추가학습이 불필요한 이미지 특징 유사도 기반 상품 식별 시스템

유영재[1,3], 윤혜정[2], 김준오[2], 박예솔[2], 장병탁[1,2,3†]

[1]서울대학교 컴퓨터공학부

[2]서울대학교 협동과정 인공지능전공

[3]투모로 로보틱스

Product Identification System based on Image Feature Similarity with Learning-free Model

Youngjae Yoo[1,3], HyeJung Yoon[2], Juno Kim[2], Yesol Park[2], Byoung-Tak Zhang[1,2,3†]

[1]Department of Computer Science and Engineering, Seoul National University

[2]Interdisciplinary Program in Artificial Intelligence, Seoul National University

[3]Tommoro Robotics

Recently, logistics centers attempts to get help from artificial intelligence robots with hard labor. To generally perform picking task for robots, it is essential to detect and identify the product through the camera. Supervised learning-based deep learning technology is suitable for recognizing objects with high accuracy. But it has a disadvantage that requires a lot of time because a human must manually label the answer of the train image. In this paper, we propose a method that minimizes manual labor and makes it easy to add products. Our algorithm identifies the product with the highest similarity by calculating the similarity between the features of the input image and the features of the images in the product database. It does not require learning and labeling to identify a new product. To verify it, we test the algorithm by photographing a test product images in an environment that simulates a logistics site.

Keywords: Object Recognition, Image Feature Matching, Learning-free Model, Logistics Automation

---

논문접수일 : 2022.10.07.
심사완료일 : 2022.12.17.
게재확정일 : 2022.12.19.
이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(2021-0-02068-AIHub/25%, 2021-0-01343-GSAI/20%, 2022-0-00951-LBA/20%, 2022-0-00166-PICA/25%, NO.2021-0-01343, 인공지능대학원지원(서울대학교)/10%)의 지원을 받아 수행되었음.
† Corresponding Author: btzhang@snu.ac.kr

1. 서론

1.1 연구배경

코로나19 및 디지털 전환으로 인해 온라인 쇼핑과 같은 전자상 거래가 폭발적으로 증가하며, 물류 센터의 필요 인력도 함께 늘고 있다(Singhdong et al., 2021). 하지만 포장, 분류, 상하차 등 물류센터의 업무는 높은 노동 강도가 요구되기에 최근 인공지능 기술을 갖춘 로봇에게 도움을 받으려는 시도가 활발히 이루어지고 있다(Karabegović et al., 2015). 고강도의 물류 작업에 적용하기 적절한 로봇은 작업대에 팔이 고정된 형태의 암 로봇이 일반적이다. <Figure 1>과 같이 로봇이 물체를 탐지하기 위해 필수적인 기술은 카메라를 통해 상품을 식별하는 것이다. 이는 로봇이 물체를 적절한 위치에 옮기거나, 분류하는 데 필수적이다.

물체의 종류를 식별하기 위해 현재 물류센터에서 사용하는 대표적인 방법은 바코드를 리더기로 읽어 파악하는 것이다. 바코드라벨은 주로 상품의 로케이션 관리, 납품처 식별, 재고관리 등에 활용되고 있다. 하지만 바코드는 정형화된 박스의 형태가 아닌 경우 배치된 면이 일관되지 않기에 위치를 파악하는 것이 쉽지 않아 사람의 도움 없이 정확한 식별을 하는 데 한계가 있다 (Grover et al., 2010).

단순 이미지만을 사용하는 경우, 이미지 해쉬와 같은 이미지 검색을 통해 부분적으로 해결할 수 있다 (Wang et al., 2015). 이미지 해쉬는 이를 이미지를 해쉬함수를 거쳐 해쉬값으로 변환한 뒤 유사도를 분석하는 방법이다. 하지만 이 방법은 동일 이미지를 찾는 데 특화되어 있기에, 물류 환경과 같이 배경이 복잡하고 구도와 조명, 촬영한 카메라가 다른 경우 유사성을 탐지하기 어렵다.

인공지능 분야에선 이를 위해 감독학습 기반 물체 탐지 연구가 활발히 이루어졌다(Cunningham et al., 2008). 감독학습 기반 모델 물체의 사진 데이터와 물체 위치 정보, 물체 종류를 데이터화하는 라벨링 작업을 거쳐 학습을 통해 개발한다. 이들은 보다 복잡한 이미지에 대해서도 높은 성능을 보이는 장점이 있지만, 학습을 위한 라벨링은 모두 사람이 수작업으로 표기하는 작업이 필요하여 많은 시간과 노동이 필요한 단점이 있다.

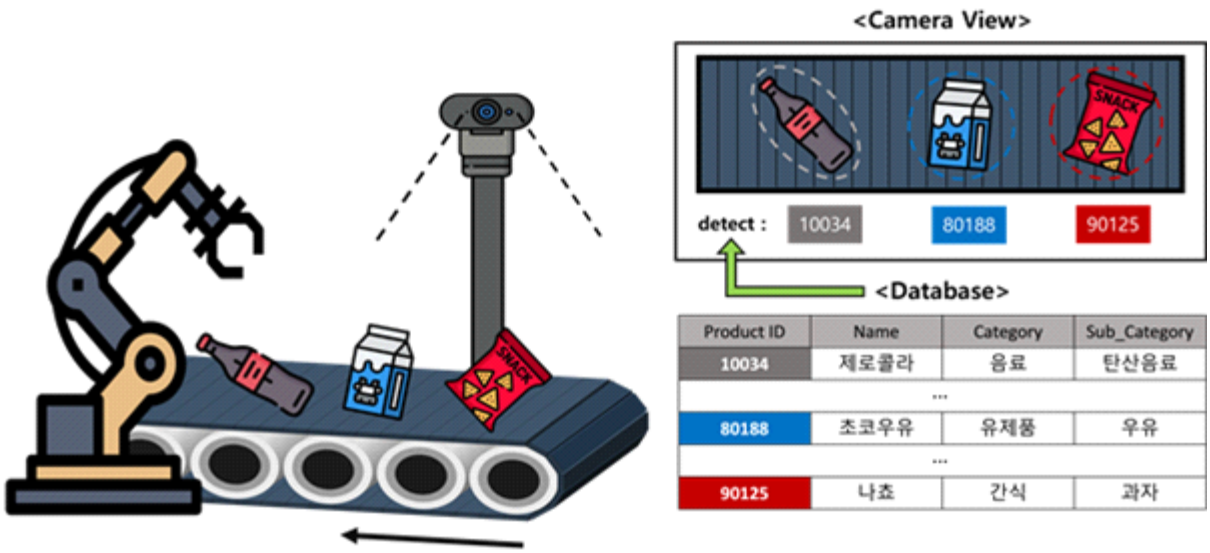


Figure 1. Object recognition for robot picking tasks

1.2 문제정의 및 논문구성

물류의 자동화를 위해 상품 이미지를 통해 종류를 식별하는 것은 필수적이나, 많은 데이터를 수집한 뒤 라벨링을 수작업으로 표기하는 작업에 많은 노동이 필요한 한계점이 있다. 또한, 새로운 물체를 추가하려면 다시 데이터 수집 및 라벨링, 학습을 수행해야 하므로 상품 추가 및 수정이 용이하지 않다. 그리고 다뤄야 하는 물체의 개수가 많아지면 이 수작업은 기하급수적으로 늘어난다. 하지만 대부분의 물류 상품의 경우 일반적으로 공업적으로 생산되어 외관이 같으며, 카메라를 통한 이미지 수집 시 작업 환경에 따라 회전, 빛 반사, 밝기 등의 제한된 변수만이 존재한다. 따라서 이와 같은 제한된 변수를 고려하면 학습을 거치지 않고, 이미지의 유사도를 비교해 동일 상품을 데이터베이스 내에서 검색하는 방식으로 식별을 수행하면 많은 양의 데이터 수집 및 라벨링이 불필요할 것이라는 발상에서 본 연구를 시작하였다.

