

휴대단말용 이벤트-기반 사진 경계 분할 및 브라우징 방법

김상철^o 낭종호

서강대학교 컴퓨터공학과

smaslayer1@sogang.ac.kr, jhnang@sogang.ac.kr

An Event-based Clustering and Browsing of Personal Photo Collections on Mobile Device

Sangchul Kim^o Jongho Nang

Dept. of Computer Science and Engineering, Sogang University

요 약

최근 모바일 기기의 저장장치 용량이 늘어나면서 사용자는 대량의 사진을 휴대하고 다닌다. 하지만 현재 대량의 사진을 한정적 크기의 화면에 효과적으로 보여줄 수 있는 인터페이스가 부족하다. 모바일 기기에서 사용자 입장에서 편의성을 제공하는 사진 브라우징을 위해서는 직관적인 탐색 방법과 탐색시간을 단축시키는 방법이 필요하다. 이를 위해 본 논문에서는 모바일 기기에 저장된 사진들에 대해 이벤트 별 사진을 자동 분류하며 이벤트 내의 객체 인식을 통해 이벤트에 자주 나오는 객체 정보들을 제공하여 직관적인 브라우징이 가능하도록 하는 방법들을 제안한다. 제안한 방법으로는 이벤트 기반의 브라우징과 객체 기반의 브라우징 방법이 있다. 이벤트 기반의 브라우징을 위해서 시간과 위치정보를 이용하여 이벤트를 군집화하고 통계적 자료에 근거한 이벤트 자동 경계 검출 방법을 사용했다. 또한 객체 기반의 브라우징을 위해서 객체 인식을 통해 사진들을 객체별로 분류하는 방법을 사용하였다. 사진내에서 객체의 인식을 위해 BoW(Bag of Word)를 사용하였으며 인식율을 높이기 위해 TF-IDF를 적용한 방법을 제안하였다. 본 방법은 기존의 방식에 비해 객체 인식이 더 높음을 확인했다.

1. 서 론

최근 태블릿PC, 스마트폰 등 카메라와 GPS가 장착된 모바일 멀티미디어 기기들이 보급 되고 있다. 이러한 모바일 기기들은 위치정보를 활용할 수 있다. 또한 저장장치의 용량이 커지면서 대량의 사진 데이터를 저장할 수 있게 됐다. 하지만 대부분 모바일 멀티미디어 기기들은 썸네일 그리드 뷰 방식의 사진 뷰어 방식을 제공한다. 이 같은 인터페이스는 대량의 데이터 탐색을 할 때 사진의 정렬 방법이 단순하기 때문에 순차적인 탐색을 할 수 밖에 없다. 따라서 사용자가 사진을 탐색할 때 시간을 낭비하게 한다. 이에 보다 효과적인 사진 뷰어 방식이 필요하다.

일반적으로 사용자는 개인용 사진 컬렉션에 대하여 이벤트(예를 들어 벚꽃놀이, 생일파티, 돌잔치) 기준으로 구성하고 접근하는 것을 선호한다[1]. 뿐만 아니라 사람들은 어떠한 기억을 떠올릴 때 장소나 어떠한 사물 위주의 연상을 하게 된다. 따라서 이벤트별로 사진을 분류하여 보여주는 기능뿐 아니라 위치별로 사진을 분류하여 보여주는 기능 그리고 객체별로 분류하여 사진을 보여주는 기능들이 있다면 사용자에게 유용할 것이다. 이러한 시스템을 구현하기 위해선 이벤트 기반 사진 자동 분류 방법과 사진 내 객체 인식 및 자동 분류 방법이 필요하다.

이벤트 기반 사진 자동 분류를 하기 위해 이벤트의

기준이 되는 시간과 이벤트가 발생한 장소를 이용하여 이벤트를 분류하며, 모바일 기기의 부족한 연산능력을 고려하는 이벤트 클러스터링 방법을 적용하였다.

