KCGS 2024
Generative Imagination

한국컴퓨터그래픽스학회

2024 학술대회

학술대회 논문집

한국컴퓨터그래픽스학회 2024 학술대회 논문집

2024년 7월 9일-12일
소노벨 경주

KCGS 2024
Generative Imagination

주관기관: 한국컴퓨터그래픽스학회

KCGS 2024 환영사

신라 천년의 역사를 품고 있는 이곳 경주에서 한국컴퓨터그래픽스학회 2024년 학술대회와 여름학교를 개최하게 됨을 매우 기쁘게 생각하며, 이 자리에 참석해 주신 모든 분을 환영합니다.

지난해 30주년을 맞이한 한국컴퓨터그래픽스학회는 학계와 산업계의 연구자들이 모여 아이디어를 공유하고 함께 성장할 기회를 제공해 왔습니다. 본 학술대회에서는 컴퓨터그래픽스와 관련된 다양한 분야인 컴퓨터게임, 영화/방송특수효과, 가상/증강현실, 인간컴퓨터상호작용, 메타버스, 미디어아트, 로보틱스 등을 포괄하여 협력과 공유, 소통을 통해 컴퓨터그래픽스 분야의 앞으로 나아갈 방향을 모색하고자 합니다.

올해 학술대회에서는 'Generative Imagination'이라는 주제로 초청 강연, 우수해외학술대회 논문발표, 창해신진연구자 및 석사논문상 후보자 발표, 디지털콘텐츠-미디어 중점 연구개발 계획 및 특성화 연구실 소개, 산업체 및 대학전시 등 다양한 프로그램이 준비되어있습니다. 또한, 이 자리에서는 본 학회의 발전을 위해서 남다른 열정과 헌신적인 노력으로 기여 해오신 중앙대학교 윤경현 교수님께서 공로패를 수상하시게 되어서 축하의 말씀을 전하고자 합니다.

끝으로 이번 행사를 준비하는데 수고해 주신 모든 분께 깊이 감사드리며, 이 행사가 성공적으로 진행될 수 있도록 지원해 주신 관계 기관과 기업체에 감사의 인사를 드립니다. 나흘 동안 진행되는 이번 행사에서 유익하고 즐겁게 지내시기를 바라며, 여러분들의 건강과 행복을 기원합니다.

2024년 7월 9일
사단법인 한국컴퓨터그래픽스학회장 최수미

학술대회 조직위원회

대회장	최수미 (세종대)		
총무	강형엽 (경희대)	김종현 (인하대)	
조직위원장	박진호 (숭실대)	최명걸 (카톨릭대)	
프로그램위원장	이성희 (KAIST)	장윤 (세종대)	
여름학교준비위원장	송오영 (세종대)	한다성 (한동대)	
조직위원	김성희 (동의대)	원정담 (서울대)	유리 (아주대)
	이정진 (숭실대)	이주호 (서강대)	허재필 (성균관대)
프로그램 위원	강형엽 (경희대)	계희원 (한성대)	고일주 (숭실대)
	권구주 (배화여대)	권태수 (한양대)	김광욱 (한양대)
	김구진 (경북대)	김덕수 (코리아텍)	김선정 (한림대)
	김영민 (서울대)	김익재 (KIST)	김종현 (인하대)
	김준호 (국민대)	김진모 (한성대)	노준용 (KAIST)
	류관희 (충북대)	문보창 (GIST)	박진아 (KAIST)
	백승환 (POSTECH)	송오영 (세종대)	송현주 (숭실대)
	신병석 (인하대)	신현준 (아주대)	안상철 (KIST)
	유리 (아주대)	윤승현 (동국대)	윤영호 (리콘랩스)
	이강훈 (광운대)	이석원 (KETI)	이영호 (목포대)
	이윤상 (한양대)	이인권 (연세대)	이종원 (세종대)
	이창하 (중앙대)	이택희 (한국산업기술대)	이환용 (아주대)
	정문열 (서강대)	정원기 (고려대)	조동식 (울산대)
	조성현 (POSTECH)	최민규 (광운대)	최유주 (SMIT)
	한다성 (한동대)	한정현 (고려대)	허재필 (성균관대)
	홍헬렌 (서울여대)		

목 차

초청강연

클로버추얼패션의 제품 개발 방법 12
오승우 대표, 클로버추얼패션 7월 10일 17:10~18:10

인공지능 시대의 음악 창작과 연구 13
남주한 교수, KAIST 7월 11일 12:40~13:40

햅틱 효과 저작 14
최승문 교수, POSTECH 7월 11일 15:50~16:50

IITP 특별세션: 디지털콘텐츠-미디어 중점 연구개발 계획 및 특성화 연구실 소개

7월 10일 14:30~17:00

메타버스 (MX: Metaverse Transformation) 중장기 연구개발 사업 기획
IITP

메타버스 융합대학원 소개
김동호 교수, 숭실대

ITRC: 초실감 XR 연구센터
최수미 교수, 세종대

방송-디지털미디어 중장기 연구개발 사업 기획
IITP

ITRC: Simulated Reality 연구센터
김영준 교수, 이화여대

창해 신진 연구자상 발표

7월 11일 13:50~15:40

창해 신진 연구자상 후보자: 이신영 박사 16
창해 신진 연구자상 후보자: 김효민 박사 17
창해 신진 연구자상 후보자: 장덕경 박사 18

석사 논문상 후보자 발표

7월 11일 13:50~15:40

석사 논문상 후보자: 최은수 20
석사 논문상 후보자: 류누리 21
석사 논문상 후보자: 윤관 22

우수 국제 학술대회 논문 발표: 애니메이션1

7월 10일 13:10~14:20

ClothCombo: Modeling Inter-Cloth Interaction for Draping Multi-Layered Clothes 24
이도해, 강현, 이인권 (연세대)

Strategy and Skill Learning for Physics-based Table Tennis Animation 26
Jiashun Wang, Jessica Hodgins (CMU) , 원정담 (서울대)

우수 국제 학술대회 논문 발표: 애니메이션2

7월 10일 14:30~15:40

SAME: Skeleton-Agnostic Motion Embedding 29
이선민, 강태호, 박정남, 이제희, 원정담 (서울대)

우수 국제 학술대회 논문 발표: 모델링/이미지/비디오

7월 10일 15:50~17:00

3Doodle: Compact Abstraction of Objects with 3D Strokes 32
최창운, 이재아, 박재식, 김영민 (서울대)

논문 발표: 시각화/렌더링

7월 11일 09:00~10:15

점진적 광자 매핑을 위한 기울기 계산 기법 (특별호)
전동희, 구정민, 문보창 (GIST)

H-ETC2: 모바일을 위한 CPU-GPU 하이브리드 텍스처 압축 기법 35
이현기, 나재호 (상명대)

Origin-Destination 기반 시각화 도구의 개발 및 사회 문제 해결을 위한 사용자 연구 (특별호)
김창기, 황성진, 김한성, 이수기, 차재혁, 김광욱 (한양대)

이기종 원격협업을 지원하는 확장현실 기반 데이터 시각화 (특별호)
하효지, 김현우, 김영서, 박상훈 (서강대)

NeRF, PBD 및 병렬 리샘플링을 결합한 실시간 3D 볼륨 변형체 시각화 (특별호)
권상민, 전소진, 박준이, 김다솔, 계희원 (한성대)

논문 발표: 애니메이션/시뮬레이션1

7월 11일 09:00~10:15

전신 모션 데이터를 이용한 실시간 얼굴 애니메이션 생성 38
장보경, 이윤상 (한양대)

인체 근골격 모델의 현실적인 보행 동작 생성을 위한 평균 상태 손실 함수를 이용하는 모델 기반 강화학습 40
김민관, 이윤상 (한양대)

시뮬레이션 되는 휴머노이드 컨트롤을 위한 잠재변수 매칭 42
김민수, 이윤상 (한양대)

두 캐릭터의 상호작용 모션에 대한 기하학을 고려한 모션 리타겟팅 44
장인서, 최수진, 홍석현, 김채린, 노준용 (KAIST)

AR과 IoT 기술을 기반으로 한 건물 화재 모니터링 및 탈출 내비게이션 시스템 (특별호)
왕문도, 이승용 (동국대), 박상훈 (서강대), 윤승현 (동국대)

논문 발표: 애니메이션/시뮬레이션2

7월 11일 10:25~11:40

멀티 터치 장치를 통한 캐릭터의 이동 동작 제어 47
박근태, 이지원, 이윤상 (한양대)

단일 2D RGB 영상을 이용한 보행주기 탐지 네트워크 49
김대용, 이연수, 김유겸 (아주대), 신정환 (서울대), 유리 (아주대)

상호작용 중심 환경 표현을 통한 데이터 기반 모션 리타게팅 프레임워크 51
이예원 (KAIST), 진태일 (KIST), 이성희 (KAIST)

유체-옷감의 상호작용에 의한 파동 난류, 확산, 주름을 표현하는 GPU 프레임워크 53
이주용, 박은수, 김종현, 박인규 (인하대)

물리기반 캐릭터의 축구 드리블 스킬 학습 55
김민수, 이윤상 (한양대)

논문 발표: 가상/증강현실1

7월 11일 10:25~11:40

확장현실 협업 환경을 위한 장면 합성 프레임워크 설계 58
나기리, 김진모 (한성대)

메타 퀘스트 프로를 활용한 혼합현실과 가상현실 기반의 몰입형 콘텐츠 제작 및 경험에 관한 연구 (특별호)
김종선, 공수민, 장문수, 김진모 (한성대)

동적 환경 적용을 위한 방향전환보행 전략의 선택적 전환 60
이호정, 전상빈 (연세대), 조용훈 (고려대), 이인권 (연세대)

HMD를 이용한 위치 기반 시공간 연동 확장현실 (특별호)
박아연, 배상준, 김미진, 이우현, 김동호 (숭실대)

색채 심리를 활용한 VR 환경에서의 스트레스 완화 프로그램 개발 62
김은지, 이채현, 권영남, 천애지, 방기석, 김선정 (한림대)

논문 발표: 가상/증강현실2

7월 12일 09:00~10:30

가상 현실 기반의 교육 목적 콘텐츠에 대한 연구 – VR 키오스크 시뮬레이션 (특별호)
이준섭, 임병일, 한지영, 이택희 (한국공학대)

A Study on Implementation of Hand Gesture in Virtual Reality (특별호)
조범준 (상명대), 김성기 (조선대)

Continuous Prediction of Pointing Target in Virtual Reality 65
정충호, 이성희 (KAIST)

몰입형 VR 환경에서 가상 반려동물과 상호작용에 관한 패시브 햅틱 요소의 영향 분석 (특별호)
김동근, 조동식 (울산대)

안구 운동을 위한 DDA 유도 VR 게임이 몰입 및 안구 피로에 미치는 영향 67
김현진, 김영조, 이영성, 김서원 (한동대), 임재호 (덱스터스튜디오), 한다성 (한동대)

메타 퀘스트 프로를 활용한 혼합현실 기반 가상 전기회로 실습교육 콘텐츠 제작 (특별호)
공수민, 김종선, 정구현, 노경빈, 박에스더, 조윤식, 김진모 (한성대)

논문 발표: 이미지/비디오

7월 12일 09:00~10:30

디퓨전 오토인코더의 시선 조작 데이터 증강을 통한 시선 추적 (특별호)
문강륜 (성균관대), 김영한 (가천대), 박용준, 김용규 (비주얼캠프)

레퍼런스 착용 스타일을 반영하는 가상 착용 이미지 합성 70
박순찬 (ETRI), 박진아 (KAIST)

RNA: ROI 기반 아틀라스 신경망을 사용한 동영상 편집 72
이재경, 김건웅, 조성현 (POSTECH)

적응형 블러 기반 비디오의 수평적 확장 여부 판별 네트워크 (특별호)
김민선 (KAIST), 서창욱 (Anigma), 윤현호, 노준용 (KAIST)

다단계 미세구조 구역화 네트워크를 활용한 3D 의료영상 구역화 74
김자연, 김지온, 신병석 (인하대)

엣지 기반 초해상도 이미지 혼합 기법 76
김동윤, 나재호 (상명대)

논문 발표: 그래픽스 응용

7월 12일 10:40~11:55

Wishbowl: 버추얼 밴드 아이돌 Verse’day 뮤직비디오 및 몰입형 인터렉티브 공연 제작 사례 연구 (특별호)
이세빈, 김경진, 김다예, 이정진 (숭실대)

시네마픽: 생성형 AI 기반 영화 컨셉 포토부스 시스템 (특별호)
정석현, 임승규, 이정진 (숭실대)

다중 피드백을 지원하는 몰입형 스마트 밸런스 보드 (특별호)
이승용, 이선호, 박준성, 신민철, 윤승현 (동국대)

동적 계획법을 통한 빠르고 직관적인 실내 장면 합성 79
조윤식, 김진모 (한성대)

퍼지 추론 시스템을 이용한 2차원 보행 영상 기반의 발 접촉 판별 네트워크 학습 81
이연수 (아주대), 신정환 (서울대), 윤기범 (디딤), 유리 (아주대)

논문 발표: 강화학습

7월 12일 10:40~11:55

IB LBM 시뮬레이션을 이용한 물고기 동작 정책 학습 84
최완욱, 이윤상 (한양대)

자율주행 차량 시뮬레이션에서의 강화학습을 위한 상태표현 성능 비교 (특별호)
안지환, 권태수 (한양대)

물리 환경 속 가상의 탄도적 스턴트 연기 동작 생성 86
김민석, 서원정 (서울대), 이성희 (KAIST), 원정담 (서울대)

강화학습을 이용한 나비의 비행 동작 생성 88
정은호 (한양대), 장이권 (이화여대), 이윤상 (한양대)

포스터 발표

7월 11일 17:30~18:30

XR 환경에서 사용자-아바타 상호작용을 위한 음성 기반 감정인식 방법 91
이민정, 신영훈, 이미란 (대구대)

사용자 악력 근활성이 반영된 VR 인터페이스의 개선 94
신영훈, 이민정, 이미란 (대구대)

깊이 감독 Dynamic NeRF: 제한된 카메라를 이용한 동적 장면 재구성 97
김규민, 박동하 (인하대), 이채은 (한양대)

위치 정보 인코딩 기반 ISP 신경망 성능 개선 (특별호)
김대연, 김우혁, 조성현 (POSTECH)

이벤트 카메라를 이용한 동적 장면에 대한 편광 반사 특성 계측 99
Ryota Maeda, 백승환 (POSTECH)

개인별 속도차를 반영한 물리 기반 군중 탈출 시뮬레이션 101
김가을, 나경민, 이성호, 김나영, 최명걸 (가톨릭대)

GAIA: 문제 해결 전략과 감정 조절 전략 제시를 통합한 게임 AI 어시스턴트 서비스 프레임워크 103
엄가람, 박은별, 채지훈, 도영임 (KAIST)

Bullet Sign: 익스트랙션 슈터 멀티게임 개발 105
윤세진, 김준호, 조인호, 최예찬, 이강윤, 김선정 (한림대)

탄성체 시뮬레이션을 위한 모멘텀 보존 요소 반전 완화 기법 107
정희조 (고려대), 김승욱 (한국외대), 이재현 (고려대), 엄기원 (텔레콤파리대학), 기민형, 한정현 (고려대)

스토리텔링 기반 장면을 이해하는 다중 인간 동작 생성 109
임동근, 배진석, 황인우, 김영민 (서울대)

캡쳐된 전신 동작에 대한 손가락 동작 생성 기술 111
이지성, 김영빈, 김윤섭, 이지나 (한동대), 임재호 (덱스터스튜디오), 한다성 (한동대)

레이-트레이싱 코어를 활용한 3차원 표면 재구성 가속화 113
김영우, 김덕수 (코리아텍)

Image-to-3D 기술을 활용한 반려동물 아바타 생성 115
황수미, 최준서, 송오영 (세종대)

실 수업내 협력 학습 연구를 위한 카메라 기반 모션 데이터 추출 방법론에 관한 예비 연구 117
조준형, 조영욱, 배정은, 신윤희, 김광욱 (한양대)

CLIP의 Value Embedding을 이용한 개방 어휘 환경 3차원 세분화 119
김상민, 이준호 (서울대), 이용현 (고등과학원), 김영민 (서울대)

과수원 자율 주행을 위한 과수 줄 인식 및 2차원 지도 생성 방법 (특별호)
윤호영, 김덕수 (코리아텍)

부정적 감정 해소를 위한 야외활동 유도 게임 애플리케이션 개발 연구 121
전아름, 권민성 (고려대)

생성형 AI를 활용한 공연예술 씬 디자인 시뮬레이션 도구 123
김명은, 박은별, 윤세린, 안재홍 (KAIST)

깊이를 색상으로써 활용한 NeRF 렌더링 125
임석현, 김민성, 이성길 (성균관대)

다중 FPGA 병렬 렌더링을 활용한 실시간 레이 트레이싱 가속화 127
한윤호, 김지영, 박우찬 (세종대)

XR 서비스를 위한 대용량 3D 도면 전처리 기술 129
서명배, 박형진 (KICT), 신승엽 (KAIST)

스케치와 점유 지점을 활용한 2차원 솔리드 경계표현법 전환 알고리즘 131
이민기, 장동수, 김영민 (서울대)

특수 목적의 웹 CAD 시스템을 위한 데이터 변환 도구의 구현 133
권초염, 김규환, 이환용 (아주대)

자유형상 곡선에 기반한 3D 텍스트 메쉬의 변형 135
권성현, 정현석, 윤승현 (동국대)

사람의 얼굴을 이용한 동물 3D 모델링 생성 서비스 137
이종혁, 소병욱, 조현태, 이준수, 백지웅, 최명걸 (가톨릭대)

3차원 공간 속 로봇 에이전트 표현을 위한 혼합 신경장 139
민철희 (서울대), 문성빈 (고려대), 김영민 (서울대)

의류의 겉감과 안감의 3D 가우시안 스플래팅 복원 방법 141
정문경, 이도해, 이인권 (연세대)

Quality Distribution Function for 3D Gaussian Splatting 143
최민서, 이도해, 이인권 (연세대)

방향전환보행 사용자를 위한 가상환경의 실시간 리셋 예측 145
전수림, 이호정, 이인권 (연세대)

GEMINI 연동 가상 에이전트를 이용한 요리 교육 VR 프로그램 147
이혜은, 이바다, 이채현, 김소희, 김선정 (한림대)

가상현실을 활용한 동물 해부 교육 콘텐츠 및 커리큘럼에 관한 연구 149
장문수, 오병택, 백종훈, 변다솔, 김진모 (한성대)

VR 사용자 경험 강화를 위한 장착형 후각 장치 개선 151
조인호, 박시연, 김선정 (한림대)

XR 환경 특수효과 제어를 위한 멀티모달 인터페이스 153
최종민 (세종대), 김민채 (고려대), 송오영 (세종대)

F-RDW: 미래 위치 예측을 통한 방향 전환 보행 155
전상빈, 정재호, 박진형, 이인권 (연세대)

무요소법을 활용한 KSTAR 중성입자빔가열 유발 플라즈마 대면 요소 열속 해석 및 가시화 157
문태욱, 윤의성 (UNIST)

ScanToSafeGuard: 공간 객체 인식 AI 기술을 활용한 건설 현장의 안전 감시 시스템 160
신승엽 (KAIST), 박형진, 서명배 (KICT), 우운택 (KAIST)

초 청 강 연

초 청 강 연

클로버추얼패션의 제품 개발 방법

오승우 대표, 클로버추얼패션

강연 내용

클로버추얼패션이 어떻게 패션과 컴퓨터그래픽 산업에서 패션 디자인 프로세스를 혁신시켜왔는지에 대해 얘기하려고 합니다. 기술 개발 자체보다는 제품 개발을 위한 효율적인 방식에 대해 얘기합니다. 클로버추얼패션의 주 목표는 사용자와 함께 성장하는 것입니다. 이를 위해 다양한 시도를 하면서 효과적인 제품 개발 방법을 발전시켜 왔습니다. 그것은 최근 사업이나 서비스 개발에 자주 쓰이는 Agile 방법론, Lean Start-up 등과 유사하지만 재미있게도 매우 공학적인 방법입니다. 본 발표에서는 이 방법에 기반한 제품 개발의 새로운 시각과 다양한 best practice에 대해 설명합니다. 이를 통해 보다 나은 소프트웨어, 서비스, 기술, 디자인, 사업 개발이 가능해 지기를 희망합니다.

오승우는 클로버추얼패션의 창업자 중 한명입니다. CLO와 Marvelous Designer 라는 의상 디자인 소프트웨어를 개발해 왔고, 현재는 CLO-SET이라는 3차원 커뮤니케이션 플랫폼 개발을 리드하고 있습니다. 창업 전에 KAIST 전산학과에서 학사, 석사, 박사 학위를 받았습니다.

