다단계 객체 추적을 통한 표시 정보의 인식 기법

최지수, 정동주, 민경식, 이병정
서울시립대학교 컴퓨터과학부
e-mail : {chlwltn214, jdj700, ksmin1710, bjlee}@uos.ac.kr

Multi-Stage Object Tracking Technique for Label Recognition

Ji-Su Choi*, Dongju Jung*, Kyeongsic Min*, Byungjeong Lee*
*Dept. of Computer Science and Engineering, University Of Seoul

요 약

건강 보조 식품, 의약품, 화장품 등 현대 제품에는 성분에 대한 제품의 구성정보가 라벨 형태로 상세히 기재 되어있다. 이러한 제품들은 실생활에서 접하기 쉽지만, 비전공자인 일반 사용자들이 이러한 성분들을 모두 기억하고 구분하여 사용하기에는 물질의 종류가 너무 많으며, 각 성분의 역할에 대해 면밀히 조사하기란 사실상 불가능하다. 하지만 제품에 대한 정확한 이해 없이는 제품을 사용 및 섭취함으로써 특정 부작용이 생길 수 있으며, 오용 및 남용할 가능성 또한 다분하다. 따라서, 제품 소비자가 사용하고 있는 제품이 어떠한 성분을 가지고 있는지를 정확히 파악할 필요가 있다. 이를 해결하기 위해, 본 논문에서는 기계 학습을 통한 객체 인식에 사용되는 실시간 객체 추적 기법을 활용하여 제품의 라벨을 1 차적으로 인식하고, 2 차적으로 라벨에 기재되어 있는 제품의 구성성분을 객체 인식하는 기법을 제안하고자 한다. 추가적으로, 해당 기법을 모바일 어플리케이션에 적용하여 건강 보조 식품 관리에 활용할 수 있는 방법에 대해 소개한다.

1. 서론

일상 생활에서 누구나 다양한 상황에서 다양한 제품을 구매하고 사용하는 상황에 직면한다. 구매자들은 일반적으로 제품을 선택할 시 디자인, 가격, 실용성 등 다양한 요소를 고려한다. 하지만 제품의 상세 구성요소를 확인해야 하는 특별한 제품들이 있다. 건강 보조 식품, 의약품, 화장품 등이 이에 해당한다. 이런 제품들은 건강과 직접적으로 연관이 있어 각별한 주의를 요하지만, 구성 물질들을 일반 사용자들이 모두 확인하기에는 무리가 있다. 이러한 상태에서 무분별하게 사용을 지속하는 과정에서 오남용의 우려가 있기에 제품의 성분들을 인지하여 구매하도록 하는 것으로 착오로 인한 부작용을 방지하여야 한다.

이미지를 텍스트로 변환하는 여러가지 도구들의 등장에 따라 많은 기법들이 활용되고 있지만, 많은 제반사항들이 존재한다. 따라서 실시간 객체 추적 기술을 통해 제품에 붙어 있는 성분 정보를 인식하여 정보들을 가공하고자 한다. 아래의 동기로부터 본 연구를 발전하였다.

- 일반적인 사용자들은 제품을 선택할 때 상세 성분까지는 파악하지 못하는 경우가 많으며, 이로 인하여 큰 부작용이 생길 여지가 존재한다.
- 일반적인 사용자가 제품에 붙어있는 구성 성분을 파악하기에는 무리가 있다. 이를 해결하기 위해 가시적이고 체계적인 분석 도구가 필요하다.
- 기존 연구들은 실시간 표시 정보 인식을 위해 현실적이고, 효과적인 기법을 제안하고 있지 못한다.

본 논문은 다음의 기여도를 갖는다.

- 제품의 구성 성분 정보를 실시간으로 파악하고 가공하여 상세한 정보를 알 수 있게 한다.
- 구성 성분 파악 시, 다단계 객체 추적을 통해 표시 정보 실시간 인식이 가능하도록 한다.
- 건강과 관련 있는 제품들을 우선하여 구성 성분과 그 함량을 추출하여 올바른 사용을 가능하게 하는 어플리케이션을 설계 및 제안한다.

