

휴대 단말용 사진 자동 분류 시스템을 위한 특징 기술자 조합에 대한 연구

김대윤[○] 송인선² 낭종호³

서강대학교

uniyoun4@naver.com, insun0813@naver.com, jhnang@sogang.ac.kr

Categorization of Personal Photo Collection in Mobile Device: Feature Selection Research

Daeyoun Kim[○] Insun Song² Jongho Nang³

Department of Computer Science and Engineering, Sogang University

요 약

최근 스마트 기기의 보급과 이러한 스마트 기기 내부의 저장 공간이 대용량화에 따라 사용자는 휴대 기기에서 촬영한 사진들을 대량으로 저장하고 관리 할 수 있게 되었다. 이에 따라 휴대기기는 사용자가 간편하고 빠르게 사진을 검색할 수 있도록 사진들을 자동 분류하는 기능이 필요하게 된다. 본 논문에서는 대량의 휴대 단말용 사진들을 자동 분류하는 시스템의 분류 정확도를 높이기 위해 진행한 카테고리 별 특징 기술자 분석 실험 및 결과에 대해 기술한다. 특히, 휴대 단말용 사진 도메인에서 추론할 수 있는 한정된 카테고리들로부터, 높은 분류 성능을 보이는 특징 기술자들에 대한 다양한 조합 실험을 진행하여 카테고리 별로 71~95%의 분류기 성능을 확인하였다. 본 논문은 연구자가 휴대 단말용 사진 분류 시스템을 구현 시, 각 카테고리 별로 휴대 단말용 사진 분류에 적합한 특징 기술자 조합을 선택하는 데 큰 지표가 될 것이라고 예상한다.

1. 서 론

최근 스마트 기기의 보급과 이러한 스마트 기기 내부의 저장 공간이 대용량화 됨에 따라 사용자는 휴대 기기에서 촬영한 사진들을 대량으로 저장하고 관리할 수 있게 되었다. 이에 따라 휴대기기는 사용자가 직접 사진 태깅을 할 수 있는 인터페이스를 제공하거나, 자동으로 사진을 분류하는 기능을 필요로 한다. 그러나 사용자가 수동으로 태깅을 하여 사진을 관리 하는 경우는 노동집약적인 작업이 불가피하므로, 휴대기기는 사용자가 간편하고 빠르게 사진을 검색할 수 있도록 휴대 단말용 사진들을 자동 분류하는 기능이 필요하다.

휴대 단말용 사진을 자동으로 분류하는 가장 기본적인 방법은 타임라인이나 GPS를 이용한 장소 기반의 클러스터링이다[1][2]. 그러나, 일반적으로 사용자는 휴대 단말용 사진을 검색할 때 사물이나 장소 위주의 연상 작용을 사용한다. 이러한 객체 기반의 휴대 단말용 사진 자동 분류를 위해서는 분류하고자 하는 카테고리를 정의하고, 각 카테고리 별로 적합한 객체 인식 기술을 적용하여 분류기를 생성하게 된다. 이 때 휴대 단말용 사진 분류기는 미리 정의한 카테고리 별로 어떠한 객체 인식 기술을 적용하느냐에 따라 성능이 좌우된다. 또한, 휴대 단말용 사진이라는 분류기의 도메인을 고려할 때, 객체 인식뿐만 아니라 설경 같은 풍경 카테고리를 분류하기 위한 적절한 특징 기술자 사용이 필요하다.

본 논문에서는 휴대 단말용 사진 자동 분류를 위한 카테고리 분류 시스템의 분류 정확도에 영향을 미치는

카테고리 별 특징 기술자들에 대한 성능 분석 실험을 진행하였다. 카테고리 별로 최근 연구 되고 있는 다양한 객체 인식 기술들과 MPEG-7에서 명시한 시각 특징 기술자 표준을 적용하여 최적의 특징 기술자 조합을 찾음으로써 휴대 단말용 사진 자동 분류 시스템의 분류 성능을 분석 비교하였다.

