한국사물인터넷학회논문지 Vol. 5, No. 2, pp. 15-20, 2019 https://doi.org/10.20465/KIOTS.2019.5.2.015

윈도우 기반의 광학문자인식을 이용한 영상 번역 시스템 구현

황선명[1], 염희균[2*]
[12]대전대학교 컴퓨터공학과 교수

An Implementation of a System for Video Translation on Window Platform Using OCR

Sun-Myung Hwang[1], Hee-Gyun Yeom[2*]
[12]Professor, Department of Computer Engineering, Daejeon University

요 약 기계학습 연구가 발달함에 따라 번역 분야 및, 광학 문자 인식(Optical Character Recognition, OCR) 등의 이미지 분석 기술은 뛰어난 발전을 보였다. 하지만 이 두 가지를 접목시킨 영상 번역은 기존의 개발에 비해 그 진척이 더딘 편이다. 본 논문에서는 기존의 OCR 기술과 번역기술을 접목시킨 이미지 번역기를 개발하고 그 효용성을 검증한다. 개발에 앞서 본 시스템을 구현하기 위하여 어떤 기능을 필요로 하는지, 기능을 구현하기 위한 방법은 어떤 것이 있는지 제시한 뒤 각기 그 성능을 시험하였다. 본 논문을 통하여 개발된 응용프로그램으로 사용자들은 좀 더 편리하게 번역에 접근할 수 있으며, 영상 번역이라는 특수한 환경으로 한정된 번역기능에서 벗어나 어떠한 환경에서라도 제공되는 편의성을 확보하는데 기여할 수 있을 것이다.

주제어 : 기계학습, 광학문자인식(OCR), 이미지 번역기, 기계번역, 영상 번역

Abstract As the machine learning research has developed, the field of translation and image analysis such as optical character recognition has made great progress. However, video translation that combines these two is slower than previous developments. In this paper, we develop an image translator that combines existing OCR technology and translation technology and verify its effectiveness. Before developing, we presented what functions are needed to implement this system and how to implement them, and then tested their performance. With the application program developed through this paper, users can access translation more conveniently, and also can contribute to ensuring the convenience provided in any environment.

Key Words : Machine learning, Optical Character Recognition, Image translator, Machine translation, Video translation

1. 서론

수많은 정보를 처리함에 따라 정보는 일정한 패턴을 가지게 됨을 인식하게 되었으며, 이를 처리하고 계산할 시스템으로서 인공지능의 개발 필요성이 대두되었다. 변화가 가져온 큰 영향으로서는 광학적 문자 인식 기술과 번역 기술이다[1,2,3,4]. 단순히 문자를 대조하던 시대에서 수많은 문자 패턴을 학습하여 문자의 인식률을 높인

*교신저자 : 염희균(yeom@dju.kr)
접수일 2019년 9월 21일 수정일 2019년 11월 15일 심사완료일 2019년 11월 24일

광학 문자 인식(Optical Character Recognition, OCR) 기술, 그리고 수많은 문장 정보를 저장하고 학습하여 이를 번역하는 신경망 번역(Neural Machine Translation, NMT) 기술은 과거 번역할 언어를 문자 단위로 번역하던 구문 기반 기계번역(Phrase Based Machine Translation, PBMT) 방식에 비해 월등히 높아졌다. 이를 나타내는 대표적인 지표가 바로 구글 번역 어플리케이션이다. 번역뿐만이 아닌 OCR기능을 지원하면서 사용자는 어디서든 원하는 화면에 초점을 맞추어 번역을 제공받을 수 있게 되었다. 하지만 OCR 기능이 지원되던 모바일의 환경과는 달리 데스크톱 환경에서는 번역 기능에 OCR의 지원을 하지 않는다. 특히 이 문제는 PDF와 같은 장문의 텍스트 이미지 또는 수많은 반복 번역이 필요한 자막 생성과 같은 상황에서는 그 문제점이 크게 나타난다[5,6,7]. 이러한 문제를 해결하기 위하여 본 논문은 기존에 존재하는 OCR 기능과 번역 기술을 접목시켜 데스크톱 기반의 이미지를 해독하여 텍스트를 추출하고, 추출한 텍스트를 번역하는 이미지 번역기 프로그램을 개발하여 사용자의 편의를 확대하고자 한다.