이처럼 상품 종류의 변동이 용이하며 사람의 수작업을 최소화하는 방안으로 본 논문에선 학습 및 라벨링이 필요하지 않고, 물체의 앞뒷면 사진만을 필요로 하는 이미지 특징 비교를 통한 상품 식별 알고리즘을 제안한다. 먼저 2장 관련 연구에선 로봇의 물체 파지를 위한 물체 탐지, 물체를 잘 탐지하기 위한 이미지 전처리, 특징 추출, 그리고 특징 간 유사도 비교를 위해 사용할 수 있는 지표 등을 설명한다. 3장에서는 제시하는 알고리즘을 설명한다. 먼저 입력 이미지를 배경 제거 및 이미지 전처리를 거쳐 특징을 추출한 뒤, 상품 데이터베이스 내 상품 이미지의 특징과 유사도를 비교하여 가장 유사한 상품을 찾아낸다. 4장에서는 이를 검증하기 위해 물류 현장을 묘사한 환경에서 상품을 촬영해 수집한 테스트 셋을 소개하고, 구축한 상품 데이터베이스를 소개한다. 그리고 실험의 결과를 보이며 제시한 알고리즘을 사례와 함께 분석한다. 5장에서는 결론을 정리하고 후속 연구 계획을 소개한다.

2. 관련 연구

2.1 로봇의 파지 작업을 위한 물체 탐지

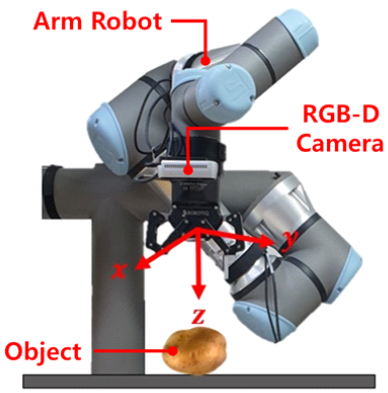


Figure 2. Arm Robot for object grasp manipulation

작업대에 고정된 암 로봇은 무거운 물체들을 파지하여 목표 지점까지 옮기는 작업을 수행할 수 있다. 이때 인공지능을 기반으로 파지할 목표 물체를 탐지하기 위해선 <Figure 2> 에서와 같이 천장이나 작업대에 고정된 rgb-d 카메라를 일반적으로 사용한다. 인공지능 기반의 목표 물체 탐지는 먼저 RGB 카메라로 물체의

2차원 이미지상 위치를 파악하고, 심도(Depth) 카메라를 활용해 3차원의 실세계 거리를 파악한다. 그리고 카메라와 로봇팔의 위치를 고려해 로봇팔이 손을 뻗을 거리를 계산한다(Sarabu et al., 2019). Zhuang et al.(2021)은 물체 위치 탐지를 위해 물체 탐지 및 분할을 수행하였다. Semantic PPF이라는 이 방법은, pointcloud 데이터를 기반으로 object-part 객체 분할을 수행한다. 그리고 현실 데이터를 다량 수집하는데 많은 시간과 제약이 있음을 고려해 가상 물리엔진을 이용해 객체 분할 데이터를 효율적으로 생성하여 데이터 생성의 비용을 줄였다. Rennie et al.(2016)은 물류창고에서 상품 파지를 위해 특화된 데이터 셋을 공개하기도 했다. 해당 데이터 셋은 약 10000장의 RGB, Depth 이미지와 3D 물체 위치를 함께 제공하였다.

2.2 이미지 유사도 분석

이미지 유사도는 이미지 데이터 간의 같은 위치의 픽셀값, 인접한 픽셀값의 변화량, 밝기, 대비, 색분포 등 다양한 기준을 비교하여 계산할 수 있다. 대표적인 이미지 유사도 분석법으로는 이미지 히스토그램과 이미지 해쉬 방법 등이 있다. 이미지 히스토그램은 가로축에 이미지 픽셀값을, 세로축에 이미지 픽셀 수를 나타내어 이미지의 특성을 비교하는 방법이다(Chapelle et al., 1999; Jia et al., 2006). 하지만 유사한 색 분포만을 가진 상품이 있거나, 상품의 방향이 바뀌는 등의 경우엔 정확성이 떨어지는 한계가 있다.

이미지 해쉬는 <Figure 3> 과 같이 이미지를 해쉬함수를 거쳐 고유한 해쉬값으로 만든 뒤 비교하는 방법으로, 크게 average hash, perceptive hash, difference hash 등이 있다(Wang et al., 2015; Zauner, 2010). 이미지 해쉬는 이미지를 낮은 차원으로 압축한 뒤 색 단순화를 거쳐 2차원 배열 혹은 해쉬값을 얻고 hamming 거리를 구해 유사도를 비교할 수 있다. 해쉬값을 사용하여 빠른 시간이 소요되는 장점이 있지만, 정해진 사이즈로 압축해야 하거나 단일 해쉬값으로 변환해야 하기에 충분한 특징을 추출하는 데 한계가 있기에 상품의 촬영 환경에 따라 정확도가 떨어지는 한계가 있다.

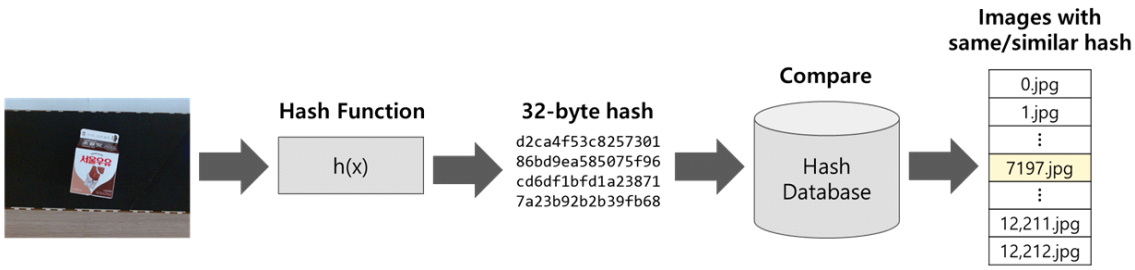


Figure 3. Similarity analysis using Image Hash

2.3 이미지 전처리 및 특징 추출

1) 이미지 전처리

이미지에서 특징을 추출하기 위해 모델에 입력하기 이전에 보다 이미지를 명확하게 가공하는 과정을 전처리라 하며, <Figure 4> 와 같이 크게 두 가지로 나눌 수 있다(Shorten and Khoshgoflaar, 2019). 첫 번째는 주어진 이미지의 픽셀 단위로 처리하는 픽셀 단계 변환(Pixel-Level Transform)으로, 흐림(Blur), 대비(Contrast), 양각(Emboss), 선명도(Sharpen) 등의 기법이 존재한다(Buslaev et al., 2020; Li et al., 2017; Singh and Kapoor, 2014). 두 번째는 이미지 공간 자체에 변형을 주는 공간 단계 변환(Spatial-Level Transform)이다. 대표적으로 뒤집기(Flip), 회전(Rotation), 그리고 이미지의 일부 영역만 이용하는 잘라내기(Crop) 등이 이에 해당한다.

