

# 지역 특징을 이용한 효과적인 다중 객체 인식 방법\*

김대훈, 황인준  
 고려대학교 전자전기전파공학부  
 e-mail : {kdh812, ehwang04}@korea.ac.kr

## Efficient Multi-Object Recognition Scheme using Local Features

Daehoon Kim, Eenjun Hwang  
 School of Electrical Engineering, Korea University

### 요 약

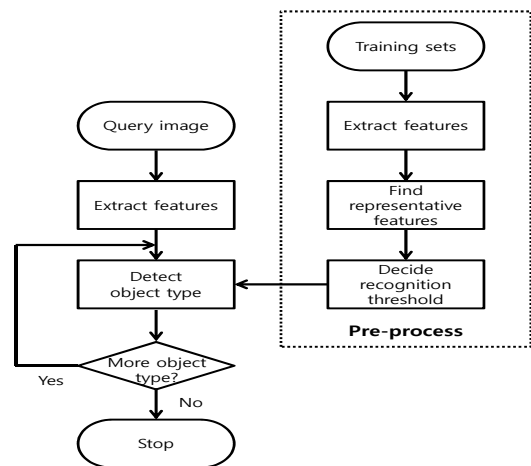
최근 스마트 폰을 비롯한 휴대용 IT 기기의 급속한 보급으로 사용자에게 유익한 서비스를 제공하는 다양한 응용 프로그램에 대한 관심이 증대되고 있다. 특히, 이러한 기기들에 기본적으로 장착되는 카메라를 통해 입력된 영상에 존재하는 객체를 효과적으로 인식할 수 있다면, 사용자에게 더욱 다양하고 유용한 서비스를 제공하는 것이 가능하다. 하지만 휴대용 기기의 경우 아직까지 성능의 제약으로 인해, 기존에 나와있는 다양한 객체 인식 방법을 그대로 적용하기에는 무리가 있다. 본 논문에서는 영상에 존재하는 객체들의 대표 지역 특징 표현자를 검출하여, 완전히 일치하는 객체뿐 아니라 유사한 객체의 경우에도 인식이 가능한 효과적인 다중 객체 인식 알고리즘을 제안하고 프로토타입 시스템을 통하여 그 성능을 평가한다.

### 1. 서론

성공적인 객체 인식 방법은 같은 객체들간의 다양성을 보장함과 동시에 다른 종류의 객체간에 존재하는 차이점을 구별할 수 있는 특성을 동시에 지녀야 한다. 이러한 특성을 보장하기 위해 이미 다양한 연구가 진행 되어 왔다. 이 중에서 bag of words 모델에 기반한 최근의 방법들은 매우 인상적인 객체 인식 및 분류 성능을 보여준다[1][2][3]. 한편, 이미지의 대표 조각들을 검출하고 이를 단계별로 구조화하여 객체 인식에 사용하는 방법도 제안 되었다[4]. 하지만 이러한 방법들은 객체 인식에 있어서는 효과적이지만, 계산할 내용이 많기 때문에 이동형 소형 기기에서 사용하기에는 적합하지 않다.

본 논문에서는 지역 특징 표현자를 추출하고 이를 기반으로 이미지뿐 아니라 동영상에까지 적용 가능한 효과적인 객체 인식 알고리즘을 제안한다. 전처리 단계에서는, 객체 종류별 이미지로부터 지역 특징을 추출하여 각 객체의 특징이라고 생각되는 부분을 자동으로 판별한다. 그리고 이 특징들을 통합하여 질의 과정에서 효과적으로 이용한다. 질의 과정에서는 특징 부분 후보들의 위치를 먼저 파악한 후에 이 부분들과 전처리 과정에서 학습된 정보를 비교하여 객체 인식을 하게 된다.

논문에서는 미리 관심 객체의 이미지 집합에 대해 대표 부분을 추출하고 이를 추후에 이용 가능하도록 정보를 가공하여 학습시키는 전처리 과정을 진행한다. 한편 질의 처리를 위해서는 주어진 질의 영상에 대해, 전처리 과정과 유사하지만 간소화된 과정을 수행하여 지역 특징들을 추출하고 전처리 과정에서 생성된 정보와 비교하여 객체의 존재 여부 및 종류를 판별하게 된다. 그림 1 은 본 논문에서 제안하는 전체 시스템의 흐름을 보여준다.



(그림 1) 전체 순서도

### 2. 지역 특징 표현자를 사용한 객체 인식

여러 종류의 객체를 신속하게 인식하기 위해서, 존

전처리 단계는 실제 이미지 질의 과정에서 필요한 여러 가지 정보를 미리 계산하여 놓는 과정이다. 가장 먼저 특징점 추출 및 표현자 생성 알고리즘을 사

\* 이 논문은 2010 년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2010-0025395).

용하여 각 종류별 이미지들의 특징을 추출하는 작업을 수행한다. 관심 지점 검출 및 표현자 생성을 위한 기법은 다양한데, 본 논문에서는 SIFT 기법[5]에 비해 정합 속도가 매우 빠르고 비슷한 정합 성능을 보이는 SURF 기법[6]을 사용하여, 훈련 데이터의 각 이미지 별로 표현자를 생성하고 대표 특징을 검출한다. 그림 2는 SURF 알고리즘에 의한 두 이미지간에 유사하다고 판단되는 부분을 보여주는 예제이다. 대표 특징 검출은 전 단계에서 검출한 관심 지점들 (interest points) 중에서 어떤 관심 지점이 해당 객체를 대표할 만한 부분인가를 판별해 주는 단계이다. 각 관심 지점 별로 해당 부분과의 유사도가 일정한 값 이상을 가진 부분들을 찾아내고 이러한 부분들의 계수가 임계 값을 넘으면 해당 부분을 대표 지점으로 결정한다. 객체 별로 대표 지점들을 검출하는 작업이 끝나게 되면, 트레이닝 셋으로 추가 학습을 통해 인식에 필요한 계수들을 자동으로 선택하는 과정을 거치게 된다. 이렇게 하면 실제 질의 처리에 필요한 전처리 과정이 끝나게 된다.



