

의미적 연관태그와 이미지 내용정보를 이용한 웹 이미지 분류[☆]

Web Image Classification using Semantically Related Tags and Image Content

조 수 선*
Soosun Cho

요 약

본 논문에서는 대용량 온라인 이미지 공유 사이트를 적용 도메인으로 하여 이미지 검색의 만족도를 높이고자 태그의 의미적 연관성과 이미지 자체의 내용 정보를 결합하는 이미지 분류 방법을 제안한다. 이미지 검색 및 분류 알고리즘이 플리커와 같은 대용량 이미지 공유 사이트에서 활용될 수 있으려면 실제 웹상의 태깅된 이미지를 대상으로 한 적용이 가능해야 한다. 제안된 알고리즘은 'bag of visual word' 기반의 이미지 내용으로 웹 이미지를 분류하기 위한 것으로서, 의미적 연관태그를 이용해 일차 검색된 이미지들을 훈련 데이터로 사용하여 카테고리 모델을 훈련하고, PLSA를 적용하여 평가 이미지들을 분류하는 것이다. 제안된 방법으로 플리커의 웹 이미지들을 대상으로 실험한 결과, 태그 정보를 이용한 기존의 방법에 비해 우수한 검색 정확도 및 재현율을 확인할 수 있었다.

ABSTRACT

In this paper, we propose an image classification which combines semantic relations of tags with contents of images to improve the satisfaction of image retrieval on application domains as huge image sharing sites as Flickr. To make good use of image retrieval or classification algorithms on huge image sharing sites as Flickr, they are applicable to real tagged Web images. To classify the Web images by 'bag of visual word' based image content, our algorithm includes training the category model by utilizing the preliminary retrieved images with semantically related tags as training data and classifying the test images based on PLSA. In the experimental results on the Flickr Web images, the proposed method produced the better precision and recall rates than those from the existing method using tag information.

☞ KeyWords : 웹 이미지 분류, 의미적 연관태그, 내용기반 분류, 플리커, Web image classification, semantically related tags, content based classification, Flickr, bag of visual word, PLSA

1. 서 론

최근 몇 년 사이에 온라인 사진 이미지 공유 사이트가 등장하여 이용자수가 급격히 증가함에 따라 태그와 같은 주석이 달린 이미지 데이터의 양이 폭발적으로 늘고 있으며 이와 같은 환경 변화는 이미지 검색 관련 연구자들에게 새로운 기회

와 도전이 되고 있다. 특히 야후의 플리커(Flickr), 구글의 피카사(Picasa)와 같은 온라인 사진 이미지 공유 사이트에서는 폭소노미로 불리는 협업적 태깅에 기반하여 다양한 태그를 주석으로 가진 이미지들을 공유하고 있다. 하지만 폭소노미로 저장된 이미지는 카테고리로 분류되어 있지 않고, 태그 또한 매우 개인적이며 특별한 의미를 가진 것이 대부분이기 때문에 실제 이미지의 내용과 많이 동떨어져 있는 것이 사실이다. 따라서 태그 기반의 이미지 검색 또는 분류 결과는 사용자의 요구수준과는 큰 차이를 보이고 있으며 이를 극

* 중신회원 : 충주대학교 컴퓨터정보공학과 부교수
sscho@cjnu.ac.kr

[2010/02/10 투고 - 2010/02/11 심사 - 2010/03/14 심사완료]

☆ 이 논문은 충주대학교 대학구조개혁지원사업비(교육과학기술부지원)의 지원을 받아 수행한 연구임

복하기 위해 태그 정보를 보다 심도있게 분석하여 이미지 검색에 이용하고자 하는 여러 가지 시도가 등장하고 있다[1, 2, 3, 4].

하지만 태그 정보만으로는 검색된 이미지의 내용상의 동질성을 높이는 데에 분명한 한계가 존재한다. 본 논문에서는 태깅된 방대한 웹 이미지 공유 사이트에서 검색결과에 대한 사용자 만족도의 향상을 위해 태그정보와 함께 이미지 고유의 내용정보를 이용한 검색 방법을 제안한다. 이미지 고유의 색상, 질감, 모양 등과 같은 저수준 특징들을 이용하여 이미지검색에 적용하고자 하는 내용 기반 이미지 검색(content based image retrieval)은 접근 가능한 디지털 이미지의 급속한 팽창과 더불어 지난 10여 년간 급속한 발전을 이루어왔다. 특히 웹 이미지의 검색 또는 분류를 위해서는 효과적인 객체 분류 알고리즘이 필요한데 최근 ‘bag of visual word’ 기법이 이 분야에서 효과적인 것으로 발표되고 있다[5]. 그러나 객체 분류 알고리즘의 성능을 개선하기 위한 대부분의 시도는 미리 수집되고 정제된 데이터 집합, 예를 들면 ‘Caltech 101[6]’, ‘UIUC car dataset[7]’ 등을 사용하여 그 효과를 입증하고 있다.

