

Visually Weighted Neighbor Voting을 이용한 이미지 태그 정제 기술

*이시형 *Wesley De Neve *노용만

한국과학기술원 영상 및 비디오 시스템 연구실

*ijiat@kaist.ac.kr

Image Tag Refinement using Visually Weighted Neighbor Voting

*Sihyoung Lee *Wesley De Neve *Yong Man Ro

Korea Advanced Institute of Science and Technology, Image and Video Systems Lab

요약

온라인을 통한 이미지 공유는 사용자들이 활발하게 이용하고 있는 분야 중 하나이다. 사용자의 활발한 참여로 거대해진 이미지 데이터 베이스 내에서 효율적으로 이미지 검색을 수행하기 위해서는 이미지를 정확하게 표현하고 있는 태그의 존재가 매우 중요하다. 하지만, 최근 이미지에 등록 태그 중에서 상당 부분이 이미지와는 직접 관련이 없는 노이즈 태그라는 조사 결과는 노이즈 태그로 인해서 이미지 검색의 정확성이 저하될 수 있다는 가능성을 암시한다. 그래서 노이즈 태그를 효과적으로 구분하기 위해서는 태그의 종류에 적합한 태그 정제 기술을 도입할 필요가 있다. 본 연구는 이를 위해서 이미지의 시각적 유사도에 기반한 Visually weighted neighbor voting 방법을 제안했다. 이를 통해서 이미지와 태그 사이의 관련성을 효과적으로 측정할 수 있었다. 그리고 기존 기술보다 안정적으로 노이즈 태그를 구분할 수 있음을 실험을 통해서 증명하였다.

1. 서론

멀티미디어 기기의 급속한 확산과 네트워크 기술의 발전은 온라인을 통한 이미지 공유를 활성화시켰다. 대표적인 온라인 이미지 공유 사이트 Flickr.com에는 2010년 9월 500만장 이상의 영상이 등록되었고, 매 분 3천장 이상의 새로운 이미지가 등록되고 있다. 이처럼 대용량 이미지가 온라인을 통해서 공유되면서, 이미지를 효율적으로 관리하고 검색하는 방법에 대한 중요성이 강조되고 있다.

현재 대부분의 이미지 검색 기술은 사용자가 키워드를 검색 엔진에 입력하면, 매칭되는 태그가 등록된 이미지를 찾아주는 태그 매칭 방법을 통해서 이루어지고 있다. 이런 환경에서 정확한 이미지 검색을 보장하기 위해서는 이미지에 정확한 태그가 충분히 등록되어 있어야 한다. 만약, 특정 이미지에 그 이미지와는 전혀 무관한 태그가 존재한다면, 태그 매칭을 통한 이미지 검색의 정확성을 감소시키는 요인이 된다. 예를 들어서, 'sky'를 표현하고 있는 이미지에 'sky'가 아닌 'water'라는 잘못된 태그가 등록되어 있다면, 사용자는 'sky'라는 태그를 쿼리로 입력해서 이미지를 검색할 수 없을 뿐만 아니라, 'water'와 관련된 이미지를 검색할 때, 잘못된 이미지를 검색 결과로 획득하게 된다.

최근 이미지와 그에 등록된 태그 사이의 연관성을 분석한 연구는 이미지에 등록된 태그 중에서 상당 부분이 이미지와는 전혀 상관없는 '노이즈 태그'라는 관찰을 보고했다[1]. 그리고 이런 노이즈 태그는 사용자가 태그를 등록할 때는 이미지 표면에 드러난 의미뿐만 아니라 사용자의 경험, 사전 정보로 이미지를 재생산해내서 이해한 의미도 태그로 등록하는 성향이 있기 때문이라는 것을 밝혀냈다[2].

최근 이런 문제를 해결하기 위해서 이미지와 상관없는 노이즈 태그를 정제하는 기술에 대한 연구가 활발하게 진행되고 있다. 그 중에서 Neighbor voting 알고리즘은 이미지와 태그 사이의 관련성 측정을 통해서, 이미지와 관련성이 낮은 노이즈 태그를 구분한다[3]. Neighbor

voting 알고리즘은 이미지에 등록된 태그와 이미지와의 관련성을 이미지와 시각적으로 유사한 특징을 가진 이미지 집합, visual neighbor set의 태그 분석을 통해서 예측한다. 입력 이미지와 데이터 베이스에 등록된 이미지 사이의 시각적 유사도(visual similarity) 비교를 통해서 입력 이미지에 대한 visual neighbor set을 생성하고, 입력 이미지에 등록된 태그가 visual neighbor set에 태그로 등록된 빈도 수와 전체 데이터 베이스에 등록된 빈도수를 고려해서 입력 이미지와 태그 사이의 관련성을 추출한다.