강연자 이력

1995-2000: KAIST 전산학과 학사졸업

2000-2002: KAIST 전자전산학과 전산학전공 석사졸업

2002-2007: KAIST 전자전산학과 전산학전공 박사졸업

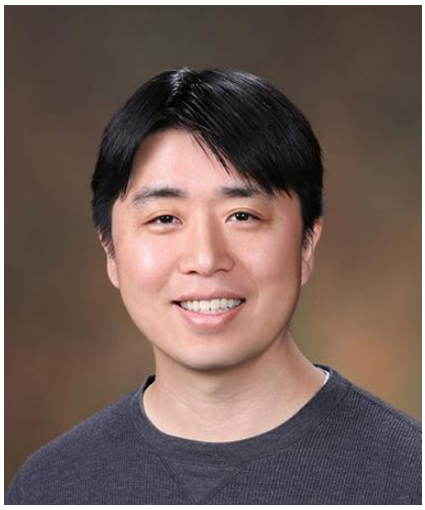
2007-2008: KAIST 엔터테인먼트공학연구소 및 문화기술연구센터 선임연구원

2009-: ㈜ 클로 버추얼 패션 대표

초 청 강 연

인공지능 시대의 음악 창작과 연주

남주한 교수, KAIST



강연 내용

우리가 음악을 창작하고 연주하는 방식은 시대에 따라 끊임없이 진화해 왔다. 특히, 새로운 기술의 등장은 예술가들의 창의력과 감정 표현을 확장시켜 새로운 장르의 음악과 공연 무대를 만들어내고, 대중 음악을 거대 산업으로 발전하는데 중추적인 역할을 해왔다. 최근 급격히 발전하고 있는 인공지능은 인간을 대신 하는 다양한 지능적 음악 기술을 통해 혁신적 변화를 기대하게 하면서도, 기존 창작 및 유통 질서 파괴에 대한 여러 가지 우려도 불러일으키고 있다. 본 강연에는, 오디오, 텍스트, 악보, 비디오 등 다양한 데이터 사이의 관계를 추론하는 멀티모달 인공지능 기술을 관점에서 음악 인공지능 분야의 최신 기술 동향을 소개한다. 또한 이러한 기술이 인간의 음악 창작과 연주에 미치는 영향, 윤리적 및 법적 이슈, 그리고 미래의 전망에 대해도 논의하고자 합니다.

강연자 이력

1994-1998:	전기공학 학사, 서울대학교
2001-2006:	소프트웨어개발자, 영창악기 전자악기연구소
2009-2010:	Electrical Engineering M.S., Stanford University
2006-2013:	Music Ph.D., Stanford University
2012-2014:	Staff Research Engineer, Qualcomm Technologies
2014-:	조교수, 부교수, KAIST 문화기술대학원

초 청 강 연

햅틱 효과 저작
(Authoring of Haptic Effects)

최승문 교수, POSTECH

강연 내용

햅틱스는 1990년대 후반에 출현한 후 약 30년 간의 연구를 통하여 많은 발전을 이루어 왔다. 본 강연에서는 컴퓨터 그래픽스를 전공하는 분들을 위해 먼저 햅틱스에 대한 소개 및 현황을 간단하게 알아보고자 한다. 그 다음에는 햅틱 효과 저작 기술에 대해 기본적인 내용과 강연자가 연구해 온 결과를 중심으로 소개한다. 햅틱 효과 저작 기술이란 다양한 상황에서 사용자에게 적절한 햅틱 효과를 제공하기 위하여 햅틱 장치에 보낼 명령을 컴퓨터를 사용하여 계산하는 것을 말한다. 전통적인 그래픽 렌더링과 대응되는 햅틱 렌더링, 인간-컴퓨터 상호작용을 위한 햅틱 효과 수동 설계, 멀티미디어의 다중감각 확장을 위한 햅틱 효과 자동 추출 및 계산 등을 포괄한다. 근래에 빠르게 발전하고 있는 메타버스에서 햅틱 효과 저작에 대해서도 간단하게 논의한다.

강연자 이력

1991-1995:	Control and Instrumentation Engineering B.S., Seoul National University
1995-1997:	Control and Instrumentation Engineering M.S., Seoul National University
1998-2003:	Electrical and Computer Engineering Ph.D., Purdue University
2015-:	Professor, Department of Computer Science and Engineering, POSTECH
2024-:	Department Head, Department of Computer Science and Engineering, POSTECH

창해 신진 연구자상 발표

2024 창해 신진 연구자상 후보

이신영 (Shinyoung Yi)

syyi@vclab.kaist.ac.kr

이신영(Shinyoung Yi)은 컴퓨터 그래픽스를 연구하는 KAIST의 박사과정생이다. 학부 시절 수학과 물리학을 복수전공하였으며, 빛의 속도가 유한함에 따른 효과가 고려된 전이(transient) 상태 렌더링, 빛의 편광 정보가 고려된 편광 렌더링 등 빛의 다양한 특성을 반영한 렌더링 이론 및 기술 개발에 주된 관심을 보이고 있다. 그의 연구 중 하나인 "미분가능 전이 상태 렌더링(differentiable transient rendering)"에서는 빛의 전파 시간이라는 추가적이라는 의존성이 계산이 포함되는 가운데에서 비편향된 미분값을 계산하였고, 이의 역렌더링 문제에 대한 응용 가능성을 제시하였다. 또한 최근에는 빛의 세기를 좌표계에 대한 특수한 의존성을 지닌 스토크스 벡터로 표현되는 편광 렌더링에서 새로운 진동수 기반 방법론을 제시하여 편광 환경광이 반영된 최초의 실시간 렌더링 기술을 선보였다.

그는 연구 실적 이외에도 컴퓨터 그래픽스 교육, 시각화, 각 논문에 산재된 지식들과 표기법의 통합 등 다양한 방면으로도 학계에 기여하고자 힘쓰고 있다.

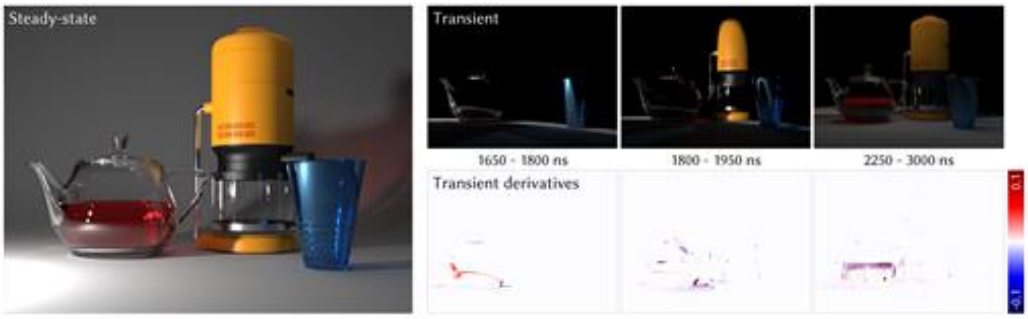


그림 1. 좌측에 표현된 장면에 대한 전이 상태 렌더링의 결과물(우측 상단)의 미분값인 미분가능 전이 상태 렌더링(우측 하단, 제안된 방법)의 결과물

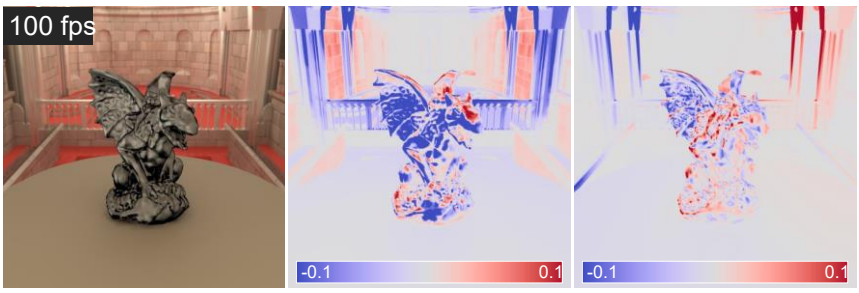


그림 2. 편광된 환경광이 고려된 실시간 편광 렌더링의 결과물

2024 창해 신진 연구자상 후보

김효민 (Hyomin Kim)

min00001@postech.ac.kr

김효민은 2024년 2월에 포항공과대학교에서 컴퓨터공학 박사 학위를 받고 현재 졸업 연구실에서 박사후 연구원으로 재직중이다. 세부 전공은 컴퓨터 그래픽스의 한 분야인 3차원 기하, 텍스처 복원으로, 이미지, 점군, 메시 등의 입력 데이터를 처리하여 정확하고 최적화된 모델로 변환하는 방법에 대해 연구하였다.

김효민이 중점적으로 수행한 연구는 한 대의 색상-깊이 카메라가 있는 상황에서 실제 사람을 촬영하고, 이를 템플릿 기반의 3차원 아바타로 복원하는 기술이다. 각 연구에서는, 피사체의 일부분만 촬영된 데이터와 템플릿 간의 매핑을 추정하고 [ICCV 2021], 부분적 데이터를 최대한 활용하여 한 시점에서 보이지 않는 영역의 텍스처 및 기하를 복원하는 [CGF 2021, TOG 2022] 방법을 제안했으며, 특히 기하 복원 연구 [TOG 2022] 에서는 입력 기하 디테일을 매우 효과적으로 보존하는 새로운 방식을 제시했다. 또한 노멀맵을 왜곡 없이 적분하는 [CVPR 2024] 방법을 제안하여, 일반적으로 깊이 이미지 대비 고해상도인 색상 이미지를 기하 복원에 추가 활용할 수 있게 하였다. 김효민은 학위 과정 동안 총 8편의 논문에 참여, 그 중 5편을 1저자로 발표하며 연구의 독창성을 인정받고 3차원 기하, 텍스처 복원 분야의 발전에 기여하였다.

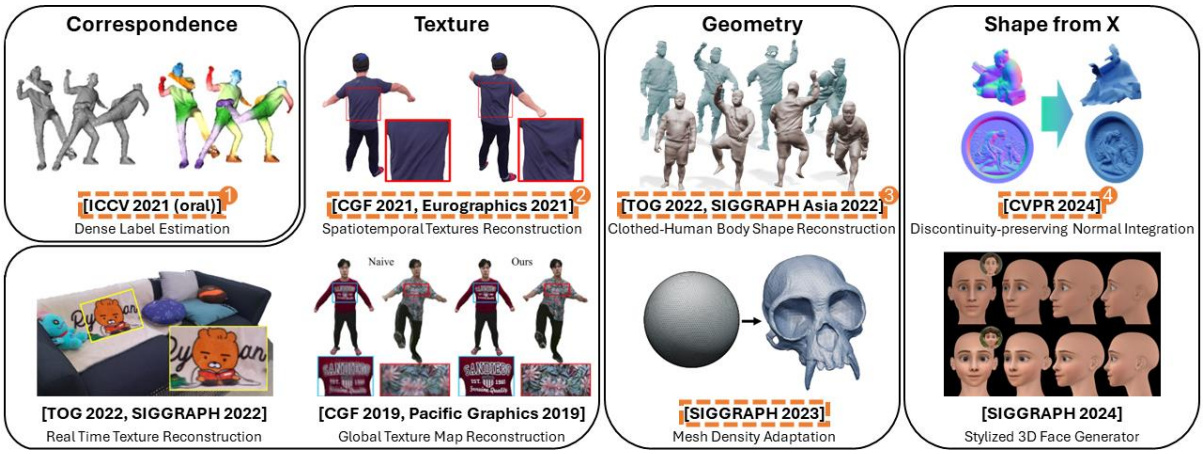


그림 1. 관심 분야 세부 분류 및 참여 논문 목록 (주황 점선 박스: 1저자 논문)

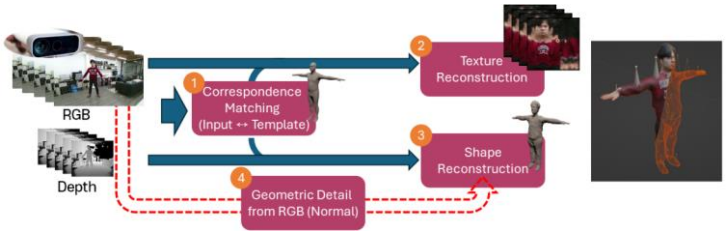


그림 2. 학위논문 "단일 색상-깊이 카메라를 이용한 3차원 사람 복원" 개요도

2024 창해 신진 연구자상 후보

장덕경 (Jang, Deok-Kyeong)

dk.jang1014@gmail.com

KAIST 이성희 교수님 아래에서 2023년 컴퓨터 그래픽스 관련 박사학위를 취득하였으며, 현재는 연구경험을 바탕으로 MOVIN Inc. ((주) 무빈) 을 창업하여 CTO로 재직 중입니다.

연구 목표는 딥러닝 기반 컴퓨터 그래픽스 및 AR/VR 시스템에서 디지털 캐릭터 모션의 품질을 향상시키는 것입니다. 현재는 실시간 3D 전신 모션 캡처 및 캐릭터 모션 특성화에 중점을 두고 연구개발을 진행하고 있습니다.

논문 실적

- MOCHA: Real-Time Motion Characterization via Context Matching (Deok-Kyeong Jang, Yuting Ye, Jungdam Won, Sung-Hee Lee / ACM SIGGRAPH ASIA 2023)
- MOVIN: Real-time Motion Capture using a Single LiDAR (Deok-Kyeong Jang, Dongseok Yang, Deok-Yun Jang, Byeoli Choi, Sung-Hee Lee / Computer Graphics Forum, Proc. Pacific Graphics 2023)
- Motion Puzzle: Arbitrary Motion Style Transfer by Body Part (Deok-Kyeong Jang, Soomin Park, Sung-Hee Lee / ACM Transactions on Graphics, Proc. SIGGRAPH 2022)
- Diverse Motion Stylization for Multiple Style Domains via Spatial-Temporal Graph-based Generative Model (Soomin Park, Deok-Kyeong Jang, Sung-Hee Lee / Proceedings of the ACM on Computer Graphics and Interactive Techniques, Proc. SCA 2021)
- Constructing Human Motion Manifold with Sequential Networks (Deok-Kyeong Jang, Sung-Hee Lee / Computer Graphics Forum, Proc. Eurographics 2020)
- Regression-Based Landmark Detection on Dynamic Human Models (Deok-Kyeong Jang, Sung-Hee Lee / Computer Graphics Forum, Proc. Pacific Graphics 2017)

연구 경험

- Postdoctoral Researcher (2023.03 – 2023.10, KAIST): LiDAR 기반의 실시간 3D 모션 캡처 및 모션 합성에 대한 연구 (지도교수: 이성희 교수)
- Research Science Intern (2022.05 – 2022.10, Meta Reality Lab, Redmond, USA): 실시간 모션 특성화, 모션스타일 전달향상 및 다양한 입력 센서를 통한 리타게팅에 대한 연구 (Advisor: Yuting Ye)

수상 실적

- 2019 한국 그래픽스 학회 우수 논문상: 순차 네트워크 기반 인간 모션 매니폴드 구성

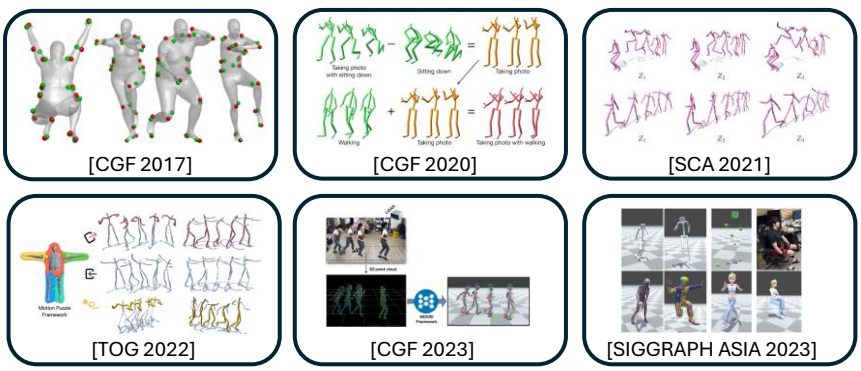


그림 1. 논문 실적

석사 논문상 후보자 발표

2024 석사 논문상 후보

최은수 (Eunsue Choi)

ches7283@postech.ac.kr

최은수는 포항공과대학교 (POSTECH) 컴퓨터공학과 석사과정 학생으로서 컴퓨터 그래픽스 연구실에서 백승환 교수의 지도를 받고 있다. 주요 연구 분야는 빛의 파동적 성질로부터 기인하는 광학 현상과 수학적 최적화 기술을 이용한 computational imaging system 개발과 Metasurface와 같은 nano-optic end-to-end optimization이다. 그는 2022년도에 포항공과대학교로부터 Magna Cum Laude 로 컴퓨터공학 학사를 취득하였다. 그가 석사 과정 중 작성한 2편의 논문은 각각 Nature Photonics에 게재되고 CVPR 2024 highlight로 선정되었다. 그는 POSTECH 나노스케일 포토닉스 & 통합생산 연구실과의 공동연구를 통해 2024년 POSTECH 학제간 공동연구 우수팀으로 선정되었다.

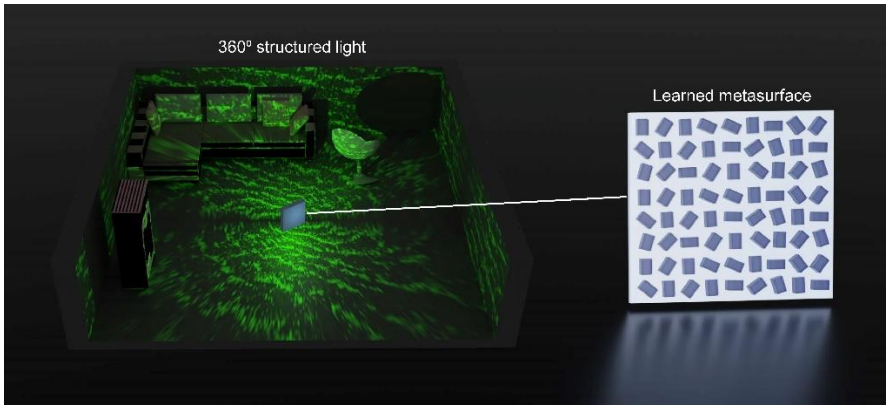


그림 1. 360° structured light with learned metasurfaces [Nature Photonics], 360° structured light 기술은 자율주행자동차, robotics, AR/VR 디바이스 등 다양한 응용 분야에 사용할 수 있다.

그림 2. Spectro-polarimetric Real-world Dataset [CVPR 2024], Object recognition, shape reconstruction, 및 여러 biological application등에 활용될 수 있다.

2024 석사 논문상 후보

류누리 (Nuri Ryu)

ryunuri@postech.ac.kr

류누리 학생(Nuri Ryu, 1999년 11월 4일 ~)은 포항공과대학교 인공지능대학원 컴퓨터 그래픽스 연구실의 석·박사 통합 과정 학생이다. 2022년 2월 포항공과대학교에서 컴퓨터공학 학사 학위를 취득했다.

류누리 학생의 주 연구 분야는 생성 모델이며, 고품질 삼차원 데이터를 생성하는 모델을 개발하는 데 집중하고 있다. 특히, 오랜 기간 난제로 여겨진 단일 영상에서 3차원 물체를 복원하는 문제의 해결에 기여한 바가 있다. 구체적으로, ACM SIGGRAPH Asia 2023에 발표된 그의 논문 "360 Reconstruction from a Single Image Using Space Carved Outpainting"은 단일 영상으로부터 물체의 다각도 영상을 생성하여 3차원 모델을 재구성하는 방법론을 제시했다 (그림 1). 이는 단일 영상에서 3차원 복원을 하는데 주류를 차지한 풀이 방식인 Score Distillation Sampling과는 전혀 다른 새로운 방안이었으며, 이를 통해 당시 학계 최고 수준 복원 방식들에 비해 뛰어난 품질의 복원을 달성했다 (그림 2).

류누리 학생은 삼차원 생성 분야에서 활발히 연구를 이어가고 있다. 최근에는 앞서 언급한 논문의 후속 연구로 단일 영상 3차원 복원 문제를 여러 소문제로 분할하고, 각 소문제에 알맞은 사전 분포를 사용하는 방식을 새로이 연구하여 기존의 논문에 비해 생성 품질과 속도를 월등히 발전시킨 모델을 개발하였으며, 본 연구의 결과는 현재 ACM SIGGRAPH Asia에 제출되어 심사 중이다.

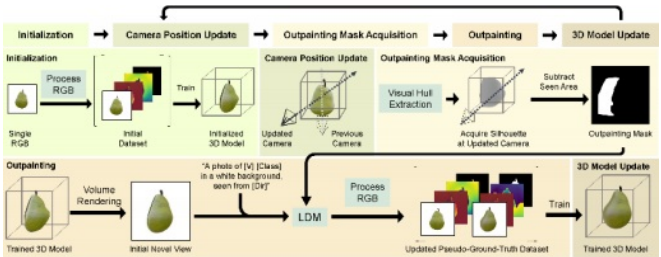


그림 1. 단일 영상으로부터 삼차원 물체를 복원하는 기법의 구조도

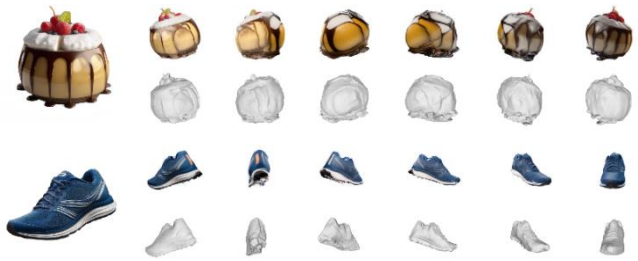


그림 2. 좌측의 입력 영상을 삼차원 복원하여 우측에 시각화한 예시

2024 석사 논문상 후보

윤관 (Kwan Yun)

yunandy@kaist.ac.kr

윤관 (영어: Kwan Yun, 1996년 11월 18일 ~)은 문화기술대학원 비주얼 미디어 연구실의 석사과정 학생으로, 그래픽스 응용 분야에서 생성 모델의 활용이라는 주제로 연구를 진행하고 있다.