본 논문의 2장에서는 사진 카테고리 분류 시스템과 관련된 연구들을 소개하고, 3장에서는 실험을 위한 시스템 구조에 대해서 상세히 설명한다. 4장에서는 진행한 실험 결과에 대한 분석을 설명하며, 5장에서는 결론 및 향후 연구로 논문을 마무리한다.

2. 관련 연구

일반적으로 사용자는 개인용 사진 컬렉션에 대하여 이벤트 기준으로 구성하고 접근하는 것을 선호한다. 이러한 이벤트 기반 사진 자동 분류 시스템을 제안하는 모델들은 이벤트 내에 존재하는 독립적인 객체 인식 방법과 비교하여 인식률이 더 낮다[3]. 또한, 객체 인식 기법을 적용하여 이벤트 경계 검출 성능을 높이기 위해 [4]에서는 BoW(Bag of Words)를 사용하였고, [5]에서 제안한 분류 기법에도 객체 인식 기법으로 BoW를 적용한다. 그러나, 이러한 BoW기법을 이용한 분류는 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.

3. 휴대 단말용 사진 자동 분류 시스템

본 논문에서 구현한 휴대 단말용 사진 자동 분류 시스템 구조도는 그림 1 과 같다. 먼저, 본 논문에서 설정한 자동 분류 시스템의 도메인은 휴대 단말용 사진이기 때문에 4.1절에서 명시한 바와 같이 8개의 카테고리를 미리 정의하였다. 학습 데이터는 카테고리 별로 긍정과 부정 이미지로 구성하고, 테스트 데이터는 무작위의 새로운 이미지들을 포함하여 앞에서 언급한 8개 카테고리의 학습 데이터에 포함되지 않는 새로운 이미지들을 조합하여 구성하였다.

N개의 카테고리 학습 데이터가 실험 툴의 입력으로 들어가면, M개의 특징 기술 모듈을 거쳐 학습기로 들어가게 되고, 그 결과로 총 $N \times M$ 개의 분류기가 학습 된다. 그 후 테스트 데이터가 특징 기술 모듈을 거치면 $N \times M$ 개의 분류기를 통해 최적의 성능을 보이는 특징 기술자가 선택된다. 즉, 카테고리 별로 최적의 성능을 보이는 특징 기술자가 선택되어 최적의 분류기가 생성되는 것이다. 휴대 단말용 사진 자동 분류기는 off-line으로 생성된 카테고리 별 최적의 분류기를 이용하여 분류기 모듈을 수행하고 최종 결과로 휴대 단말용 사진들의 카테고리를 분류하게 된다.

category#1을 검출하기 위한 CLC(Consistent Line Cluster)이다. 이 특징 기술자는 건물 객체가 가지는 직선 성분들을 색 조합 정보, 소실점(vanishing point) 존재 여부 그리고 이미지 도메인에서의 직선 평행성을 이용하여 clustering하고, 이 cluster들간의 inter/intra relationship으로 특징 기술자를 정의한다.

4.1절의 category#3에 해당하는 특징 기술자로는 Haar-like feature를 사용하였다. 또한 category#4에 해당하는 특징 기술자로는 메모 카테고리가 갖는 수직 또는 수평 성분의 많은 빈도수를 가지고 새로이 정의하였다. Hough Transform을 이용하여 직선 성분을 검출한 후, 검출된 직선이 수직 또는 수평인 빈도수를 가지고 확률 값을 특징 기술자로 정의하였다.

객체를 포함하지 않는 풍경 이미지들(category#5,6,8) 분류를 위해서는 MPEG-7의 시각 특징 기술자 중 CSD(Color Structure Descriptor), CLD(Color Layout Descriptor), DCD(Dominant Color Descriptor) 그리고 EHD(Edge Histogram Descriptor)를 사용하였다. 위의 4가지 특징 기술자는 global feature로서 사용자의 전역적 분석은 사람의 시각 인지 과정과 상이한 한계를 보이므로 category#4를 위해 정의한 특징 기술자와 조합하여 총 26개($=2^{(1+4)} - (1+4) - 1$)의 특징 기술자를 추가로 정의하였다.

