

객체 인식을 위한 SIFT 템플릿 매칭 방법

Part I: 클러스터링 기반 SIFT 템플릿 생성 방법

김상철^o 김소현, 김유진, 남종호
서강대학교 컴퓨터공학과

smaslayer1@nate.com, thgus4013@nate.com, jean703@nate.com, jhnang@sogang.ac.kr

SIFT Template Matching for Object Recognition

Part I: Clustering Based SIFT-Template Generation

Sangchul Kim^o Sohyun Kim, Yoojin Kim, Jongho Nang
Dept. of Computer Science and Engineering, Sogang University

요 약

최근 카메라를 장착한 하드웨어 기기의 보급으로 영상에서의 객체 인식을 기반으로 한 응용이 늘어나고 있다. 매칭 기반 객체 인식 방법인 SIFT(Scale Invariant Feature Transform)의 경우 1:1 매칭 기반의 방법으로써 비교하고자 하는 대상이 완전히 같은 것이 아닐 경우에 객체의 인식률이 현저히 나빠지는 단점이 존재한다. 이를 해결하기 위해서 해당 객체를 대표하는 템플릿 피처를 준비한다면 객체의 인식률이 좋아질 것이다. 이러한 템플릿 피처를 제공하기 위하여 샘플 이미지의 피처를 클러스터링하여 대표 클러스터를 템플릿 피처로 선정한다. 이 때 정확도를 높이기 위하여 TF/IDF 방법을 적용하였으며 인식을 위한 Adaptive Threshold를 적용한 방법을 제안하였다. 실험을 통하여 단순한 SIFT 매칭 방법이나 BoVW방법에 비하여 interclass similarity가 낮은 객체들의 경우 성능이 높게 나옴을 알 수 있었다.

1. 서 론

블랙박스나 핸드폰처럼 카메라가 내장된 하드웨어 기기의 보급이 늘어남에 따라 영상처리를 통한 응용의 요구가 늘어나고 있다. 예를 들면, 자동차의 자동 주행 중 여러 상황을 인지하기 위한 표지판 인식방법[1]이나 실제 이미지 위에 증강 현실을 구현할 때 마커 대신 실제 영상안에 인식된 객체에 적용하는 방법[2]등에서 가장 중요한 이슈는 영상 내의 객체 인식 방법이다.

최근 영상 내의 객체를 인식 하는 방법으로 가장 각광받고 있는 방법은 SIFT [3]이다. 하지만 이러한 SIFT는 1:1 매칭 기반의 방법으로써 영상에서 찾고자 하는 객체 외의 영역, 예를 들면 배경 등에서 추출된 피처가 객체 인식 성능 저하를 일으키는 원인이 된다. 또한, 밝기 변화나 색상, 크기 변화에 의해서도 약간의 성능이 저하된다. 따라서, 이러한 성능 저하에 미치는 요인들을 제외할 수 있는 대표 피처들을 수집하여 이를 매칭할 때에 사용한다면 보다 강건한 객체 인식 성능을 얻을 수 있을 것이다.

본 논문에서는 대표 피처 DB를 구축하기 위해 성능 저하를 시키는 요소들을 제외하기 위하여 TF/IDF 방법을 적용하여 피처들을 추출하였고, 이러한 피처들을 수집하여 각각 이미지 별로 클러스터링을 하여 템플릿 피처를 생성하였다. 최종적으로 이 템플릿 피처를 해당 이미지의 피처 DB 로 구축하였다. 제안한 방법으로 17개의 객체에 대하여 인식 실험을 한 결과 기존 SIFT나 BoVW기반 객체 인식 방법에

비하여 성능이 좋음을 보였다. 본 논문의 구성은 다음과 같다. 2장에서 관련된 연구에 대해 소개하고 3장에서 본 논문에서 제안한 객체 인식 방법에 대해 설명하며 4장에서 실험 및 분석 결과에 대해 소개하고, 5장에서 결론 및 향후 연구로 맺는다.

2. 관련 연구

2.1 영상 객체 인식

최근 이미지 객체 인식 분야는 이미지 전체를 고려하는 방법과 국소 부위를 고려한 방법들이 연구되고 있다[4]. 이 중 국소 부위를 고려한 방법은 국소 영역에 대하여 특징을 정의하고 분별력 있는 국소 영역에 대해서만 매칭 시에 비교하기 때문에 영상 전체를 추상화하여 하나의 피처로 정의하는 방법에 비하여 정확도가 높은 장점이 있다.