객체 자동 분류를 위한 객체 인식 방법으로는 최근 이미지 객체 인식 분야에서 활발히 연구가 진행되고 있고 다른 방법에 비해 인식률도 좋은 BoW방법을 적용하였다. BoW방법은 자연어 처리에서의 텍스트 자동 분류방법이다. 이미지 분야에서는 이미지 한장을 하나의 문서, 이미지의 로컬패치(Local patch)를 하나의 단어로 생각하여 이미지를 분류하는 방법이다. 또한 BoW는 다른 객체 인식 방법들에 비해 객체의 인식률이 크게 높아짐을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 연구에 대해 소개 하고 3장에서 본 논문에서 제안한 브라우징 방법들에 대해 설명하며 4장에서 실험 분석 및 구현한 데모시스템을 소개하며 5장에서 결론 및 향후 연구로 맺는다.

2. 관련 연구

2.1 이벤트 클러스터링

이벤트 별로 분류하여 보여주기 위해서는 사진 집합에서 이벤트 경계를 구분해주는 방법이 필요하다. [2]에서 제안한 방법은 시간 간격만을 고려하여 계층적으로 클러스터링을 하여 사진을 관리하는

방법으로 속도가 빠르지만 오프-라인 모드만 지원하는 단점이 있다. [3]에서 제안한 방법은 윈도우를 이동하여 이전 단계들의 정보를 기반으로 이벤트 경계를 검출하는 방법은 윈도우 크기에 민감하고 이 크기를 지정하는 데에 한계가 있다. [4]에서 제안한 Mean-shift 클러스터링을 한 뒤에 GPS 정보를 이용하여 보정하는 방법은 위의 방법들보다 정확하지만 Mean-shift 클러스터링 시에 연산 소요 시간이 많이 필요하다는 단점이 있다. [5]에서 제안한 순차적 분할 이벤트 클러스터링을 위해 경계를 통계적으로 찾아내는 방법은 연산 소요시간이 적으며 위치정보와 시각정보를 함께 고려했다. 하지만 통계적 수치를 얻어내기 위해 많은 샘플 데이터가 필요하다.

2.2 객체 인식

최근 이미지 객체 인식 분야는 이미지의 전체를 고려하는 것이 아니라 국소부위를 고려한 방법들이 각광받고 있다. 이러한 종류중에 파트 기반 모델은 국소영역에 포함된 객체들간의 위치 구조를 모델링하여 전체객체를 인식하는 방법이다. 국소 영역의 객체를 인식하는 과정과 각 객체들간의 구조를 따로따로 모델링 하는 작업이 너무 많은 단점이 있다. 이에 반해 BoW 는 로컬패치가 나타난 빈도에 따라 코드북(codebook)을 참조해 객체를 인식하는 방법으로 모델링이 단순하며 객체 인식률이 더 높지만 이미지 분야에서는 사전역할을 하는 코드북을 제작해야 하는 단점이 있다.[6]

3. 휴대단말용 이벤트-기반 사진 자동 분류 및 검색/브라우징

그림 1은 본 논문에서 제안한 시스템의 구조도이다.

본 시스템은 이벤트 자동 분류, 객체 인식을 위한 코드북 생성 그리고 객체 인식을 위한 분류기 학습의 3가지 방법을 통해 브라우징 인터페이스를 제공한다. 이벤트 분류를 위해서 먼저 시간정보와 위치정보를

