

부분 훼손 족적 검색을 위한 MLLM 기반 비대칭 리랭킹 및 보수적 특징 포함 전략

이창엽, 권현섭, 김동주, 서영주*

포항공과대학교 인공지능연구원

lcy8417@postech.ac.kr, kert07@postech.ac.kr, kkb0320@postech.ac.kr, yjsuh@postech.ac.kr

Asymmetric Reranking and Conservative Feature Inclusion Strategy based on MLLM for Partially Occluded Shoeprint Retrieval

Chang Yeop Lee, Hyeon Sub Kwon, Dong Ju Kim, Young Joo Suh*

POSTECH Institute of Artificial Intelligence

요약

범죄 현장에서 수집된 족적 이미지는 심한 부분적 훼손과 배경 노이즈로 인해, 전역적 특징에 의존하는 기존 시각 기반 검색 시스템에서 저조한 식별 성능을 보인다. 본 논문은 이러한 한계를 극복하기 위해 멀티모달 LLM(MLLM)을 활용한 새로운 비대칭 리랭킹 프레임워크를 제안한다. 제안 모델은 MLLM을 통해 족적 이미지에서 시각적 추론 텍스트와 계층적 구조 정보(Macro, Micro, Unique)를 추출한다. 이후, 비대칭적 특징 포함(Asymmetric Feature Inclusion)에 기반한 스코어링 함수를 적용한다. 특히, 식별력이 높은 희소 특징에 대해 설계된 이 전략은 환각 현상을 억제하고 정답을 강력하게 특정하는 핵심 메커니즘으로 작용한다.

I. 서론

범죄 현장에서 수집된 족적은 용의자 추적의 핵심 단서이나, 심각한 배경 노이즈와 부분 훼손으로 인해 전역적 시각 특징에 의존하는 기존 검색 시스템에서 저조한 성능을 보인다. 특히, 기존 심층 거리 학습 방식은, 훼손된 족적과 매칭되는 신발의 국소적인 패턴에 대한 포함 관계를 반영하지 못하는 근본적인 한계가 있다[1].

본 논문은 이를 극복하기 위해 멀티모달 LLM(MLLM) 기반의 비대칭 리랭킹 프레임워크를 제안한다. 제안 기법은 MLLM의 시각적 추론 능력을 활용해 족적을 Macro(전체 형태), Micro(세부 문양), Unique(특이점)의 계층적 구조 정보로 변환한다. 이후, 단순 유사도가 아닌 비대칭적 특징 포함여부를 평가하여 순위를 제조정한다. 특히 희소 특징에 대해서는 보수적 부분집합 전략을 적용하여 MLLM의 고질적인 환각(Hallucination) 현상을 억제하고 검색 신뢰성을 확보했다. 이러한 기법을 통해 열악한 데이터 환경에서의 강건성을 입증하였다.

II. MLLM 기반 비대칭 리랭킹 및 보수적 특징 포함 전략

본 논문은 부분 훼손된 족적(Q)이 온전한 신발 데이터(D)의 부분집합이라는 특성에 기반하여, MLLM을 활용한 비대칭 리랭킹 프레임워크를 그림 1과 같이 제안한다. 우선 멀티모달 LLM을 통해 족적 이미지로부터 문양의 공간적 배열을 서술하는 시각적 추론 텍스트(Visual Reasoning)와 3단계 계층적 구조 정보(Macro, Micro, Unique)를 추출한다. 이를 바탕으로 기존 검색모델의 결과인 베이스라인 점수(S_{base})를 보정한다. 결과적으로, 최종 스코어링 함수($S_{final} = S_{base} + \lambda S_{rerank}$)로 기존 랭킹을 재조정한다.