2.2 BoVW(Bag of Visual Word)

Bag of Word 방법은 자연어 처리에서의 텍스트 자동 분류 방법으로서, 이미지 분야에서는 이미지 한 장을 하나의 문서, 이미지의 로컬 패치(local patch)를 하나의 단어(Visual Word)로 생각하여 이미지를 분류하는 방법이다[6]. 로컬 패치들의 피처들을 클러스터링 하여 사전(Codebook)을 생성한 뒤 객체 인식 단계에서 로컬 패치들의 피처의 빈도수, 즉 사전에서 얻을 수 있는 단어(Codeword)들이 얼마나 자주 출현하는가에 따라 객체를 인식하는 방법이다. 이미지

객체인식에서 전역적 특징이 가진 한계를 극복하기 위하여 국소 영역의 특징을 이용하여 객체의 인식률을 높였다.

3. 클러스터 기반 SIFT 템플릿 매칭

그림 1 은 제안한 전체 방법의 구조도 이다. 먼저 각 객체별 이미지 set 을 모은 뒤 이미지별로 SIFT 피처를 추출하여 객체에서만 존재하는 피처만 남기기 위하여 TF/IDF like Outlier Removal 방법을 적용한다. 그 후에 남은 피처들로 K-means Clustering 을 하여 클러스터의 대표값들을 해당 객체를 대표하는 SIFT-Template 피처로 정의하여 DB 로 구축한다. Image Query 가 들어오면 DB(SIFT-Template Features DB)에 존재하는 각각의 SIFT Template Feature 와 매칭하여 점수가 가장 높은 객체를 결과로 하는 방법이다.

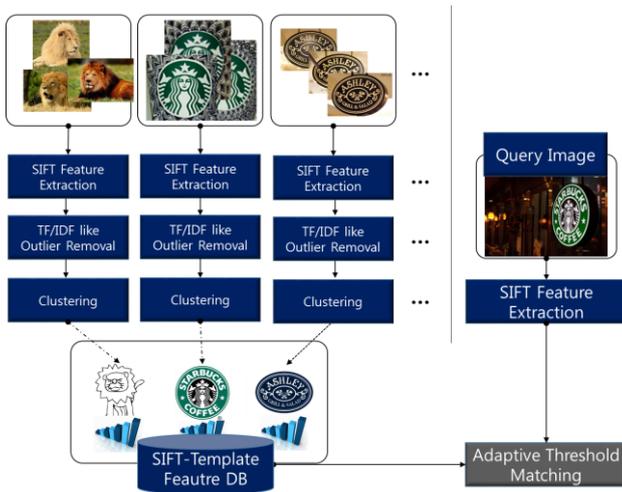


그림 1. Cluster based SIFT-Template Matching 구조도

3.1 TF/IDF 적용 오정합점 제거



그림 2. 배경이 클러스터 결과에 영향을 미치는 예시

그림 2 는 배경이 클러스터 결과에 영향을 미치는 예시를 나타내고 있다. 객체 별 이미지 set 에서 추출된 SIFT 피처는 배경에서도 뽑힐 수 있다. 배경이 유사한 경우에 클러스터 대표값이 어떤 객체를 대표하는지 알 수 없기 때문에 객체 인식 성능 저하의 영향을 미친다. 이러한 배경에서 추출된 피처를 언어모델의 IDF 로 보고 객체에서 추출된 피처를 TF 로 가정할 수 있다. 따라서 클러스터링 과정을 거치기 전에 배경에서 뽑히는 특징점을 제외하면 클러스터의 성능을

향상시킬 수 있다. 모든 객체별 이미지 set 은 객체는 갖지만 배경이 다양한 경우이기 때문에 두개 이상의 이미지를 SIFT matching 을 수행 하여 매칭된 SIFT 피처만 이용할 경우에 객체에서 추출된 SIFT 피처라고 할 수 있다. 하지만 잘못 매칭된 SIFT 피처는 객체에서 추출된 SIFT 피처라고 신뢰할 수 없기 때문에 Outlier 피처 를 제거 한 뒤에 클러스터링을 수행하게 해야 한다. SIFT Matching 이후에 추출된 피처들 중 Outlier 피처를 제거하기 위하여 SIFT 피처의 방향성을 고려한 통계적 기법을 적용하여 제거하는 방법[1]을 적용하였다. 이렇게 추출된 피처를 K-means 클러스터링을 하여 대표 피처를 결정하였다.

