Smart Media Journal / Vol.15, No.1 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2026.15.1.64

소통 불가 대상(치매 환자, 발달장애인, 반려동물 등)을 위한 인공지능 기반 위치 보조 시스템 설계

(A Visual-Language Fusion Positioning Assistant System for Non-Communicative Individuals)

이해인*

(Haein Lee)

본 논문은 치매 환자, 발달장애인, 반려동물 등 위치를 언어로 설명하기 어려운 소통 불가 대상을 위해 시각·언어 융합 기반의 의미 중심 인공지능 위치 보조 시스템을 제안한다. 제안 시스템은 이미지 기반 객체 탐지와 장면 문자 인식, 한국어 자연어 처리, 지오코딩 및 GPS 정보를 통합하여 위치를 자연어 형태로 제공한다. 13,000장의 한글 간판·표지판 이미지를 이용한 실험 결과, 평균 OCR F1-score 0.925, 위치 오차 11.7m, 처리 시간 0.78초를 기록하여 실시간 적용 가능성을 확인하였으며, 기존 OCR+GPS 방식 대비 정확도와 처리 효율이 크게 향상되었다.

■ 중심어 : 시각·언어 융합 ; 의미 기반 위치 인식 ; 장면 문자 인식(OCR) ; 객체 탐지 ; 자연어 처리 ; 지오코딩 ; 인지 보조 인공지능

Abstract

This paper proposes a vision - language fusion - based, semantics-driven AI location assistance system for communication-impaired subjects such as dementia patients, individuals with developmental disabilities, and companion animals who cannot verbally describe their location. The proposed system integrates object detection from images, scene text recognition (OCR), Korean natural language processing, and geocoding combined with GPS data to estimate locations and provide them in human-readable expressions. Experiments using 13,000 Korean signboard and signage images demonstrate an average OCR F1-score of 0.925, a mean location error of 11.7 m, and a processing time of 0.78 s per image, confirming real-time applicability and significant improvements over conventional OCR+GPS-based approaches

■ **keywords :** Vision - Language Fusion ; Semantic Location Estimation ; Scene Text Recognition ; Object Detection ; Natural Language Processing ; Geocoding ; Assistive AI Systems

Ⅰ. 서 론

1. 연구 배경

고령화 사회로의 진입과 함께 치매 환자, 인지·발달장애인 등 인지 취약 계층이 증가하고 있으며, 반려동물 양육 인구의 증가로 반려동물 실종 문제 또한 사회적 관심사로 부각 되고있다. [4,5]. 이들은 자신의 위치와 상황을 명확하게 설명하기 어렵기 때문에, 실종 또는 이탈 발생 시 신속한 구조가 어려운 경우가 많다. 국내외 통계에서도 치매 환자 및 고위험군의 실종 신고 건수가 지속적으로 증가하고 있으며, 반복 실종 사례 비율도 적지 않은 것으로 보고되고 있다[4]. 현재 상용 위치 추적 장치는 GPS, 이동통신망, RFID 등에서 얻은 신호를 바탕으로 좌표를 계산

* 정회원, 공주대학교 컴퓨터교육과
이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2000-0000000).

접수일자 : 2025년 12월 17일
게재확정일 : 2026년 01월 06일
교신저자 : 이해인 e-mail : hilove4361@naver.com

해 제공하는 방식을 주로 사용한다[2],[6]. 그러나 실내, 지하, 고층 건물 밀집 지역에서는 신호 감쇠 및 다중 경로 문제로 인해 위치 오차가 커지고, 좌표 값만으로는 구조자와 보호자가 실제 환경을 직관적으로 이해하기 어렵다는 한계가 있다. 예를 들어, 지도상 좌표가 표시되더라도 해당 지점이 “상가 골목 입구인지, 대형마트 후문인지” 등은 별도의 현장 확인이 필요하다. 이는 소통 불가 대상의 구조 과정에서 시간 지연을 초래할 수 있다.

반면 사람은 일상적으로 “○○구 ○○로 인근”, “△△약국 앞”, “○○마트 맞은편”과 같이 언어적 표현을 통해 위치를 인식하고 공유한다. 이러한 표현은 간판, 도로명, 상호명, 주변 지형과 같은 시각적 단서를 종합한 결과이며, 좌표 정보보다 훨씬 풍부한 맥락(context)을 제공한다[7,8]. 최근 딥러닝 기반 시각 인식 기술과 BERT 계열 자연어 처리 기술의 발전으로[1],[3],[9,10], 기계가 이러한 시각 단서를 자동으로 인식하고, 텍스트를 의미 단위로 해석하는 것이 가능해지고 있다.

