장면 분류를 위한 클래스 기반 클러스터링

김준형 류승철 김승룡 손광훈

연세대학교

khsohn@yonsei.ac.kr

Bag-of-Words Scene Classification based on Supervised K-means Clustering

Kim, Junhyung Ryu, Seungchul Kim, Seungryong Sohn, Kwanghoon

Yonsei University

요약

컴퓨터 비전에서 BoW를 이용한 장면 분류 기법에 대한 연구가 활발히 진행되고 있다. BoW 기법의 장면 분류는 K-means 클러스터링을 통하여 코드북을 생성하는 과정에서 트레이닝 이미지의 클래스 정보를 활용하지 않기 때문에 성능이 제한적이라는 문제점을 가지고 있다. 본 논문에서는 BoW를 이용한 장면 분류 과정에서 코드북 생성을 위하여 각각 특징 기술자들의 유클리디안 거리뿐만이 아니라 클래스 확률 밀도 함수들의 히스토그램 교차값을 최소화 하는 최적화 K-means 클러스터링 기법을 제안한다. 장면의 SIFT 특징 기술자 정보뿐만 아니라 장면이 속해있는 클래스 정보를 결합하여 클러스터링을 수행함으로써 장면 분류의 정확도를 높일 수 있다. 장면 분류 정확도 실험에서 제안하는 클러스터링을 사용한 BoW 장면 분류 기법은 기존의 K-means을 사용한 BoW 장면 분류 기법보다 높은 정확도를 보여준다.

1. 서론

최근 컴퓨터 비전 분야에서 주어진 장면의 의미론적 클래스를 결정하는 장면 분류 분야의 연구가 활발하게 진행되고 있다 [1]. 장면 분류는 주어진 장면의 정보를 추출하여 후보 클래스 장면들이 어느 클래스에 해당하는지 분류해 내는 과정이다.

초기의 장면 분류 기법은 장면에서 색이나 텍스쳐의 히스토그램을 사용하거나 GIST와 같은 장면 전체 기술자를 사용하는 기법이 많이 연구 되었다 [2]. 최근에 텍스트 분류에서 널리 사용되고 있는 알고리즘인 BoW(Bag-of-Words) 기법에서 영감을 얻어 장면의 특징점 기술자의 BoW를 사용하는 장면 분류 기법이 널리 연구되었다 [1]. SIFT 특징점 기술자는 장면의 지역적인 특성을 그래디언트(Gradient) 히스토그램으로 표현한다. 이 SIFT 특징점 기술자가 BoW 장면 분류에서 장면의 코드북(Code Book)을 생성하는데 사용된다 [3]. 최근에 장면의 구조 정보를 사용하여 장면 분류를 수행하기 위하여 장면을 피라미드로 나누어 BoW 기법을 수행하는 SPM(Spatial Pyramid Matching) 기법이 연구되었다 [4].

장면 분류 기법에서는 주로 이미지의 특징 기술자를 바탕으로 클러스터링을 통하여 코드북을 생성하는 방식을 따른다. 하지만 장면의 특징 기술자 정보만으로 코드북을 생성하는 경우 장면 분류를 위한 기계 학습 과정에서 훈련 장면의 클래스 정보를 활용하지 않기 때문에 성능의 한계를 가질 수 있다. 본 논문에서는 BoW 기법을 활용하여 장면 분류를 수행하는 과정에서 각 장면의 클래스 정보를 사용하여 클러스터링을 수행한다. 이 클러스터링 방법을 사용해 장면 분류 성능을 개선시키는 알고리즘을 제안한다.