3.2 학습 모듈

특징 기술자의 학습에 가장 널리 쓰이는 학습기는 SVM(Support Vector Machine)이다. SVM은 회귀(Regression)과 지도 학습(Supervised Learning)에 이용될 수 있는 분류기의 일종으로, 입력 벡터 x 를 고차원의 특징 공간으로 매핑시킨 후 두 클래스 사이의 마진(margin)을 최대화 시키는 분리 경계면을 찾는 것을 목적으로 한다.

본 논문에서 사용한 또 다른 학습기는 바로 cascade classifier이다. Cascade classifier는 약한 분류기들을 boosting method를 적용하여 강한 분류기로 변형시켜 학습하는 방법이다. 특히 이 학습기는 얼굴 인식에 강건한 특징 기술자인 Haar-like feature와 함께 널리 사용되므로 셀프 카메라 카테고리에 대해서는 cascade classifier 학습기를 사용하였다.

위에서 언급했듯이, 본 논문에서는 셀프 카메라 카테고리 분류를 위한 Haar-like feature를 제외한 나머지 특징 기술자들에 대해서는 SVM을 사용하였다. 특히, SVM에서는 특징 벡터의 고차원에의 매핑을 위해 kernel function이 적용된다. 본 논문에서는 다양한 kernel function 중 RBF kernel을 사용하였고, learning module에서 최적의 kernel function 파라미터 조합은 [5]에서 보여주는 실험 결과를 토대로 일정 범위 내의 파라미터 조합을 실험하여 최적의 결과를 파라미터 수치로 설정하였다.

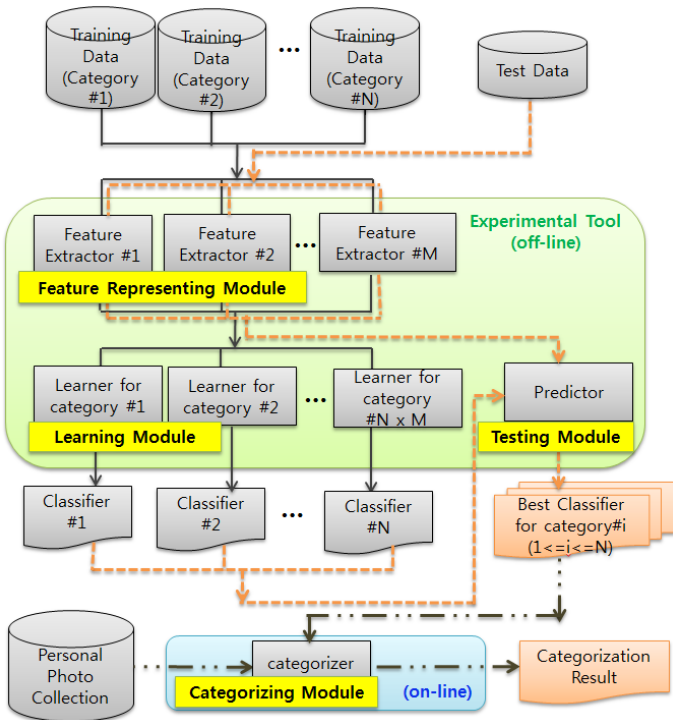


그림 1 - 휴대 단말용 사진 자동 분류 시스템 구조도

3.1 특징 기술자 구현 모듈

특징 기술자는 미리 정의된 각 카테고리의 인식 분별력을 높이는 기술자들을 도입 및 구현하였다.

실험에 도입한 특징 기술자들로는 먼저, 4.1절의

4. 실험 및 결과 분석

4.1 실험 데이터

본 논문에서 정의한 카테고리과 학습, 테스트 데이터로 수집한 이미지들의 수는 표 1 과 같다.

| | 카테고리 명 | 학습 데이터 수 | | 테스트 데이터 수 |
|------------|--------|----------|----------|-----------|
| | | 긍정 이미지 수 | 부정 이미지 수 | |
| Category#1 | 건물 | 320 | 980 | 200 |
| Category#1 | 음식 | 380 | 1100 | 210 |
| Category#3 | 셀프 카메라 | 270 | 800 | 180 |
| Category#4 | 메모 | 300 | 850 | 160 |
| Category#5 | 야경 | 250 | 780 | 180 |
| Category#6 | 설경 | 260 | 780 | 185 |
| Category#7 | 경기장 | 300 | 810 | 200 |
| Category#8 | 물가 | 300 | 900 | 180 |