리랭킹 점수(S_{rerank})는 다음 세 가지 요소로 구성된다. 첫째, 족적의 불완전성을 고려하여 자카드 유사도와 같은 대칭적 지표 대신, 쿼리의 거시

(Macro)-미시(Micro) 특징이 타겟 신발 데이터에 포함되는 비율을 측정하는 비대칭 포함도를 계산한다.

$$S_{\text{include}}(q, r) = \sum_{k \in \{\text{micro, macro}\}} w_k \cdot \frac{|F_k^q \cap F_k^r|}{|F_k^q + \epsilon|} \quad (1)$$

둘째, MLLM의 환각 현상을 억제하기 위해 로고와 같은 희소 특징(Unique)에 대해서는 쿼리의 특징 집합이 신발 데이터에 완벽히 포함될 때만 점수를 부여하는 보수적 부분집합 전략을 적용한다.

$$S_{\text{unique}}(q, r) = I(F_u^q \subseteq F_u^r) \cdot \sum_{f \in F_u^q} w(f) \quad (2)$$

여기서 I 는 조건이 참일 때 1, 거짓일 때 0을 반환하는 지시 함수이다.

셋째, 시각적 추론 텍스트의 임베딩 벡터 간 코사인 유사도를 계산하여 문맥적 일치도를 반영한다. 이러한 전략은 단순한 형상 매칭을 넘어 논리적 포함 관계를 규명함으로써 검색의 강건성을 확보한다.

$$S_{\text{sem}}(q, r) = w_{\text{text}} \cdot \text{sim}(v_q, v_r) \cdot I(\text{sim}(v_q, v_r) > \tau) \quad (3)$$

여기서 τ 는 단순 키워드 매칭의 한계를 넘어선 문맥의 유사도를 반영하기 위한 임계값이다. 너무 낮은 임계값은 노이즈로 간주하고 점수 반영을 배제하여 안정성을 확보한다.

결과적으로, 위와 같은 결과들로 제안하는 리랭킹 함수의 최종 결과 식은 아래와 같고, 부분적으로 훼손된 족적에도 강건한 검색 성능 달성이 가능

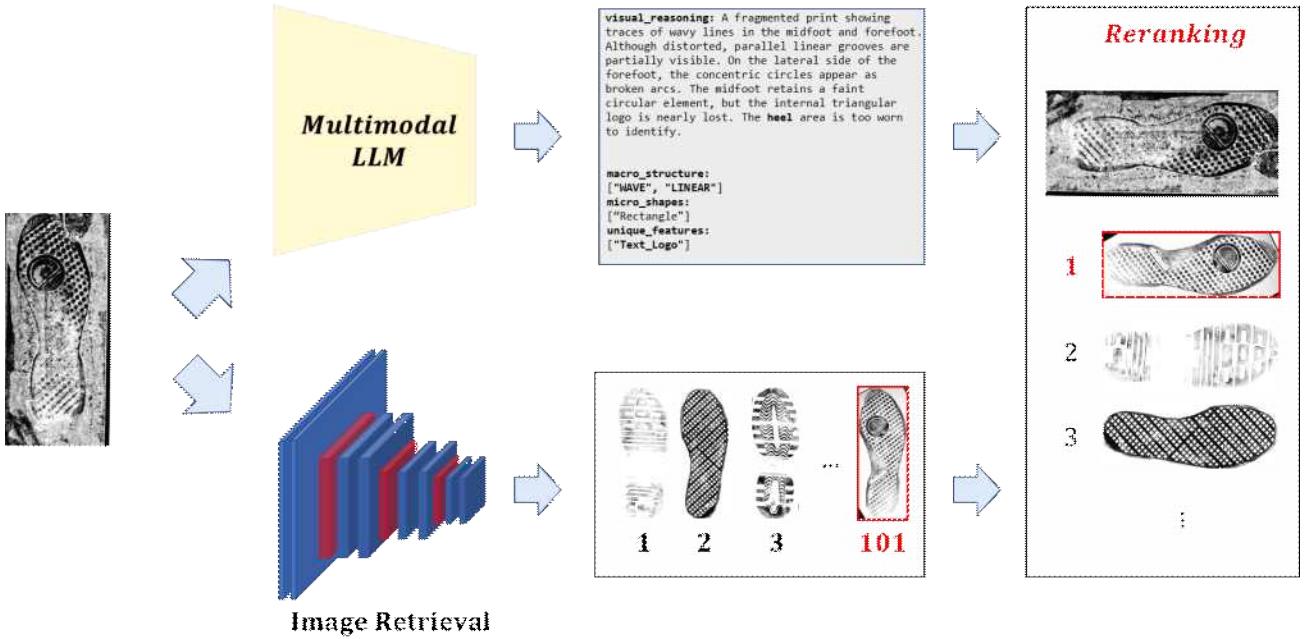


그림 1 Multimodal LLM을 활용한 비대칭 리랭킹 프레임워크

하다.