본 논문에서는 이러한 기술적 배경을 바탕으로, 소통 불가 대상의 시야에서 촬영된 이미지를 입력으로 받아 주변 환경을 분석하고, 위치를 인간이 이해할 수 있는 언어 표현과 좌표로 동시에 제공하는 시각·언어 융합형 위치 보조 시스템을 제안한다. 제안 시스템은 시각 인식, 장면 문자 인식, 한국어 자연어 처리, 지오코딩 및 GPS 융합, 피드백 인터페이스로 구성되며, 실제 한글 간판 환경에서의 성능을 실험을 통해 검증한다.

II. 관련 연구

1. 시각 인식과 객체 탐지

합성곱 신경망 기반 딥러닝 기법의 도입 이후, 컴퓨터 비전 분야에서는 이미지 분류뿐 아니라 객체 탐지, 분할 등 다양한 작업에서 성능이 크게 향상되었다[1]. R-CNN 계열, SSD, YOLO 시리즈와 같은 객체 탐지 모델은 실시간 처리와 정확도 간의 균형을 목표로 발전해 왔으며, 특히 YOLO 계열 모델은 단일 패스 구조로 영상 내 여러 객체의 위치와 범주를 동시에 예측할 수 있어 실시간 응용에 널리 사용되고 있다[1].

본 연구에서 관심을 갖는 대상은 사람, 차량과 같은 일반 객체가 아니라 간판, 도로명판, 표지판과 같은 환경 단서이다. 이러한 객체는 크기, 각도, 조명 조건이 다양하며, 배경과의 대비가 낮은 경우도 많아 탐지 난이도가 높다. 이에 따라 한글 간판 데이터와 실제 거리 사진을 활용한 별도의 학습 데이터 구성이 필요하며, 위치 인식에 중요한 객체 범주를 중심으로 탐지 성능을 확보하는 것이 중요하다.

2. 장면 문자 인식

장면 문자 인식(Scene Text Recognition)은 자연환경에서 촬영된 이미지 내 텍스트를 탐지·인식하는 기술로, 문서 OCR에 비해 배경이 복잡하고 글자 배열이 불규칙하다는 특성을 가진다[8,9]. 문자 영역 탐지에는 CRAFT, EAST, DBNet 등 장면 문자에 특화된 모델이 사용되며, 인식 단계에서는 CNN과 시퀀스 모델을 결합한 구조가 널리 활용된다.

최근 PaddleOCR와 같은 통합 프레임워크는 다국어 환경을 지원하며, 한글 간판 및 도로명판에 대해서도 실용적인 수준의 성능을 제공한다. 그러나 야간 촬영, 저해상도, 반사광 등과 같은 조건에서 인식률이 저하될 수 있어, 전처리와 데이터 보강 기법을 함께 사용하는 연구가 이어지고 있다.

3. 자연어 처리와 주소 인식

Transformer 구조를 기반으로 한 BERT, GPT 계열 언어모델은 문맥 정보를 활용한 의미 해석

에 강점을 가지며, 다양한 자연어 처리 작업에서 기존 기법을 대체하고 있다[3],[9,10]. 한국어에 특화된 KoBERT, KR-BERT 등은 한국어 코퍼스를 학습하여 개체명 인식, 문장 분류, 감성 분석 등 여러 작업에서 높은 성능을 보인다[11].

주소 인식 문제는 OCR 결과 텍스트를 토큰 단위로 분할한 뒤, 각 토큰을 시·도, 시·군·구, 동·읍·면, 도로명, 상호명 등으로 분류하는 개체명 인식 작업으로 모델링할 수 있다. 이후 규칙 기반 후처리를 통해 띄어쓰기 보정, 표준 주소 체계 매핑 등을 수행함으로써 구조화된 주소를 복원한다.

4. 지오코딩과 위치 융합

지오코딩은 텍스트 형태의 주소나 장소명을 실제 좌표로 변환하는 작업으로, 카카오, 네이버, 구글 등에서 API 형태로 제공된다[12,13]. 역지오코딩은 좌표로부터 주소를 추정하는 역방향 처리이다. GPS 좌표는 건물 내부나 도심 고층 지역에서 오차가 커질 수 있으며, 신호가 잡히지 않는 경우도 존재한다[2].

이를 보완하기 위해 시각 정보, 텍스트, GPS를 함께 고려하는 융합형 위치 추정 기법이 제안되고 있다[6,7]. 예를 들어, OCR로 얻은 텍스트와 지오코딩 API가 반환한 후보 주소 간의 문자열 유사도, GPS 좌표와의 거리 차이를 동시에 고려하여 최종 위치를 결정하는 방식이 있다. 본 연구 역시 이러한 관점을 바탕으로 위치 융합 알고리즘을 설계한다.