따라서 플리커와 같은 이미지 공유 사이트에서 활용될 수 있으려면 실제 웹상의 이미지를 대상으로 한 객체 분류 알고리즘의 개발 및 적용이 필요하며, 특히 본 논문에서 제안하는 바와 같이 웹 이미지의 태그 정보와 함께 내용 정보를 이용함으로써 더욱 효과적인 결과를 얻을 수 있다. 본 논문에서는 대용량의 웹 이미지 공유 사이트인 플리커 상의 실제 웹 이미지들을 대상으로 하여 그 태그정보와 내용정보를 동시에 이용한 이미지 분류 방법을 제안하고 실험을 통해 그 효과를 입증하고자 한다. 논문의 구성은 이어지는 제2절에서 관련 연구를 소개하고, 제3절에서 본 논문에서 제안하는 분류 알고리즘을 소개하며, 제4절에서 실험 및 평가를 통해 제안된 알고리즘의 효과를 입증한다. 마지막으로 제5절에서 결론을 맺는다.

2. 관련연구

2.1. 태그기반의 이미지 검색

단순한 태그기반의 검색은 태그들이 동일한 이미지에 달려있다 하더라도 그들이 어떻게 연관되어 있는지를 찾아낼 수 없으므로 많은 한계를 드러낸다. 예를 들어 검색어 ‘mammal’로는 ‘lion’, ‘cow’, ‘cat’과 같은 태그를 포함한 자료를 검색할 수 없다. 이와 같은 한계를 극복하고자 태그들의 동시 출현에서 서로 연관된 태그들(inter-related tags)을 찾아내어 검색 또는 분류에 활용하고자 하는 시도들이 있었다[1, 8]. 더불어 태그의 의미 정보를 활용하기 위해 시맨틱 웹을 이용하여 태그들의 의미적 연관성(semantic relation)을 찾아내어 온톨로지 매칭으로 검색 또는 분류하고자 하는 시도들도 있었다[2, 9]. 첫 번째 방법인 연관된 태그들을 이용하여 이미지를 분류하는 것은 현재 플리커에서 클러스터 보기 기능으로 제공되고 있으나 동일 클러스터로 분류된 태그들 간에 어떠한 공통 의미를 가진 것이 아니므로 서로 다른 클러스터로 분류된 이미지들 사이에 뚜렷한 차이를 보이지 않는 경우가 많다. 두 번째 방법인 온톨로지를 접목하여 태그 의미 정보를 찾아내고 이를 분류하고자 하는 시도는 태그들의 특성상 특정 온톨로지 상에 나타나는 의미적 연관성들을 찾아내기 힘들어 실험 결과 분류에 포함되는 범위가 매우 제한적인 것으로 알려졌다[2].

연구[4]에서는 의미정보를 얻기 위해 엄격한 온톨로지를 적용하는 대신 간편한 온라인 어휘사전을 적용함으로써 그 효과를 높이려는 시도를 전개하였다. 이 연구에서는 프린스턴대학에서 개발한 온라인 어휘사전인 ‘워드넷[10]’을 사용하여 검색어 및 이미지 태그들의 의미정보를 획득한 후, 이를 바탕으로 한 웹 이미지 검색 결과를 보여주는데, 플리커 검색 결과에 비해 정확도의 향상이 뚜렷한 것으로 나타났다. 워드넷 상의 동의어 및 상위어 집합을 이용하여 태그들 간의 연관성을 계산한 후, 이를 이용하여 우선 태그를 추출

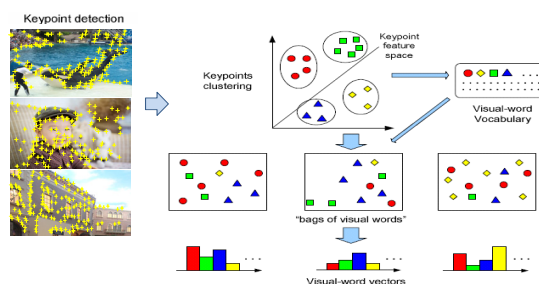
하고 검색에 적용하는 것인데, 동시 출현만으로 획득되는 연관성 정보대신 워드넷 온라인 어휘사전의 적용을 통해 의미적 연관성 정보를 획득하여 검색에 사용하는 방법이다. 하지만 워드넷 자체의 한계로 말미암아 제한적인 의미정보만 사용할 수 있을 뿐이고, 따라서 첫 한, 두 페이지에서 월등히 높았던 정확도가 검색 페이지가 증가할수록 빠르게 줄어드는 것이 단점이라 할 수 있다.