그래픽스 분야에 활용되는 다양한 딥러닝 기반 접근들의 문제점중 하나인 데이터 부족 문제, 더 나아가 데이터가 있는 경우에만 적용할 수 있었던 접근법들을 사전 학습된 네트워크를 활용해 사용자가 원하는 방식대로 활용할 수 있도록 하는 연구를 진행하였다. 그림 1의 LeGO 연구의 경우 메쉬 변형을 통해 스타일화되고 애니메이션 가능한 얼굴 생성 이라는 그래픽스 응용분야에서 중요한 주제를 다루고 있고 특히 기존 딥러닝 기반 접근 방식들이 해결하지 못했던 텍스쳐 적용과 애니메이션에 필수적인 일관된 메쉬구조 문제를 하나의 예시 데이터만으로 해결하여 CVPR 논문중 Highlights에 선정되었다. 또한 그림 2의 StyleSketch 연구는 사전 학습된 생성모델의 특징맵을 이용해 얼굴 스케치를 추출하거나 스케치를 에디팅하는 연구를 진행하였으며, 생성 모델을 처음으로 스케치 분야에 적용하여 매우 적은 데이터로도 원하는 스타일로 스케치 추출이 가능함을 처음으로 보였다.

그림 1. 하나의 데이터로 학습하여 생성한 애니메이션 가능한 얼굴 메쉬 (CVPR 2024)

그림 2. 적은 데이터로 학습하여 추출한 얼굴 스케치 (EUROGRAPHICS 2024)

우수 국제 학술대회 논문 발표

애니메이션1

ClothCombo: 여러 겹의 3D 의류 가상피팅을 위한 의류 간 상호작용 모델링

이도해[0], 강현, 이인권
연세대학교 컴퓨터과학과
dlehgo1414@gmail.com, kdkh0508@yonsei.ac.kr, iklee@yonsei.ac.kr

ClothCombo: Modeling Inter-Cloth Interaction for 3D Virtual Try-on of Multi-Layered Garments

Dohae Lee[0], Hyun Kang, In-Kwon Lee
Dept. of Computer Science, Yonsei University

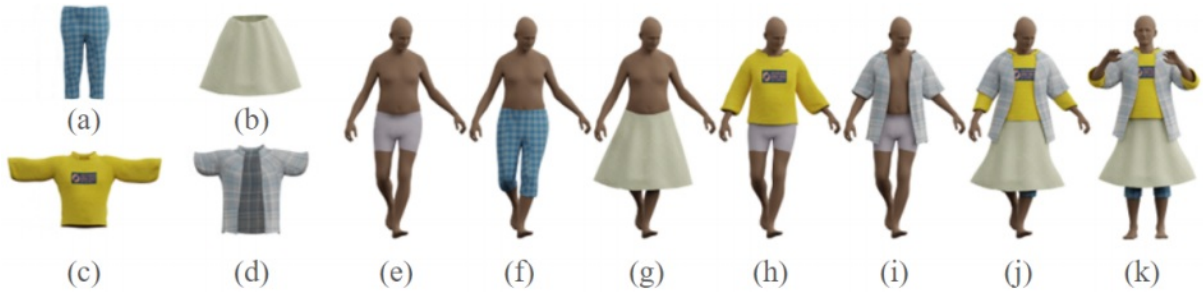


그림 1: (a-d) 대상 의류, (e) 대상 체형, (f-i) 한벌씩 착용한 결과, (j) 겹쳐 착용한 결과, (k) 다른 자세에서의 결과.

요약

본 논문은 임의의 3D 의류를 임의의 3D 아바타에 여러 겹 겹쳐 착용하는 AI 기반 가상피팅 시뮬레이션 기법(ClothCombo)을 제안한다. 기존 3D 가상피팅 방법은 대상 의류 하나 당 하나의 신경망 모델 학습이 필요하여 효율적이지 못하고, 두벌 이상의 의류 여러 개를 겹쳐 착용하는 것이 불가능하였다. 제안하는 기법은 임의의 토폴로지를 가지는 3D 의류를 동일한 잠재공간에 임베딩하여 의류 종류에 상관없이 모든 의류에 일반적으로 적용 가능하며, 의류 간의 상호작용을 모델링하여 여러 벌의 의류를 임의의 체형에 겹쳐 착용시키는 것이 가능하다. ClothCombo 모델은 ground-truth 데이터 없이 물리 기반 손실함수를 활용하여 자기지도 방식으로 학습되며, 여러 종류의 실험을 통해 기존 방법 대비 우수한 성능으로 여러 겹의 가상피팅이 가능함을 보였다. 또한, 가상 데이터 뿐만 아니라 현실 의류 스캔 데이터를 이용한 실험을 통해 제안한 기법이 현실 의류의 가상피팅에도 적용 가능함을 보였다.

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract)으로서, 원본 논문은 SIGGRAPH Asia 2023에 발표 및 ACM Transactions on Graphics 42(6). 2023에 게재되었음.
* 이 연구는 정부 (과학기술정보통신부)의 재원으로 한국연구재단 (No. RS-2024-00348094) 및 한국전파진흥협회 (No. RNIX20230200)의 지원으로 수행되었음.

1. 서론

의류 시뮬레이션은 현실 의류의 움직임을 모방하여 컴퓨터 프로그램으로 구현하는 것으로, 영화, 게임 등의 콘텐츠 제작에 필수적으로 활용되어 컴퓨터 그래픽스 분야에서 오랫동안 중요한 주제로 연구되고 있다. 그 세부 분야로 현실의 의류를 현실의 사람이 착용한 모습을 예측하는 가상피팅 기술 또한 활발히 연구되고 있는데, 사람의 체형, 자세, 그리고 의류의 종류와 조합의 다양성이 매우 커 오랜 난제로 남아있다. 최근, 인공지능을 활용하여 임의의 체형과 자세의 가상 아바타에 의류를 착용하는 기법들이 제안되었으나[1], 의류 마다 신경망 모델을 학습해야 하는 한계가 존재한다. 또한, 가상피팅 응용에서 여러 겹의 의류를 레이어링 하는 것이 중요한데, 의류 간의 상호작용은 매우 복잡하여 연구의 진척도가 미흡하다. 몇몇 연구에서 여러 겹의 의류를 겹쳐 착용하는 방법을 제안한 바 있으나[2], 상의-바지 등 제한된 상황에서만 적용 가능하거나, 의류마다 별도의 최적화 과정이 필요하고 정해진 자세에서만 가능한 한계가 있다[3].

반면, 제안하는 방법은 의류마다 별도 학습이 필요 없이 모든 의류에 광범위하게 적용 가능하며, 다양한 자세 및 체형 위에 여러 겹을 겹쳐 가상 피팅하는 것이 가능하다 (그림 1). 이것은 서로 다른 토폴로지의 3D 의류를 동일한 잠재공간에 임베딩하고, 의류 간의 상호작용을 모델링함으로써 가능하였다. 의류 3D 데이터를 이용한 다양한 실험을 통해 제안한 모델이 종래 기술에 비해

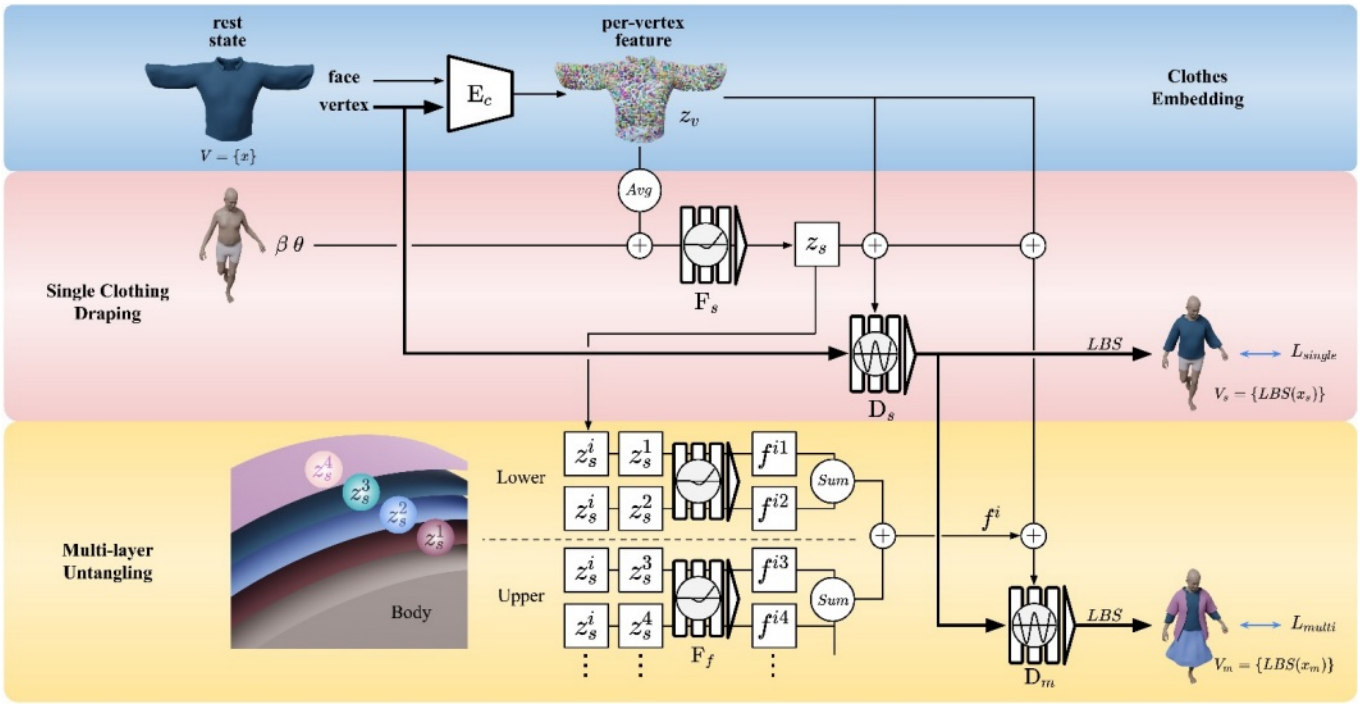


그림 2: ClothCombo 모델 개요도

더 나은 성능을 보임을 확인할 수 있었고, 최초로 여러 체형 및 자세의 휴먼 아바타에 임의의 의류를 여러 벌 겹쳐 착용하는 것이 가능함을 보였다.

2. 방법

ClothCombo는 세 단계로 구성된다 (그림 2). 첫번째 단계인 Clothes Embedding 단계에서는 mesh의 토폴로지에 robust한 DiffusionNet[4]을 활용하여 의류를 잠재공간에 임베딩 한다. 두번째 단계인 Single Clothing Draping 단계에서는 사람의 체형과 자세에 의한 대상 의류의 변형을 vertex의 displacement로 계산한다. 마지막 단계인 Multi-layer Untangling 단계에서는 의류의 순서에 따른 의류 간 주고받는 힘을 Graph Neural Network 기반의 방법으로 모델링하여, 여러 의류를 겹쳐 착용하였을 때의 모든 레이어 상에 있는 의류의 변형을 계산한다.

ClothCombo를 구성하는 모든 세부 네트워크는 비지도 학습 방식으로 학습되며 물리적 제약조건을 손실함수로 사용한다. 이전 연구에서 제안된 바 있는 strain loss, bending loss, collision loss 등을 활용하였으며, 추가적으로 여러 겹의 의류 간 상호작용을 모델링하기 위해 의류 사이의 충돌을 제어하는 multi-layer loss 등을 제안하여 사용하였다.

3. 실험 및 결과

제안한 모델은 그림 1에서처럼 임의의 체형과 자세의 인체 (그림 1, e) 위에 임의의 의류 (그림 1,f-i)를 착용시킬 수 있으며, 그림 1의 (j-k)에서와 같이 여러 벌의

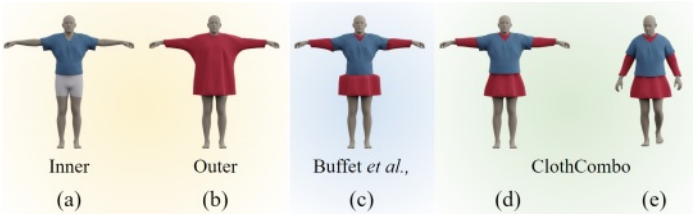


그림 3: 비교모델(c)과 제안한 모델(d-e)의 결과 비교.

의류를 겹쳐 착용시킬 수 있다. ClothCombo는 그림 3에서처럼 기존 기술 (그림 3, c)에 비해 우수한 성능 (그림 3, d-e)으로 의류 간 상호작용을 모델링하여 가상피팅할 수 있었다.

4. 토의 및 결론

본 연구에서는 임의의 의류를 여러 벌 겹쳐서 3D 가상피팅할 수 있는 AI 기반의 방법을 제안하였다. 실험을 통해 기존 기술보다 우수하게 여러 겹 의류들의 가상피팅이 가능함을 보였다. 본 기술은 가상 의류 데이터 뿐만 아니라, 현실 의류 3D 스캔 데이터에도 적용 가능하여 그 응용 가능성이 크다.

참고문헌

[1] Igor Santesteban, et al. SNUG: Self-Supervised Neural Dynamic Garments. IEEE/CVF CVPR, 2022.
[2] Luca De Luigi, et al. DrapeNet: Generating Garments and Draping them with Self-Supervision. ArXiv, 2022.
[3] Igor Santesteban, et al. ULNeF: Untangled Layered Neural Fields for Mix-and-Match Virtual Try-On. NeurIPS, 2022.
[4] Nicholas Sharp, et al. DiffusionNet: Discretization Agnostic Learning on Surfaces, SIGGRPAH 2022.

물리 기반 탁구 애니메이션을 위한 전략과 스킬 학습 방법론

Jiashun Wang[0,a], Jessica Hodgins[a], 원정담[*]
[a]Carnegie Mellon University, [*]서울대학교
[a]{jiashunw, jkh}@cmu.edu, [*]jungdam@imo.snu.ac.kr

Strategy and Skill Learning for Physics-based Table Tennis Animation

Jiashun Wang[0,a], Jessica Hodgins[a], Jungdam Won[*]
[a]Carnegie Mellon University, [*]Seoul National University

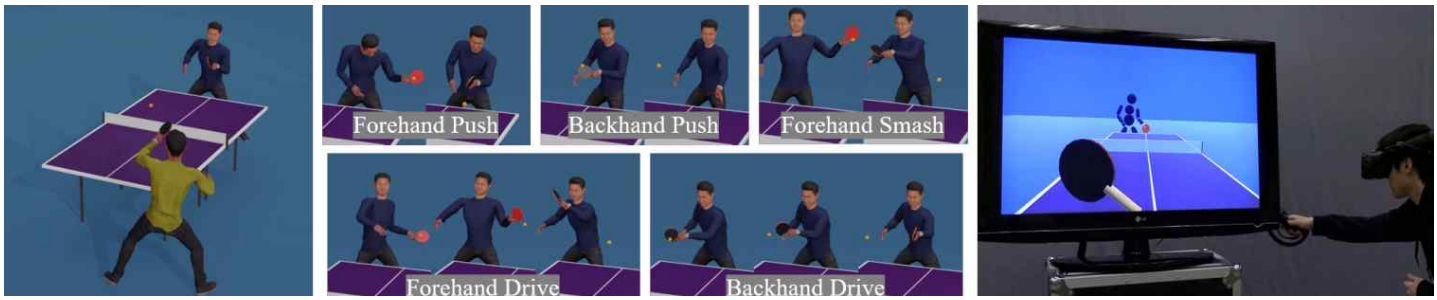


그림 1: 본 연구에서 제안된 제어기를 이용하여 두 에이전트가 탁구 경기를 수행하는 모습. 왼쪽부터 두 에이전트 간의 랠리, 각 스킬의 모습, 인간이 시뮬레이션 된 에이전트와 상호작용하는 모습을 확인할 수 있다.

요약

물리 기반 캐릭터 애니메이션은 딥러닝 기법의 많은 발전으로 인해 공중제비와 같은 현란한 동작이 표현 가능해졌다. 하지만, 탁구 경기와 같이 어려운 목표를 달성하기 위해 다양한 동작을 표현해야 하는 경우 네트워크의 모드 붕괴(mode collapse) 문제로 인해 캐릭터는 제한된 움직임만 표현하고, 좋은 성과로 목표를 달성하지 못한다. 본 연구는 탁구 경기를 사례로, 모드 붕괴를 효과적으로 제거하여 풍부한 동작을 생성하는 심층 강화학습 방법론을 제시한다. 제어기를 전략 단계 제어기와 스킬 단계 제어기로 나누고 각 단계에 맞춘 효과적인 학습 방법을 제시하였다. 또한, 시뮬레이션 상에서 인공지능-인공지능 상호작용과 VR 환경을 사용한 인간-인공지능 상호작용에 대해 제시한 방법론을 검증하였고, 최신의 스킬 기반 동작 생성 방법론 연구들과 본 연구에서 제시한 방법론을 비교, 검증하였다.

1. 서론

최근 물리 기반 캐릭터 애니메이션에 딥러닝 기법을 적용하여 자연스럽고 다양한 동작을 생성하는 연구들이 제시되었다. 이러한 연구들은 캐릭터가 물리 시뮬레이션에서 모션 캡쳐 데이터와 유사한 움직임을 수행하도록 하는 네트워크 학습 방법론을 제시한다. 대표적인 방법으로 적대적 모방 학습(Generative Adversarial Imitation Learning, GAIL) 방법론[1]에서는 분별기(Discriminator)를 사용하여 캐릭터 동작과 모션 캡쳐 데이터 사이의 차이를 줄이는 방향으로 네트워크가 학습된다. 다양한 동작을 캐릭터가 구사하기 위해서는 학습된 동작이 잠재 공간으로 표현되어야 하고, 최근 연구들은 잠재 공간에서 재사용 가능한 스킬을 학습하는 방법에 주목하고 있다[2,3].

스킬 사이에 동작의 차이가 크지 않으면, 데이터들 사이의 차이를 학습하지 못하고 제한된 동작만 표현하는 모드 붕괴가 발생할 수 있다. 운동 경기와 같이 다양한 동작으로 복잡한 문제를 해결해야 하는 상황에서 모드 붕괴가 발생하면, 강화 학습의 탐색이 제대로 이루어지지 않아 주어진 상황에 대해 최적의 캐릭터 동작을 표현하지 못하게 된다.

탁구 경기의 동작은 상대 선수를 비롯하여 환경과 기민한 상호작용이 필요하여 모션 캡쳐 데이터에 명시적으로 나타나 있지 않는 자연스러운 동작을 포함되어야 한다. 따라서 물리 시뮬레이션을 사용하여 동작을 생성하는 것이 필요하다. 또한 각 타법 사이의 차이는 탁구채의 방향 뿐으로 모드 붕괴가 발생하여 캐릭터가 상황에 맞는 적절한 스킬을 사용하지 못할 수 있다. 따라서 탁구 경기를 사례로 본 연구에서 제시한 학습 방법론을 검증한다. 먼저, 전문가 혼합 방식(Mixture-of-Experts, MoE)에서 영감을 받아 다양한 탁구 동작을 풍부하게 생성하는 스킬 제어기를 학습하였다. 스킬 제어기는 모드 붕괴에 효과적으로 대응하고 스킬 사이의 전환을 매끄럽게 표현한다. 다음으로, 추상적인 전략을 학습하는 방법론을 제시하여 인공지능이 경기 상대방과 경쟁과 협동과 같은 다양한 상호작용을 학습할 수 있게 하였다.

2. 학습 방법론

본 연구에서 제안하는 제어기는 각 선수의 상태와 탁구공의 착지 지점과 같이 추상적인 목표를 다루는 전략 단계와 캐릭터 관절 개별의 PD(Proportional-Derivative) 제어 목표를 계산하는 스킬 단계로 나눌 수 있다 (그림2 참조). 스킬 단계의 제어기는 모방 정책, 공 제어 정책, 혼합 정책으로 나뉜다. 전략 단계의 제어기는 전략과 경쟁

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 SIGGRAPH 2024 Conference Paper에 게재 확정 되었음

상호작용을 학습하기 위해 후술할 두 방법론으로 학습된다. 모든 정책은 PPO(Proximal Policy Optimization)을 사용하여 학습하였다.

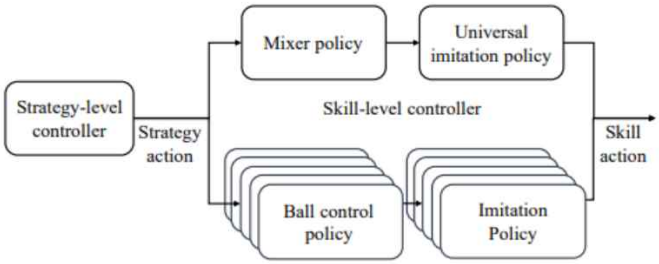


그림 2: 전략 단계와 스킬 단계로 표현한 제어기 구조

2.1 스킬 단계

탁구 모션 캡쳐 데이터를 포핸드 드라이브, 커트, 스매시, 백핸드 드라이브, 커트로 나누어 개별 데이터에 대해 모방 학습을 진행하여 $\pi^i(a^i|s,z^i)$의 다섯 정책($i \in 1,\dots,5$)을 학습하였다. 또한 모든 데이터를 사용하여 통합 정책(π^u)도 학습하였다. 개별 모방 정책은 ASE[2]에서 제시한 방법론을 사용하여 분별기 보상, 잠재 변수와 액션 사이의 상호 정보 보상을 최대화하고 다양성 Loss를 최소화하게 학습하였다.

공 제어 정책은 모방 정책에 입력으로 들어가는 잠재 변수를 출력하는 정책으로 공의 정확한 제어를 담당한다. 이 정책은 탁구채 보상과 공 보상, 스타일 보상을 최대화 하게 학습된다. 탁구채 보상은 캐릭터가 공과 탁구채를 가까이 이동하게 한다. 공 보상은 목표 지점에 탁구공이 도달하도록 하는 보상이다. 스타일 보상은 ASE 학습 과정에서 사용한 보상과 유사하다.