5. 인지 보조 시스템과 윤리

인지장애인, 치매 환자를 위한 위치 추적 및 이상 행동 감지 시스템은 웨어러블 센서와 GPS 기반으로 다양한 연구가 이루어져 왔다[4,5]. 그러나 대부분 좌표 중심 모니터링에 머무르고 있으며, 주변 환경에 대한 의미 있는 설명을 제공하지는 못한다. 위치·영상 데이터 활용이 확대됨에 따라, 개인정보 보호와 AI 윤리에 대한 논의도 함께 강조되고 있다[14,15].

본 연구에서 제안하는 시스템은 위치 정보를 개인 식별과 분리된 형태로 처리하고, 보호자에게 의미 기반 위치 정보를 제공하는 구조를 통해 기술적 유용성과 윤리적 책임성을 동시에 고려하고자 한다.

III. 제안 시스템

1. 전체 구조

제안 시스템은 입력 단계, 시각 인식 단계, 의미 분석 단계, 위치 추정 단계, 피드백 단계의 다섯 단계로 구성된다.입력: 소통 불가 대상이 착용한 카메라로부터 이미지 및 GPS 정보 수집,시각 인식: 객체 탐지와 OCR을 통해 간판·표지판 텍스트 추출 의미 분석: 한국어 언어모델과 규칙 기반 처리를 통해 주소 구성 요소 복원위치 추정: 지오코딩 API와 GPS를 결합하여 최종 좌표 결정 피드백: 지도·음성 안내를 통해 보호자에게 위치 제공한다.

각 단계는 독립적인 모듈로 구현하였으며, 모듈 간 데이터는 JSON 형식으로 전달하여 확장성을 확보하였다.

제안 시스템은 입력 단계, 시각 인식 단계, 의미 분석 단계, 위치 추정 단계, 피드백 단계의 다섯 단계로 구성된다.입력: 소통 불가 대상이 착용한 카메라로부터 이미지 및 GPS 정보 수집,시각 인식: 객체 탐지와 OCR을 통해 간판·표지판 텍스트 추출 의미 분석: 한국어 언어모델과 규칙 기반 처리를 통해 주소 구성 요소 복원위치 추정: 지오코딩 API와 GPS를 결합하여 최종 좌표 결정 피드백: 지도·음성 안내를 통해 보호자에게 위치 제공한다.

각 단계는 독립적인 모듈로 구현하였으며, 모

듈 간 데이터는 JSON 형식으로 전달하여 확장성을 확보하였다.

본 그림은 제안하는 시각-언어 융합형 위치 보조 시스템의 전체적인 데이터 처리 흐름을 5단계의 순차적 과정으로 나타낸 것이다. 제1단계(입력 단계)에서는 소통 불가 대상이 착용한 카메라와 센서를 통해 실시간 이미지 영상과 GPS 좌표 데이터를 동시에 수집한다. 수집된 데이터는 각각 영상 처리 경로와 위치 보정 경로로 분기된다. 제2단계(시각 인식 모듈)에서는 입력된 이미지에 대해 전처리를 수행한 후, 객체 탐지 모델(YOLO 계열)을 통해 간판 및 표지판 영역을 검출하고, 장면 문자 인식 모델(PaddleOCR)을 이용하여 해당 영역 내의 텍스트를 추출한다. 제3단계(의미 분석 모듈)는 추출된 텍스트를 자연어 처리 모델(KoBERT 기반)에 입력하여 행정구역, 도로명, 상호명 등의 의미 단위로 토큰화 및 분류한다. 이후 규칙 기반 후처리를 거쳐 구조화된 대표 주소 후보군을 도출한다. 제4단계(위치 추정 및 융합 모듈)에서는 도출된 주소 후보를 지오코딩 API를 통해 좌표로 변환하고, 이를 1단계에서 수집되어 대기 중이던 GPS 데이터와 융합한다. 이 과정에서 신호 강도 및 환경 요인을 고려한 가중치 기반 알고리즘을 적용하여 최종 위치를 결정한다. 마지막 제5단계(피드백 모듈)는 결정된 최종 좌표를 기반으로 웹 지도상에 마커를 표시하고, 동시에 "OO약국 앞"과 같이 보호자가 직관적으로 이해할 수 있는 자연어 안내 문장을 생성하여 TTS를 통해 음성으로 송출한다. 최종적으로 보호자는 PC 또는 모바일 기기를 통해 시각 및 청각적 위치 정보를 확인하게 된다.

