

특이점 가중치 기반 PLSA를 이용한 객체 범주화

송현철* 최광남**

중앙대학교 컴퓨터공학과

shon@vim.cau.ac.kr, knchoi@cau.ac.kr

Object Categorization Using PLSA Based on Weighting Distinctions

Hyun Chul Song, Kwang Nam Choi

School of Computer Science & Engineering, Chung-Ang University

요 약

영상 내 사물들의 카테고리를 인식하는 연구는 시각적 영상처리와 연관된 다양한 분야에서 활발히 진행되고 있다. 객체 범주화(Object Categorization)는 가정과 같은 실내에서 책상, 의자, 컵, 주전자 등의 다양한 사물들을 구분하여 인식하는데 중요한 역할을 할 수 있다. 본 논문에서는 최근 영상 내 객체들의 카테고리 분석을 위해 연구된 PLSA를 기반으로 특이점에 가중치를 부여하여, 보다 유사한 카테고리 간에 인식 성능을 향상시키는 접근법에 대하여 연구하였다. PLSA는 문서기반의 정보검색 분야로부터 소개된 기법으로, 약한 수준의 비감독 방법임에도 불구하고 인상적인 인식성능을 보여준다. 그러나 비슷한 특징점 분포를 보이는 유사한 카테고리 간의 객체 카테고리 인식에 대해서는 비교적 낮은 성능을 보인다. 본 연구에서는 카테고리간의 비교실험을 통해 각 특징점에 대하여 가중치를 부여한 PLSA를 적용하여 유사한 객체 간의 카테고리 인식 가능성을 살펴보았다. 실험에서는 기존의 PLSA 기법과 제안한 가중치 부여 PLSA 기법을 각각 적용하여 그 성능을 비교하였다. 본 연구에서는 기존 PLSA 기법에서는 비교적 낮은 인식률을 보인 유사한 카테고리 인식에 대하여 실험 결과를 통해 가중치를 부여한 PLSA 기법이 보다 향상된 성능을 보임을 확인하였다.

1. 서 론

영상 내 객체의 카테고리를 식별하는 문제는 수십년 간 많은 연구자들의 노력에도 불구하고 대부분의 부분들이 해결되지 않은 채 남아 있다. 사람의 얼굴 인식과 같은 특정 부분에 대해서는 납득할 만큼 성공적인 성과들이 이루어졌으나 일반적인 사물의 카테고리 식별에 대해 만족할 만한 방법론은 없다. 때문에 영상 학습을 통해 영상 내 사물을 인식해 내고 영상간의 카테고리를 추출하려는 연구가 활발히 진행되고 있다[1].

사물의 카테고리를 학습하고 인식하기 위한 다양한 모델들이 다음과 같이 제시되어 왔다. 첫째, Bag of Words 모델은 이미지를 하나의 문서로 가정하고 문서의 키워드들을 추출하는 것과 같이 영상 내의 특이점들을 각각의 Visual Words로 보고 이미지의 카테고리를 학습한다. 현재 이 모델의 가장 주도적인 두

접근법으로는 PLSA(Probabilistic Latent Semantic Analysis)[2]와 PLSA 모델의 위계적 형태인 LDA(Latent Dirichlet Allocation)[3]가 있다. 둘째, Part-Based 모델은 위치 정보에 대한 고려가 부족한 Bag of Words 모델과는 달리 사물 각 부분의 상대적 위치정보를 바탕으로 카테고리를 판별한다. 셋째, Discriminative Methods는 기존에 알려진 다양한 분류기(Nearest Neighbor, Neural Network, Support Vector Machines, Conditional Random Fields 등)를 통해 대상을 판별한다.

