2025 한양대학교

SW/ICT/AI 종합학술대회

[분야 : Paper To Code Challenge]

제목: RAG 기반 LLM 에이전트를 이용한 로봇 시공간 메모리 시스템

이정호(23학번), 안선우(25학번)

2025. 11. 28

한양대학교 ERICA 소프트웨어융합대학 / 한양대학교 ERICA SW중심대학사업단

연구 배경 및 문제 정의

현재 로봇 현황



- 로봇은 시간 단위(hours)로 지속적으로 운용
- 다양한 객체, 동적 이벤트, 현상을 지속적으로 관찰
- 기존 맵(metric/semantic)으로 표현 불가능한 정보 축적

기존 접근법의 한계

접근법	처리 가능 시간	문제점
Transformer 기반	1~2분	메모리 증가, 추론 시간 폭발적 증가
대규모 Context Window	고정 길이	무한한 히스토리 처리 불가
기존 Episodic Memory	짧은 구간	실시간 행동 정보 미제공

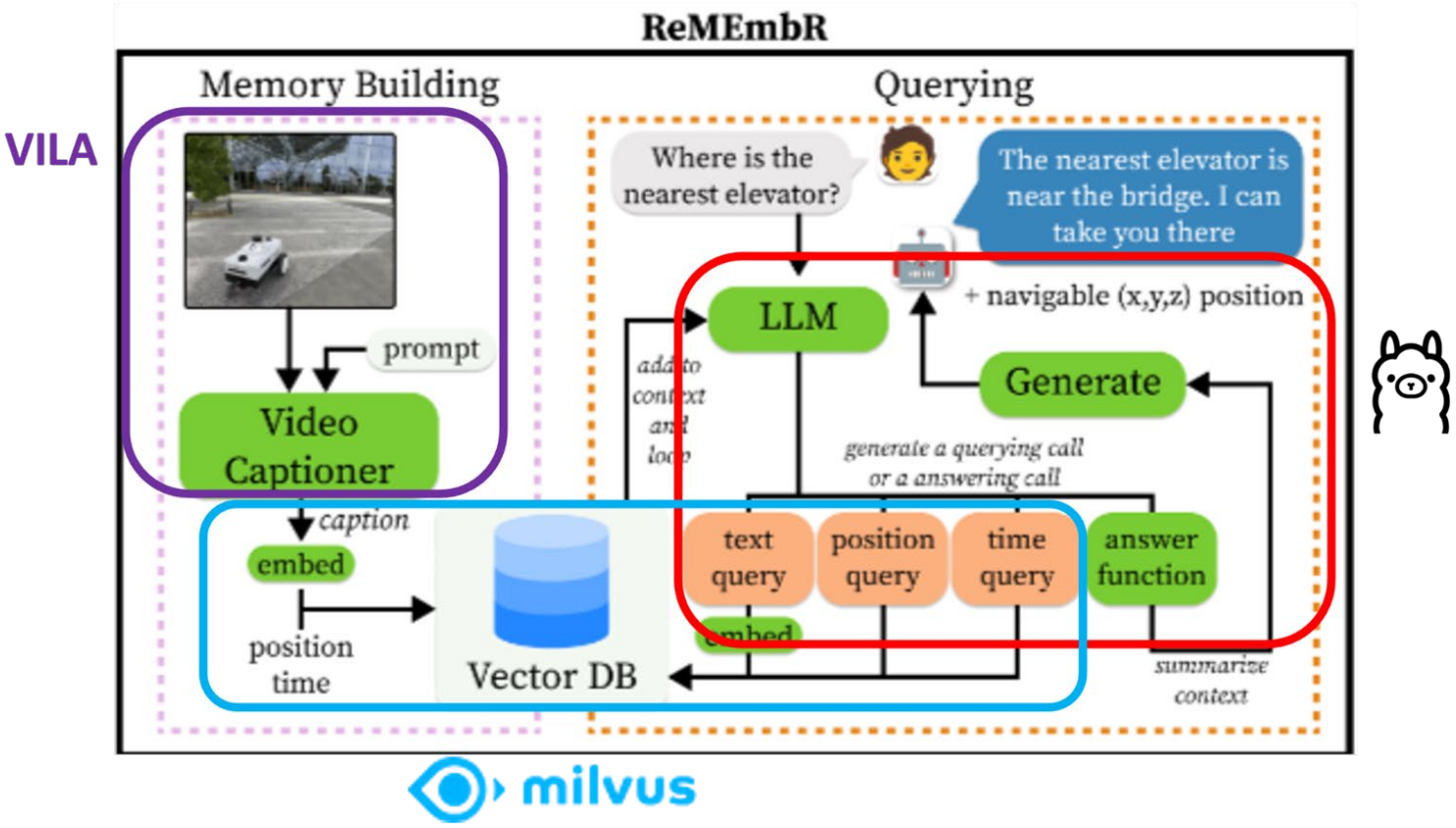
핵심 연구

- 🕐 **Temporal**: "When did you see the boxes fall?" → "15 minutes ago"
- 📍 **Spatial**: "Where is the closest bathroom?" → (x,y) coordinates
- 📝 **Descriptive**: "Was the sidewalk busy today?" → Yes/No or detailed text