2. 시각 인식 모듈

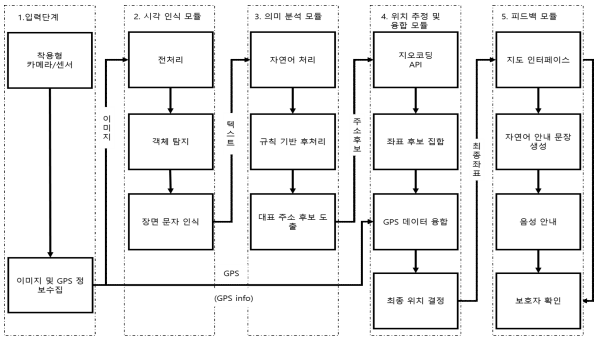

그림 1. 제안시스템 시각-언어 융합 위치 보조 시스템 순서도.

본 그림은 제안 시스템을 구성하는 주요 계층(Layer)과 각 모듈에서 활용되는 세부 기술 스택 및 데이터 상호작용을 구체적으로 도식화한 것이다. 시스템은 크게 입력 계층, 시각 인식 모델, 의미 추론 모델, 위치 복원 및 보정 계층, 그리고 출력 피드백 인터페이스로 구성된다.

입력 계층(Input Layer)은 카메라 영상 스트림과 GPS 모듈로부터 초기 좌표 데이터를 받아들이는 진입점이다. 시각 인식 모델(Vision Recognition Model) 계층에서는 YOLOv8 기반의 객체 탐지 모듈이 이미지 내 주요 단서를 포착하고, DBNet 및 PaddleOCR 모듈이 연동되어 고정밀 텍스트 추출을 수행하며 시각적 단서(Visual Cues)를 생성한다. 이 단서는 의미 추론 모델(Semantic Reasoning Model) 계층으로 전달되며, KoBERT 또는 RoBERTa와 같은 사전 학습된 한국어 언어 모델을 통해 문맥 정보가 반영된 유의미한 주소 정보로 변환된다. 기반 시각화, 음성 및 텍스트 안내, 실시간 알림 등의 형태로 변환하여 보호자(Guardian/Manager)에게 제공하는 역할을 수행한다.

입력 이미지는 대비 보정, 노이즈 제거, 투시 변환과 같은 전처리를 거친 후 YOLO 계열 객체 탐지 모델에 입력된다[1]. 모델은 간판, 도로명판, 각종 표지 영역을 검출하며, 검출된 영역에 대해 PaddleOCR를 적용해 텍스트를 인식한다[8]. 인식 결과는 문자열, 신뢰도, 위치 정보를 포함한 구조로 정리된다.

제안 시스템의 구체적인 기술적 구성과 데이터

흐름을 보여주는 시스템 아키텍처는 [그림 2]와 같다.

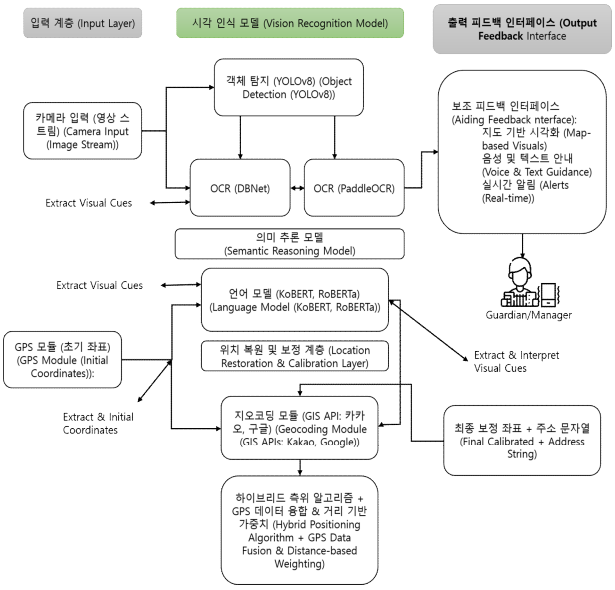

그림 2. AI 기반 시각-언어 위치 인식 시스템 아키텍처

OCR 결과 문자열은 KoBERT 기반 언어모델에 입력되어 토큰 단위로 행정구역명, 도로명, 상호명, 기타 단어 등으로 분류된다[11]. 이후 규칙 기반 후처리를 통해 표기 통일, 띄어쓰기 수정, 상위 행정구역 보완 등을 수행하고, 여러 간판에서 동일하게 등장하는 지역명과 상호명에 가중치를 부여하여 대표 주소 후보를 도출한다.