본 연구의 동기는 Bag of Words Model에서 가장 주도적인 접근법의 하나인 PLSA를 기반으로 각기 다른 카테고리에 속한 사물들의 차이점을 주도적으로 나타내는 특징점에 가중치를 부여하여 주도적인 Visual Word를 강조함으로써 영상의 카테고리 인식성능을 향상시킬 수 있는 접근법에 대해 연구한다. 이 논문에서의 접근법은 특정 카테고리에 차별적으로 등장하는 특징점에 대해서는 해당 카테고리가 선택될 가능성을 높여주고 전체 카테고리에 대하여 중복되어 나타나는 특징점에 대해서 특정 카테고리로 인식될 가능성을 줄여 줌으로써 유사한 카테고리 간의 변별력을 높여준다. 본 논문에서는 불특정 분류에 속한 사물들의 예제 사이에서 자동적으로 시각적인 일관성을 학습하여 새로운 예제를 인식하는 방법론을 소개하고

※ 이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (R01-2005-000-10940-0)

* 학생회원 : 중앙대학교 컴퓨터공학과

** 종신회원 : 중앙대학교 컴퓨터공학과 교수

접수일 : 2007년 4월 17일,

완료일 : 2007년 5월 17일

분류에서 주도적으로 드러나는 특징에 대한 가중치 부여를 통해 기존의 방법을 보다 개선하고자 한다.

객체범주화는 수많은 가능한 분류 중에서 해당 이미지가 속한 카테고리가 어떤 것인지 결정하는 과제이다. 따라서 분류기의 경우는 두 개의 카테고리를 분류하게 되지만 객체 범주화의 경우는 n개의 카테고리 분류 과제를 갖게 된다. 본 논문에서는 하나의 객체가 하나의 카테고리에 속한다고 가정한다.

이전까지의 연구에서는 특정 사물에 대한 인식과는 다르게 사물의 카테고리를 인식하는 과제는 더 이상 한 분류에 속한 예제 각각이 별개의 대상이 아니기 때문에 더 큰 어려움을 갖는다. 자동차나 얼굴에 대한 성공적인 연구결과가 있음에도 불구하고, 이들의 연구는 이러한 특정 카테고리에 적합하도록 디자인 되어 있다. 또한 그들의 표현들은 얼굴과 자동차에 대해서 쉽게 구분 지을 수 있는 특징점과 제약조건들이 주어진다. 현재 이러한 연구들은 특정 조정을 요구하지 않고 모든 카테고리에 적용할 수 있는 새로운 알고리즘을 개발하는데 중점이 모아지고 있다. 특히 각 카테고리의 핵심점을 추출함으로써 분류 내 변화들을 모델링하고 있다. 보다 최근의 연구들은 인식대상 자체나 그것의 놓여있는 방향, 크기는 윤곽선 정보를 무시하고 영상으로부터 특징 있는 영역들의 특징점의 집합을 추출한다. 본 논문에서는 Bag of Words 모델을 기반으로 가중치 부여 PLSA에 대하여 연구한다.

Bag of Words 모델은 모든 문서(열)에 대해 각 단어(행)의 빈도수를 대응시킨 관찰행렬로 자료를 표현함으로써 분석을 효과적으로 단순화시킨다. Bag of Words 모델은 각 특징점들 간의 어떠한 공간적 관계성을 무시하는 표현법을 가지고 있음에도 불구하고 문서기반 영역에서 놀라운 정도의 높은 변별력으로 성공적이었다.

문서기반의 주어진 대량의 어떤 주석도 없는 훈련 데이터로 비감독 주제 발견에 주목할 만한 성공을 거두었다. 이 점에 착안하여 객체의 카테고리를 발견하기 위해 문서에서 주제를 찾기 위한 통계적 자연어 처리에서 사용되는 모델이 적용되었다. 이 고안된 모델에서 문서는 영상으로 단어는 영상의 지역적 특징점 기술자(Descriptor)를 양자화한 시각적 단어(Visual Word)에 대응된다. 가장 주요하게 연구되는 두 모델은 Hofman의 PLSA(probabilistic Latent Semantic Analysis)[2]와 Blei의 LSA(Latent Dirichlet Allocation)[3][4]. Sicvic은 PLSA를 비감독 객체범주화에 Fei-Fei와 Perona는 LDA를 장면 분류에 각각 이 두 모델을 적용하였다[5][6]. 두 모델은 비슷한 결과를 보인다. 이 논문에서는 특이점에 대한 가중치 부여의 효과를 보이기 위해 보다 단순한 형태의 PLSA 기법에 초점을 두고자 한다. 두 모델은 모두 특징점들 간의 상대적 위치관계를 고려하지 않는 Bag of Words 표현법을 사용한다.