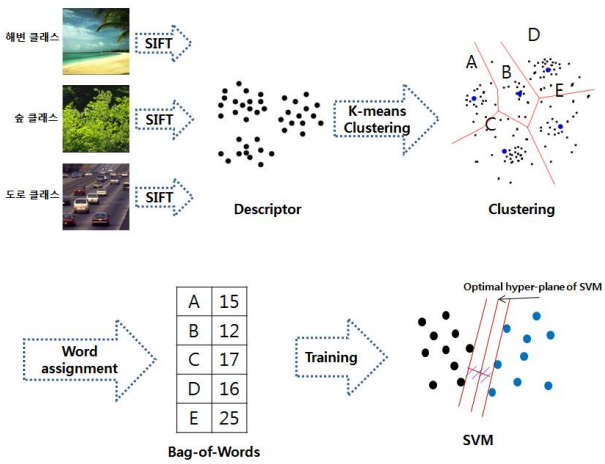

그림1. BoW 장면 분류 과정

장면 분류를 위한 특징 기술자의 코드북을 생성하는 과정에서 기존의 K-means 클러스터링 기법이 아닌 각 장면의 클래스 정보의 히스토그램 정보를 활용한 클러스터링 기법을 사용하여 장면 분류 성능을 높인다.

본 논문의 구성은 다음과 같다. 2장에서 BoW 기법을 이용한 장면 분류에 대한 과정을 설명하고 3장에서 클래스 정보를 활용하여 클러스터링을 수행하는 BoW 기법을 제안하고, 제안된 기법의 실험 결과를 4장에서 기술한다. 마지막으로 5장에서 본 논문의 결론을 맺는다.

2. Bag-of-Words 기반의 장면 분류

BoW 기법을 통한 장면 분류 기법은 4가지 파트로 구성된다. 그림 1에서와 같이 각각의 장면에서 특징 기술자를 추출하고 클러스터링을 통하여 코드북을 생성하고 장면의 BoW를 구한 후 기계 학습을 통하여 장면 분류를 수행하게 된다. 첫 번째로 장면의 특징 기술자를 추출하기 위해서 주로 SIFT 특징 기술자 [3]를 사용한다. SIFT 특징 기술자는 주변 픽셀의 방향을 구한 후 방향 히스토그램을 사용하여 현재 특징점을 표현하는 방식으로 주로 128 벡터를 갖는다. 하나의 장면에서 생성되는 특징점의 수가 N 이라 할 때 SIFT 특징 기술자를 사용하는 경우 $N\times 128$차의 차원을 필요로 한다. 두 번째로 장면의 특징 기술자를 적절한 수로 클러스터링을 한 다음 영역별 분류를 하여 코드북을 생성한다. 이 때 K-means 클러스터링 기법을 사용하여 각각의 특징점 기술자들의 유클리디안 거리에 기반하여 코드북을 생성한다 [5]. 세 번째로 결정된 코드북에 따라서 각각의 장면의 코드북의 히스토그램인 BoW를 형성하여 각 장면을 표현한다. 즉, BoW는 장면을 나타내는 전체 장면 기술자라 할 수 있다. 마지막으로 BoW를 기계 학습하는 과정을 거치게 되는데 일반적으로 SVM 알고리즘을 사용한다 [6]. SVM은 두 그룹에서 각각의 데이터간 거리를 측정하여 두 개의 중심을 구해서 그 가운데에서 최적의 초평면을 구함으로써 그룹을 나누는 방법이다.

이러한 BoW를 사용한 장면 분류 기법은 각각 장면의 특징 기술자들을 클러스터링 하여 생성된 코드북의 분별력에 따라서 성능이 좌우된다 [7]. 기존의 BoW 장면 분류 기법에서는 훈련 장면들에서 얻어진 특징 기술자들을 단순한 유클리디안 거리에 기반한 K-means 클러스터링 기법으로 코드북을 생성하기 때문에 장면 분류 성능에 한계가 존재한다. 따라서 훈련 장면의 클래스 정보까지 고려한 클러스터링 기법으로 코드북을 생성하여 BoW를 구하여 장면 분류를 수행한다면 정확도가 향상될 수 있다.

3. 클래스 정보 활용 클러스터링 기반의 BoW 기법

제안하는 장면 분류 기법은 장면의 클래스 정보를 사용하여 클러스터링을 수행하여 오차에 강인한 코드북을 생성한다. 그림 2에서와 같이 제안하는 방식은 두 단계에 걸쳐서 클러스터링을 수행하여 성능을 높인다. 먼저 특징 기술자를 K-means 클러스터링 기법을 사용하여 원하는 코드북의 수보다 오버 클러스터링을 수행하여 각각의 클래스 확률 밀도 함수를 구한다. 두 번째로 구해진 클래스 확률 밀도 함수의 히스토그램 교차값과 특징점 기술자의 유클리디안 거리를 최소화 하는 클러스터링을 수행한다. 이러한 방식으로 결정된 코드북은 기존의 K-means 클러스터링 기법에 비하여 클래스에 따라 특징 기술자들이 클러스터링 됨으로써 분별력이 높아지고 장면 분류 성능을 높을 수 있는 장점을 가진다.

