

Design of the 3D Object Recognition System with Hierarchical Feature Learning

Joohee Kim[†] · Dongha Kim[†] · Incheol Kim^{**}

ABSTRACT

In this paper, we propose an object recognition system that can effectively find out its category, its instance name, and several attributes from the color and depth images of an object with hierarchical feature learning. In the preprocessing stage, our system transforms the depth images of the object into the surface normal vectors, which can represent the shape information of the object more precisely. In the feature learning stage, it extracts a set of patch features and image features from a pair of the color image and the surface normal vector through two-layered learning. And then the system trains a set of independent classification models with a set of labeled feature vectors and the SVM learning algorithm. Through experiments with UW RGB-D Object Dataset, we verify the performance of the proposed object recognition system.

Keywords : Feature Learning, RGB-D Images, Object Recognition, Attribute Recognition, Surface Normal Vector

계층적 특징 학습을 이용한 3차원 물체 인식 시스템의 설계

김 주 희[†] · 김 동 하[†] · 김 인 철^{**}

요 약

본 논문에서는 계층적 특징 학습을 이용하여 물체의 컬러 영상과 깊이 영상으로부터 해당 물체가 속한 범주와 개체, 그리고 다양한 속성들을 효과적으로 인식할 수 있는 시스템을 제안한다. 본 시스템의 전처리 단계에서는 물체의 깊이 영상을 물체의 모양 정보를 좀 더 효과적으로 표현할 수 있는 표면 법선 벡터 데이터로 변환하고, 특징 학습 단계에서는 물체의 컬러 영상과 표면 법선 벡터 데이터로부터 두 단계에 걸쳐 패치 단위 특징과 이미지 단위의 특징을 추출해낸다. 그리고 추출된 특징 벡터들과 SVM 학습 알고리즘을 이용하여 각기 독립적인 다수의 분류 모델들을 학습한다. 미국 워싱턴 대학의 RGB-D 물체 데이터 집합을 이용한 실험을 통해, 본 논문에서 제안하는 물체 인식 시스템의 높은 성능을 확인할 수 있었다.

키워드 : 특징 학습, RGB-D 영상, 물체 인식, 속성 인식, 표면 법선 벡터

1. 서 론

최근 다양한 영상 센서들과 영상 인식 기술의 발전에 힘입어, 컬러 영상(Color Image)과 깊이 영상(Depth Image)을 이용한 3차원 물체 인식(Object Recognition) 연구[1, 2, 3]와 물체의 속성 인식(Attribute Recognition) 연구[4, 5]가 매우 활발하다. 특히 RGB-D 영상 기반의 물체 인식과 속성 인식

기술은 물체의 컬러 영상과 깊이 영상을 함께 분석함으로써, 물체의 범주(Category)나 개체(Instance), 그리고 물체의 색상(Color)이나 모양(Shape), 재질(Material) 등의 속성을 판별해내는 기술을 말한다. 지능 로봇이나 자율 주행과 같은 많은 응용 분야에서는 카메라 영상에 포착된 물체들에 대한 심층 이해를 위해서 물체가 속한 범주나 개체 정보 외에도, 물체의 색상, 모양, 재질 등과 같은 속성 정보들을 파악하는 작업도 매우 중요한 역할을 한다.

물체의 컬러 영상은 일반적으로 복수 채널의 RGB 값을 갖기 때문에 단일 채널의 깊이 영상보다는 정보가 풍부하지만, 같은 범주의 물체라도 다른 무늬이거나 색상 차가 클 경우 인식 성능이 낮아질 수 있다. 또한 기하학 정보가 포함되어 있지 않는 2차원 컬러 영상만으로는 실세계에 존재하는 3차원 물체들을 판별해내는 일은 쉽지 않다. 따라서

※ 본 연구는 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음.
※ 이 논문은 2015년도 한국정보처리학회 춘계학술발표대회에서 '계층적 특징 학습을 이용한 3차원 물체 인식'의 제목으로 발표된 논문을 확장한 것임.

[†] 준 회원 : 경기대학교 컴퓨터과학과 석사과정

^{**} 종신회원 : 경기대학교 컴퓨터과학과 교수

Manuscript Received : June 22, 2015

First Revision : October 26, 2015

Accepted : November 14, 2015

* Corresponding Author : Incheol Kim(kic@kyonggi.ac.kr)

물체 표면의 깊이 정보를 포함하고 있는 깊이 영상을 컬러 영상과 함께 사용한다면 3차원 물체 인식에 유리할 수 있다. 다만, 깊이 영상은 특정 위치에 놓인 깊이 센서에서 물체 표면까지의 단순 거리 측정치만을 포함하고 있기 때문에, 센서 위치에 독립적으로 해당 물체의 3차원 모양 정보를 표현하기에는 부족함이 있다. 따라서 깊이 영상을 물체의 모양 정보를 직접 표현할 수 있는 표면 법선 벡터(Surface Normal Vector) 데이터로 변환 후, 인식에 이용한다면 물체 인식 성능을 한층 더 높일 수 있을 것이다.

