

객체 인식을 위한 SIFT 템플릿 매칭 방법

Part II: 브랜드 로고 인식 성능 분석¹⁾

김소현^o 김유진 김상철 유재석 낭종호
서강대학교 컴퓨터공학과

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

SIFT Template Matching for Object Recognition

Part II: Analysis Performance of Brand Logo Recognition

Sohyun Kim^o Yoojin Kim Sangchul Kim Jaesug Ryu Jongho Nang

Dept. of Computer Science and Engineering, Sogang University

요 약

본 논문의 Part I에서 제안한 SIFT-Template 피쳐의 인식 성능을 높이기 위하여 Adaptive Threshold Matching을 제안하였고, 이에 대한 성능을 분석하기 위하여 inter class similarity가 적은 브랜드 로고들에 대하여 실험하였다. 기존의 SIFT 매칭 방법에 비하여 피쳐를 클러스터링 하여 대표 클러스터를 Template 피쳐로 선정할 경우와 이 때 정확도를 높이기 위하여 TF/IDF 방법을 적용한 경우, 그리고 인식을 위한 Adaptive Threshold를 적용한 경우에 대하여 분석하였다. 실험을 통하여 단순히 클러스터링만을 하여 객체를 매칭하여 인식할 때에 비해 유사한 수준의 Recall과 5%의 Precision 상승폭을 보였다.

1. 서 론

최근 명함, 광고지 등에 QR코드를 많이 활용하고 있다. 예를 들어 QR코드를 광고지에 삽입하여 해당 웹 사이트로 이동할 수 있고, 명함의 경우 휴대폰에 개인정보를 더 쉽게 저장할 수 있다. 하지만 QR코드는 비슷한 모양과 무늬를 가지고 있기 때문에 알아보기 힘들고, 오히려 사용자에게 익숙하지 않다. 따라서 QR코드를 보다 친숙한 이미지로 대체할 수 있다면 긍정적 효과를 야기할 수 있다. 예를 들어 길을 걷다가 간판을 보고 브랜드 로고를 촬영할 수 있고, 머그잔에 그려진 로고를 촬영할 수 있다. 이로써 사용자에게 로고를 직접 찍게 하여 손쉬운 검색이 가능하게 하고 각인의 효과를 기대할 수 있다. 따라서 QR코드를 이미지로 대체할 수 있다면 사용자에게 편리함을 광고주에게 광고의 효과를 극대화 할 수 있다. 바코드가 아닌 이미지로 대체하기 위해서는 객체 인식 기술이 필요하다[1]. 객체 인식 기술에는 매칭과 학습의 방법이 있다. 매칭은 정형화된 이미지를 찾는 방법으로 단순하고 신속한 특징이 있다. 학습은 비슷한 모양의 이미지를 찾는 방법으로 샘플을 수집하는데 신속성이 떨어지고 학습기의 성능에 따라 성능이 좌우되는 특징이 있다. 학습은 그 모양이 비슷하지만 같지 않은 경우, 즉 사람의 얼굴을 찾는 데에 이용된다[2]. 브랜드 로고는 정형화된 이미지이므로 매칭 방법을 사용하는 것이 적절하다[3]. 따라서, 이미지의 크기, 회전 등에 강건한 SIFT(Scale Invariant

Feature Transform)[4] 매칭 방법을 이용한다. 하지만 이러한 SIFT는 1:1 매칭 기반의 방법으로서 찾고자 하는 객체 외의 영역, 예를 들면 배경 등에서 추출된 피쳐가 객체 인식 성능 저하를 일으키는 원인이 된다[5]. 또한, 밝기 변화나 색상, 크기 변화에 의해서도 약간의 성능이 저하된다[6]. 따라서 이러한 성능 저하에 미치는 요인들을 제외할 수 있는 대표 피쳐들을 수집하여 이를 매칭할 때에 사용한다면 보다 강건한 객체 인식 성능을 얻을 수 있을 것이다.

