

## 각도 피라미드를 이용한 Bag-of-features 를 통한 회전에 강한 물체 인식

권보준, 김선아, 이경준, \*윤일동, 이상욱  
서울대학교, \*한국외국어대학교

baukwon@spl.snu.ac.kr sunakim@spl.snu.ac.kr kjoon@spl.snu.ac.kr yun@hufs.ac.kr  
sanguk@spl.snu.ac.kr

### Rotation-invariant Object Categorization using Bag-of-features with Angular Pyramid

Bojun Kwon Seona Kim Kyong Joon Lee \*Il Dong Yun Sang Uk Lee  
Seoul National University \*Hankuk University of Foreign Studies

#### 요 약

본 논문에서는 영상에서의 물체 인식에 주로 사용되었던 공간 피라미드를 변형시킨 각도 피라미드를 이용한 bag-of-features 방법으로 회전 변화에도 강한 물체 인식에도 적용할 수 있도록 한다. 기존의 공간 피라미드에서 수직의 격자 모양으로 영상을 나누었던 것에 비해 각도 피라미드는 영상의 중심을 기준으로 동일한 각도로 영상을 분할하여 피라미드로 만든다. 각 영역 안에서 특징들의 히스토그램으로 영상을 표현하고 영상간의 유사도는 각도 피라미드를 단계별로 순환적 자리옮김을 통해 회전시켜가며 히스토그램 교집합을 구하여 측정한다. 이 방법을 Caltech-101 데이터베이스에 적용해본 결과 회전 변환을 준 테스트 영상에 대해 기존의 공간 피라미드를 사용한 방법에 비해 높은 성능을 보이는 것을 확인하였다. 따라서 이 방법을 통하여 다양한 상황의 일반적인 물체 분류할 수 있을 것으로 기대한다.

#### 1. 서론

영상에서 물체를 인식하고 구별하는 것은 인간에게 있어서 가장 중요하고 기본적인 능력이다. 하지만 여기에는 변형이나 시점 변화, 밝기 변화, 가림 등 기계로 처리하기에는 아직 한계점이 많이 존재하여 많은 연구가 진행되고 있다.

이런 연구 가운데 하나로 bag-of-feature 를 이용하여 물체를 인식하는 방법[1]이 있다. 이 방법에서는 글을 보고 어떤 단어가 주로 쓰이는지 봐도 글의 내용을 추측할 수 있듯이 영상을 국소적인 특징들의 집합으로 간주한다. 즉, 영상에서 어떤 특징들이 주로 나왔는지를 살펴봄으로써 영상 내의 물체를 분류하는 방법이다. 이 방법은 특징점들을 사용하기 때문에 시점 변화나 변형 등에서 자유롭고 영상을 압축적으로 표현할 수 있다. 하지만 특징점들이 물체와 배경 중 어디서 뿔힐 지 알 수 없고 특징점들 사이의 기하학적 모양을 무시하는 것이 단점으로 지적된다.

이를 보완하기 위한 방법이 공간 피라미드를 이용한 bag-of-feature 표현[2]이다. 이 방법에서는 영상을 격자 모양으로 잘라서 피라미드를 형성하고 그 영역 안에서 히스토그램을 만들어 서로 비교한다. 이를 통해 영상을 표현하는데 필요한 차원은 늘어나지만 간접적으로 위치 정보를 사용하여 더 향상된 인식률을 얻을 수 있다.

하지만 입력 영상이 위성 사진이나 항공 사진처럼 일정한

방향으로 정렬되어 있지 않은 경우에는 분류에 실패하기 쉽다. 공간 피라미드의 경우는 같은 영역끼리만 비교하기 때문에 회전된 물체를 잘 분류하지 못하기 때문이다. 따라서 본 논문에서는 이를 해결하고자 각도 피라미드를 도입하여 회전 변환에도 적합한 물체 인식을 구현한다.

#### 2. 각도 피라미드를 이용한 매칭

각도 피라미드는 기존의 공간 피라미드를 회전 변환에 강하게 만들기 위해 변형시킨 것이다. 공간 피라미드에서 영상을 수직의 격자 형태로 나눈다면, 각도 피라미드에서는 영상의 중심을 기준으로 일정한 각도로 나누어서 히스토그램을 만든다.

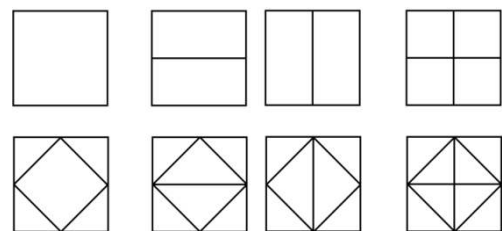


그림 1 L=2, A=4 인 각도 피라미드(위)와 방사 정보를 포함한 각도 피라미드(아래)의 모습. 왼쪽이 0 단계, 가운데 2 개가 1 단계, 오른쪽이 2 단계를 나타낸다. 방사정보를 포함한 경우 0 단계에 왼쪽 아래와 위 영역이 모두 사용된다.