3.2 SIFT 템플릿 피처 DB 매칭

$$ResultObject = \arg \min_i (Similarity_{Obj_i}) \quad (1)$$

$$Similarity_{Obj_i} = \sum_{k=0}^n dist(matches_{i_k}) / n \quad (2)$$

식 (1)은 Query Image 에 대해 객체 인식 결과를 나타내기 위한 식이다. $Similarity_{Obj_i}$ 의 값이 가장 작은 객체가 Query Image 의 객체를 나타낸다. 식 (2)는 $Similarity_{Obj_i}$ 를 구하기 위한 식으로 $matches_{i_k}$ 는 SIFT-Template DB 의 i 번째 피처와 쿼리 이미지와 매칭된 n 개 중 k 번째 피처를 나타낸다. 매칭된 k 번째 피처간의 거리를 모두 합한 것을 $Similarity_{Obj_i}$ 로 정의한다.

4. 실험 결과 및 분석

4.1 실험 환경

제안한 방법을 테스트하기 위하여 표 1 과 같이 4 개의 카테고리 총 17 개의 객체에 대하여 각각 약 50 장씩의 테스트 이미지를 준비하여 SIFT 템플릿 DB 를 구축하였다.

표 1. SIFT 템플릿 피처 DB 구축을 위한 객체 종류

카테고리	객체
브랜드로고	스타벅스, 애슬리, 엔젤리너스, 맥도날드, 파스구찌
교통표지판	주차금지, 공사중, 정지, 진입금지, 철도건널목
사물	자전거, 오토바이, 음료수병, 사과
동물	사자, 호랑이, 개, 고양이

제안한 방법의 성능을 비교하기 위하여 기존의 SIFT 매칭 방법[3]과 BoVW 를 이용한 객체 인식 방법[]에 대하여 17 개의 객체를 인식하여 Recall 과 Precision 을 평가하였다. SIFT 매칭 방법의 경우 템플릿 이미지 한 개를 이용하여 여러 개의 사진에 대해 매칭하는 방법으로 실험을 진행하였다.

4.2 성능 분석

그림 3, 4는 제안한 방법, BoVW 그리고 단순 SIFT 매칭에 대한 객체 인식의 Recall 과 Precision 그래프이다. SIFT 템플릿 이미지를 사용할 경우에 빛변화나 뷰포인트 변화에 의하여 차이가 나는 BoVW 의 경우 코드북을 생성할 때에 고차원 공간상에 나타나는 차원의 저주문제 때문에 클러스터링으로 수집되는 코드워드 자체의 신뢰성을 보장할 수 없다. 또한, 모든 훈련 이미지 셀에서 SIFT Feature 를 추출하기 때문에 3.1 절에서 설명한 것과 같이 객체를 대표하지 않는 Feature 들이 코드워드가 될 수 있고 이러한 코드워드들의 워드 빈도수에 의해서 결과가 다르게 나올 수 있는 단점이 있다. 하지만 제안한 논문에서는 각 객체의 템플릿들을 각각 생성해내는 과정에서 객체를 대표하는 템플릿 피처를 얻을 수 있다. 또한 이 때문에 BoVW 방법에 비해 높은 성능을 얻을 수 있다.