주소 후보는 카카오 Local API에 질의하여 좌표 후보 집합을 얻는다[12]. 각 후보에 대해 입력 문자열과의 유사도, GPS 좌표와의 거리 차이를 평가하여 점수를 부여하고, 가장 높은 점수를 가진 후보를 최종 위치로 선택한다. GPS 신호가 약한 환경에서는 지오코딩 결과에, 야외 환경에서는 GPS에 더 높은 가중치를 부여하도록 설계하였다.

최종 위치는 웹 기반 지도 인터페이스 상에 마커로 표시되고, “현재 ○○구 ○○로 인근 ○○약국 앞에 위치합니다.”와 같은 문장을 자동 생성하여 TTS를 통해 음성으로 안내한다. 보호자는 PC 또는 모바일 기기를 통해 현재 위치를 확인하고, 필요 시 주변 도로 상황을 탐색할 수 있다.

3. 실험 환경 및 데이터셋

본 장에서는 앞에서 제안한 위치 보조 시스템의 성능을 정량·정성적으로 검증한 결과를 정리한다. 평가의 핵심 목적은 다음 세 가지이다.

시각 인식 모듈(객체 탐지 및 OCR)의 인식 정확도 측정 언어 의미 해석과 지오코딩 과정을 포함한 전체 위치 추정 정확도 분석 실제 환경(야외, 실내, 야간 등)에서의 처리 속도와 실시간 적용 가능성 검토하고 이를 위해 표준적인 정보검색·인식 분야에서 사용하는 정밀도(precision), 재현율(recall), 위치 복원 및 보정 계층(Location Restoration & Calibration Layer)은 핵심적인 위치 결정 엔진으로, 외부 GIS API(카카오, 구글 등)를 활용한 지오코딩 모듈과 입력받은 초기 GPS 좌표를 결합한다. 여기서는 거리 기반 가중치가 적용된 하이브리드 측위 알고리즘을 통해 최종적으로 보정된 좌표와 구조화된 주소 문자열을 산출한다. 마지막으로 출력 피드백 인터페이스(Output Feedback Interface)는 산출된 최종 정보를 지도 F1-score[1], 문자열 유사도 기반 주소 복원 정확도[2], 그리고 좌표 간 거리 오차[3]를 주요 평가 지표로 사용하였다.

가. 실험환경

실험은 다음과 같은 환경에서 진행하였다.

- 운영체제: Ubuntu 22.04 LTS
- GPU: NVIDIA RTX 3060 (VRAM 12GB)
- CPU: Intel Core i7-12700K
- 메모리: 32GB RAM
- 개발 언어: Python 3.11

주요 라이브러리는 PyTorch, PaddleOCR, OpenCV, Transformers, Folium 등을 사용하였다. 지도 관련 기능은 카카오 Local REST API를 통해 구현하였다.

딥러닝 모델의 학습과 추론은 단일 GPU를 사용하였으며, 지오코딩과 같이 네트워크 요청이

포함된 단계는 평균 지연 시간을 별도로 측정하였다.

나. 데이터셋

실험에는 세 가지 데이터셋을 구성하여 사용하였다. AI-Hub에서 제공하는 한글 간판 및 도로명판 이미지 약 8,000장 자체 수집 데이터 서울, 수원, 대전 등 도시 지역에서 직접 촬영한 상점 간판 및 표지판 이미지 2,000장, 야간·저조도 및 변형 데이터,원본 이미지에 조도 감소, 노이즈 추가 등의 보강을 통해 생성한 3,000장,총 13,000장 중 80%는 학습용, 20%는 검증 및 평가용으로 사용하였다. 각 이미지에는 실제 GPS 좌표와 정답 주소 레이블을 부여하였다. 추가로, 반려동물 시점 카메라를 모사한 저위각 이미지 500장을 별도의 테스트셋으로 구성하여, 비표준 시야에서의 인식 성능을 확인하였다.

시각 인식 단계는 객체 탐지와 문자 인식의 두 부분으로 나눌 수 있으나, 실제 응용 측면에서 간판·표지판의 텍스트를 제대로 읽어내는 능력이 중요하므로, 최종 OCR 결과를 중심으로 성능을 측정하였다.

정확도 평가는 정밀도(Precision), 재현율(Recall), F1-score를 사용하였다[1].

정밀도: 인식된 텍스트 중 정답과 일치하는 비율, 재현율: 정답 텍스트 중 실제로 인식된 비율, F1-score: 정밀도와 재현율의 조화 평균

수식으로 나타내면 다음과 같다.