본 논문에서는 성능 저하를 시키는 요소들을 제외하기 위하여 TF/IDF 방법을 적용한 뒤 객체를 인식하는 Adaptive Thresholding Matching 방법을 제안하였다. 본 논문의 Part I 방법과 본 논문에서 제안한 방법으로 15개의 브랜드로고 인식실험을 한 결과 기존 클러스터링만을 사용하여 피쳐 DB를 구축하는 경우보다 TF/IDF 방법을 적용하였을 경우 높은 Precision을 나타냄을 보였다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 소개 하고 3장에서 본 논문에서 제안한 객체 인식 방법에 대해 설명하며 4장에서는 실험 및 분석 결과에 대해 소개하고, 5장에서 결론 및 향후 연구로 맺는다.

2. 관련 연구

2.1 SIFT-Template Matching

본 논문의 Part 1에서 제안한 SIFT-Template Matching은 그림 1과 같이 각 개체별로 훈련 이미지 셋의 Feature를 클러

1) 본 연구는 미래창조과학부 및 한국산업기술평가관리원의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음. [10044615, 클라우드 기반 개방형 소셜 방송미디어 콘텐츠 융합 생성, 편집 및 재생을 위한 미디어 제작 및 전송 시스템 개발]

스터링 하여 객체를 대표하는 템플릿 피쳐들을 구한 뒤, 이를 매칭 기반의 방법으로 인식하는 방법이다.

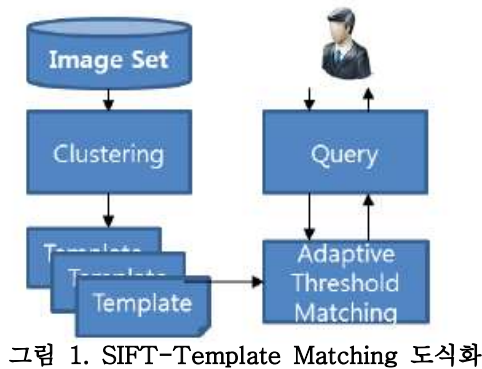


그림 1. SIFT-Template Matching 도식화

3. Adaptive Threshold Matching

본 논문에서는 DB와 Query Image 의 SIFT 피쳐간 매칭을 통하여 객체를 판별한다.

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

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

$$\begin{cases} \text{if}(dis(matches_{i_k}) > \alpha_i) \\ \text{then } dis(matches_{i_k}) = 0, n = n - 1 \end{cases} \quad (3)$$

식 (1)은 Query Image에 대해 객체 인식 결과를 나타내기 위한 식이다. $Similarity_{Obj_i}$ 의 값이 가장 작은 객체가 Query Image의 객체를 나타낸다. 식 (2)는 $Similarity_{Obj_i}$ 를 구하기 위한 식으로 $matches_{i_k}$ 는 SIFT-Template DB의 i 번째 피쳐와 Query Image 와 매칭된 n 개 중 k 번째 피쳐를 나타낸다. 매칭된 k 번째 피쳐간의 거리를 모두 합한 것을 $Similarity_{Obj_i}$ 로 정의한다. 하지만 SIFT template 피쳐의 개수가 Query Image 에서 추출된 피쳐의 개수보다 작을 때 객체가 아닌 영역에서 추출된 피쳐가 항상 매칭을 시도하게 된다. 이런 경우에 매칭된 피쳐의 실제 Distance는 매우 크기 때문에 이러한 매칭이 일어난 경우에 결과에 영향을 미칠 수 있다. 따라서, 식(3)과 같이 매칭 된 피쳐간 distance가 일정 Threshold α_i 보다 클 경우에는 배제하여 객체 인식 성능을 높인다. 모든 테스트 이미지가 SIFT Template 피쳐와 매칭되는 평균 거리를 Threshold α_i 로 결정했다.