앞으로의 본 논문은 다음 구성으로 이루어진다. 2 장에서는 본 논문을 이해하기 위한 배경지식을 설명한다. 그 후, 3 장에서 관련 연구를 소개하고, 4 장에서는 표시 정보 다단계 실시간 인식 기법을 상세하게 서술한다. 5 장은 서술된 인식 기법을 사례연구를 통해 그 유효성을 보인다. 6 장에서는 본 연구에서 특정하고자 하는 건강 보조 식품에서 활용할 방법을 논의한다. 그리고 7 장에서는 기존 연구와 본 연구의 특징을 비교하고 본 연구의 한계점에 대해 토의한다. 마지막으로 8 장에서는 앞으로의 연구 계획 및 결론으로 마무리한다.

2. 배경 지식

객체 탐지(Object Detection)란 영상 속에서 탐지를 원하는 객체(Label)가 어디에(x,y) 어느 사이즈(w,h)로 존재하는 지를 파악하는 기법을 의미한다. 영상을 객체화 하여 정보를 인식하려는 연구는 지속적으로 이루어졌고, 빛 번짐, 굴절, 왜곡 등의 요인으로 발생한 낮은 품질의 영상에서도 지연 없이 빠르고 정확한 인식을 목표로 하는 연구가 진행되고 있다. 일반적인 객체 탐지 기법들은 영상에서 객체를 감지하여 바운딩 박스(Bounding Box) 형태로 어느 객체인지를 표시한다.

활용되는 객체 탐지 기법으로는 R-CNN, Fast/Faster R-CNN, 그리고 최근 각광받고 있는 YOLO(You Only Look Once)가 있다. R-CNN 은 input 이미지를 2000 개 정도의 sub image 로 추출하여 CNN 을 통해 분류한다. 그 후에

SVM(Support Vector Machine)을 통해 각 Object 를 분류한다. Fast/Faster R-CNN 은 R-CNN 의 병목현상을 개선하고자, RPN(Region Proposal Networks)를 사용하여 모든 Proposal 들이 CNN 을 거치지 않고 전체 이미지를 한번에 CNN 을 거치게 한다[5].

최근에 등장한 YOLO 는 현재 가장 널리 쓰이는 빠르고 정확한 실시간 객체 인식 기법이다[1, 3]. YOLO 는 기존의 분석 기법들과 달리 하나의 신경망을 전체 이미지에 적용한다. 이 신경망은 이미지를 영역으로 분할하고, 각 영역의 바운딩 박스와 확률을 예측한다. 이 바운딩 박스는 예측된 확률에 의해 가중치가 적용된다.

YOLO 는 이미지의 전체를 보고 예측 정보를 알려주며, R-CNN 보다 1000 배 이상 빠르고, Fast R-CNN 보다 100 배 빠르다. 최대 약 45FPS (Frame Per Second)까지 처리가 가능하여 영상에서 실시간 객체 탐지가 가능하다. 하지만 YOLO 는 실시간 객체 인식에 특화되어 있어, 이미지 내의 텍스트를 추출하는 과정에는 좋은 성능을 발휘하지 못한다.

또한, 제품의 라벨을 인식하는 과정에 사용될 수 있는 다양한 광학 문자 인식(Optical Character Recognition; OCR) 관련 연구가 진행 중에 있다[2, 4, 6]. OCR 이란 실제 이미지 혹은 기계에서 인쇄한 이미지를 스캔하여 기계가 읽을 수 있는 문자로 변환하는 것을 뜻한다. 이상적으로 OCR 을 처리한 출력 형태는 입력한 값과 같아야 한다.

현재 ABBYY 의 FineReader, Adobe 의 Acrobat Pro/DC, IRIS 의 Readiris, Google 의 Tesseract OCR 프로그램 등이 널리 사용되고 있다. 문서 분석과 자연 영상에서의 문자열 추출은 문서 및 영상을 이해하기 위한 가장 기초적이고 중요한 문제이다. 문자 인식은 이미 많은 상용화된 프로그램들이 많이 있는 반면, 복잡한 문서나 자연 영상 등을 분석하고 인식하는 분야는 아직 추가 연구가 필요하다.

