



가중치된 이웃 태그 투표를 사용한 이미지 태깅

Image Tagging using Weighted Neighbor Tag Voting

저자 (Authors)	조선영, 차재성, 변혜란 Sunyoung Cho, Jaeseong Cha, Hyeran Byun
출처 (Source)	정보과학회논문지 : 소프트웨어 및 응용 38(2) , 2011.2, 69-76 (8 pages) Journal of KISS : Software and Applications 38(2) , 2011.2, 69-76 (8 pages)
발행처 (Publisher)	한국정보과학회 KOREA INFORMATION SCIENCE SOCIETY
URL	http://www.dbpia.co.kr/Article/NODE01602079
APA Style	조선영, 차재성, 변혜란 (2011). 가중치된 이웃 태그 투표를 사용한 이미지 태깅. 정보과학회논문지 : 소프트웨어 및 응용, 38(2), 69-76.
이용정보 (Accessed)	대구가톨릭대학교 203.250.33.*** 2018/01/12 22:21 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

가중치된 이웃 태그 투표를 사용한 이미지 태깅

(Image Tagging using Weighted Neighbor Tag Voting)

조 선 영 ^{*} 차 재 성 ^{*} 변 혜 란 ^{**}
(Sunyoung Cho) (Jaeseong Cha) (Hyeran Byun)

요 약 최근 Flickr, Facebook과 같은 사진 공유 기반의 소셜 미디어 공유 사이트의 발전으로 인해 이미지의 양이 폭발적으로 증가하면서, 효율적인 이미지 검색을 위한 연구가 활발히 진행되고 있다. 이와 함께 이미지에 자동으로 관련된 태그를 어노테이션하는 이미지 태깅 연구가 진행되고 있으며, 이미지 뿐 아니라 태그와 같은 이미지에 달린 컨텍스트 정보까지도 함께 고려함으로써 태깅 성능을 높이려는 시도를 하고 있다. 본 논문에서는 이미지에 관련된 태그 추출의 정확도를 높이기 위해 가중치된 이웃 태그 투표(voting) 기법을 이용한 이미지 태깅 방법을 제안한다. 제안하는 방법은 데이터 기반(data-driven) 방식이며, 전체 과정은 다음과 같다: 1) 관심영역 기반의 이미지 특징을 이용하여 쿼리의 이웃 이미지를 검색한다, 2) 검색된 이웃 이미지의 태그로부터 이미지 내용 유사도 기반 가중치를 결합한 투표 기법을 통해 쿼리와 관련된 태그를 추출한다, 3) 추출된 태그의 순서를 관련성 순으로 결정하여 태깅한다. Flickr로부터 수집한 태그가 달린 이미지 데이터셋에 대해 실험하였고, 일반적인 이웃 투표 기법과 제안하는 방법을 비교함으로써 태깅 성능을 평가한 결과 제안하는 방법이 기존의 방법보다 더 높은 태깅 정확도를 가지고 있음을 보였다.

키워드 : 이미지 태깅, 가중치된 이웃 태그 투표 기법

Abstract Recently, the amount of images are explosively increasing with the advent of social media sharing sites such as Flickr or Facebook, and consequently many works for efficient image search have actively studied. For this, image tagging works have studied to automatically annotate a tag to the image, and they improve the performance by considering the textual descriptions of the image as well as visual contents. This paper proposes a weighted neighbor tag voting method for image tagging. The proposed method is data-driven approach and the overall procedure is follows: 1) Search the neighbor images of query using ROI-based image feature, 2) Extract the relevant tags by voting method with weight based on the content similarity, and finally 3) Tag by determining the tag order in order of relevance. We evaluate the performance by comparing the proposed method with voting method in the image dataset collected from Flickr. We show that the proposed method gives more high tagging accuracy than the previous method.

Key words : Image Tagging, Weighted neighbor tag voting

· 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2010-0017260)

^{*} 학생회원 : 연세대학교 컴퓨터과학과

sycho22@yonsei.ac.kr

cjs135@yonsei.ac.kr

^{**} 종신회원 : 연세대학교 컴퓨터과학과 교수

hrbyun@yonsei.ac.kr

논문접수 : 2010년 11월 18일

심사완료 : 2010년 11월 25일

Copyright©2011 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제38권 제2호(2011.2)

1. 서 론

최근 인터넷이 발달하고 디지털 카메라가 보편화되면서, Flickr, Facebook, YouTube와 같은 소셜 미디어 공유 사이트가 급격히 성장하고 있다. 그 결과, 멀티미디어 양이 폭발적으로 증가하면서 효율적인 멀티미디어 검색 기술은 중요한 이슈가 되고 있다. 이를 위해 이미지에 관련된 태그를 자동으로 달기 위한 이미지 태깅(Image Tagging)이나 태그 추천(Tag Suggestion) 기술이 활발히 연구되고 있다.