2. 관련연구

2.1 Google Cloud Vision API

Google Cloud Vision API는 구글에서 제공하는 이미지 분석 서비스로서 클라우드 기반으로 동작하며 머신러닝 학습 모델을 이용하여 이미지 안의 개별 객체를 인식하여 수천 가지 카테고리로 분류하거나 성인 콘텐츠에서부터 폭력적인 콘텐츠에 이르기까지 다양한 유형의 부적절한 콘텐츠를 감시하는 기능을 제공한다. REST API를 제공하기 때문에 사용자가 이미지 파일에 대한 정보를 로컬 또는 리모트 상에서 제공할 수 있으며 이에 대한 메타 파일을 JSON 구조로 받을 수 있어 정보에 대한 추출 및 가공이 쉽다는 특징을 가지고 있다. 또한, Vision API는 OCR을 사용해 이미지에서 50개가 넘는 언어와 다양한 파일 형식의 텍스트를 감지하는 장점이 있다. 〈Table 1〉은 Google Cloud Vision API에서 제공하는 서비스 정보를 나타낸다.

OCR 서비스는 이미지에서 문자(텍스트)를 감지하고 추출한다. 간판이나 표지판이 찍힌 사진을 예로 들 수 있다. JSON 은 추출된 전체 문자열과 함께 개별 단어와 해당 경계 상자를 포함한다.

이미지 속성 감지 API는 이미지의 주요 색상과 같은 일반적인 속성을 감지하는 기능이다.

얼굴 감지 API는 이미지에서 여러 개의 얼굴을 감정 상태와 같은 주요 얼굴 관련 속성과 함께 감지하는 기능이다. 안면 인식은 지원되지 않는다.

로고 감지 API는 이미지에서 인기 제품 로고를 감지하는 기능이다.

라벨 감지 API는 라벨을 통해 물체, 장소, 활동, 동물종, 상품 등을 식별할 수 있다.

〈Table 1〉 Google Vision API Feature

Feature	Description
Optical Character Recognition	Detect and extract text within an image, with support for a broad range of languages, along with support for automatic language identification
Image Attributes	Detect general attributes of the image, such as dominant colors and appropriate crop hints
Face Detection	Detect multiple faces within an image, along with the associated key facial attributes like emotional state or wearing head wear
Logo Detection	Detect popular product logos within an image
Label Detection	Detect broad sets of categories within an image, ranging from modes of transportation to animals
Explicit Content Detection	Detect explicit content like adult content or violent content within an image
Landmark Detection	Detect popular natural and man-made structures within an image
Web Detection	Search the internet for similar images

2.2 신경망 기반 기계번역 모델

딥러닝이 성공적으로 적용되는 대표적인 자연어 처리 분야가 기계번역이라고 할 수 있는데, 신경망(Neural Networks) 기반 기계번역(Neural Machine Translation, NMT)은 하나의 신경망으로 번역 모델이 구성되고 학습된다는 측면에서 기존 여러 모듈에 기반한 기계 번역과 다른 패러다임을 제시하고 있다[8,9,10]. 일반적으로 NMT는 인코더(Encoder)와 디코더(Decoder)로 구성되는데, 단어들로 구성된 입력문장을 인코더가 벡터공간에 표현하고, 이를 디코더가 다시 출력 문장의 단어들을 하나씩 순차적으로 만들어 내는 것으로 번역 과정이 진행된다. 이러한 과정은 전통적인 기계번역 시스템이 단어들을 심볼(Symbols) 수준에서 직접 다루는 것과 상반된다[11].

NMT 모델은 단일 신경망 구조를 사용하는 End-to-end 방식의 신경망 번역 모델로, 언어별, 텍스트 유형별로 학습 정도에 따라 차등이 있으나, 두 줄 정도 길이의 문장 내에서는 문맥 파악이 이루어져 결과물의 의미적, 통사적

완성도가 크게 높아진 것을 확인 할 수 있다. 그 외 전반적으로 보이는 문제점으로는 누락, 부적절한 직역으로 인한 의미 전달 오류, 불필요한 표현의 반복, 문장 단위를 벗어난 전문용어의 불일치, 단복수 오류 등이 있다.

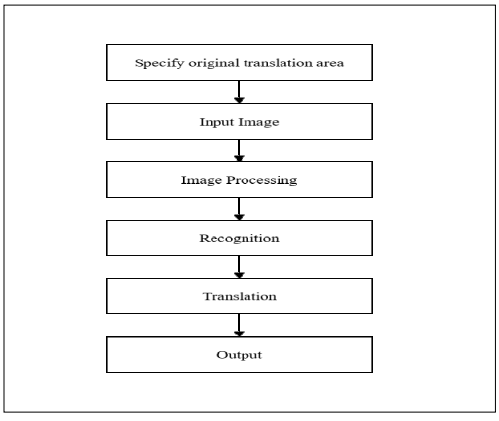
2.3 OCR 엔진

OCR(광학문자인식) 분야는 오랫동안 관심 있게 연구가 진행되어온 분야로서 ABBYY의 TextGrabber의 Google의 Cloud Vision API등을 대표적인 예로 들 수 있다. Google Cloud Vision API는 학습된 기계 학습 모델을 사용해 이미지의 내용을 파악하고 이미지 안의 개별 객체를 감지하고 인쇄된 단어를 찾아주는 등 이미지 처리에 유리한 이점을 갖고 있는 OCR 엔진이다 [12,13,14,15].