혼합 정책은 앞서 학습된 $\pi^i(a^i|s,z^i)$와 π^u의 출력인 캐릭터 관절 각도에 대한 혼합 가중치를 출력하여 MoE의 Gating Network 역할을 수행한다. 공 제어 정책과 동일한 보상을 사용하여 혼합 정책을 학습하였다.

2.2 전략 단계

전략 단계의 정책은 스킬 단계의 정책에서 사용할 스킬의 종류와 탁구공의 낙하 지점을 출력한다. 경기 비디오에서 추출한 전문가 데모를 이용하여 모방학습을 하였다. 추가로 인공지능 사이의 경기를 이용하여 협동과 경쟁의 상호작용을 학습하였으며, 협동은 인공지능이 친 공을 상대가 성공적으로 반환한 것을, 경쟁은 승리한 결과를 선별하여 학습에 이용하였다. 조건부 VAE(Conditional Variational Auto Encoder, CVAE)를 네트워크로 사용해 학습하였다.

3. 결과

본 연구에서 제안한 방법론은 인공지능-인공지능이 상호작용 환경과 인공지능-인간이 상호작용하는 환경에서 확인되었다. 인공지능-인간의 상호작용 환경은 유니티 엔진을 활용하여 VR 장치에 표시하였고, 물리 시뮬레이션인 아이작 짐(Isaac Gym)에서 제어기를 구동하였다. 구체적으로 VR 제어기의 위치를 탁구채의 위치로 설정하여 인간의 동작을 유니티 엔진에 삽입하였다.

학습 결과는 앞서 제시된 방법으로 학습된 스킬 단계 제어기와 최신 작업인 ASE[2], CASE[3] 그리고 우리의 방법론에서 혼합 정책을 제외한 정책(explicit transition model, ET)을 비교하여 평가하였다. 평가 지표로는 동작의 유려함과 작업 성능으로 나누었고, 동작의 유려함은 분별기 지표와 다양성 지표로 나누었다. 분별기 지표는 분별기를 사용하여 데이터와 제어기 출력 사이의 차이를 평가하는 척도이고, 다양성 지표는 포핸드 드라이브와 커트와 같이 시각적으로는 유사하지만, 기능적으로 다른 스킬이 얼마나 잘 학습되었는지 평가하는 척도이다. 작업 성능은 정확도와 평균 반환 능력으로 나누어 평가하였다.

정확도를 제외한 모든 지표에서 본 연구에서 제시된 방법론이 우수한 성능을 보였다. 정확도는 ET가 가장 우수한 성능을 보였지만, 혼합 정책이 있는 제어기와 성능의 차이가 크지 않았으며, ET는 되돌아 나온 공에 대응하는 능력이 부족했다. 따라서 정확한 스킬의 구사를 위해 혼합 정책이 필요하다는 것을 알 수 있다. 전략 단계 제어기는 강화학습을 사용한 정책과 비교하였고, 협동과 경쟁 정책 모두 본 연구에서 제안한 방법이 우수한 결과를 보였다.

VR 환경의 인간의 동작과 제어기가 이제까지 학습한 시뮬레이션의 데이터 사이에는 미묘한 차이가 있다. 따라서, 인간-인공지능 상호작용 환경에서 제어기를 평가하기 전에 스킬 단계 제어기를 VR 경기 데이터를 이용해 미세 조정하였다. 미세 조정의 결과 경쟁 정책은 10% point 이상의 더 높은 승률을, 협동 정책은 20% 이상의 더 긴 랠리를 인간과 수행할 수 있었다.

4. 결론 및 제언

본 연구에서는 탁구 경기를 물리 기반 캐릭터 애니메이션으로 구현하기 위해 계층적인 제어기를 제안하고, 그 구조가 모드 붕괴를 효과적으로 극복할 수 있음을 확인하였다. 또한 최신 다른 연구와 비교하였을 때, 유려한 동작을 생성하고 좋은 작업 성능을 가진다는 것을 확인하였다.

하지만, 본 연구에서 제안한 방법론의 혼합 정책은 수백 개의 스킬로 확장하기에는 무리가 있고, 강체 시뮬레이션만 사용하여 탁구공의 마그누스 효과와 같은 유체역학적 현상을 표현할 수 없다는 한계를 가지고 있다. 이러한 사항들은 향후 연구에서 보완되어야 할 것이다.

참고문헌

[1] X.B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, AMP: Adversarial motion priors for stylized physics-based character control, *ACM Transactions on Graphics (TOG)*, 40(4):1-15, 2021.

[2] X.B. Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler, ASE: Large-scale reusable adversarial skill embeddings for physically simulated characters, *ACM Transactions On Graphics (TOG)*, 41(4):1-17, 2022.

[3] Z. Dou, X. Chen, Q. Fan, T. Komura, and W. Wang, C· ASE: Learning conditional adversarial skill embeddings for physics-based characters, In *SIGGRAPH Asia 2023 Conference Papers*, pp. 1-11, December 2023.

우수 국제 학술대회 논문 발표

애니메이션2

SAME: 캐릭터골격 비의존적 동작 잠재공간*

이선민[0a], 강태호[a], 박정남[a], 이제희[b,a], 원정담[a]
a: 서울대학교, b: NC Research
{smileeesun, passerby98, namjohn10, jeheel, nonaxis}@ snu.ac.kr

SAME: Skeleton-Agnostic Motion Embedding for Character Animation

Sunmin Lee[0,a], Taeho Kang[a], Jungnam Park[a], Jehee Lee[b,a], Jungdam Won[a]
a: Seoul National University, b: NC Research

요약

컴퓨터 그래픽스 분야에서 모션 데이터를 활용한 심층 인공신경망 기반 연구가 보편화되고 있다. 학습에 활용할 수 있는 모션 데이터의 양이 증가하고 있지만, 데이터셋마다 캐릭터골격이 모두 다르며 이를 통일하기 위해서는 번거로운 비선형 최적화가 필요하다. 본 논문에서는 다양한 캐릭터 애니메이션 작업을 캐릭터의 골격과 무관하게 수행할 수 있는 프레임워크를 제안한다. 연구의 핵심은 모션데이터로부터 골격관련 정보를 분리하고 움직임의 의미를 보존하는 잠재공간을 학습하는 것이다. 캐릭터 비의존적 동작 잠재공간을 학습하는 효과적인 오토인코더 구조와 잠재공간을 통해 모션 분류, 검색, 복원, 실시간 제어 등의 캐릭터 애니메이션 작업을 수행하는 방법을 제안한다.

1. 서론

최근 대량의 데이터를 활용한 대규모 모델 학습이 여러 분야에서 큰 성과를 보이고 있다. 학습에 사용할 수 있는 모션캡처 및 키프레임 애니메이션 데이터의 양이 빠르게 증가하였다. 하지만 데이터셋마다 사용하는 캐릭터가 모두 다르다는 점이 대규모 모델 학습에 걸림돌이 되고 있다. 최근 AMASS[1] 데이터셋은 인체 골격 매개변수 모델 SMPL[2]로 표준화하여 다량의 모션 데이터를 활용한 딥러닝에 크게 기여하였다. 하지만, 인체에서 벗어나거나 SMPL과 다른 토폴로지 골격에 적용할 수 없다. 본 연구는 골격 구조와 무관하게 모션 데이터를 수집, 학습, 및 적용할 수 있는 유연한 프레임워크를 제안한다.

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 ACM SIGGRAPH Asia 2023 Conference Proceedings 에 게재되었음.
* 본 연구는 NCSOFT AI Center under Project Number NCSOFT-2023- 1283-K의 지원에 의하여 수행되었음.
* 이선민, 박정남, 원정담 연구자는 서울대학교 신임교수 연구정착금으로 지원되는 연구비, 서울대학교 컴퓨터연구소의 지원과 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 지원을 받아 연구를 수행하였음 (IITP-2023-2020-0-01460)

2. 시스템 개요

우리는 캐릭터 모션을 골격 비의존적 잠재공간으로 투영하는 인코더, 잠재공간을 임의의 캐릭터 골격의 모션으로 실현하는 디코더, 그리고 잠재공간 상에서 수행하는 캐릭터골격 비의존적 애니메이션 작업으로 구성된 프레임워크를 제안한다 (그림1).

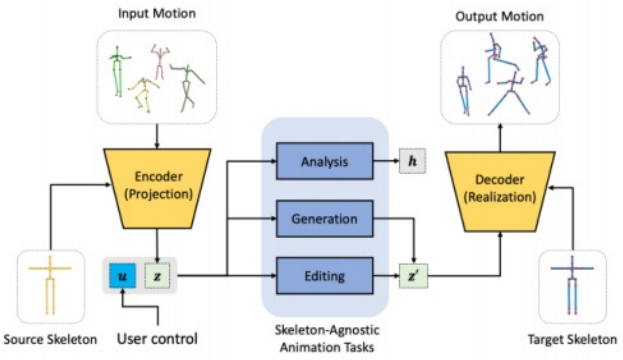


그림 1 시스템 개요

3. 캐릭터골격 비의존적 동작 공간 오토 인코더

3.1. 그래프 심층 신경망

우리는 임의의 골격 구조의 캐릭터 모션을 효과적으로 포착하기 위해 관절간 연결관계를 그래프로 표현하였다. 그래프로 표현된 모션 데이터를 잠재공간으로 투영하고 다시 복원하기 위해 Graph Convolution과 Graph pooling의 두 연산을 활용한 그래프 심층 신경망 기반 오토인코더를 제안한다 (그림2, 그림3).

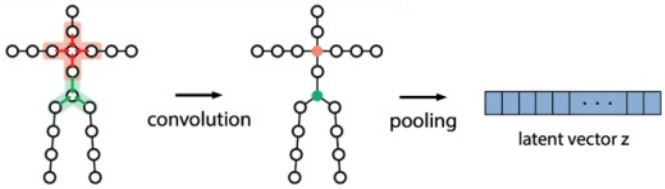


그림 2 Graph convolution, graph pooling 연산

Graph Attention[3]은 Graph convolution의 일종으로 이웃 노드와의 관계에 따라 동적 가중치를 집계하고

모든 노드의 공통 파라미터를 학습하는 방법으로 임의의 캐릭터 골격 토폴로지를 하나의 네트워크로 다루기에 적합하다. Graph pooling 은 모든 노드의 데이터를 하나의 고정길이 벡터로 집계하는 연산으로, 임의의 캐릭터 모션을 골격 구조 및 관절 개수 등에 관계없이 공통의 잠재공간으로 투영하기 위해 사용하였다.

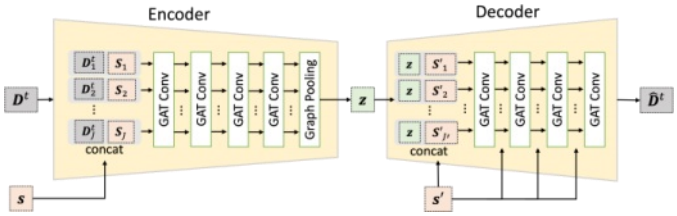


그림 3 그래프 심층 신경망 기반 골격 비의존적 모션 오토인코더 구조

3.2. 데이터셋 및 학습

다양한 골격구조의 모션 데이터를 통해 골격 비의존적 모션 오토인코더를 지도학습한다. 효율적인 학습을 위해 모션 데이터를 임의의 골격구조로 리타기팅하여 데이터를 증강한다. 즉, 움직임의 의미가 같지만 캐릭터 골격이 서로 다른 모션 데이터 쌍으로 오토인코더를 지도학습한다. 총 약 130분, 16가지 캐릭터가 포함된 데이터를 수집하였으며, 이를 약 13시간, 170여개의 캐릭터를 포함한 데이터로 증강하였다. 데이터 증강은 MotionBuilder의 리타기팅 기능을 사용하였다.

4. 캐릭터골격 비의존적 애니메이션 작업

4.1 모션 리타기팅

우리가 제안한 골격-비의존적 모션 인코더와 디코더에 서로 다른 캐릭터 골격을 입력하여 캐릭터 모션을 리타기팅할 수 있다. Aberman et al.[4]는 골격 구조마다 오토인코더를 트레이닝해야 하는 것에 비해, 우리는 하나의 오토인코더 모델로 임의의 캐릭터로 리타기팅을 수행할 수 있는 장점이 있다.

4.2 모션 분류

다양한 골격 구조의 이족 캐릭터 모션을 골격-비의존적 잠재공간으로 투영하고, 잠재공간 벡터에 따라 모션을 분류하는 분류기를 학습할 수 있다. 95% 성능의 분류기를 학습하였다.

4.3 비슷한 모션 검색

임의의 골격 구조를 가진 모션 데이터 간의 유사도를 정의하는 것은 까다롭다. 우리는 골격-비의존적 모션 인코더를 통해 서로 다른 골격의 모션 데이터셋을 공통된 잠재공간으로 투영하고, 잠재공간상의 거리를 비교해 비슷한 동작을 검색하였다.

4.4 누락된 일부 신체의 모션 복원

골격-비의존적 모션 오토인코더로 주어진 모션을 임의의 캐릭터로 재현하는 것을 넘어, 누락된 일부 신체의 모션 또한 복원할 수 있다. 상체 움직임만이 주어졌을 때 하체를 복원하거나, 손목까지 주어진 움직임에서 손가락을 복원하는 등의 특정 신체의 복원을 위해 모델을 추후 미세조정하였다.

4.5 실시간 캐릭터 모션 제어

골격-비의존적 잠재공간 상에서 모션매칭을 통해 사용자의 입력에 따라 캐릭터를 실시간으로 제어할 수 있다. 서로 다른 캐릭터 골격구조가 섞여있는 여러 출처의 모션 데이터셋을 잠재공간 데이터베이스로 변환하고, 그 공간상에서 모션매칭을 수행하였다. 또한, 골격-비의존적 디코더의 입력 캐릭터를 바꿈으로써 제어하는 캐릭터를 실시간으로 변경할 수 있다.

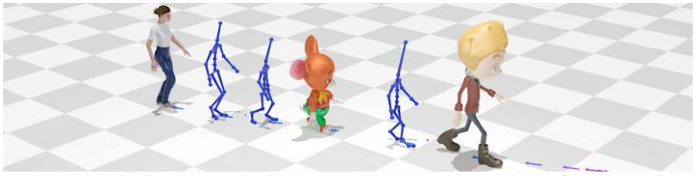


그림 2 잠재공간 상의 모션매칭 결과. 캐릭터를 실시간으로 변경할 수 있다.

5. 결론

우리는 캐릭터의 골격과 무관한 동작의 의미를 담는 캐릭터골격 비의존적 잠재공간을 학습하고 임의의 캐릭터 모션과 이 잠재공간을 연결해주는 오토인코더를 제안하였다. 이를 통해 모션 분류, 모션 검색, 실시간 제어 등의 캐릭터 애니메이션 작업을 캐릭터 골격에 비의존적으로 수행할 수 있음을 보였다.

하지만 다른 데이터 기반 모델 학습과 마찬가지로, 우리가 제안한 모델 또한 학습에 사용한 데이터셋에 큰 영향을 받으며, 그 너머의 일반화에는 어려움이 있다. 우리의 작업은 현재 이족 캐릭터를 가정하고 있으며, 학습과정에서 본 적 없는 신체(꼬리, 옷, 장신구 등)가 있을 경우, 이에 대해 적절한 대처를 하지 못한다. 더 넓은 범위의 캐릭터 골격 구조로 본 논문을 확장하는 것은 큰 의의가 있을 것이다.

우리는 본 논문의 접근법이 다양한 캐릭터의 모션이 섞여있는 비정제된 데이터셋을 이용한 거대 모션 모델학습 및 사용에 기여할 것으로 기대한다.

참고문헌

[1] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. 2019. AMASS: Archive of motion capture as surface shapes. In Proceedings of the IEEE/CVF international conference on computer vision (CVPR). 5442–5451

[2] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. ACM Trans. Graph. 34, 6 (Oct. 2015), 248:1–248:16

[3] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. arXiv preprint arXiv:1710.10903 (2017).

[4] Kfir Aberman, Peizhuo Li, Dani Lischinski, Olga Sorkine-Hornung, Daniel Cohen-Or, and Baoquan Chen. 2020. Skeleton-aware networks for deep motion retargeting. ACM Transactions on Graphics (TOG) 39, 4 (2020), 62–1.

우수 국제 학술대회 논문 발표

모델링/이미지/비디오

3차원 스케치를 이용한 물체의 함축적 표현형*

최창운[0], 이재아, 박재식, 김영민
서울대학교
changwoon.choi00@gmail.com, {hayanz, jaesik.park, youngmin.kim}@snu.ac.kr

3Doodle: Compact Abstraction of Objects with 3D Strokes

Changwoon Choi[0], Jaeah Lee, Jaesik Park, Young Min Kim
Seoul National University

Abstract

While free-hand sketching has long served as an efficient representation to convey characteristics of an object, they are often subjective, deviating significantly from realistic representations. Moreover, sketches are not consistent for arbitrary viewpoints. We propose 3Doodle, generating descriptive and view-consistent sketch images given multi-view images of the target object. We express 2D sketches as a union of view-independent and view-dependent components. Our pipeline directly optimizes the parameters of 3D stroke primitives to minimize perceptual losses in a fully differentiable manner. We demonstrate that 3Doodle can faithfully express concepts of the original images compared with recent sketch generation approaches.

1. Introduction

Free-hand sketching is an effective tool for visual communications. Previous works explore the connection between 3D structure and conventional sketch lines and show promising results in extracting sketch lines from detailed 3D geometry [1]. However, it is not trivial to formulate the exact mapping from artistic sketch curves on-screen space to the underlying 3D structure of the object.

In this work, we present a method to obtain 3D geometric primitives from multi-view images which can be rendered to 2D sketch lines from arbitrary views.

2. Sketch Representation

We propose a coherent set of 3D strokes which can be rendered into 2D sketch lines. Parts of sketch lines can be rendered from 3D geometric feature lines which can be represented directly from curves that reside in 3D space. However, 3D curves with fixed 3D locations cannot convey the geometric structure completely (e.g. contour lines of smooth surfaces).

To embrace different structural elements, we separately model view-dependent and view-independent components:

$$\mathcal{S}^{3D} = \mathcal{S}^{3D}_{ind} \cup \mathcal{S}^{3D}_{dep}$$

The view-independent components $\mathcal{S}^{3D}_{ind}$ are represented as a set of 3D cubic Bézier curves:

$$\mathcal{S}^{3D}_{ind} = \{B^{3D}(p_i)\}_{i=1}^{N_{ind}}, p_i = (p_i^0, p_i^1, p_i^2, p_i^3)$$

where N_{ind} is number of strokes.

Additionally, view-dependent components $\mathcal{S}^{3D}_{dep}$ encapsulate the 3D volume of the given object. We use the composition of superquadrics as a compact parametric representation that can express various 3D shapes.

3. Differentiable Rendering

3.1 Differentiable Rendering of 3D Bézier curves

We utilize differentiable rasterization library[2] which is able to render 2D Bézier curve. We first project the 4 control points to image plane, then we approximate the rational Bézier curve to cubic Bézier curve and draw the 2D Bézier curve with the projected

---

* 구두(포스터) 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 ACM Transactions on Graphics, Vol. 43, No.4에 게재확정 되었음.

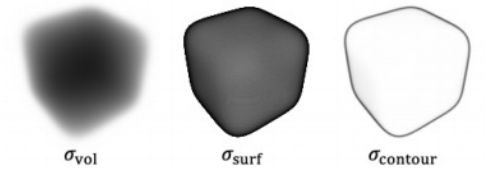


Figure 1. Visual comparison of the volume rendered results of different volume density function.

control points.

3.2 Differentiable Rendering of Contours

We propose a method to differentiable render the contours given volume density. We define a surface volume density function and view-dependent contour volume density function and we can render the contour by volume render the defined density. We illustrate the rendered volume densities in Figure 1.

4. Optimization

Given the differentiable rendering pipeline, we optimize the parameters of our sparse 3D primitives to generate sketch images. While there exists a significant domain gap between photo-realistic input images and the abstract sketch, we design a loss function to encourage a balance between structural components and semantic perception compared to input multi-view images. Specifically, we use LPIPS loss [3] to capture rough geometric layout and CLIP score [4] to maintain the high-level semantic meaning. Then, we optimize the sketch parameters to minimize the loss function.

5. Result

We show the reconstructed 3D sketch lines from multi-view images in Figure 2. Both in synthetic images and real-world images (last two rows), 3Doodle can reconstruct the feature curves that convey both geometric structures (fine structures of sails in the boat scenes) and semantic meanings (a wave pattern in the boat scene). Our view-dependent components successfully represent the smooth bounding surfaces that envelop the objects.

Figure 2. Qualitative results of 3Doodle.

References

[1] Doug DeCarlo, Adam Finkelstein, SzymonRusinkiewicz, and Antyony Santella. 2003. Suggestive contours for conveying shape. *ACM Transactions on Graphics (TOG)* 22, 3 (2003), 848-855.

[2] Tzu-Mao Li, Michael Lukac, Gharbi Michael, and Jonathan Ragan-Kelly 2020. Differentiable Vector Graphics Rasterization for Editing and Learning. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 39, 6 (2020), 193:1-193:15.

[3] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6012-6021.

[4] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748-8763.