대부분의 초기 이미지 태깅 및 어노테이션 연구는 색상, 텍스트 및 형태와 같은 시각적 특징을 이용하여 이

미지의 내용을 분석한다[1-3]. 이 방법은 다양한 패턴인식 기법을 적용하여 이미지 특징과 태그 간 관계를 사상(mapping)하는 모델을 정의한다. Mori의 알고리즘[1]에서는 이미지를 여러 서브이미지로 나누고, 양자화한 각 서브이미지의 특징 벡터와 태그를 사상한다. Duygulu가 제안한 방법[2]에서는 이미지를 영역단위로 분할한 후, EM(Expectation Maximization) 알고리즘을 이용하여 영역 타입과 키워드 간 사상관계를 학습한다. Blei가 제안한 방법[3]은 CORR-LDA(Correspondence Latent Dirichlet Allocation)를 이용하여 이미지 영역과 워드 집합의 latent 변수 표현 간 관계를 찾는 모델을 정의한다. 비록 이러한 방법은 정의하는 태그의 개수가 적을 경우에는 잘 작동하지만, 데이터셋이 커지고 태그의 종류가 다양해짐에 따라 성능이 떨어지게 된다. 또한, 이러한 모델기반 방법은 저차원 이미지 특징과 고차원 이미지 시맨틱 간의 차이로 인한 시맨틱 갭 문제(Semantic gap problem)를 갖는다.

따라서 이러한 시맨틱 갭을 줄이고 이미지 태깅 성능을 향상하기 위해 제목, 태그나 설명 같은 이미지에 달린 여러 컨텍스트 정보를 활용하는 연구들이 많아지고 있다[4-7]. Yeh의 알고리즘[4]에서는 이미지와 키워드를 혼합한 검색 방법을 제안한다. 이 방법은 웹 페이지로부터 관련된 키워드를 추출하여, 쿼리 이미지와 가장 비슷한 이미지를 찾는데 이용한다. Weinberger의 알고리즘[5]에서는 태그 분포의 가중치된 KL(Kullback-Leibler) divergence에 기반하여 태그를 분석함으로써 이미지에 대한 태그를 추천한다.

특히, 최근에는 웹의 발전과 함께 이미지의 양이 폭발적으로 증가하면서 웹에서의 활용성을 고려한 데이터 기반(data-driven) 방식의 연구들이 제안되고 있다[6,7]. 데이터 기반 방식은 먼저 대용량의 이미지 데이터셋으로부터 이미지 특징을 이용하여 시각적으로 유사한 이웃 이미지 집합을 검색한다. 검색된 이웃 이미지 집합에 대해, 각 이미지의 컨텍스트 정보로부터 다양한 기법을 이용하여 쿼리 이미지와 관련된 태그만을 추출한다. 마지막으로 추출된 태그들의 관련성 순으로 태깅되는 순서를 결정한다. Wang이 제안한 방법[6]에서는 컨텍스트 정보로부터 마이닝 기법을 적용하여 관련된 용어나 구절을 추출한다. Li의 알고리즘[7]에서는 유사한 이미지 집합에 존재하는 공통적 태그들은 쿼리 이미지와 관련된 태그일 것이라는 가정을 가지고, 이웃 투표(neighbor voting) 기법을 제안함으로써 관련된 태그를 추출한다. 이러한 방법들은 컨텍스트 정보를 활용함으로써 사용자의 태깅 경향을 반영하고 효율적으로 관련된 태그를 추출한다. 본 논문에서도 이미지 내용뿐만 아니라 태그 정보도 함께 이용하며, 데이터 기반 방식에 의해 이미지의

태그를 추출한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 이미지 태깅 방법의 문제점을 지적하고 이로부터 제안하는 이미지 태깅 방법의 아이디어를 유도한다. 3장에서는 이웃 이미지 검색, 태그 추출, 태그 랭킹의 3단계로 나눠서 제안하는 이미지 태깅 방법을 서술한다. 또한 4장에서는 다른 이미지 태깅 방법과의 비교를 통해 제안하는 방법의 우수성을 입증한다. 마지막으로 5장에서는 결론 및 향후 연구 방향에 대해 기술하겠다.

2. 기존의 문제점 및 제안하는 방법 유도

데이터 기반 방식의 이미지 태깅에서 태깅 성능은 크게 두 가지 단계에 의해 결정된다고 볼 수 있다. 먼저 첫 번째 단계인 쿼리와 유사한 이미지 집합을 검색하는 단계이다. 이 단계는 이미지의 시각적 특징을 이용하여 쿼리와 유사한 이웃 이미지를 찾음으로써, 두 번째 단계에서 분석될 후보 태그들을 필터링하는 역할을 한다. 따라서 쿼리와 얼마나 유사한 이미지들이 잘 검색되었는지에 따라 후보 태그들이 달라지며 결과적으로 추출되는 태그의 정확성이 결정된다. 즉, 만약 쿼리와 관련이 없는 이미지들이 검색된 경우, 그 이미지들의 태그를 이용하여 추출한 태그 역시 관련 없는 태그가 될 수 있다. 태깅 성능을 결정하는 또 다른 단계는 이웃 이미지 집합의 태그를 이용하여 쿼리와 관련된 태그를 추출하는 단계이다. 이는 첫 번째의 이미지 검색 단계는 오랫동안 연구되어온 분야이지만 성능적인 한계가 존재하기 때문에 태그 정보를 함께 사용함으로써 태깅 성능을 향상시키기 위함이다. 그러나 일반적으로 이미지에 달린 태그는 각 개인 사용자에게 의해 달린 태그이기 때문에 주관적이고 개인화된 잡음 태그를 포함하고 있다. 따라서 대부분의 이미지 태깅 연구들은 다양한 방식을 통해 잡음 태그를 포함한 태그들로부터 쿼리와 관련된 객관적인 태그만을 추출하는 방법을 제안함으로써 태깅 성능을 높이려는 시도를 하고 있다.