3. 관련 연구

[4]는 제품의 라벨 인식을 위해 OCR 기법을 사용하는 연구를 진행하였다. 해당 연구에서는 기존의 OCR 기법을 확장하여, 모바일 이미지와 라벨을 대상으로 전처리를 통해 인식 정확도를 제고하였다. 하지만 해당 연구는 실시간 영상이 아닌 단일 이미지를 대상으로 연구를 진행하였다. 따라서 이 기법을 활용하여 응용 프로그램을 개발한다면, 제품의 구성정보를 인식하기 위해 이미지를 별도로 촬영하는 추가 과정이 필요하여 그 사용이 번거로울 수 있다.

[7]은 제품의 라벨 인식을 위해 실시간 객체 추적과 OCR 기법을 사용하며, 연구의 목적이 본 논문과 대동소이하다. 다만, 해당 연구는 기존 OCR 기법을 통해 문자를 인식하는데 걸리는 시간을 간과했다는 단점이 있다.

실시간 OCR 를 위한 기법을 제안한 [6]에서 표준 PC 환경에서 평균 0.3 초가 소요되었다. 특히, 본 연구의 목적은 모바일 환경에서의 실시간 표시 정보 인식이므로, OCR 예상 소요 시간은 더욱 증가할 것이다. 즉, 비록 YOLO 를 통한 객체 인식은 실시간으로 가능하지만, 제품의 표시정보에 인쇄된 모든 문자를 인식하는데 필요한 OCR 처리 시간이 증가하기 때문에 표시정보의 모든 문자를 실시간 인식하기란 사실상 불가능하다.

따라서 본 논문에서는 이 현실적 한계를 보완하고자, 다단계 실시간 객체 추적을 활용하여 OCR 이 필요한 문자를 최소화하고, 이를 통해 OCR 처리시간을 줄이는 방식으로 실시간 표시 정보 인식이 가능하도록 하였다.

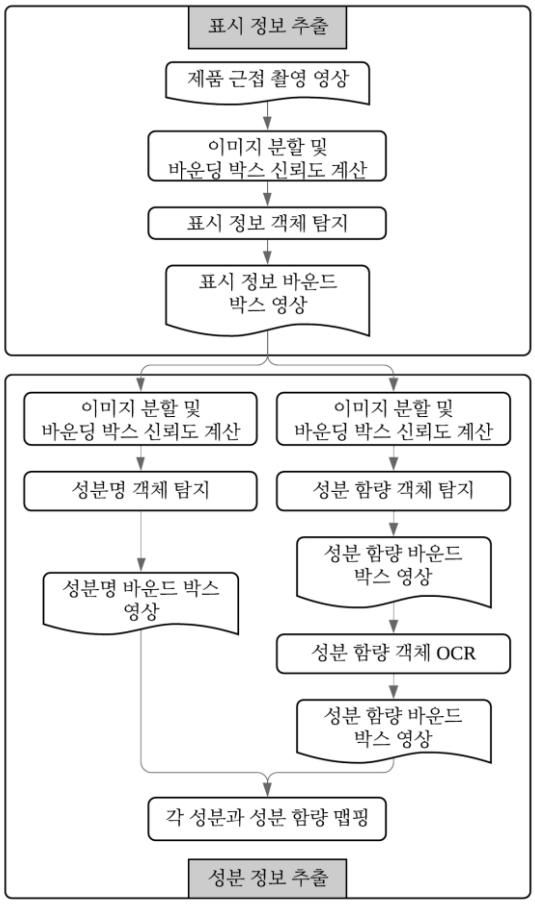
4. 표시 정보 다단계 실시간 인식 기법

4.1 개요

그림 1 은 본 논문에서 제안하는 표시 정보의 다단계 실시간 인식 기법의 개요를 보인다. 이는 제품 촬영 영상에서 표시 정보를 추출하는 단계와 추출된 표시 정보에서 성분 정보와 성분 함량 정보를 추출하는 단계로 나뉜다. 각 단계의 세부 단계는 4.2 장과 4.3 장에서 상세히 서술한다.