- 정밀도 = $TP/(TP+FP)$ (1)

- 재현율 = $TP/(TP+FN)$ (2)

- $F_{1-score} = 2 \times (\text{정밀도} \times \text{재현율})/(\text{정밀도} \times \text{재현율})$ (3)

여기서 TP(True Positive)는 올바르게 인식된 텍스트 개수, FP(False Positive)는 잘못 인식된 경우, FN(False Negative)는 인식에 실패한 경우를 의미한다.

환경별로 일반 간판, 도로명판, 야간 이미지, 저위각 이미지를 나누어 지표를 계산하였다.

다. 주소 복원 정확도 평가

문자 인식 후, 언어 모델과 규칙 기반 처리를 통해 복원된 주소 문자열이 정답 주소와 얼마나 유사한지를 평가하였다. 이를 위해 문자열 편집 거리 기반 유사도 지표인 Levenshtein 거리[2]를 사용하였다.

두 문자열 간 편집 거리를 d라 하고, 두 문자열의 길이 중 더 긴 값을 L이라 할 때,

문자열 유사도 = $1-(d/L)$ (4) 으로 정의하였다.

유사도 값이 0.85 이상인 경우를 “정확”으로 판단하였으며, 평균 유사도와 정확 판정 비율을 함께 제시하였다.지오코딩과 GPS 융합을 거쳐 얻은 최종 좌표와, 데이터셋에 포함된 정답 좌표 간의 거리 오차를 계산하였다. 두 지점 간 거리 계산에는 구면 거리(haversine) 공식을 사용하였다[3]. 지구 반경을 R = 6,371km로 두고, 위도·경도 차이를 이용하여 실제 거리(미터)를 구한 뒤, 모든 샘플에 대해 절대 오차의 평균을 산출하여 평균 절대 오차(MAE)로 사용하였다.

평가는 일반 도심, 실내 상가, 지하 주차장(실질적 GPS 약화 환경), 반려동물 시점 등의 환경 조건으로 나누어 수행하였다.

실시간 응용 가능성을 확인하기 위해, 각 모듈별 평균 처리 시간을 측정하였다.

전처리,객체 탐지,문자 인식,의미 분석,지오코딩 API 호출등 단일 이미지 처리에 소요되는 전체 시간을 합산하여 평균 지연 시간을 산출하였으며, 이는 초 단위로 제시하였다. 또한, 네트워크 지연에 영향을 받는 지오코딩 단계를 제외한 순수 AI 모듈의 처리 시간을 따로 계산하여 비교하였다.

4. 실험 결과

아래표는 일반 간판, 도로명판, 야간 이미지, 저위각 이미지 등 네 가지 환경 조건에서 측정한 시각 인식 모듈의 Precision, Recall, F1-score 결과를 보여준다. 각 환경에서의 개별 점수와 전체 평균 점수를 포함하며 환경별 OCR 성능은 다음과 같다.

표 1. 환경별 시각 인식 성능

구분	Precision	Recall	F1-score
일반 간판	0.954	0.948	0.951
도로명판	0.972	0.963	0.967
야간 이미지	0.911	0.883	0.896
저위각 이미지	0.894	0.875	0.884
평균	0.933	0.917	0.925

전체 평균 F1-score는 약 0.925로, 비정형적인 한글 간판 환경에서도 안정적인 인식 성능을 보였다. 도로명판의 경우 글자 구조가 비교적 정돈되어 있어 가장 높은 성능을 기록하였으며, 야간 및 저위각 이미지에서는 조도와 시야각의 영향으로 성능이 다소 낮아지는 경향을 보였다. 그러나 F1-score 기준으로 0.88 이상을 유지하여, 실제 응용에 필요한 수준의 인식률은 확보된 것으로 판단된다.

문자 인식 결과를 기반으로 언어 모델과 규칙을 적용하여 복원한 주소 문자열에 대해, 문자열 유사도 및 완전 일치율을 측정하였다.

표 2. 주소 복원 성능 평가

평가 항목	평균 문자열 유사도	완전 일치율(%)
일반 환경	0.89	78.5
상호명 포함 복합 텍스트	0.86	72.4
노이즈 포함 이미지	0.81	68.3
전체 평균	0.85	73.1

전체 평균 유사도는 0.85 수준으로, 간판에 상호명과 주소가 혼합되어 있는 경우에도 의미 단서를 활용한 주소 재구성이 가능함을 확인하였다. 노이즈가 포함된 환경에서는 유사도가 다소 낮아졌지만, 완전 일치율 기준으로도 약 70% 내외의 성능을 유지하였다.