본 논문에서는 태깅 성능을 결정하는 두 단계의 성능을 높이는 방법을 제안한다. 먼저, 첫 번째 단계의 이미지 검색 성능을 높이기 위해, 이미지 전체가 아닌 관심 영역의 이미지 특징을 이용하여 이웃 이미지 집합을 검색한다. 대부분의 사람들은 일반적으로 피사체를 중심으로 사진을 찍고 태그를 다는 경향이 있다. 이러한 사용자 태깅 경향을 반영하여 이미지의 피사체에 해당하는 관심 영역을 추출하고 이들의 이미지 특징을 기반으로 유사한 이웃 이미지 집합을 검색한다. 두 번째의 태그 추출 단계는 가중치된 이웃 투표 기법을 적용함으로써 태깅 성능을 높인다. Li의 알고리즘[7]에서 제안된 이웃 투표 기법은 이웃 이미지의 태그들 중에서 빈도수가 큰

태그는 객관적일 것이라는 직관을 가지고 간단하면서도 효율적으로 관련된 태그를 추출한다. 그러나 이 방법은 기본적으로 여러 개의 이웃 이미지를 검색하기 때문에, 가장 유사한 이미지 뿐만 아니라 유사도가 적은 이미지까지 이웃 이미지로 검색된다. 따라서 이웃 이미지 집합에 여러개의 같은 카테고리의 유사도가 적은 이웃 이미지가 포함되는 경우 관련없는 태그가 추출되는 경우가 발생한다. 그림 1은 이러한 상황의 예를 보여준다. 그림에서 볼 수 있듯이, 쿼리와 관련된 book 태그를 추출하는 반면, 관련없는 building 태그도 함께 추출하는 것을 볼 수 있다. 따라서 본 논문에서는 이미지 유사성을 고려한 이웃 투표 기법을 통해 관련없는 이웃 이미지로부터 관련없는 태그가 추출되는 것을 막는다. 즉, 쿼리 이미지와의 유사도에 따라 가중치를 부여하여 유사도가 적은 이웃 이미지의 태그에 대한 관련성을 낮춰줌으로써 태깅 성능을 향상시킨다.

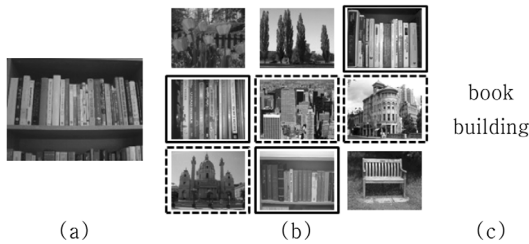


그림 1 Li의 알고리즘[7]의 문제점을 보여주는 예시. (a) 쿼리 이미지, (b) 이웃 이미지 집합, (c) 태그 추출 결과. 실선박스의 이미지들은 쿼리와 관련된 이미지이며, 나머지 이미지들은 관련없는 이미지이다. 이웃 투표 기법은 실선박스의 이미지 태그로부터 book 태그를, 점선박스의 이미지 태그로부터 building 태그를 추출한다.

3. 제안하는 이미지 태깅 알고리즘

제안하는 이미지 태깅 알고리즘은 데이터 기반 방식으로 다음의 3가지 과정으로 구성된다. 먼저, 주어진 쿼리 이미지로부터 피사체 영역으로 추정되는 관심영역을 추출한다. 관심영역에 대해 이미지 특징을 추출하고 이를 통한 내용 기반 검색을 통해 쿼리 이미지와 유사한 이웃 이미지 집합을 찾는다. 다음은 이웃 이미지 집합에 포함된 각 이미지의 태그들에 대해, 가중치된 이웃 태그 투표 기법을 적용하여 쿼리 이미지와 관련된 태그를 추출한다. 마지막으로 앞에서 추출된 태그들에 대해 관련성 순으로 태깅되는 순서를 결정한다. 그림 2는 제안하는 알고리즘의 흐름도를 보여준다.

3.1 ROI 기반 이웃 이미지 검색

주어진 쿼리 이미지에 대해, 이미지의 피사체라고 여겨지는 관심영역을 추출하기 위해 GBVS(Graph-Based Visual Saliency) 모델을 적용한다[8]. 이 모델은 그래프 이론을 이용하여 추출된 특징 맵(Feature map)에서의 활성화 지도(Activation map)를 만들고, 마코브(Markovian) 알고리즘을 이용하여 활성화 지도를 정규화함으로써 관심영역을 강조한다. 이 방법은 간단하면서도 효율적으로 인간의 관심과 주의를 끄는 이미지 영역을 추출한다.

추출된 이미지의 관심영역에 대해, 색 현저도(Color saliency)와 지역 형태(Local shape)의 두 가지 이미지 특징을 추출한다. 먼저, 색 현저도 특징을 추출하기 위해 색 현저도 부스팅(Color saliency boosting) 알고리즘을 이용한다[9]. 이 알고리즘은 해리스 검출기(Harris detector)에 의해 검출된 현저한 점(Salient point)에 대해 색 현저도를 추출한다. 영역 전체가 아닌 현저한 점 위치에서의 색 정보만을 특징으로 추출하기 때문에 간단하면서도 효율적으로 특징을 추출한다. 이미지의 지역