표 1 문서와 시각 영역에 적용된 PLSA에서의 서로 다른 용어

기호	영역	
	문자기반	영상기반
d	문서	영상
w	단어	Visual Word
z	주제	주제/객체

2. Visual Words

Visual Words는 다양한 특징점 탐지기를 통하여 찾은 점을 영상특징기술자로 기술하고 마지막으로 다양한 영상에서 얻은 많은 수의 특징벡터들을 클러스터링 알고리즘을 통해 군집화하여 벡터 양자화한 결과물이다. 따라서 Visual Words는 Bag of Words 모델의 기반이 된다. 본 논문에서는 특징점 탐지기로 Canny Edge Detector를 적용하였고, 영상특징기술자로 SIFT 기술자를 이용하였다. 마지막으로 클러스터링 알고리즘으로는 K-means 알고리즘을 사용하여 훈련데이터를 바탕으로 Visual Wordbook을 형성하였다.

2.1. Canny 경계선 탐지기

Canny 경계선 검출기는 1986년 John Canny 에 의해서 고안되었으며 경계선 검출을 위한 최선의 방법 중 하나로 알려져 있다. Canny Edge를 기반으로 한 Canny Edge detector는 최소한의 윤곽선의 화소들을, 가장 세부적인 Canny edge map은 가장 큰 숫자의 윤곽선의 화소 수를 갖는다. 관계없는 윤곽선들로 인해 생길 수 있는 부작용을 줄이기 위해, 기본적인 추적은 입력 이미지에서 이미지들 사이의 큰 차이를 보이는 값들 통해 얻어진 Canny edge를 사용한다.

2.2. SIFT 서술자

영상특징기술자(Image feature descriptor)는 시점의 변화, 물체의 크기, 이동 및 회전, 일부 조도변화, 어파 인이나 3D투영에 변하지 않는 영상 내 특징점을 추출하고 기술할 수 있는 방법으로써 물체인식, 영상복구와 같은 다양한 환경에서의 비전문제를 위해 사용된다. Lowe가 제안한 SIFT는 피라미드 크기 공간에서 가우시안 차 (Difference of Gaussian: DoG)의 최대값들을 핵심점(Keypoint)으로 정의하고 방향성분의 히스토그램으로써 기술자를 정의함으로써 물체나 전경의 서로 다른 시점사이의 신뢰성 있는 매칭을 할 수 있는 영상특징기술자로 잘 알려진 방법이다. 이는 핵심점 추출과 기술자가 배경의 변화에 따라 오차를 갖지만 연속적으로 프레임에서의 무늬(texture)를 갖는 물체들에게는 좋은 반복성(Repeatibility)을 갖는다. 또한