3.1 클래스 인지 클러스터링 에너지 함수

기존에 코드북 생성을 위한 K-means 클러스터링은 식 (1)과 같은 식을 만족하는 중심 특징 기술자 집합을 찾아내는 것이다. 클러스터링을 위한 전체 특징 기술자들에 대하여 반복법을 통하여 유클리디안 거

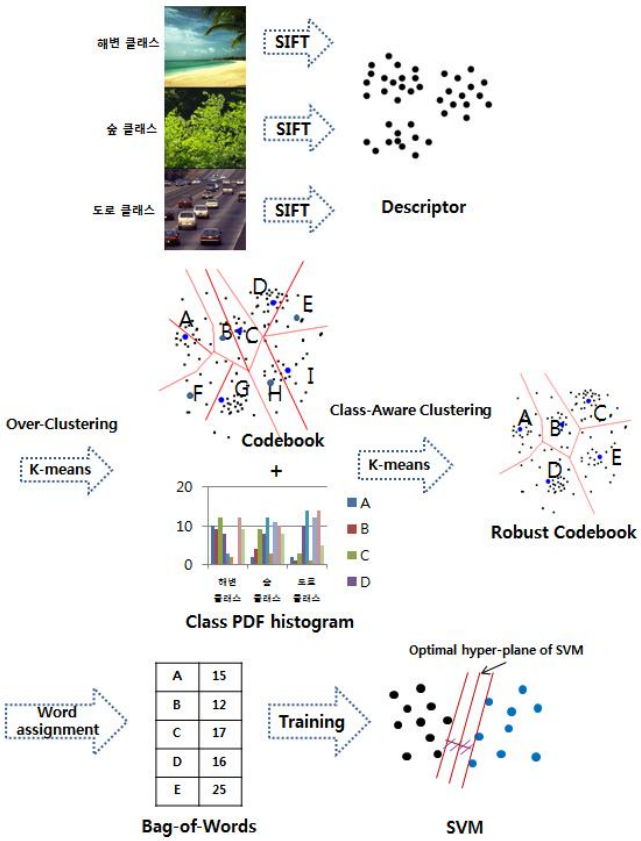

그림 2. 제안하는 클러스터링 방법을 통한 BoW 장면 분류 과정

리가 가장 최소화 되는 최적의 분할점을 찾는다.

$$\min_V \sum_{m=1}^{M} \min_{k=1,\dots M} \|x_m - V_k\|^2 \qquad (1)$$

x_m은 각각의 특징 기술자이고 V_k는 클러스터링 되는 중심점을 의미한다. K는 클러스터링 개수를 의미하고 M은 분할하고자 하는 특징 기술자의 개수를 말한다. K-means 클러스터링은 유클리디안 거리를 최적화하는 클러스터링 중심 벡터 집합 V를 찾아낸다. 이러한 K-means 클러스터링 기법은 각각의 특징 기술자들이 어떠한 클래스에서 생성되었는지에 대한 정보를 사용하지 않기 때문에 성능에 한계가 있다.

따라서 제안하는 클러스터링 기법은 식 (2)와 같이 각각의 특징 기술자들의 유클리디안 거리 뿐만 아니라 각각의 특정 기술자들의 클래스 히스토그램의 교차값을 최소화 하는 방향으로 최적의 분할점을 찾는다.

$$\min_V \sum_{m=1}^{K} \min_k (\|x_m - V_k\|^2 + \lambda D_c(x_m, V_k)) \qquad (2)$$

λ는 가중치 지수이고 $D_c(x_m, V_k)$는 각각의 특징 기술자의 클래스 히스토그램 교차값을 의미한다. 제안하는 방식은 K-means와는 다르게 유클리디안 거리와 교차값이 최소화되는 클러스터링 중심 벡터 집합 V를 찾아낸다.