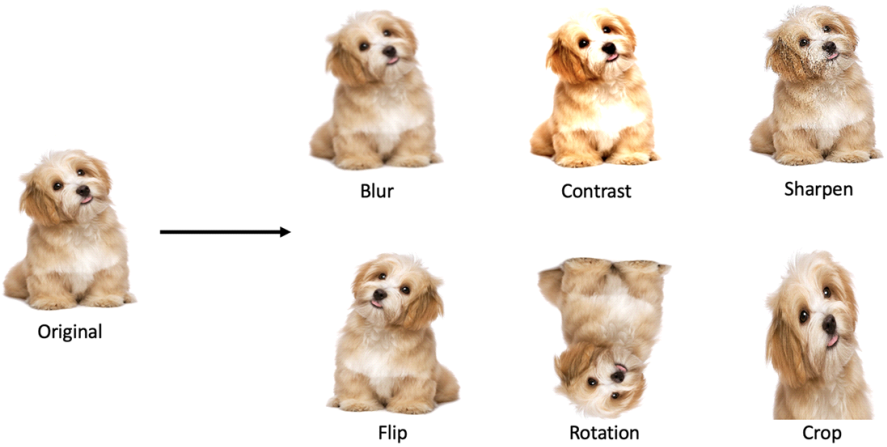


Figure 4. Image preprocessing

2) 이미지 객체 분할

이미지 객체 분할(Instance Segmentation)이란, 이미지 내에서 픽셀 단위로 영역을 분리해 객체를 추출하는 방법을 말한다(Zaitoun and Aqel, 2015). U-Net은 객체 분할 모델 중 하나로, 모델의 모양이 U자를 띠고 있다(Ronneberger et al., 2015). <Figure 5> (a)에서와 같이 U-Net은 크게 수축 경로(Contracting path)와 전환 구간(Bottle Neck), 그리고 확장 경로(Expansive path)로 이루어져 있다. 수축 경로를 통해 점진적으로 넓은 범위의 이미지 픽셀을 보며 의미 정보를 추출하고, 전환 구간에서 확장 경로로 바뀐다. 확장 경로에서 해당 정보를 위치 정보와 결합해 각 픽셀이 어떤 객체에 속하는지 구분한다. 이 과정을 거쳐 입력된 넓은 범위의 이미지에서 의미 있는 객체를 가져온다. Qin et al., (2020)에서 소개한 U2-Net (U-Square Net)은 Ronneberger et al.,(2015)의 U-Net 구조를 가진 블록(residual U-block)들이 또 하나의 U-Net를 이루고 있는 모델이다. Rembg는 이러한 U2-Net을 기반으로 사전 학습한 모델을 제공하는 배경 제거 도구로, <Figure 5> (b)처럼 입력된 이미지에서 배경을 제거하고 객체에 해당하는 부분만 <Figure 5> (c)와 같이 얻어낼 수 있다.

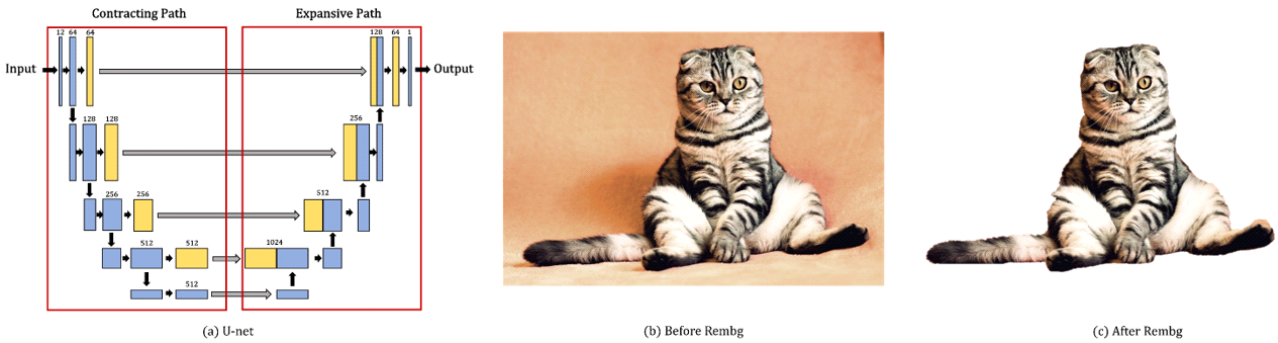


Figure 5. U2-Net based remove background(rembg) example

3) 이미지 특징 추출

이미지 데이터는 RGB의 3차원 정보를 저장한 픽셀로 이루어져 있다. 이미지의 특징 추출에 특화된 인공지능 모델인 합성곱 신경망은 이미지의 공간정보를 유지한 채 학습을 진행한다. 합성곱 층을 더 깊이 쌓아 향상된 성능을 제공하는 모델로, <Figure 6>와 같이 ResNet과 EfficientNet이 있다(He et al., 2016; Tan and Le, 2019).

ResNet은 모델의 깊이(depth)를 키울수록 과적합이 발생하는 문제를 Residual learning이라는 개념을 도입하며, 입력값과 출력값의 차이(Residual)를 학습에 이용해 늘어나는 레이어의 수에 따라 성능을 개선할 수 있게 하였다. EfficientNet은 최근 높은 성능을 보이는 모델 중 하나이며 모델의 깊이뿐만 아니라 너비(width), 해상도(resolution) 간 관계를 효율적으로 조절할 수 있는 Compounding scaling 방법을 제안하여 뛰어난 성능 향상을 이뤄내었다.

딥러닝 라이브러리인 PyTorch에서는 위의 ResNet 및 EfficienNet의 사전 학습 모델을 제공한다. 사전 학습 모델이란, Deng et al.(2009)의 ImageNet과 같은 방대한 양의 데이터를 이용해 학습을 사전에 완료한 모델이다. 따라서 사전 학습 모델을 활용할 경우, 정교한 특징 추출 모형을 처음부터 직접 구축하는 것보다 효율적으로 사용할 수 있다.

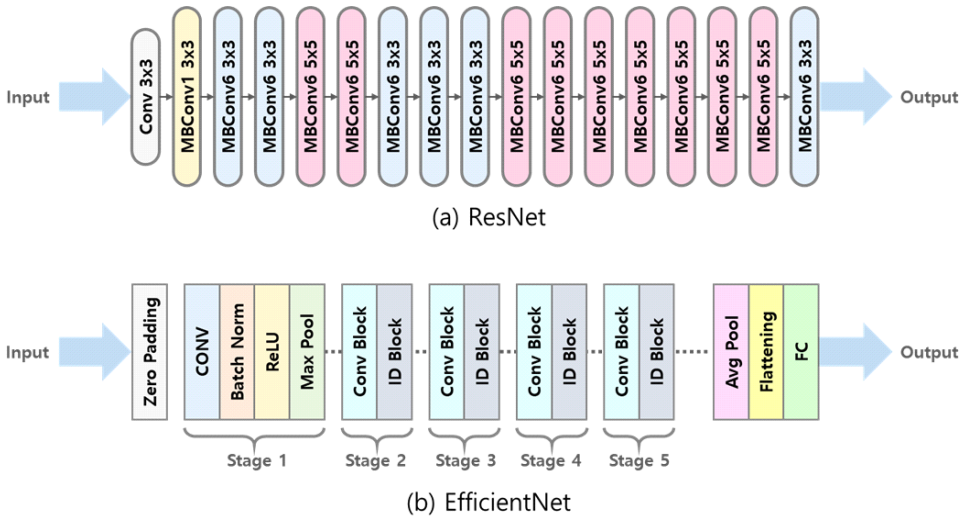
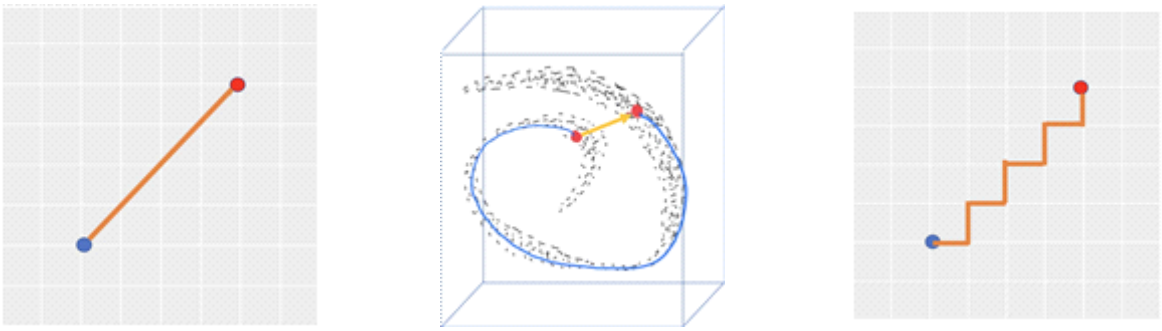