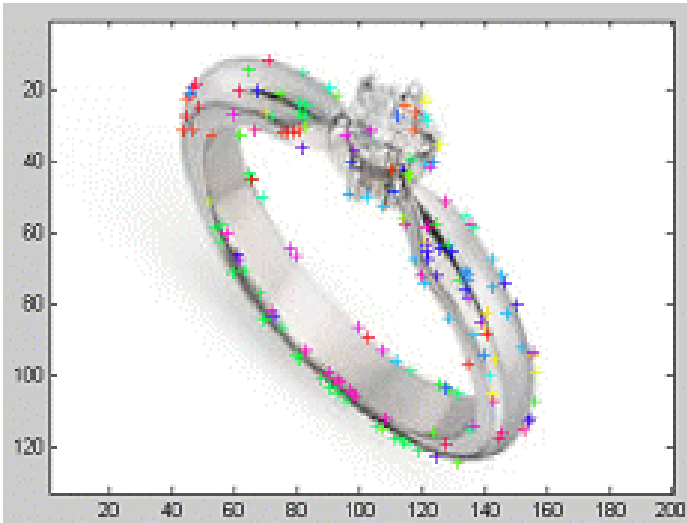


그림 1 영상에서 추출된 Visual Words

상태에 정의되는 관심영역을 설정함으로써 처리시간에 대한 부담을 줄임으로써 전체관측가능성의 융합을 위한 변별력 있는 영상특징으로 사용할 수 있다[7]. 본 논문에서는 특징점을 기술하여 Visual Word를 형성하는 서술자로 사용되었다.

2.3. Visual Word의 생성

Sivic은 영상으로부터 추출된 영역들의 집합들로부터 Visual Words를 생성하기 위해 양자화된 외형정보를 담고있는 같은 벡터를 추출하였다[8]. 어떤 위치정보도 없이 단지 영상은 Visual Words의 히스토그램에 의해서 표현된다. 시각적 어휘집은 다양한 주제의 서로 독립적인 훈련 영상들의 집합으로부터 추출된 큰 특징점 영역에 대한 집합에 대하여 K-means를 실행함으로써 생성된다. 어휘집의 크기는 시스템의 파라미터로 주어지게 된다. Sivic과 Fei-Fei의 연구에서 200~2000개 범위의 다양한 어휘집 크기가 적용하였다. 본 논문에서 사용된 바와 같이 Sivic은 각각의 특징점 영역들을 128차원의 SIFT 기술자를 사용하였고, Fei-Fei와 Perona는 각 영역에 대하여 서로 다른 두 표현법을 사용하였다[9]. SIFT는 관심영역의 기술자로서 두 연구에서 뛰어난 성능을 보였다.

K-means 알고리즘에서는 각각의 Cluster들은 오직 그 안의 Code-vector에 의해 특징 지워지고, 모든 훈련 벡터는 자신과 가장 가까운 Code-vector가 있는 Cluster에 속하게 된다. 이 방법은 Clustering하는데 있어서 매우 효과적이며 많이 쓰이는 방법이다.[10] 이러한 Code-vector들을 모아 Codebook을 생성할 수 있다. Codebook 작성 후, 각 영상에서 추출된 SIFT 기술자들을 벡터 양자화한다. 즉, Codebook을 기준으로 가장 가까운 거리에 위치한 Code-vector에 의해

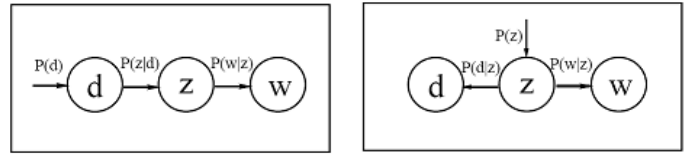


그림 2 주제 Z에 대한 매개변수화

라벨링하는 것이다. 이 과정에서 각 영상에 대한 Visual Words를 얻을 수 있다. 각 영상을 특징점을 기반으로 Visual Words의 집합으로 표현하는 방법은 Bag of Words 모델의 대표적인 표현법이다. 이를 바탕으로 베이지안 네트워크, PLSA, LDA 등의 다양한 기법을 적용할 수 있으며, 본 연구에서는 가중치 부여 PLSA를 제안하여 실험하였다. 그림1은 각 카테고리 별 영상에서 추출된 Visual Words가 300개의 프로토타입에 따라 각기 다른 색으로 라벨링 된 결과를 보여준다. 같은 색은 같은 Visual Words로 구분된 특징점을 뜻한다.

