

안드로이드 환경에서 화면 캡처와 OCR을 활용한 실시간 번역

애플리케이션 개발

이승우⁰, 김성진*, 윤영현**, 백재순(교신저자)*

⁰명지전문대학 ICT융합공학과,

*명지전문대학 ICT융합공학과,

**명지전문대학 정보통신공학과

e-mail: {2022821013⁰, ict214548*, yhyoon**, hisoon99*}@mj.ac.kr

Development of a Real-time Translation Application using Screen Capture and OCR in Android Environment

Seung-Woo Lee⁰, Sung Jin Kim*, Young Hyun Yoon**, Jai Soon Baik(Corresponding Author)*

⁰Dept. of ICT Convergence Engineering, Myongii College,

*Dept. of ICT Convergence Engineering, Myongii College,

**Dept. of Information Technology communication, Myongii College

● 요약 ●

본 논문은 안드로이드에서 화면 캡처와 OCR을 통한 실시간 번역 애플리케이션 개발을 주제로 한다. 코틀린으로 개발된 애플리케이션은 사용자가 원하는 화면 영역을 캡처하여 해당 텍스트를 OCR로 추출하고, 구글 Cloud Vision API와 Cloud Translation API를 활용해 번역한다. 이를 통해 외국어 애플리케이션 사용의 편의성을 향상시키고, 정보의 이해와 공유를 도울 수 있음을 제시한다. 이 기술은 더욱 다양한 분야에서 활용 가능성을 열어놓고 있다.

키워드: 실시간 번역 (Real-time Translation), 화면 캡처 (Screen Capture), OCR (Optical Character Recognition), 안드로이드 어플리케이션 (Android Application), 클라우드 API (Cloud API)

I. Introduction

본 논문은 사용자가 자유롭게 지정한 화면 영역의 캡처와 이를 OCR을 통한 텍스트 추출, 그리고 해당 텍스트의 실시간 번역을 가능하게 하는 안드로이드 어플리케이션 개발에 관한 것이다. 현대 사회에서는 스마트폰이 일상생활의 필수품이 되었고, 다양한 정보를 얻기 위해 수많은 앱을 사용하게 된다. 그러나 외국어로 제공되는 앱이나 정보에 접근하는 것은 언어장벽으로 인해 쉽지 않다. 이러한 문제를 해결하기 위해 본 연구에서는 실시간으로 특정 화면을 캡처하고, 해당 내용을 OCR과 번역 API를 통해 사용자가 이해할 수 있는 언어로 번역하는 어플리케이션을 구현한다. 이 어플리케이션은 안드로이드 11 이상에서 구동되며, 사용자는 번역 대상 언어와 번역될 언어를 설정할 수 있다.

II. Preliminaries

1. Target Research

"딥러닝 기반 open OCR API 성능 비교 분석" 연구는 구글, 네이버, 카카오의 OCR API를 비교 분석하였다. 이 논문에서는 각 API의 언어 지원 범위, 인식 가능한 최대 회전각도, 글자수에 따른 인식 속도 등을 비교하였으며, 구글의 OCR API가 가장 뛰어난 성능을 보였다[1]. 다른 관련 기술로는 Google Translate와 네이버 파파고 같은 서비스가 사용자가 입력한 텍스트를 실시간으로 다른 언어로 번역해주는 기능을 제공하고 있다. 그리고 번역한 글자를 이미지에 자연스럽게 합성하는 수준 높은 기능을 제공하고 있다. 또한, 삼성의 갤럭시S엔 번역 기능은 사용자가 S엔으로 단어를 선택하면 그 단어를 다른 언어로 번역해주는 기능을 제공한다. 본 논문에서 개발하는 어플리케이션의 차별점은 화면의 글자를 복사하거나 캡처하여 번역하는 기존의 번역 방식과는 달리, 현재 화면의 글씨를 실시간으로 인식하여 번역한다는 점이다. 사용자는 복잡한 단계를 거치지

않고도 어플리케이션 화면의 외국어를 이해할 수 있다. 이는 복잡한 인터페이스나, 화면 복사 및 캡처가 어려운 어플리케이션에서 유용하고, 사용자가 다양한 언어로 제공되는 어플리케이션의 접근성을 높일 수 있다.