4. 실험 결과 및 데모 시스템

4.1 실험 환경

제안한 방법을 테스트 하기 위해서는 증강현실에 적합한 카테고리인 브랜드 로고를 중심으로 테스트를 수행하였고, 이를 위하여 15 개의 브랜드 로고를 정하여, 클러스터링을 위한 샘플 데이터 약 300 장과 매칭 결과를 분석하기 위한 테스트 이미지 약 750장을 선별하였다.

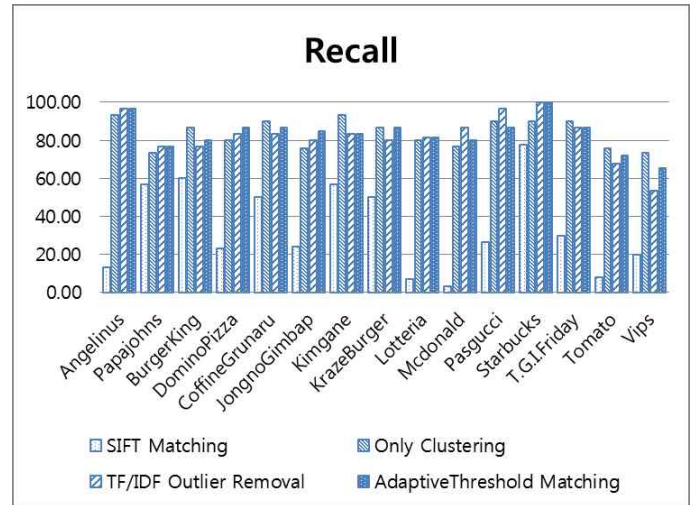


그림 2. 제안한 방법의 Recall 분석 그래프

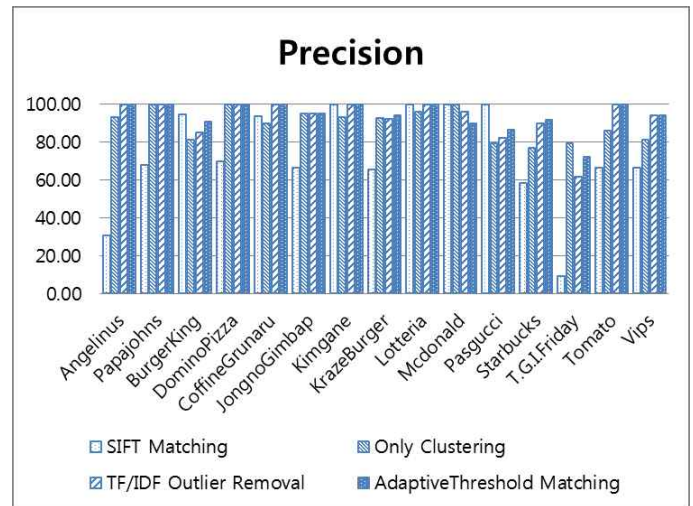


그림 3. 제안한 방법의 Precision 분석 그래프

4.2 Clustering의 효과 분석

그림 2,3는 제안한 방법의 매칭 결과에 대한 Recall과 Precision의 그래프이다. SIFT Matching 항목과 Only Clustering 항목을 비교해보면 기본적으로 클러스터링만 함으로써 Recall/Precision이 높아지는 것을 확인할 수 있다. 그 이유는 여러 이미지의 feature를 모아 clustering 하여 그 대푯값과 비교하였기 때문이다.

하지만 BugerKing, CoffineGrunaru, Kimgane, Lotteria같은 경우는 Precision에서 SIFT 매칭이 더 높다. 우선 Lotteria와 Mcdonald같은 경우는 Recall과 비교해서 보면 알 수 있듯이 워낙 답으로 찾은 결과가 적어 Precision 이 높아진 것으로 보인다. 답으로 찾은 결과가 적은 이유는 브랜드 마크가 매우 간단하기 때문에 뽑힌 피쳐 개수가 적기 때문이다. 또한 BugerKing, CoffineGrunaru, Kimgane 같은 경우에는 임의로 뽑은 test sample 들 중에 실제 간판이 아닌 대표 템플릿과 같은 일반 이미지인 것이 많아 다음과 같은 결과가 나온다.

