

SIFT 알고리즘을 이용한 플리커 이미지 자동분류

장현웅*, 조수선**

**한국교통대학교 컴퓨터정보공학과

e-mail:jhwsorg@gmail.com

Flickr Image Classification using SIFT Algorithm

Hyun-Woong Jang*, Soo-Sun Cho**

Dept. of Computer Science & Information Engineering

Korea National University of Transportation

요 약

플리커와 같은 대용량 영상저장 및 공유 사이트가 인기를 끌면서 이미지 정보의 양은 점점 늘어나고 있고 사용자들은 정확한 이미지 정보 검색을 요구하고 있다. 태그기반의 이미지 검색에서 정확도를 높이기 위하여 태그들의 의미적 연관성을 이용하는 등 다양한 연구가 진행되고 있다. 본 논문에서는 특징점 추출에 기반하여 이미지를 분류하는데 뛰어난 성능을 가진 SIFT알고리즘을 사용하여 플리커 이미지를 분류하는 방법을 제안한다. 위키피디아 의미 연관성을 이용해 태그 정보로 1차 분류된 데이터베이스에 SIFT알고리즘을 사용해본 결과 기존의 SURF를 사용한 연구보다 높은 정확성을 보이는 것을 확인하였다. 따라서 이 방법을 통하여 다양한 이미지를 더욱 정확하게 분류할 수 있을 것으로 기대한다.

1. 서론

이미지 자동 분류에는 색상, 모양, 질감 등 내용정보를 이용하거나 외곽선을 추출하는 방식 등이 사용되어 왔다. 그러나 많은 양의 이미지 데이터가 웹 공간에 저장됨에 따라 사용자들은 점점 더 직관적으로 정확한 정보를 얻기를 바라게 되었다. 그로인해 이미지의 분류기술도 발전하게 되었고, 최근에는 이미지의 특징을 추출하고 이를 이용하여 색상, 위치, 크기, 회전 등의 변화에 강한 이미지 분류방법[1]이 주로 이용되고 있다.

기존에 SURF를 사용하여 이미지 시각단어를 만들어 자동 분류 및 태깅을 지원하고자 하는 연구가 있었다[2]. 본 논문에서는 이미지 매칭과 변화에 대해 강한 특징을 가지고 있는 SIFT(Scale Invariant Feature Transform)를 이용하여 보다 더 정확하게 분류하려고 한다. 이미지의 특징점을 추출하여 분류하는 SIFT는 많은 연구와 응용을 통해 그 효과가 입증 되었다[1].

또한 본 논문에서는 잘 정제된 이미지 데이터베이스를 분류 대상으로 하지 않고, 인터넷 상의 대용량 영상정보 저장 및 공유 사이트인 플리커(Flickr)의 검색 이미지를 분류대상으로 한다. 이 때, 분류 알고리즘인 SIFT를 훈련하기 위하여 위키피디아 기반의 의미 연관성 검색방법을 사용하여 플리커 검색 결과에서 정확도를 더 높인 결과

이미지들을 훈련 데이터로 사용하였다.

본 논문의 구성은 다음과 같다. 2절에서는 관련된 연구에 대해 소개하고 3절에서는 본 논문에서 제안한 구현방법들에 대해 설명하며 4절에서 실험 분석 및 평가를 다루며 5절에서는 결론을 맺는다.

2. 관련연구

2.1 태그 또는 내용기반의 이미지 분류

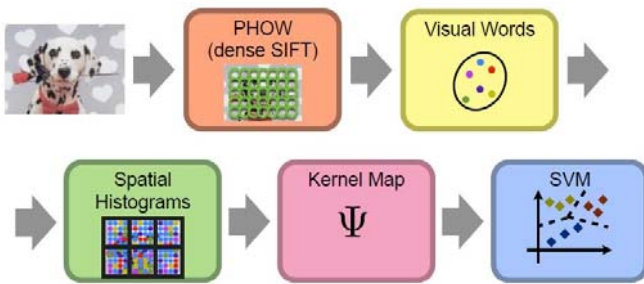
하나의 이미지에 수많은 태그를 붙일 수 있는 태깅은 일반 사용자들에 의해 만들어지는데, 주관적이면서도 수십 개의 태그를 붙이기 때문에 적당한 이미지를 찾는 데 한계가 있다. 이를 극복하기 위해 태그와 검색어의 의미 연관성을 이용한 연구가 있었다[3, 4]. 이 연구들은 태그의 단순 매칭 방법에 비해 의미 연관성을 이용한 검색 방법이 정확도에서 훨씬 효과가 있음을 보여주고 있다.