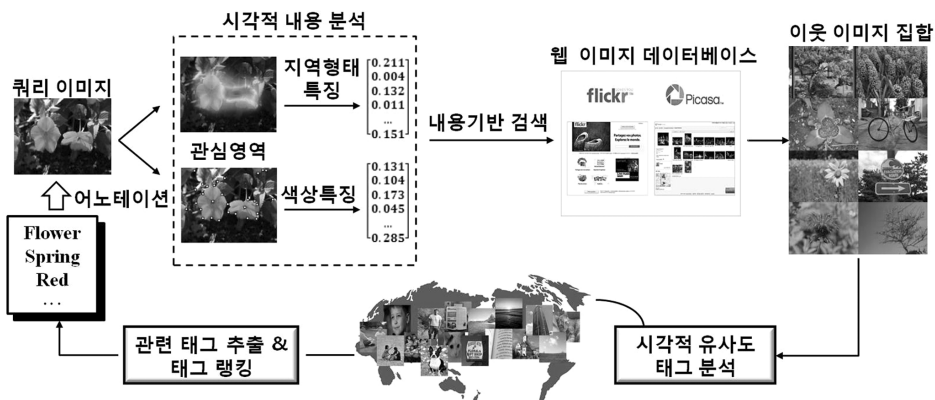


그림 2 제안하는 이미지 태깅 방법의 흐름도

형태 특징을 추출하기 위해서는 PHOG(Pyramid Histogram of Oriented Gradients)를 이용한다[10]. PHOG는 지역 형태와 그에 대한 공간적 레이아웃을 이용하여 이미지를 표현한다. 지역 형태는 에지 방향의 히스토그램에 의해 표현되며, 작은 회전에 강인하다는 장점을 갖는다.

위에서 기술한 두 가지 이미지 특징은 이미지 전체 영역이 아닌 관심영역에 대해서만 추출함으로써 피사체가 아닌 다른 영역에 의해 유사하지 않은 이웃 이미지가 검색되는 경우를 줄여준다. 전체 데이터베이스에 있는 이미지로부터 유사한 이웃 이미지 집합을 찾기 위해서는 k-최근 이웃(k-nearest neighbor) 알고리즘을 적용한다. 이때 이미지 특징 간 시각적 유사도를 판단하기 위해서는 유클리디안 거리(Euclidean distance)를 이용한다.

3.2 시각적 유사도 가중치 기반 태그 분석

위 과정에서 검색된 이웃 이미지 집합에 대해, 이웃 이미지에 달린 태그들간의 가중치된 이웃 투표 기법을 이용하여 쿼리 이미지에 관련된 태그를 추출한다. 2절에서 언급하였듯이, Li의 알고리즘[7]에서 제안한 이웃 투표 기법은 간단하면서도 효율적인 알고리즘을 통해 이전 방법보다 더 좋은 성능을 가지지만, 이미지 검색의 성능적 한계로 인해 획득되는 관련없는 이웃 이미지로부터 관련없는 태그를 추출할 수 있다. 본 절에서는 이러한 문제를 해결하기 위한 방법으로 제안한 이미지 내용의 유사도 정도를 고려하는 가중치된 이웃 태그 투표 기법을 기술한다. 제안하는 방법은 먼저 이웃 이미지들의 태그중에서 일정 임계치 투표 이상의 값을 갖는 태그들을 찾는다. 그 다음에 이미지 유사도를 가중치로 고려하여 쿼리 이미지와 각 태그간의 관련성을 계산한다.

주어진 쿼리 이미지 I_q 에 대해, k 개의 유사한 이웃 이미지를 포함하는 집합을 $N(I_q, k)$ 라 하고, 그 집합에 포함된 각 이웃 이미지에 달린 태그들의 집합을 $W(I_q, k)$ 라 한다. $W(I_q, k)$ 에 포함된 각 태그 w 에 대해, $WT(I_q, w)$ 는 w 를 포함하는 $N(I_q, k)$ 내의 이미지 집합이다. 먼저, 각 태그를 포함하는 이미지들의 개수를 고려하여 $W(I_q, k)$ 에 있는 태그들 중에서 일정 임계치 이상의 이미지 개수를 갖는 태그들을 찾는다. 이 때 일정 임계치 값은 $k/10$ 으로 결정하였고, $\hat{W}(I_q, k)$ 를 일정 임계치 이상의 투표값을 갖는 태그들의 집합이라고 한다.

다음에 추가적으로 $\hat{W}(I_q, k)$ 에 포함된 관련없는 태그를 필터링하기 위해서 태그 관련성을 측정한다. 태그 관련성 측정을 위해 시각적 유사도를 고려한다. 즉, 시각적 유사도가 더 큰 이미지의 태그는 더 관련있는 태그일 확률이 높다고 판단한다. $\hat{W}(I_q, k)$ 의 각 태그에 대해

시각적 유사도는 앞 단계에서 계산된 이미지 특징 간 유클리디안 거리를 사용한다. 이를 가중치로 하여 투표 기법과 결합함으로써 각 태그의 관련성을 계산하는 식을 정의한다. 쿼리 이미지 I_q 와 태그 w 와의 관련성은 다음의 식 (1)을 이용하여 계산한다.

$$R(w, I_q) = 1 - \frac{1}{|WT(I_q, w)|} \sum_{I_i \in WT(I_q, w)} S(I_i, I_q) \quad (1)$$

위 식에서 $|WT(I_q, w)|$ 는 $WT(I_q, w)$ 에 포함된 이미지의 개수를 의미한다. 즉, 태그 w 가 받은 투표값으로써, 값이 클수록 태그 관련성이 커지게 된다. $S(I_i, I_q)$ 는 앞 단계의 이미지 내용 분석을 통해 획득된 이미지 간 유사도 점수를 의미한다. 태그 w 를 갖는 각 이미지와 쿼리 이미지와의 시각적 유사도들의 합을 투표값 항에 가중치로 곱함으로써, 잘못된 이미지 검색으로 인해 획득되는 시각적 유사도가 적은 이미지들의 태그들에 대한 관련성을 낮춰주는 효과를 갖는다.

