AI 에이전트 기반 문서 왜곡 및 조명 변화에 강건한 OCR 기법

여수향, 한요섭
숭실대학교

yyy1247@gmail.com, yoseob.han@ssu.ac.kr

A Study on AI Agent-Based Robust OCR Technique for Document Distortion and Illumination Variations

Suhyang Yeo, Yoseob Han
Soongsil University

요 약

비대면 모바일 서비스가 활발해지면서 스마트폰을 이용한 문서 촬영 및 제출이 증가하고 있다. 그러나 사용자 촬영 환경의 조명 불균형과 문서 왜곡 등이 광학 문자 인식 (Optical Character Recognition, OCR) 성능을 저하시킨다. 이러한 문제를 인간의 행동을 모방한 AI 에이전트 기반 OCR 기법을 통해 개선하고자 한다. 경량화된 멀티모달 (Multimodal Large Language Models)을 추론 엔진으로 사용하여 이미지 상태를 진단하고 적절한 전처리 도구를 동적으로 선택하여 수행한다. 제안 기법은 최신 단일 OCR 모델 대비 인식 정확도를 향상시켰으며 문자 단위 오인식률 (Character Error Rate, CER)과 단어 단위 오인식률(Word Error Rate, WER)을 개선하여 저품질 문서에서의 강건성을 입증하였다.

Ⅰ. 서 론

자동차 산업은 제조 중심 산업에서 벗어나 모빌리티 서비스 플랫폼으로 진화함에 따라 커넥티드 카, 차량 공유, 보험 연계 등 파생 서비스가 확대되고 있다. 차량 등록증과 같은 필수 문서 제출이 모바일 플랫폼으로 전환되며 스마트폰으로 직접 촬영한 문서 이미지가 입력 데이터의 주를 이루게 되었다.
그러나 통제되지 않은 촬영 환경은 조명 반사, 그림자, 문서 구겨짐 등 다양한 노이즈를 유발하며 이는 OCR 엔진의 인식률을 저하시켜 고객센터의 수기 입력 검증이라는 운영 비효율과 비용 증가로 이어진다.

기존에는 딥러닝 기반의 GAN(Generative Adversarial Networks)이나 Transformer 기반 모델을 활용한 다양한 연구가 수행되었으나 데이터의 상태와 무관하게 학습 시 정해진 동일한 연산을 적용하여 데이터 특성에 따른 적응적 처리와 연산 효율성 측면에서 한계를 보인다. 최근 주목받는 LLM 은 문맥에 의존하여 존재하지 않는 텍스트를 생성하는 환각(Hallucination) 현상이 있어 정확성이 필요한 문서 처리 업무에 적용하기 어렵다.[1]

이에 본 연구는 사람이 잘 보이지 않는 문서를 읽을 때 취하는 인지적 행동(확대, 밝기 조절 등)을 모사한 AI 에이전트(Human-mimetic AI Agent) 기반 OCR 기법을 제안한다. 제안하는 에이전트는 이미지 상태를 분석하고 필요한 전처리 도구만을 선택적으로 수행하는 명시적인 이미지 개선 과정을 거친다. 이를 통해 환각 현상을 원천적으로 배제하고 인식 성능의 강건성과 결과의 설명 가능성을 확보하고자 한다.

Ⅱ. 본론

A. 제안 방법

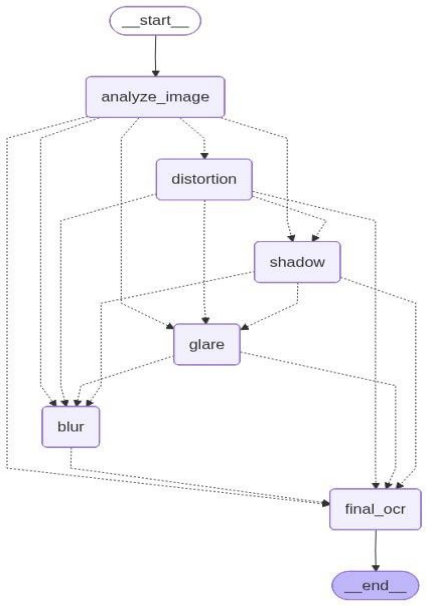

그림 1. 아키텍처

그림 1 은 AI 에이전트 기반 OCR 프레임워크의 전체 처리 흐름을 나타낸다. 입력 이미지의 물리적 상태를 진단하고 상황에 맞는 전처리 파이프라인을 동적으로 구성한다. 전체 구조는 (1) 분석 에이전트, (2) 도구

에이전트, (3) OCR 엔진의 3 단계로 구성되며 LangGraph 를 활용하여 각 단계의 상태를 관리한다.