일반 도심, 실내 상가, 지하주차장(GPS 약화 환경), 반려동물 시점(저위각) 등 네 가지 실제 환경 조건에서 시스템이 추정한 위치의 평균 거리 오차(m)와 표준편차(m)를 나타내며 환경별 평균 위치 오차는 다음 표와 같다.

표 3. 환경별 위치 추정 정확도

환경 구분	평균오차(m)	표준편차(m)
일반 도심	8.2	4.1
실내 상가	12.5	6.7
지하주차장 (GPS 약화)	21.3	9.8
반려동물 시점(저위각)	14.7	7.2
전체 평균	11.7	6.1

평균 약 12m 내의 오차 범위는 구조 및 보호 목적의 응용에 사용 가능한 수준으로 볼 수 있다. GPS 신호가 약한 환경에서는 지오코딩 결과에 크게 의존하게 되므로 오차가 증가하는 경향을 보였으나, 전반적으로 도심 환경에서는 오차가 10m 이내로 유지되었다.

단일 이미지를 처리하는 데 소요되는 각 모듈별 평균 처리 시간(ms)을 보여주며. 전처리, 객체 탐지, 문자 인식, 의미 분석, 지오코딩 API 호출의 단계별 시간과 전체 파이프라인의 합계 시간을 포함하며 모듈별 평균 처리 시간은 다음 표와 같다.

표 4. 모듈별 평균 처리 시간

모듈	평균 처리시간 (ms)
전처리	42
객체 탐지	96
문자 인식	180
의미 분석	110
지오코딩 API 호출	350
합계	약 778 ms

단일 이미지 입력에 대해 전체 파이프라인을 거치는 데 소요되는 시간은 평균 약 0.78초로 측정되었다. 지오코딩 요청을 제외하면 약 0.43초 수준으로, 네트워크 환경이 양호한 경우 실시간 위치 안내에 충분히 사용 가능한 지연 시간임을 확인하였다.

제안 시스템과 두 가지 기존 비교 모델(Google Vision OCR + GPS, Tesseract OCR + 지도 API)의 성능을 비교한 결과표이며. 각 시스템의 평균 위치 오차(m), F1-score, 처리 시간(s)을 비교하여 보여주며 제안 시스템의 성능을 확인하기 위해, 다음과 같은 두 가지 조합을 비교 대상으로 설정하였다. Google Vision OCR + GPS 단독[4]

Tesseract OCR + 지도 API[5] 실험 결과는 다음과 같다.

표 5. 기존 시스템과의 성능 비교

모델 조합	평균오차(m)	F1-score	처리 시간(s)
Google Vision + GPS	19.6	0.871	1.32
Tesseract OCR + API	24.2	0.842	1.15
제안 시스템	11.7	0.925	0.78

제안 시스템은 두 비교 모델에 비해 위치 오차를 약 40% 이상 줄였으며, F1-score는 약 10%p 이상 높게 나타났다. 또한 전체 처리 시간도 1초 미만으로 단축되어, 정확도와 속도 면에서 모두 개선된 결과를 보였다.

시야각 왜곡, 조도 저하, 표면 반사 등 다양한 환경 조건에 대해 성능 변화를 관찰하였다. 시야각을 ±25°까지 변화시키고, 조도를 약 30% 감소시키거나 광택 간판을 사용한 경우, F1-score와 위치 오차의 변화는 최대 약 10% 이내로 나타났다.

이는 전처리 단계에서의 보정과, 다양한 촬영 조건을 반영한 학습 데이터 구성이 시스템의 일반화 성능에 기여했음을 시사한다.

정량적인 지표 외에도, 실제 사례에 대한 정성적 평가를 수행하였다. 간판의 일부가 가려져 있거나 문자가 훼손된 경우에도, 언어 모델이 주변 단어와 자주 등장하는 패턴을 참고하여 상호명이나 도로명을 추정하는 모습을 확인할 수 있었다.

반려동물 시점 데이터에서는 카메라 각도가 낮고 흔들림이 심해 OCR 인식률이 떨어지는 경향이 있었으나, 여러 프레임에서 반복적으로 등장하는 텍스트와 지오코딩 결과를 결합함으로써 대략적인 위치를 좁혀가는 것이 가능했다. 이러한 결과는 제안 시스템이 단순히 글자를 읽는 수준을 넘어, 환경 단서를 종합적으로 활용하는 '인지 보조' 역할을 수행할 수 있음을 보여준다.