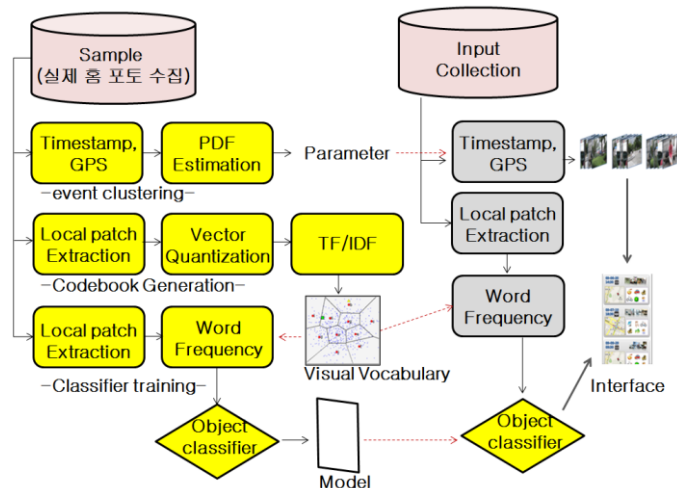


그림 1 시스템 구조도

이용하여 이벤트들을 군집화한 뒤 샘플의 통계적 자료를 통해 이벤트를 자동 분류하였다. 그리고 BoW방법의 객체 인식을 위해서 이미지에서 로컬패치를 추출하여 코드북을 생성을 한다. 이때 벡터 양자화 하게 되면서 발생하는 정보손실 문제를 해결하기 위해서 코드북에 TF-IDF를 적용시킨다. 분류기 학습을 위해서 로컬패치들을 코드북에 참조하여 출현빈도를 계산한다. 이렇게 3가지의 과정을 수행하여 만들어진 학습데이터를 토대로 이벤트 자동 분할 브라우징 인터페이스를 제공한다.

3.1 이벤트 분류

모바일 기기의 특성상 연산량이 적어야 하며 GPS가 내장되어 위치정보를 함께 고려할 수 있다. 따라서 본 논문에서는 [5]에서 제안한 휴대폰 상에서의 이벤트 분할 방법을 적용하여 이벤트를 분류하였다.

3.2 객체 인식

개인용 사진집합은 객체들이 복잡하게 나타난다. 종류, 위치, 크기, 방향이 모두 제 각각이다. 따라서 종류에 둔감하기 위해 BoW를 이용하였다. BoW의 로컬 패치로서는 위치, 크기, 방향에 둔감한 SIFT-descriptor를 적용하였다.[7] 하지만 BoW를 적용할 때 벡터 양자화 과정에서 정보의 손실이 발생하여 모든 로컬 패치들이 객체 분별에 도움이 주지 않는다. 오히려 객체 분별에 방해가 되는 로컬 패치들도 생기기 때문에 객체 분별에서 의미 있는 로컬 패치들을 분류하는 방법이 필요하다.

따라서 언어모델에서의 TF-IDF를 적용하였다. 그림 2를 보면 TF에 해당하는 오브젝트의 로컬패치는 분별력이 있는 반면 IDF에 해당하는 배경은 분별력에 방해가 됨을 알 수 있다. (1)에서 $tf_{i,c}$ 는 이미지n에 있는 로컬

$$tf_{i,c_j} = \frac{n_{i,c_j}}{\sum_k n_{k,c_j}} \quad (1)$$

$$idf_{i,c_j} = \frac{1}{\sum_{k,k \neq j} tf_{i,c_k}} \quad (2)$$

패치가 클래스 c에서 얼마나 자주 나오는지 측정하고 (2)는 $idf_{i,c}$ 는 로컬 패치 i가 클래스 c외의 다른 클래스에서 얼마나 희박한지를 측정하였다. 또한 그림 3을 보면 특정 객체내의 유사한 로컬 패치들은 중첩되는 경향이 나타난다. 따라서 이런 로컬 패치들이 동시에 연달아서



Low IDF High TF
그림 2 TF-IDF 효과 예시



그림 3 로컬패치들의 중첩 예시

나올 확률을 계산해야 한다. 언어 모델에서는 n-gram을 적용하여 계산한다. 본 논문에서는 모바일 기기의 특성을 고려해 계산비용이 적은 Bi-gram을 적용하였다. (3)은 Bi-gram을 적용해 수식화 한 것이다.

