

# 시각 장애인을 위한 음성 책 읽기 기능을 갖춘 스마트 안경 시스템 개발

명지훈, 신수용\*

국립금오공과대학교 전자공학과, 국립금오공과대학교 IT융복합공학과\*

금오공과대학교

[20220402@kumoh.ac.kr](mailto:20220402@kumoh.ac.kr), [wdragon@kumoh.ac.kr](mailto:wdragon@kumoh.ac.kr)

## Development of a Smart Glasses System with Audio Book Reading Capabilities for the Visually Impaired

Myoung-Ji Hoon, Shin-Soo Young\*

Kumoh National Institute of Technology

### 요약

본 논문은 시각적 정보 접근에 어려움을 겪는 이용자가 별도의 대체 자료 없이도 일반 인쇄물을 즉각적으로 이해할 수 있도록 지원하는 NVIDIA Jetson 기반의 휴대용 End-to-End OCR - TTS 스마트 안경 시스템을 제안한다. 제안된 장치는 카메라를 통해 입력된 텍스트 이미지를 Transformer 기반 OCR 모델로 실시간 인식하고, 그 결과를 딥러닝 기반 TTS 시스템을 통해 자연스러운 음성으로 변환한다. 모든 처리 과정은 단일 파이프라인으로 통합되어 있으며, 고성능 임베디드 프로세서인 NVIDIA Jetson에서 온디바이스 방식으로 구동되므로 외부 네트워크 연결 없이도 빠르고 안정적인 정보 제공이 가능하다.

### I. 서론

시각적 정보 접근이 어려운 이용자는 책, 안내문, 메뉴판, 제품 설명서 등 일상에서 접하는 다양한 인쇄물을 실시간으로 이해하는 데 큰 제약을 겪고 있다. 현재 제공되는 점자책, 오디오북, 디지털 문서 등의 대체 자료는 제작 과정이 복잡하고 시간과 비용이 많이 들며, 모든 인쇄물을 즉각적으로 대응할 수 없다는 근본적인 한계를 가진다. 또한 스마트폰 기반의 OCR 애플리케이션은 네트워크 의존성, 처리 속도의 불안정성, 정확도의 변동, 사용 환경 제약 등의 문제로 인해 실생활에서 안정적으로 사용하기 어렵다는 단점이 지적되어 왔다. 이러한 기술적, 구조적 한계를 해결하기 위해서는 어떤 인쇄물이라도 사용자가 직접 촬영하고 곧바로 청각 정보로 변환하는 즉시적 휴대형 접근성 기술이 반드시 필요하다.

본 논문에서 제안하는 End-to-End OCR - TTS 기반 휴대용 보조 디바이스는 텍스트 인식과 음성합성을 하나의 파이프라인으로 통합하여, 점자나 별도의 음성 자료 없이도 실시간으로 인쇄물의 내용을 이해할 수 있도록 설계되었다. 특히 고성능 연산이 가능한 NVIDIA Jetson을 메인 처리 유닛으로 활용한 온디바이스 처리 구조를 통해 네트워크 환경과 관계없이 빠르고 안정적인 구동이 가능하며, 민감한 정보가 외부 서버로 전송되지 않아 개인정보 유출 위험을 최소화할 수 있다.

### II. 본론

본 연구에서 제안하는 스마트 안경 시스템은 사용자가 인쇄물을 촬영하는 단순한 입력 동작만으로 실시간에 가까운 음성 정보를 제공받을 수 있도록 최적화된 하드웨어 아키텍처를 갖추고 있다. 시스템의 하드웨어는

시각 장애 이용자가 직관적으로 조작할 수 있도록 안경 측면에 배치된 물리 버튼과 전면의 카메라 모듈, 그리고 데이터 처리를 담당하는 메인 유닛인 NVIDIA Jetson으로 구성된다. 사용자가 인쇄물을 향해 안경을 위치시키고 버튼을 클릭하면 카메라 모듈이 즉시 이미지를 캡처하며, 이 데이터는 고속 버스를 통해 메인 유닛으로 전달된다. Jetson은 강력한 GPU 연산 자원을 활용하여 이미지 분석, 글자 추출, 음성 정보 생성 과정을 온디바이스에서 수행하며, 최종적으로 생성된 음성 정보는 이어폰을 통해 사용자에게 직접 전달된다. 이러한 구성은 공공장소에서도 타인의 방해 없이 정보를 습득할 수 있게 하며 사용자의 편의성을 극대화한다.

시각 장애인용 스마트 안경 하드웨어 구성도

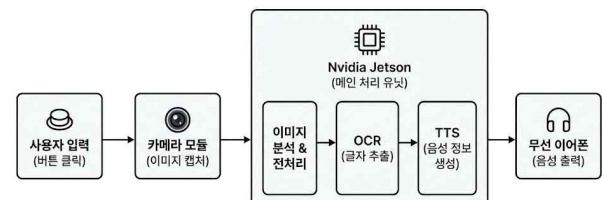


Figure 1 하드웨어 구성도

## Automated Processing of Documents

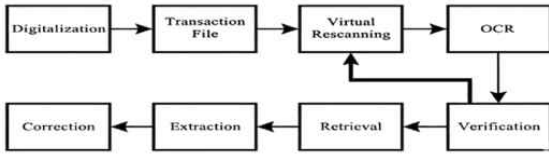


Figure 2 소프트웨어 알고리즘

### I. 영상 입력 및 지능형 전처리 파이프라인

카메라를 통해 획득된 원본 이미지는 OCR 인식률을 보장하기 위해 고도화된 영상 전처리 과정을 거친다. 이 단계에서는 OpenCV 기반의 알고리즘을 사용하여 그레이스케일 변환, 노이즈 억제 필터링, 대비 향상 및 이진화 과정을 수행함으로써 문자와 배경 간의 명암 차이를 극대화한다. 특히 촬영 시 발생할 수 있는 이미지의 기울기를 자동으로 감지하고 수평에 가깝도록 정렬하는 보정 알고리즘을 적용하여 다양한 각도와 조도 조건에서도 안정적인 인식 기반을 마련하였다. 이러한 전처리 기술은 이후 단계인 OCR 엔진의 인식 정확도를 확보하고 특히 한글 문서에서의 인식 효율을 높이는 데 결정적인 역할을 수행한다.

