

# SIFT Bag of Word를 이용한 이미지 카테고리 분류

정현종<sup>○</sup>, 이제민, 낭종호

서강대학교 컴퓨터공학과

fona@sogang.ac.kr, geminix@sogang.ac.kr, jhnang@sogang.ac.kr

## Image Categorization Using SIFT Bag of Word

Hyunjong Jeong<sup>○</sup>, Jemin Lee, Jongho Nang

Dept. of Computer Science & Engineering, Sogang University

### 요 약

디지털 카메라가 보편화되고 스마트폰의 카메라 기능도 빠르게 발전하면서 개인이 촬영한 데이터의 증가와 웹 상에 업로드 하는 경우가 많아짐에 따라 여러 회사에서 대량의 이미지, 영상 데이터베이스를 구축하여 서비스하고 있다. 이에 따라 방대한 이미지, 영상 데이터베이스에서 원하는 이미지를 찾아낼 수 있도록, 유사한 이미지끼리 분류 할 수 있는 시스템에 대한 요구가 증대되고 있다. 따라서 본 논문에서는 효과적으로 이미지를 분류할 수 있는 방법을 제안한다. 제안한 분류 방법은 scale, orientation, luminance의 변화에 강건한 특징을 가지는 SIFT 기술자를 이용한 BoW(Bag-of-Word)를 사용하였다.

### 1. 서 론

여러 장의 이미지 중 원하는 이미지를 효과적으로 검색하기 위한 방법으로 이미지를 카테고리 별로 세분화 한 뒤 검색하는 방법이 사용되고 있다. 이러한 분류 방법으로는 색상 정보를 이용한 DCD, CSD, CLD, Color-Histogram 등이 있다[4]. 하지만 이러한 방법으로 유사 이미지에 대해 카테고리 분류를 수행 할 경우 비슷한 형태의 이미지인 경우에도 오브젝트의 색상, 위치, 크기, 회전 등의 차이 때문에 유사이미지 카테고리의 정확도가 떨어지게 된다.

본 논문에서는 오브젝트의 변화에 대해 강건한 특징을 가지는 SIFT를 이용하고자 한다[1,3]. 본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 연구에 대해 소개하고 3장에서는 본 논문에서 제안한 방법들에 대해 설명하며 4장에서 실험 분석 및 데모 시스템을 소개하고 5장에서는 결론 및 향후 연구로 맺는다.

### 2. 관련 연구

#### 2.1. 내용기반 이미지 검색 기법

내용 기반 이미지 검색을 위해서는 이미지를 대표 할 수 있는 특징을 추출해야 하며 이를 기반으로 색인과 검색을 수행한다. 일반적으로 사용되고 있는 영상의 주요 특징으로는 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 등이 있다.

내용 기반 영상 검색의 초기부터 사용된 영상의 색상 정보는 영상의 내용을 가장 잘 대표하는 특징이라 할 수 있는데 대표적인 기술자로는 Dominant Color Descriptor(DCD), Scalable Color Descriptor(SCD),

Color Structure Descriptor (CSD), Color Layout Descriptor(CLD) 네 가지가 있다.

하지만 색상 정보를 바탕으로 추출되는 기술자 들은 오브젝트의 색상, 밝기, 크기 등에 차이에 따라서 카테고리 분류의 성능이 떨어지게 되고 사용자가 관심을 갖는 부분인 오브젝트에 대한 고려가 없다. 따라서 최근 이미지 객체 인식 분야는 이미지의 전체보다 관심영역인 국소 부위를 고려한 방법들이 각광받고 있다. 이러한 방법들 중에 로컬 패치가 나타난 빈도에 따라 코드북(codebook)을 참조하여 객체를 정의하고 인식하는 방법인 BoW는 다른 객체 인식 방법들에 비해 객체의 인식률이 높은 성능을 보인다.

#### 2.2. Bow이용한 유사도 검색

BoW 방법은 자연어 처리에서의 텍스트 자동 분류 방법이다. 이미지 분야에서는 이미지 한 장을 하나의 문서, 이미지의 로컬 패치(local patch)를 하나의 단어로 생각하여 이미지를 분류하는 방법이다.

