

[붙임5] 2026년 소·중·대 산학협력프로젝트(캡스톤디자인) 결과보고서

2026년 전남대학교 소프트웨어중심대학사업 소·중·대 산학협력프로젝트(캡스톤디자인) 결과보고서

프로젝트명	단안 영상 기반 3D 공간 분석 AI 챗봇					
Github url 주소	(없을 시 공란)					
팀 명	원황찬란			과제수행기간	2026. 3. 25. ~ 6. 23.	
지도교수	학 과	컴퓨터정보통신공학과		성 명	김명진	
프로젝트 수행인원 (※팀장은 첫줄에 기입)	학과(부·복수전공)	학년	학번	성명	연락처	E-Mail
	컴퓨터정보통신공학과	4	214732	정주현	01067887585	7585sy@naver.com
	컴퓨터정보통신공학과	4	204637	황찬희	01052920688	hwagchan@gmail.com
	컴퓨터정보통신공학과	4	214803	김성찬	01046908377	awdr1057@naver.com
참여 기업	기업명	멘토명		직위	연락처(HP)	E-Mail
	티제이에코시스	황민		이사	-	-

위와 같이 2026년 전남대학교 소프트웨어중심대학사업
산학협력프로젝트(캡스톤디자인) 지원 프로그램 결과보고서를 제출합니다.

2026년 6월 23일

신청자명(대표학생) : 정주현
지도교수 : 김명진

정주현
(인)

전남대학교 소프트웨어중심대학사업단장 귀하

산학협력프로젝트(캡스톤디자인) 결과보고서(요약)

프로젝트명	단안 영상 기반 3D 공간 분석 AI 챗봇		
수행기간	2026. 3. 25. ~ 6. 23.	소요예산	267,519원
소요예산 세부내역	-회의비 197,700 -SW활용비 69,819		
참여인원	구분	인원수	성명(모두 기재)
	교수	1	김명진
	석박사과정	-	
	학부생	3	황찬희, 정주현, 김성찬
	기업체	1	황민
계			
계			
추진배경	단안 카메라로 촬영한 실내 영상에 대해 자연어 질의로 3차원 공간 정보를 이해하고 응답하는 인공지능 챗봇 서비스를 개발하였다. 시각언어모델 VLM-3R에 3차원 공간 인코더 CUT3R를 결합한 멀티모달 추론 파이프라인을 구축하고, 이를 메시지 큐 기반의 비동기 백엔드로 구현하여 가구의 위치, 개수, 공간 구조 등 다양한 질의에 응답하도록 하였다. 별도의 깊이 센서 없이 일반 스마트폰 영상만으로 동작하며, 수십 초가 소요되는 무거운 AI 추론을 안정적으로 서비스하기 위해 비동기 처리, 데이터 정합성, 캐시 기반 부하 분산을 적용한 백엔드를 설계하고 부하테스트로 검증하였다.		
목표 및 내용	첫째, 단안 영상 기반 멀티모달 추론 파이프라인을 완성하였다. LLaVA- NeXT-Video 기반 7B 시각언어모델과 CUT3R 공간 인코더를 분리 로드 구조로 설계하고 4비트 양자화를 적용하여 8GB VRAM 환경에서 동작하도록 하였으며, 영상당 특징 추출 약 25초, 채팅 응답 약 5초로 측정되었다. 둘째, AI 추론 지연을 사용자 경험과 분리하는 비동기 아키텍처를 구현하였다. Spring 백엔드는 업로드 시 작업 식별자를 즉시 반환하고 분석 요청을 RabbitMQ로 전달하며, 완료 결과를 SSE로 실시간 전달한다. 이를 통해 45초 이상의 추론 중에도 사용자 요청 처리가 지연 없이 유지되고 외부 AI 장애가 격리된다. 셋째, 분산 환경의 데이터 정합성과 메시지 신뢰성을 보장하였다. 애플리케이션 검증과 DB UNIQUE 제약을 결합한 이중 먹등성으로 중복 요청 시에도 단일 레코드만 생성됨을 검증하였고, 외부 호출을 트랜잭션 경계 밖으로 분리하여 AI 장애가 정합성을 훼손하지 않도록 하였다. 넷째, 제한된 자원에서 시스템 한계를 부하테스트로 정량 규명하였다. 1GB, 2코어 환경의 k6 테스트에서 Redis 캐싱 적용 시 처리량이 약 2.4배 향상되고 DB CPU가 0%로 유지되어 캐시의 DB 보호 효과를 확인하였다. 또한 커넥션 풀 권장 공식이 본 환경에 부적합함을 측정으로 확인해 재조정함으로써 최악 지연을 약 80% 감소시켰으며, 실시간 연결의 병목이 메모리 자원임을 규명하였다.		
기대효과	본 성과는 일반 영상만으로 3차원 공간 이해와 자연어 응답이 가능하다는 점에서 부동산 매물 안내, 인테리어 설계 보조, 시각장애인 환경 인식 지원, 무인 매장 모니터링, 실내 안전 점검 등으로 확장 가능하다. 깊이 센서 장비 없이 촬영 영상만으로 동작하므로 서비스 진입 장벽을 낮춘다. 모델 분리 로드 구조와 비동기 파이프라인은 더 큰 모델로의 교체나 다중 사용자 확장 시에도 재사용 가능하며, 메시지 큐 기반 계약 구조는 다른 멀티모달 AI 서비스의 참고 패턴이 된다. 특히 본 연구는 AI로 구현 속도는 빨라졌으나 안정성 검증은 여전히 사람의 영역이라는 문제의식에서, 제한된 자원에서 한계를 측정하고 데이터에 근거해 설정을 결정하는 방법론을 실증하였다. 이 결과는 향후 확장 시 용량 산정과 장애 대응의 근거가 되며, 후속 연구에 기여할 것으로 기대된다.		