2.2. Bag of Visual Word 기반의 이미지 분류

내용기반 이미지 검색은 색상, 질감 등과 같은 저수준의 이미지 특징들을 이용하여 이미지를 표현하고 유사한 특징을 가진 이미지를 검색하는 기술이다. 칼라 히스토그램 및 가버 필터 등이 이미지 특징을 나타내는데 많이 사용되며, 지역정보를 포함하기 위해 이미지를 여러 개의 작은 영역으로 나누거나 객체와 배경을 분리하여 특징 벡터를 구한 후 이를 다시 결합하여 사용하기도 한다. 이와 같은 기존의 이미지 특징 표현 방법 외에 최근 이미지의 키포인트(keypoint)들로 그 특징을 표현하고자 하는 시도들이 주목을 받고 있다. 특히 이미지 내용정보를 이용하는 웹 이미지 분류를 위해서는 효과적인 객체 분류 알고리즘이 필요한데 이를 위해 이미지 키포인트를 이용한 방법들이 최근 객체 또는 장면 분류에서 그 효과가 입증되고 있다.[11, 12]. 'Bag of visual word' 기법은 이미지의 키포인트를 뽑아내어 이를 벡터 양자화(vector quantization)한 특징 값으로 이미지를 표현하는 것이다. 이것은 텍스트 검색에서 문서를 하나의 'bag of word'로 보고 처리하는 것을 이미지에 적용한 방법으로 최근 다양한 연구 결과들이 발표되고 있다[9, 13].

그림1은 키포인트 특징을 이용하여 이미지들이 벡터 양자화된 'bag of visual word'로 표현되는 과정을 보여준다. 키포인트는 이미지에서 상대적으로 풍부한 지역정보를 포함하고 있는 돌출영역(salient area)으로서 DoG(Difference of Gaussian)와 같은 알고리즘을 사용하여 자동으로 추출될 수

있다. 그림 1에서 3개의 이미지 상에 표현된 작은 +기호들이 이미지 키포인트인데 이들은 주로 객체의 경계선 근처에서 많이 나타난다. 자동 추출된 키포인트를 표현하는 방법으로는 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[14].



(그림 1) 키포인트 특징을 이용한 이미지의 벡터 양자화 과정

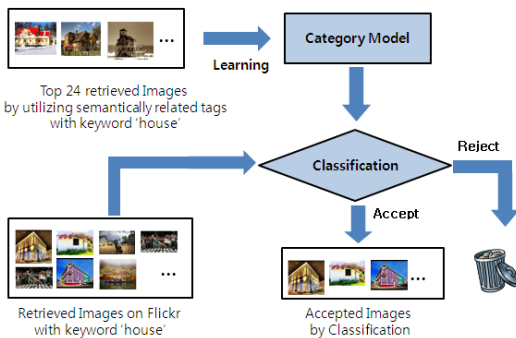
분류하고자하는 이미지는 이 키포인트들의 집합으로 표현될 수 있지만 분류를 위해서는 고정된 차원의 이미지 특징 벡터가 필요하므로 키포인트들의 집합 형태로는 활용이 어렵다. 이 문제를 해결하기 위해 벡터 양자화 기법을 사용하는데 그림 1에서 보이는 것처럼 키포인트 클러스터링을 통한 시각적 단어(visual word)의 추출과 이를 이용한 이미지의 벡터 표현이 그 방법이다. 벡터 양자화 기법은 먼저 이미지로부터 키포인트를 추출하여 이를 K-means 알고리즘 등을 통해 클러스터링하고 각 키포인트를 그것이 속한 클러스터의 인덱스로 표현한다. 이 때, 각각의 클러스터가 하나의 시각적 단어가 되어 이들의 집합으로 시각 어휘집(visual word vocabulary)이 구성되며, 시각적 단어의 개수로 어휘집의 크기가 결정된다. 고정된 크기의 어휘집에 속한 시각적 단어들로 각각의 이미지를 표현할 수 있게 되어 비로소 하나의 이미지를 고정된 차원의 벡터 양자화된 값으로 취급할 수 있게 된다. 그림 1은 이와 같은 방

법으로 각각의 이미지가 시각적 단어의 누적 개수에 따라 히스토그램과 같은 형태의 양자화된 값으로 표현될 수 있음을 나타낸다.

웹 이미지 분류에서 검색어로 주어지는 카테고리(예를 들면, 'house')에 속하는 이미지들을 'bag of visual word' 기법으로 찾아내기 위해서는 순도 높은 이미지 데이터들로부터 시각 어휘집을 구성하고 카테고리 모델을 훈련시켜야 한다. 본 논문에서는 이를 위해 의미적 연관태그를 이용한 일차 검색 결과를 카테고리 모델의 훈련 데이터로 이용하는 방법을 제안한다. 즉, 태깅된 이미지에서 얻을 수 있는 텍스트 정보(태그)와 이미지 내용 정보(이미지 키포인트)를 모두 활용한 분류 방법인 것이다.