그동안 물체 인식 분야에서 주로 이용되어온 영상 특징들(Image Features)은 크게 사람에 의해 미리 설계되어진 특징들(Human-Designed Features)과 학습을 통해 자동으로 추출되는 특징들(Learned Features)로 나눌 수 있다. 사람에 의해 미리 설계되어진 특징들은 설계된 특정 인식 대상들이나 인식 데이터에 대해서는 높은 인식 성능을 얻을 수 있다. 하지만, 새로운 인식 대상들이나 인식 데이터들에 대해서는 그 특징들을 그대로 적용할 수 있는 가능성도 낮고 높은 인식 성능도 기대하기 어렵다는 한계가 있다. 또한, 사람에 의해 미리 설계된 방식으로 추출된 특징 벡터는 대부분 저수준(low-level)의 특징 정보들을 제공하는 반면, 많은 물체 인식 문제들에서는 보다 고수준(high-level)의 특징 정보들을 요구하는 경우가 많다. 물체 인식을 위해 미리 설계된 대표적인 영상 특징들로는 SIFT(Scale Invariant Feature Transform)[6]나 HOG(Histogram of Oriented Gradients)[7], FPFH(Fast Point Feature Histogram)[8] 등이 있다. 반면, 학습을 통해 자동으로 추출되는 특징들은 특별히 특정 인식 대상이나 인식 데이터 유형에 국한되지 않고 적용이 가능하므로, 확장성과 일반성이 뛰어나다. 또한, 특징 학습을 여러 단계들에 거쳐 계층적으로 수행함으로써 물체에 대한 저수준의 특징 정보들뿐만 아니라 고수준의 특징 정보들도 함께 얻을 수 있다는 장점이 있다. 학습을 통해 자동으로 물체 인식을 위한 특징들을 추출한 대표적인 사례들로는 CNN(Convolutional Neural Network)[9], HMP(Hierarchical Matching Pursuit)[10], 커널 서술자(Kernel Descriptors)[11] 등이 있다.

본 논문에서는 계층적 특징 학습을 이용하여 물체의 컬러 영상과 깊이 영상으로부터 해당 물체가 속한 범주와 개체, 그리고 다양한 속성들을 효과적으로 인식할 수 있는 시스템을 제안한다. 본 시스템에서는 전처리 단계를 통해 물체의 깊이 영상을 물체의 모양 정보를 좀 더 효과적으로 표현할 수 있는 물체 표면의 법선 벡터 데이터로 변환한다. 특징 학습 단계에서는 물체의 컬러 영상과 표면 법선 벡터 데이터로부터 두 단계에 걸쳐 패치 단위 특징들(Patch Features)과 이미지 단위 특징들(Image Features)을 추출해낸다. 그리고 추출된 특징 벡터들과 SVM 학습 알고리즘을 이용하여 각기 독립적인 다수의 분류 모델들을 학습한다. 본 논문에서 제안하는 물체 인식 시스템의 성능을 분석하기 위해, 미국 워싱턴 대학(University of Washington)의 RGB-D 물체 데이터 집합을 이용한 실험을 수행하고, 그 결과를 소개한다.

2. 관련 연구

Silberman의 연구[12]에서는 컬러 영상과 깊이 영상으로부터 SIFT 특징을 추출하여, 물체 인식(Object Recognition)에 이용하였다. SIFT 특징은 코너점 등으로 특징 점을 선택한 후, 특징 점 주변의 지역 패치(Local Patch) 내 픽셀들의 그라디언트(Gradient) 방향과 크기에 대한 히스토그램(Orientation Histograms)을 서술자로 사용하는 특징이다. 이 특징은 크기와 형태, 회전 등의 변화에 강건하면서도 분별력이 뛰어나기 때문에 영상처리 분야에서 매우 널리 사용된다. 또한, 이 연구에서는 입력데이터로 물체의 색상 정보를 포함한 컬러 영상뿐만 아니라 물체의 모양 정보를 포함한 깊이 영상도 함께 사용함으로써, 물체 인식 성능을 향상시켰다. Rusu의 연구[13]에서는 물체 표면의 법선 벡터 데이터로부터 FPFH 특징을 추출하여, RGB-D 영상 내 물체 검출(Object Detection)에 이용하였다. FPFH 특징은 3차원 물체의 법선 벡터 데이터를 계산하여, 각 법선 벡터와 그 이웃의 법선 벡터들의 방향 차를 표현한 특징이다. 이 특징은 3차원 물체의 모양을 효과적으로 표현할 수 있는 물체 표면의 법선 벡터를 사용하였으며, 계산량이 적고 센서의 위치나 방향에 강건한 장점이 있다. 하지만 SIFT 특징과 FPFH 특징은 모두 각각 고유한 인식 문제 유형에 맞게 사람이 미리 설계한 특징들로서, 앞서 언급한대로 새로운 유형의 인식 문제들에서도 높은 성능을 유지하려면 필연적으로 확장과 변경이 필요하다.

한편, Liefeng Bo의 연구[11]에서는 컬러 영상과 깊이 영상으로부터 학습을 통해 커널 서술자(Kernel Descriptors)들을 추출하고 이들을 물체 인식(Object Recognition)에 이용하였다. 커널 서술자 학습에서는 특정 영상 패치의 명암 변화, 컬러, 바이너리 패턴(Binary Pattern) 등의 픽셀 속성을 사용하여 커널을 설계하고, 커널 PCA(Kernel Principle Component Analysis)를 적용하여 기저 벡터를 추출하였다. 이 방법은 RGB-D 영상으로부터 자동으로 저수준의 특징 정보를 학습할 수 있는 가능성은 열어놓았지만, 고수준의 특징 정보를 얻을 수 있는 계층적 특징 학습 방법을 제시하지는 못하였다. Girshick의 연구[14]에서는 심층 학습(Deep Learning)의 하나인 컨볼루션 신경망(Convolutional Neural Network) 학습을 적용하여 컬러 영상으로부터 다양한 수준의 특징들을 자동 추출하고, 이들을 영상 내 물체 검출(Object Detection)에 이용하였다. 컨볼루션 신경망을 이용한 이 특징 학습 방법은 완전 연결 방식의 기존 신경망보다 학습해야 할 파라미터 수가 적어 학습의 효율성도 높고, 다양한 수준의 특징 정보들을 함께 얻을 수 있어 인식 성능을 높일 수 있었다. 하지만 다양한 수준의 특징 학습을 위해 신경망의 은닉 단계(Hidden Layer)들이 늘어나면 학습에 필요한 전체 계산량도 증가하기 때문에, 특징 추출을 위해 많은 계산 시간과 고성능 계산 자원이 요구된다는 문제점이 있다. 또한, Girshick의 연구[14]에서는 특징 학습을 위해 컬러 영상만을

이용하였을 뿐, 물체의 3차원 모양 정보를 포함한 깊이 영상을 이용하지는 못하였다는 한계점이 있다.