3. 제안 시스템 구조

3.1 개요

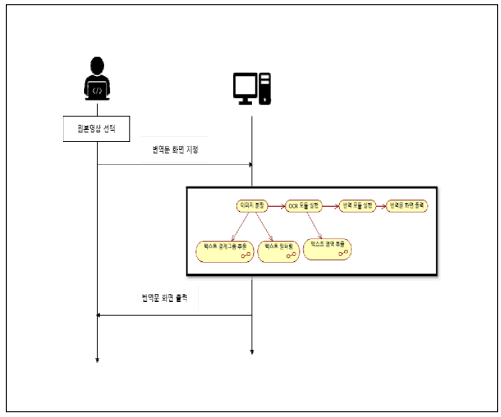
제안 시스템은 크게 원본 영상에서 번역 영역을 지정한 후 이미지로 저장 한 후, 이미지 처리 과정, 인식 과정 그리고 번역 과정을 거쳐 마지막 출력 과정으로 이루어지며, 각각의 과정을 통해 최종적으로 원하는 영상의 번역 결과를 화면에 출력하게 된다. [Fig. 1] 은 제안 하는 시스템의 흐름도이다.



[Fig. 1] Flowchart of the proposed system

영상 화면에서 번역 영역을 지정하여 이미지로 변환한 후, 이미지에서 텍스트를 추출하고, 이를 번역하여 출력하는 영상 번역 시스템이다. 사용자는 영상 화면에서 번역 화면 영역 이미지를 지정하고, 그 번역 결과를 확인한다. 먼저 이미지 분할과정은 번역 화면 이미지에서 문자로 인식되는 경계 그룹을 추출하고, 필터링은 비텍스트 영역을 제거하여 경계 그룹의 텍스트 인식률을 높인다. OCR 모듈은 이미지 분할을 통해 추출된 텍스트를 인식하여 정확한 텍스트 영역을 추출한다.

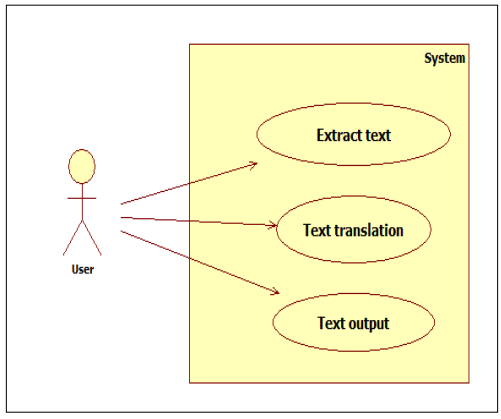
번역 모듈은 추출된 텍스트를 이용하여 지정된 언어로 번역을 한다. 화면 출력은 번역된 텍스트를 지정된 영역에 출력한다. [Fig. 2]는 제안하는 시스템의 구조이다.



[Fig. 2] System Architecture

3.2 주요 기능

[Fig. 3]과 같이 유즈케이스 모델의 주요 기능 3가지로, 첫 번째 원본 영상에서 번역을 원하는 텍스트 영역 추출 기능, 텍스트 번역 기능, 텍스트 출력 기능으로 나누어 설명한다.



[Fig. 3] UseCase Model

3.2.1 텍스트 추출 기능

텍스트 추출 기능은 원본 영상에서 번역 화면 영역을 지정 하여 이미지 분할을 통해 텍스트 영역을 추출 하고 OCR 기능을 수행하여 정확한 텍스트를 추출하는 기능이다.

3.2.2 텍스트 번역 기능

텍스트 번역 기능은 텍스트 추출 기능에서 추출된 텍스트를 입력 받아 번역 API 모듈을 이용하여 번역되는 기능이다.

3.2.3 텍스트 출력 기능

텍스트 출력 기능은 번역된 텍스트를 입력 받아 원래의 지정된 영역에 출력 하는 기능이다. 번역된 텍스트는 기존 화면에 Blur 처리를 하여 기존 텍스트와의 겹침 현상을 방지한다.