먼저 데이터베이스 이미지들의 SIFT 기술자들을 추출하고 이들을 K-means 알고리즘으로 클러스터링(clustering)하여 이미지들을 대표하는 기술자들 K개의 집합으로 이루어진 codebook을 생성한다.

생성 된 codebook을 참조하여 이미지가 가지는 SIFT 기술자를 quantization한다. 이때 생성되는 histogram을 한 이미지의 기술자로 하여 이미지 유사도 검색에 사용한다. 모델링이 단순하며 객체 인식률이 높지만 코드 북을 제작 하는 시간 비용이 크다는 단점이 있다.

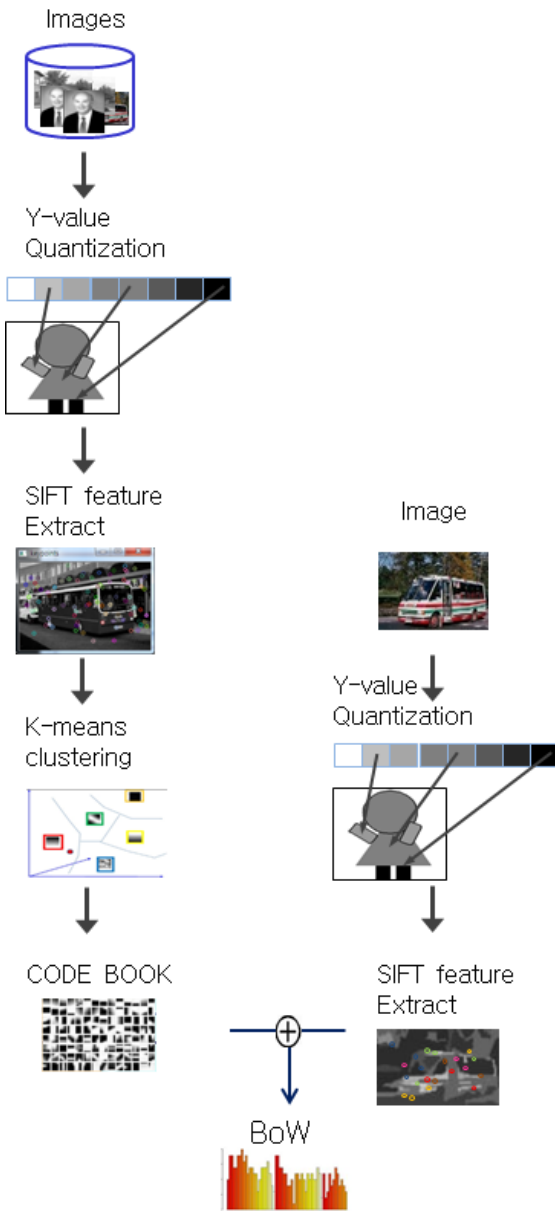
#### 2.3. SIFT(Scale-Invariant Feature Transform)

SIFT 는 object matching 에 사용하는 알고리즘으로 이미지의 scale, orient, luminance 변화에 강건한 특징을

갖고 있다. 먼저 이미지를 grayscale 로 처리한 뒤, 스케일 스페이스에서 DOG 영상을 형성하여 특징 점을 추출하고 이 특징 점에서의 국소영역의 gradient 의 방향과 크기를 128 차원으로 나타낸 기술자이다. 빛이 나아가는 방향을 통해서 회전에 강건하고 스케일 스페이스에서의 특징 점 추출을 통해 크기변화에 강건하다. 또한 전역 기술자가 아닌 국소 기술자로서 휘손이나 잡음에 강건한 특징을 갖는다[1,2].