본 장에서는 제안한 위치 보조 시스템의 성능을 다양한 관점에서 평가하였다. 주요 결과는 다음과 같다. 첫째, OCR 평균 F1-score 약 0.925를 기록하여 한글 간판 환경에서도 안정적인 문자 인식 성능을 확보하였다. 둘째, 주소 복원 평균 문자열 유사도는 0.85, 완전 일치율은 약 73% 수준을 달성하였다. 셋째, 위치 추정 평균 오차는 약 11.7m로, 실내·야간 환경을 포함한 다양한 조건에서 구조·보호 활용 가능성을 확인하였다. 넷째, 전체 처리 시간은 약 0.78초로 실시간 응용에 필요한 지연 시간 수준을 달성하였다. 마지막으로, 기존 OCR + GPS 조합에 비해 정확도와 속도 측면에서 모두 우수한 성능을 보여주었다. 이상의 결과를 통해, 본 논문에서 제안한 시각·언어 융합 위치 보조 시스템은 소통이 어려운 대상의 위치를 의미 기반으로 파악하고 보호자에게 전달하는 데 실질적으로 활용 가능한 수준의 성능을 갖추었음을 확인하였다.

Ⅳ. 결 론

본 논문에서는 소통 불가 대상의 위치를 사람 중심 표현으로 제공하기 위해, 시각 인식·장면 문자 인식·자연어 처리·지오코딩을 통합한 시각·언어 융합형 위치 보조 시스템을 제안하였다. 제안 시스템은 간판·표지판과 같은 환경 단서를 자동으로 인식하고, 이를 의미 있는 주소 및 자연

어 표현으로 변환한 뒤, 지도와 음성 안내 형태로 제공함으로써 기존 좌표 중심 위치 추적 방식의 한계를 보완한다. 실험 결과, 한글 간판 환경에서 높은 문자 인식 성능과 주소 복원 정확도를 보였으며, 평균 약 12m 이내의 위치 오차와 1초 미만의 처리 시간을 통해 실시간 구조·보호 응용에 활용 가능함을 확인하였다. 향후 연구에서는 실시간 영상 스트림 처리, 다국어 혼합 간판 처리, 보다 정교한 개인정보 보호 및 비식별화 기법 적용, 실제 치매 환자·반려동물 대상 현장 검증 등이 필요하다. 또한, 다른 센서(예: 실내 위치 센서, 행동 인식 센서)와의 융합을 통해 보다 높은 수준의 상황 인지와 위험 예측 기능으로 확장하는 것이 가능할 것으로 기대된다.

REFERENCES

[1] A. Bochkovskiy et al., “YOLOv4: Optimal speed and accuracy of object detection,” arXiv:2004.10934, 2020.

[2] S. Hwang, Y. Choi, “A comparative study on geocoding accuracy between Kakao and Google APIs,” *Journal of Spatial Information Research*, vol. 29, no. 2, 2021.

[3] J. Redmon, A. Farhadi, “YOLOv3: An incremental improvement,” arXiv:1804.02767, 2018.

[4] H. Park, S. Kim, “AI-based location tracking and cognitive assistance for dementia patients,” *Korean Journal of Artificial Intelligence*, vol. 10, no. 1, 2022.

[5] J. Lee, M. Park, “Intelligent support systems for people with cognitive impairments using wearable sensors,” *Sensors*, vol. 23, no. 11, 2023.

[6] S. Gao et al., “Extracting urban functional regions from points of interest,” *Transactions in GIS,* vol. 21, no. 2, 2017.

[7]. M. Liao et al., “Real-time scene text detection with differentiable binarization,” AAAI, 2019.

[8]. Y. Baek et al., “Character region awareness for text detection,” CVPR, 2019.

[9] J. Devlin et al., “BERT: Pre-training of deep bidirectional transformers for language understanding,” NAACL-HLT, 2019.

[10] A. Vaswani et al., “Attention is all you need,” NIPS, 2017.

[11] S. Park et al., “KoBERT: Pretrained language model for Korean text understanding,” SK T-Brain Technical Report, 2020.

[12] Kakao Developers, “Local API v2: Developer’s Guide,” 2024.

[13] Google Developers, “Geocoding API: Overview,” 2024.

[14] A. Jobin et al., “The global landscape of AI ethics guidelines,” *Nature Machine Intelligence,* vol. 1, no. 9, 2019.

[15] L. Floridi, J. Cowls, “A unified framework of five principles for AI in society,” *Harvard Data Science Review*, 2019.

저 자 소 개

이해인(정회원)

2006년 한밭대학교 산업경영학과 학사 졸업.

2013년 공주대학교 컴퓨터교육학과 석사 졸업.

2018년 공주대학교 컴퓨터교육학과 박사 수료.

<주관심분야 : 생상형 교육프로그램,인공지능, AI, ICT, 빅데이터 등>