1. 프로젝트 개요

프로젝트명	단안 영상 기반 3D 공간 분석 AI 챗봇
주제영역	☐ 생활 ☒ 업무 ☐ 공공/교통 ☐ 금융/핀테크 ☐ 의료 ☐ 교육 ☐ 유통/쇼핑 ☐ 엔터테인먼트
기술분야	☐ IoT ☐ 모바일 ☐ 데스크톱 SW ☒ 인공지능 ☐ 보안 ☐ 가상현실 ☐ 빅데이터 ☐ 자동제어기술 ☐ 블록체인 ☐ 영상처리 ☐ 기타()
성과목표	☐ 논문게재 및 포스터발표 ☐ 앱등록 ☐ 프로그램등록 ☐ 특허 ☐ 기술이전 ☒ 실용화 ☐ 공모전(<i>공모전명</i>) ☐ 기타()

2. 프로젝트 추진배경

자율주행차, 서비스 로봇, 드론, AR/VR로 대표되는 '피지컬 AI(Physical AI)' 수요가 빠르게 증가하면서, 기계가 현실 공간을 이해하는 능력의 중요성이 커지고 있다. 글로벌 3D 비전 시장은 2024년 약 150억 달러에서 2030년 400억 달러 이상으로 연평균 15%를 상회하는 성장이 전망된다. 자율 시스템이 복잡한 실제 환경에서 안전하게 동작하기 위해서는 2차원 객체 인식을 넘어 사물 간 거리와 공간 구조를 파악하는 3차원 공간 지능(3D Spatial Intelligence)이 필수적이다. 그러나 현재의 3D 인식 기술은 다음 세 가지 구조적 한계에 직면해 있다.

첫째, 고가 하드웨어 의존성이다. 기존 3D 인식은 LiDAR, 스테레오 카메라, RGB-D 센서 등 전용 장비를 전제로 하며, 자율주행용 LiDAR는 수백만 원에서 수천만 원에 이른다. 반면 단안 카메라는 수만 원대 USB 카메라나 스마트폰으로 대체할 수 있어 약 100배 이상의 비용 절감이 가능하다. 이러한 비용 부담은 소형 로봇, 드론, 소비자용 디바이스의 상용화를 가로막는 핵심 장벽으로 작용한다.

둘째, 환경 민감성이다. 강한 햇빛이나 야간 조건에서는 깊이 추정 정확도가 저하되고, 비, 안개, 눈에서는 LiDAR 신호가 산란되어 거리 측정 오류가 발생한다. 동적 객체가 많은 환경에서는 센서 간 시간 동기화와 데이터 불일치 문제로 인해 안정적인 3차원 구조 추정이 어려워 실제 운용 환경에서의 신뢰성이 저하된다.