3. 인식 시스템 설계

본 논문에서는 계층적 특징 학습을 이용하여 물체의 컬러 영상과 깊이 영상으로부터 해당 물체가 속한 범주와 개체, 그리고 다양한 속성들을 효과적으로 인식할 수 있는 시스템을 제안한다. 제안하는 인식 시스템의 전체 구조는 Fig. 1과 같다.

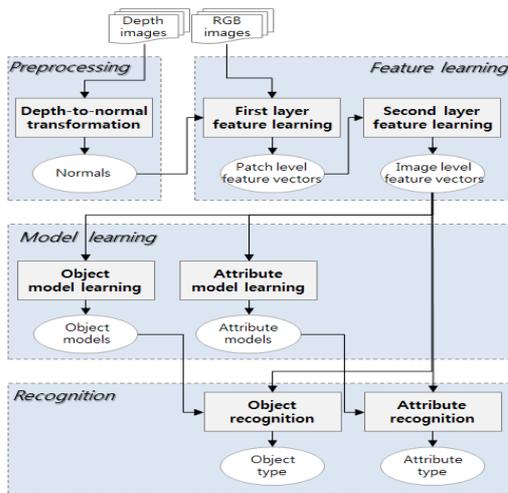


Fig. 1. Overall System Architecture

본 시스템은 크게 전처리(Pre-Processing) 단계, 특징 학습 (Feature Learning) 단계, 분류 모델 학습(Model Learning) 단계, 인식(Recognition) 단계들로 이루어진다. 전처리 단계에서는 입력 영상들 중 깊이 영상(Depth Image)을 물체의 모양 정보를 효과적으로 표현할 수 있는 물체 표면의 법선 벡터(Normal Vector) 데이터로 변환한다. 특징 학습 단계에서는 전처리 결과물로 얻어진 물체 표면의 법선 벡터 데이터와 컬러 영상으로부터 패치 단위 특징들(Patch Features)과 이미지 단위 특징들(Image Features)을 자동으로 학습해낸다. 훈련 단계에서는 표시자(Label)가 붙은 특징 벡터들을 이용하여 인식하고자 하는 물체의 범주, 개체, 색상, 모양, 재질 별로 독립적인 분류 모델들을 학습한다. 인식 단계에서는 학습된 분류 모델들을 새로운 RGB-D 영상의 특징 벡터에 각기 독립적으로 적용함으로써, 해당 영상에 포함된 물체의 범주와 개체, 그리고 색상, 모양, 재질 등과 같은 물체의 속성들을 자동으로 판별해낸다.

3.1 인식 목표

본 시스템의 인식 목표는 물체가 속한 범주(Category)와 개체(Instance), 그리고 색상, 모양, 재질 등 해당 물체의 속성들(Attribute)이다. 각 인식 목표에 해당하는 예시는 Fig. 2와 같다. 범주는 머그컵, 휴대폰, 치약 등 물체의 종류를 말

하고, 개체는 머그컵 1, 머그컵 2와 같이 동일 범주에 속한 각기 독립적인 물체명이나 식별번호를 말한다. 속성은 각 물체의 색상이나 모양, 재질 등과 같이 해당 물체의 특성을 설명할 수 있는 요소들이다.

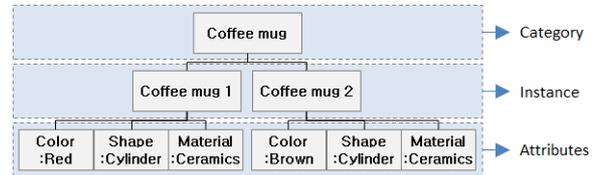
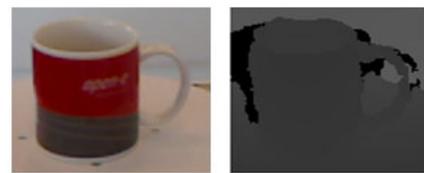


Fig. 2. Categories, Instances, Attributes

본 시스템에서는 물체에 관한 서로 다른 인식 목표를 달성하기 위해 각기 독립적인 분류 모델들을 이용하지만, 분류 모델 학습과 검증에 필요한 특징 벡터들은 3.3절에서 소개할 동일한 특징 학습 방법을 적용하여 추출해낸다.

3.2 전처리

제안하는 물체 인식 시스템의 입력 데이터는 Fig. 3의 (a)와 같은 한 물체의 컬러 영상과 (b)와 같은 깊이 영상의 쌍들로 구성된다. 이와 같이 각 물체의 동기화된 컬러 영상과 깊이 영상들은 마이크로소프트사(Microsoft)의 Kinect와 같은 RGB-D 센서들을 이용함으로써 얻을 수 있다.



(a) Color image (b) Depth image

Fig. 3. Input Images

전처리 단계에서는 물체의 깊이 영상을 3차원 물체 모양을 좀 더 효과적으로 표현할 수 있는 물체 표면의 법선 벡터 데이터로 변환하는 Fig. 4와 같은 변환 과정을 수행한다.