핵심 과제: 로봇의 장시간 히스토리 $H_{1:k}$를 효율적으로 구축하고, 사용자 질문 Q에 대해 실시간으로 답변 A를 생성하는 시스템 필요

제안 시스템: 시공간 메모리 기반 RAG-LLM 에이전트

System Architecture

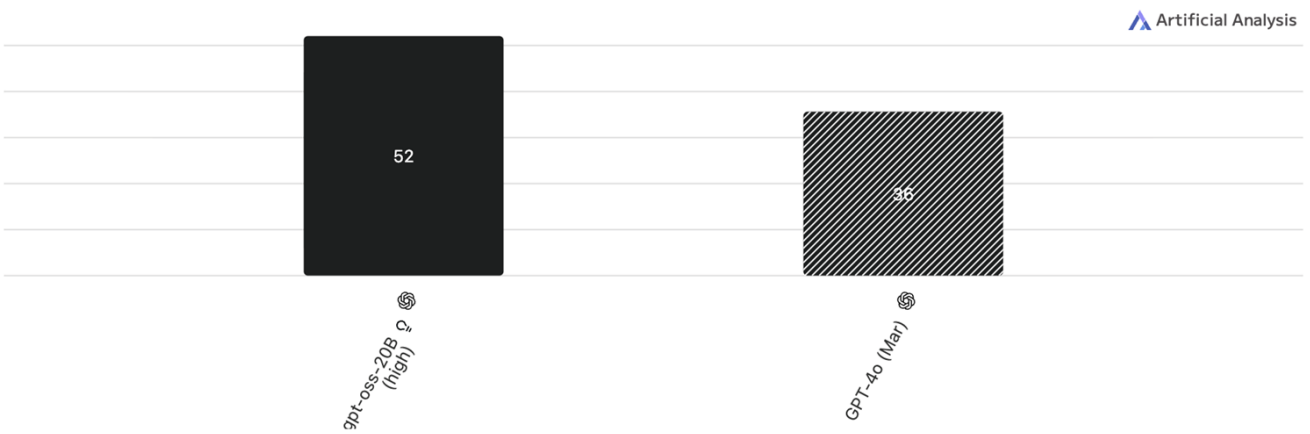


확장성: 무한한 길이의 히스토리 처리 가능
효율성: 필요한 정보만 선택적 검색 (25초/질문)
실용성: 로봇 네비게이션에 직접 활용 가능한 (x,y) 좌표 출력
다중 모달: 시각, 공간, 시간 정보 통합 처리

구현 시스템 아키텍처 (OURS 비교)

구현 시스템 아키텍처

구성 요소	📄 논문 (ReMEmbR)	🔧 우리 구현 (OURS)	💡 개선점
LLM 모델	• GPT-4o (클라우드) • Codestral/Command-R • Llama3.1:8b	**Ollama + GPT-oss 20** • 로컬 실행 • API 비용 없음	비용 절감 지연시간 감소 데이터 보안
비디오 캡셔닝	VILA1.5-13b (3초당 6프레임)	VILA-3b (양자화) (3초당 4프레임)	메모리 효율 실시간 처리
벡터 DB	일반 Vector DB	**Chroma DB** • 로컬 저장 • 영구 보존	오프라인 동작 빠른 검색
임베딩 모델	mxbai-embed-large-v1	**sentence-transformers** all-MiniLM-L6-v2	경량화 빠른 임베딩
하드웨어	Nova Carter Jetson Orin 32GB	**일반 PC/노트북** • RTX 3060 12GB • RAM 16GB	접근성 향상 개발 용이