3. 제안하는 시스템

3.1 SIFT BoW 를 이용한 이미지 유사도 검색



[그림 1] 시스템 구성도

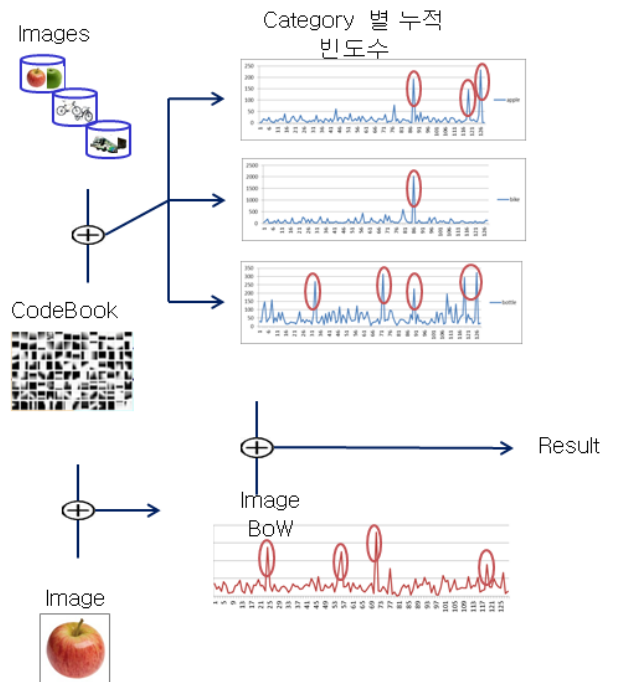
대량의 데이터베이스 내의 이미지들은 객체들의 특징이 다양하고 복잡하게 나타난다. 따라서 종류에 둔감한 BoW를 이용하였고 위치, 크기, 방향, 색 변화에 강건한 SIFT 기술자를 로컬패치로 사용하였다[5].

하지만 BoW 모델을 위해 codebook 을 생성할 때 이미지의 오브젝트 이외의 영역에서 생기는 SIFT 기술자들이 객체 분별에 방해가 되는 요소가 된다는 문제점이 있다. 따라서 이를 해결하기 위하여 BoW 카테고리 정보를 수집할 때 배경이 없는 이미지를 사용하였다.

BoW 를 생성하는 과정은 다음과 같다. 먼저 이미지들을 카테고리 별로 분류한 뒤 이미지의 YUV 밝기 정보 값인 Y 값을 8 단계의 단계로 나눠 quantization 한다. 그 다음 SIFT feature extraction 을 통해 카테고리 별 전체 이미지들에 대해 SIFT feature 들을 추출한다. 이렇게 추출된 SIFT feature 들을 다시 k-means 알고리즘을 통해 최종적으로 k 개의 BoW 로 생성한다.

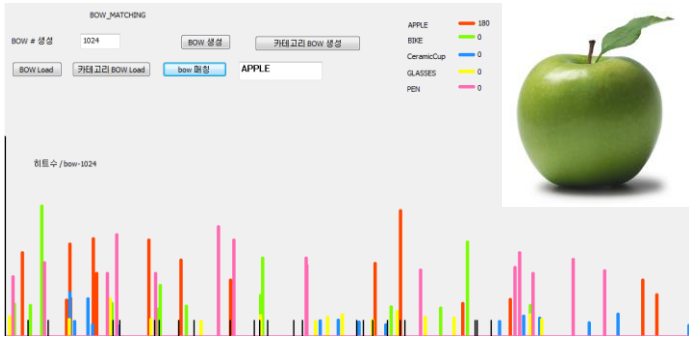
3.2 매칭 알고리즘

카테고리별로 k 개의 BoW 정보에서 가장 두드러진 특징을 나타내는 SIFT 특징점을 상위 n 개 추출한다. 마찬가지로 카테고리가 지정되지 않은 input이미지에 대해서도 quantization을 적용한 뒤에 input이미지에서 SIFT feature extraction을 수행하여 가장 두드러지게 나타나는 특징 점을 상위 m개 추출하였다. 이렇게 카테고리 별로 추출된 BoW의 상위 n개의 점과, input이미지에 대한 상위 m개의 점에 대해 BoW의 어떤 점에 매칭되었는지를 통해 input이미지에 대해 카테고리를 정의 할 수 있다. 이때 카테고리 별 점수를 측정하는 방법은 다음과 같다. 카테고리 별 BoW에서 상위 n개의 BoW를 합한 SoMB(Sum of Max BoW)를 구하고 상위값들을 각각 SoMB값으로 나누어 상위 점수를 정의 하였다[6,7].