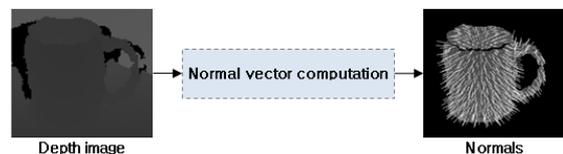


Fig. 4. Preprocessing

단일 채널로 이루어진 물체의 깊이 영상은 특정 위치의 깊이 센서에서 측정된 물체 표면까지의 거리 정보만을 포함하고 있는 데이터이다. 따라서 하나의 물체 깊이 영상은 깊이 센서와 물체가 놓인 상대적 위치에 강한 의존성을 갖기 때문에, 물체 고유의 3차원 모양 정보를 효과적으로 나타낸다고 보기 어렵다. 따라서 본 시스템에서는 보다 효과적으로 3차원 물체를 인식하기 위해 센서의 위치 변화에 강건한

면서 좀 더 풍부한 물체의 모양 정보를 포함할 수 있는 표면 법선 벡터 데이터로 깊이 영상을 변환한다. 깊이 영상으로부터 얻어지는 3차원 물체의 표면 법선 벡터 데이터는 Fig. 4의 우측 영상의 예와 같이 물체의 표면 기울기 변화를 잘 나타낸다.

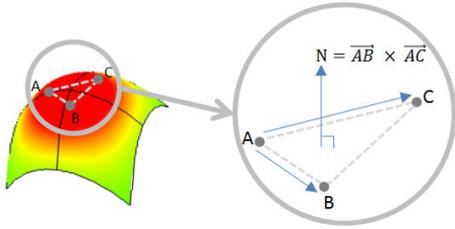


Fig. 5. Normal Vector Computation

물체의 표면 법선 벡터는 Fig. 5와 같이 물체 표면 위의 이웃한 세 점 A, B, C을 이용하여 평행하지 않는 두 벡터 \overrightarrow{AB} , \overrightarrow{AC} 를 각각 구하고, 이 두 방향 벡터들의 외적(Cross Product)을 계산함으로써 얻을 수 있다. 본 논문에서는 먼저 센서 위치 정보와 그 위치에서 측정된 깊이 영상을 토대로 3차원 물체의 표면을 나타내는 점군(Point Cloud) 데이터를 얻은 후, 점군 상의 이웃한 3 개의 점들을 이용하여 Fig. 5와 같은 방법으로 해당 물체의 표면 법선 벡터 데이터를 계산해낸다.

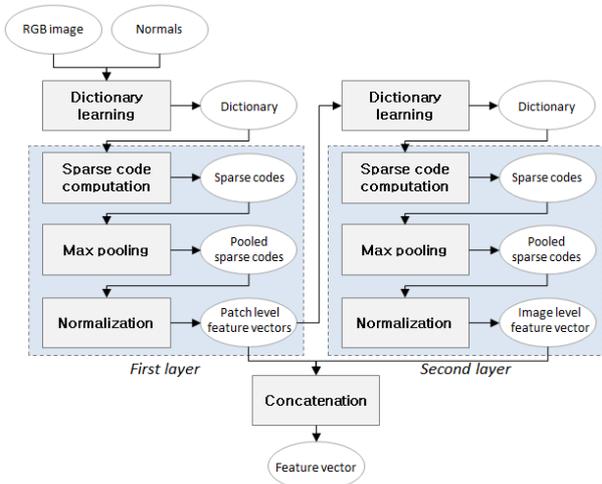


Fig. 6. Feature Learning

3.3 특징 학습

특징 학습 단계에서는 물체의 컬러 영상과 표면 법선 벡터들에 Fig. 6과 같은 계층적인 특징 학습 과정을 적용함으로써 물체 인식을 위한 효과적인 특징 벡터들을 얻어낸다. 특징 학습 과정은 크게 패치(Patch) 단위의 특징들을 학습하는 단계와 이미지(Image) 단위의 특징들을 학습하는 단계 등 총 두 단계로 구성된다. 그리고 각 단계는 다시 사전 학습(Dictionary Learning), 희소 코드 계산(Sparse Code Com-

putation), 피라미드 최대값 통일(Pyramid Max Pooling), 정규화(Normalization)와 같은 세부 과정들로 구성된다.

1) 사전 학습(Dictionary Learning)

본 시스템에서는 물체 표면의 법선 벡터 데이터와 컬러 영상에서 샘플링한 5x5 크기의 영상 패치들에 K-SVD 알고리즘을 적용함으로써, 각 영상 패치들을 희소 코드화(Sparse Coding)하기 위한 사전(Dictionary)을 학습한다. K-SVD (Singular Value Decomposition) 알고리즘은 아래의 Equation (1)과 같이 재현 오류(Reconstruction Error)를 최소화할 수 있는 사전 $D=[d_1, \dots, d_m, \dots, d_M]$ 와 희소 코드 벡터 $X=[x_1, \dots, x_n, \dots, x_N]$ 를 관측 데이터 행렬 Y로부터 학습해낸다.

$$\min_{D, X} \|Y - DX\|^2 \tag{1}$$

2) 다단계 특징 학습(Multi-Layer Feature Learning)

위에서 설명한 바와 같이 사전 학습이 이루어지고 나면, 이를 토대로 총 두 단계로 구성된 본격적인 특징 학습 과정이 시작된다. 특징 학습 첫 번째 단계(First Layer)에서는 앞서 학습한 사전을 이용하여 패치 단위의 특징들을 추출하고, 두 번째 단계(Second Layer)에서는 첫 번째 단계의 결과물을 토대로 이미지 단위의 특징들을 추출해낸다. 그리고 각 단계는 다시 Fig. 6과 같이 희소 코드 계산(Sparse Coding), 피라미드 최대값 통일(Pyramid Max Pooling), 정규화(Normalization)와 같은 세부 학습 과정들로 구성된다. 한편, 두 번째 단계에서 사용할 사전은 첫 번째 단계의 학습 결과로 얻어진 특징 벡터들을 이용해 학습해낸다.