논문 발표

시각화/렌더링

H-ETC2: 모바일을 위한 CPU-GPU 하이브리드 텍스처 압축 기법*

이현기, 나재호[0]†
상명대학교 컴퓨터과학과
gusrl3204@gmail.com, jaeho.nah@smu.ac.kr

H-ETC2: A Hybrid CPU-GPU Texture Compression Methodology for Mobile Devices

Hyeon-ki Lee, Jae-Ho Nah[0]†
Dept. of Computer Science, Sangmyung University

요약

본 논문은 모바일 플랫폼의 표준 텍스처 압축 코덱인 ETC2 형식의 인코딩을 위한 새로운 CPU-GPU 하이브리드 방법을 제안한다. 본 연구에서 설계한 파이프라인에서는 CPU 인코더를 통해 고속으로 1차 인코딩을 완료한 후, 여기에서 문제가 있는 픽셀 블록들을 식별하여 이에 대해서만 병렬 연산에 특화된 GPU 인코더로 2차 인코딩을 수행하여 압축 품질을 향상시킨다. 그 결과, 기존 오프라인 ETC2 인코더와 비견할 만한 품질을 보이면서도 수십~수만 배 빠른 속도를 달성하였다.

1. 고속 텍스처 압축의 필요성

실시간 애플리케이션에서 텍스처 맵들은 보통 실시간으로 디코딩 가능하도록 압축된 형태로 저장된다. ETC2[1]는 OpenGL ES 3.0의 표준 텍스처 압축 코덱으로, 120억대 이상의 모바일 기기에서 이를 지원한다. 다른 압축 코덱과 마찬가지로 ETC2의 인코딩은 보통 오프라인 상에서 수행이 되는데, 이 경우 압축 속도와 품질 사이에 상충관계(trade-off)가 존재한다. 즉, 압축 품질을 올리기 위해서 복잡한 알고리즘을 사용하면 할수록 그만큼 앱 개발시 소요되는 시간이 증가한다. 특히, 수십~수백GB의 텍스처를 사용하는 AAA 게임 개발시에는 고속 텍스처 압축 기술의 중요성이 더욱 증가한다.

2. 제안하는 H-ETC2 인코딩 기법

본 논문에서는 고품질 고속 ETC2 인코딩을 달성하기 위해 새롭게 디자인한 CPU-GPU 하이브리드 ETC2 (H-ETC2) 인코딩 방법을 제안한다. 이 구조는 분기에 뛰어난 CPU 구조와 병렬 연산 성능이 높은 GPU 구조의 장점을 모두 활용하기 위해 etcpak[2] CPU 인코더와 Betsy[3] GPU 인코더를 [그림 1]과 같이 결합한다.

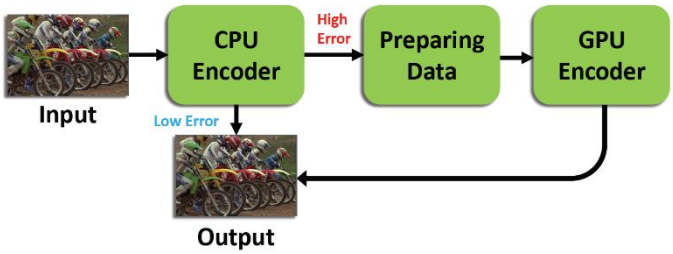


그림 1: H-ETC2 파이프라인의 전체 실행 과정

각 단계에 대해 살펴보면 다음과 같다. 먼저, CPU 파트에서는 QuickETC2 알고리즘[4]을 통합한 etcpak 정식 버전[2]을 활용, 고속으로 ETC2 인코딩을 수행한다. 다만 QuickETC2와 달리 최종 블록의 오류 계산시 luma 단위가 아닌 RGB 단위를 사용하며, 이후 각 채널의 오류 중 가장 높은 값을 취하여, 이를 기준으로 PSNR 쓰레시홀드(35.68 dB)보다 낮은 픽셀 블록들을 별도의 쓰레드 로컬 버퍼에 저장한다. 이 블록들은 압축이 끝난 후 하나의 버퍼로 병합되며, 설정된 압축 모드(best, normal, fast)에 따라 전부, 오류 상위 40%, 또는 오류 상위 10%의 블록들을 2D 텍스처 형태의 GPU 입력 데이터로 재구성한다. 즉, 전체 이미지가 아닌 문제가 있는 일부 블록들만 GPU의 입력 데이터로 구성되기 때문에, 기존 Betsy 대비 CPU에서 GPU로의 통신 비용 및 GPU 상에서 요구되는 연산량을 감소시킬 수 있다.

GPU 인코딩은 Betsy[3]의 흐름을 그대로 따르지만, 앞서 언급한 입력 데이터 크기의 감소와 함께 GPU 인코딩 파트 자체의 품질 및 속도도 크게 개선하였다. 먼저 8비트의 RGB 범위를 벗어나는 Betsy의 양자화 오류를 수정하였고, 아울러 RGB 채널에 각각 다른 비중을 부여하는 지각 오류 메트릭(perceptual error metric)

---

* 구두발표논문
* 본 논문은 요약논문(Extended Abstract)으로, 원본 논문[9]은 Computer Graphics Forum 저널 및 Pacific Graphics 2023 학회에서 발표되었음
* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1G1A1093404)
† 교신저자

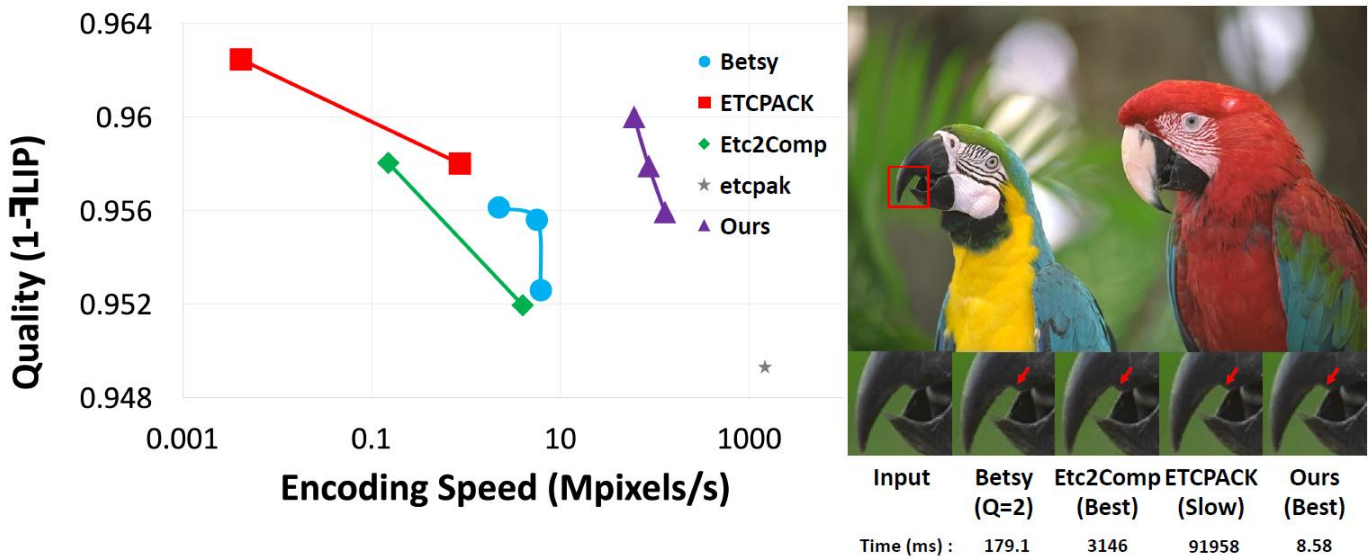


그림 2: QuickETC2 테스트 세트[4]에서의 품질과 인코딩 속도 비교 결과 (좌), *kodim23*에서의 비교 결과 (우)

을 적용, 눈에 띄는 블록 아티팩트(block artifact)를 크게 개선하였다. 또한, Betsy에서 4x4 블록 내 픽셀들의 조합으로 만들어지는 총 120쌍 모두에 대해 K-평균 클러스터링을 수행하는 것을 개선, 대각선 부분의 4개 쌍에 대해서만 클러스터링을 수행한 후 이를 ETC2의 T/H모드에서 이용하도록 함으로써 큰 품질 저하 없이 GPU 파트의 압축 속도를 6배 이상 높였다.

3. 실험 및 결과

실험은 QuickETC2[4]의 64개 텍스처 세트를 이용하여, 인텔 코어 i5 12400 CPU와 엔비디아 지포스 RTX3060 GPU로 구성된 컴퓨터 상에서 수행되었다. [그림 2]와 같이 인코딩 속도는 초당 처리 픽셀 수(Mpixels/s)로, 품질은 FLIP[7] 수치로 비교하였다. H-ETC2의 best 모드는 품질 측면에서 레퍼런스 인코더인 ETCPACK[5]의 slow 모드 다음의 매우 우수한 결과를 나타냈다. 속도 측면에서 H-ETC2는 기존 오프라인 인코더 (ETCPACK[5], Etc2Comp[6], Betsy[3]) 대비 수십~수만 배 빠른 결과를 나타냈다. etcpak[2,4]와 비교 시에는 H-ETC2가 추가 GPU 인코딩으로 인해 느린 속도를 보여주지만, etcpak에서 발견되는 블록 아티팩트나 색상 왜곡 등의 품질 저하 요인은 대부분 해결되었다.

4. 결론 및 향후 연구

본 논문에서는 고품질, 고속의 CPU-GPU 하이브리드 ETC2 인코딩 파이프라인인 H-ETC2를 제안하였다. 이는 최고품질 인코더와 비교해도 큰 차이가 없는 결과를 출력해 주면서도 텍스처 한 장당 수~수십 ms 안에 인코딩을 완료해 주기 때문에, 기존 게임 엔진(Unity 등)에서 오랜 기간 사용되어 왔던 오프라인 ETC 인코더[5,6]를 대체할 수 있을 것으로 기대한다. 향후 연구로는 본 연구에서 적용한 방법들이 다른 표준 코덱(BC7, ASTC 등)에도 적용 가능한지 조사해 보고자 하며, 또한 개선된 CPU 인코딩 알고리즘[8]의 적용을 통해 GPU 인코딩 부담 경감과 품질 개선을 도모하고자 한다.

참고문헌

[1] J. Störm and M. Pettersson, “ETC2: Texture Compression using Invalid Combinations,” in *Proc. of SIGGRAPH/Eurographics Workshop on Graphics Hardware*, 2007, pp. 49-54.
[2] B. Taudul, etcpak: The fastest ETC compressor on the planet, 2023. URL: https://github.com/wolfpld/etcpak
[3] M. N. Goldberg, besty: Betsy GPU compressor, 2022. URL: https://github.com/darksylinc/betsy
[4] J.-H. Nah, “QuickETC2: Fast ETC2 texture compression using luma differences,” *ACM Transactions on Graphics*, vol. 39, no. 6, Article 270, 2020.
[5] Ericsson, ETCPACK: Provides code for compression of textures for ETC1, ETC2 and EAC textures, 2018. URL: https://github.com/Ericsson/ETCPACK
[6] Google Inc. and Blue Shift Inc., Etc2Comp –Texture to ETC2 compressor, 2017. URL: https://github.com/google/etc2comp
[7] P. Andersson, J. Nilsson, T. Akenine-Möller, M. Oskarsson, K. Åström, and M. D. Fairchild, “FLIP: A Difference Evaluator for Alternating Images.” *Proc. of ACM Computer Graphics and Interactive Techniques*, vol. 3, no. 2, Article 15, 2020.
[8] J.-H. Nah, “QuickETC2-HQ: Improved ETC2 encoding techniques for real-time, high-quality texture compression,” *Computers & Graphics*, vol. 116, 2023, pp. 308-316.
[9] H. Lee and J.-H. Nah, “H-ETC2: Design of a CPU-GPU Hybrid ETC2 Encoder,” *Computer Graphics Forum*, vol. 42, no. 7, 2023.

논문 발표

애니메이션/시뮬레이션1

전신 모션 데이터를 이용한 실시간 얼굴 애니메이션 생성

장보경[0], 이윤상
한양대학교 컴퓨터소프트웨어학과
{jbok6825, yoonsanglee}@hanyang.ac.kr

Real-Time Facial Motion Synthesis based on Body Movements

Bokyeong Jang[0], Yoonsang Lee
Department of Computer Science, Hanyang University

요약

본 연구에서는 캐릭터의 전신 모션 데이터를 활용하여 해당 데이터에 부합하는 얼굴 애니메이션을 실시간으로 생성하는 방법을 제안하며, 이를 통해 사용자가 캐릭터의 전신 모션이 자연스럽다고 느끼는 정도를 향상시키는 데 기여할 수 있다.

1. 서론

캐릭터의 움직임과 얼굴 표정이 어울리지 않거나 무표정으로 일관한다면 관찰자는 캐릭터의 움직임을 자연스럽지 못하다고 느낄 수 있다. 이는 우리가 항상 일상생활에서 다른 사람의 움직임을 관찰할 때 움직임 자체뿐만 아니라 그와 동시에 변화하는 표정도 함께 관찰해왔기 때문일 것이다. 즉, 얼굴 표정은 감정의 상태 뿐만 아니라 움직임의 자연스러움에 대한 인식에도 영향을 미친다고 얘기할 수 있다. 실제로 최근의 전신 모션을 결과로 보여주는 로코모션 동작 생성[1]이나 제스처 생성[2] 등을 다루는 많은 연구들은 캐릭터가 무표정한 얼굴로 움직이는 것을 보여주는 대신, 얼굴 표정이 드러나지 않는 캐릭터 메시를 사용하여 결과를 보여주고 있다.

본 연구에서는 사용자가 캐릭터의 전신 움직임이 더 자연스럽다고 느낄 수 있도록 전신 움직임에 부합하는 얼굴 애니메이션을 실시간으로 생성하는 방법을 제안한다. 이 시스템은 두 단계의 학습 과정을 거친다. 첫 번째 단계에서는 자기회귀 조건부 변이형 오토인코더(VAE) 구조를 활용하여 얼굴 애니메이션의 운동학적 전이 정보를 학습한다. 두 번째 단계에서는 학습된 VAE의 디코더 부분을 이용하여, 현재 및 과거의 연속된 전신 포즈를 입력으로 받아 전신 동작에 적합한 얼굴 애니메이션을 프레임 단위로 생성할 수 있는 컨트롤러를 학습한다.

* 구두 발표논문, 요약논문 (Extended Abstract)
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 및 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

2. 네트워크 구조

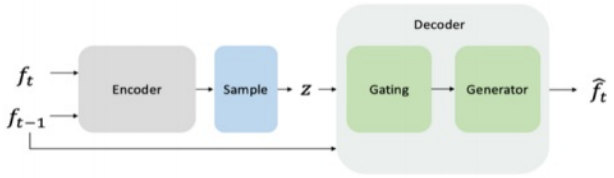


그림 1: Facial Motion VAE 네트워크

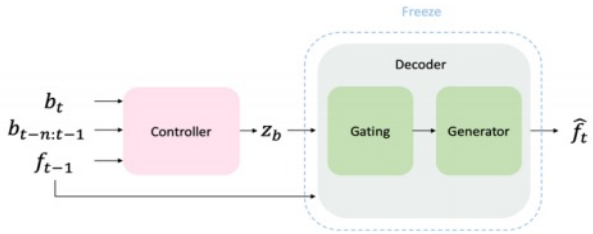


그림 2: Controller 네트워크

우리의 시스템은 크게 Facial Motion VAE 부분과, Controller 부분으로 구성된다. 런타임에는, Controller와 Facial Motion VAE의 디코더 부분을 사용해 전신 모션에 적절한 표정 정보를 한 프레임씩 생성하게 된다.

2.1. Facial Motion VAE

본 연구에서는 Motion VAEs[3]로부터 영감을 받아 자기회귀 조건부 변이형 오토인코더, 즉 Facial Motion VAE를 사용하여 자연스러운 얼굴 동작 전환 정보를 지도학습 방식으로 학습하였다. Facial Motion VAE는 인코더와 디코더로 구성되며, 인코더는 현재 및 직전 프레임의 표정 정보 f_t 와 f_{t-1} 을 입력으로 받아 잠재공간에서 추출된 잠재 벡터 z를 출력한다. 이 잠재 벡터는 얼굴 표정의 전이 정보를 나타낸다. 디코더는 직전 표정 정보 f_{t-1}와 인코더로부터 생성된 잠재 백터 z를 입력으로 받아 현재 프레임의 표정 $\hat{f}_t$을 예측한다. 각 시점의 표정 정보 f_k는 FLAME[4] 포맷으로 표현되며, 이는 100개의 shape key 값과 턱의 상태를 나타내는 회전 값을 포함한다.

2.2. Controller

전신 모션 정보가 주어졌을 때, 현재 모션에 적합한 표정을 생성할 수 있는 Controller를 지도 학습 방식으로 학습하였다. Controller는 두 개의 은닉층을 갖는 인코더로, 각 층은 128개의 유닛과 ELU 활성화 함수를 사용한다. 이 네트워크는 현재의 전신 포즈 정보 b_t, 연속된 과거의 전신 포즈 정보 $b_{t-n:t-1}$, 그리고 직전 표정 정보 f_{t-1}를 입력으로 받아, 이 입력들을 바탕으로 얼굴의 표정 전환 정보를 나타내는 잠재 벡터 z_b를 출력한다. 출력된 잠재 벡터 z_b는 사전 학습되어 고정된 Facial Motion VAE의 디코더 부분에 입력으로 들어가며, 디코더는 Controller가 생성한 잠재 벡터를 입력으로 사용하여 전신 모션 정보에 적합한 표정 정보를 생성하게 된다. 각 시점의 전신 포즈 정보 b_k는 현재 시점의 캐릭터 프레임에 대해 표현된 각 관절의 위치와 회전 및 속도로 표현된다.

3. 네트워크 학습

우리의 시스템은 두 단계의 학습과정을 거친다. 첫 번째 단계에서는 Facial Motion VAE를 학습하고, 두 번째 단계에서는 Controller와 사전학습된 Facial Motion VAE의 디코더 부분을 고정시켜 학습한다. Facial Motion VAE의 학습에는 재구성 손실과 KL 발산 손실을 가중합하여 사용한다. 반면, Controller의 학습에는 재구성 손실만을 사용한다.

재구성 손실은 Facial Motion VAE의 디코더가 출력한 표정 정보 $\hat{f}_t$와 원본 데이터 f_t 사이의 차이를 측정하며, FLAME 포맷의 각 shape key 값의 평균 제곱 오차와 턱의 회전 값 차이를 가중합하여 계산한다. KL 발산 손실은 Facial Motion VAE의 인코더가 출력하는 잠재 벡터 z가 샘플링되는 잠재 공간에 적용되며, 이 잠재 공간이 다변량 표준 정규 분포에 가까워지도록 돕는다.

4. 실험 결과

우리의 시스템은 Motion-X[5] 데이터셋의 일부를 사용하여 학습되었으며, Facial Motion VAE의 학습에는 약 30시간이, Controller의 학습에는 약 2시간이 소요되었다. 실험에서는 학습 데이터셋에 포함되지 않은 Motion-X의 전신 모션 데이터 2개를 입력으로 제공하여 생성된 표정 동작의 자연스러움을 평가하였다. 각각의 데이터는 춤을 추는 경우와 무거운 물건을 드는 경우의 전신 모션에 해당된다.

그림 3은 각각의 데이터에 대해 생성된 표정 모션을 입력으로 넣어준 전신 모션과 함께 시각화 해주었을 때의 결과이다. 춤을 추는 전신 모션을 입력으로 넣어주었을 때는 웃는 표정이 전신 모션에 동기화되어 나타났다. 반면 무거운 물건을 드는 전신 모션을 입력으로 넣어주었을 때는 찡그리고 힘들어하는 표정이 전신 모션에 동기화되어 나타나는 것을 확인할 수 있었다.

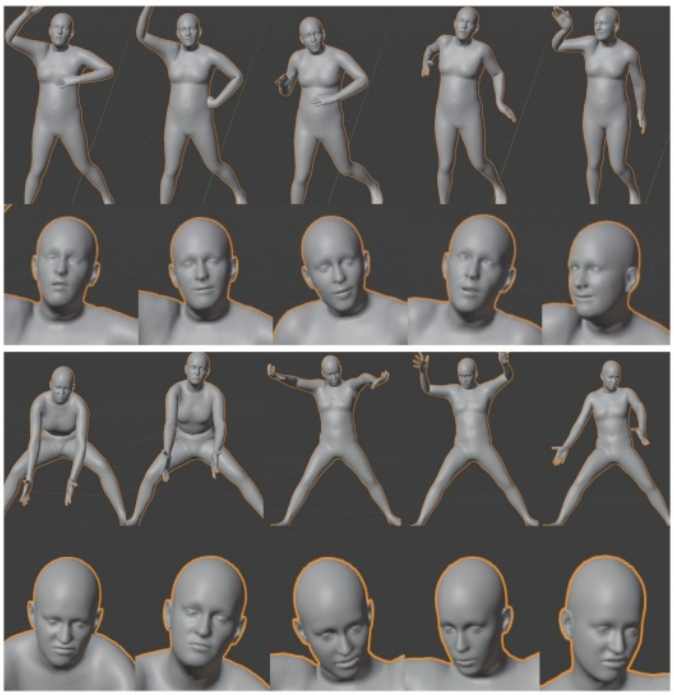


그림 3:
춤을 추는 동작(위)과
무거운 물건을 드는 동작(아래)에
대해 생성된 얼굴 모션 결과

5. 결론

우리는 캐릭터의 전신 모션 데이터를 활용하여 해당 데이터에 부합하는 얼굴 애니메이션을 실시간으로 생성하는 방법을 제안하였다. 실험을 통해 입력된 전신 모션에 적합하고 동기화된 부드러운 표정 모션이 생성되는 것을 확인할 수 있었다. 이는 우리 시스템이 전신 모션 데이터에 대응하여 자연스러운 얼굴 애니메이션을 효과적으로 생성할 수 있음을 보여준다.