번역 기능을 제공함으로써, 사용자가 외국어로 제공되는 어플리케이션의 정보를 빠르게 이해하는 데 도움을 줄 수 있었다. 미래에는 이 기술을 활용해 더욱 다양한 분야에서 응용할 수 있는 가능성이 열려 있으며, 이를 통해 전세계 사람들이 정보를 효과적으로 공유하고 이해하는 데 기여할 수 있을 것으로 기대된다.

III. The Proposed Scheme

1. Implement

본 논문에서는 사용자가 지정한 영역의 화면 콘텐츠를 캡처하고 이를 실시간으로 번역하는 안드로이드 어플리케이션을 구현하였다. 이를 위해 MediaProjection API를 사용하여 사용자의 화면을 캡처한다. 이 API는 안드로이드 5.0 (API 레벨 21) 이상에서 동작하며, 사용자의 동의를 통해 화면 콘텐츠에 접근할 수 있다. 사용자로부터 동의를 받은 후에는 startCapture() 메서드를 호출하여 화면 캡처를 시작하였다.

두 번째로, WindowManager를 사용하여 다른 앱 위에 표시할 수 있는 플로팅 버튼을 추가하였다. 이 플로팅 버튼은 WindowManager.LayoutParams를 통해 위치와 크기를 설정하였고, 사용자가 버튼을 드래그하여 원하는 위치로 이동할 수 있게 구현하였다.

세 번째로, Foreground Services를 사용하여 플로팅 버튼이 백그라운드에서도 계속 동작하고, 사용자의 화면 캡처 요청을 기다릴 수 있도록 하였다. 이를 통해 다른 어플리케이션을 사용하는 동안에도 어플리케이션이 동작을 중지하지 않고 화면 캡처와 번역 요청을 처리할 수 있다.

네 번째로, Google Cloud Vision API를 사용하여 캡처된 화면 이미지에서 텍스트를 추출한다. 이 API는 높은 정확도로 이미지 안의 텍스트를 인식하고 추출할 수 있다.

마지막으로, OCR을 통해 추출된 텍스트를 Google Cloud Translation API를 사용하여 한국어로 번역한다. 이 API는 머신러닝 기반의 번역 기술을 제공하며, 사용자가 어플리케이션에서 설정한 대상 언어를 정확하게 인식하여 해당 언어를 한국어로 번역할 수 있다.

이렇게 구현된 기능들을 조합하여 사용자는 외국어로 작성된 어플리케이션 화면을 실시간으로 한국어로 번역받을 수 있는 어플리케이션을 이용할 수 있다. 본 논문에서 개발하는 앱은 안드로이드 11 이상의 버전에서 동작하며, 코틀린으로 개발되었다.

IV. Conclusions

본 논문에서는 코틀린을 이용하여 안드로이드 환경에서 작동하는 실시간 캡처 및 번역 어플리케이션을 제안하였다. 이 어플리케이션은 사용자가 화면에서 원하는 영역을 캡처하여 해당 영역의 텍스트를 OCR로 추출하고, 번역하는 기능을 제공한다. 사용자는 자유롭게 번역하고자 하는 언어와 번역된 언어를 선택할 수 있으며, 이를 통해 외국어 어플리케이션 사용의 편의성을 높일 수 있다. Google의 Cloud Vision API와 Cloud Translation API를 활용하여 고품질의 OCR과

REFERENCES

- [1] Dong-Ik Jang, Eun-Jung Choi, Sung-Min Woo.(2022). Performance Comparison Analysis of Deep learning-based open OCR API. 대한전자공학회 학술대회,(),1189-1191.
- [2] MediaProjection - Android Developers, <https://developer.android.com/reference/android/media/projection/MediaProjection>
- [3] WindowManager - Android Developers, <https://developer.android.com/reference/android/view/WindowManager>
- [4] Foreground services - Android Developers, <https://developer.android.com/guide/components/foreground-services>