여주었다. 단, 전화기 분류기는 61.22%로 낮은 성능을 보여주는데 이는 학습량이 충분치 않았던 것으로 학습량이 충분하면 다른 영역과 마찬가지로 90%이상의 인식률을 보여주었다. 화장실 분류기와 엘리베이터 분류기에서 나타나는 오 분류는 타 픽토그램을 배경영역으로 학습시켜서 해결할 수 있었다. [표 4-4]처럼 엘리베이터 분류기를 다시 학습한 결과 화장실 픽토그램은 22.49에서 0% 그리고 전화기 픽토그램은 30.61%에서 2.04%등으로 높은 분류율을 보여주었다.

표 4-4. 타 픽토그램을 배경으로 학습한 인식결과(단위: %)

화장실	엘리베이터	에스컬레이터	전화기	매표소	휠체어
0	98.22	0	2.04	5.84	0

V. 결론

본 논문에서는 조명에 대한 영향을 제거하기 위하여 적응적 이진화를 적용한 이진 영상에서 객체의 범위를 설정하고 각 객체별로 HOG특징을 추출하였다. 또한 그 특징들을 SVM을 사용하여 분류하는 방법에 대해 설명하였다. 또한 HOG와 SVM을 사용하여 픽토그램 인식하는 방법이 좋은 결과를 보여준다는 것도 확인할 수 있었다. 테스트 집합에 대해서 좋은 모습을 보이면서 일반화의 측면에서도 분류가 잘 된다는 것을 확인할 수 있었다. 이러한 픽토그램 인식에 대한 실험결과는 픽토그램 뿐만 아니라 특정한 객체의 인식에도 사용될 수 있을 것이며 시각장애인의 보행도우미 같은 응용프로그램으로도 사용될 수 있을 것으로 생각된다.

참고문헌

[1] Dalal and Bill Triggs, “Histograms of oriented gradients for human detection”, CVPR 2005.

[2] T. Joachims, Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

[3] 오일석, “pattern recognition”, p137-139, 2008.