참고문헌

[1] Kwon, T. Gu, J. Ahn, and Y. Lee, "Adaptive Tracking of a Single-Rigid-Body Character in Various Environments," in *SIGGRAPH Asia 2023 Conference Papers*, 2023

[2] T. Ao, Z. Zhang, and L. Liu, "GestureDiffuCLIP: Gesture Diffusion Model with CLIP Latents," *ACM Transactions on Graphics (TOG)* 42, 4(2023), 1-18

[3] H. Y. Ling, F. Zinno, G. Cheng, and M. van de Panne, "Character Controllers Using Motion VAEs," *ACM Transactions on Graphics(TOG)* 39, 4(2020), 1-12

[4] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, "Learning a model of facial shape and expression from 4D scans," *ACM Transactions on Graphics(TOG)* 36, 6(2017), 1-17

[5] J. Lin, A. Zeng, S. Lu, Y. Cai, R. Zhang, H. Wang, and L. Zhang, "Motion-X: A Large-scale 3D Expressive Whole-body Human Motion Dataset," in *NeurIPS 2023 Conference Papers*, 2023

인체 근골격 모델의 현실적인 보행 동작 생성을 위한 평균 상태 손실 함수를 이용하는 모델 기반 강화 학습*

김민관[0], 이윤상
한양대학교 컴퓨터소프트웨어학과
{palkan21,yoonsanglee}@hanyang.ac.kr

Generating Realistic Locomotion of Musculoskeletal Humanoid with Model Based Reinforcement Learning using Average State Loss Function

Minkwan Kim[0], Yoonsang Lee
Department of Computer Software, Hanyang University

요약

본 연구는 최근 캐릭터 컨트롤러 학습 연구의 경향성과 다르게, 레퍼런스 모션 데이터의 도움 없이 근골격 모델의 자연스러운 보행 동작을 학습하는 방법을 제안한다. 이를 위해 사람의 구동을 모델링한 근육 시뮬레이션, 보행의 특성을 표현하기에 적합한 손실 함수, 그리고 시뮬레이션을 예측하는 월드 모델 등을 사용했다.

1. 서론

최근 심층 강화 학습으로 물리 시뮬레이션 되는 캐릭터가 모션 캡쳐 데이터를 재현하는 정책을 찾는 연구가 많이 발표되고 있다. 특히, 데이터셋에 포함된 움직임을 낮은 차원의 잠재 공간에 매핑하고, 이를 이용해서 특정 스킬을 선택하고, 움직임의 방향 및 속도 등을 조종하며 더 어려운 작업을 수행하는 능력을 여러 레벨에 걸쳐 학습하는 연구 [1]도 발표된 바 있다.

이러한 데이터 기반의 접근 방식은 다양한 동작 생성에 좋은 모습을 보이나, 학습 도중에 사용한 데이터의 특성이 결과 동작에 내재된다는 한계점이 존재한다. 또한, 모션 캡쳐가 어려운 동물이나 데이터가 희소한 동작, 혹은 가상의 생물 등에는 적용이 불가능하다.

모션 데이터를 사용하지 않고 동작 제어 정책을 학습하려는 시도도 있었으나, 보상 함수나 각 관절의 토크 제한 등에 대한 세밀한 조절을 필요로 하거나 [2], 모든 각도에서 봤을 때 자연스러운 동작을 만들지 못하고 오랫동안 보행을 지속하지 못하거나 [3], 학습된 정책이 선형 피드백 규칙과 함께 사용되어야 한다는 [4] 한계점이 있었다. 최근 6년간 모션 데이터에 기반하지 않는 다른 관련 연구를 찾기 힘들 정도로, 데이터 기반 방식에 비해서 동작 생성 정책의 학습에 어려움을 갖는다.

이러한 학습의 어려움을 극복하기 위해, 우리는 시뮬레이션을 예측하는 월드 모델 기반의 엔드 투 엔드 강화 학습으로 시뮬레이션의 내재된 원리를 제어 정책에게 전달했다. 또한, 보행 내의 특정 시간 구간 동안 캐릭터의 평균 상태를 계산해서 이를 바탕으로 손실 함수를 계산했으며, 실제 생물이 보행에서 최소화한다고 알려진 근육에 의해 계산된 에너지 소모 수식을 포함한다.

위와 같은 요소들에 힘 입어 본 연구는 데이터의 도움 없이 실제 사람과 비슷한 보행 동작을 생성했으며, 이는 근육 기반의 모델이 존재하는 모든 캐릭터의 보행 동작 생성에 확장될 수 있는 잠재력을 가진다.

2. 시스템 요약

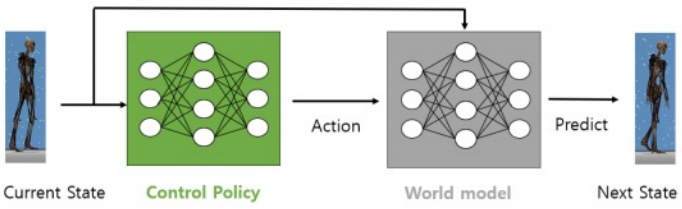


Figure 1: Overview of the proposed system

본 연구는 레퍼런스 모션의 도움 없이 현실적인 사람의 보행 동작을 생성하는 방법을 제안하며, 그림 1은 전체 학습 과정을 나타낸다. 제어 정책은 캐릭터의 현재 상태를 입력 받아서 각 근육의 활성화 값을 출력하고, 월드 모델을 거쳐서 캐릭터의 다음 상태가 결정된다.

시뮬레이션 단계가 미분이 불가능한 문제를 해결하기 위해 모델 기반의 강화 학습을 사용하였으며, 월드 모델은 학습 도중에 물리 엔진의 근육 시뮬레이션 결과를

* 구두 발표논문, 요약논문 (Extended Abstract)
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 및 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

예측하고 기울기를 역 전파하는 역할을 하고, 런 타임에는 물리 엔진으로 근육 시뮬레이션을 진행했다.

또한, 매 스텝마다 캐릭터의 상태를 평가하는 대신, 특정 시간 구간 동안의 평균 상태를 기반으로 보행에 대한 손실 함수를 계산하였다. 이는 실제 사람의 보행 주기에서 각 관절의 회전이나 루트의 움직임이 특정 값을 중심으로 편차를 갖는 것에서 기반했으며, 실제로 자연스러운 움직임을 학습하는 것에 중요한 역할을 했다.

3. 학습 과정 및 시뮬레이션 상세

제안하는 시스템은 우선 시뮬레이션 샘플을 수집하고, 해당 샘플로 월드 모델 및 제어 정책을 번갈아 가며 학습하는 과정을 반복한다.

3.1. 근육 시뮬레이션 샘플 수집

해당 단계에서는 현 시점의 제어 정책이 출력한 근육 활성화 값으로 물리 시뮬레이션을 진행하며 캐릭터의 이전 상태 및 다음 상태를 샘플로 버퍼에 저장한다.

제어 정책이 출력하는 0에서 1사이의 근육 활성화 값은, 각 근육의 이완 및 수축의 정도를 표현한다. 근육의 힘은 근육의 길이, 속도 및 활성화 값 기반으로 미리 정의된 그래프로부터 계산되며, 그 후 각 근육 별로 정의된 경로 점 사이에서 당기는 방향으로 힘이 적용해서 최종적으로 관절을 회전시킨다.

근육 시뮬레이션에는 MuJoCo [5]가 사용되었고, 각 근육이 낼 수 있는 힘, 경로 점의 위치 등 근육 관련 파라미터가 포함된 근골격 모델은 기존 연구 [6]의 것을 사용하였다.

3.2. 시뮬레이션을 예측하는 월드 모델 학습

월드 모델은 현재 캐릭터의 상태와, 제어 정책이 출력한 근육 활성화 값을 입력 받아서 다음 캐릭터의 상태를 출력한다. 월드 모델이 예측한 다음 상태는 버퍼에 저장된 실제 시뮬레이션 결과와 오차가 적어지도록 학습된다. 캐릭터의 상태는 각 근육의 길이, 속도와 각 관절의 위치, 속도, 회전, 각 속도 등의 정보로 구성된다.

3.3. 보행 동작 생성을 위한 제어 정책 학습

제어 정책은 캐릭터가 자연스러운 동작으로 앞으로 넘어지지 않고 이동하는 것을 목표로 학습된다. 정책은 액션을 출력하고 앞서 학습된 월드 모델로 캐릭터 상태를 업데이트하는 것을 24 스텝동안 반복한다. 그 뒤, 이 시간 구간 동안의 평균 상태를 이용해서 보행 동작의 만족 정도를 평가하며. 손실 함수의 구성은 아래와 같다.

$$loss = l_{vel} + l_{height} + l_{dir} + l_{pose} + l_{energy}$$

l_{vel}, l_{height}, l_{dir}, l_{pose}는 각각 시뮬레이션 구간 동안 루트의 속도, 높이, 회전 및 캐릭터의 자세의 평균 값과 목표 값 사이의 차이를 의미하며, 이 차이를 줄이는 방향으로 학습된다. 이는 실제 사람이 보행 시 속도, 높이, 각 관절의 회전 등이 편차를 보이지만, 보행의 한 주기에 해당하는 시간 구간의 평균을 구하면 특정 값으로 수렴한다는 관찰에서 기반한 손실 함수 계산 방식이다.

실험을 통해 찾은 목표 속도는 일반적인 보행의 평균 속도 범위에 포함되는 1.2m/s, 목표 높이는 캐릭터의 초기 높이보다 약간 낮은 0.91m, 목표 자세는 가만히 서 있는 자세로 일반적인 예측과 비슷하다. l_{energy}는 목표 값이 0으로, 근육의 에너지 소모 수식 계산 결과 값을 최소화하며 자연스럽고 안정적인 동작을 학습하는데 중요한 역할을 한다.

4. 실험 결과 및 결론

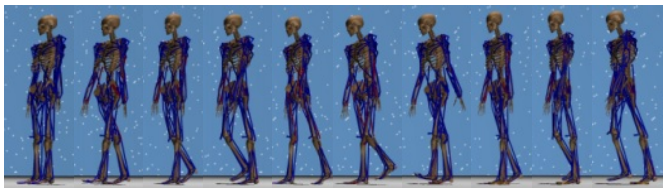


Figure 2: Realistic locomotion from proposed system

우리는 앞서 언급한 3가지 단계를 거쳐서 월드 모델과 제어 정책을 번갈아 가며 학습했고, 최종적으로 그림 2와 같이 실제 사람의 보행에서 보일법한 모습을 갖춘 현실적인 동작을 근육 시뮬레이션으로 생성했다. 제어 정책과 월드 모델은 모두 완전 연결 신경망이며 학습은 NVIDIA 4070 GPU에서 약 15시간이 소요되었다.

또한 섹션 2에서 제안한 요소들 대신 단순히 근육 활성화 값을 최소화하거나 혹은 매 스텝마다 손실 함수로 보행 정도를 평가하도록 학습할 경우, 무릎을 굽힌 채로 걷는 부자연스러운 동작이 생성되는 것을 확인했다.

결론적으로, 우리의 연구는 모델 기반의 강화 학습과 평균 상태 기반의 손실 함수, 그리고 근육에 의해 계산되는 에너지 최소화 요소를 포함하는 구조로 레퍼런스 모션의 도움 없이 자연스러운 보행 동작을 생성할 수 있음을 보였다. 추후에는 다양한 속도의 동작 생성 혹은 사람 이외의 근골격 모델에 대한 확장을 목표로 한다.

참고문헌

[1] Y Feng, X Xu, L Liu, MuscleVAE: Model-Based Controllers of Muscle-Actuated Characters, SA '23: SIGGRAPH Asia 2023 Conference Papers, December 2023, Article No.: 3, Pages 1–11

[2] W.Yu, G.Turk, K.Liu, Learning Symmetric and Low-energy Locomotion, ACM Transactions on Graphics 37(4), 2018.

[3] M.K.Kim and Y.S.Lee. Learning Human-like Locomotion Based on Biological Actuation and Rewards, ACM SIGGRAPH 2023 Posters July 2023 Article No.: 5 Pages 1–2

[4] G.C.Kang and Y.S.Lee. “Finite State Machine-Based Motion-Free Learning of Biped Walking”. IEEE Access 9 (2021), 20662–72.

[5] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012.

[6] Y.S.Lee, M.S.Park, T.S.Kwon, J.H.Lee, Locomotion control for many-muscle humanoid, ACM Transactions on Graphics, 33(6), 2014.

시뮬레이션 되는 휴머노이드 컨트롤을 위한 잠재변수 매칭*

김민수 [0], 이윤상
한양대학교 인공지능학과, 한양대학교 컴퓨터 소프트웨어학과
igotaspot426@gmail.com, yoonsanglee@hanyang.ac.kr

Latent Matching for Simulated Humanoid Control

Minsu Kim[0], Yoonsang Lee
Dept. of Artificial Intelligence, Hanyang University, Dept. of Computer Science, Hanyang University

요약

본 연구에서는 참조 모션 데이터셋을 추종하도록 사전 학습된 물리기반 캐릭터의 저수준 정책 모델을 이용해, 사용자의 실시간 입력을 따르도록 캐릭터를 제어하는 잠재변수 매칭 기법을 제안한다. 이 기법을 통해 추가적인 고수준 정책의 학습 없이도 사용자가 시뮬레이션 되는 캐릭터의 이동 속력, 방향 등을 실시간으로 제어할 수 있다.

1. 서론

물리기반 환경에서 시뮬레이션 되는 휴머노이드 캐릭터를 자유롭게 움직이도록 제어하는 방법은 게임, 가상현실, 시뮬레이션 분야에서 중요한 주제이다. 물리법칙에 의해 캐릭터의 움직임이 제약을 받기 때문에 물리기반 캐릭터 컨트롤은 최적화가 어렵다. 한편 최근 심층강화학습의 발전으로 인해, 물리 환경의 캐릭터의 보행 뿐만 아니라 난이도가 높은 묘기까지 보여줄 수 있게 되었다. 최근에는 적은 양의 모션이 아닌 많은 양의 모션 시퀀스를 잠재공간에 임베딩하는 사전학습 모델이 등장하면서 물리기반 캐릭터 연구가 가속되었다. 이 사전학습 모델에 추가적인 모델을 연결해 특정 태스크를 수행하는 계층적 구조의 에이전트의 구조가 많이 연구되었고, 기존에는 볼 수 없었던 자연스러우면서 다양한 높은 품질의 움직임을 보여줄 수 있었다. 그러나 이러한 계층적 구조는 추가적인 심층강화학습 에이전트의 학습과 모델, 그리고 복잡한 보상 설계가 요구된다.

우리는 운동학적 캐릭터 애니메이션 제어를 위해 많이 사용되는 모션 매칭 기법에 영감을 받아 [1, 2], 추가적인 모델과 학습없이 기존에 사전학습된 저수준 정책 모델만으로 시뮬레이션 되는 캐릭터를 실시간으로 제어할 수 있는 방법인 잠재변수 매칭 기법을 제안한다.

2. 물리기반 모션 임베딩 모델

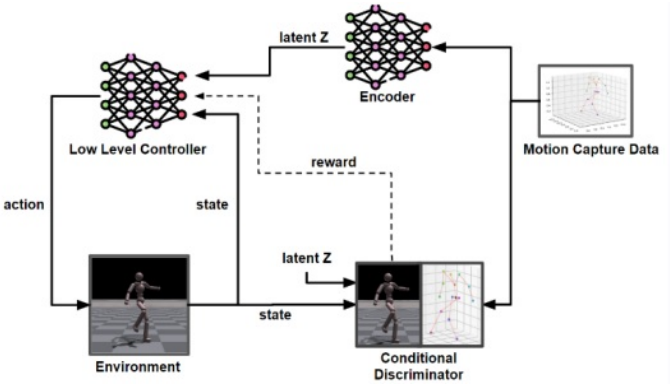


그림 1: 물리기반 모션 임베딩 모델 구조

물리기반 모션 임베딩 모델은 모션 데이터셋의 임의의 모션 시퀀스를 잠재공간의 특정 포인트에 임베딩하는 모델로서, 임베딩 된 포인트가 캐릭터의 상태와 함께 저수준 정책 네트워크에 입력되면 해당 모션 시퀀스를 재현할 수 있는 액션을 출력하도록 임베딩을 담당하는 인코더와 저수준 정책이 함께 학습된다. 본 연구에서는 물리기반 모션 임베딩 모델의 학습을 위해 CALM [3]에서 제안한 방법을 사용하였다. 이 모델은 동시에 학습되는 판별자가 물리환경에서 시뮬레이션 된 모션과 원본 모션 시퀀스와의 차이를 구분을 하지 못할 수록 높은 보상을 얻도록 강화학습 프레임워크에서 학습이 된다. 인코더, 판별자, 정책은 한꺼번에 종단간 학습이 된다. 저수준 정책 모델은 상태와 잠재변수를 입력받고 액션을 출력한다. 잠재변수는 인코더 모델의 출력으로부터 얻을 수 있다. 인코더는 모션 시퀀스를 단일 잠재변수로

---

* 구두 발표논문, 요약논문 (Extended Abstract)
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 과 정보통신기획평가원의 지원(No.RS-2020-II201373,인공지능대학원지원(한양대학교)) 과 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

변환하는 역할을 하기 때문에 상태의 시퀀스를 입력으로 받는다. 결과적으로 원본 모션 시퀀스는 인코더에 의해 특정 잠재변수로 변환이 되고, 학습이 완료되면 인코더 없이 학습된 잠재변수만으로 저수준 정책 모델의 입력으로 사용될 수 있다. 그림 1 은 물리기반 모션 임베딩 모델의 구조를 보여주고 있다.

4. 잠재변수 매칭

본 연구에서 제안하는 잠재변수 매칭 방법은 우선 현재 캐릭터의 상태와 유저의 입력을 조합하여 쿼리 특성(feature)을 만들고, 모션 데이터베이스의 각 프레임에 해당하는 특성을 미리 계산하여 저장해놓은 특성 데이터베이스에 이를 쿼리하여 가장 가까운 특성에 해당하는 모션 프레임을 찾는다. 모션 매칭에서는 해당 모션 프레임부터 캐릭터 모션을 재생하지만, 잠재변수 매칭 방법에서는 해당 모션 프레임에 해당하는 잠재변수를 계산하여 이를 저수준 정책에 입력으로 제공하여 캐릭터의 동작을 제어하게 된다. 쿼리에 사용하는 특성은 다음과 같은 27 차원의 벡터이다.

$$\mathbf{x}_i = \{\mathbf{p}_i^{LH}, \mathbf{p}_i^{RH}, \mathbf{p}_i^{LF}, \mathbf{p}_i^{RF}, \mathbf{v}_i^{root}, \mathbf{t}_i, \mathbf{d}_i\} \in \mathbb{R}^{27}$$

$\mathbf{p}_i^{LH}$, $\mathbf{p}_i^{RH}$, $\mathbf{p}_i^{LF}$, $\mathbf{p}_i^{RF}$ 는 각각 양 손과 양 발의 위치를 의미하는 3 차원 벡터이다. $\mathbf{v}_i^{root}$ 는 캐릭터의 루트(골반)의 속도를 의미하는 3 차원 벡터이고, $\mathbf{t}_i$ 는 루트의 미래 세 시점에 대한 평면 위치인 6 차원 벡터이며, $\mathbf{d}_i$ 는 루트의 미래 세 시점에서의 수평면 상의 루트 정면 방향을 의미하는 단위 벡터들의 집합인 6 차원 벡터이다. 쿼리할 때는 특성 생성기라는 간단한 함수에 의해 특성이 만들어지는데, 양 손과 양 발의 위치 정보는 캐릭터 상태로부터 얻고, 미래 시점의 위치와 정면방향은 유저입력으로부터 얻어서 만들어진다. 그림 2 에서 잠재변수 매칭의 구조를 볼 수 있다.

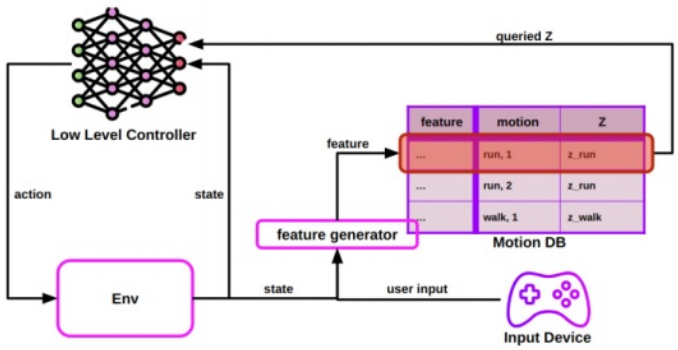


그림 2: 잠재변수 매칭 구조

5. 실험

잠재변수 매칭의 정성적 평가를 위해 매칭 주기를 변화시켜 가며 시뮬레이션 결과를 확인했고 그림 3 은 그 결과를 4 개의 프레임에 대해 캡처한 이미지다. 매칭 주기 n은 n=1, 5, 10 으로 3 가지로 했는데, 여기서 n은 프레임 단위이고 n=1 일 때는 매 프레임마다 쿼리를 수행한다는 의미이다.
n=10 일 때는 캐릭터의 움직임이 부드럽지 못한 결과를 보여주었다. n=5 일 때는 부드러우면서 충분히 민첩한 움직임을 볼 수 있었다. n=1 때는 비교적 더 민첩하고 부드럽게 움직이지만, 매 프레임마다 매칭하기 때문에 연산 오버헤드가 발생해 fps가 떨어지는 것을 확인했다.

그림은 잠재변수 매칭 기법이 보행뿐만 아니라 게임패드의 특정 버튼을 눌러 점프 모션까지 수행할 수 있음을 보여준다. 이 때 버튼을 누르면 점프 모션을 하도록 명령하는 것이 아니라, 그 순간 만들어진 특성값을 쿼리한 결과가 점프 모션의 특정 프레임과 가장 가깝기 때문에 점프 모션에 대한 잠재변수를 반환하고 저수준 정책이 점프 모션을 시뮬레이션 하게 되는 것이다.