Figure 6. ResNet & EfficientNet Architecture

2.4 이미지 특징벡터 간 유사도 분석을 위한 지표

2.3의 이미지 특징 추출을 이용해 이미지 특징 벡터 간 유사도 분석을 위해선 인공지능 특징 추출 모델을 통과하여 나온 벡터들을 비교하는 유사도 비교 기법이 필요하다(Choi et al., 2010). 벡터 유사도 비교 기법에는 크게 두 가지 유형이 있다. 첫 번째는 벡터 간의 거리를 비교하는 유사도 기법으로 이에 유클리디언 유사도, 맨해튼 유사도, 민코스키 유사도가 해당한다. 두 번째로는 벡터 간의 각도를 비교하는 유사도 기법으로 이에 코사인 유사도가 해당한다.

1) 거리 기반 유사도 비교 기법



(a) Euclidean in 2D (b) Euclidean in high dimension (c) Minkowski in 2D

Figure 7. Visualization of distance based similarity matching metrics

벡터 간의 거리를 비교하는 유사도 비교 기법 중 대표적인 방법은 유클리디안 거리로 식 (1)과 같다. 결과 값이 작을수록 두 벡터 간의 거리가 가까운 것이므로 유사하다고 판단한다. 하지만, 유클리디안 거리는 벡터 차원에 대해 독립적이지 못하므로, 벡터 차원의 크기에 따라 거리의 계산값이 왜곡될 수 있다. 따라서 유클리디언 거리는 차원의 저주 이론에 따라 벡터의 차원이 커질수록 부정확한 유사도를 제공한다(Keogh and Mueen, 2017). <Figure 7> (a)와 같은 2차원 공간의 경우 두 점 간의 거리가 최단 거리로 계산되어 유사도가 정확하게 비교되는 것을 볼 수 있다. 하지만, 높은 차원을 시각화한 <Figure 7> (b)를 보면 실제 거리는 직선이 아닌 왜곡된 거리이다. 이처럼 유클리디안 거리는 높은 차원에서는 신뢰성이 떨어지기 쉽다. 따라서 유클리디안 거리는 낮은 차원에서 보다 신뢰성 있는 결과를 제공한다.

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

높은 차원의 벡터를 비교하는 경우 맨해튼 거리는 유클리디언 거리보다 더 신뢰 있는 유사도를 도출할 수 있다(Aggarwal et al., 2001). 맨해튼 거리 기법은 식 (2)와 같이 두 벡터 간의 1차원적인 정보를 비교하여 거리를 계산하는 기법이다. 유클리디언 거리와 유사하지만, 차이점은 2차원 정보를 사용하지 않고 1차원 정보를 사용한다는 것이다. 유클리디언 거리는 <Figure 7> (a)와 같이 체스판에서 대각선으로 움직여 거리를 계산한다면 맨해튼 거리는 <Figure 7> (c)와 같이 대각선으로는 움직이지 않고 거리를 계산한다. 따라서, 맨해튼 거리는 최단 거리를 계산하지 않기 때문에 유클리디언 거리와 달리 높은 차원에서 비교적 신뢰도 높은 유사도를 보일 수 있다.

$$D(x,y) = \sum_{i=1}^{n}|x_i - y_i| \tag{2}$$

민코스키 거리는 식 (3)과 같이 거리 기반 유사도를 매개변수에 따라 일반화한 유사도 기법이다. 식 3의 매개변수 p에 따라 p를 1로 설정하면 식 (1)과 같고, 2로 설정하면 식 (2)와 같다. p를 큰 값으로 설정하여 n 차원의 거리를 계산할 수 있다. 하지만, Aggarwal at el.(2001)에서 소개된 것과 같이 높은 차원에서 p 값을 크게 설정할수록 정확한 거리를 얻기 힘들고 신뢰 있는 유사도를 얻기 힘들다. 이 경향은 벡터 간에 가장 멀리 떨어져 있는 차원만 보기 때문에 관찰된다. 고차원 공간에서 이는 가까운 벡터에 대해 의미 있는 거리 표현을 보장하지 못한다. 따라서 고차원에서 노름 공간을 바탕으로 하는 거리 계산 기법은 정확한 유사도를 도출하기 힘들다. Christian(2019)과 Hennig(2020)의 연구에서는 높은 차원에서 p 값이 1일 경우의 가능성에 더 주목해 볼 필요가 있음을 주장하였다.

$$D(x,y) = \sum_{i=1}^{n} (|x_i - y_i|^p)^{\frac{1}{p}} \tag{3}$$

2) 각도 기반 유사도 비교 기법

벡터 간의 각도를 비교하는 유사도 기법인 코사인 유사도는 식 (4)와 같다. 코사인 유사도는 유클리디안 거리의 단점인 차원의 저주를 보완하기 위해 사용할 수 있다. 코사인 유사도는 두 벡터 간의 코사인 각도를 계산하고 내적 값으로 정규화를 하여 두 벡터 간의 각도를 비교하는 유사도 기법이다. 각도 계산의 결괏값은 -1부터 1의 값을 출력하고 1에 가까울수록 서로 유사한 벡터이며 -1에 가까우면 서로 유사하지 않은 벡터라 판단한다. 위의 유클리디안 거리와 달리 코사인 유사도는 벡터의 크기에 대한 값은 정규화 하므로, 만약 벡터 간의 크기가 중요한 특징이라면 코사인 유사도는 적용이 부적절하다. 즉, <Figure 8> (b)와 같이 벡터 간의 방향성이 유사하지만 크기가 다르다면 잘못된 결과를 도출할 수 있다.

$$D(x,y) = \cos\theta = \frac{x \bullet y}{|x||y|} \tag{4}$$

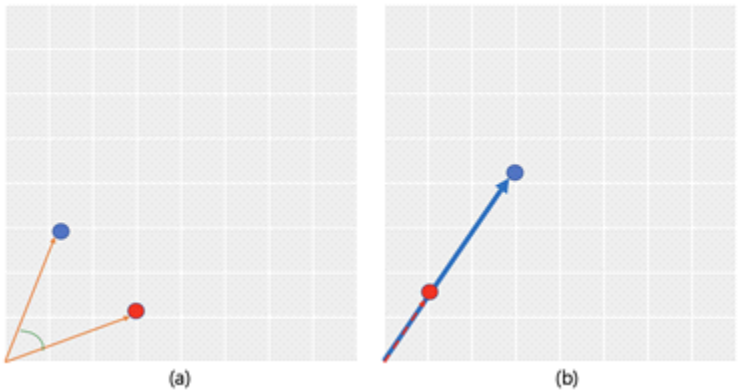


Figure 8. Cosine similarity

하지만, 코사인 유사도 기법은 두 벡터 간의 각도를 바탕으로 계산하기 때문에 차원의 저주를 받지 않아 높은 차원에서 정확한 정보를 도출할 수 있다는 장점이 있다. 특히 이미지 특징 벡터와 같이 차원이 높지만, 벡터들의 크기 정보가 중요하지 않은 경우 코사인 유사도 기법이 사용 가능하다.