4.2 표시 정보 추출

표시 정보 추출 단계는 제품의 표시 정보의 객체를 인식하여 영상에서 제품 표시 정보 영상만을 추적하고, 추출하는 단계이다. 이를 위해, 제품 근접 촬영 영상을 입력받아 이미지를 그리드 형태로 분할하고 그리드에 해당하는 각 바운딩 박스의 신뢰도를 계산한다. 이후, YOLO 네트워크의 합성곱 신경망 수행을 통해 계산된 분류 신뢰도 점수로 표시 정보의 최종 바운딩 박스를 결정한다. 이 과정을 통해 표시 정보 객체를 탐지하고, 최종적으로 표시 정보 영상만을 추출하여 결과물로 생성하고 성분 정보 추출 단계로 반환한다.



(그림 1) 표시 정보 다단계 실시간 인식 기법 개요

4.3 성분 정보 추출

성분 정보 추출 단계는 이전 단계에서 결과물로 생성된 표시 정보 영상을 입력 받아 독립된 두 추출 프로세스로 나누어 성분의 정보를 추출한다. 독립된 두 추출 성분 프로세스는 각각 성분명 정보 추출, 성분 함량 정보 추출 프로세스로 구분된다.

먼저, 성분명 정보 추출 프로세스는 다양한 성분명 영상을 학습시킨 모델을 통해 성분명 객체를 분류한다. 해당 프로세스는 표시 정보 추출 단계와 동일한 기법을 사용하여, 입력 받은 표시 정보 영상에서 이미지 분할과 바운딩 박스의 신뢰도 계산을 통해 성분명 객체만을 탐지한다.

성분 함량 정보 추출 프로세스에서 성분 함량 객체는 '숫자%'의 형태로 표현된 함량으로 학습시켜, 동일한 형태를 가진 객체를 판별하게 된다. 해당 프로세스 또한 YOLO 네트워크를 사용하여 입력 받은 표시 정보 영상에서 이미지 분할과 바운딩 박스의 신뢰도 계산을 통해 성분 함량 객체만을 탐지한다. 그 후, 성분 함량 바운드 박스 영상을 대상으로 OCR 을 적용하여 성분의 함량에 해당하는 숫자를 추출한다.

마지막 단계로, 각각의 추출 프로세스에서 인식된 성분 객체와 성분 함량 객체의 영상 내 위치 정보를 구하고, 유사한 y 좌표를 가진 두 객체를 맵핑하여 추출된 데이터들을 갈무리한다.

최종적으로, 위의 단계들을 통해 실시간 객체 탐지와 OCR 기법을 활용하여 제품 표시 정보 추출 및 표시 정보의 정보추출이 가능해진다.

5. 사례 연구

그림 2 는 제품의 표시 정보를 촬영한 영상 이미지이다. Darkflow[1]를 통해 학습한 모델을 이용하여 이 영상을 대상으로 YOLO 를 사용하여 표시 정보 추출 단계를 진행하게 되면, 그리드 형태의 세부 바운딩 박스의 신뢰도 계산을 통해 가장 높은 신뢰도를 가진 최종 바운딩 박스를 결정하게 된다. 따라서, 그림 3 의 'Label'로 표시된 보라색 박스와 같이 바운딩 박스를 통해, 표시 정보의 인식을 확인할 수 있다.

(그림 2) 제품 표시 정보 객체 인식 결과

그 다음, 성분 정보 추출 단계를 진행하기 위해, 원본 영상에서 바운딩 박스 부분만을 잘라낸다. 바운딩 박스 부분만을 추출한 영상이 그림 3 이다. 그림 3 을 대상으로 성분 정보 추출 단계를 진행하게 되며, 성분명 객체와 성분 함량 객체를 독립적으로 인지할 수 있도록 한다. 이 과정은 표시 정보 추출 단계와 같이 세부 바운딩 박스의 신뢰도 계산을 통하여 최종 바운딩 박스를 결정한다. 그 결과는 그림 3 의 보라색 바운딩 박스로 표기되어 있으며, 성분명과 성분 함량이 각각의 다른 객체로 구분되어 있는 것을 확인할 수 있다.