셋째, 센서 융합의 높은 복잡도이다. 다수 센서의 데이터를 통합 처리하기 위해서는 고성능 GPU와 병렬 처리 구조가 필요하여 전력 소모가 크고, 설계, 유지보수, 확장 단계의 복잡도가 상승하여 산업 현장 적용과 서비스화에 막대한 인프라 비용이 발생한다.

3. 프로젝트(주제) 목표 및 내용

본 연구는 단안 카메라 영상으로부터 3차원 공간 정보를 추정하고, 이를 자연어 인터페이스를 통해 사용자에게 제공하는 통합 서비스 'MonoSpace AI'를 개발하는 것을 목표로 한다. 이를 위해 본 연구는 이론적, 실험적 접근을 결합한 네 단계의 수행 절차를 채택하였다. 먼저 단안 영상 기반 3D 분석 모델과 경량화 기법, 분산 시스템 설계 방법론에 대한 이론적 사전 조사를 수행하고, 이를 바탕으로 전체 시스템 아키텍처와 통신 구조를 설계하였다. 이어서 AI 추론 파이프라인과 백엔드 서비스의 핵심 모듈을 구현하였으며, 마지막으로 제한된

자원 환경에서 부하테스트를 통해 모든 설계 결정의 타당성을 정량적으로 검증하였다.

각 단계는 독립적으로 분리되지 않고 하나의 파이프라인으로 유기적으로 연결된다. 본 연구가 다른 학습 프로젝트와 구별되는 지점은 구현 자체가 아니라 검증의 비중에 있다. 오늘날 대규모 모델과 풍부한 프레임워크 덕분에 기능을 동작시키는 일은 비교적 빠르게 가능해졌으나 그렇게 만들어진 시스템이 제한된 자원과 외부 장애, 동시성 충돌 속에서도 안정적으로 동작함을 보장하는 일은 전혀 다른 차원의 과제이다. 본 연구는 '누구나 빠르게 만들 수 있는 시대에 안정성 검증은 별개의 문제'라는 문제의식 아래, 구현과 측정 기반 검증을 동등한 비중으로 다루었으며, 모든 설계 결정을 직관이 아닌 정량적 측정 결과에 근거하여 확정하였다

4. 시스템 구성 및 내용

이론적 조사 결과를 토대로 본 연구는 다음과 같은 계층형 분산 마이크로서비스 아키텍처를 설계하였다. 단일 서버에 모든 기능을 통합하는 모놀리식 구조 대신 마이크로서비스 구조를 택한 이유는 무거운 GPU 연산을 수행하는 AI 서버와 데이터 정합성을 책임지는 백엔드 서버의 자원 특성과 장애 특성이 근본적으로 다르기 때문이다. 두 영역을 분리함으로써 한쪽의 장애나 부하가 다른 쪽으로 전파되는 것을 막고, 각 영역을 독립적으로 개발, 배포, 확장할 수 있도록 하였다.

클라이언트 계층은 React 기반 웹 클라이언트(Vercel 배포)와 React Native 기반 모바일 클라이언트로 구성되며, REST API와 SSE를 통해 백엔드와 통신한다. API 게이트웨이 계층은 Spring Boot 기반 백엔드 서버가 담당하며, 사용자 요청의 진입점이자 데이터 정합성과 트랜잭션 경계를 책임진다. AI 추론 계층은 FastAPI 기반의 비동기 AI 서버로, 양자화된 VLM-3R과 CUT3R을 이용한 무거운 멀티모달 추론을 전담한다. 메시지 큐 계층은 RabbitMQ가 담당하여 AI 추론 완료 결과를 백엔드로 비동기 전달하며, 캐시 계층의 Redis는 반복적인 조회 트래픽을 흡수하여 데이터베이스를 보호한다. 데이터 저장 계층은 영상, 결과 파일과 특징벡터를 보관하는 AWS S3, 사용자, 공간 정보를 저장하는 MariaDB, 대화 이력을 저장하는 MongoDB로 구성된다. 아울러 분석 완료 시점에 사용자에게 결과를 알리기 위한 FCM 기반 푸시 알림 채널을 두어 사용자가 분석 완료를 기다리며 화면을 지키지 않아도 되도록 하였다. 각 계층은 RESTful API와 메시지 큐를 통해 통신하며, 전체 시스템은 AWS EC2 인스턴스 위에서 Docker/Compose로 컨테이너화되어 운영되고, 배포는 GitHub Actions를 통해 자동화하였다. 외부 트래픽은 Nginx 리버스 프록시를 통해 유입된다.