3. 이미지 특징 유사도 기반 상품 식별 시스템

3.1 시스템 개요

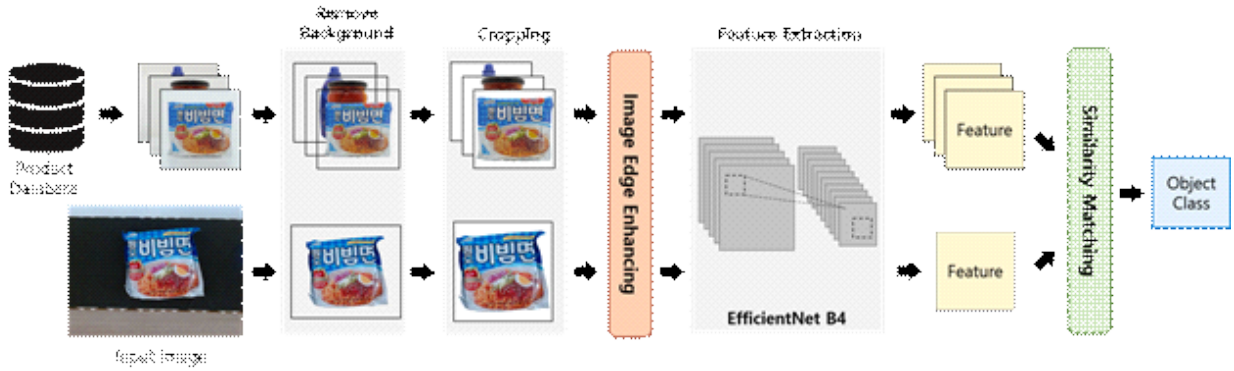


Figure 9. System architecture

본 연구의 시스템 구조는 <Figure 9>와 같다. 먼저 상품 데이터베이스 내 모든 물체의 사진을 데이터베이스에서 순차적으로 열어 3차원 배열화한다. 그리고 U2-Net 기반 배경 제거 모델을 활용해 이미지들의 배경을 제거한 뒤 흰 배경으로 채우는 이미지 전처리를 수행한다. 그리고 카메라를 통해 찍은 이미지를 동일하게 배경 제거한 뒤, 이미지 윤곽선 강화(Image Edge Enhancing)를 거쳐 윤곽선이 두드러지도록 전처리를 적용한다. 전처리를 거친 각 이미지를 특징 추출기인 EfficientNet B4를 거쳐 크기 1,792인 특징을 추출한다. 그리고 유사도 매칭 방법인 코사인 유사도를 통해 입력 상품의 특징과 상품 데이터베이스의 상품별 특징과 유사도를 구한 뒤, 가장 높은 유사도의 상품 번호를 찾아낸다.

3.2 이미지 전처리

상품 데이터베이스 이미지 및 입력 이미지에서 상품의 특징을 추출하기 이전, 이미지 내에서 상품에 보다 집중하기 위한 전처리 과정을 거친다. 먼저 2.3.2절에서 소개한 배경 제거 모델인 Rembg를 사용해 원본 이미지에서 배경을 제거하고 상품 객체만을 얻기 위해 잘라내기(cropping)를 수행한다. 배경 제거 시엔 배경을 투명하게 표현하기 위해 투명도(alpha) 값을 포함하는 RGBA 이미지로 변환된다. RGBA 이미지는 이미지 윤곽선 강화를 적용하기에 용이하지 않으므로, 투명한 픽셀은 흰색 값(255, 255, 255)으로 채우며 RGB 이미지로 다시 변환한다. 다음으로 잘라낸 이미지에 대표적인 윤곽선 강화 처리 기법인 Embossing과 Sharpening을 <Figure 10>과 같이 적용한다. 또한 비교실험을 위해 RandomBrightnessContrast도 함께 적용해본다. 이미지 전처리 과정을 모두 거친 이미지는 크기를 600 x 600으로 조정한 뒤 사전학습 된 특징 추출 모델인 EfficientNet B4에 입력해 특징을 추출한다.

Figure 10. Image preprocessing type of our system

3.3 특징 유사도 측정

본 연구에서는 총 4가지 특징 벡터 유사도 측정 기법을 적용한다. EfficientNet B4를 통해 추출된 특징 벡터의 크기는 1,792로, 차원이 크다. 2.4장에서 소개한 항목들 중, 높은 차원에 적절한 유사도 기법인 코사인 유사도와 맨해튼 거리 유사도를 우선 적용한다. 또한 비교실험을 위해 유클리디안 거리와 민코스키 거리 지표도 함께 적용해 본다. 이때 민코스키의 경우 매개변수 p를 3으로 설정한다.

4. 시스템 도입을 위한 실험

4.1 실험 데이터 수집 환경

실험 데이터 수집을 위한 촬영은 물류센터를 모사한 <Figure 11>의 실험 환경에서 진행하였다. 환경은 상품이 혼재되어 있지 않고, 개별 상품이 분리되어 컨베이어 벨트를 통해 이동하는 특정 상황임을 가정 하였다. 촬영엔 컨베이어벨트 위에 거치대를 통해 바닥면과 수평으로 고정된 로지텍 C920 카메라를 사용하였으며, 생활 물체가 모두 화면에 담기도록 카메라 높이를 40cm로 고정해 주었다.

Figure 11. Experimental environment for data collection

4.2 실험 데이터 구성

Table 1. Data statistics for AIHub product image dataset

Category	Amount	Heights(3) + Angles(24)	
		Singular	Plural
과자	1,693	121,896	121,896
디저트	77	5,544	5,544
면류	208	14,976	14,976
상온HMR	1,093	78,696	78,696
생활용품	1,112	80,064	80,064
소스	736	52,992	52,992
유제품	291	20,952	20,952
음료	1,130	81,360	81,360
의약외품	203	14,616	14,616
이/미용	2,019	145,368	145,368
주류	496	35,712	35,712
커피차	508	36,576	36,576
통조림/안주	322	23,184	23,184
홈클린	392	28,224	28,224
Total	10,280	740,160	740,160

실험을 위한 상품 데이터베이스는 한국지능정보사회진흥원이 운영하는 AI 통합 플랫폼인 AIHub의 '상품 이미지' 데이터 셋(<Table 1>)을 활용하여 구축하였다. 상품 이미지 데이터는 10,280개의 상품을 하나씩 촬영한 단수 상품 이미지와 2개 묶음을 촬영한 복수 상품 이미지로 나누어 제공한다. 한 상품에 대해 다양한 구조에서 촬영하였으며 3가지의 높이(0도, 30도, 60도)와 15도씩 24번의 회전된 이미지를 제공한다. 이처럼 상품 한 개에 144개, 전체 상품에 대해 총 약 144만 장의 이미지를 제공한다. 제공되는 이미지 예시는 <Figure 12>과 같다.