$$p(w_{ij} | w_{i,j-1}, C_k) = \frac{Freq(w_{i,j-1}, w_{ij} | C_k)}{Freq(w_{i,j-1} | C_k)} \quad (3)$$

$p(w_{ij} | w_{i,j-1}, C_k)$ 는 i번째 이미지에서 j-1번째 로컬패치 w 와 특정객체 C_k 군집이 함께 나왔을 때, w_i 의 j번째 단어가 나올 확률로서 중첩된 로컬 패치들이 연달아 나온 확률로 볼 수 있다. 이를 계산 하기 위해선 C_k 가 나왔을 때 $w_{i,j-1}$ 까지 출현빈도에서 $w_{i,j-1}$ 와 w_{ij} 가 함께 나온 확률을 계산하면 구할 수 있다.

위에서 계산한 TF-IDF와 Bi-gram 값을 사용하여 식 (4)와 같이 각 로컬 패치들의 가중치 값을 계산하였다. 특히 Bi-gram의 경우 0일 때 전체 가중치가 0이되 버리는 것을 막기 위하여 SIGMOID함수를 적용시켰으며, TF-IDF의 경우 비 선형적인 스케일링을 위하여 마찬가지로 SIGMOID함수를 적용시켰다.

$$F_{ij,C_k} = \frac{1}{1 + e^{-a(tf_idf_{i,C_k} - b)}} \times \frac{1}{1 + e^{-c(p(w_{ij}|w_{i,j-1}, C_k) - d)}} \quad (4)$$

앞서 설명한 바와 같이 영상에서 추출된 로컬 패치들의 가중치는 해당 패치가 어떤 클래스에 속할 지를 나타내는 확률과 같은 역할을 한다고 볼 수 있다. 따라서 식 (5)와 같이 각 패치들의 가중치 F_{ij,C_k} 를 곱하여서 가장 높은 값의 클래스를 선택하는 방식으로 분류하였다.

$$\hat{C} = \arg \max_{C_k} p(C_k | I) = \arg \max_{C_k} \prod_{w_{i,j} \in I} F_{ij,C_k} p(C_k) \approx \arg \max_{C_k} \prod_{w_{i,j} \in I} F_{ij,C_k} \quad (5)$$

3.3 브라우징 인터페이스

본 논문에서는 썸네일 그리드 뷰 방식보다 보다 효과적으로 사용자에게 사진 탐색을 할 수 있도록 하는 브라우징 방법을 제안하였다. 서론에서도 밝혔듯이 사람들은 이벤트 기준으로 사진을 분류하는 것을 가장 선호 한다. 따라서 이벤트 기반으로 브라우징 하는 것을 기본 인터페이스로 하였다. 이벤트 기반으로 분류한 영역별로 나타나는 그 이벤트내의 대표 사진들을 썸네일 형태로 보여주었다. 이 대표사진을 클릭함으로써 이벤트내의 전체 사진을 볼 수 있게 했다. 또한

이벤트는 시간 뿐만 아니라 공간적 특성에 의해서도 분류되어 있으므로 분류된 사진들과 함께 이벤트가 일어난 시간과 지역을 지도로 표시해주었다. 이 지도를 클릭함으로써 위치별로 분류되어 있는 사진들을 볼 수 있다. 또 이벤트 내에 자주 나타나는 객체들을 보여줌으로써 이벤트에 대한 직관적 이해를 돕게 하였다. 또한 이 객체를 클릭함으로써 객체가 포함되어 있는 사진 전체를 브라우징 할 수 있게 하였다.

4. 실험 및 구현

[5]의 이벤트 클러스터링 방법을 위해선 많은 양의 샘플이 필요하다. 또한 모바일 기기를 이용하여 찍을 경우에는 자주 나올 법한 객체들을 정해져 있을 것이다. (예를 들면 자동차, 얼굴, 애완동물 등) 따라서 본 논문에서 수집한 데이터 집합의 양과 인식을 위해 학습한 객체들은 그림 4와 같다.