식 (1)을 이용하여 $\hat{W}(I_q, k)$ 에 포함된 각 태그마다 쿼리 이미지 I_q 와의 관련성을 계산하고, 가장 큰 관련성을 갖는 n 개의 태그를 추출한다. n 을 각각 1, 3, 5로 하여 실험한 결과는 4장에서 기술하겠다.

3.3 태그 랭킹을 통한 이미지 태깅

이미지 태깅 시, 어노테이션 하는 태그의 순서를 결정하는 태그 랭킹은 효율적인 이미지 검색을 위해 중요하다. 따라서 이미지와 관련된 태그일수록 먼저 달아주는 것이 더 좋은 이미지 태깅 알고리즘이 된다. 본 논문에서는 식 (1)에서 계산되는 태그의 관련성과 그 태그를 포함하는 이미지에서의 순서를 함께 고려하여 태그 랭킹 함수를 정의한다.

앞의 과정에서 추출된 쿼리 이미지 I_q 와 관련된 태그들을 $W^*(I_q)$ 라 하자. 각 $W^*(I_q)$ 에 포함된 w^* 에 대해, 태그 랭킹 함수는 다음의 식 (2)와 같이 정의한다.

$$TR(w^*, I_q) = R(w^*, I_q) + R(w^*, I_q) \left\{ 1 - \frac{1}{|WT(w^*)|} \sum_{I_i \in WT(w^*)} \frac{o(w^*, W_{I_i})}{|W_{I_i}|} \right\} \quad (2)$$

식 (2)에서 W_{I_i} 는 이미지 I_i 에 대한 태그 리스트이며, $o(w^*, W_{I_i})$ 는 태그 리스트 W_{I_i} 에서의 w^* 의 순서를 의미한다. 이는 기본적으로 관련성이 높은 태그일수록 태그 랭킹 함수 값을 높게 만들되, 이미지의 태그 리스트에서 앞쪽에 달린 태그일수록 관련성에 더 가중치를 준다. 결과적으로 쿼리 이미지에는 식 (2)를 이용하여 계산된 태그 랭킹 함수 값이 높은 순서대로 태그를 달게 된다.

4. 실험

본 실험에서는 제안하는 이미지 태깅 알고리즘의 성



그림 3 실험에 사용된 10개 카테고리 데이터셋 예

능을 평가하기 위해 Flickr로부터 직접 수집한 태그를 포함하는 이미지 데이터셋을 사용하였다. 데이터셋 수집 시, 다음의 10개의 키워드를 쿼리로 검색하였다: animal, book, building, chair, flower, people, road, sign, tree, vehicle. 쿼리 검색 결과로부터 각 카테고리당 30개씩 무작위로 총 300개의 이미지를 수집하였다(그림 3 참고). 데이터 셋의 각 이미지 당 태그의 수는 1부터 21까지 다양하였고, 평균적으로 5.8개의 태그를 가지고 있었다. 실험은 태깅 성능에 영향을 미치는 이미지 검색의 정확도와 실제 관련 태그 추출의 성능을 평가하였다. 제안하는 방법은 학습 단계가 필요하지 않으므로, 전체 데이터셋의 이미지를 테스트 이미지로 사용하였다. 각 이미지에 대한 ground truth 태그는 이미지를 보며 관련된 태그를 일일이 달아줌으로써 생성하였다.

4.1 이미지 검색 정확도 측정

제안하는 이웃 이미지 검색의 성능을 평가하기 위해 이미지의 전체영역에 기반한 검색 정확도와 제안하는 관심영역에 기반한 검색 정확도의 성능을 비교하였다. 일반적으로 이미지 태깅 연구들의 이미지를 검색하는 단계에서는 다양한 시각적 특징들이 사용될 수 있다. 따라서 성능 평가는 어떠한 시각적 특징을 사용하던지 이미지의 전체영역 기반에 비해 제안하는 관심영역 기반의 이웃 이미지 검색이 더 높은 정확도를 갖는다는 것을 보여야 한다. 본 논문에서는 이미지 검색을 위한 시각적 특징으로써 PHOG와 색 현저도를 함께 사용하였다. 따라서 성능 평가는 이미지 특징으로 색 현저도만 이용한 경우, PHOG만 이용한 경우, 색 현저도와 PHOG를 함께 이용한 경우의 3가지 경우에 대해 이미지 전체영역 기반과 제안한 관심영역 기반의 이미지 검색의 정확도를 측정함으로써 평가하였다. 정확도 측정 시 검색하는 이웃이미지는 전체 데이터셋의 1/10 크기인 30개로 하였고, 검색된 30개의 이웃 이미지 중에서 쿼리 이미지와 같은 카테고리를 가진 이미지들의 비율을 통해 정확도를 계산하였다.

