

이미지 시각단어를 이용한 배경포함 이미지의 자동분류

조성우, 이성재, 조수선
한국교통대학교 컴퓨터정보공학과
e-mail:f2@live.co.kr

Visual word-based Classification of Images Including Background Objects

Sungwoo Cho, Seongjae Lee, Soosun Cho
Dept. of Computer Science & Information Engineering
Korea National University of Transportation

요 약

이미지의 시각단어를 이용한 이미지의 자동분류 및 태깅에 관련된 연구가 다양하게 진행되고 있지만, 기존의 연구는 특징점 추출과 이미지 비교를 위하여 비슷한 구도의 객체에만 적용하거나 배경을 제거한 객체를 대상으로 하는 등 선별된 이미지를 주로 사용하고 있다.

본 논문에서는 사용자가 특징점의 비교를 의도하지 않고 배경을 포함하여 촬영한 이미지를 대상으로 하여 이미지 시각단어를 이용한 자동 분류 및 태깅의 정확도를 향상시키는 방법을 소개하고자 한다.

1. 서론

과거 이미지 인식기술은 OpenCV[1]를 이용해 색상비교나 객체의 외곽선을 추출하는 방식이 주로 이용되었다. 최근에는 이미지의 회전이나 크기 변환에 영향을 덜 받는 특징점을 이용한 방법이 주로 이용되고 있다[2].

그러나 특징점 기반의 이미지 인식 및 분류에서도 객체가 인식되기 용이한 방법으로 이미지의 배경을 설정하거나, 분류할 객체에 배경이 들어가지 않도록 정제된 이미지를 이용하는 연구가 대부분이었다. 이는 배경과 같은 필요하지 않은 영역에서 특징점이 추출되는 것을 방지하고 해당 이미지의 일치가능성을 높이는 결과를 보였다. 하지만 최근 SNS의 확산으로 상품과 같은 정형화된 이미지뿐만 아니라 배경을 포함한 일반 사진에서도 사용자가 원하는 대상을 찾을 수 있는 방법이 요구되고 있다.

본 연구에서는 이러한 문제를 해결하기 위해 플리커와 같은 대용량 이미지 공유사이트의 이미지들을 자동으로 분류하여 트레이닝 이미지 집합으로 구축하고 사용자가 디지털카메라나 스마트 기기로 촬영한 이미지의 시각단어를 이용하여 자동 분류 및

태깅을 지원하는 기술을 개발하고자 한다.

2. 관련연구

트레이닝 이미지의 자동구축은 이미지의 태깅 정보 등을 이용하여 이미지와 실제 의미의 연관성 분석에 관한 연구결과를 이용할 수 있다. 관련 연구로 사용자가 이미지에 작성한 태그들의 우선순위를 위키피디아와 워드넷을 이용하여 조정하는 연구가 있었다[3,4]. 이 연구는 이미지 검색의 성능을 개선하기 위해 사용자가 이미지에 작성한 태그들 중 검색어 또는 다른 태그들과의 의미 연관성이 높은 주요 태그들을 선별하고, 이를 이용하여 태깅된 웹 이미지의 검색순위를 재조정하였다.

테스트 이미지의 자동분류는 이미지 시각단어를 이용한 연구 결과를 이용할 수 있다. 관련 연구로 SURF(Speeded Up Robust Feature), SIFT(Scale Invariant Feature Transform)등의 특징점 추출 알고리즘을 상품 이미지 자동 태깅 등에 적용한 연구가 있다[5].

SIFT와 SURF의 성능에 있어서 특징점을 이용한 타 연구[6]에서 SIFT 방식의 정확도가 SURF보

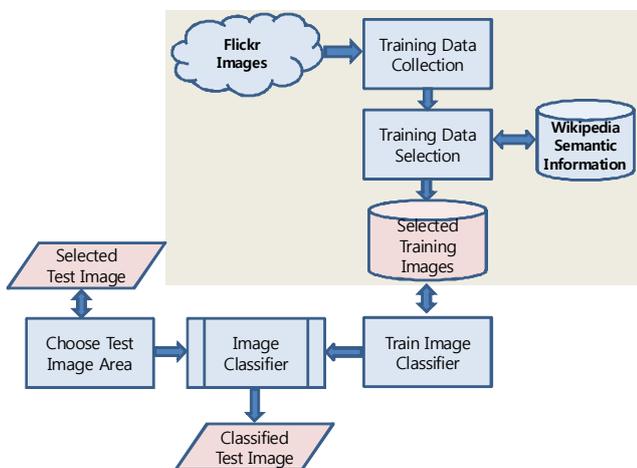
다 높지만 근소한 차이이며, 그에 비해 속도가 SURF 방식과 비교하면 수배 이상 느리다는 단점이 밝혀졌기 때문에 SIFT 가 아닌 SURF 방식을 이용하기로 결정하였다. 이들을 이용한 이미지 분류 연구는 'bag of visual words' 모델[7]을 이용하였다. 이는 기존의 텍스트 문서에서 이용되던 'bag of word' 모델을 그대로 영상처리에 적용한 것으로 이미지의 특징점을 단어로 취급하고 이미지를 문서로 취급하여 분류하는 방법이다.