3. Object Categorization 모델

N개의 문서가 있고 각 문서에 M개의 주요한 어휘군이 있다고 할 때, 문서는 $D = \{d_1, d_2, \dots, d_N\}$, 어휘군은 $W = \{w_1, w_2, \dots, w_M\}$ 로 표현할 수 있다. 문서 내의 순차적 위치관계를 무시한다면 $N = (n(d_i, w_j))_{ij}$, ($n(d, w) \in \mathbb{N}$) $N \times M$ 의 상호교차표로 정리할 수 있다. 즉 'Bag of Words'로 표현됨으로써 벡터공간으로 표현이 가능한 것이다.

3.1. Probabilistic Latent Semantic Analysis (PLSA)

PLSA는 $P(d, w)$ 에 대하여 카테고리에 해당하는 z에 대한 매개변수화하면 그림 2와 같은 모델을 갖는다. 다음 수식은 그림 2에 표현된 모델에 대한 수식표현이다. $P(d)$, $P(w)$ 는 서로 독립사건이라는 가정 하에 $P(d, w)$ 를 $P(w|d)$ 에 대하여 전개하였고, 최종적으로 주제 z에 대하여 매개 변수화 시키게 된다.

$$P(d, w) = P(d)P(w|d) \quad (1)$$

$$P(w|d) = \sum_{z \in Z} P(w|z)P(z|d) \quad (2)$$

$$P(d, w) = \sum_{z \in Z} P(z)P(d|z)P(w|z) \quad (3)$$

PLSA는 공기(co-occurrence) 데이터 분석을 위한 통계적 기법으로서 모델링에 적용된다. PLSA

기법에서는 공기 데이터에 대한 Aspect 모델에 기반하여 데이터 각 항목에 대해 잠재 변수 $z \in Z = \{z_1, z_2, \dots, z_k\}$ 를 이용한다. 영상의 경우 각 데이터 항목은 Visual Words, w 와 Image, d 의 쌍 (w, d)로 주어지며, 각 z_i 는 하나의 의미자질(Semantic Topic) 즉 영상의 카테고리를 나타낸다. PLSA는 확률적 mixture decomposition에 기반한 방법으로 데이터에 대한 생성 모델을 다항분포로 정의하고 이에 대한 유사도값을 최대화함으로써 모델을 학습한다. 모델의 학습은 훈련이미지에 대하여 topic z 에 대한 파라미터를 EM(Expectation Maximization) 알고리즘을 통해 지역최적해를 구한다. PLSA 훈련단계에서 훈련 데이터 집합에서 얻은 확률모델을 얻게 된다. 확률모델은 EM(Expectation Maximization) 알고리즘을 이용해 얻는다.

3.2. 제안한 가중치 부여 PLSA 알고리즘

PLSA로부터 얻은 결과를 통하여 각 카테고리에 해당하는 Visual Words에 대한 히스토그램을 생성한다. 각 카테고리에서 히스토그램의 차이가 큰 Visual Words에 차이 히스토그램의 빈도를 가중치로 적용하게 된다. PLSA의 과정을 이용하기 때문에 PLSA가 가지는 비감독 범주화의 특성을 유지시킬 수 있다. 본 시스템에 적용한 가중치 획득과 가중치 부여하는 과정은 다음과 같다.

확률적 모델로서 얻은 $P(w|z)$ 와 $P(z)$ 를 이용하여 각 테스트 영상이 특정 카테고리에 속할 확률 $P(z|d)$ 을 얻게 된다. $P(z|d)$ 를 얻기 위해서 간단한 베이저안 정리를 이용한다. 이를 위한 수식은 다음과 같다.

$$\begin{aligned}
 P(z|d) &= \frac{P(d \cdot z)}{P(d)} = \frac{P(d \cdot z)}{\sum_{i=1} P(d \cdot z_i)} \\
 &= \frac{\frac{P(d \cdot z)P(z)}{P(z_i)}}{\sum_{i=1} \frac{P(d \cdot z_i) \cdot P(z_i)}{P(z_i)}} \\
 &= \frac{P(d \cdot z)P(z)}{\sum_{i=1} P(d \cdot z_i) \cdot P(z_i)} \quad (4)
 \end{aligned}$$