그림 4는 각 이미지 특징 별로 전체영역 기반과 관심영역 기반에서의 이미지 검색의 정확도를 비교함으로써 성능을 평가한 결과를 보여준다. 색 현저도를 이미지 특

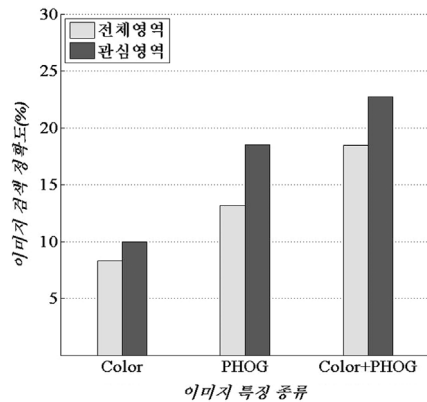


그림 4 이미지 특징 별 이웃 이미지 검색 정확도 비교

징으로 사용한 경우에는 전체영역 기반은 8.3%, 관심영역 기반은 10.0%로 평균 1.7%의 정확도 향상이 있었고, PHOG를 이미지 특징으로 사용한 경우 전체영역 기반은 13.6%, 관심영역 기반은 18.53%로 평균 5.37%의 정확도 향상이 있었다. 마지막으로 결합한 특징을 사용한 경우에는 전체영역 기반은 18.45%, 관심영역 기반은 22.73%로 평균 4.28%의 정확도 향상이 있었다. 색 현저도와 같은 이미지 검색을 위한 좋은 특징이 아닌 경우나 PHOG 또는 두 특징간 결합 같은 비교적 좋은 특징을 사용한 모두의 경우에서, 제안하는 관심영역 기반 방법이 전체영역 기반의 방법에 비해 이미지 검색 정확도가 더 높은 것을 알 수 있다. 앞에서 언급했던 것처럼, 이미지 검색에는 다양한 시각적 특징이 이용될 수 있으므로 향후 더 좋은 특징을 이용함으로써 제안하는 방법의 성능 향상 효과는 더욱 증가될 수 있다.

그림 5는 이미지 특징으로써 PHOG와 색 현저도를 결합하여 사용했을 경우 각 카테고리 별 이미지 검색 정확도이다. 그 결과 chair와 vehicle의 두 카테고리를 제외한 나머지 카테고리에 대해서는 관심영역에 기반한 이미지 검색 정확도가 더 높은 것을 알 수 있다.

그림 6은 이웃 이미지 집합을 검색한 결과의 예를 보

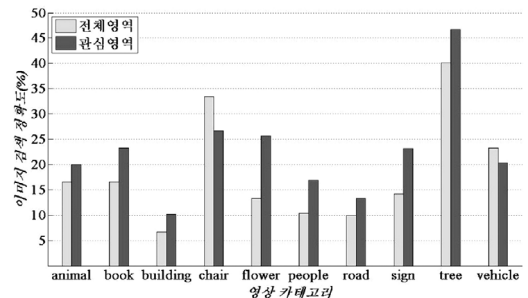


그림 5 각 카테고리 별 이미지 검색 정확도 비교

쿼리 이미지	이웃 이미지 집합			
				
				
				

그림 6 이웃 이미지 검색 결과의 예

여준다. 그림에서 볼 수 있듯이, 이웃 이미지 집합에는 쿼리 이미지와 관련된 이미지도 포함하는 반면에, 관련 없는 이미지도 포함하고 있다. 첫 번째 쿼리 이미지는 chair 카테고리으로써, 이웃 이미지 집합에는 chair 카테고리의 이미지 뿐만 아니라 sign과 book 카테고리의 이미지도 포함하고 있다. 이는 쿼리 이미지와 비슷한 색 분포와 형태를 가지고 있기 때문에 검색된 것으로 판단된다. 이러한 관련 없는 이웃 이미지는 결국 관련 없는 태그 추출을 가능하게 한다. 이 경우에, sign 관련 태그와 book 관련 태그들이 추출 가능할 수 있다. 두 번째 쿼리 이미지 역시 people 뿐 아니라 animal, road, book 카테고리 이미지를 포함하고 있으며, 세 번째 쿼리 이미지도 관련없는 building 카테고리 이미지를 포함하고 있다.

4.2 이미지 태깅 성능 비교

Li의 알고리즘[7]에서 제안한 이웃 투표 기법은 이미지 검색 성능의 한계로 인해 관련없는 태그가 추출될 수 있다고 하였다. 본 논문에서는 이를 해결하고 성능을 향상시키기 위해 시각적 유사도를 고려한 이웃 태그 투표기법을 제안하였다. 따라서 본 절에서는 제안한 방법의 성능을 평가하기 위해 Li의 알고리즘[7]에서 제안한 투표 기법과의 태그 추출 성능을 비교하였다. 그 이유는 Li의 알고리즘[7]은 간단하면서도 효율적인 방법으로 이전의 방법들[11,12]보다 더 높은 정확도를 가지며, 최근의 이미지 검색이나 태깅 관련 논문에서 많이 참조되는 알고리즘이기 때문이다. 표 1은 각 카테고리 별로 태그 추출 성능을 측정된 것이다. 두 방법 중에서 성능이 더 좋은 경우는 굵게 표시하였다. 성능을 측정하기 위해 Li의 알고리즘[7]에서 사용한 다음의 2가지 평가 척도를 이용하였다.

- **Precision at n ($P@n$):** n 개의 추출된 태그 중에서 관련있는 태그의 비율이다. 본 논문에서는 각 카테고리의 쿼리 이미지에 대해, $P@1$, $P@3$, $P@5$ 를 측정하였다. 관련있는 태그인지의 여부를 판단하기 위해서는 테스트 시 추출된 태그가 쿼리 이미지의 ground truth 태그와 일치하면 관련있는 태그로 판단하였다.
- **Average precision (AP):** 추출된 전체 태그 리스트에서의 각 태그 순으로 관련성을 측정함으로써, 이미지 태깅의 세 번째 단계인 태그 랭킹의 성능을 측정한다. 일반적으로 AP는 precision-recall 커브의 아래 영역에 대한 근사값이기 때문에 precision과 recall이 잘 결합된 평가 척도라고 알려져 있다. 본 논문에서는 각 카테고리마다 AP를 측정하고, 전체 성능을 측정하기 위해 MAP(Mean Average Precision)를 측정하였다. AP 값은 다음의 식 (3)을 통해 계산된다.

$$\frac{1}{R} \sum_{i=1}^l \frac{R_i}{i} \delta_i \quad (3)$$

이때 R_i 는 i 개 추출된 태그로부터 관련있는 태그의 수이며, R 은 전체 추출된 태그 리스트에서의 관련있는 태그의 개수이다. δ_i 는 i 번째 태그가 관련있는 경우 1값을, 관련없는 경우 0값을 갖는다. l 은 전체 추출된 태그 리스트에 있는 태그의 개수로서, 5로 하였다.