시스템의 의사결정을 담당하는 두뇌 역할로 Google 의 Gemma3[2]모델을 채택하였다. Gemma3 는 한국어 이해도와 시각적 처리 능력을 갖추면서 온디바이스(On-device) 구동이 가능한 경량 모델이다. Ollama 프레임워크 기반의 로컬 환경에 구축하여 고객의 민감한 개인정보가 외부 서버로 유출되는 것을 방지하였다.
분석 에이전트는 프롬프트 엔지니어링을 통해 인간의 인지 모델을 모방한다. 단순히 "이미지를 개선하라"는 명령 대신 "이미지에 그림자가 심하므로 조명 보정과 텍스트 선명도를 높인다"와 같은 논리적 추론을 수행한다. 그 결과로 ['shadow', 'blur']와 같은 실행 계획을 생성하며 이미지 상태가 양호할 경우 Skip 을 결정하여 불필요한 보정을 하지 않도록 정보 손실을 방지한다.

분석 에이전트가 호출하는 도구 에이전트는 문서 이미지에서 빈번하게 발생하는 문제 유형들을 중점적으로 해결하기 위해 다음과 같은 세부 모듈로 구성된다:

1. Distortion Tool: UVDoc[3]을 활용하여 구겨짐, 곡률, 원근 왜곡 등 기하학적 변형을 보정한다.

2. Shadow, Glare Tool: OpenCV 기반의 영상 처리 기법을 적용하여 그림자 영역을 밝히고 과도한 반사광(Glare)을 감소시켜 문자 대비를 높인다.

3. Blur Tool: 초점이 맞지 않거나 흔들린 이미지에 엣지 강조(Edge Enhancement) 필터를 적용하여 문자 획의 경계를 선명하게 복원한다.

B. 실험

제안 시스템은 개인정보가 포함된 실제 자동차 등록증 대신 실제와 유사한 환경의 데이터를 생성하여 평가를 진행하였다. 비교 대상으로는 최신 OCR 엔진인 PaddleOCR(PP-OCRv5)[4] 단독 모델을 사용하였으며 평가 지표로는 문자 단위 오인식률(CER), 단어 단위 오인식률(WER), 전체 정확도(Accuracy)를 사용하였다.

표 1. 합성 데이터 기반 성능 비교

Method	*CER ↓*	*WER ↓*	*Accuracy(%)*
PaddleOCR	0.2039	0.4105	79.61
Proposed	**0.1725**	**0.4016**	**82.75**

표 1 의 실험 결과, 제안하는 에이전트 기반 OCR 시스템은 베이스라인인 PaddleOCR 대비 약 3.14%p 향상된 정확도를 기록하였다. CER 이 0.2039 에서 0.1725 로, WER 이 0.4105 에서 0.4016 으로 모두 감소한 것은 에이전트가 합성 이미지에 포함된 기하학적 왜곡과 인위적인 조명 노이즈를 효과적으로 분석하여 최적의 전처리 도구를 적용했음을 의미한다.

표 2 는 사용자가 직접 스마트폰으로 촬영한 이미지인 그림 2 를 대상으로 ChatGPT 모델과 비교 분석한 실제 환경 평가 결과이다.

표 2. 실제 촬영 환경 이미지 기반 성능 비교

Method	*CER ↓*	*WER ↓*	*Accuracy(%)*
Paddle-OCR	0.2536	0.5430	74.64%
ChatGPT	0.1845	0.5415	81.55%
Proposed	**0.1762**	**0.5117**	**82.38%**

그림 2. 실제 문서 이미지에 대한 전처리 및 OCR 인식 결과

실제 환경 실험 결과, 82.38%의 정확도를 기록하여 PaddleOCR 대비 7.74%p, ChatGPT 대비 0.83%p 높은 성능을 달성하였다. 분석 결과 ChatGPT 는 고품질 영역에서는 높은 성능을 보였으나 저품질 영역에서는 텍스트 자체를 누락하거나 강력한 문맥 추론을 바탕으로 존재하지 않는 정보를 임의로 생성하는 Hallucination 현상을 보였다. 제안 모델은 이미지 자체의 물리적 복원을 통해 가독성을 확보하고 비정형 데이터에서도 누락 없이 인식이 가능함을 입증하였다.

Ⅲ. 결론

본 연구에서는 모바일 촬영 환경에서 발생하는 저품질 문서를 대상으로 인간의 인지 과정을 모사한 AI 에이전트 기반 OCR 기법을 제안하였다. 입력 문서의 특성에 따라 전처리를 선택적으로 적용하여 실제 촬영 환경에서도 안정적인 인식 성능을 확인하였다. 전처리 도구에 대한 과정이 기록되어 설명 가능성과 신뢰성을 확인할 수 있었으며 오픈소스 기반 온디바이스 구축을 통해 민감 데이터에 대한 보안성을 강화하고 운영 측면에서의 경제성도 확보하였다.

참 고 문 헌

[1] Y. Zhang et al., “Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models,” Computational Linguistics, pp. 1– 46, Sep. 2025, doi: 10.1162/COLI.a.16.

[2] R. Merhej et al., “Gemma 3 technical report,” arXiv:2503.19786, Mar. 2025. [Online]. Available: https://arxiv.org/abs/2503.19786

[3] F. Verhoeven, T. Magne, and O. Sorkine-Hornung, “UVDoc: Neural Grid-based Document Unwarping,” in ACM Conferences. New York, NY, USA: ACM, Dec. 2023, pp. 1– 11.

[4] C. Cui et al., “PaddleOCR 3.0 Technical Report.” July 08, 2025.