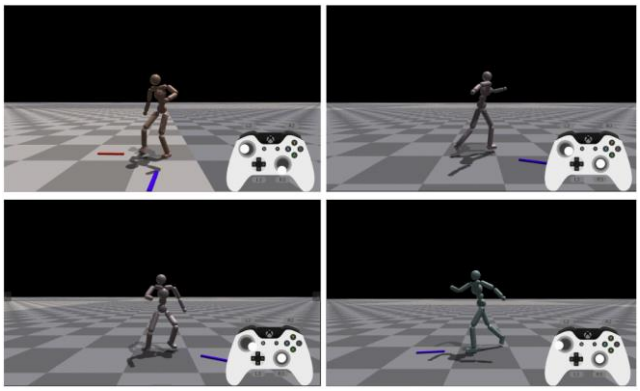


그림 3: n=5 일 때의 잠재변수 매칭 결과. 다양한 유저 입력에 대해 평면상에서 이동하는 모습을 보여주고 있음.

6. 결론

잠재변수 매칭 기법은, 사전학습한 물리기반 모션 임베딩 모델에 추가적인 학습없이 사용자의 입력에 기반한 컨트롤을 가능하도록 한다. 우리는 CALM [3]에 기반한 물리기반 모션 임베딩 모델을 통해 잠재공간을 사전학습했다. 우리의 제안은 상황에 맞는 적절한 잠재변수들을 찾아내는 방법으로 그 어떤 물리기반 모션 임베딩 모델에도 적용될 수 있는 방법이며 앞으로 다양한 방향의 추가 연구가 가능할 것이다.

참고문헌

[1] Simon Clavet, Motion matching and the road to next-gen animation, In Proc. of Game Developers Conference, 2016
[2] Jeongmin Lee, Taesoo Kwon, Hyunju Shin, Yoonsang Lee, Utilizing Motion Matching with Deep Reinforcement Learning for Target Location Tasks, Eurographics 2024 Short Papers, 2024
[3] Chen Tessler, Yoni Kasten, Yunrong Guo, Shie Mannor, Gal Chechik, and Xue Bin Peng, CALM: Conditional Adversarial Latent Models for Directable Virtual Characters, In ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23). Association for Computing Machinery, Article 37, 1–9, 2023

두 캐릭터의 상호작용 모션에 대한 기하학을 고려한 모션 리타겟팅*

장인서[1,2], 최수진[1], 홍석현[1], 김채린[1], 노준용[1]
KAIST[1], KAI Inc.[2]

Geometry-Aware Retargeting for Two Character Interaction

Inseo Jang, Soojin Choi, Seokhyeon Hong, Chaelin Kim, Junyong Noh

요약

여러 캐릭터 간의 상호작용 모션은 게임과 영화에서 널리 사용된다. 하지만 캐릭터의 다양한 메쉬 형태를 고려한 상호작용 모션 생성에 대한 연구는 아직 이루어지지 않았다. 우리는 임의의 메쉬 형태를 가진 두 캐릭터의 상호작용 모션을 리타겟팅하기 위해 Spatio Cooperative Transformer(SCT)를 제안한다. SCT는 상호작용하는 차이를 원본과 대상 캐릭터들 사이의 차이를 고려하여 루트의 위치와 관절의 회전 잔차를 예측한다. 또한, 우리는 상호작용하는 캐릭터들이 리타겟팅 될 때 원본 동작의 기하학적 거리를 유지하기 위해 앵커 손실 함수를 도입했다. 균일한 메쉬 형태를 가진 원본-대상 캐릭터 페어 데이터셋을 준비하기 위해 메쉬 변형 기반 모션 적응 방법을 이용한 모션 증강 방법을 제안한다.

1. 서론

모션 리타게팅은 소스 캐릭터의 모션을 타겟 캐릭터로 전달하면서 원래의 의미적 맥락을 유지하는 기술로, 보유한 모션 데이터를 재사용하여, 많은 비용이 소요되고 후처리 작업이 필요한 모션 캡쳐 작업을 줄여준다. 특히, 상호작용 동작은 영화와 게임 등 다양한 분야에서 빈번하게 등장하며, 악수 같이 단순한 모션부터 춤, 레슬링 같이 복잡한 모션에 이르기까지 다양한 종류가 존재한다. 이러한 상호작용 동작들이 현실적으로 보일수록 시청자나 플레이어의 몰입감이 높아진다.

다중 캐릭터 사이 상호작용 모션의 리타게팅을 위해 상호작용 메쉬 및 그래프[1, 2], 볼류메트릭 메쉬[3]를 활용한 방법들이 제안되었다. 하지만, 다양한 체형과 비율을 가진 캐릭터 간의 상호작용 모션 리타게팅은 원본 모션의 접촉을 유지하지 못하거나 상호침투와 같은 아티팩트가 발생할 수 있기 때문에 메쉬의 구조에 상관없는 상호작용 모션을 리타게팅하는 것은 아직 해결되지 않은 어려운 과제이다. 또한, 두 캐릭터의 상호작용 동작에 대한 리타게팅 시 상호작용하는 파트너를 고려하지 않고 본인의 모션만 고려하는 경우, 캐릭터 간의 침투가 발생할 수 있기 때문에 두 캐릭터의 형태를 모두 고려한 적절한 리타게팅 과정이 필요하다. 우리는 그림 1과 같이 다양한 메쉬 토폴로지를 가진 캐릭터 간 상호작용 모션을 리타게팅할 수 있는 Spatio Cooperative Transformer(SCT)를 제안하고 SCT을 학습하기 위한 데이터 준비 과정 및 추론 과정을 제안한다.

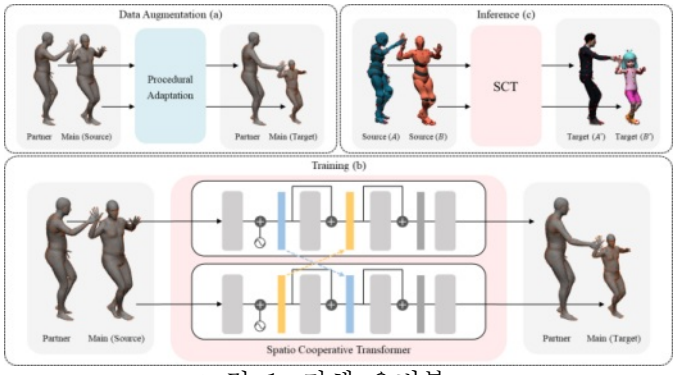


그림 1. 전체 오버뷰

SCT는 소스 모션을 타겟 캐릭터로 리타게팅할 때 소스와 타겟 캐릭터의 골격 구조와 메쉬 정보를 고려하여 소스 모션으로부터의 잔차를 예측한다. 우리는 캐릭터의 포즈를 표현하는 메쉬 비의존적인 표현인 앵커를 도입하고 상호작용하는 캐릭터 사이 앵커 간의 거리를 일치시켜, 상호작용의 기하학적 의미를 유지하는 앵커 거리 손실을 도입했다. 데이터 준비 단계에서는 변형 기반 적응을 이용한 모션 증강 방법을 사용하여 학습 데이터셋을 확장했다.

2. 방법

2.1. 데이터 증강

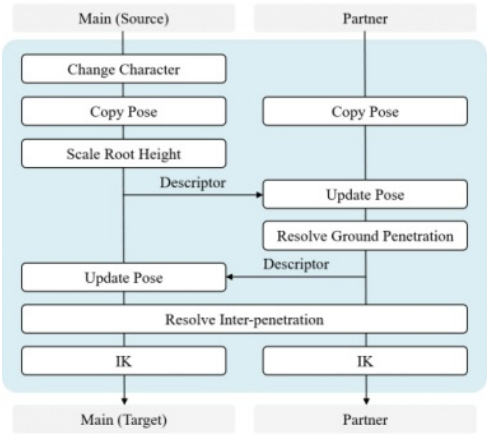


그림 2. 절차적 데이터 증강 방법

---

* 구두 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출중
* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00333478).

이 섹션에서는 학습 데이터셋 준비 절차를 설명한다. 우리는 SMPL 모델[4]을 템플릿 캐릭터로 사용하여 동일한 골격 구조와 메쉬 토폴로지를 공유하는 여러 인스턴스를 생성한다. 이 방법은 소스에서 타겟으로의 변형에 따른 포즈 조정을 쉽게 할 수 있게 하고, 뼈 길이와 메쉬 형태를 변경하여 다양한 인스턴스를 쉽게 증강할 수 있게 해준다. 우리는 캐릭터의 뼈 길이와 형태를 다양하게 조정해 여러 타겟 캐릭터를 생성하고, Relational descriptor[5]를 사용해 소스 캐릭터의 변형으로부터 파트너 캐릭터의 포즈를 조정한다. 데이터 준비 단계에서는 메인 캐릭터만 변경되고 파트너 캐릭터는 변하지 않으며, 메인 캐릭터와 파트너 캐릭터는 동일한 Descriptor point을 공유한다. 그림 2와 같이, 변형 기반 적응은 소스와 타겟 캐릭터의 형태 차이를 이용해 파트너 캐릭터의 관절 위치를 조정하며, 각 프레임에서 메인과 파트너 캐릭터의 모션을 절차적으로 조정한다.

2.2. Spatio Temporal Transformer

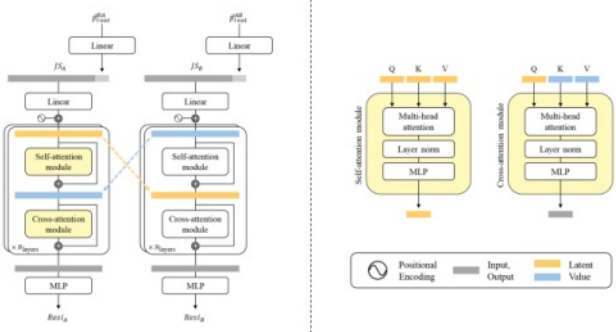


그림 3. SCT 구조(a) 및 어텐션 모듈(b)

SCT는 준비된 데이터셋을 기반으로 학습된 두 캐릭터 사이의 상호작용 모션을 리타게팅하는 네트워크이다. 두 캐릭터의 상호작용 모션을 리타게팅할 때 네 개의 캐릭터가 관련되며, 각각 소스 캐릭터와 그에 대응하는 타겟 캐릭터로 구성된다. 그림 3(a)와 같이, SCT는 각 캐릭터의 모션 잔여값을 예측하기 위해 협력하는 두 개의 대칭적 트랜스포머 기반 네트워크로 구성되어 있다. 각 트랜스포머 브랜치는 N+1개의 관절 상태를 입력으로 받고, 선형 계층을 통해 관절 토큰을 추출한 후, 자기 및 교차 어텐션 모듈을 통해 상호작용 특징을 추출한다. 최종적으로 MLP를 통해 상호작용 특징을 디코딩하여 관절 회전 및 루트 위치의 잔차값을 얻는다. 그림 3(b)와 같이, 어텐션 모듈은 다중 헤드 어텐션, 정규화 계층, MLP로 구성되며, 트랜스포머 기반 다중 헤드 어텐션을 통해 쿼리, 키, 값에서 특징을 도출한다. SCT는 두 캐릭터의 모션 시퀀스 간 관계를 모델링하기 위해 상호작용 특징을 공유하며, 첫 번째 어텐션 계층은 자기 어텐션 연산을 통해 포즈의 공간적 특징을 추출하고, 교차 어텐션을 통해 상호작용하는 캐릭터의 공간적 특징을 추출한다.

2.3. 앵커 및 손실함수

그림 4과 같이, 앵커는 캐릭터의 메시 상에 드문드문 배치된 대표적인 정점으로, 다른 메시 토폴로지를 가진 캐릭터 간에도 유사한 신체 부위에 위치한다. 이에 따라 상호작용하는 소스 캐릭터들의 앵커 거리와 타겟 캐릭터들의 앵커 거리를 일치시켜, 리타게팅된 캐릭터들 사이의 침투를 방지하고 접촉을 유지하는데 활용한다. SCT의 학습에는 재구성 손실과 앵커 손실이 사용되며, 재구성 손실은 루트 위치와 관절 회전, 관절 위치를 복원하고, 앵커 손실은 소스 캐릭터의 앵커 거리 행렬과 타겟 캐릭터의 앵커 거리 행렬 간의 차이를 최소화하여 중요한 상호작용 의미를 유지한다.

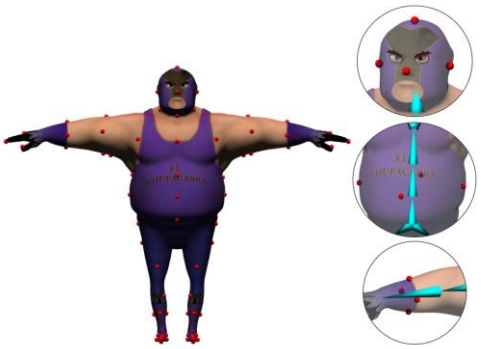


그림 4. 캐릭터의 메쉬 위 앵커 위치

3. 결론

SCT는 다른 메쉬 토폴로지를 가진 두 스킨 캐릭터 사이의 상호작용 동작을 재조정하는 새로운 네트워크로, 소스 및 타겟 캐릭터의 뼈대 구조와 형태 정보를 사용하여 모션 잔차를 추정한다. 이를 위해 새로운 표현인 앵커를 활용하며 캐릭터의 형태를 메쉬에 무관하게 표현 가능하게 한다. 모션 리타게팅의 성능을 향상시키기 위해 메쉬의 변형으로부터 모션 적응에 기반한 데이터 증가 방법을 제안한다. 한계점으로 리타게팅된 모션이 물리 법칙을 위반할 수 있고, 다른 뼈대 구조를 가진 캐릭터에 대해 동작을 리타게팅 할 수 없다. 이러한 한계를 극복하기 위해 물리 시뮬레이션을 동반하는 접근 방식과 다양한 뼈대에 대응 가능한 리타게팅 방법을 활용할 것을 기대한다.

참고문헌

[1] Edmond S. L. Ho, Taku Komura, and Chiew-Lan Tai. 2010. Spatial relationship preserving character motion adaptation. ACM Trans. Graph.
[2] Yunbo Zhang, Deepak Gopinath, Yuting Ye, Jessica Hodgins, Greg Turk, and Jungdam Won. 2023a. Simulation and Retargeting of Complex Multi-Character Interactions. In ACM SIGGRAPH 2023 Conference Proceedings.
[3] Taeil Jin, Meekyoung Kim, and Sung-Hee Lee. 2018. Aura Mesh: Motion Retargeting to Preserve the Spatial Relationships between Skinned Characters. Computer Graphics Forum.
[4] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: a skinned multi-person linear model. ACM Trans.
[5] Rami Ali Al-Asqhar, Taku Komura, and Myung Geol Choi. 2013. Relationship descriptors for interactive motion adaptation. In Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '13).

논문 발표

애니메이션/시뮬레이션2

멀티 터치 장치를 통한 캐릭터의 이동 동작 제어*

박근태[0], 이지원, 이윤상
한양대학교 컴퓨터소프트웨어공학과
qkrrmsxo01@hanyang.ac.kr, babap8514@gmail.com, yoonsanglee@hanyang.ac.kr

Interactive Character Locomotion Control via Multi-touch Device

Geuntae Park[0], Jiwon Yi, Yoonsang Lee
Dept. of Computer Science, Hanyang University

요약

멀티 터치 장치상에서 둘 또는 세 손가락의 터치 정보를 이용하여 캐릭터의 이동 동작을 제어하는 방법을 제안한다. 시스템은 기본적으로 사용자가 두 손가락으로 입력한 위치 정보의 변화로부터 적절한 전신 동작을 생성하며 추가로 다른 한 손가락의 입력 정보에 따라 캐릭터의 정면 방향 회전을 결정한다.

1. 서론

Finger-walking은 두 손가락을 이용하여 사람의 발과 다리의 움직임을 묘사하는 방법이다. 이는 걷기, 달리기 등 다양한 종류의 움직임뿐만 아니라 보폭, 발걸음 속도 등 발동작의 스타일도 직관적으로 표현할 수 있다.

우리는 캐릭터의 이동 및 동작을 제어할 때 자주 사용되는 속도 및 목표 경로 등의 정보 대신 지면에 접촉 중인 발의 위치 정보로부터 전신 동작을 생성하는 방법을 제안한다. 먼저, 사용자는 멀티 터치 장치상에서 Finger-walking을 수행하여 매 시점 지면에 접촉 중인 발의 위치 정보를 시스템에 입력한다. 시스템에 입력된 발의 위치는 각 시점에서의 캐릭터에 상대적인 위치 변화로 해석된다. 그런 다음 시스템에서는 위치 변화 정보를 통해 사용자가 입력한 발의 상대 위치 정보를 지키면서도 이전까지의 동작과 이어지는 다음 자세를 만들어낸다. 따라서 사용자는 캐릭터의 자세한 발걸음을 직접적으로 제어 및 조작할 수 있다.

두 발의 위치 변화 정보만으로는 캐릭터의 정면 방향을 제어하기 어려우므로 이를 제어하기 위한 하나의 터치 입력을 시스템에 추가했다. 사용자는 다른 한 손가락을 수평으로 이동시켜 캐릭터의 왼쪽 및 오른쪽 회전을 제어할 수 있다.

* 구두 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 및 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

2. 네트워크

제안하는 시스템의 네트워크는 크게 이동 네트워크와 포즈 네트워크로 구성되며 각각 다음 시점에서의 캐릭터의 이동과 자세를 출력한다. 자연스럽게 이어지는 동작을 만들어내기 위해 이동 및 포즈 네트워크 모두 Gated Recurrent Units (GRU)를 사용하여 과거의 정보들도 효과적으로 활용하도록 했다.

2.1. 이동 네트워크

이동 네트워크는 2개의 GRU 계층으로 구성된다.

이동 네트워크는 최근 $n+1$개 프레임에서의 루트의 수평 위치 이력 $p_{t-n:t}$와 최근 n개 프레임과 다음 프레임에서의 발 접촉 정보 이력 $f_{t+1-n:t+1}$을 입력 받아 다음 프레임에서의 루트의 수평 위치 p_{n+1}을 출력한다. $p_{t-n:t}$ 및 p_{n+1}은 모두 현재 시점의 캐릭터 좌표계 기준으로 표현되며 $f_{t+1-n:t+1}$은 각 정보마다 대응되는 시점에서의 캐릭터 좌표계를 기준으로 표현된다.

2.2. 포즈 네트워크

포즈 네트워크는 [1]과 같이 Mixture of Experts (MoE) 구조를 가지나 과거 정보를 활용하기 위해 gating 네트워크에 GRU를 적용했다. Gating 네트워크는 1개의 GRU 계층으로 구성되며 generator는 2개의 완전 연결 은닉 계층을 가진 K개 expert들로 구성된다 (실험에서는 $K=8$ 사용).

Gating 네트워크는 발 접촉 정보 이력 $f_{t+1-n:t+1}$만을 입력으로 받아 expert들에 대한 혼합 가중치를 출력한다. Generator의 매개변수는 gating 네트워크가 출력한 가중치에 따라 expert들의 매개변수들을 합한 값으로 결정되며 다음 프레임에서의 발 접촉 정보 f_{t+1}과 정면 방향 d_{t+1} 및 캐릭터 상태 이력 $s_{t-n:t}$을 입력 받아 다음 프레임에서의 캐릭터 상태 정보 $\tilde{s}_{t+1}$를 출력한다. d_{t+1}, $s_{t-n:t}$ 및 $\tilde{s}_{t+1}$은 현재 시점의 캐릭터 좌표계 기준으로 표현되며 나머지는 이동 네트워크와 동일하게 표현된다.

3. 학습

학습 시 이동 및 포즈 네트워크가 동시에 학습된다.

학습에 사용된 전체 손실 함수는 다음과 같다.

$$L = w_1 \cdot L_r + w_2 \cdot L_{FK} + w_3 \cdot L_d + w_4 \cdot L_c + w_5 \cdot L_{ct} \quad (1)$$

위 식에서 L_r은 네트워크의 출력과 원본 데이터 간의 평균 제곱 오차로 계산된다.

L_{FK}는 forward kinematics(FK) 계층[2]을 사용하여 네트워크가 예측한 포즈 및 이동으로부터 각 관절의 다음 시점에서의 예측 위치를 구하고 실제 다음 시점 관절들의 위치와 평균 제곱 오차를 계산한 값이다.

L_d은 생성된 자세의 정면 방향이 입력 받은 목표 방향을 따르도록 학습시키는 손실함수로 입력으로 들어온 목표 정면 방향과 네트워크가 예측한 포즈의 정면 방향 사이의 평균 제곱 오차를 계산한다.

앞서 설명한 손실함수들을 통해 입력에 따른 모션 생성을 학습시킬 수 있으나 생성되는 모션의 발 미끄러짐 현상이 심각한 것을 확인했다. 이에 우리는 발 미끄러짐을 방지하기 위해 추가로 L_c와 L_{ct}를 학습에 적용했다.

L_c는 다음과 같이 표현된다.

$$L_c = \left\| (\mathbf{f}_t^c \odot \mathbf{f}_{t+1}^c) \odot \mathbf{MSE}_{\mathbf{feet}}(\mathbf{f}_{t+1}^{FK}, \mathbf{f}_t^p) \right\|_1 \quad (2)$$

여기서 $\odot$은 요소별 곱을 나타내며 $\mathbf{MSE}_{\mathbf{feet}}$은 두 발 각각에 대한 2차원 수평 위치 정보를 받아 각 발에 대한 평균 제곱 오차를 출력한다. $\mathbf{f}_{t+1}^{FK}$은 FK layer를 통해 얻은 현재 시점의 캐릭터 좌표계로 표현된 다음 시점에서의 예상 발 위치를 나타내며 $\mathbf{f}_t^c$와 $\mathbf{f}_t^p$는 각각 입력 받은 두 발의 접촉 상태와 위치를 나타낸다.