$P(w|z)$ 를 통해 구분된 라벨에 따라 각 이미지에서 속한 Visual Words의 빈도수를 누적하여 카테고리별 Visual Words에 따른 히스토그램을 구한다. 각 카테고리에 속한 이미지의 수에 따라 평균 히스토그램을 구한다. 최종적으로 두 히스토그램의 차이를 기록한 차이 히스토그램을 얻게 된다.

$$\begin{aligned}
 &(z_1 \text{의 차이 히스토그램}) \\
 &= (z_1 \text{의 평균 히스토그램}) - (z_2 \text{의 평균 히스토그램}) \quad (5)
 \end{aligned}$$

z 의 차이 히스토그램은 해당 카테고리과 다른 카테고리 간의 Visual Words의 분포차이를 그대로 반영한다. z_1 에서 주도적으로 등장하는 Visual Words에 대해 높은 빈도를 갖게 되고 카테고리과 상관없이 공통적으로 등장하는 Visual Words에 대해서는 상대적으로 낮은 빈도를 갖게 된다. 이러한 특성 때문에 차이 히스토그램의 빈도수를 가중치로 사용하여 각 Visual Words가 특정카테고리에 속할 확률에 적용할 경우 특이점이 되는 Visual Words에 대한 변별력을 높이고 빈도가 높음에도 불구하고 공통된 Visual Words에 대한 변별력은 낮춰지게 된다. 따라서 Visual Words의 분포가 유사한 카테고리간의 변별력을 높여주는 역할을 하게 된다.

차이 히스토그램을 가중치고 반영하는 과정은 다음과 같다. EM의 파라미터인 $P(w|z)$ 와 차이히스토그램의 빈도수를 각각의 Visual Words에 따라 적용시킨다.

$$P(w|z) = P(w|z) \times (z_1 \text{의 차이 히스토그램}) \quad (6)$$

이때 확률적 정의에 의해 상대적으로 각 카테고리에 속할 확률의 합이 1이 되어야 한다. 이에 대한 정규화 작업은 다음과 같다.

$$P(w|z) = \frac{P(w|z)}{\sum_{i=1} p(w|z_i)} \quad (7)$$

각각의 Visual Words에 대해 가중치 적용된 $P(w|z)$ 를 EM알고리즘의 파라미터로 적용하여 $P(w|z)$ 와 $P(z)$ 에 대한 지역 최적해를 구하게 된다. 적용되는 목적함수는 다음과 같다.

$$P(z|d, w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z' \in Z} P(z')P(d|z')P(w|z')} \quad (8)$$

가중치 부여 PLSA 평가 과정을 통해 $P(w|z)$ 에 특이점에 대한 가중치가 반영하여 테스트 이미지에 대해 가중치가 반영된 $P(d|z)$ 와 $P(z)$ 의 확률적 모델을 정의한다. 마지막으로 $P(d|z)$ 와 $P(z)$ 를 이용하여 베이저안 정리로써 $P(z|d)$ 의 확률모델을 구한다. 본 시스템은 최종적으로 얻게 되는 $P(z|d)$ 값이 최대값을 갖는 z 에 대하여 해당 이미지의 카테고리로 최종평가를 내린다.

표 3 PLSA와 Weighted PLSA의 카테고리 별 인식을 비교

	오토바이		자동차		비행기		반지	
	PLSA	제한한 PLSA	PLSA	제한한 PLSA	PLSA	제한한 PLSA	PLSA	제한한 PLSA
오토바이			67.4%	69.0%	75.0%	76.1%	69.8%	70.0%
자동차	67.4%	69.0%			83.3%	83.3%	93.0%	93.0%
비행기	75.0%	76.1%	83.3%	83.3%			92.1%	92.1%
반지	69.8%	70.0%	93.0%	93.0%	92.1%	92.1%		