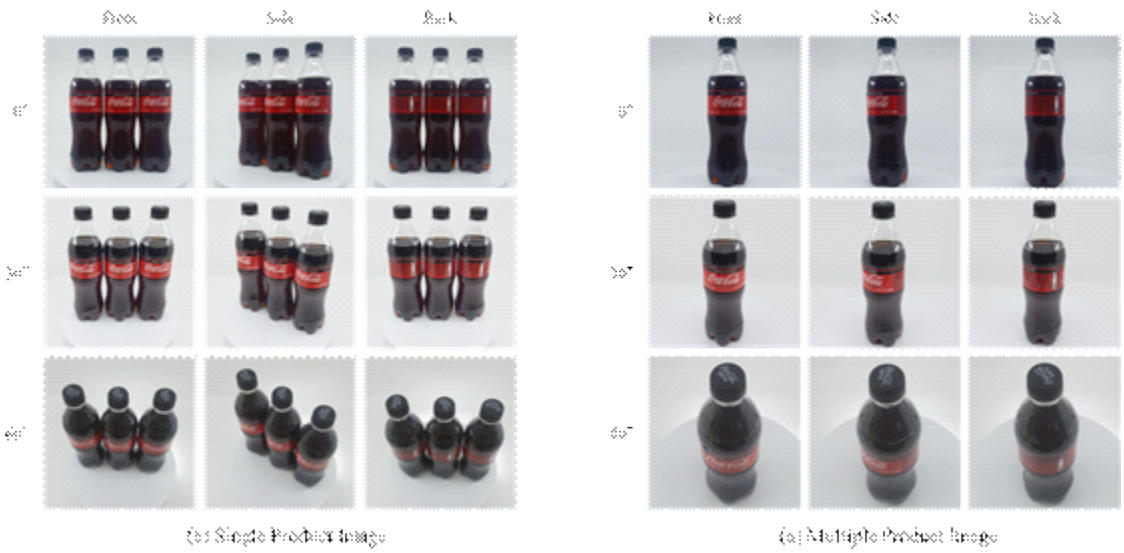


Figure 12. Example of AIHub product image dataset

본 논문에선 위의 데이터 셋에서 단수 상품 이미지만을 사용하였으며 상품당 0도 높이에서 앞면과 뒷면을 찍은 이미지 2장만을 사용하여 자체 상품 데이터베이스를 구축하였다. 이는 총 14개 카테고리의 6,106개 상품에 대해 상품당 앞, 뒷면 이미지를 가져와 총 12,212장을 저장하였다. AIHub의 상품 이미지 데이터 셋은 사람이 수작업으로 제작하여 이미지가 깨진 채 저장된 오류가 있거나 파일명이 일관되지 않아 앞, 뒷면 이미지가 존재하지 않는 상품이 다수 존재하였다. 따라서 일관된 상품 데이터베이스 구축을 위해 이처럼 오류가 발생할 수 있는 상품은 제외하여 저장해 주었다.

Figure 13. Product database structure

상품 데이터베이스(Product DB) 구축을 위한 구조는 <Figure 13>과 같다. 먼저 이미지 저장 시에 0번부터 순차적으로 숫자를 이름으로 붙여주어 기존의 복잡한 폴더 구조를 단순화하였다. 파일명(id)과 함께 앞서 데이터를 선별하며 파싱해온 정보인 상품 이름(name), 상품 고유번호(product_id), 이미지의 앞/뒷면 여부(fb), 카테고리(category) 등의 메타 데이터를 스프레드시트 파일에 함께 저장하였다. 최종적으로 생성된 데이터베이스의 구조와 데이터베이스의 메타정보를 저장한 메타 데이터 구조는 <Table 2>와 같다.

Table 2. Structure of meta data for product database (data.csv)

id	name	product_id	fb	category
0	오뚜기도나스믹스500G	10232	0	소스
1	오뚜기도나스믹스500G	10232	1	소스
2	청정원12나트륨솔트	10243	0	소스
3	청정원12나트륨솔트	10243	1	소스
4	대상청정원명란크림파스타소스350G	10293	0	소스
5	대상청정원명란크림파스타소스350G	10293	1	소스
6	폰타나카르니아베이컨앤머쉬룸크림430G	15329	0	소스
7	폰타나카르니아베이컨앤머쉬룸크림430G	15329	1	소스
8	청정원)연탄불고기양념140G	25214	0	소스
9	청정원)연탄불고기양념140G	25214	1	소스
⋮	⋮	⋮	⋮	⋮

또한 실시간으로 물류 환경에서 식별이 필요한 상품 데이터 셋은 다음과 같이 구축하였다. 먼저, 데이터베이스에 있는 6,106개의 상품 중 시중에서 쉽게 구할 수 있는 상품을 50개 선정하였다. 그리고 각 상품을 실험 환경의 컨베이어벨트에 올려 앞면과 뒷면 이미지를 직접 촬영하여 100개의 테스트 이미지 데이터 셋을 구축하였다. 촬영 시 각 물체 당 이미지 촬영 시 3초 이내로 제한을 두어, 최대한 실제 산업 현장과 비

슷하게 정제되지 않은 이미지 데이터를 구축하도록 하였다. 테스트 이미지는 800x600 해상도로 저장하였다. 최종적으로 생성된 테스트용 이미지 예시는 <Figure 14>와 같고 상품 정보를 저장한 메타 데이터 구조는 <Table 3>과 같다.

Table 3. Structure of meta data for test product dataset (test_label.csv)

id	name	product_id
0	롯데)제크오리지날100G	30100
1	롯데)제크오리지날100G	30100
2	영남코프레이션레이지레몬주스(노란통)	65349
3	영남코프레이션레이지레몬주스(노란통)	65349
4	오뚜기스위트콘340G	15037
5	오뚜기스위트콘340G	15037
6	오뚜기옛날볶음참깨100G	20298
7	오뚜기옛날볶음참깨100G	20298
8	금강깨끗한표백제(바르는타입)	80213
9	금강깨끗한표백제(바르는타입)	80213
⋮	⋮	⋮

Figure 14. Example of image in test dataset

4.3 실험 결과 및 분석

1) 이미지 윤곽선 강화, 밝기 조절 및 특징 추출 모델에 따른 비교 실험 결과

Table 4. Experimental results by image edging enhancing filter

	ResNet		ResNet + EfficientNet		EfficientNet (Ours)	
	Top-1 Acc.	Top-3 Acc.	Top-1 Acc.	Top-3 Acc.	Top-1 Acc.	Top-3 Acc.
No Filter	51%	71%	60%	**78%**	53%	77%
Emboss Filter	**59%**	**75%**	54%	74%	58%	74%
Sharpen Filter	52%	69%	**62%**	77%	54%	72%
Random Brightness Contrast Filter	43%	62%	59%	75%	**60%**	76%
Emboss + Sharpen Filter (Ours)	53%	66%	54%	72%	58%	**80%**

<Table 4>는 이미지 윤곽선 강화 및 이미지 특징 추출 모델의 비교실험 결과이다. 본 연구의 시스템과의 비교를 위한 특징 추출 모델은 크게 ResNet과 ResNet + EfficientNet 두 개의 모델을 동시에 사용하는 앙상블 모델로 선정하여 비교실험을 하였다. 사용한 성능 지표는 Top-1 Accuracy와 Top-3 Accuracy로, Top-1 Accuracy는 100장의 테스트 이미지에 대해 특징 유사도 기법을 이용해 상품 이미지 데이터베이스 내 상품을 검색했을 때, 가장 유사한 상품으로 식별되는 상품이 실제 테스트 이미지의 상품과 일치하는 정도를 수치화 한 것이고, Top-3 Accuracy는 가장 유사한 상품으로 판단되는 3가지 상품 중에 실제 상품이 존재하는 지를 수치화 한 것이다. 먼저 본 논문의 시스템에서 사용한 특징 추출 모델인 EfficientNet에서, 이미지 강화 처리하지 않은 경우보다 Emboss와 Sharpen을 모두 적용한 경우 더 높은 성능을 보였다. 또한 이미지 강화 처리를 적용한 경우, 다른 특징 추출 비교모델보다 EfficientNet이 가장 높은 성능을 보였다.