5. 프로젝트 결과물에 대한 기술

본 연구는 단안 카메라로 촬영한 실내 영상에 대해 자연어 질의를 통해 3차원 공간 정보를 이해하고 응답하는 인공지능 챗봇 서비스 'MonoSpace AI'를 완성하였다. 별도의 LiDAR나 깊이 센서, 3차원 스캐너 없이 일반 스마트폰으로 촬영한 영상 한 건만으로 공간 분석과 자연어 질의응답이 가능한 통합 서비스를 End-to-End로 구축하였으며, 연구 착수 시 계획한 AI 추론, 백엔드/인프라, 통합/검증의 세 영역에서 다음과 같은 정성적 성과를 도출하였다.

첫째, 멀티모달 3D 추론 파이프라인을 완성하였다.

시각-언어 모델 VLM-3R과 3차원 공간 인코더 CUT3R을 결합하여, 단안 영상으로부터 공간 토큰과 카메라 토큰을 추출하고, 이를 LLM의 임베딩 차원에 정렬하여 시각, 공간, 텍스트를 하나의 통합 입력 시퀀스로 처리하는 추론 구조를 구현하였다. 그 결과 단순 깊이 추정으로는 불가능한, 공간 크기, 가구 위치, 객체 개수 등 관계적, 의미론적 질의에 대한 자연어 응답이 실제로 동작함을 확인하였다. 실서비스 환경에서 “촬영한 공간의 크기를 알려줘”라는 질의에 대해 “촬영된 공간의 크기는 10.24m²입니다”와 같이 정량적 수치를 포함한 응답을 생성하였다.

둘째, 제한된 GPU 자원에서 대규모 모델을 구동하는 경량화 기술을 확보하였다. 4비트 양자화와 Pipeline Split(전처리, 추론 분리) 구조를 적용하여 70억 파라미터 모델을 8GB VRAM 환경에서 구동 가능하도록 최적화하였으며, 추론 시 CUT3R을 GPU에 적재하지 않는 구조를 통해 약 1.5GB의 VRAM을 절감하였다. 이로써 고가의 대용량 GPU 없이도 대규모 시각-언어 모델을 실제 서비스에 투입할 수 있음을 실증하였다.

셋째, 자원 제약 환경에서의 안정적 서비스 아키텍처를 구축하였다. React 클라이언트, Spring 백엔드, FastAPI AI 서버, RabbitMQ, Redis, MariaDB/MongoDB/S3로 구성된 분산 마이크로서비스 구조를 설계하고, 무거운 AI 추론은 비동기 채널로, 가벼운 조회, 채팅은 동기 채널로 분리하였다. 또한 이중 먹등성에 의한 중복 요청 방어, 트랜잭션 경계 분리에 의한 외부 장애 격리, Redis 캐싱에 의한 조회 부하 완화, 메시지 신뢰성 설계 등 운영 안정성을 위한 핵심 설계를 모두 구현하였다.

넷째, 모든 설계 결정을 정량적으로 검증하였다. EC2 1GB RAM, 2vCPU의 제한된 자원 환경에서 k6 부하테스트를 수행하고, 평균값이 아닌 재현 가능한 p95/p99 지표를 기준으로 캐싱 효과, 커넥션 풀 최적화, S3 우회 업로드 안정성, SSE 동시 연결 한계를 측정하였다. 이를 통해 직관이 아닌 측정에 근거하여 병목을 규명하고 설정값을 재결정하는 검증 방법론을 확립하였다.