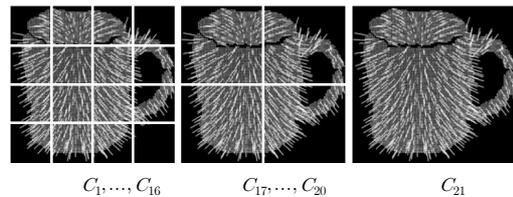


Fig. 7. Spatial Cells of Image Patch

Fig. 8은 특징 학습의 각 단계에서 생성되는 특징 벡터들의 길이와 세부 구성을 나타내고 있다. 특징 학습의 첫 번째 단계에서는 16x16 패치 단위의 특징 벡터를 생성한다. 이를 위해 먼저 패치 내의 모든 픽셀에 대해 해당 픽셀 주변 5x5의 이웃 픽셀들을 데이터 값을 이용해 d차원의 희소 코드를 계산한다(Sparse Coding). 그런 다음, 피라미드 최대값 통일(Pyramid Max Pooling) 과정을 수행함으로써 각 픽셀의 희소 코드들을 이용해 패치 단위의 특징 벡터를 생성한다. 피라미드 최대값 통일 과정에서는 Fig. 7과 같이 영상 패치를 여러 레벨의 공간 셀(Cell)들로 나눈 다음, Equation (2)와 같이 각 셀 C에 포함된 모든 희소 코드 벡터의 j번째 요소들 x_{jm} 중 최대값 $\max_{j \in C} |x_{jm}|$ 을 찾아 이것으로

이들을 통합한다.

$$F(C) = [\max|x_{j1}|, \dots, \max|x_{jd}|, \dots, \max|x_{jd}|] \quad (2)$$

그리고 각 셀별로 구해진 희소 코드 벡터 $F(C)$ 들을 (식 3)과 같이 결합함으로써, 패치 단위의 특징 벡터 F_p 를 생성한다.

$$F_p = [F(C_1^p), \dots, F(C_{21}^p)] \quad (3)$$

일반적으로 조명(Illumination) 변화와 겹침(Occlusion) 현상 등으로 인해 희소 코드 값 크기의 편차가 크게 나타날 수 있다. 이러한 환경 변화에 강건한 특징 벡터를 학습하기 위해, 패치 단위의 특징 벡터 F_p 를 L2 정규화 $\sqrt{\|F_p\|^2 + \epsilon}$ 으로 정규화(Normalization)한다.

특징 학습 두 번째 단계에서는 첫 번째 단계에서 구해진 패치 단위의 특징 벡터들에 대해 다시 희소 코드 계산(Sparse Code Computation), 피라미드 최대값 통일(Pyramid Max Pooling), 정규화(Normalization) 과정들을 차례대로 적용함으로써, 이미지 전체를 나타내는 특징 벡터를 생성한다. 특징 학습 두 번째 단계에서 이용할 사전(Dictionary)은 첫 번째 단계에서 구해진 패치 단위의 특징 벡터들을 샘플링하여 이들로부터 학습한다.

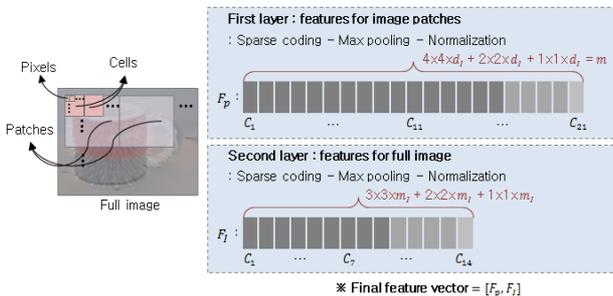


Fig. 8. Feature Vectors

분류 모델 학습과 인식에 이용할 최종 특징 벡터는 Fig. 8과 같이, 첫 번째 단계에서 구해진 특징 벡터와 두 번째 단계에서 구해진 특징 벡터를 하나로 결합(Concatenation)함으로써 완성된다.

3.4 분류 모델 학습과 인식

분류 모델 학습 단계에서는 부류 표시(Class Label)가 부여된 각 물체의 특징 벡터 집합에 선형 SVM(Linear Support Vector Machines) 알고리즘을 적용하여, 범주(Category), 개체(Instance), 색상(Color), 모양(Shape), 재질(Material) 등 인식 목표가 다른 총 5 개의 분류 모델들을 독립적으로 학습한다. SVM 알고리즘은 이미 문자 인식, 얼굴 인식, 물체 인식 등 다양한 분야의 인식 문제들에서 매우 높은 성능을 보여준 학습 알고리즘이다. 특히 SVM 알고리즘은 특징 공간(Feature Space)에서 부류 간 여백을 최대화(Margin Maximization)할 수 있는 최적의 분류 경계면(Optimal Hyperplane)을 찾

아춤으로써, 과적합 문제(Overfitting)를 피할 수 있고 일반화 능력(Generalization)이 뛰어난 알고리즘으로 알려져 있다. 본 시스템에서는 3.3절에서 소개한 특징 학습 방법으로 추출한 특징 벡터들과 SVM 학습 알고리즘을 공통적으로 이용하여, Fig. 9와 같이 각기 다른 총 5 개의 분류 모델들을 독립적으로 학습한다.

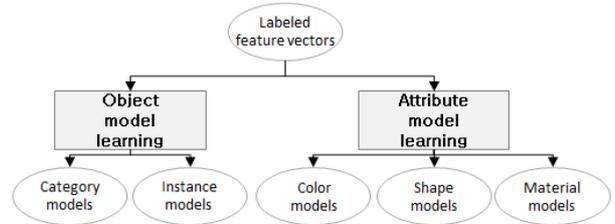


Fig. 9. Category, Instance, Attribute Models

인식 단계에서는 학습된 분류 모델들을 새로운 RGB-D 영상의 특징 벡터에 각기 독립적으로 적용함으로써, 해당 영상에 포함된 물체의 범주와 개체, 그리고 색상, 모양, 재질 등과 같은 물체의 속성들을 자동으로 판별해낸다.

4. 실험 및 평가

4.1 실험 데이터 집합

본 논문에서 제안하는 물체 인식 시스템의 성능을 분석하기 위해, MATLAB을 이용하여 제안 시스템을 구현하고 성능 실험을 수행하였다. 본 실험에서는 미국 워싱턴대학(University of Washington)에서 제공하는 RGB-D 물체 데이터 집합[15]을 사용하였다.