표 1에서 볼 수 있듯이, 제안하는 방법이 Li의 알고리즘[7]보다 $P@1$ 에서는 18%, $P@3$ 에서는 9.6%, $P@5$ 에서는 7%가 더 높은 정확도를 갖는다. 대체적으로 이웃 이미지 검색 성능이 좋은 카테고리일수록 태그 추출 성능도 좋았다. 그러나 building 카테고리의 경우 이미지 검색 성능이 10개의 카테고리 중 가장 좋지 않으면서도 세 번째로 높은 태그 추출 성능을 보였는데, 이는 각 building 카테고리의 이미지 태그를 조사한 결과 다른 카테고리에 비해 잡음 태그들의 비중이 낮기 때문인 것으로 판단된다. AP 역시 제안하는 방법이 대부분 더 높은 값을 가졌고, 따라서 제안하는 방법의 MAP가 8.2% 더 높은 성능을 가졌다.

그림 7은 제안하는 태깅 방법을 이용하여 태그를 추출한 결과의 예를 보여준다. 제안하는 방법은 Li의 알고리즘[7]에 비해 쿼리 이미지와 관련된 태그를 더 많이 추출하며 태그의 순서 또한 관련된 것이 먼저 나온 것을 볼 수 있다. 특히, 2번째 행의 chair 카테고리 이미지의 경우를 보면, Li의 알고리즘[7]은 관련된 chair 태그 외에도 book과 sign 태그를 추출하는 것을 볼 수 있다. 이는 그림 6에서 나온 첫 번째 쿼리 이미지와 동일한 이미지인데, 그림 6의 이웃 이미지 집합을 보면 chair 카테고리 외에도 book과 sign 카테고리의 이미지를 포함하고 있다. 이 경우 Li의 알고리즘[7]은 이러한 관련없는 이미지의 태그로부터 투표 기법을 이용하여

표 1 각 카테고리 별 태그 추출 성능.

Query	Precision at 1		Precision at 3		Precision at 5		Average precision	
	Li et al. [7] method	Proposed method	Li et al. [7] method	Proposed method	Li et al. [7] method	Proposed method	Li et al. [7] method	Proposed method
animal	0.100	0.333	0.111	0.256	0.080	0.200	0.027	0.142
book	0.367	0.400	0.267	0.211	0.167	0.167	0.140	0.125
building	0.767	0.567	0.400	0.567	0.420	0.493	0.312	0.386
chair	0.600	0.733	0.411	0.567	0.373	0.453	0.275	0.361
flower	0.533	0.400	0.322	0.311	0.207	0.293	0.162	0.198
people	0.100	0.400	0.178	0.244	0.207	0.200	0.114	0.163
road	0.233	0.467	0.322	0.422	0.267	0.400	0.146	0.275
sign	0.033	0.400	0.189	0.267	0.227	0.240	0.088	0.152
tree	0.067	0.900	0.478	0.756	0.320	0.507	0.195	0.472
vehicle	0.200	0.100	0.133	0.167	0.120	0.133	0.075	0.073
average	0.290	0.470	0.281	0.377	0.239	0.309	0.153	0.235

관련없는 태그를 추출하는 것이다. 본 논문에서 제안한 방법은 태그 추출을 위한 투표 기법 적용시, 이미지 유사도를 고려하기 때문에 chair 카테고리 이미지보다 더 낮은 유사도를 갖는 book과 sign 카테고리 이미지의 태그는 가중치가 낮아져서 태그로 추출되지 않는다. 그림 7의 결과에서 볼 수 있듯이 쿼리와 관련된 태그만 추출된 것을 볼 수 있다. 나머지 쿼리 이미지 역시 이러한 이유로 Li의 알고리즘[7]은 관련없는 이미지의 태그가 고려됨으로써 잘못된 태그 추출이 가능하나, 제안하는 방법은 이 문제를 효과적으로 해결한다.

여기서 주목할 점은 제안하는 태그 추출 방법에서 이미지 유사도 점수로써 이용된 이미지 검색의 성능이다. 비록 제안하는 이미지 검색 방법이 전체영역 기반보다 더 높은 정확도를 보였지만, 정확도 평균이 22.73%로 만족할 만한 정도는 아니다. 이러한 정확도를 갖는 이미지 검색 방법을 유사도 점수로써 활용했는데도 불구하고 기존의 방법보다 더 높은 태깅 성능을 갖는 것을 볼 수 있다. 이는 향후 더 좋은 이미지 특징을 이용한 향상된 이미지 검색 방법을 이용하면, 제안하는 방법을 이용한 태그 추출의 성능을 더 높일 수 있다고 볼 수 있다.