[그림 2] BoW 매칭 알고리즘

#### 4. 실험 분석 및 데모 시스템



[그림 3] 실험 tool 그래프 및 매칭 결과

category \ BoW	1024	2048
Apple	98.6%(69/70)	98.6%
Bike	96.1%(75/78)	96.1%
Glasses	86.9%(73/84)	85.7%(72/84)
MugCup	81%(47/58)	81%(47/58)
Pen	85%(187/222)	85%(187/222)

[표 1] BoW별 실험 결과

[그림 3]은 실험 tool이다. 사용된 카테고리는 사과, 자전거, 세라믹컵, 안경, 펜을 사용하였다. 실험에서 사용된 이미지의 카테고리 매칭 확률은 맞춘개수/테스트 이미지셋으로 표현했다. 그리고 그래프에서 각각의 색이 카테고리 별 BoW에서 상위 n개의 SIFT 특징점을 나타낸다. 검은색으로 나타나는 색은 현재 input으로 주어진 이미지에서 추출된 SIFT 특징점 중 상위 m개를 표현한다. 결과는 각각의 카테고리 별 BoW에 검은색으로 나타나는 input값의 존재 유무로 점수를 측정 하였다.

[표 1]은 BoW별 실험 결과다. BoW의 개수에 따른 정확도는 많은 차이가 나지 않음을 확인할 수 있다.

#### 5. 결론 및 향후 연구

원래 BoW 알고리즘은 자연어 처리에서 어떠한 단어가 가장 많이 나왔는지를 알 수 있는데 사용되는 알고리즘이다. 하지만 우리는 이러한 알고리즘을 바탕으로 카테고리 별로 SIFT feature들을 BoW로 생성 하여 해당 카테고리를 대표할 수 있는 feature들을 생성 하였으며 input이미지들에 대해 카테고리 구분을 할 수 있었다. 또한 BoW 알고리즘은 색상의 정보를 사용하지 않기 때문에 기본적으로 색상의 정보를 사용하는 DCD, CSD, CLD, Color-Histogram에 비해 성능은 효과적이지 못했으나 흑백의 이미지를 사용할 경우에도 사용될 수 있다는 장점이 있었다. 추가로 수행해야 할 점은 BoW의 개수 및 카테고리과 input이미지에 대해서 상위 몇 개의 n, m개를 추출할 것인가에 대한 최적화이다. 이러한 결

과를 바탕으로 사용될 수 있는 것으로는 흑백의 이미지에 대해 카테고리를 구분할 수 있으며 또한 하나의 이미지에 대해 여러 오브젝트들이 들어 있을 경우 이러한 SIFT BoW 알고리즘을 통해 하나의 이미지에 어떠한 카테고리에 있는 오브젝트들이 있는지를 한번에 파악할 수 있을 것이다.

#### 참고문헌

[1] D. Lowe, "Distinctive image features from scale invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, Nov 2004.

[2] D. Lowe, "Object Recognition from Local Scale-Invariant Features", Proc. Of the International Conference on Computer Vision, Corfu, 1999.

[3] B. S. Manjunath, Member, IEEE, Jens-Rainer Ohm, Member, IEEE, Vinod V. Vasudevan, Member, IEEE, and Akio Yamada, "Color and Texture Descriptors", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 6, JUNE 2001.

[4] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin, "An efficient color representation for image retrieval," IEEE Trans. Image Processing, vol. 10, pp. 140-147, Jan. 2001.

[5] G. M. Haley and B. S. Manjunath, "Rotation invariant texture classification using a complete space-frequency model," IEEE Trans. Image Processing, vol. 8, pp. 255-269, Feb. 1999.

[6] W. Y. Ma and B. S. Manjunath, "A texture thesaurus for browsing large aerial photographs," J. Amer. Soc. Inform. Sci., vol. 49, no. 7, pp. 633-648, May 1998.

[7] D. K. Park, Y. S. Jeon, C. S. Won, S.-J. Park, and S.-J. Yoo, "A composite histogram for image retrieval," in Proc. ICME 2000, vol. 1, July 2000, pp. 355-358.