Table 1. Dataset

Cellphone	Coffee mug	Soda can	Stapler	Toothpaste

RGB-D 물체 데이터 집합 중에서 Table 1과 같이 일상 생활에서 흔히 볼 수 있는 휴대폰, 머그컵, 음료캔, 스탬플러, 치약 등 총 5개의 물체 범주(Category)와 이 범주에 속한 총 32개의 물체 개체(Instance)들의 RGB-D 영상 데이터들을 실험에 이용하였다.

Table 2. Attributes

Attribute	Words
Color	Red, Yellow, Green, Black, Blue
Shape	Rectangle, Cylinder, Circle, Stick
Material	Ceramic, Plastic, Paper, Food, Metal

또한, Table 2와 같이 색상, 모양, 재질 등의 물체 속성을 총 14개의 속성 값에 따라 나누어 속성 인식 실험에 이용하였다. 각 입력 영상의 크기는 320×240 픽셀(Pixel)로 이루어져있고, 본 실험에서는 마스크(Mask)를 이용해 각 영상에 포함된 배경 부분을 제거하고 물체 부분만을 잘라내어 실험에 이용하였다. 또한, 본 실험에서는 성능 평가를 위해 전체 데이터 집합의 2/3는 분류 모델 학습을 위한 훈련용(Training)으로, 1/3은 인식 테스트를 위한 검증용(Testing)으로 사용하였다.

4.2 물체 범주 인식 실험

첫 번째 실험은 본 인식 시스템의 물체 범주 인식 성능을 분석하기 위한 실험이었다. 이 실험에서는 본 논문에서 적용한 특징 학습 방법(HMP, Hierarchical Matching Pursuit)의 효과를 확인하기 위해, 사람이 설계한 대표적인 특징 추출 방법들인 SIFT, FPFH 등과 비교 성능 실험을 진행하였다. 또한, 특징 학습 및 추출에 컬러 영상만을 이용한 경우(RGB), 깊이 영상만을 이용한 경우(Depth), 표면 법선 벡터를 이용한 경우(Normals), 본 논문에서 제안한 방식처럼 컬러 영상과 표면 법선 벡터를 함께 이용한 경우(RGB+Normals)의 인식 성능들을 서로 비교함으로써, 3차원 물체 범주 인식에 미치는 효과를 분석해보았다. 분류 모델 생성을 위한 학습 알고리즘으로는 SVM[16]과 K-NN(Nearest Neighbors)를 적용하였다.

Table 3. Comparison of Category Recognition Accuracy

	Features	Models	
		SVM	K-NN
Human-designed features	SIFT - RGB	36.5%	69.78%
	SIFT - Depth	69.8%	66.8%
	FPFH - Normals	91.8%	78.1%
Learned features	HMP - RGB	89.3%	82.5%
	HMP - Depth	94.1%	40.4%
	HMP - Normals	94.5%	83.3%
	HMP - RGB+Normals	90.6%	98.0%

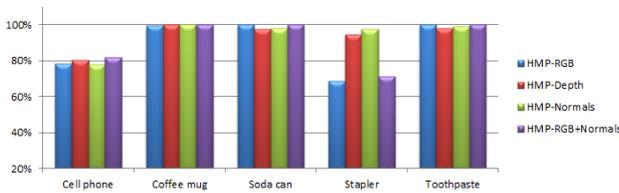


Fig. 10. Category Recognition Accuracy for Each Category

Table 3은 물체 범주 인식 성능 비교 실험 결과를 보여준다. 먼저, 특징 학습 및 추출 방법에 따른 인식 성능을 확인해 보면, 사람이 미리 설계한 특징 추출 방법들(SIFT, FPFH)에 비해, 학습을 통해 자동으로 특징을 추출한 경우(HMP)의 범주 인식 성능이 훨씬 더 높다는 것을 확인할 수 있다. 또한, 컬러 영상만을 특징 학습과 추출에 이용한 경우(RGB)에 비해, 깊이 영상(Depth)과 물체 표면의 법선 벡터를 이용한 경우(Normals)가 대체적으로 더 높은 인식 성능

을 보였다. 또, 깊이 영상을 이용하는 경우(Depth)에 비해서 물체 표면의 법선 벡터를 이용한 경우(Normals)가 좀 더 높은 인식 성능을 보여주었다. 한편, SVM 학습 알고리즘을 이용했을 때는 표면 법선 벡터로부터 특징을 자동 학습한 경우(HMP-Normals)가 인식 성능 94.5%, K-NN 알고리즘을 이용했을 때는 컬러 영상과 표면 법선 벡터 모두로부터 특징을 자동 학습한 경우(HMP-RGB+Normals)가 인식 성능 98.0%로 가장 높은 인식 성능을 보였다.

Fig. 10은 각 물체 범주별 인식 성능을 나타낸 그래프이다. 이 실험에서는 HMP 특징 학습과 SVM 분류 모델 학습 알고리즘을 이용하였다. Coffee Mug, Soda Can, Toothpaste 등의 범주 인식 성능은 컬러 영상, 깊이 영상, 법선 벡터 등 어떤 경우에도 비교적 높고 안정적이다. 하지만, Stapler의 범주 인식 성능은 컬러 영상을 이용한 경우(HMP-RGB, HMP-RGB+Normals) 각각 68.4%, 71.2%로, 깊이 영상(HMP-Depth)과 표면 법선 벡터(HMP-Normals)를 이용한 경우들에 비해 확연히 낮은 것을 알 수 있다. 이러한 결과의 원인으로서는 다른 범주의 물체들에 비해, 본 실험 데이터 집합 안에서는 Stapler 범주에 속한 물체들이 모양은 거의 유사하지만 색상이 많이 다르기 때문인 것으로 추측된다. 이와 같은 실험 결과를 통해서 3차원 물체 범주 인식의 경우, 물체의 모양 정보가 매우 중요한 역할을 수행한다는 것을 확인할 수 있었다.