6. 프로젝트 진행내용

1) 참여인원 및 담당 역할

연번	소속학과	성명	수행역할 분담내용
1	컴퓨터정보통신공학과	황찬희	AI 추론
2	컴퓨터정보통신공학과	정주현	AI 서버 패키징 및 프론트엔드
3	컴퓨터정보통신공학과	김성찬	백엔드

2) 회의 및 SW멘토링 진행

번호	일시/장소	회의/멘토링 내용(상세히 작성)	관련 사진
1	2026. 4. 27. (16:00~17:00) / 공대7호관 팬도로시 2층	- 개발 대상 정확히 세분화하여 목표 다시 잡기	
2	2026. 6. 11. (17:00~18:00) / 공대7호관 팬도로시 2층	- 경쟁사 조사 철저히 하여 사업화에 대한 명분 설명 보완	

7. 프로젝트 세부일정 및 내용

No.	작업 내용	4월				5월				6월				담당자	비고
		1	2	3	4	1	2	3	4	1	2	3	4		
1	계획서 작성													황찬희/ 정주환/ 김성찬	
2	요구사항 분석 및 시스템 설계													황찬희/ 정주환/ 김성찬	
3	페이지 UI 설계													황찬희/ 정주환/ 김성찬	
4	시스템 개발													황찬희/ 정주환/ 김성찬	
5	시스템 개발													황찬희/ 정주환/ 김성찬	
6	테스트 및 검증 (실 데이터 적용)													황찬희/ 정주환/ 김성찬	
7	결과보고서 작성/제출													황찬희/ 정주환/ 김성찬	

8. 결과물에 대한 향후 활용계획

1) 성과 관리 추진체계

본 연구의 성과물은 소스 코드, 시스템 아키텍처 문서, API 명세서(Scalar), GPU 환경 셋업 가이드, 부하테스트 결과 데이터로 구성된다. 모든 소스 코드는 Git 기반 버전 관리 체계로 통합 관리하며, GitHub Actions를 통한 자동화 배포 파이프라인을 유지하여 코드 변경 이력과 배포 이력을 일관되게 추적한다. AI 추론 서버와 백엔드 서버, 클라이언트는 각각 독립된 저장소로 분리 관리하되, 시스템 간 인터페이스 계약은 API 명세서를 단일 기준 문서로 삼아 동기화한다. 특히 본 연구에서 통합 단계 문제의 상당수가 인터페이스 계약 불일치에서 비롯되었던 경험을 반영하여 명세서의 갱신과 코드의 변경을 항상 연동하여 관리한다. 또한 부하테스트 결과와 운영 과정에서 축적된 의존성 충돌 사례 및 트러블슈팅 기록은 문서로 정리하여, 향후 유지보수와 후속 연구의 참조 자산으로 보존한다.

2) 예상되는 활용 분야 및 활용 방안

본 연구의 성과는 단안 영상만으로 3차원 공간 이해와 자연어 응답이 가능하다는 특성상 다양한 분야로 확장 활용이 가능하다. 실내 자율주행 분야에서는 LiDAR 없이 RGB 카메라만으로 공간을 인식하고 장애물을 회피하는 배달, 방역 로봇 등 보급형 기기에 적용할 수 있다. 산업 안전 관리 분야에서는 CCTV 영상을 분석하여 작업자와 위험 기계 간 거리를 정량화하고 위험을 자동 경고함으로써 비전문가도 즉각 위험을 인지할 수 있는 모니터링 체계를 구축할 수 있다. 부동산, 인테리어 분야에서는 스마트폰 영상으로 공간 치수와 객체 정보를 추출하여 가구 배치 시뮬레이션과 매물 정보 자동 등록에 활용할 수 있으며, 시각 보조 분야에서는 주변 환경을 음성으로 안내하는 보조 공학 서비스로 활용할 수 있다. 공공 안전, 재난 분야에서는 드론, CCTV 영상을 기반으로 진입 경로와 위험 요소를 분석하여 구조 의사결정을 지원할 수 있다. 이러한 활용은 모두 별도의 깊이 센서나 BIM 데이터 없이 일반 사용자가 촬영한 영상만으로 동작하므로, 서비스 진입 장벽이 낮다는 공통된 강점을 갖는다.

9. 참고자료 및 문헌

- 1) <https://vlm-3r.github.io/>
- 2) <https://cut3r.github.io/>