(그림 2) SURF 기법을 사용한 대표 특징 검출 예제

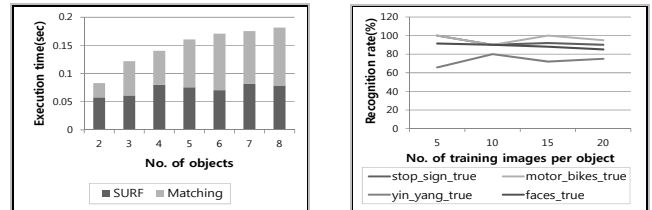
질의 처리 단계에서는 이러한 특징들을 이용하여 질의 이미지가 어떠한 객체를 포함하고 있는지를 결정한다. 질의 처리 단계는 전처리 과정과 유사하지만 중간에 몇 가지 과정이 생략 되어있다. 먼저 질의 이미지에 대해 관심 지점 및 범위 검출을 수행하여 관심 지점을 추출한다. 그 후에 각 관심 지점에 대해 전처리 과정에서 얻어진 대표 특징들과 비교를 하여 대표 특징들과 현재 질의 이미지의 지점들이 얼마나 일치하는지를 계산한다. 질의 이미지의 관심 지점들에 대해, 어떤 객체의 분류 중 일정 비율 이상의 분류가 존재한다고 판별되면 최종적으로 질의 이미지는 해당 객체를 포함하고 있다고 판별한다.

### 3. 실험 결과

본 실험에 사용된 훈련 이미지 셋은 Caltech101[25]로서, 총 8500 여개의 이미지가 101 개의 분류로 나뉘어 있으며 각 분류 당 최소 50 개의 다양한 해상도와 다양한 객체의 크기를 가진 이미지로 구성되어있다. 실험은 Intel Core 2 Duo 2.67Ghz, 4GB 메모리 사양의 Windows 7 환경에서 수행되었으며, 전처리 및 질의 처리를 포함한 모든 과정은 MATLAB 을 기반으로 작성 되었다.

그림 3은 각각 객체 인식 시간과 인식률에 대한 실험의 결과이다. 그림 3(a)는 최대 여덟 개의 객체에 대해 객체 인식에 걸리는 평균 시간을 측정한 것으로, SURF 와 Matching 라고 표기 되어있는 그래프는 각각

SURF 특징 추출과 객체 인식에 걸린 시간이다. SURF 특징 추출은 객체 이미지의 복잡도에 따라 약간씩 변하지만 거의 일정한 시간이 걸리는 것을 알 수 있다. 그에 비해 객체수가 늘어날 수록 객체 인식에 걸리는 시간은 늘어나고 있다. 하지만 전체 인식 시간에서 객체 인식 시간이 차지하는 비율을 살펴보면, 효과적으로 객체 인식이 가능한 것을 확인할 수 있다. 그림 3(b)는 네 개 객체에 대해 각각의 인식 정확도를 측정 한 실험으로, 트레이닝 이미지의 숫자가 많지 않더라도 일정한 수준 이상의 정확도를 유지할 수 있음을 보여준다.



(a) 검색 시간

(b) 검색 정확도

(그림 3) 검색 결과

### 4. 결론

본 논문에서는 지역 특징 표현자를 사용한 효과적인 객체 인식 방법을 제안하였다. 제안된 기법은 검출된 관심 지점 및 범위를 기반으로 대표 조각을 결정하고, 실제 질의 처리 때 이 정보를 효과적으로 이용하여 효과적인 객체 인식이 가능하다. 특히 기존의 특징 추출 기술에 비해 가볍고 간단한 특징 추출 기술을 이용하여 객체 인식 시 필요한 비용도 줄일 수 있다. 실험을 통하여 제안된 기법이 성능의 제약이 따르는 휴대용 기기에서도 효과적으로 객체인식을 수행할 수 있음을 보였다.

### 참고문헌

- [1] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints", IN WORKSHOP ON STATISTICAL LEARNING IN COMPUTER VISION, ECCV, 2004, p. 1--22.
- [2] J. Sivic, B.C. Russell, A.A. Efros, A. Zisserman, and W.T. Freeman, "Discovering object categories in image collections", In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2005.
- [3] L.J. Li and L. Fei-Fei, "Optimol: automatic online picture collection via incremental model learning", International Journal of Computer Vision, vol. 88, 2010, p. 147- 168.
- [4] S. Ullman, "Object recognition and segmentation by a fragment-based hierarchy", Trends in Cognitive Sciences, vol. 11, 2007, p. 58-64.
- [5] D. Lowe, "Object recognition from local scale-invariant features", In Proc. International Conference on Computer Vision, 1999.
- [6] S. Ullman, "Object recognition and segmentation by a fragment-based hierarchy", Trends in Cognitive Sciences, vol. 11, 2007, p. 58-64.