3. 의미적 연관태그와 이미지 내용을 이용한 분류 방법

3.1 전체 구조



(그림 2) 검색어 'house'를 이용한 분류방법 적용 예

연구[4]에서는 검색어와 웹 이미지 태그들의 의미적 연관성을 이용하는 방법으로 플리커 이미지들을 검색한 결과, 첫 페이지(24개 이미지)의 검색 정확도를 약 79%로 보고하고 있다. 이 비율은 일반적인 플리커 검색의 정확도가 40%에 못 미치는 것과 비교해볼 때 매우 높은 값을 알 수 있다. 이것은 검색어와 의미적 연관성이 높은 태그들

(이후 '의미적 연관태그')을 많이 달고 있는 이미지들을 검색 순위의 앞쪽에 배치함으로써 얻을 수 있는 효과이다.

본 논문에서 해결하고자 하는 문제는 플리커와 같은 대용량의 온라인 이미지 공유 사이트에서 자동화된 방법으로 사용자의 만족을 높일 수 있는 검색 방법을 찾는 것이다. 이를 위해 기왕에 태깅된 이미지들이 가지고 있는 태그 정보를 이용하여 순도 높은 참조 이미지들을 뽑아내고 이와 유사한 이미지들을 내용기반으로 분류하는 알고리즘을 제안한다. 즉, 태그의 의미적 연관성을 이용하면 정확도가 높은 이미지들을 일차로 분류할 수 있으므로 이들을 내용기반 분류를 위한 훈련 데이터로 사용하는 것이다. 이 과정에서 사용자는 검색어를 입력하는 것 외에 부수적인 상호작용 없이 원하는 검색 결과의 정확도를 높일 수 있다. 따라서 대부분의 내용기반 이미지 검색에서 수차례의 사용자 피드백을 통해 참조 이미지를 선택하게 하는데 비해, 검색어만 입력하게 하는 one-pass 방식으로서의 차별성을 가진다. 그림 2는 본 논문에서 제안하는 이미지 분류 방법을 검색어 'house'를 사용한 예를 통해 간략하게 표현한 것이다. 그림의 상단에는 검색어 'house'를 이용한 의미적 연관태그 기반의 검색으로 랭킹 24위까지의 이미지들을 검색한 후, 이를 이용하여 'house' 카테고리 모델을 훈련시키는 과정을 보여준다. 또한, 훈련된 카테고리 모델을 이용하여 플리커에서 1차 검색된 이미지들을 다시 분류함으로써 분류기에서 선택된 이미지들로만 결과 페이지를 구성할 수 있다. 그 결과, 40%에 못 미치는 플리커의 검색 정확도를 향상시키는 것은 물론, 검색 순위가 낮아질수록 정확도가 급격히 떨어지는 의미적 연관태그기반의 검색 방법 자체의 한계도 극복할 수 있게 된다.

3.2 의미적 연관태그와 가중치 계산

기존의 연구에서는 연관된 태그(inter-related tag)를 찾아서 그에 따른 분류를 클러스터로 나타

내고 검색에 활용할 수 있도록 하고 있다. 연관 태그를 찾아내는 기존의 방법은 하나의 오브젝트에 동시에 출현하는 태그들끼리의 포함 관계를 이용하여 클러스터링하거나[15], 태그들을 동시에 출현하는 횟수를 가중치로 가지는 에지로 연결함으로써 태그 공간을 그래프로 구성한 후, 그 결과를 클러스터링하는 것[8] 등이 있다. 그러나 이와 같은 연관 태그 구성만으로는 태그들 간의 의미적 차이는 여전히 구분하기가 어렵다. 본 논문에서는 온라인 어휘 사전인 워드넷을 이용하여 이 연관 태그들을 의미적 연관태그(semanticly related tags)로 확장한 연구[4]의 방법을 채용한다. 이 방법은 앞에서 언급된 바와 같이 검색 순위가 낮아질수록 정확도가 급격히 떨어지는 단점이 있으나 본 논문에서 제안하는 방법은 상위 랭킹에 해당하는 첫 번째 검색 페이지의 이미지들만 사용하므로 문제가 안 된다.

의미적 연관태그의 구성을 위해서는 워드넷의 의미 정보를 이용하는데, 프린스턴 대학에서 개발된 워드넷은 어휘들 간의 관계로 동의어(synonym), 반의어(antonym), 하위어(hyponym), 상위어(hypernym) 등을 표현함으로써 어휘의 의미에 대한 카테고리 분류가 잘 정의되어 있으며, 어휘의 계층구조와 연관관계가 잘 표현되어져 있다[16]. 특히 워드넷은 명사, 동사, 형용사, 부사들이 동의어 집합인 synset (set of synonyms)으로 나뉘어져 있고, synset과 synset 사이의 의미적인 관계를 표현한 링크로 이루어져 있다. 의미적 연관태그 기반의 검색 방법에서 사용하는 연관성 점수 계산 방법은 워드넷 상에서 단어들의 포함관계를 고려한다. 이 방법은 어떤 이미지 D_i 에 달린 하나의 태그 T_j 가 검색어 K 의 워드넷 synset이나 K 의 상위어의 synset에 포함될 때마다 일정한 가중치를 주고, 이미지 D_i 의 모든 태그들에 대해 이 가중치를 합하여 검색어 K 에 대한 이미지 D_i 의 의미 가중치 W_i 를 계산하는 것이다. 따라서 검색어와 의미적으로 연관성이 높은 태그들을 많이 가진 이미지가일수록 그 값이 커지게 된다. 검색 결과로는 검색어 K 에 대한 각 이미지의 의미 가중치 W_i 가 높