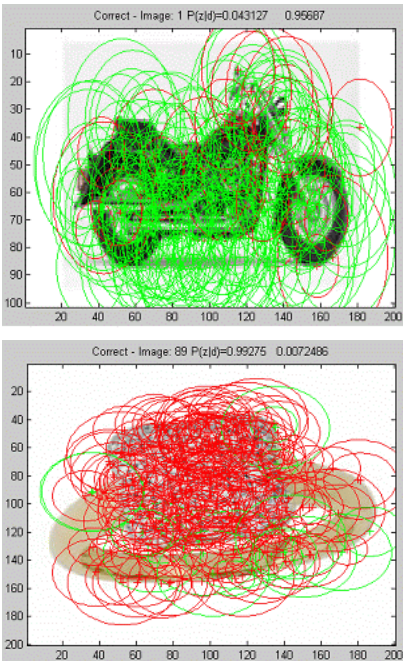


그림 3 오토바이와 반지 영상에서 추출된 특징점

표 4 오토바이와 자동차 카테고리 대한 인식률 평균 및 표준편차

	PLSA	제한한 PLSA
평균인식률	67.40%	69.00%
표준편차	0.4983	0

4. 실험 및 결과

본 논문에서는 제안한 가중치 부여 PLSA를 통해 영상 내 사물의 카테고리 인식을 위한 시스템을 구현하고 이를 통해 영상의 카테고리를 인식하고자 한다. 실험에 사용된 영상의 카테고리는 다음과 같이 오토바이, 자동차, 비행기, 반지이다. 각 카테고리에 대해서 100개의 영상 중 50%는 훈련데이터로 50%는 평가 데이터로 사용하였다. 각 실험에 두 개의 카테고리에 포함된 200개의 영상데이터가 사용되었고, 카테고리에 따라 영상의 색인의 순서를 정하였다.

본 논문의 실험은 다음과 같은 절차를 따른다. 첫째, 컬러 이미지를 영상처리를 통해 흑백과 일정한 크기의 이미지로 변환하고, Canny Edge 탐지기를 특징점 탐지기로 사용하여 특징점을 추출한다. 둘째, 추출된 특징점을 SIFT 기술자로 요약하여 K-mean를 통해 300개의 Visual Words에 대한 Codebook를 생성한다. 추출된 Codebook으로 각 영상을 Vector 양자화하여 Bag of Word 모델로 표현한다. 셋째, PLSA기법을 적용하여 EM 알고리즘을 통해 학습된 매개변수로 평가영상의 카테고리를 분류하고 카테고리에 해당하는 Visual Words의 히스토그램을 추출하여 차별적인 Visual Words에 대하여 카테고리 간의 차이에 따라 가중치를 부여한다. 마지막으로 가중치 부여된 파라미터를 재평가하여 비슷한 카테고리 간에 차별적인 인식을 하게 된다.

실험과정 중에서 EM 알고리즘에서 초기값이 임의로 설정됨에 실행 시 마다 서로 다른 지역최적화에 수렴하기 때문에 30회에 걸친 반복 실험으로 인식성능을 평가하였다. 각 반복회수에 따른 결과를 살펴보면 수렴된 값이 대부분 1~2%의 작은 범위 내의 변화를 확인할 수 있다.

표 4는 오토바이와 자동차 카테고리에 대한 인식률과 30회의 각 실험에 따른 결과의 표준편차를 보여준다. 이 표에서 확인할 수 있듯이 기존의 PLSA는 실험에 따라 낮은 수준의 결과차이를 보인다. 제안한 가중치 부여 PLSA의 경우 반복실험에 대해 안정된 결과값을 보여준다.

그림 3에서 오토바이와 반지에서 추출된 특징점의 주제별 구분이 명확하지 않고 곡선을 이루는 부분에서 서로 다른 카테고리의 특징점으로 인식되는 것을 보여준다. 특징점을 중심으로 표시된 원은 해당 Visual Word가 대표하는 영역을 보여준다.