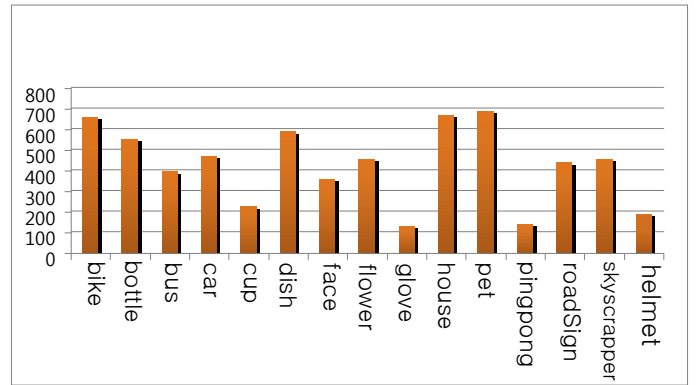


그림 4 데이터 집합과 학습한 객체들

4.1 BoW에 TF-IDF와 Bi-gram 적용

본 논문에서는 BoW를 통한 객체인식의 문제점을 보완하기 위해 언어모델에서의 Bi-gram과 TF-IDF 방법을 적용시켜보았다. 그림 5는 기본 BoW를 적용시켰을 때의 정확도이다. 전체적으로 정확도가 낮아 하지만 그림 6은 Bi-gram과 TF-IDF를 적용 한 뒤에 단일 객체만 있는 사진에서의 정확도이다. 기본 BoW만 적용했을 때 보다 좋은 성능을 나타냄을 알 수 있다. 이는 보통의 경우 배경의 로컬패치들을 TF-IDF로 걸러주고 객체 자체의 로컬패치들만 분별할 수 있게 되기 때문이다. 하지만 bike객체의 경우에는 분별력이 낮아진 것을 볼 수 있다. 샘플데이터중 버스나 자동차는 아스팔트 도로가 배경의 대부분에 해당한다면 자전거는 잔디밭, 도로위, 담벼락 앞등 배경이 다양했다. 따라서 TF-IDF의 효과가 없어 오히려 분별력에 도움이 되지 않았다.

4.2 인터페이스

그림 7은 본 논문에서 제안한 방법을 적용하여 만든 데모 시스템이다. 적용 환경은 모바일 기기는 갤럭시 K

를 사용하였고 OS는 Google Android 2.2 Froyo가 탑재되어 있다. 그림 7은 좌측 순서대로 이벤트 브라우징과 위치-기반 브라우징을 나타냈다. 맨 우측은 오인식한 객체를 수정하기 위한 인터페이스다. 그림 8은 수정 인터페이스로 버튼 드래그를 통해 구간을 분할, 합병해주는 인터페이스다.

5. 결론 및 향후 연구

본 논문에서는 기존의 휴대용 개인 단말기의 그리드 뷰 형식의 문제점을 극복하고자 다양한 검색 및 브라우징 방식을 제안하였다. 본 논문의 방법은 사용자의 선호도에 맞는 방법으로 이벤트 기반으로 자동 분류를 하였으며 객체 인식을 통해 사진에 객체를 자동으로 태그 해주었다. 또한 위치 기반과 객체 기반의 탐색을 제공 하여 사용자에게 편의를 제공하였다. 하지만 본 논문에서는 분류된 이벤트 내에서 대표 사진을 보여주는 방법은 단순한 리스트형태이다. 향후엔 이 대표 사진을 선택하는 방법과 보다 직관적인 사진 미리 보기 방법을 연구할 계획이다.

6. 참고 문헌

[1] Lim J.H., Tian Q., Mulhem P., "Home photo content modeling for personalized event-based retrieval," *ACM Transaction on Multimedia Computing*, vol. 2, no.1, pp.1-19, 2006.