특징 기술자의 클래스 히스토그램의 교차값은 식 (3)과 같이 계산된다.

$$D_c(x_m, V_k) = 1 - \sum_{i=1}^{I} \min(p(c_i|x_m), p(c_i|V_k)) \quad (3)$$

$p(c_i|x_m)$는 특징 기술자 x_m의 클래스 밀도 함수를 말하고, $p(c_i|V_k)$는 클러스터링 중심 V_k의 클래스 밀도 함수를 말한다. 클래스 밀도 함수의 유사도가 높을수록 특징 기술자는 비슷하게 분할되어야 하므로 $D_c(x_m, V_k)$은 클래스 히스토그램 교차값을 나타낸다. 따라서 $D_c(x_m, V_k)$ 값이 작을수록 x_m과 V_k는 유사하다.

3.2 오버 클러스터링 및 클래스 확률 밀도 함수 생성

오버 클러스터링은 K-means 기법을 사용하여 원하는 코드북 크기보다 많은 수로 클러스터링을 수행해 주는 과정이다. 이러한 오버 클러스터링은 무한 가짓수가 나오는 특징 기술자를 유한 개의 각각 특징 기술자의 클래스 히스토그램을 생성을 가능하게 한다. 즉, 각각의 특징 기술자에 대하여 확률 밀도 함수 $p(c_i|x_m)$를 생성한다.

3.3 클래스 정보 활용 클러스터링

오버 클러스터링 후 식 (2)에서와 같이 특징 기술자가 속한 확률 밀도 함수를 결합한 데이터간의 거리를 구하여 클러스터링을 수행하는 과정을 거친다. 즉, 클래스 정보를 특징 기술자와 결합해서 인접하는 데이터의 위치를 재정립시켜서 정확성을 높이려는 것이다. 여기에서 특징 기술자 간의 거리는 기존의 방식과 마찬가지로 유클리디안 거리를 사용하고, 클래스의 확률 밀도 함수의 거리는 히스토그램 교차값 거리 계산법을 수행한다. 각 클래스 값의 최소값을 합해서 거리 값을 구한다. 두 거리 값을 합해서 최소가 되는 방향으로 클러스터링을 수행한다. 단순히 특징 기술자 간의 거리만을 고려하는 것이 아니라, 특징 기술자가 속하게 되는 클래스의 확률 정보도 포함시킴으로써 성능의 향상을 얻을 수 있다.

그림 3. 2688장의 8개의 클래스 장면 [8]

4. 실험 결과 및 분석

4.1 실험 조건

제안하는 기법의 장면 분류 성능을 평가하기 위하여 [7]에서 사용된 데이터 베이스의 테스트 장면(2688장)을 활용하였다. 8개의 클래스로 구분되는 2688장의 테스트 이미지 중 2000장을 트레이닝 하여 668장을 정확성을 확인하기 위한 분류 대상 이미지로 사용하였다. 프로그램이 작동하는 컴퓨터의 성능은 Windows 7, Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz, 4.00GB RAM, 64비트 운영체제이다. 구현한 소프트웨어 프로그램은 MATLAB R2009b 버전이다.

제안하는 클러스터링을 수행하기 위하여 SIFT 특징 기술자를 사용하였고 200개의 코드북을 생성하였다. 각각의 특징 기술자는 500개의 오버 클러스터링을 통하여 클래스 확률 밀도 함수를 구하였다. 장면 이미지 분류를 위해서 각각의 8개의 클래스에 해당하는 이미지 2688장 중 2000장을 SVM [6] 훈련 과정을 수행하고 688장을 테스트 장면으로 사용하였다.

4.2 장면 분류 정확도 성능 평가

장면 분류 정확도 실험을 위하여 제안하는 클러스터링 기법을 사용한 BoW 기법과 K-means 클러스터링 기법을 사용한 BoW, 그리고 SPM 기법을 비교하였다. 표 1은 각각의 장면 분류 정확도 실험을 수행한 결과를 보여준다. 장면 분류 정확도는 SVM 트레이닝을 통하여 테스트 장면을 분류를 수행하였을 경우 총 몇 개의 장면이 정확히 클래스에 분류되는지를 측정하였다.