4.3 물체 개체 인식 실험

두 번째 실험은 본 시스템의 물체 개체 인식 성능을 분석하기 위한 실험이었다. 개체(Instance)는 같은 범주에 속한 서로 다른 물체를 의미하며, 본 실험 데이터 집합에서는 물체 범주에 따라 각각 5~8 개의 서로 다른 개체 데이터를 포함하고 있다. 첫 번째 물체 범주 인식 실험과 마찬가지로, 이 실험에서도 서로 다른 특징 추출 방법들과 분류 모델 학습 알고리즘들을 적용하여 인식 성능들을 서로 비교해 보았다.

Table 4. Comparison of Instance Recognition Accuracy

	Features	Models	
		SVM	K-NN
Human-designed features	SIFT - RGB	27.2%	49.3%
	SIFT - Depth	21.8%	15.0%
	FPFH - Normals	36.7%	28.6%
Learned features	HMP - RGB	91.5%	82.5%
	HMP - Depth	55.7%	41.2%
	HMP - Normals	55.9%	43.6%
	HMP - RGB+Normals	94.4%	83.0%

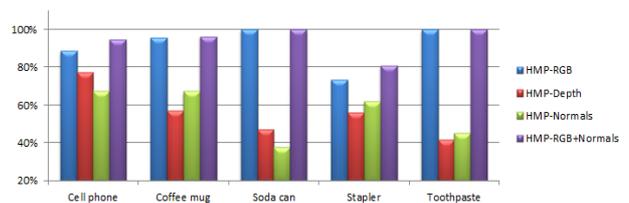


Fig. 11. Instance Recognition Accuracy for Each Category

Table 4는 물체 개체 인식 성능 비교 실험 결과를 보여 준다. 물체 범주 인식 실험의 경우와 마찬가지로, 사람이 미리 설계한 특징 추출 방법들(SIFT, FPFH)에 비해, 학습을 통해 특징을 추출한 경우(HMP)의 범주 인식 성능이 훨씬 더 높다는 것을 Table 4를 통해 확인할 수 있다. 하지만, 주목할 점은, 범주 인식 실험의 경우와는 달리, 개체 인식 실험에서는 컬러 영상을 이용한 경우(RGB, RGB+Normals)가 깊이 영상(Depth)이나 표면 법선 벡터(Normals)만을 이용한 경우들에 비해 대체적으로 더 높은 인식 성능을 보였다. 일반적으로 같은 범주에 속한 3차원 물체들은 대부분 모양은 유사하다. 따라서 같은 범주에 속한 서로 다른 개체들을 판별해야 하는 개체 인식 실험에서는 깊이 영상이나 표면 법선 벡터에 비해 컬러 영상이 더 중요한 역할을 수행한 것으로 판단된다. 하지만, 컬러 영상과 물체 표면의 법선 벡터를 함께 사용한 경우(HMP-RGB+Normals)의 개체 인식 성능이 각각 94.4%, 83.0%로, 컬러 영상만을 사용한 경우들(SIFT-RGB, HMP-RGB)에 비해 높게 나타난 것을 볼 때, 물체의 색상 정보뿐만 아니라 모양 정보도 개체 인식에 도움을 준다는 것을 알 수 있었다.

Fig. 11은 각 물체 범주별 개체 인식 성능을 나타낸 그래프이다. 이 실험에서는 HMP 특징 학습과 SVM 분류 모델 학습 알고리즘을 이용하였다. 물체 범주별로 약간의 차이는 있으나, 이 그래프에서도 컬러 영상을 이용한 경우들이 깊이 영상이나 표면 법선 벡터만을 이용한 경우들에 비해 훨씬 높은 개체 인식 성능을 보인 것을 확인할 수 있다.

4.4 물체 속성 인식 실험

세 번째 실험은 본 시스템의 속성 인식 성능을 분석하기 위한 실험이었다. 이 실험에서도 범주 및 개체 인식 실험의 경우와 같이, 서로 다른 특징 추출 방법들과 분류 모델 학습 알고리즘들을 적용하여 속성 인식 성능들을 서로 비교해 보았다. 특히 이 실험에서는 좀 더 정확한 속성 인식 성능 확인을 위해, 각 속성의 유형마다 최소 5개 이상의 개체 데이터들이 포함되도록 실험 데이터 집합을 확장하였다.

Table 5는 속성 인식 성능 비교 실험 결과를 보여준다. 물체 범주 및 개체 인식 실험들과 마찬가지로, 이 실험에서도 사람이 미리 설계한 특징 추출 방법들(SIFT, FPFH)에 비해 학습을 통해 특징을 추출한 경우(HMP)가 훨씬 더 높은 속성 인식 성능을 보인 것을 확인할 수 있다. 또한, 속성 인식에서는 색상 정보와 모양 정보를 함께 사용한 경우(RGB+Normals)가 전체적으로 더 좋은 인식 성능을 보여주었다는 것을 알 수 있다. Fig. 12는 색상(Color), 모양(Shape), 재질(Material) 등 속성 유형별 속성 인식 성능을 나타낸 그래프이다. 이 그래프에서 컬러 영상과 표면 법선 벡터를 모두 이용한 경우(HMP-RGB+Normals)가 색상, 모양, 재질 등 모든 속성 유형을 인식하는데 가장 높은 성능을 보였다는 것을 알 수 있다. 한편, 속성 유형별로는 색상 인식 성능이 모양이나 재질 인식 성능에 비해 대체적으로 낮게 나타났다. 일반적으로

로 조명 변화, 음영 정도에 따라 영상에 포착된 동일 물체의 색상들이라도 해당 물체의 기준 색상과는 크게 달라질 수 있는데, 이것이 낮은 색상 인식 성능의 원인으로 판단된다.