은 순으로 이미지들을 배치한다. 의미 가중치 W_i 의 계산식은 다음과 같다.

$$W_i = \sum_{j=1}^{N_i} w_j$$

where N_i is the number of tags on D_i ,

$$w_j = \begin{cases} \alpha & \text{if each tag } T_j \in \text{SigSet}(K) \\ 0 & \text{else} \end{cases},$$

$$\text{SigSet}(K) = \text{synset of } K \cup \text{synset of } (\text{hypernym of } K)$$
(1)

3.3 PLSA 기반의 카테고리 모델과 분류기

'Bag of visual word'로 표현된 이미지의 분류를 위해서는 카테고리 모델의 학습과 이진 분류기(binary classifier)에 의한 카테고리의 결정이 필요하다. 이를 위해 나이브 베이지안 분류기와 같은 생성 기법(generative method)과 SVM과 같은 분류/식별 기법(discriminative method)이 사용되고 있는데, 본 논문에서는 나이브 베이지안 분류기보다 일반적으로 성능이 더 나은 것으로 알려진 생성 기법인 PLSA(Probabilistic Latent Semantic Analysis)[17] 모델을 사용한다.

연구[18]에서는 이미지의 비감독 객체 분류에 PLSA 모델을 이용하였다. 문서를 'bag of word'로 취급하는 텍스트 분류 및 검색에서는 각 문서에 대해 단어들의 출현 횟수를 공기 행렬(co-occurrence matrix)로 표현하여 사용한다. 단어 수가 W , 문서의 수가 D 일 때, $W \times D$ 공기행렬은 각 요소값으로 문서 d 에 나타난 단어 w 의 출현 횟수, $n(w, d)$ 를 가진다. 이미지 분류에서 문서 d 는 이미지에, 단어 w 는 시각적 단어에 대응된다. PLSA 모델은 주어진 단어(시각적 단어)의 출현 횟수로 표시되는 문서(이미지)가 있을 때, 그 문서(이미지)가 가지고 있는 잠재된 주제(latent topic), z 을 찾아내는 것이다. 이 때, 문서의 주제 z 는 이미지의 카테고리에 해당한다.

문서 d 가 주어졌을 때 단어 w 의 확률 $P(w|d)$ 는 관찰된 데이터로부터 구할 수 있는데, PLSA에서

는 이것을 다음 식과 같이 주어진 주제에 대한 단어의 조건 확률, $P(w|z)$ 와 주어진 문서에 대한 주제의 조건 확률, $P(z|d)$ 에 대하여 전개함으로써 $W \times D$ 행렬을 $W \times Z$ 행렬과 $Z \times D$ 행렬로 나눈다.

$$P(w|d) = \sum_{z=1}^Z P(w|z)P(z|d) \quad (2)$$

훈련 데이터를 이용한 모델의 비감독 학습은 EM(Expectation Maximization) 알고리즘을 통하여 이루어진다. E-step에서는 각 주제에 대한 사후 확률 $P(z|w, d)$ 를 계산하고, M-step에서는 이를 이용하여 모델의 매개변수, $P(w|z)$ 와 $P(z|d)$ 를 갱신한다. 이 때, 다음 식과 같은 우도함수(likelihood function)를 최대화하는 지역 최적해로 모델 매개변수를 찾는다.

$$L = \prod_{d=1}^D \prod_{w=1}^W P(w|d)^{n(w,d)} \quad (3)$$

학습된 분류기에서 평가 데이터의 카테고리를 결정하기 위해, 즉 주어진 이미지 d 에 대하여 z 를 결정하기 위해서는 아래 식과 같이 조건 확률 $P(z|d)$ 의 값이 최대가 되게 하는 z 를 선택하면 된다.

$$z^* = \arg_z \max P(z|d) \quad (4)$$

이미지를 분류하기 위해 $P(w|z)$ 를 고정시키고 EM 과정을 반복함으로써 주어진 이미지에 대한 주제의 조건 확률 $P(z|d)$ 를 추정할 수 있다. 따라서 이 확률 값을 최대로 만들어주는 z 를 선택함으로써 카테고리의 결정, 즉 이미지의 분류가 이루어진다.

4. 실험 및 평가

4.1 실험 데이터 및 구현 환경

실험을 위해 먼저 검색어 'house'를 이용하여 플리커에서 검색한 결과 이미지와 그에 달린 태

그들을 데이터베이스에 저장하였다. 또한 실시간으로 변하는 플리커에서, 이후 실험에 사용한 것과 똑같은 데이터를 사용한 검색 결과를 그대로 보존하기 위해 플리커의 모든 결과 페이지는 화면을 그대로 캡처하여 저장하였다. 한 페이지당 24개씩, 모두 50개 페이지의 1,200개 이미지를 데이터베이스에 저장한 후, 제3.2절에서 설명한 의미적 연관태그 기반의 검색과 제안된 방법의 검색을 각각 수행하였다.