### II. 딥러닝 기반 OCR 인식 및 텍스트 정제

전처리된 영상 정보는 NVIDIA Jetson 내부의 Transformer 기반 딥러닝 OCR 모듈로 전달되어 디지털 글자 추출 과정을 거친다. 본 시스템은 이미지 내에서 텍스트 영역을 탐지하고 이를 행 단위로 분할한 뒤, 각 영역의 문자 시퀀스를 순차적으로 디코딩하여 유니코드 문자열을 생성한다. 특히 한글과 영어 혼합 문서에서도 안정적인 성능을 위해 언어 모델 분리 처리 기법을 적용하였으며, 인식 과정에서 각 문자에 대한 신뢰도를 산출하여 관리한다. 추출된 텍스트는 곧바로 음성으로 출력되지 않고 텍스트 정규화 과정을 거치는데, 여기서는 자주 발생하는 오인식 패턴을 사전 기반 규칙으로 자동 교정하고 맞춤법 및 띄어쓰기 오류를 수정하여 음성 합성 단계에서의 부자연스러움을 사전에 차단한다. [1]

### III. 온디바이스 TTS 합성 및 음성 최적화

정제된 텍스트 데이터는 TTS 모듈을 통해 청각 정보로 변환되어 이어폰으로 출력된다. 입력된 텍스트는 엔진 내부에서 숫자, 기호, 약어 등이 정규화되고 문장 단위로 분할되어 발음 기호 시퀀스로 변환된 후, 자연발화 기반 합성 기법을 통해 억양과 강세가 반영된 음성 파형으로 생성된다. 특히 운율 보정 알고리즘을 적용하여 문장부호에 따른 적절한 일시정지를 삽입함으로써 기계적인 느낌을 최소화하였으며, 긴 문서의 경우 인식이 끝난 부분부터 즉시 재생하는 스트리밍 방식을 채택하여 전체 지연 시간을 단축하였다. 모든 연산은 외부 서버 없이 NVIDIA Jetson 내부에서 완결되는 온디바이스 방식으로 구현되어 네트워크 환경과 무관한 안정성을 확보하였다. [2][3]

## III. 결론

본 연구에서는 시각 정보 접근이 어려운 이용자가 별도의 대체 자료 없이도 일반 인쇄물을 즉시 이해할 수 있도록 지원하는 NVIDIA Jetson 기

반의 스마트 안경 시스템을 설계하고 구현하였다. 제안된 시스템은 버튼 조작을 통한 이미지 캡처부터 OCR 기반의 글자 추출, 그리고 이어폰을 통한 음성 출력 과정을 하나의 일관된 파이프라인으로 통합함으로써 사용자의 최소 조작만으로 인쇄 정보를 실시간 청각 정보로 전환하는 성과를 거두었다. 특히 고성능 임베디드 프로세서를 활용한 이미지 전처리 단계와 운율 보정 설계를 통해 실제 환경의 변수에도 안정적인 동작을 구현하였으며, 모든 연산이 기기 내부에서 수행되는 온디바이스 방식을 채택하여 네트워크 독립성과 개인정보 보호라는 두 가지 핵심 가치를 동시에 달성하였다. 이러한 시스템은 사용자가 별도의 보조 자료 제작을 기다리지 않고 주변의 정보를 스스로 확인하게 함으로써 정보 접근의 자율성을 향상시킨다.

나아가 본 시스템은 단순히 기술적 구현을 넘어 시각 장애인의 사회적 자립과 삶의 질 향상에 실질적으로 기여한다는 점에 있다. 점자 자료나 오디오북 제작에 수반되는 막대한 비용과 인력 부담 같은 기존 체계의 구조적 한계를 완화할 수 있으며, 공공기관이나 교육기관에서 투입되는 자원을 절감하는 동시에 보다 다양한 콘텐츠를 빠르게 제공할 수 있는 기반을 마련한다. 또한 본 시스템은 실시간 처리와 유연한 파이프라인 구조를 바탕으로 교육 현장, 도서관, 공공시설 등 다양한 실생활 환경에 즉시 응용될 수 있는 높은 확장성을 지닌다. 향후 다국어 번역 기능이나 상황 인지 인터페이스를 결합한다면, 보다 지능화된 보조공학 플랫폼으로 발전하여 정보 소외 계층을 위한 포용적 정보 이용 환경을 조성하는 데 핵심적인 역할을 수행할 것으로 전망된다.

## ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음" (IITP-2025-RS-2024-00437190, 50%)

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2025-00553810, 50%)

## 참 고 문 헌

- [1] 서제원, 지영석, "설치형 오픈소스 OCR 엔진 성능 비교", 한국통신학회 2024 학술대회 논문집, 2024.
- [2] 성동현, 양수원, "Zero-shot TTS를 활용한 TTS 데이터 증강 및 성능 향상 기법", 대한전자공학회 학술대회 논문집, 2025.
- [3] 김규림 외, "Tesseract와 Tacotron2 모델을 이용한 TTS 시스템", 한국통신학회 학술발표회 논문집, 2021.