4.3 TF/IDF 적용의 효과 분석

Only Clustering 항목과 TF/IDF Outlier Removal 항목을 비교해보면 이것 역시 outlier removal을 했을 때 Recall/Precision 이 높아진다. 배경이나 빛의 변화에 의해 잘못 추출된 feature들을 어느 정도 지움으로써 정확도가 더 높아지는 것이다. 하지만 T.G.I Friday와 McDonald 의 경우 단순 클러스터링만 하였을 때 보다 Precision 이 낮아지는 현상이 나타났는데, 이들 브랜드 로고들은 글자로 구성되어 로고 자체에서 추출되는 SIFT 피처가 적어서 Outlier Removal의 의미 있는 피처들이 제외되기 때문에 분석된다. 하지만 전반적으로 기본적으로 클러스터링 방법에 비해 약 0.1%의 Recall 성능 하락과 5.2%의 Precision 상승률을 보여주고 있으며, 성능이 좋지 않은 McDonald와 T.G.I Friday를 제외할 경우 0.01%의 Recall 하락과 7.5%의 Precision 상승률을 보여주고 있다.

4.4 Adaptive Thresholding 매칭 적용의 효과 분석

3장에서 제안한 Adaptive Thresholding 매칭은 TF/IDF에서 완전히 제거 못한 객체에 포함되지 않는 Template Feature들에 의하여 매칭된 Feature간의 거리가 멀게 되는 것을 제외하기 때문에 전반적으로 약간의 성능 향상이 나타나는 것을 볼 수 있다. 하지만 McDonald 처럼 단순한 이미지 형태의 브랜드 로고의 경우 Adaptive Thresholding 매칭에 의하여 다른 객체 중 간판이나 브랜드 로고의 매끈한 부분에 존재하는 곳에 있는 피처들이 포함되어 성능 하락에 영향을 미침을 알 수 있다.

5. 결론 및 향후 연구

본 논문에서는 영상에서 효과적으로 객체 인식을 위하여 Cluster based SIFT-Template matching 기법을 제안하였다. 본 논문의 방법은 SIFT의 장점인 1:1 매칭 뿐만 아니라 객체 인식에도 적합하도록 클러스터링으로 대표 피처를 Template 으로 두어 인식률을 높였다. 또한 클러스터링의 대푯값의 신뢰도를 높이기 위하여 TF/IDF 아이디어를 적용하여서 Precision 을 높임을 알 수 있었다. 하지만 단순한 모양의 객체에서 갖는 SIFT의 한계점 때문에 나타나는 성능 저하를 보완하기 위한 전처리 기법과 후처리 기법에 대한 연구가 필요하다. 또한 제안한 Adaptive Threshold를 결정하는 방법은 휴리스틱 하기 때문에 학습 방법 등을 통하여 성능을 높일 수 있는 방법에 대한 연구를 할 계획이다.

6. 참고 문헌

[1] D.W. Kim, D.S. Hwang, "Non-Marker based Mobile AugmentedReality and its Applications using Object Recognition," *Journal of Universal Computer Science*, Vol. 18, No. 20, pp. 2832-2850, 2012

[2] Y.K. Jang, J.W. Kim, S.G. Moon, T.J. Nam, D.S. Kwon and W.T. Woo. "Object Recognition utilizing Complementary Feature-point-based descriptor containing

color information," in Proc. of KCC 2012, Vol. 39, No. 1(C), pp. 341-343, 2012.

[3] 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)

[4] Lowe, D. G., "Distinctive Image Feature from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol.60, No.2, pp.99-110, 2004.

[5] 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)

[6] 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.