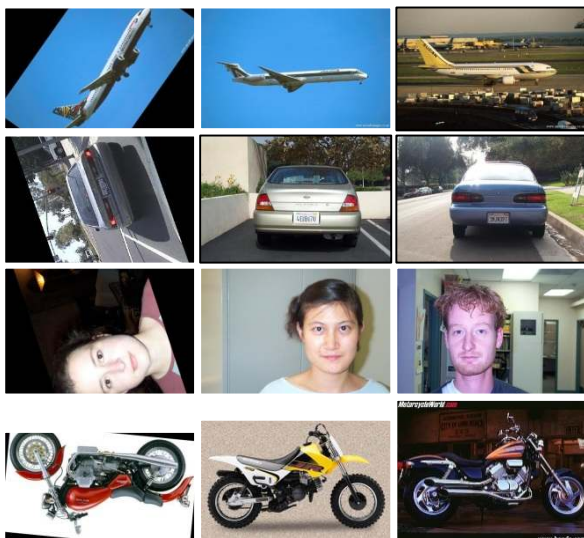
영상을 A 개의 각도로 나누어 L+1 개의 단계로 피라미드를 형성한다고 하면 **그림 1** 과 같이 0 단계는 360°, 1 단계는 180°, 2 단계에서는 90° 로 나뉘어진 영역으로 구성된다. 각 단계는 0 단계를 제외하고 360° / A 의 각도로 회전해야 하기 때문에 A 개의 영역들로 이루어지게 된다. 각각의 영역들에서 추출된 특징들로 히스토그램을 만들게 되므로 필요한 히스토그램은 AL+1 개가 된다. 따라서 K 개의 영상적 단어를 사용한다면 한 장의 영상을 표현하는 것에는 (AL+1)K 차원의 벡터가 필요하다. 여기서 공간 피라미드에 비해 잃어버리게 되는 방사상의 정보를 추가하기 위해 **그림 1** 의 아래처럼 각 면의 중점을 이은 선들로 다시 영상의 영역을 나누었다. 이 경우에는 영상이 총 (2AL+3)K 차원의 벡터로 표현된다.

각각의 히스토그램은 영상의 크기나 위치 별로 뽑힌 특징들의 수를 보정하기 위해 1 로 정규화된다. 이 때 특정 영역에 뽑힌 특징의 수가 너무 적은 경우에는 특징이 없는 지역이라고 보고 균일분포로 채워준다. 마지막으로 각 단계에서 영역의 크기가 커질수록 이상치가 포함될 가능성이 높아지므로 2 단계부터 단계당 2 배의 가중치를 준다.

영상을 각도 피라미드를 이용한 bag-of-feature 로 표현한 후에는 이를 통해 서포트 벡터 머신(support vector machine, SVM)을 훈련시킨다. 이 때 커널로 히스토그램 교집합이 사용되는데 회전 변환 아래서도 매칭을 하기 위해 구해놓은 A 개의 각도에 대한 히스토그램을 순환적으로 자리를 옮겨가면서 교집합을 구한 후 최대값을 두 벡터간의 커널 함수값으로 사용한다.

### 3. 실험 결과

실험에는 Caltech-101 [3] 데이터베이스에서 비행기, 자동차, 얼굴, 오토바이에서 종류별로 100 개의 영상을 골라 훈련에, 30 개의 영상을 0~360° 로 무작위적으로 회전시켜 테스트에 사용하였다. 특징점 추출과 기술자로는 SIFT 검출기와 기술자를 사용하였고 영상적 사전은 200 개의 단어들로 k-평균 군집화 알고리즘을 통해 구성되었다.



**그림 2** 실험에 사용된 데이터베이스. 각 행들이 물체 분류를 나타내고 왼쪽 열은 테스트로 사용한 회전된 영상, 가운데와 오른쪽 열이 훈련에 사용된 회전되지 않은 영상들이다.

L	공간 피라미드	각도 피라미드	각도 피라미드 & 방사 정보
0	77.5	77.5	77.5
1	<b>82.5</b>	84.2	85.0
2	75.0	<b>88.3</b>	<b>85.8</b>
3	-	85.8	70.0

**표 1** 사용한 피라미드와 단계에 따른 분류 정확도를 퍼센트로 나타낸 표. L 단계에서 공간 피라미드는 (L+1)×(L+1)개의 영역을 만들고 각도 피라미드는 2<sup>L</sup>개의 영역을 만든다.

물체	1	2	3	4
1	96.7	3.3	0.0	0.0
2	6.7	80.0	6.7	6.7
3	3.3	0.0	86.7	10.0
4	6.7	0.0	3.3	90.0

**표 2** 테스트 결과에 따른 오분류표. 물체 1은 airplane, 2는 car, 3은 face, 4는 motorbike 를 나타낸다. (i, j)번째 값은 i 번째 물체가 j 번째 물체로 분류된 경우를 퍼센트로 나타낸 것이다.

실험 결과는 **표 1** 과 같다. 공간 피라미드를 사용하는 경우(82.5%)보다 각도 피라미드를 사용하는 경우에 88.3%로 성능이 향상된다. 각도 피라미드에 방사정보를 추가한 경우에는 성능이 향상되지 못하거나 저하된다. 이는 [2]에서 피라미드 단계를 높였을 때 발생한 결과와 마찬가지로이다. 공간을 너무 잘게 분할하여 의미 있는 정보를 잃어서 높은 레벨에서 매칭이 제대로 수행되지 않았지만, 가중치가 높아 전체적으로 잘못된 값이 들어가기 때문이다. **표 2** 는 가장 좋은 결과를 보인 각도 피라미드에 L=2인 경우의 오분류표이다.

### 4. 결론

본 논문에서는 기존의 공간 피라미드를 변형하여 회전 변화에도 변하지 않는 물체 인식을 위하여 각도 피라미드를 제안하였다. 이를 통해 회전된 데이터에 적용해본 결과 기존의 방법보다 성능이 향상된 것을 볼 수 있었다. 이 방법은 좀 더 일반적인 상황에서 물체 분류에 이용될 수 있을 것이다.

### 5. 참고문헌

[1] G.Dorko and C. Schmid. Object class recognition using discriminative local features. PAMI, 2005.

[2] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006.

[3] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In IEEE CVPR Workshop on Generative-Model Based Vision, 2004. [www.vision.caltech.edu/Image\\_Datasets/Caltech101](http://www.vision.caltech.edu/Image_Datasets/Caltech101)