쿼리 이미지	Li et al. [7] method	Proposed method	쿼리 이미지	Li et al. [7] method	Proposed method
	building face horse	horse <i>animal</i> <i>grass</i>		<i>building</i> face person	<i>artificial</i> <i>building</i> <i>grass</i>
	flower tree grass	chair <i>book</i> spring		<i>chair</i> book sign	<i>chair</i> artificial grass
	book <i>flower</i> sign	<i>flower</i> spring grass		building <i>face</i> horse	<i>face</i> <i>person</i> building
	artificial grass road	<i>road</i> <i>highway</i> building		book flower <i>sign</i>	<i>sign</i> book flower
	book flower sign	<i>tree</i> flower spring		<i>car</i> mediaday 2010	artificial <i>car</i> <i>racing</i>

그림 7 제안하는 방법을 이용한 태그 추출 결과의 예. 이탤릭체는 관련된 태그를 의미하며, 볼드체는 제안하는 방법에 의해 추출된 관련된 태그를 의미한다.

5. 결 론

본 논문에서는 데이터 기반 방식의 가중치된 이웃 태그 투표 기법을 통한 이미지 태깅 방법을 제안하였다. 먼저 이미지 태깅 성능에 중요한 영향을 미치는 이웃 이미지 검색의 성능을 높이기 위해 이미지의 관심 영역을 추출하고 그 영역에 대한 내용 기반 이미지 검색을 실시하였다. 검색된 이웃 이미지 집합에 대해 이미지 유사도를 고려한 가중치된 투표 기법을 통해 태그의 관련성을 계산하였고, 이는 앞 단계에서 관련없는 이미지가 추출된 경우 발생할 수 있는 관련 없는 태그 추출의 가능성을 막는 다는 것을 실험을 통해 입증하였다. 실험은 이미지 검색 정확도 및 이웃 투표 기법과 제안하는 방법과의 태그 추출 정확도를 비교하였고, 이를 통해 제안하는 이미지 태깅 방법의 우수성을 입증하였다.

비록 제안하는 방법이 이전 방법에 비해 우수하지만, 논문에서 기술했던 것처럼 이미지 검색의 성능적 한계로 인해 이미지 유사도가 잘못 측정되는 경우가 발생할 수 있다. 이로 인해 관련없는 이웃 이미지 검색과 태그 추출이 가능할 수 있다. 이는 더 좋은 이미지 특징을 이용하여 이미지 검색의 성능을 높이거나, 태그 관련성 계산시 다른 여러가지 요인도 함께 고려하여 추출함으로써 해결이 가능할 것이다.

참 고 문 헌

- [1] Y. Mori, H. Takahashi and R. Oka, "Image-to-Word Transformation Based on Dividing and Vector Quantizing Images with Words," *Proc. First Int'l Workshop Multimedia Intelligent Storage and Retrieval Management*, 1999.
- [2] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," *Proc. European Conf. Computer Vision*, pp.97-112, 2002.
- [3] D. Blei and M.I. Jordan, "Modeling Annotated Data," *Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, pp.127-134, 2003.
- [4] T. Yeh, K. Tollmar and T. Darrell, "Searching the Web with Mobile Images for Location Recognition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, no.2, pp.76-81, 2004.
- [5] K. Weinberger, M. Slaney, and R. van Zwol, "Resolving Tag Ambiguity," *Proc. ACM Multimedia*, pp.111-119, 2008.
- [6] X.-J. Wang, L. Zhang, X. Li, and W.-Y. Ma, "Annotating Images by Mining Image Search Results," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol.20, no.11, pp.1-14, 2008.
- [7] X. Li, C. G. M. Snoek, and M. Worring, "Learning

- Social Tag Relevance by Neighbor Voting," *IEEE Trans. Multimedia*, vol.11, no.7, pp.1310-1322, 2009.
- [8] J. Harel, C. Koch and P. Perona, "Graph-Based Visual Saliency," *Advances in NIPS 19*, pp.545-552, 2007.
- [9] J. van de Weijer, T. Gevers and A.D. Bagdanov, "Boosting Color Saliency in Image Feature Detection," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol.28, no.1, pp.150-156, 2006.
- [10] A. Bosch, A. Zisserman and X. Munoz, "Representing shape with a spatial pyramid kernel," *Proc. ACM Int'l Conf. Image and Video Retrieval*, pp.401-408, 2007.
- [11] A. Torralba, R. Fergus and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol.30, no.11, pp.1958-1970, 2008.
- [12] C. Wang, F. Jing, L. Zhang and H.-J. Zhang, "Scalable search-based image annotation," *Multimedia Systems*, vol.14, no.4, pp.205-220, 2008.



조 선 영

2007년 숙명여자대학교 컴퓨터과학과 졸업. 2009년 연세대학교 컴퓨터과학과 석사 졸업. 2009년~현재 연세대학교 컴퓨터과학과 박사과정. 관심분야는 영상인식, 패턴인식



차 재 성

2010년 광운대학교 컴퓨터공학과 졸업. 2010년~현재 연세대학교 컴퓨터과학과 석사과정. 관심분야는 패턴인식, 영상처리



변 혜 란

1980년 연세대학교 수학과 졸업(이학사). 1983년 연세대학교 대학원 수학과 졸업(이학석사). 1987년 University of Illinois, Computer Science(M.S.). 1993년 Purdue University, Computer Science(Ph.D.). 1994년~1995년 한림대학교 정보공학과 조교수. 1995년~1998년 연세대학교 컴퓨터과학과 조교수. 1998년~2003년 연세대학교 컴퓨터과학과 부교수. 2003년~현재 연세대학교 컴퓨터과학과 교수. 관심분야는 패턴인식, 영상처리, 영상인식