3. 구현 및 개발환경

본 연구의 전체적인 구현 알고리즘은 그림 1과 같다. 전체적인 흐름은 트레이닝 이미지를 수집/선별한 후, 이를 이용하여 이미지 분류기를 훈련시키고, 훈련된 분류기를 통해 사용자가 선택한 테스트 이미지를 분류하는 과정이다. 이 때, 이미지 데이터는 모두 'bag of visual words' 모델을 적용한 시각 단어 벡터로 표현된다.

그림 1의 알고리즘 중, 플리커 이미지를 대상으로 트레이닝 이미지를 수집하고, 위키피디아 의미정보를 이용한 이미지 태그들의 우선순위를 찾아내어 이를 기반으로 트레이닝 데이터를 선택하는 부분은 기존 연구[3, 4]를 활용하였다. 이 부분은 그림 1의 위상단에 표시되어 있다.

본 논문에서 시도하는 분류 정확도 향상을 위한 방법은 테스트 이미지 내에서, 분류하고자 하는 객체를 포함하는 영역을 사용자가 직접 선택하는 방법이다. 수집된 트레이닝 데이터로 훈련된 이미지 분류기는 선택된 테스트 이미지를 자동 분류한다. 이 부분은 그림 1의 좌하단에 표시되어 있다.



(그림 1) 시스템 구현 알고리즘

본 연구는 표 1의 환경을 이용하여 1,500개의 트레이닝 이미지를 플리커에서 수집하였고, 이미지 태그들의 순위 재조정을 전처리 과정으로 처리하여 상위 500개의 트레이닝 이미지를 카테고리별로 선택하였다.

또한 SURF 알고리즘을 이용하여 이미지의 특징점을 추출하였고, Accord 프레임워크[8]를 이용하여 UI 및 분석 모듈을 개발해 이미지의 시각단어를 기반으로 테스트 이미지를 분류하였다.

<표 1> 개발에 사용된 사양

CPU	intel i5 3.3Ghz
RAM	8GB
OS	WINDOWS 7 64bit
개발도구	VISUAL STUDIO 2010
개발언어	PHP, C#, JAVA

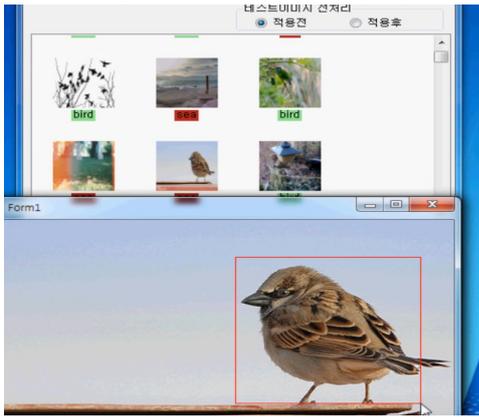
4. 실험 및 결과



(그림 2) 사용자가 촬영한 이미지

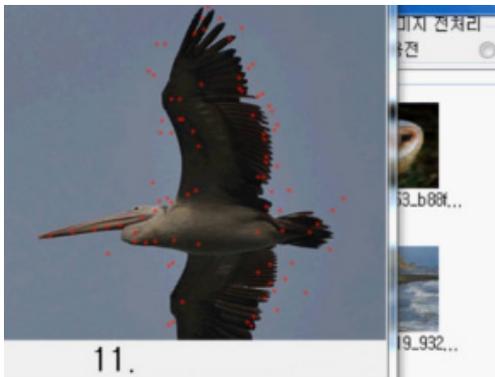
그림 2는 테스트에 사용된 일부 이미지이다. 인식을 원하는 자동차(붉은박스 영역) 외에 었드린 사람과 차량진입 방지기둥, 이륜차량 등이 같이 등장하는 것을 알 수 있으며, 이 경우 자동차를 제외한 기타 영역에서도 특징점이 동시에 추출되어 시각단어 비교 시에 정확도가 낮아질 가능성이 높다.

그림 3은 선택 지점 두 곳을 터치하면 선택한 영역에서만 시각단어를 추출하는 UI를 보여준다. 이와 같은 테스트 이미지의 영역 선택을 통한 전처리는 불필요한 특징점이 추출되지 않아 정확도를 향상시키고 동시에 계산을 줄여 처리시간을 단축시키고, 비교 분석을 위해 서버에 전송할 이미지의 용량을 줄여주는 이점을 갖기 때문에 모바일 환경에서 특히 유용할 것이다.