시스템 구현을 위해 Linux환경에서 MySql Server 5.0와 Apache Server 2.2를 사용하였으며 PHP 5.2.4를 활용하여 웹프로그래밍으로 구현하였다. 한편, PLSA 구현을 위해서는 MatLab 과 함께 MIT 컴퓨터과학과에서 제공하는 Matlab 스크립트와 Linux 바이너리로 구성된 라이브러리[19]를 다운받아 사용하였다.

4.2 시각 어휘집의 구성과 카테고리 분류

키폰트 검출기로는 MIT 라이브러리에서 제공하는 Canny Edge 탐지기를 그대로 사용하였고, 추출된 키폰트들은 SIFT 기술자를 통해 128차원의 특징 벡터로 표현하였다. 이 특징 벡터들로부터 K-means 클러스터링을 통해 시각적 단어를 생성하여 이것으로 시각 어휘집을 구성하였다. 시각 어휘집을 구성하기 위한 훈련 데이터로는 3.2절에서 언급한 바와 같이 의미적 연관태그 기반의 검색 결과 랭킹 순위가 높은 이미지들을 사용한다. 실험에서는 검색어 'house'를 사용한 검색 결과 첫 페이지의 24개 이미지를 사용하였으며, 이들의 정확도는 79%를 넘는다. 시각 어휘집의 크기에 있어서는 연구[5]에서 비용 대비 효과적인 것으로 밝힌 1,000을 채택하였다.

카테고리 분류를 위한 평가 이미지들은 앞 절에서 설명한 대로 1,200개의 수집된 이미지 중에서 훈련에 사용된 24개 이미지를 제외한 나머지가 된다. 앞에서 구성된 시각 어휘집을 사용하여 각 평가 이미지들을 벡터 양자화하여 'bag of visual word'로 표현한다. 실험에서 사용된 이진 분류기

에서는 **PLSA** 기법을 적용하여 **EM** 알고리즘을 통해 학습된 매개변수로 평가 이미지의 카테고리를 결정한다. **EM** 알고리즘의 수행에서 지역 최적해를 구하기 위해 매개변수를 임의로 초기화하여 20번 반복한 후 수렴되는 값으로 하였다.

요약하면, 카테고리 모델은 의미적 연관태그 기법으로 검색된 이미지 데이터를 사용하여 학습되고, 학습된 카테고리 모델에 기반하여 플리커로 검색된 이미지에 대하여 이진 분류('house' vs. 'background')가 수행된다. 이 때, 바탕(background) 카테고리 모델은 기존 연구[20]에서 사용된 바탕 이미지 데이터로 학습된다.

4.3 성능 평가

실험을 위한 프로토타입 시스템 구현으로 제안된 분류 방법을 평가한 결과는 다음과 같다

표 1은 플리커 검색, 의미적 연관태그를 이용한 검색, 그리고 본 논문에서 제안된 분류 방법으로 검색한 결과를 비교하고 있다. 3가지 방법으로 검색된 결과를 첫 페이지부터 20번째 페이지까지 조사하여 각 페이지에 포함되어 있는 24개의 검색 이미지들에 대해 검색어 'house'에 적합한지 여부를 조사한 것이다. 판정 방법은 연구[4]에서와 같이 검색어 'house'를 집으로 보이는 건축물의 외형으로 판단하고 이에 해당되면 '적합' 해당되지 않으면 '부적합'으로 판정하였다. 또한 'house'는 그 의미단어집합에서 알 수 있듯이 'home'과 동의어로도 사용될 수 있으므로 집으로 보이는 건축물의 내부, 즉 거실, 침실, 주방과 같은 공간을 나타내면 역시 '적합'으로 판정하였다.

표 1에서 첫 번째 페이지에 대한 제안된 방법의 적용 결과는 실제로 내용기반 분류 방법이 적

(표 1) 20개 페이지의 검색결과 비교

페이지	구분	플리커			의미적 연관태그 이용			제안된 분류방법 이용		
		적합	부적합	비율(%)	적합	부적합	비율(%)	적합	부적합	비율(%)
1		7	17	29.17	19	5	79.17	19	5	79.17
2		9	15	37.50	17	7	70.83	18	6	75.00
3		10	14	41.67	16	8	66.67	16	8	66.67
4		11	13	45.83	10	14	41.67	18	6	75.00
5		5	19	20.83	15	9	62.50	20	4	83.33
6		8	16	33.33	6	18	25.00	19	5	79.17
7		7	17	29.17	12	12	50.00	19	5	79.17
8		12	12	50.00	9	15	37.50	21	3	87.50
9		6	18	25.00	7	17	29.17	19	5	79.17
10		10	14	41.67	10	14	41.67	16	8	66.67
11		8	16	33.33	15	9	62.50	17	7	70.83
12		10	14	41.67	13	11	54.17	17	7	70.83
13		12	12	50.00	11	13	45.83	17	7	70.83
14		8	16	33.33	7	17	29.17	16	8	66.67
15		9	15	37.50	12	12	50.00	19	5	79.17
16		12	12	50.00	14	10	58.33	18	6	75.00
17		12	12	50.00	14	10	58.33	20	4	83.33
18		6	18	25.00	9	15	37.50	21	3	87.50
19		12	12	50.00	11	13	45.83	22	2	91.67
20		9	15	37.50	7	17	29.17	19	5	79.17
합계		183	297	38.13	234	246	48.75	371	109	77.29