$$S_{\text{final}}(q, r) = S_{\text{base}}(q, r) + \lambda \cdot S_{\text{rerank}}(q, r) \quad (4)$$

III. 실험

본 장에서는 제안하는 하이브리드 리랭킹 프레임워크의 유효성을 검증하기 위해 경찰청으로부터 제공받은 실제 죽적 데이터를 활용하여 실험을 수행하였다. 베이스라인으로는 Token-based Retrieval 모델을 [2], 텍스트 특징 추출을 위해 Gemini Pro와 text-embedding-004 모델을 활용하였으며 [3], 리랭킹 가중치는 의미론적 텍스트 $w_{txt} = 0.9$, 회소 특징 $w_{uni} = 0.5$, 미세 및 거시 구조 $w_{mic} = w_{mac} = 0.1$, 노이즈 필터링 임계값 $\tau = 0.4$, 글로벌 가중치 $\lambda = 0.1$ 로 설정하였다. 실험 결과, 제안 모델은 모든 평가지표에서 우월한 성능을 입증하였으며, 22.09%인 베이스라인 대비 R@5에서 6.3%, R@10에서 8.8%의 성능 향상을 기록하였다. 특히 R@1에서는 12.5%의 성능 향상을 기록하였으며, 이는 죽적의 전체 형상이 훼손되어 시각적 정보만으로는 정답을 확신하기 어려운 경우에도, 비대칭적 구조 포함 전략과 시각적 추론 텍스트의 의미적 매칭이 상호 보완적으로 작용하여 오탐을 억제하고 정답을 안정적으로 상위권에 진입시켰음을 시사한다.

IV. 결론

본 논문에서는 범죄 현장의 죽적이 갖는 불완전성과 기존 시각 기반 검색 시스템의 대칭적 유사도 측정 한계를 극복하기 위해, 멀티모달 LLM을 활용한 새로운 비대칭 리랭킹 프레임워크를 제안하였다. 제안 기법은 죽적 이미지를 시각적 추론 텍스트와 계층적 구조 정보로 변환하고, 이를 바탕으로 쿼리가 신발 데이터의 부분집합이 되는지를 평가하는 비대칭 특징 포함 스코어링 함수를 적용하였다. 특히, 식별력이 높은 회소 특징에 대해 설계된 보수적 부분집합 전략은 MLLM의 환각 현상을 효과적으로 제어하며 검색의 신뢰성을 크게 향상시켰다. 실험 결과, 제안 모델은 베이스라인 대비 R@1 성능을 12.5% 개선하고 nDCG 등 주요 지표에서도 유의미

한 성과를 거두어, 열악한 데이터 환경에서의 강건성을 입증하였다. 향후 연구에서는 죽적 도메인에 특화된 경량화 MLLM을 미세 조정 (Fine-tuning)하여 추론 속도를 개선하고, 다양한 환경 변수에 대한 적응력을 강화하는 방향으로 연구를 확장할 계획이다.

ACKNOWLEDGMENT

본 논문은 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(RS-2022-NR070870).

본 논문은 과학기술정보통신부·경찰청이 공동 지원한 ‘폴리스랩 2.0 사업’의(RS-2023-00281072) 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] W. J. Bodziak, *Footwear impression evidence: Detection, recovery and examination*, CRC Press, 2017.
- [2] H. Wu et al., “Learning Token-based Representation for Image Retrieval,” Proc. of the AAAI Conf. on Artificial Intelligence, vol. 36, no. 3, pp. 2703–2711, Feb. 2022.
- [3] G. Comanici et al., “Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities,” arxiv preprint, July 2025.