(그림 3) 터치를 활용하는 UI

이미지 자동 분류에는 K-means 알고리즘[9]이 이용되었다. K-means는 데이터를 k개로 클러스터링 하는 단순하고 빠르다는 특징을 가진 알고리즘이다. 받아들인 데이터(점)에 k의 개수만큼 임의로 중심점을 지정하고, 해당 중심점들에 가까운 점들을 규합하여 클러스터를 만들고 다시 해당 점들의 평균(무계중심)을 구한다. 이를 반복하여 나누어지는 클러스터에 더 이상 변동이 없으면 종료한다. 본 연구에서는 SURF 특징점을 이용하여 각각의 이미지에서 추출한 시각단어 벡터를 데이터로 이용하였다.



(그림 4) 추출된 SURF 특징점(붉은 점)

테스트 이미지의 자동분류 실험에는 새(bird), 자동차(car), 바다(sea) 세 개의 카테고리로서 약 500~600장씩 총 1675장의 트레이닝 이미지가 사용되었다. 테스트 이미지로는 각 카테고리의 30개씩의 이미지를 합쳐 총 90개의 이미지를 대상으로 분류작업을 하였다. 실험은 트레이닝 이미지를 토대로 학습한 결과를 기반으로 테스트 이미지들이 제대로 자신의 카테고리로 분류되었는가를 판정하였다. 먼저 가공되지 않은 그대로의 테스트 이미지로 실험하고 테스트 이미지를 그림 3과 같이 잘라내어 다시 분류하고 기존의 결과와 비교하였다.

<표 2> 분류 실험 결과

카테고리	영역 지정 전		영역 지정 후	
	적합 이미지 수	정확도	적합 이미지 수	정확도
bird	16	53.3%	14	46.7%
car	23	76.7%	20	56.7%
sea	19	63.3%	28	93%
합계	58	65.6%	62	68.9%

실험 결과, 일차 1,500장의 트레이닝 이미지를 사용한 실험에서는 영역 지정 전보다 지정 후의 정확도가 떨어지는 모습을 보였다. 이를 트레이닝 데이터의 부족에서 기인한 문제로 판단하고 약 150여장의 트레이닝 이미지를 추가하여 총 1,675장의 이미지를 학습, 분류에 이용한 결과, 표 2와 같이 영역 지정 후의 평균 정확도가 향상되는 결과가 나타났다. 그러나 카테고리별 정확도에선 극단적인 정확도 향상을 보인 sea를 제외한 두 카테고리 모두 지정 전보다 정확도가 낮아져 영역의 지정이 전반적으로 분류 정확도를 향상시킨다고 보기는 어렵다.

5. 결론 및 배운점

테스트 이미지의 전처리를 통해 이미지 분류 및 태깅의 정확도가 전반적으로 향상되는 결과를 얻었으나 기대했던 정도에는 미치지 못하였다.

그 원인을 짐작해보면, 실험에 사용된 카테고리가 총 3개로 같은 테스트 이미지 내에 다른 카테고리의 객체가 동시에 등장하는 경우가 거의 없다는 점, 그리고 원본 이미지의 해상도가 낮은 상태이기 때문에 해당 이미지에서 다시 객체를 잘라낼 경우 과도하게 해상도가 작아진다는 점이 분류에 악영향을 미치는 것으로 추정된다. 따라서 다음 연구에서는 카테고리 수와 이미지의 해상도, 그리고 트레이닝 이미지의 수를 늘려서 분류의 정확도를 향상시키고자 한다.

본 연구는 방송통신위원회의 방송통신인프라원천기술개발사업의 연구결과로 수행되었음 (KCA-2012-12-911-04-005)

참고문헌

- [1] openCV 홈페이지 <http://opencv.org/>
- [2] 권기원, 이해연, 오득환, “크기 및 회전 불변 특징점을 이용한 파노라마 영상 합성 알고리즘” 한국정보처리학회 논문지, 제17-B권 제 5호, 2010년 10

월

[3] 이성재, 조수선 “위키피디아 의미정보를 이용한 태깅된 웹 이미지 검색” 한국정보처리학회 추계학술 발표대회 논문지, 제18권 제 2호, 2011년, 11월

[4] 권대현, 홍준혁, 조수선 “워드넷 의미정보로 선별된 우선 태그와 이를 이용한 웹 이미지의 검색” 한국멀티미디어학회지, 제12권, 제7호, 2009년, 7월

[5] 하정우, 김병희, 이바도, 장병탁 “잡지기사 관련 상품 연계추천 서비스를 위한 하이퍼네트워크 기반의 상품이미지 자동 태깅 기법” 정보과학회 논문지, 제16권 제10호, 2010년. 10월

[6] 복윤수, 황영배, 권인소 “영상 매칭 및 자세 추정을 이용한 무인 차량의 위치 추정” 한국군사과학기술학회 종합학술대회 논문집 2007년. 8월

[7] “Recognizing and Learning Object Categories”
<http://people.csail.mit.edu/torr/alba/shortCourseRLOC/index.html>

[8] Accord.NET, <http://code.google.com/p/accord/>

[9] J. B. MacQueen (1967): “Some Methods for classification and Analysis of Multivariate Observations” Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297