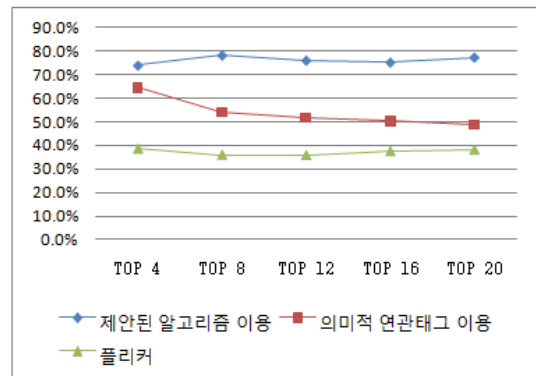
용된 것이 아니라 의미적 연관태그 기반의 검색 결과를 그대로 사용한 것이다. 즉, 제안된 분류 방법으로 구현되는 웹 이미지 검색 환경에서, 첫 페이지로는 의미적 연관태그 기반 검색 결과를, 이어지는 나머지 페이지들은 내용기반 검색 결과를 보게 된다. 표 1의 마지막 행은 20개 페이지의 정확도(precision) 평균을 나타내는데, 플리커는 38.13%, 의미적 연관태그 이용방법은 48.75%를 보인 반면에 제안된 알고리즘 이용방법은 77.29%로 월등히 높아진 것을 알 수 있다.

표 2는 각 검색 결과의 정확도 및 재현율을 비교해보기 위해 20개 검색 결과 페이지를 첫 4개 페이지, 8개 페이지, 12개 페이지 등으로 묶어서 그 결과를 조사한 것이다. 그림 3, 그림 4는 각각 검색 방법에 따른 정확도 및 재현율을 그래프로 나타낸 것이다. 그림 3에서 보이는 것처럼 의미적 연관태그 기반의 검색은 정확도에서 플리커에 비해 우수한 것으로 나타났지만 검색 페이지 수가 증가할수록 정확도가 플리커와 비슷한 수준으로 떨어지는 것을 알 수 있다. 반면에 제안된 분류 방법을 이용한 검색 결과는 정확도가 나머지 두 가지 방법보다 우수하면서도 70%대에서 일정하게 유지되고 있다. 그 결과, 재현율도 검색 페이지 수가 많아질수록 나머지 두 가지 방법보다 빠르게 증가하고 있음을 알 수 있다.

5. 결론

본 논문에서는 플리커와 같은 대용량의 웹 이미지 공유 사이트에서 검색 결과에 대한 사용자

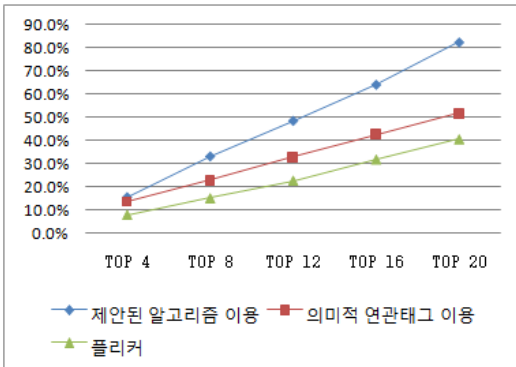
만족도를 높이기 위해 태그 정보와 이미지 내용 정보를 동시에 활용하는 웹 이미지 분류 방법을 제안하였다. 웹 2.0 환경에서 태깅된 이미지가 급속히 증가하고 있지만 이미지 태그만을 이용한 검색에서는 이미지 내용상의 동질성을 보장해주지 못하므로 많은 한계를 드러내고 있으며, 이미지 내용기반 검색은 대부분 제한된 이미지 데이터베이스를 테스트 대상으로 함으로써 실제 웹상의 태깅된 방대한 이미지들을 대상으로 적용하기에는 제약이 따른다. 본 논문에서는 이와 같은 문제를 해결하기 위하여 의미적 연관태그와 ‘bag of visual word’기반의 이미지 내용 분류 기법을 함께 사용함으로써 사용자 피드백 없는 one-pass 방식으로 웹 이미지 검색 결과의 정확도를 높일 수 있는 알고리즘을 소개하였다. 또한 실험 및 평가를 통해 그 효과를 입증함에 따라 향후 실질적인 검색 시스템의 구현으로 활용될 가능성을 높였다.