4. 기능 구현 및 평가

4.1 이미지 분할 기능 구현

이미지 분할 기능은 OpenCV라이브러리로 구현 하였다. 〈Table 2〉 와 같이 이미지 분할 과정을 통해 텍스트를 추출한다. 첫 번째는 컬러 이미지를 회색조 이미지로 변환한다. 두 번째는 외곽선 추출을 위해 팽창 및 침식의 모폴로지 연산(MorphologyEx)을 수행한다. 세 번째는 처리된 이미지에 임계값을 적용하여 흑백 이미지를 얻어낸다. 네 번째는 모폴로지 팽창 연산을 적용한 이미지에 침식 연산을 다시 적용하여 이미지의 경계를 강조시킨다.

〈Table 2〉 Boundary Extraction Process

Step 1	Step 2	Step 3	Step 4
Gray scale transformation	Outline extraction	Apply threshold	Close operation

4.2 OCR 구현

윈도우 PC 환경에서 OCR, 즉 광학문자 인식 기능은 Tesseract-OCR과 Vision API 라이브러리로 구현 하였다. 이미지 상에 존재하는 텍스트를 추출해내는 단계로서 처리속도뿐 만이 아닌, 정확도 역시 높일 수 있는 방법으로 2개의 라이브러리를 사용한다. Vision API는 단순히 문자뿐만이 아닌 이미지를 분석하는 기능으로서 화면의 문자 인식률이 높을 뿐만 아니라, 문자의 위치 좌표, 문자의 언어 등 다양한 정보를 제공 해준다.

4.3 텍스트 번역 구현

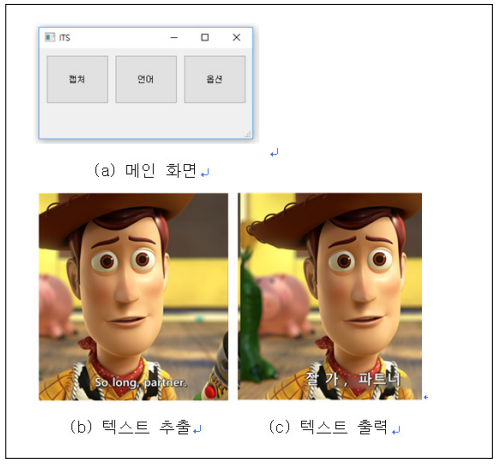
텍스트 번역은 구글 번역 API로 구현 하였다. 구글 번역은 NMT 방식을 이용하며 문장 전체를 학습, 분석하여 문장 전체의 문맥을 파악하고 불안정한 번역을 바로 잡는 기능을 제공한다.

4.4 텍스트 출력 구현

번역된 문장을 디스플레이에 출력시키기 위해 기존 화면에 Blur 처리를 하여 기존 텍스트와의 겹침 현상을 방지하였다. 기존 화면에 Blur 처리를 하여 기존 텍스트와의 겹침 현상을 방지하는 방법으로 기존 화면에 자연스러운 출력이 가능하다.

4.5 인터페이스 구현

윈도우 기반의 인터페이스 [Fig. 4]의 (a)로 이루어져 있다. [Fig. 4]는 윈도우 기반 제안 시스템의 메인 실행화면부터 순차적으로 나타낸다. [Fig. 4](a)는 메뉴를 나타내고, '캡쳐' 버튼을 클릭하면 [Fig. 4](b) 하나의 이미지에서 텍스트를 잘라내어 '언어' 버튼을 클릭한다. [Fig. 4](c)는 텍스트 번역 출력 결과를 나타낸다.



[Fig. 4] System User Interface

4.6 제안 시스템 평가

제안한 시스템의 예상 기대 시간은 약 1~2초 정도로 기대되며, 텍스트 인식 정확도는 98퍼센트로 예상된다. 이 시스템의 가장 큰 장점은 바로 편의성과 범용성이다. 윈도우 PC 환경에서 사용자가 원하는 이미지 영역을 지정하여 번역을 제공받을 수 있으며, 별도의 출력 영역이 할당되지 않더라도 기존 화면에 직접적으로 확인이 가능하다. 또한 영상을 그대로 분석하여 번역하기 때문에 문자가 존재하는 어떠한 환경에서라도 번역이 가능하다.

기존 시스템과의 평가는 다음 〈Table 3〉과 같다. 제안 시스템은 출력 속도는 느리나 문자 인식률이 높다. 기존 시스템 출력은 별도 출력 영역에 하는 반면, 제안 시스템은 기존 원본의 텍스트 영역에 출력을 하게 된다. 따라서 보다 자연스러운 번역 결과를 확인 할 수 있다.