Table 5. Comparison of Attribute Recognition Accuracy

	Features	Models	
		SVM	K-NN
Human-designed features	SIFT - RGB	56.0%	50.8%
	SIFT - Depth	60.4%	54.5%
	FPFH - Normals	45.4%	39.1%
Learned features	HMP - RGB	81.4%	72.5%
	HMP - Depth	71.5%	70.5%
	HMP - Normals	75.8%	75.4%
	HMP - RGB+Normals	81.9%	81.0%

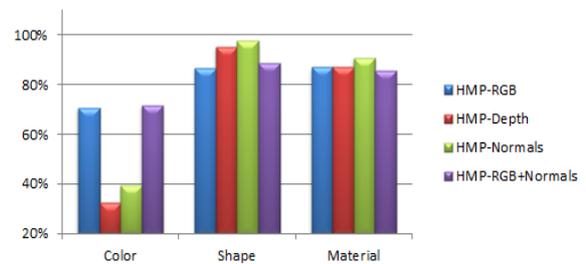


Fig. 12. Attribute Recognition Accuracy for Each Attribute

앞선 세 가지 실험 결과들을 종합해볼 때, 물체의 모양 정보와 색상 정보를 모두 활용하여 계층적 특징 학습을 수행하는 것이 3차원 물체의 범주와 개체, 그리고 속성을 인식하는데 매우 효과적이라는 것을 알 수 있다.

5. 결론

본 논문에서는 계층적 특징 학습을 이용하여 물체의 컬러 영상과 깊이 영상으로부터 해당 물체가 속한 범주와 개체, 그리고 다양한 속성들을 효과적으로 인식할 수 있는 시스템을 제안하였다. 본 시스템의 전처리 단계에서는 물체의 깊이 영상을 표면 법선 벡터 데이터로 변환하고, 특징 학습 단계에서는 물체의 컬러 영상과 표면 법선 벡터 데이터로부터 두 단계에 걸쳐 패치 단위 특징과 이미지 단위의 특징을 추출해낸다. 그리고 추출된 특징 벡터들과 SVM 학습 알고리즘을 이용하여 각기 독립적인 다수의 분류 모델들을 학습한다. 미국 워싱턴 대학의 RGB-D 물체 데이터 집합을 이용한 실험을 통해, 본 논문에서 제안한 물체 인식 시스템의 높은 성능을 확인할 수 있었다. 계획하고 있는 향후 연구로는 물체 인식 대상 범위를 확대할 수 있도록 시스템을 확장하는 일과 컨볼루션 신경망(CNN)과 같은 심층 학습 기반의 특징 추출 방법과 성능 비교 실험을 수행해보는 일 등이다.

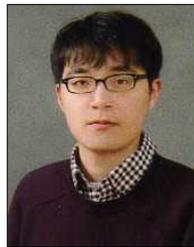
References

- [1] D. Paulk, V. Metsis, C. McMurrough, and F. Makedon, "A supervised learning approach for fast object recognition from RGB-D data," *Proc. of ACM Intl. Conf. on Pervasive Technologies Related to Assistive Environments*, p.5, 2014.
- [2] Y. Cheng, X. Zhao, K. Huang, and T. Tan, "Semi-supervised Learning For RGB-D Object Recognition," *Proc. of IEEE Intl. Conf. on Pattern Recognition*, pp.2377-2382, 2014.
- [3] L. Bo, X. Ren, and D. Fox, "Learning hierarchical sparse features for RGB-(D) object recognition," *International Journal of Robotics Research*, Vol.33, No.4, pp.581-599, 2014.
- [4] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *Proc. of IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.36, pp.453-465, 2014.
- [5] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, pp.1778-1785, 2009.
- [6] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110, 2004.
- [7] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, Vol.1, pp.886-893, 2005.
- [8] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," *Proc. of IEEE Intl. Conf. on Robotics and Automation*, pp.3212-3217, 2009.
- [9] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Transactions on Neural Networks*, Vol.8, No.1, pp.98-113, 1997.
- [10] L. Bo, X. Ren, and D. Fox, "Hierarchical matching pursuit for image classification: architecture and fast algorithms," *Advances in Neural Information Processing Systems*, pp. 2115-2123, 2011.
- [11] L. Bo, X. Ren, and D. Fox, "Kernel descriptors for visual recognition," *Proc. of Neural Information Processing Systems*, pp.244-252, 2010.
- [12] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," *Proc. of IEEE Intl. Conf. on Computer Vision*, pp.601-608, 2011.
- [13] R. B. Rusu, A. Holzbach, M. Beetz, and G. Bradski, "Detecting and segmenting objects for mobile manipulation," *Proc. of IEEE Intl. Conf. on Computer Vision*, pp.47-54, 2009.
- [14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, pp.580-587, 2014.
- [15] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," *Proc. of IEEE Intl. Conf. on Robotics and Automation*, pp.1817-1824, 2011.
- [16] LIBSVM [Internet], <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.



김 주 희

e-mail : jh.k@kyonggi.ac.kr
 2014년 경기대학교 컴퓨터과학과(학사)
 2014년~현 재 경기대학교 컴퓨터과학과
 석사과정
 관심분야 : 인공지능, 컴퓨터비전



김 동 하

e-mail : kdh2040@kyonggi.ac.kr
 2015년 경기대학교 컴퓨터과학과(학사)
 2015년~현 재 경기대학교 컴퓨터과학과
 석사과정
 관심분야 : 인공지능, 컴퓨터비전



김 인 철

e-mail : kic@kyonggi.ac.kr
 1985년 서울대학교 수학과(이학사)
 1987년 서울대학교 전산학과(이학석사)
 1995년 서울대학교 전산학과(이학박사)
 1996년~현 재 경기대학교 컴퓨터과학과
 교수
 관심분야 : 인공지능, 기계학습