(그림 3) 정확도 비교 그래프

(표 2) 검색결과에의 정확도 및 재현율 비교

구분 페이지	플리커					의미적 연관태그 이용					제안된 분류방법 이용				
	적합	부적합	합계	정확도	재현율	적합	부적합	합계	정확도	재현율	적합	부적합	합계	정확도	재현율
TOP 4	37	59	96	0.39	0.08	62	34	96	0.65	0.14	71	25	96	0.74	0.16
TOP 8	69	123	192	0.36	0.15	104	88	192	0.54	0.23	150	42	192	0.78	0.33
TOP 12	103	185	288	0.36	0.23	149	139	288	0.52	0.33	219	69	288	0.76	0.49
TOP 16	144	240	384	0.38	0.32	193	191	384	0.50	0.43	289	95	384	0.75	0.64
TOP 20	183	297	480	0.38	0.41	234	246	480	0.49	0.52	371	109	480	0.77	0.82



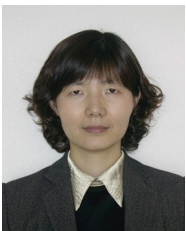
(그림 4) 재현율 비교 그래프

참 고 문 헌

- [1] A. Hotho, R. Jaschke, C. Schmitz, and G. Stumme, "Information Retrieval in Folksonomies: Search and Ranking," In Proc. of ESWC'06, 2006.
- [2] S. Angeletou, et al., "Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report," In Proc. of Workshop: Bridging the Gap between Semantic Web and Web 2.0 at European Semantic Web Conference, 2007.
- [3] 이시화, 이만형, 황대훈, "Web2.0 환경에서의 효율적인 이미지 검색을 위한 태그 클러스터링 시스템의 설계 및 구현," 멀티미디어학회 논문지, 제11권, 제8호, 2008.
- [4] 권대현, 홍준혁, 조수선, "워드넷 의미정보로 선별된 우선 태그와 이를 이용한 웹 이미지의 검색," 멀티미디어학회 논문지, 제12권, 제7호, 2009.
- [5] J. Yang, C-W. Ngo, A. Hauptmann, and Y-G. Jiang, "Evaluating Bag-of-Visual-Words Representations in Scene Classification," In Proc. of the ACM Multimedia Information Retrieval Workshop (MIR 2007) at ACM Multimedia 2007, 2007.
- [6] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories," In Proc. of Workshop on Generative-Model Based Vision, 2004.
- [7] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 11, 2004.
- [8] G. Begelman, P. Keller, and F. Smadja, "Automated Tag Clustering: Improving search and exploration in the tag space," In Proc. of the Collaborative Web Tagging Workshop at WWW'06, 2006.
- [9] C-V. Damme, M. Hepp, and K. Siorpaes, "Folksontology: An integrated approach for turning folksonomies into ontologies," In Proc. of Workshop: Bridging the Gap between Semantic Web and Web 2.0 at European Semantic Web Conference, 2007.
- [10] WordNet 3.0, "WordNet, a lexical database for the English language," <http://wordnet.princeton.edu/>, 2006.
- [11] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman, "Discovering object categories in image collections," In Proc. of International Conference on Computer Vision, 2005.
- [12] L. Fei-Fei, and P. Perona, "A bayesian hierarchical model for learning natural scene categories," In Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005.
- [13] Y-G. Jiang, C-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," In Proc. of ACM Int' Conf. on Image and Video Retrieval, 2007.

- [14] D-G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, Vol. 20, No.2, 2004.
- [15] P. Schmitz, "Inducing Ontology from Flickr Tags," In Proc. of the Collaborative Web Tagging Workshop at WWW'06, 2006.
- [16] G-A. Miller, "WordNet: An On-line Lexical Database," International Journal of Lexicography, Vol. 3, No. 4, 1990.
- [17] T. Hofmann, "Probabilistic latent semantic analysis," In Proc. of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99), 1999.
- [18] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. "Discovering object categories in image collections," Technical Report A. I. Memo 2005-005, Massachusetts Institute of Technology, 2005.
- [19] <http://people.csail.mit.edu/fergus/iccv2005/bagwords.html>
- [20] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," In Proc. Computer Vision and Pattern Recognition, 2003.

● 저 자 소 개 ●



조 수 선

1987년 서울대학교 계산통계학과 졸업(학사)
1989년 서울대학교 대학원 계산통계학과 졸업(석사)
2004년 충남대학교 대학원 컴퓨터학과 졸업(박사)
1994년~2004년 한국전자통신연구원 컴퓨터·소프트웨어연구소 선임연구원
2006년~2007년 미시간대학교(앤아버) 방문연구원
2004년~현재 충주대학교 컴퓨터정보공학과 부교수
관심분야 : 데이터마이닝, 기계학습, 센서 데이터 처리
E-mail : sscho@cjnu.ac.kr