# RGB-D 기반 문자인식과 GPT 문장 복원을 활용한 시각장애인 안내 시스템

김내경, 현장훈*
국립한밭대학교, *국립한밭대학교
20221044@edu.hanbat.ac.kr, *jhhyeon@hanbat.ac.kr

# A Real-Time RGB-D Text Guidance System with GPT-Based Sentence Restoration for Visually Impaired Users

Kim Nae Kyung, Hyeon Jang Hun*
Hanbat National Univ., *Hanbat National Univ.

## 요 약

본 논문은 시각장애인이 보행 환경에서 문자 정보를 인식하기 어려운 문제를 개선하기 위해 문자 인식의 불안정성(가림, 거리 변화, 조도 변화)을 완화하기 위한 RGB-D 기반 안내 시스템을 제안한다. 제안한 시스템은 다중 전처리 기반 OCR을 통해 인식률을 향상 시키고, DBSCAN 기반 문장 클러스터링으로 단어 단위 결과를 문장 단위로 결합하며, 깊이(Depth) 기반 거리, 방향 추정을 통해 문장의 상대적 위치 정보를 산출한다. 또한 최근 프레임 히스토리를 사용하여 누락되거나 분절된 텍스트를 복구하며, GPT 기반 문장 복원을 통해 인식된 문장을 안내 문장으로 재구성하여 TTS로 제공한다. 추가적으로 YOLO 기반 동적, 정적 객체 인식 결과를 사용해 접근 방향에 따라 비프 경고음을 제공한다. 세 가지 실제 환경 실험에서 제안된 시스템은 문장 품질(Perplexity)을 77.61-96.13% 감소, 군집 품질(Silhouette)은 529.7-1316.6% 향상, Davies-Bouldin 지수는 76.3-91.6% 감소, 프레임 안정성(IoU)은 4.7-25.8% 향상시켜 실제 보행 환경에서의 안정성 향상을 확인하였다.

## Ⅰ. 서 론

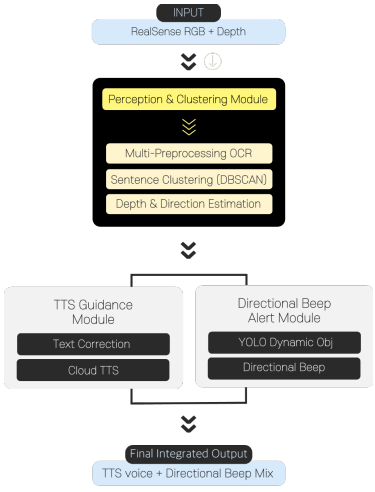
시각장애인은 안내문이나 표지판, 경고문 등과 같은 문자 정보를 시각적으로 인식하기 어렵다. 이로 인해 일상생활 속에서 중요한 정보를 놓치거나 주변 환경을 제대로 파악하지 못해 위험한 상황에 놓일 수 있다.

이전 연구에서는 사용자의 시야와 정지 상태를 기반으로 한 음성 안내 시스템을 제안했다. 하지만 배경이 복잡한 실외 환경이나 동적 객체가 많은 환경에서는 여전히 인식 혼선이 발생하는 문제가 있었다.

본 연구에서는 이러한 문제를 해결하기 위해 문자 인식(OCR)과 깊이(Depth) 카메라를 활용한 새로운 음성 안내 시스템을 제안한다. Depth 카메라를 통해 문자와 사용자 간의 실제 거리를 계산하고, 인식된 문자를 GPT 기반 교정 과정을 거쳐 자연스러운 음성 안내(TTS)로 제공한다. 또한 사용자를 향해 접근하는 동적 객체가 있을 경우 접근 방향에 따라 비프음(beep sound)을 제공하여 위험을 미리 인지하고 회피할 수 있도록 설계하였다. 제안된 시스템은 세 가지 실제 환경에서의 실험을 통해 기존 연구보다 다양한 환경에서 안정적인 인식과 안내가 가능함을 확인하였고, 시각장애인이 보다 안전하게 주변 정보를 파악할 수 있도록 돕는 것을 목표로 한다.

## Ⅱ. 시스템 구성

본 연구에서는 시각장애인을 위한 문자 인식 기반 안내 시스템을 구현하였다. 제안된 시스템은 [Fig. 1]과 같이 RealSense 카메라로부터 RGB-D 영상을 받아서 OCR, 클러스터링, TTS, 동적 객체 경고음을 한 흐름으로 처리하도록 구성되어 있다. 실제 환경에서는 사람이 문자 앞을 지나가거나, 조명과 해상도에 따라 인식이 불안정해지는 경우가 많아서 이를 보완하기 위해 본 시스템에서는 최근 10프레임을 이용한 문장 복원 기능을 적용하였다.




[Fig. 1] Overall architecture of the proposed OCR-TTS -Directional Beep guidance system using RGB-D input

### 2.1 Perception & Clustering Module

단일 영상만으로는 조도나 대비에 따라 인식률이 떨어질 수 있기 때문에, 본 연구에서는 원본 영상과 함께 감마 보정, 밝기 정규화, 대비 강화 등을 적용한 여러 버전의 이미지를 생성하여 OCR에 입력하였다[1]. 여러 OCR 결과 중 텍스트가 더 정확하게 인식된 항목을 자동으로 선택하여 병합하는 방식으로 인식률을 향상시켰으며, 문자 일부가 가려져 사라지는 문제는 최근 10프레임 중 가장 신뢰도가 높은 클러스터를 가져와 복원하도록 하였다.