Neighbor voting 알고리즘은 데이터 베이스로부터 구성된 visual neighbor set에 포함된 이미지에 동일한 가중치를 두고 태그 관련성을 추출한다. 만일 visual neighbor set를 크기를 증가해서, 시각적으로 덜 유사한 이미지가 visual neighbor set에 포함되더라도 시각적으로 매우 유사한 이미지와 같은 가중치를 갖게 된다. 그래서 [3]의 저자가 밝혔듯이, Neighbor voting 알고리즘은 visual neighbor set의 크기에 따라서 태그 정제 성능 변화가 심하게 된다.

본 논문에서는 visual neighbor set의 크기에 따른 태그 정제 성능 변화에 대한 Neighbor Voting 알고리즘의 단점을 보완하기 위해서, visual neighbor set에 포함된 이미지에 시각적 유사도를 가중치로 두는 Visually weighted neighbor voting을 제안한다. Visually weighted neighbor voting은 입력 이미지와 시각적으로 더 유사한 이미지에 많은 가중치를 주고, 입력 이미지와 시각적으로 덜 유사한 이미지에 적은 가중치를 줄여서, 효과적으로 입력 이미지와 태그 사이의 관련성을 측정한다.

이를 위해서 2장에서는 본 논문에서 제안한 Visually weighted neighbor voting 알고리즘에 대한 설명을 하고, 3장에서는 실제 온라인 이미지 공유 사이트에서 획득한 데이터를 통해서 이전 방법에 비한 성능 향상을 실험적으로 증명한다. 마지막으로 4장에서는 논문의 결론을 맺는다.

2. Visually Weighted Neighbor Voting

입력 이미지 i 에 등록된 태그 집합을 T_i 라 하자. 이 때, 태그 정제 과정은 입력 이미지에 등록된 태그 $t \in T_i$ 중에서 태그 관련성 값이 임계치 이상인 태그를 선별하는 과정으로 정의한다. 태그 정제 과정은 다음 수식으로 표현된다[4].

$$T_i^{refined} = \{t \in T_i, R(i, t) \geq \text{threshold}\}, \quad (1)$$

여기서, $R(i, t)$ 는 입력 이미지 I 와 태그 t 사이의 태그 연관성을 측정하는 함수를 나타내고, threshold 는 태그 t 를 노이즈 구분할지에 대한 임계 태그 연관성 값을 의미한다.

본 논문에서는 Visually weighted neighbor voting 알고리즘을 함수 $R(\cdot)$ 로 제안한다.

$$R(i, t) = \frac{\sum_{j \in N(i, k)} \text{sim}(i, j) \cdot \text{vote}(j, t) - k}{|F|} \quad (2)$$

여기서, $N(i, k)$ 는 이미지 i 에 대해서 생성한 크기가 k 인 visual neighbor set을 의미한다. $\text{sim}(i, j)$ 는 입력 이미지 i 와 visual neighbor set에 포함된 이미지 j 사이의 시각적 유사도를 의미하고, $\text{vote}(j, t)$ 는 이미지 j 에 태그 t 가 등록되어 있으면 1, 등록되어 있지 않으면 0을 갖는다. 그리고 F 는 이미지 폭소노미를 나타내고, $|F|$ 는 이미지 폭소노미에 등록된 이미지의 수를 의미한다.

수식 (2)에서 확인할 수 있는 것처럼, $\text{vote}(j, t)$ 에 $\text{sim}(i, j)$ 를 곱해서 가중치로 뚫으로써, visual neighbor set에 속한 이미지라도 시각적 유사도에 따라서 서로 다른 가중치를 가진다. 그래서 제안하는 방법은 visual neighbor set에 있는 모든 이미지가 같은 가중치를 가지는 기존 방법에 비해서 visual neighbor set의 크기가 증가하더라도 태그와 이미지의 관련성을 측정할 때 발생하는 오차를 감소시킨다.

3. 실험 및 분석

논문에서 제안한 Visually weighted neighbor voting 알고리즘의 효율성을 증명하기 위해서 웹을 통해서 공개된 MIRFLICKR-25000 데이터 베이스[5]를 사용하였다. MIRFLICKR-25000 데이터 베이스는 9,862명의 사용자가 Flickr.com에 등록한 25,000장의 이미지와 그 이미지에 등록된 223,537개의 태그로 구성되어 있다. 그 중에서 실험을 위해서 4,862개의 태그가 등록된 200장의 이미지를 선택했고, 4,862개의 태그는 수동으로 정확한 태그와 노이즈 태그로 구분되었다. 이를 통해서 3,615개의 태그가 이미지와 상관없는 노이즈 태그로, 1,247개의 태그가 이미지와 관련된 태그로 분류되었다.