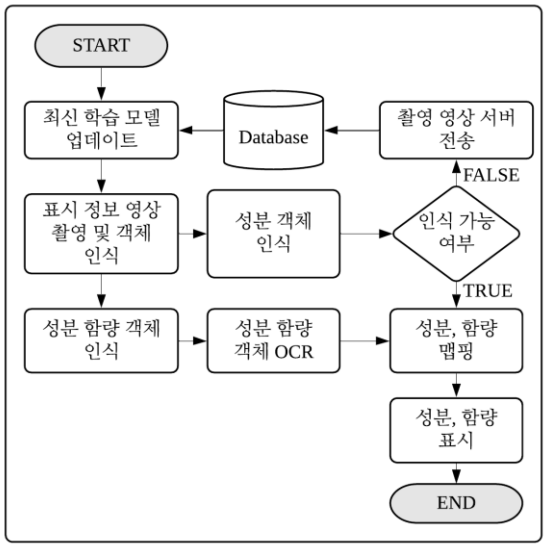
성분명의 경우, 각 성분을 "Calories", "Total Fat"과 같이 독립된 객체로 인지하게 됨을 확인할 수 있고, 성분 함량의 경우에는 "%" 문자를 포함한 "content"라는 공통된 객체로 인지하게 됨을 보인다. 그 중 'content' 객체들은 별로의 영상으로 추출되어 그림 1 의 '성분 함량 객체 OCR' 과정을 거치게 된다. 이를 통해 각 성분 함량 영상에 들어있는 문자들인 '0%', '4%'와 같은 문자열을 추출한다.

마지막으로, 각각 인지된 객체는 영상에서의 위치 정보를 저장하여, 각 성분명에 대응하는 성분 함량을 맵핑하게 된다. 예를 들어, 'Total Fat' 성분 객체는 content 객체들 중 유사 y 좌표를 가진 '0%' 와 맵핑되고, 'Dietary Fiber' 성분 객체 또한 content 객체들 중 유사 y 좌표를 가진 '4%'와 맵핑된다. 하지만 해당 맵핑 과정에서 'Calories'의 경우 유사 y 좌표를 가진 content 객체가 존재하지 않으므로, 그 성분 함량을 확인할 수 없다.

최종적으로 'Total Fat', 'Saturated Fat', 'Cholesterol' 등의 성분과 그 함량 데이터를 표시 정보로부터 추출할 수 있다.

(그림 3) 표시 정보 성분 정보 추출 결과

6. 건강 보조 식품 관리 응용 프로그램



(그림 4) 제안 기법의 응용 개요

본 연구의 최종 목적인 건강 보조 식품 관리 응용 프로그램은 다음과 같은 구조를 가진다. 미리 객체 탐지 기법과 OCR 처리를 한 데이터를 학습시켜 모델을 준비한다. 사용자가 모바일 환경에서 실시간으로 영상을 촬영하면, 준비

[1] https://github.com/thtrieu/darkflow

된 학습 모델을 통해 신경망에서 1 차적으로 성분 객체 인식을 시도한다. 정상적인 경우에는 2 차적으로 성분 함량 객체를 추가로 인식하여 OCR 을 진행한다. 이후에 성분 객체와 맵핑을 진행하여 사용자의 모바일 환경에서 가시적으로 확인할 수 있도록 한다. 하지만 성분 객체 인식에 실패할 경우에는 촬영 영상을 서버로 전송하여 Database 에 저장하고, 모델에서 추가로 학습을 진행하여 모델의 성능을 개선시킨 후에 업데이트를 하여 실패했던 영상의 객체 인식을 성공할 수 있도록 진행한다.

7. 토의

(표 1) 관련 연구와의 비교

	본 연구	[4]	[7]	[6]
실시간 정보 추출	○	×	△	△
OCR	○	○	○	○
객체 인식	○	×	○	×
다단계 객체 인식	○	×	×	×
대상 영상	실시간 영상	이미지 영상	실시간 영상	실시간 영상

해당 장에서는 제품 표시 정보로부터 유의미한 정보를 추출하는 것을 목표로 하였을 때, 본 연구와 기존 연구 간의 정성적 비교와 본 연구의 한계점과 보완점에 대해 토의한다. 표 1 은 본 연구와 기존 연구와의 정성적 비교 표이다. 본 연구는 실시간 정보 추출 목표를 충족하는 반면, [4] 연구는 실시간 정보 추출이 불가능하고, [6]과 [7]은 제한적으로만 가능하다. 이 목표를 이루기 위해 본 연구를 포함한 모든 연구는 공통적으로 OCR 기법을 적용하였다. 하지만 본 연구와 [7]은 추가적으로 실시간 객체 인식을 사용한다. 특히, 본 연구는 다단계 객체 인식을 사용한다. 마지막으로, 이미지 영상을 정보 추출 대상으로 삼는 [4]를 제외하고는 모두 실시간 영상을 대상으로 삼는다.