한편, 내용 기반의 이미지 분류에서 최근 많은 관심을 받고 있는 특징점 기반의 이미지 분류는 이미지 데이터에서 이미지의 내용 구성 요소들을 분리하고, 분리된 구성 요소들의 특징점을 추출하여 군집화 하여 비교 또는 검색을 실시하는 방법이다. 이 방법은 사람들이 직접 태그를 만들어야 하고 만들어진 태그들이 주관적이라는 단점을

극복할 수 있다. 특징점 기반의 이미지 분류는 이미지에서 상대적으로 풍부한 정보를 포함하고 있는 돌출영역(salient area)인 이미지 키포인트를 벡터 양자화(vector quantization)한 특징 값으로 뽑아내어 이용하는 것이다. 이것은 텍스트기반 검색에서 문서를 하나의 'bag of word'로 보는 것과 유사하게 이미지를 하나의 'bag of visual word'로 보고 미리 구성된 시각 어휘집(visual-word vocabulary)을 이용하여 이미지를 분류하고자 하는 것이다.

2.2 SIFT알고리즘을 이용한 이미지 분류

특징점 기반의 이미지 분류에서 키포인트를 벡터 양자화 하는 방법으로는 DoG(Difference of Gaussian) 디텍터와 SIFT(Scale Invariant Feature Transform) 알고리즘이 주로 채택되고 있는데 SIFT를 이미지 키포인트의 기술자(descriptor)로 사용함으로써 크기와 회전에 강건한 특징을 추출할 수 있는 것으로 알려져 있다. 이미지 시각단어를 구성하기 위한 특징점 추출 알고리즘으로 SURF(Speeded Up Robust Feature), SIFT등을 이용한 이미지 카테고리 분류 연구가 있다[5]. SIFT와 SURF의 이미지 분류 성능에 있어 SURF가 SIFT방식에 비해 속도가 빠르지만 정확도가 SIFT방식이 높기 때문에 본 연구에서는 SIFT방식을 이용하기로 하였다. SIFT를 이용한 특징점 매칭 과정은 회전, 크기, 변화 등에 강한 이미지 매칭을 수행할 수 있기 때문에 이미지 분류에 유리하다.



(그림 1) 이미지 검출과정[1]

그림 1은 본 논문의 시스템 구현에서 사용한 오픈 라이브러리인 VLFeat의 SIFT 기반 이미지 분류과정을 보여준다. 우선 특징점을 추출하기 위해 가우시안 커널을 이용하여 스케일을 줄이고, SIFT 기법을 사용해 시각 언어를 추출하며 추출된 시각 언어를 히스토그램에 집합으로 구성한다. 이렇게 선택된 특징점들의 집합을 단어로 취급하고 이미지를 문서로 취급하여 이미지를 분류한다.

3. 구현 및 실험

본 연구에서는 VLFeat 라이브러리[1]를 이용하여 분류

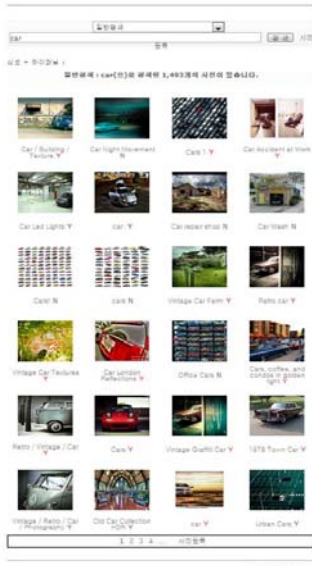
엔진을 구현하였다. VLFeat는 컴퓨터 시각처리 분야에서 유명한 오픈소스 라이브러리로 HOG, SIFT, MSER, k-means, hierarchical k-means, agglomerative information bottleneck, SLIC superpixels, and quick shift 등의 알고리즘을 맷랩에서 쉽게 구현할 수 있도록 지원하고 있다.

먼저 보편적인 이미지 분류를 테스트할 때 사용되는 데이터 집합인 Caltech101 dataset[6]을 이용하여 SIFT의 성능을 평가했다. Caltech101 dataset은 California Institute of Technology에서 2003년 9월에 Computer Vision 기술 개발에 기여하기 위해 만들어진 정제된 이미지 집합이다. Caltech101에는 총 9146개의 이미지와 101개로 구별되는 카테고리 포함되어 있다. 표1에서 보이는 것처럼 이미 정제된 Caltech101 이미지는 94.00%의 높은 정확도를 보이고 있다.

<표 1> Caltech101 dataset을 이용한 30개 훈련, 10개 테스트 이미지의 분류 (정확도: 94.00%)

카테고리	적합이미지 수	정확도
accordion	10	100%
airplanes	10	100%
anchor	7	70%
ant	10	100%
barrel	10	100%
합계	47	94.00%

본 논문에서는 실제 인터넷 상에서 검색된 이미지를 분류 대상으로 하기 위해서 플리커 검색 이미지들을 훈련 및 테스트 이미지로 사용한다. 이 때, 좀 더 정확하게 검색된 이미지들을 훈련 데이터로 사용하기 위해 위키피디아 기반의 의미정보를 이용하여 이미지 태그들의 우선순위를 찾아 이미지를 수집했던 기존 연구[4]를 활용하였다. 기존 연구[4]에서 얻어진 이미지 데이터를 사용해 총 3,003개의 이미지를 수집하였고, bird(새), car(자동차) sea(바다) 세 개의 카테고리별로 선행 연구[2]에서 사용한 이미지 보다 200장씩 많은 각 700개씩 총 2,100장을 임의로 뽑아 훈련 이미지로 사용하였다. 테스트 이미지로는 각 카테고리에서 임의로 30개씩을 뽑아 총 90개의 이미지를 대상으로 분류하는 실험을 하였다.