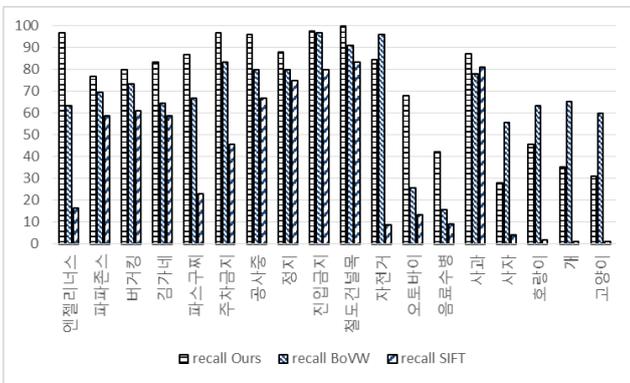


그림 3. 객체 인식 방법에 대한 Recall 그래프

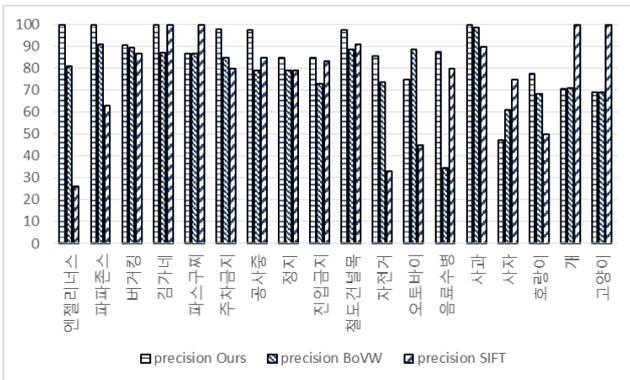


그림 4. 객체 인식 방법에 대한 Precision 그래프

객체들의 inter class similarity 가 작은 브랜드로거나 교통표지판의 경우 BoVW 보다 SIFT 매칭과 제안한 방법의 성능이 매우 높게 나왔는데 이는 위에 언급한 내용처럼 BoVW 에서 객체를 대표하지 않는 코드워드들에 의하여 분별력이 저하되는 특징으로 나타난다. 또한 단순 SIFT 매칭에 비하여 제안한 방법의 성능이 좋게 나오는 이유는 첫째로 매칭하는 템플릿 이미지 자체와 뷰 포인트나 밝기 변화가 매우 큰 이미지의 경우 1:1 매칭에 적합한 SIFT 가 유사객체 매칭에 약한 특징 때문이다. 반면, 제안한 방법의 경우 이런 모든 경우를 고려하여 추출된 템플릿 피처들을 매칭함으로써 이에 대한 단점을 보완하기 때문에 성능이 높게 나오는 것을 알 수 있다. 하지만 사물이나 동물의 경우

BoVW 에 비해 성능이 낮음을 알 수 있는데 이는 제안한 방법이 매칭 기반의 인식 방법으로서 inter class similarity 가 낮은 객체들의 경우 다양한 사진에서 서로 매칭되는 SIFT Feature 가 부족하여 클러스터링 된 템플릿 피처들 자체의 신뢰성이 현저히 떨어지는 것으로 분석되었다.

5. 결론 및 향후 연구

본 논문에서는 영상에서 효과적으로 객체 인식을 위한 SIFT 매칭 방법을 위하여 Cluster based SIFT-Template 을 생성하는 기법을 제안하였다. 본 논문의 방법은 SIFT 의 장점인 1:1 매칭 뿐만 아니라 객체 인식에도 적합하도록 클러스터링으로 대표 피처를 Template 으로 두어 인식률을 높였다. 또한 클러스터링의 대표값의 신뢰도를 높이기 위하여 TF/IDF 아이디어를 적용하여서 전반적인 성능이 높아짐을 알 수 있었다. 하지만 매칭 기반의 SIFT Feature 가 갖는 한계점 때문에 나타나는 성능 저하를 보완하기 위한 전처리 기법과 후처리 기법에 대한 연구가 필요하다. 본 논문의 Part II 에서는 매칭시의 성능을 높이기 위한 Adaptive Threshold Matching 방법을 제안하고 inter class similarity 가 작은 브랜드 로고 인식에 대한 성능을 분석한다.

6. 참고 문헌

- [1] S. C. Kim, J. M. Lee, D. Y. Kim, J. H. Nang, "Applying SIFT Feature to Occlusion, Damage and Rotation Invariant Traffic Sign Recognition," in *Proc. of the KCC 2012*, Vol. 39, No. 1, pp. 351-353, 2012 (in Korean).
- [2] D.W. Kim, D.S. Hwang, "Non-Marker based Mobile Augmented Reality and its Applications using Object Recognition," *Journal of Universal Computer Science*, Vol. 18, No. 20, pp. 2832-2850, 2012.
- [3] Lowe, D. G., "Distinctive Image Feature from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol.60, No.2, pp. 99-110, 2004.
- [4] S. C. Kim, J.H. 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] Mikolajczyk, K., Schmid, C., "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 27, pp. 1615-1630, 2005.
- [6] H. J. Jung, J. M. Lee, J. H. Nang, "Image Categorization Using SIFT Bag of Word," in *Proc. of KCC 2013*, pp.1277-1279, 2013 (in Korean).