### 2.2 Sentence Clustering (DBSCAN)

OCR 출력은 단어 단위 박스가 서로 따로 잡히는 경우가 빈번하여 이를 해결하기 위해 본 연구에서는 DBSCAN을 사용하여 단어들을 하나의 문장으로 묶었다[4]. eps 값은 여러 후보를 시험하여, 클러스터링 품질을 나타내는 Silhoutte Score 값을 자동으로 선택하도록 하였다. 이 과정은 다양한 거리와 각도 그리고 문자 크기에서도 문장이 안정적으로 형성되도록 돕는다.

### 2.3 Depth & Directional Estimation

문장이 인식되면 RealSense Depth 정보를 이용해 해당 문장이 사용자로부터 얼마나 떨어져 있는지 계산한다. 박스 내부의 깊이 값 중에 노이즈가 적은 중앙값을 사용해 거리 추정을 수행했다. 또한 화면을 좌·정면·우 세 영역으로 나누어 문장이 영상 어느 방향에 위치하는지 판단하여 이를 TTS 안내에 활용하였다[3].

### 2.4 TTS Gudiance Module

본 연구에서는 OCR로 인식된 문장을 그대로 읽는 대신, gpt-5-mini를 활용하여 시각장애인에게 적합한 안내형 문장으로 재구성하도록 설계하였다. GPT에는 문장의 의미를 크게 변경하지 않는 범위에서 인식된 텍스트가 어떤 종류의 안내문인지, 사용자와의 거리, 좌·정면·우 중 어느 방향에 있는지의 정보를 반영해 자연스러운 음성 안내 멘트를 생성하도록 프롬프트를 구성하였다.

### 2.5 Directional Beep Alert Module

YOLO 기반 추적을 사용해 검출된 동적 객체의 중심점 위치와 깊이 정보를 기반으로 어느 방향에 장애물이 많은지를 계산하고, 가까이에 접근하는 객체가 있을 경우 해당 방향으로 비프음을 울려 사용자에게 위협을 알려준다[2]. 사람과 같은 객체는 상대적으로 위험도가 높아서 거리 기반 가중치를 더 크게 적용하여 추천 경로 판단에 반영하였다.

[Fig. 2] [Fig. 2] Example outputs of the proposed system including text recognition, distance estimation, and directional guidance

## Ⅲ. 성능 평가 및 분석

본 논문에서의 실험 환경은 실제 보행 환경을 반영하여 다음과 같은 세 가지 시나리오로 구성하였다. (1) 근거리에서 두 안내문이 인접한 경우, (2) 단일 안내문을 사람이 자주 가리는 경우, (3) 원거리에서 인식 거리가 큰 환경.

평가 지표는 (i) 문장 품질(Perplexity, 낮을수록 우수), (ii) 클러스터링 품질(Silhouette, 높을수록 우수; Davies-Bouldin, 낮을수록 우수), (iii) 프레임 안정성(IoU, 높을수록 우수)으로 설정하였다[5],[6],[7]. [Table 1]은 세 가지 시나리에오에 대한 Baseline 대비 Proposed 개선율(%)을 요약한 것이다.

[Table 1] Quantitative performance comparison of text quality, clustering quality, and stability metrics

| Data | PPL(↓) | Silhouette(↑) | DB Index(↓) | IoU(↑) |
|---|---|---|---|---|
| (1) | -94.73% | +994% | -76.3% | +13.9% |
| (2) | -96.13% | +529.7% | -86.2% | +25.8% |
| (3) | -77.61% | +1316.6% | -91.6% | +4.7% |

전체적으로 다중 전처리 OCR, 깊이 기반 문장 필터링, 최근 프레임 복원을 결합한 방식이 서로 보완적으로 작용하면서, 근거리·가림·원거리 등 다양한 환경에서 안정적인 문자 인식과 문장 단위 클러스터링을 가능하게 하였다. 실험 결과는 제안된 시스템이 시각장애인 보행 환경에서 발생하는 실제 어려움(가려짐, 거리 변화, 조도)에 효과적으로 대응함을 보여준다.


## ACKNOWLEDGMENT

본 연구성과는 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업과(No. RS-2025-25432454) "2025년 과학기술정통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"과(2022-0-01068) 2025년도 교육부 및 세종특별자치시의 재원으로 세종RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다.(2025-RISE-08-004)


## 참 고 문 헌


[1] Y. Du, C. Li, R. Guo, C. Cui, W. Liu, J. Zhou, B. Lu, Y. Yang, Q. Liu, X. Hu, D. Yu, and Y. Ma, “PP-OCRv2: Bag of Tricks for Ultra Lightweight OCR System,” arXiv preprint arXiv:2109.03144, 2021.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. https://arxiv.org/abs/1506.02640

[3] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerry-Ryan, R. A. Saurous, Y. Agiomyrgiannakis, and Y. Wu, “Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions,” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018. https://arxiv.org/abs/1712.05884

[4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1996. https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

[5] J. Goodman, “A Bit of Progress in Language Modeling,” Computer Speech & Language, vol. 15, no. 4, pp. 403-434, 2001. https://doi.org/10.1006/csla.2001.0161

[6] P. J. Rousseeuw, “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis,” Journal of Computational and Applied Mathematics, vol. 20, pp. 53-65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7

[7] D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, pp. 224-227, 1979. https://doi.org/10.1109/TPAMI.1979.4766909