[2] PLATT J., CZERWINSKI M., AND FIELD B., "Simplifying the management of large photo collections," *Proceeding of IEEE Pacific Rim Conference on Multimedia*, pp. 6-10, 2003.

[3] Cooper M., Foote J., Girgensohn A., Wilcox L., "Temporal event clustering for digital photo collections," *Proceeding of ACM International Conference*, vol. 1, no. 3, pp. 269-288, 2005.

[4] Liangliang Cao., Jiebo Luo., Kautz H., Huang T.S., "Image Annotation Within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models," *IEEE Transaction on Multimedia*, vol 11, no. 2, pp. 208-219, 2009.

[5] Jeongsoo Yu., Jongho Nang., "Automatic Event Clustering Method for Personal Photo Collection on Mobile Phone," *Journal of KIISE : Computing Practices and Letters*, vol. 16, no. 12, pp.1141~1273, 2010.

[6] J. Yang., Y. G. Jiang, A. G. Hauptmann, and C. W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," *Proceeding of ACM MIR*, 2007.

[7] D. Lowe., "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp.99~110. 2004.

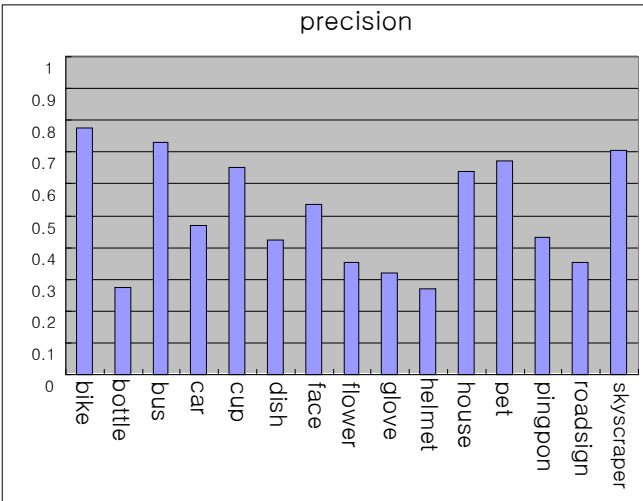


그림 5 기본 BoW 적용

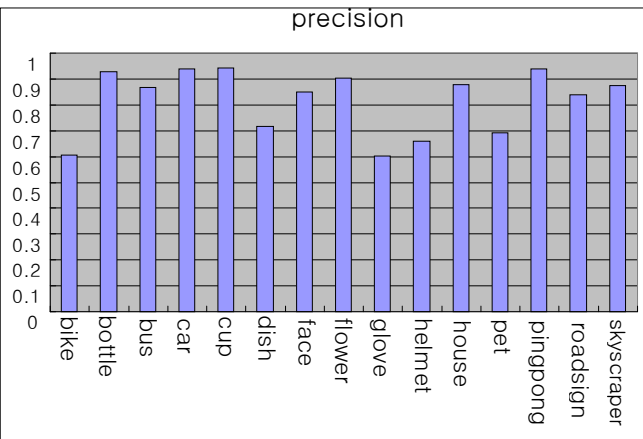


그림 6 TF-IDF, Bi-gram을 적용



그림 7 브라우징 인터페이스

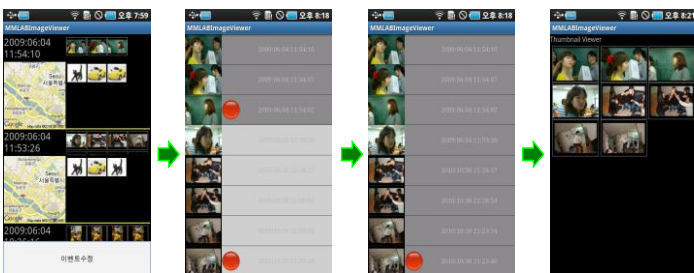


그림 8 이벤트 수정 인터페이스