실험 구성

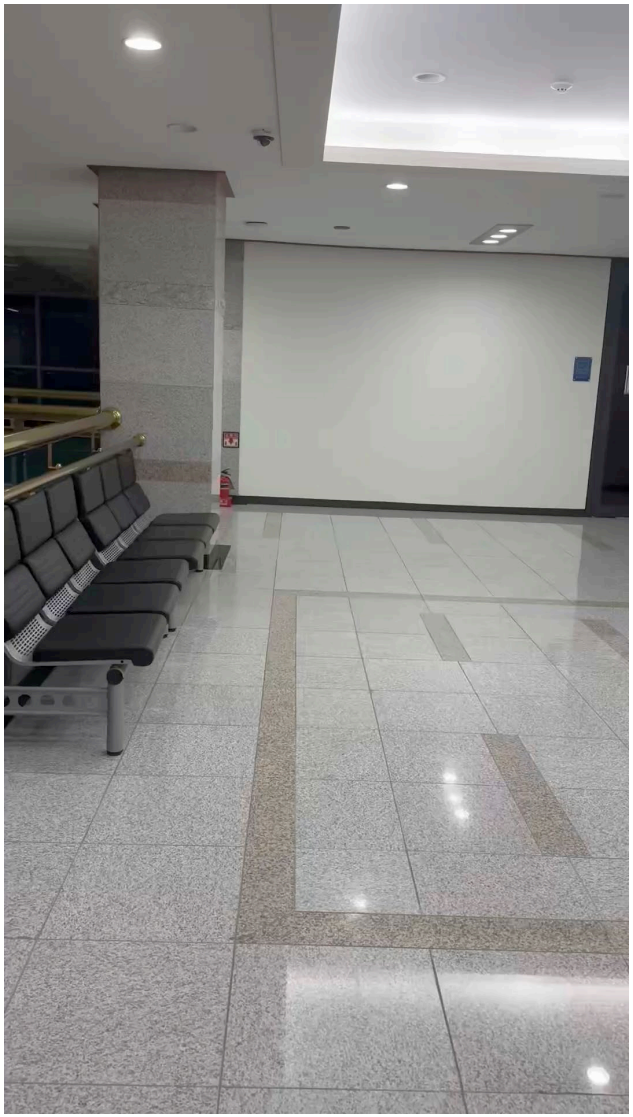
데이터셋

학연산 클러스터 2층 1분 영상

테스트한 모델

VLM (VILA-8B)

촬영 영상



실제 적용 영상

VILA 캡션 결과



```
Enter your question (or 'quit' to exit):
> what is in this video?

[Processing...]

=== Response ===
The video shows a long, empty hallway with white walls and a shiny, reflective floor. There are several doors along the hallway, and at the end, there's a large, white wall with a black stripe running across it. The hallway is well-lit wit
h overhead lights, and there are some signs and a fire extinguisher visible. The overall atmosphere is clean and sterile, typical of a hospital or office building.
============================================================
```

실험 결과 및 분석

실험 결과

영상 캡션 모델은 연속된 장면의 객체·행동·상황을 이해해

자연스러운 설명 문장을 자동 생성한다



기대효과

🔍 Finding 1: 확장성

- 긴 비디오에서도 일정한 성능
- 메모리 증가율: 4MB/분
- 무제한 길이 처리 가능

🚀 Finding 2: 실용성

- 완전 오프라인 동작
- 일반 하드웨어에서 구동
- 개인정보 보호 완벽

💡 Finding 3: 최적화 가능성

- 양자화로 추가 경량화 가능
- 캐싱으로 반복 질문 가속
- 멀티모달 LLM 통합 가능

결론 및 향후 연구

결론

- ReMEmbR는 장시간 로봇 관찰 데이터를 RAG 방식으로 효율적 관리
- 시공간 질의에 대해 빠르고 정밀한 답변 가능
- 실제 로봇 환경에서도 **실시간 수준의 속도·정확도** 입증
- LLM을 로봇의 장기 메모리 모듈로 사용할 수 있음을 보여줌

향후 연구 방향

ReMEmbR Robot Navigation

- **여러 정답이 가능한 상황 대응 (Ambiguous QA)**
- **비디오 캡션 외 추가 정보 활용**
  - 방 번호 / 표지판 / 장비 라벨
  - Semantic Map / Scene Graph 등
  - 음성 데이터
- **중복·불필요 기억 자동 압축**
- **더 정교한 메모리 선택 및 강화학습 기반 Retrieval**
- **온디바이스 LLM 최적화로 더 빠른 응답 가능**