본 실험에서 visual neighbor set을 구성하고 이미지 사이의 시각적 유사도를 측정하기 위해서 MPEG-7 Scalable Color Descriptor[6]를 사용하였다. 그리고 visual neighbor set은 입력 이미지 i 에 대해서 콘텐츠 기반 k-nearest neighbor(k-NN)을 통해서 구성하였다[3]. 수식 (2)의 $\text{sim}(i, j)$ 는 cosine similarity를 통해서 측정하였다.

그림 1은 visual neighbor set의 크기에 따른 본 논문에서 제안한 방법과 Neighbor voting 방법에 따른 Noise Level의 차이를 나타내고 있다. Noise Level은 전체 태그 중에서 노이즈 태그가 차지하고 있는 비율을 수치화한 것이다. Noise Level이 낮을수록 노이즈 태그가 차지하는 비율이 낮아져서 태그 정제가 성공적으로 이루어졌다는 것을 의미한다. 그림에서 확인할 수 있는 것처럼, 제안하는 방법이 visual neighbor set의 크기를 증가하더라도, Neighbor voting 방법에 비해서 더 좋은 성능을 보이는 것을 확인할 수 있었다. 뿐만 아니라, 제안하는

방법이 visual neighbor set의 크기 변화에 상대적으로 안정된 성능을 보장하는 것을 확인할 수 있다.

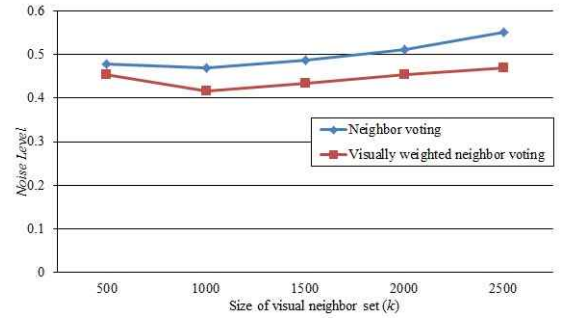


그림 1 Visual neighbor set의 사이즈에 따른 제안한 방법과 Neighbor voting 방법의 성능 비교

4. 결론

이미지에 등록된 태그 중에서 상당수가 노이즈 태그라는 최근 관찰은 태그 매칭을 통한 현재 이미지 검색 환경에서 많은 이미지 검색 오류 가능성을 암시한다. 이런 오류를 줄이기 위해서 노이즈 태그를 효과적으로 구분할 필요성이 대두되고 있다. 본 연구는 이를 위해서 이미지의 시각적 유사도에 기반한 Visually weighted neighbor voting 방법을 제안했다. 이를 통해서 이미지와 태그 사이의 관련성을 효과적으로 측정할 수 있었다. 그리고 기존 기술보다 visual neighbor set의 크기에 영향을 받지 않고 안정적으로 노이즈 태그를 구분할 수 있음을 실험을 통해서 증명하였다.

Acknowledgement

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0012495)

참고문헌

- [1] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, "NUS-WIDE: A Real-World Web Image Database from National University of Singapore," Proc. of ACM International Conference on Image and Video Retrieval, pp. 223-232, July 2009.
- [2] S. Lindstaedt, R. Mörzinger, R. Sorschag, V. Pammer, and G. Thallinger, "Automatic Image Annotation using Visual Content and Folksonomies," Multimedia Tools and Applications, vol. 42, no. 1, pp. 97-113, March 2009.
- [3] X. Li, C. G. M. Snoek, M. Worring, "Learning Social Tag Relevance by Neighbor Voting," IEEE Trans. Multimedia, vol. 11, no. 7, pp. 1310-1322, Aug. 2009.
- [4] S. Lee, W. De Neve, and Y. M. Ro, "Tag Refinement in an Image Folksonomy using Visual Similarity and Tag Co-occurrence Statistics," Signal Processing: Image, vol. 25, no. 10, pp. 761-773, 2010.
- [5] M. J. Huiskes and M. S. Lew, "The MIR Flickr Retrieval Evaluation," Proc. ACM International Conference on Multimedia Information Retrieval, pp. 39-43, Oct. 2008.
- [6] B. S. Manjunath et al., "Introduction to MPEG-7: Multimedia Content Description Language," John Wiley and Sons, 2002.