ResNet과 EfficientNet의 특징을 연결한 앙상블 모델에서는 전반적으로 좋은 성능을 보였으나, 윤곽선 강화를 적용한 EfficientNet에 비해 높은 성능을 보이진 않았다. 이미지 전처리와 특징 추출모델을 복합적으로 분석해 보면, 일반적으로 ResNet의 경우 색 정보를, EfficientNet의 경우 윤곽선 정보를 중시하는 경향을 보였다. 동일한 테스트 이미지에 대해서 <Figure 15>과 같이 전자의 경우 색이 유사한 물체를, 후자의 경우 모양이 비슷한 물체를 우선하여 선택하는 경향을 보임을 확인할 수 있다.

Figure 15. Example of results by ResNet and EfficientNet

2) 이미지 특징 유사도 비교 지표에 따른 결과

Table 5. Experimental results by similarity comparison metric

Emboss + Sharpen Filter (Ours)	ResNet		ResNet + EfficientNet		EfficientNet (Ours)	
	Top-1 Acc.	Top-3 Acc.	Top-1 Acc.	Top-3 Acc.	Top-1 Acc.	Top-3 Acc.
Cosine Similarity (Ours)	**53%**	**66%**	51%	**72%**	**58%**	**80%**
Euclidean Distance	48%	60%	52%	71%	56%	77%
Manhattan Distance	50%	61%	**54%**	71%	55%	72%
Minkowski Distance	20%	30%	32%	46%	36%	51%

이미지 특징 유사도 비교 지표에 따른 실험 결과는 <Table 5> 와 같다. 먼저 대부분의 수치에서 각도 유사도 기법인 코사인 유사도가 거리 유사도 기법들에 비해 높은 정확도를 보여주었다. 그리고 유클리디안 거리와 맨해튼 거리가 비슷한 수치로 그다음의 성능을 보였다. 본 연구의 특징 추출모델을 통해 얻은 특징 벡터들은 1,000 이상의 고차원이므로 코사인 유사도가 가장 높은 정확도를 보여주는 것으로 분석된다. 반면, Christian(2019)의 연구를 근거로 거리 기반 유사도 기법 중 맨해튼 거리 방법이 가장 높은 정확도를 보일 것으로 추측하였지만, 유클리디안 거리와 유사하거나 오히려 좋지 못한 결과를 보였다. 민코스키 유사도는 매개변수 p가 3일 경우 Hennig(2020)의 연구에서와 같이 좋지 못한 정확도를 보여주었다.

3) 이미지 해쉬 검색 방법과의 비교 실험

Table 6. Comparison experiment with Image hash

Model	Top-1 Acc.	Top-3 Acc.
Ours	62%	80%
Average Hash	0%	0%
Perceptual Hash	0%	3%
Diffetential Hash	1%	1%

<Table 6>은 본 논문의 시스템과 이미지 검색 기법인 이미지 해쉬와의 비교실험 결과이다. Ours는 <Table 4>의 Top-1 정확도와 Top-3 정확도의 최댓값을 기준으로 작성하였다. 이미지 해쉬의 경우 전반적으로 상당히 낮은 정확도를 보였다. Top-3 정확도에 대해 Perceptual hash 사용 시 3%를, Differential hash 사용 시 1%를 보였다. 이는 테스트 이미지가 상품 데이터베이스의 이미지와 비교하여 회전되었거나, 초점이 맞지 않아 지문 문자열을 유사하게 찾기 어려웠을 것이라 분석된다. 또한 이미지 해싱 기술 적용을 위해 이미지 해상도를 낮게 조정하여 특징 정보가 많이 소실되고, 회색조 이미지로 색 단순화 변환을 거치며 색 정보가 고려되지 않아 더욱 유사 패턴을 찾기 어려웠을 것으로 분석된다. 따라서 물류환경과 같은 동적인 환경에선 본 논문의 이미지 특징 유사도 비교 방법이 더욱 적합함을 확인하였다.

Figure 16. Experimental result for test images

4) 테스트 이미지의 상품 식별 결과 분석

<Figure 16>은 테스트 상품의 다양한 상황에 따른 상품 식별 결과 예시이다. <Figure 16> (a) 상품의 경우 과자 봉지 재질의 특성상 빛 반사가 잘 발생하였다. 빛 반사로 일부 픽셀이 소실되지만, 특징적인 로고나 색감 등이 뚜렷하여 식별에 성공하였다. <Figure 16> (b)는 움직이는 물체를 촬영해 초점이 정확히 맞지

않았다. 이미지의 특징이 왜곡되기 쉬움에도 형태와 주요 색감 특징이 드러나 성공적으로 식별되었다. <Figure 16> (c)는 같은 회사의 같은 상호인데 디자인이 달라진 경우이다. 이에 따라 동일 상품을 2번째로 식별하였지만, 가장 높은 유사도의 상품도 동일 상품군으로 식별하였다. 이는 사람과 유사한 방식으로 물체를 찾는 것을 보여주었다. 또한 투명한 용기 재질로 인해 내부 물체가 차 있는 정도가 미세하게 다른 것도 변수로 작용했을 것으로 보인다. <Figure 16> (d)는 상품의 방향이 기울어졌지만, 상품 데이터베이스 내의 정방향 이미지를 잘 식별하였다. 이는 이미지 특징 추출 시 합성곱 연산이 다양한 방향에서 특징을 추출하기에 동적 환경에서 적응적인 활용에 용이함을 보여준다. <Figure 16> (e)는 상품이 이동하며 찍히는 와중에 초점이 맞지 않았고 상품명 및 로고가 보이지 않아 쉽게 구분하기 어려운 경우이다. 하지만 상품 데이터베이스에서 해당 상품의 옆면 디자인이 테스트 이미지와 유사하며 용기의 모양이 특징적이었기 때문에 잘 식별된 것으로 분석된다. 이를 통해 상품의 이미지를 다양한 구도에서 촬영하여 상품 데이터베이스를 구축하면 연산량이 증가하지만 보다 동적인 환경의 상품에 대해 강건한 식별을 수행할 수 있을 것으로 분석된다. <Figure 16> (f)는 앞면에 상품의 주요 정보가 집중되어 있지만 뒷면이 입력되어 마치 사람도 정확한 식별을 수행하기 어려운 경우이다. 이처럼 상품의 특징이 매우 적게 드러나는 경우엔 식별의 한계가 있어 추가적인 메타정보와 함께 고려되는 것이 필요해 보인다. <Figure 16> (g)는 유사 상품 간 혼동되기 쉬운 상황에서 동일한 재질 및 형태의 포장을 상호보다 더 높은 유사도로 탐지한 결과이다. 사람의 경우 상호와 내부 내용물의 유사성을 우선으로 판단하나 모델은 미미하게 포장 형태 및 포장 재질을 우선으로 탐지하였다.