〈Table 3〉 System Evaluation

Evaluation item	Existing system	Suggestion system
Character recognition rate	middle	height
Translation rate	middle	height
Output speed	0.5~1s	1~2s
Output method	Other area	Original area

5. 결론

본 논문에서는 OCR 기술과 번역 기술을 접목시킨 이미지 번역 시스템을 개발하였다. 사용자가 프로그램을 실행시키면 일련의 프로세스를 통하여 번역문을 제공받게 된다. 원하는 영역을 설정한 후 화면 인식 단계로 넘어가며, 텍스트 위치 분석이 끝난 뒤에는 OCR 단계로 넘어간다. OCR로 파악된 텍스트는 번역되며 텍스트 출력 단계에서 지정된 디스플레이 영역에 출력된다. 이는 사용자에게 큰 편의성을 제공하며, 동시에 어디서든 사용할 수 있게 하는 범용성 역시 제공한다.

REFERENCES

[1] K.H.Cho, et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” arXiv preprint arXiv: 1406.1078, 2014.

[2] B.Dzmitry, K.H.Cho, and Y.Bengio, “Neural machine translation by jointly learning to align and translate,” arXiv preprint arXiv:1409.0473, 2014.

[3] Tu, Zhaopeng, et al., “Context gates for neural machine translation,” Transactions of the Association for Computational Linguistics 5, pp.87-99, 2017.

[4] V.Ashish, et al., “Attention is all you need,” Advances in Neural Information Processing Systems, 2017.

[5] Ma, Mingbo, et al., “Osu multimodal machine translation system report,” arXiv preprint arXiv:1710.02718, 2017.

[6] Madhyastha, P.Swaroop, J.Wang, and L.Specia, “Sheffield multimt: Using object posterior predictions for multimodal machine translation,” Proc. of the Second Conference on Machine Translation, 2017.

[7] Caglayan, Ozan, et al., “Lium-cvc submissions for wmt17 multimodal translation task,” arXiv preprint arXiv:1707.04481, 2017.

[8] N.Kalchbrenner and P.Blunsom, “Recurrent continuous translation models,” EMNLP, 2013.

[9] I.Sutskever, O.Vinyals, Q.V.Le, “Sequence to Sequence Learning with Neural Networks,” Advances in Neural Information Processing Systems (NIPS), 2014.

[10] D.Bahdanau, K.Cho and Y.Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” Int'l Conf. on Learning Representations (ICLR), 2015.

[11] P.Koehn, “Statistical Machine Translation. Statistical Machine Translation,” Cambridge University Press, ISBN 9780521874151, 2010.

[12] R.Mithe, S.Indalkar, and N.Divekar, “Optical character recognition,” International Journal of Recent Technology and Engineering, Vol.2, pp.72-75, 2013.

[13] E.B.Go, Y.J.Ha, S.R.Choi, K.H.Lee, and Y.H.Park, “An implementation of an android mobile system for extracting and retrieving texts from images,” Journal of Digital Contents Society, Vol.12, No.1, pp.57-67, 2011.

[14] M.H.Cho, “A study on character recognition using wavelet transformation and moment,” Journal of The Korea Society of Computer and Information, Vol.15, No.10, pp.49-57, 2010.

[15] J.W.Song, N.R.Jung, and H.S.Kang, “Container BIC-code region extraction and recognition method using multiple thresholding,” Journal of the Korea Institute of Information and Communication Engineering, Vol.19, No.6, pp.1462-1470, 2015.

황 선 명(Sun-Myung Hwang) [정회원]

- 1984년 2월 : 중앙대학교 전자계산학과 (이학석사)
- 1987년 2월 : 중앙대학교 전자계산학과 (이학박사)
- 1989년 3월 ~ 현재 : 대전대학교 컴퓨터공학과 교수

〈관심분야〉
인공지능, 차세대 로봇, 클라우드 컴퓨팅

염 희 균(Hee-Gyun Yeom) [정회원]

- 2002년 2월 : 대전대학교 컴퓨터공학과 (공학석사)
- 2007년 8월 : 대전대학교 컴퓨터공학과 (공학박사)
- 2012년 3월 ~ 2018년 2월 : 대전대학교 강의전담
- 2018년 3월 ~ 현재 : 대전대학교 컴퓨터공학과 강사

〈관심분야〉
사물인터넷, 정보통신, 빅데이터, 기계학습