표 3은 기존 PLSA와 가중치부여 PLSA를 각 카테고리에 대해 교차실행한 시험결과를 표로 정리한 것이다. 각 실험에 대한 결과를 보면 가중치를 부여한 PLSA는 두 카테고리의 구분이 명확하지 않은 경우에 대해 인식성능개선을 보였다. 그러나 카테고리 간의 시각적 차이가 있어 기존 PLSA에 의해 높은 인식률을 갖는 경우 가중치 부여 PLSA의 성능도 비슷한 결과를 얻었다. 이는 이미 카테고리에 빈번하게 등장하는 Visual Words의 분포가 기존 PLSA로 85% 이상 변별할 수 있는 경우 이미 특이점 간의 구분이

명확하기 때문에 가중치 부여의 영향이 적기 때문이라 보여진다.

표 3에서 오토바이와 자동차의 카테고리 인식실험에서는 기존의 PLSA보다 가중치를 부여한 PLSA가 1.6%의 성능이 향상되었고 오토바이와 비행기의 카테고리 인식실험에서는 1.1%의 성능향상 결과를 보였다. 기존의 PLSA에서 비교적 결과가 낮은 카테고리에서 성능 향상을 보였다. 이유는 공통된 특이점에 대해서는 낮은 가중치를 부여하고 주도적으로 차이를 보이는 특이점에 대해서는 높은 가중치를 부여함으로써 높아진 변별력이 반영된 결과라 할 수 있다.

5. 결론

본 논문에서는 사물의 카테고리를 인식하기 위한 여러 가지 다양한 방법 중 하나인 Bag of Words의 주도적 확률 모델 중 하나인 PLSA에 대하여 가중치를 부여하는 시스템을 구현하였다. 가중치 부여 알고리즘을 적용시켜 보다 정확한 객체의 카테고리 인식이 가능한지를 실험하였다.

객체의 카테고리 인식에 주된 작용을 하는 특징점에 대하여 가중치를 부여함으로써 기존의 PLSA기법으로 카테고리인식률이 낮은 유사한 카테고리 간의 인식성능을 향상시켰다. 제안한 알고리즘은 사물을 인식하고, 유사한 시각적 특징을 갖는 사물들을 보다 효율적으로 인식하기 위한 기법으로 사용될 수 있다. 또한 LDA와 같은 다른 방법을 이용해 가중치 부여하는 연구도 가능할 것이며 동영상에서의 인식이나 장면을 인식시키는 연구로 발전시킬 수 있을 것이다.

참 고 문 헌

- [1] R. Fergus, L. Fei-Fei, P. Perona and A. Zisserman, Learning Object Categories from Google's Image Search. In Proc. ICCV, 2005
- [2] T. Hofmann, Probabilistic latent semantic indexing. In SIGIR, 1999.
- [3] A. Ng, Blei, and M. Jordan, Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993-1022, Jan 2003.
- [4] L. Fei-Fei and P. Perona, A Bayesian Hierarchical Model for Learning Natural Scene Categories. In Proc. CVPR, 2005.
- [5] T. Hofmann, Unsupervised Learning by Probabilistic Latent Semantic Analysis. Machine Learning, 43:177-196, 2001.
- [6] J. Sivic, B. C. Russell, A. Efros, A. Zisserman and W. T. Freeman, Discovering Object Categories in Image Collections. MIT AI Lab Memo AIM-2005-005, MIT, 2005.
- [7] D. Lowe, Local Feature View Clustering for 3D Object Recognition. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, Springer, December 2001.
- [8] G. Csurka, C. Bray, C. Dance and L. Fan, Visual Categorization with Bags of Keypoints. In Workshop on Statistical Learning in Computer Vision, ECCV, 2004.
- [9] D. Lowe, Object Recognition from Local Scale-invariant Features. In Proc. of the 7th International Conference on Computer Vision, Kerkyra, Greece, September 1999.