(그림 2) 위키피디아 기반의 의미 연관성을 이용한 검색으로 추출한 1,001개의 car 이미지

특징점 추출에는 SIFT검출기를 사용하였고 이미지 분류를 위해 K-means 군집화 알고리즘[7]을 사용하였다. 추출한 데이터에 K개의 중심축을 계산하고, 해당 중심축에 가까운 점들을 군집화를 반복한다. 추출한 시각 언어 히스토그램[8]을 만들고 이미지 데이터를 분류하는데 사용한다.

4. 분석 및 평가

<표 2> SIFT알고리즘을 사용한 실험 결과

카테고리	적합이미지 수	정확도
bird	21	70%
car	23	76.66%
sea	18	60.00%
합계	62	68.89%

표 2는 각 카테고리의 1,001개의 이미지 중 700개를 훈련 데이터로 사용하고 30개씩 임의의 이미지를 합쳐 90개의 이미지를 대상으로 분류 실험을 한 결과이다. bird 카테고리는 30개중 21개를 분류하여 70%의 정확도를 나타냈다. car는 30개중 23개를 분류하여 76.66%의 정확도를 나타냈고 sea는 30개중 18개를 분류하여 60.00%의 정확도를 나타냈다. 그 결과 평균 68.89%의 정확도를 나타냈다.

실험결과 68.89%의 정확도는 Caltech101을 이용해 이미지를 분류했던 94.00%의 정확도보다 낮아졌다. 그 이유는 Caltech101은 미리 분류기에 맞게 정제되었던 이미지 데이터 집합이고 위의 실험에서는 인터넷상에서 실제로 검색된 이미지를 사용하여 분류기를 훈련했기 때문에 Caltech101로 실험했던 정확도보다 떨어질 수밖에 없다. 하지만 선행된 연구의 실험 결과인 65.6% 보다 약 3.4% 높은 정확도를 기록했다. 이는 SURF 알고리즘을 이용하

는 것보다 SIFT알고리즘을 이용하는 것이 더 좋은 정확도를 나타낸다는 것을 알았다. 3.4%의 차이는 작게 보일 수 있지만 찾으려는 이미지의 개수가 많아질수록 차이는 커질 것이라 예상된다.

5. 결론

본 연구에서는 더 정확도가 높은 이미지들로 분류기를 훈련시키기 위해 플리커로 바로 검색된 이미지 대신 위키 피디아 기반의 의미 연관성을 이용하여 검색된 이미지들을 훈련 데이터로 사용하였다. 또, 이미지의 특징을 추출하기 위해 SIFT 방법을 이용함으로써 이미지 분류의 정확도를 높일 수 있다는 것을 입증하였다. 기존 연구[2]와 다르게 SIFT알고리즘을 사용하고, 카테고리 수를 늘리고, 훈련 이미지의 수를 늘려서 정확도를 더 향상시킬 수 있었다. 앞으로 의미적 연관성과 이미지 내용 정보를 이용하여 더 효과적으로 이미지를 분류할 수 있는 연구를 계속할 계획이다.

참고문헌

[1] VLFeat 홈페이지 <http://www.vlfeat.org/>
 [2] 조성우, 이성재, 조수선, “이미지 시각언어를 이용한 배경포함 이미지의 자동분류” 한국정보처리학회 추계학술발표대회, 2012년, 11월, 제주대학교
 [3] 권대현, 홍준혁, 조수선, “워드넷 의미정보로 선별된 우선 태그와 이를 이용한 웹 이미지의 검색” 한국멀티미디어학회지, 2009년,7월
 [4] 이성재, 조수선, “위키피디아 의미정보를 이용한 태깅된 웹 이미지 검색,” 한국정보처리학회 추계학술발표대회, 2011년, 11월, 연세대학교
 [5] 김승률, 유훈재, 손종인, 오창범, 손광훈 “SIFT 알고리즘의 강인한 특징점 검출을 위한 양방향 필터 기반 스케일 공간” 한국방송공학회 하계학술대회, 2012년, 7월
 [6] Caltech 101 Dataset 홈페이지 http://www.vision.caltech.edu/Image_Datasets/Caltech101/
 [7] Andrea Vedaldi and Brian Fulkerson, “VLFeat - An open and portable library of computer vision algorithms”, <http://www.ros.org/workshops/eccv/2010/Vedaldi.pdf/>
 [8] Bag of visual words model : recognizing object categories, http://www.robots.ox.ac.uk/~az/icvss08_az_bow.pdf/