표 1 - 카테고리과 데이터 집합

4.2 실험 결과 및 분석

| 카테고리 명 | 선택된 특징 기술자 | Precision | Recall | F1 |
|--------|---------------|-----------|----------|----------|
| 건물 | CLC | 0.884615 | 0.92 | 0.901961 |
| 음식 | CSD+DCD | 0.78022 | 0.865854 | 0.820809 |
| 셀프 카메라 | Haar-like | 0.968032 | 0.93333 | 0.953634 |
| 메모 | Hough+CSD+CLD | 0.824561 | 0.817391 | 0.820961 |
| 야경 | DCD+CSD+CLD | 0.657143 | 0.851852 | 0.741935 |
| 설경 | CSD+CLD | 0.643312 | 0.834711 | 0.726619 |
| 경기장 | DCD+CLD+EHD | 0.666667 | 0.775862 | 0.717131 |
| 물가 | CSD+CLD | 0.770492 | 0.859756 | 0.81268 |

표 2 - 실험 결과

본 논문에서 선택한 특징 기술자 후보는 3.1절에서 소개한 바와 같이 총 32가지이다. 미리 정의한 카테고리 수는 4.1절에서와 같이 8개 이므로, learning phase에서 생성된 후보 분류기는 총 256개가 되고, 각 분류기의 성능을 측정하는 실험을 통하여 결과를 정리하였다. 표 2와 같이 실험 결과는 공간 제약의 이유로 카테고리 별로 가장 분류 성능이 높은 특징 기술자만을 선택하여 작성하였다. 이 때, 표 2의 F1은 분류 성능을 측정하는 척도로서, 아래와 같은 식으로 정의된다.

$$F1 = Precision \times Recall \times 2 / (Precision + Recall)$$

표 2의 실험 결과에서 알 수 있듯이, 객체 인식 기술을 이용하여 90%이상의 높은 분류 성능을 보이는 카테고리는 건물과 셀프 카메라다. 두 카테고리 모두 카테고리를 대표하는 객체의 특징이 명확하기 때문인데, 표 2에는 명시하지 않았지만, 셀프 카메라 같은 경우는 OpenCV library에서 제공하는 cascade classifier를 이용하였을 때보다 7~8%정도 분류 성능 향상을 보였다.

휴대 단말용 사진이라는 도메인 안에서 셀프 카메라 카테고리는 일반 얼굴 인식용 데이터와 다르게 촬영 기기의 촬영 각도, 측면으로 치우친 얼굴의 회전 각도 때문에 OpenCV의 cascade classifier에는 적합하지 않기 때문이다.

5. 결론 및 향후 연구

일반적으로 본 논문에서 진행한 실험 결과의 핵심은 높은 인식 성능을 보이는 특징 기술자 조합을 찾음으로써 분류기 전체 성능을 높이는 데 있다. 그러나, 이번 실험에서 적용한 몇몇의 특징 기술자들은 휴대 단말에 적용하기에는 다소 높은 시간 복잡도를 가진다. 이를 해결하기 위해 최근 활발하게 연구되고 있는 CUDA와 같은 가속화 프로그래밍 기술을 이용하여 최적화 하는 방안들을 향후 연구 과제로 남겨둔다.

6. 참고문헌

[1] 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.

[2] Liangliang C., Jiebo L., Kautz H. and 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.

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

[4] SangChul K. and JongHo Nang., "An Event-based Clustering and Browsing of Personal Photo Collections on Mobile Device," *Journal of KIISE : Computer Systems and Theory*, Vol. 39 , No. 1, pp.59-65, 2012. (in Korean)

[5] 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.

[6] Liangliang Cao., Jiebo Luo., Kautz H. and 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.

[7] Yi Li. and L.G. Shapiro., "Consistent Line Clusters for Building Recognition in CBIR", *Proceedings of the International Conference on Pattern Recognition.*, 2002.