본 연구의 한계점과 보완점

본 연구가 갖는 한계점들이 있으며, 그에 따른 보완점은 다음과 같다.

- ✓ 성분명이 영어가 아닌 한국어나 기타 언어로되어 있는 경우가 존재할 수 있음 : 동일한 성분이 다른 언어로 쓰인 데이터셋을 따로 학습시키고, 사용할 때 언어에 따라 다른 학습 모델을 가져와 작동시킬 수 있도록 한다.
- ✓ %표시만 인식이 되고, g 으로 표기된 함량은 인식이 되지 않음 : g 으로 표기된 것들은 추후 연구에서 추가적인 객체탐지 보완 기법을 통해 보완하도록 한다.
- ✓ 모든 성분명을 학습시키기 위해서는 많은 비용과 시간이 듦 : 데이터 셋 가공은 실시간이 아니어도 되므로 기존 OCR 기법을 활용하여 성분명에 바운딩 박스를 그리는 기법을 추가 연구한다.

8. 결론 및 향후 연구

본 논문에서는 실시간 영상 객체 추적 기법인 YOLO 를 통해 제품들의 라벨 표시 정보 객체를 인식하고 성분 객체와 성분 함량 객체로 분리한 후, 성분 함량 객체를 OCR 기법을 통해 원하는 텍스트를 추출하여 각 성분과 성분 함량을 맵핑하는 기법을 제안한다.

기존 OCR 은 다량의 데이터를 실시간으로 처리하지 못하여 그 사용이 불편하다는 점을 해결하고자, 단면적인 사진 대신 영상으로 인식한 데이터를 다단계 객체 인식을 통해 전처리하여, 성분 표시 정보로부터 정보 추출을 최적화하였다. 이를 통해 실시간으로 정보 추출이 가능하도록 하였다.

또한, 사용자에게 편리함을 제공하고자 모바일 어플리케이션에서 촬영한 영상을 가공하여 필요한 정보를 확인할 수 있는 응용 프로그램을 제시하였다.

모바일 어플리케이션에서 서버-클라이언트 방식을 이용할 때 전송 과정 중의 지연 시간으로 인하여 소요시간이 증가되는 문제점을 보완하고자, 객체를 다단계로 인식하여 OCR 처리과정을 필수 부분으로 최소화하여 사용자가 최소 시간으로 필요한 정보를 수집할 수 있도록 한다. 향후에는 최소 정보로 최소 시간에 OCR 처리를 할 수 있도록 연구하고자 한다.

Acknowledgement

이 논문은 2019 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 -현장맞춤형 이공계 인재양성 지원사업의 지원을 받아 수행된 연구임(NRF-2017H1D8A1030582).

참고문헌

[1] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., "You Only Look Once: Unified, Real-Time Object Detection," In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[2] Coates, A., Carpenter, B., Case, C., Satheesh, S., Suresh, B., Wang, T., Wu, D.J., Ng, A.Y., "Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning," In Proc. of International Conference on Document Analysis and Recognition (ICDAR), Oct. 2011.

[3] Redmon, J., Farhadi, A., "YOLOv3: An Incremental Improvement," Technical Report, University of Washington, Apr. 2018, arXiv:1804.02767.

[4] Grubert, O., Gao, L., "Recognition of Nutrition Facts, Labels from Mobile Images," Technical Report, Stanford University, Apr. 2014.

[5] 주미소, "딥 러닝 및 영상 처리 기법을 활용한 실시간 객체 분리 연구," 고려대학교 대학원, 석사학위 논문, 2018.

[6] Neumann, L., Matas, J., "Real-Time Scene Text Localization and Recognition", In Proc. of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[7] 이철훈, 민경식, 이병정, "객체 추적과 OCR 을 통한 표시 정보의 실시간 인식", 한국정보과학회 한국컴퓨터종합학술대회, 2019.