본 연구의 시스템은 색 정보와 윤곽선 정보를 복합적으로 고려해 상품을 인식한다. 그런데, 위의 <Figure 16> (e)에서 분석했듯, 실험 결과에서는 색 정보보다 윤곽선(모양) 정보를 더 중시하여 상품을 식별하는 경향을 볼 수 있다. 이는 <Figure 16> (b), <Figure 16>(e) 등 다른 제품들에서도 유사한 경향성을 확인할 수 있다. <Figure 16> (b)의 상위 1 유사도는 색과 모양이 동일하여 식별하였지만 뒤이어 오는 상위 2, 3 유사도의 상품들은 색은 다르지만, 모양이 비슷한 제품들로 탐지하였다. 이는 이미지 윤곽선 강화 적용으로 상품의 모양 정보가 두드러지게 강조되었기 때문으로 분석된다.

5. 결론

본 논문에선 학습 및 라벨링이 필요하지 않고 물체의 앞뒷면 사진을 사용해 데이터베이스 내 이미지와 특징 비교를 하여 상품을 식별해내는 알고리즘을 제안하였다. 제안한 알고리즘은 배경 제거 및 이미지 전처리를 거쳐 특징을 추출한 뒤 특징 간 유사도를 비교하는 방식으로 이미지 검색알고리즘 중 하나인 이미지 해쉬와 비교해 더 우수한 성능을 보였다. 그뿐만 아니라 초점이 잘 맞추어지지 않았거나, 빛 반사가 있는 경우 등 촬영 환경이 동적인 상황에도 상품을 잘 식별해냄을 결과로 확인하였다.

하지만 상품이 특정 부분이 가려져 있지 않고 모두 보여야 하는 시스템의 제약이 있어, 상품이 겹쳐 있거나 혼재된 상황이 아닌 개별상품이 분리되어 촬영된 특정상황에서만 실험을 적용한 한계점이 있다. 또한 실험 결과에서 보였듯, 일부 유사품이 있거나 같은 표지의 상품임에도 다른 용량이거나 디자인이 바뀐 경우에는 잘 잡아내지 못한 한계점도 있었다. 이의 경우엔 물체의 크기 및 무게와 같은 메타 정보를 함께 고려하면 성능이 향상될 것으로 보인다.

본 연구의 기술은 물류센터, 마트, 제조업 등에서 사용 가능할 것으로 기대되며 로봇팔을 통한 동적 환경에서의 자율 작업뿐만 아니라 상품 분류 및 검수 작업에도 활용 가능할 것이다. 향후 후속 연구에선 가로, 세로, 높이 정보, 무게 정보 등을 함께 수집해 유사도 분석 시 함께 고려하여 성능을 향상하고, 이를 실제 로봇의 픽-앤-플레이스 작업에 적용하여 작업 수행 효율을 최적화해보려 한다.

참고문헌

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001), On the surprising behavior of distance metrics in high dimensional space, In International conference on database theory ., pp, 420~434, Springer, Berlin, Heidelberg.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020), Albumentations: fast and flexible image augmentations. Information, 11(2), 125.

Chapelle, O., Haffner, P., and Vapnik, V. N. (1999), Support vector machines for histogram-based image classification, IEEE transactions on Neural Networks, 10(5), 1055-1064.

Choi, S. S., Cha, S. H., and Tappert, C. C. (2010), A survey of binary similarity and distance measures, Journal of systemics, cybernetics and informatics, 8(1), 43-48.

Cunningham, P., Cord, M., and Delany, S. J. (2008), Supervised learning, In Machine learning techniques for multimedia., pp, 21~49, Springer, Berlin, Heidelberg.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009), Imagenet: A large-scale hierarchical image database, In 2009 IEEE conference on computer vision and pattern recognition., pp, 248~255, Ieee.

Grover, A., Braeckel, P., Lindgren, K., Berghel, H., and Cobb, D. (2010), Parameters effecting 2D barcode scanning reliability, In Advances in Computers ., Vol, 80, pp, 209~235, Elsevier.

He, K., Zhang, X., Ren, S., and Sun, J. (2016), Deep residual learning for image recognition, In Proceedings of the IEEE conference on computer vision and pattern recognition., pp, 770~778.

Hennig, C. (2020), Minkowski Distances and Standardisation for Clustering and Classification on High-Dimensional Data, In Advanced Studies in Behaviormetrics and Data Science ., pp, 103~118, Springer, Singapore.

Jia, W., Zhang, H., He, X., and Wu, Q. (2006), A comparison on histogram based image matching methods, In Proceedings-IEEE International Conference on Video and Signal Based Surveillance 2006, AVSS 2006.

Karabegović, I., Karabegović, E., Mahmić, M., and Husak, E. (2015), The application of service robots for logistics in manufacturing processes, Advances in Production Engineering & Management, 10(4).

Keogh, E. J., and Mueen, A. (2017), Curse of dimensionality, Encyclopedia of machine learning and data mining, 314-315.

Li, S., Kang, X., Fang, L., Hu, J., and Yin, H. (2017), Pixel-level image fusion: A survey of the state of the art, information Fusion, 33, 100-112.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., and Jagersand, M. (2020), U2-Net: Going deeper with nested U-structure for salient object detection, Pattern recognition, 106, 107404.

Rennie, C., Shome, R., Bekris, K. E., and De Souza, A. F. (2016), A dataset for improved rgbd-based object detection and pose estimation for warehouse pick-and-place, IEEE Robotics and Automation Letters, 1(2), 1179-1185.

Ronneberger, O., Fischer, P., and Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, In International Conference on Medical image computing and computer-assisted intervention., pp. 234~241, Springer, Cham.

Sarabu, H., Ahlin, K., and Hu, A. P. (2019), Leveraging deep learning and rgb-d cameras for cooperative apple-picking robot arms. In 2019 ASABE Annual International Meeting ., pp. 1, American Society of Agricultural and Biological Engineers.

Shorten, C., and Khoshgoftaar, T. M. (2019), A survey on image data augmentation for deep learning, Journal of big data, 6(1), 1-48.

Singh, K., and Kapoor, R. (2014), Image enhancement using exposure based sub image histogram equalization, Pattern Recognition Letters, 36, 10-14.

Singhdong, P., Suthiwartnarueput, K., and Pornchaiwiseskul, P. (2021), Factors Influencing Digital Transformation of Logistics Service Providers: A Case Study in Thailand, The Journal of Asian Finance, Economics and Business, 8(5), 241-251.

Tan, M., and Le, Q. (2019), Efficientnet: Rethinking model scaling for convolutional neural networks, In International conference on machine learning,. pp, 6105~6114, PMLR.

Wang, X., Pang, K., Zhou, X., Zhou, Y., Li, L., and Xue, J. (2015), A visual model-based perceptual image hash for content authentication, IEEE Transactions on Information Forensics and Security, 10(7), 1336-1349.

Zaitoun, N. M., and Aqel, M. J. (2015), Survey on image segmentation techniques, Procedia Computer Science, 65, 797-806.

Zauner, C. (2010), Implementation and benchmarking of perceptual image hash functions.

Zhuang, C., Wang, Z., Zhao, H., and Ding, H. (2021), Semantic part segmentation method based 3D object pose estimation with RGB-D images for bin-picking. Robotics and Computer-Integrate

d Manufacturing, 68, 102086.