표 1을 통해서 확인할 수 있듯이 기존의 방식으로 BoW를 수행했을 때와 비교하면 제안된 방법으로 수행했을 때, 11%의 정확성이 향상되었음을 확인할 수 있다. 제안하는 방식은 최근 가장 성능이 높은 장면 분류 기법으로 알려져 있는 SPM보다도 높은 정확도를 보였다.

표 1. 장면 분류 정확도 실험 결과

	BoW	제안 방법
분류 정확성(%)	40.3436	51.2384
비율 (매치/테스트)	277/688	352/688

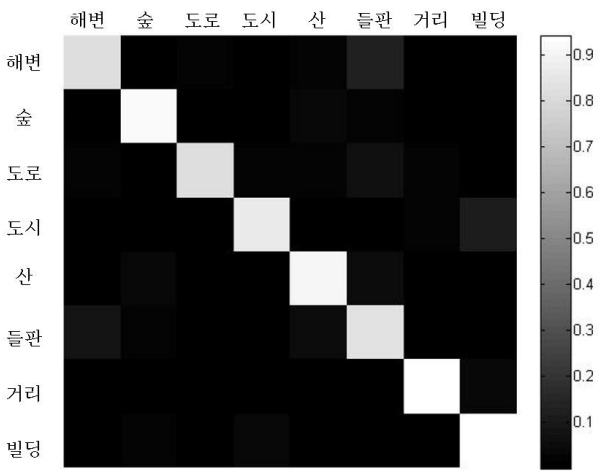

그림 4. 장면 분류 테스트 장면의 오차 행렬

그림 4는 위의 실험에서 수행했던 장면 분류의 매치되는 각각의 정확도를 확인할 수 있는 오차행렬(Confusion Matrix)이다 [8]. 각각의 클래스가 평균적으로 분류되는 정도를 대각선의 수치에서 확인할 수 있으며 장면 분류 정확도가 높을수록 대각선 수치가 높아진다. 그림 4에서 보는 바와 같이 제안하는 장면 분류 기법은 기존의 알고리즘에 비하여 높은 성능을 나타냄을 확인할 수 있다.

5. 결론

본 논문에서는 BoW를 이용한 장면 분류 과정에서 정확성을 높이기 위하여 클래스 정보를 고려한 클러스터링 과정을 제안하였다. 장면을 분류하기 위해서는 SIFT 특징 기술자를 추출하여 오버 K-means 클러스터링을 거쳐서 각각의 특징 기술자의 클래스 확률 밀도 함수를 구한다. 다음으로 각각의 특징 기술자의 유클리디안 거리 뿐만 아니라 클래스 확률 밀도 함수의 교차값을 최소화 하는 클러스터링 중심을 구하여 코드북을 생성한 후 BoW를 수행하게 된다. 제안하는 방식은 기존의 BoW 장면 분류 기법보다 11%의 정확성이 향상되는 것을 확인할 수 있다. 제안하는 방식은 장면 분류 분야에서 뿐만이 아니라 K-means 클러스터링을 사용하는 분야에서 성능을 향상시킬 수 있을 것이다.

참고문헌

[1] Li, Li-Jia, et al. "Objects as attributes for scene classification." Trends and Topics in Computer Vision. Springer Berlin Heidelberg, 2012. 57-69.

[2] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145 - 175, 2001.

[3] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60.2 (2004): 91-110.

[4] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." *Computer Vision and Pattern Recognition,* 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.

[5] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1 (1979): 100-108.

[6] Cortes, C., and V. Vapnik. "Support vector machine." *Machine learning* 20.3 (1995): 273-297.

[7] Csurka, Gabriella, et al. "Visual categorization with bags of keypoints." *Workshop on statistical learning in computer vision, ECCV.* Vol. 1. 2004.

[8] Dataset : Oliva, Aude, and Antonio Torralba. "Modeling the shape of the scene: A holistic representation of the spatial envelope." *International journal of computer vision* 42.3 (2001): 145-175.

[9] Yang, Jun, et al. “Evaluating bag-of-visual-words representations in scene classification." *Proceedings of the international workshop on Workshop on multimedia information retrieval.* ACM, 2007.