L_{ct}는 다음과 같다.

$$L_{ct} = \left\| (\mathbf{f}_t^c \odot \mathbf{f}_{t+1}^c) \odot \mathbf{MSE}_{\mathbf{feet}}(\hat{\mathbf{f}}_{t+1}^p, \mathbf{f}_t^p) \right\|_1 \quad (3)$$

여기서 $\hat{\mathbf{f}}_{t+1}^p$은 $\mathbf{f}_{t+1}^p$를 네트워크가 예측한 이동 및 포즈로부터 구한 캐릭터 좌표계를 역연산하여 t 프레임에서의 캐릭터 좌표계 기준으로 변환한 값이다. L_c와 L_{ct}는 이동 네트워크 및 포즈 네트워크의 출력이 입력된 발의 상태 변화를 보다 잘 반영하도록 학습시킨다.

4. 실험 결과

실험을 위해 LaFAN[3] 데이터들 중 걷기, 달리기, 점프 같은 평평한 지면에서의 동작이 담긴 데이터를 선정하여 약 3시간 네트워크를 학습했다.

실험에서 멀티 터치 모니터를 사용하여 다양한 스타일의 finger-walking들과 목표 방향 정보를 입력했으며 시스템에서 장치 상에 입력한 발의 위치 변화 및 목표 방향 입력을 반영하여 동작들을 생성하는 것을 확인했다 (그림 1 참조).

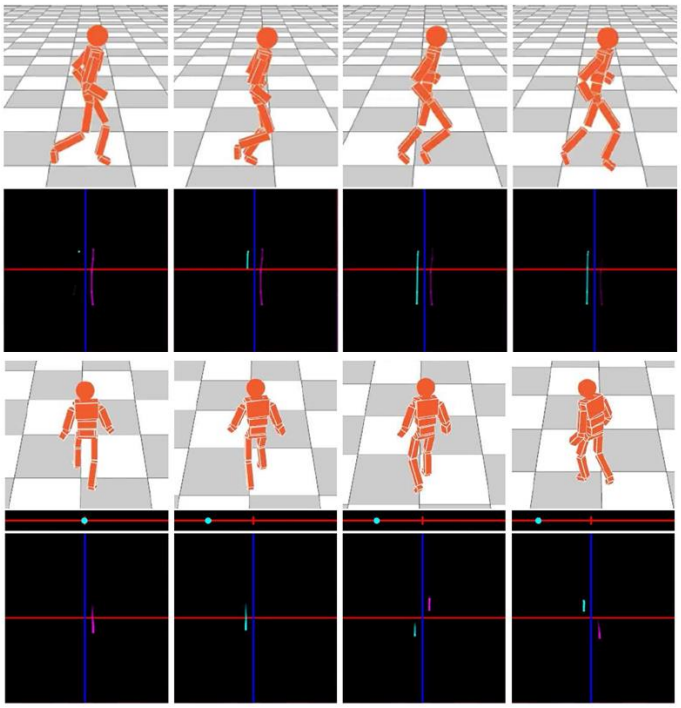


그림 1: 터치 입력에 따라 생성된 결과 모션. 접촉 해제 시에도 입력 경로의 시각화를 위해 이전 입력 위치가 점차 연한 색으로 렌더링 됨. 초록색은 왼발, 붉은 색은 오른발 터치 입력 궤적을 의미. (상단: 달리는 동작, 하단: 목표 방향 입력에 따라 회전하며 걷는 동작)

5. 결론

우리는 멀티 터치 장치상에 사용자가 입력한 finger-walking으로부터 캐릭터의 전신 동작을 생성하는 방법을 제안하였다. Finger-walking으로 전신 동작을 생성하는 기존 연구[4]는 장치상의 터치 위치를 전역 공간의 좌표 값으로 지정했기 때문에 캐릭터의 활동 영역이 터치 장치의 넓이에 제한되는 반면 우리의 연구는 캐릭터 좌표계에 대한 상대적인 발 위치로 표현함으로써 작은 영역에서도 캐릭터의 이동을 제어할 수 있다.

참고문헌

[1] He Zhang, Sebastian Starke, Taku Komura, and Jun Saito. 2018. Mode-adaptive Neural Networks for Quadruped Motion Control. ACM Trans. Graph. 37, 4, Article 145 (July 2018), 11 pages.

[2] Ruben Villegas, Jimei Yang, Duygu Ceylan, and Honglak Lee. 2018. Neural kinematic networks for unsupervised motion retargetting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '18).

[3] Félix G. Harvey, Mike Yurick, Derek Nowrouzezahrai, and Christopher Pal. 2020. Robust Motion In-Betweening. ACM Trans. Graph. 39, 4, Article 60 (July 2020), 12 pages.

[4] Noah Lockwood and Karan Singh. 2012. Finger walking: motion editing with contact-based hand performance. In Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '12).

단일 2D RGB 영상을 이용한 보행주기 탐지 네트워크 *

김대용[0,1], 이연수[2], 김유겸[2], 신정환[3], 유리[1,2,*]
아주대학교 인공지능학과[1], 아주대학교 소프트웨어학과[2], 서울대학교병원 신경과[3]
so155s@ajou.ac.kr, dldustnajoo@ajou.ac.kr, dbrua1222@ajou.ac.kr, neo2003@snu.ac.kr, riyu@ajou.ac.kr

The Novel Gait Cycle Detection Network (GCDN) for Video Based Gait Analysis Based on Single 2D RGB Video

Dae-Yong Kim[0,1], Yeon-Su Lee[2], Yu-Kyeom Kim[2], Jung-Hwan Shin[3], Ri Yu[1,2,*]
Dept. of Artificial Intelligence, Ajou University[1],
Dept. of Software and Computer Engineering[2],
Dept. of Neurology, Seoul National University College of Medicine[3]

요약

최근 단안 비디오로부터 3차원 보행 동작을 재건 및 분석하는 연구가 진행되고 있다. 보행 분석을 하기 위해서는 정확한 보행 주기를 추정해야 하고, 이를 위해서는 발과 지면 간의 접촉 여부를 정확히 얻어내야 한다. 본 연구는 지도학습을 사용하여 학습한 보행주기 탐지 네트워크 (GCDN)를 제안한다. 보행 동작을 모션 캡쳐 시스템과 RGB 카메라로 동시에 촬영한 데이터셋[1]을 사용하여 학습데이터를 구축하였다. 모션 캡쳐 데이터로부터 발 관절 접촉여부에 대한 ground-truth 값을 생성하고 네트워크를 학습하였다. 학습된 GCDN은 비디오에서 얻어낸 관절 좌표를 입력 받으면 각 발의 뒤꿈치와 엄지발가락의 지면 접촉 여부를 출력한다. 존재하지 않는 레이블을 추정해서 네트워크의 학습에 이용하는 해당 방법에 대한 결과의 유의성을 확인할 수 있었다.

1. 서론

보행 분석은 다양한 분야에서 필수적이다. 예를 들어, 파킨슨병과 같은 신경계와 관련된 질환의 경우에는 보행장애와 같이 운동기능과 관련한 증상들이 주로 나타나는데, 병의 진전에 대한 판단이나 치료 계획을 수립하는 데 있어서 보행주기의 정량적인 평가는 필수적이다. 보행 분석 시, 발과 지면이 접촉하는 양상에 따라 구분된 보행 주기를 파악하는 것은 보폭과 같은 보행 지표를 추출하기 위해 필수적인 작업이다. 이전 연구[2]에서는 포즈 추정 기법을 사용하여 얻어낸 발 관절 위치를 입력으로 하여 신호 처리 방식을 통해서 발과 지면의 접촉 정보를 추정하였다. 이 방식은 2D 영상에서 추출한 신호에 노이즈가 존재하기 때문에 상대적으로 낮은 정확도를 보인다.

본 연구는 단안 영상에서 발과 지면 사이의 접촉 여부를 추정하는 Gait Cycle Detection Network (GCDN)을 제안한다. 발과 지면 간의 접촉 여부를 탐지하는 네트워크를 학습하는 데 있어서 양질의 레이블이 존재하는 학습 데이터는 필수적이다. 동일 동작에 대해 모션 캡쳐 데이터와 영상 데이터를 함께 수집한 데이터셋[1]을 사용하여 GCDN의 학습데이터를 구축하였다.

본 연구는 크게 두 단계로 나뉜다. 첫 번째는 데이터셋 전처리 단계이다. 3D 모션 캡쳐 데이터에서 가속도 기반 알고리즘을 통해 레이블 값을 얻어내고, 2D 영상에서 목표가 되는 사람의 2D 관절 좌표 데이터를 OpenPose[3]를 이용하여 추출한다. 두 번째 단계에서는 첫 번째 단계에서 구축한 학습데이터를 기반으로 GCDN 학습을 진행한다.

성능 평가를 위해 GCDN의 결과와 GT값을 비교하여 일치도를 측정하였다. 본 연구는 단안 비디오로부터 정확한 보행주기를 얻어내기 위한 초석으로 비디오 기반 보행분석 연구에 기여할 수 있을 것이라 기대한다.

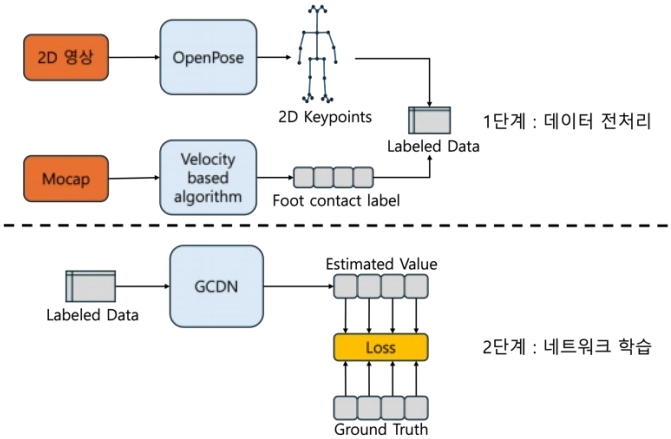


그림 1: 전체적인 흐름도

2. 방법

전체적인 과정에 대해서 데이터 전처리와 네트워크 학습의 2단계에 걸쳐서 기술한다. 첫번째 단계는 데이터 전처리 과정으로 네트워크의 입력에 적절한 형태로 데이터를 가공하고 데이터에 대한 라벨링을 하는 것이 목표이다. 두번째 단계에서는 가공한 데이터를 이용하여 RNN기반의 GCDN의 학습을 진행한다 (그림 1 참조).

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, MDS 2024에 게재확정 되었음.
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구결과로 수행되었음(IITP-2024-RS-2023-00255968)

2.1. 학습 데이터셋

학습으로 이용하는 데이터셋[1]의 일부를 활용하는데, 하나의 동일한 동작에 대해서 3D 모션 캡쳐 데이터와 서로 다른 시점에서 촬영된 2개의 영상이 존재한다. 해당 연구에 사용한 영상의 fps는 30이고, 전체 영상의 길이는 25.46분으로 모두 합쳐서 45834개의 프레임으로 구성되어 있다. 해당 데이터셋에는 발과 지면 사이의 접촉에 대한 데이터가 존재하지 않기 때문에, 3D 모션 캡쳐 데이터를 이용하여 가속도를 기반으로 레이블 값을 구한다. 가속도가 일정 이하인 경우에 목표가 되는 부분과 지면이 접촉했다고 판단한다 (그림 2 참조).

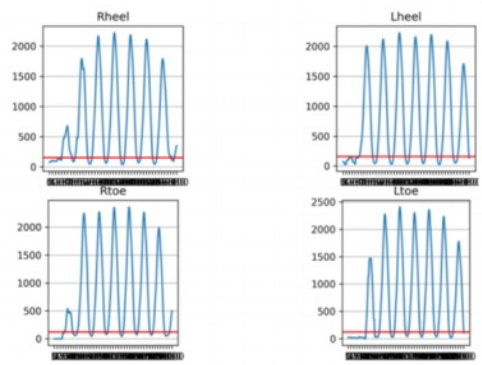


그림 2: GT 값 습득 방식

OpenPose[3]를 이용하여 목표가 되는 사람에 대한 2가지 시점의 비디오에서 2D 관절 좌푯값을 습득한다. 비디오에서 얻은 관절의 2D 좌푯값 중에서 하체에 해당하는 관절 좌푯값과 GT 값을 이용하여 네트워크의 학습에 사용한다.

2.2. 네트워크 학습

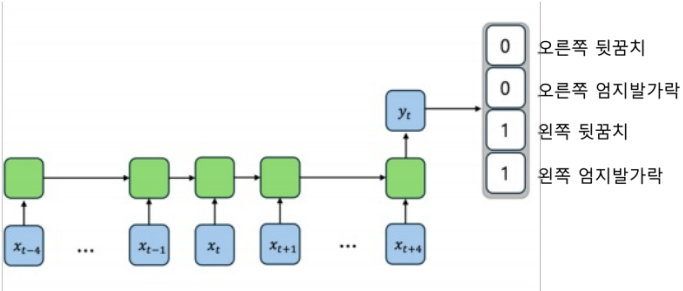


그림 3: 네트워크 구조

이 단계에서는 입력 값으로 하체 부분의 2차원 관절 좌표 $\boldsymbol{P} \in \mathbb{R}^{(lower\ joints \times 2) \times 9}$ 로 모델 f를 학습시켜 각 발의 엄지 발가락과 뒷꿈치의 접촉 여부 $\hat{Y} \in \mathbb{R}^4$에 대한 결과값을 얻게 된다. 영상의 t 프레임에 해당하는 발과 지면 사이의 접촉 여부를 판단하기 위해서 [t−4, t+4] 범위의 연속적인 프레임 데이터를 네트워크의 입력값으로 사용한다 (그림 3 참조).

3. 결과

표 1: 성능 평가표

video_name	Left toe	Left heel	Right toe	Right heel
gait1_off_rear	0.76	0.76	0.79	0.78
gait2_off_front	0.75	0.73	0.68	0.67
gait2_off_rear	0.61	0.59	0.52	0.52
gait2_on_front	0.42	0.45	0.48	0.49
gait2_on_rear	0.38	0.37	0.40	0.39
walk_fast_on_rear	0.64	0.62	0.42	0.43
walk_preferred_on_front	0.67	0.68	0.68	0.71
walk_preferred_on_rear	0.61	0.62	0.65	0.61
walk_slow_on_rear	0.70	0.69	0.58	0.61
walk_stroop_on_front	0.55	0.53	0.60	0.56
walk_stroop_on_rear	0.54	0.54	0.55	0.54
walk_turn_off_front	0.56	0.58	0.54	0.53
walk_turn_off_rear	0.62	0.63	0.62	0.61
Mean	**0.60**	**0.60**	**0.58**	**0.57**
Standard deviation	**0.11**	**0.11**	**0.11**	**0.11**

보행 장애가 있는 환자 영상에 대해서 네트워크 성능 평가를 진행했다. 평가 지표로는 GCDN의 결과값과 모션 캡쳐 데이터를 통해서 얻은 GT값 사이의 일치도를 측정했다. GCDN의 결과값과 GT값 사이에는 58.8%의 일치도를 확인할 수 있었다. 각 키포인트(왼쪽 엄지발가락, 왼쪽 발뒤꿈치, 오른쪽 엄지발가락, 오른쪽 발뒤꿈치)에 대한 일치도의 값은 0.60 ± 0.11, 0.60 ± 0.11, 0.58 ± 0.11, 0.57 ± 0.11 이다 (표 1 참조).

4. 결론 및 한계

보행주기를 판별하는 네트워크인 GCDN을 소개하였다. 본 연구는 보행 운동 장애가 나타나는 환자의 보행주기를 정량적으로 평가하고 적절한 치료 계획을 수립하여 치료 효과에 대한 모니터링에 기여할 수 있을 것이라 생각한다. 의료 분야 뿐만 아니라 캐릭터 애니메이션, 로보틱스와 같이 다양한 분야에서 활용 가능하다.

본 연구는 몇 가지 한계점이 존재한다. 해당 연구에서는 지면 접촉에 대한 ground-truth (GT)를 구하기 위해 가속도 기반의 휴리스틱한 방법론을 사용하였다. 비교적 정확한 센서 데이터로부터 GT를 추정하긴 하였으나 최종 정확도에 영향을 미칠 수 있다. 또한, 네트워크 학습에 사용된 데이터 양이 적었고, 따라서 다양한 보행에 대한 성능을 검증하는 데 한계가 있었다.

후속 연구에서는 보다 정확한 레이블 추정 방법을 이용해서 GT의 신뢰성을 높이고, 다양하고 큰 규모의 보행 데이터를 사용하여 GCDN의 성능을 향상시킬 예정이다.

참고문헌

[1] WARMERDAM, Elke, et al. “Full-Body Mobility Data to Validate Inertial Measurement Unit Algorithms in Healthy and Neurological Cohorts,” *Data*, 7(10):136, 2022.

[2] Shin JH, Yu R, Ong JN, et al. “Quantitative Gait Analysis Using a Pose-Estimation Algorithm with a Single 2D-Video of Parkinson's Disease Patients,” *J Parkinsons Dis,* 11(3):1271-1283, 2021.

[3] CAO, Zhe, et al. “Realtime multi-person 2d pose estimation using part affinity fields,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7291-7299, 2017.

상호작용 중심 환경 표현을 통한 데이터 기반 모션 리타게팅 프레임워크*

이예원[1], 진태일[2], 이성희[1]
[1]한국과학기술원 문화기술대학원, [2]한국과학기술연구원
yeone22@kaist.ac.kr, jin219219@kist.re.kr, sunghee.lee@kaist.ac.kr

A Data-Driven Framework for Motion Retargeting with Interaction-Oriented Environment Representation

Yewon Lee[1], Taeil Jin[2], Sung-Hee Lee[1]
[1]Graduate School of Culture Technology, KAIST, [2]Korea Institute of Science and Technology

요약

본 연구는 데이터 기반 방법을 통해 환경에 적응하는 모션을 효과적으로 생성하는 새로운 방식의 리타게팅 프레임워크를 제안한다. 이를 위해 환경을 이해하고 모션과 환경 간의 상호작용을 설명하는 상호작용 중심 환경 표현을 도입한다. 제안된 프레임워크는 소스 모션과 소스 가구와의 상호작용을 타겟 가구에 적응하여 타겟 모션으로 리타게팅한다. 두 개의 조건부 가변 자동 인코더 네트워크를 사용하여 환경의 중요도를 학습하고, 예측된 중요도에 따라 리타게팅된 모션을 예측한다. 본 프레임워크는 타겟 장면이 소스 장면과 일치하지 않거나 타겟 환경에 가구가 없는 경우에도 효과적으로 모션을 리타게팅 한다.

1. 서론

환경과 상호작용하는 모션을 리타게팅하는 작업은 다양한 가구에 적응, 접촉 조건 만족, 원본 모션의 상호작용 보존 등의 도전 과제가 있다. 많은 연구에서 최적화 기법[1-3]을 사용하여 환경 적응 리타게팅을 시도하지만, 소스 환경과 타겟 환경이 크게 다를 경우 이러한 시도는 종종 어려움을 겪는다. 이 문제를 해결하기 위해 본 연구는 데이터 기반 접근 방식을 채택한다. 환경과 상호작용하는 모션을 생성하는 기존 데이터 기반 모델[4-6]의 환경 표현은 캐릭터의 신체 부위별 환경 정보를 제공하지 않기 때문에 세분화된 리타게팅이 어렵다. 따라서 본 연구에서는 모션과 환경의 상호작용을 더욱 정확하게 설명하는 새로운 환경 표현을 제안한다.

* 구두 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 타 학술대회(논문지)에 발표 예정.
* 본 연구는 한국전파진흥협회 메타버스랩지원사업 지원으로 수행되었음.

2. 방법론

2.1. 상호작용 중심 환경 표현

상호작용 중심 환경 표현은 환경을 위치, 라벨, 중요도로 총 3가지의 구성 요소로 정의한다. 캐릭터의 주변에 가상의 원통형 공간을 가정하고, 이 원통형 공간의 반지름을 7등분, 높이를 20등분, 각도를 20등분으로 나눈다. 그림 1(a)와 같이, 원통형 공간의 상단 원 평면에서 각도를 20등분하여 20개의 반지름을 얻는다. 이 반지름을 다시 7등분하여 총 140개의 위치에서 지면을 향해 수직으로 광선을 방출한다. 또한, 원통형의 높이를 20등분하여 각 높이에서 상단 원의 반지름과 평행한 방향으로 20개의 광선을 방출한다. 이 과정을 통해 평행으로 400개의 광선을 방출하게 된다.

총 540개의 광선을 통해 프레임마다 환경과 교차하는 위치를 탐색한다. 환경과 교차하지 않는 광선에는 임의의 값을 할당한다. 그림 1(b)는 탐색 된 지점들을 주황색 구로 시각화한다. 각 지점에 지면, 가구의 종류, 교차하지 않는 지점 순으로 one-hot 벡터의 형태로 라벨을 지정한다. 상호작용을 모델링하기 위해 신체를 여섯 부위, 즉 손, 발, 척추, 골반으로 나눈다. 신체 부위와의 근접성을 기준으로 각 지점의 중요도 가중치를 계산하는데, 특정 지점과 특정 신체 부위 사이의 거리가 캐릭터 키의 4분의 1보다 멀면 중요도가 0이 되고 가까울수록 1에 가까운 값을 갖는다.

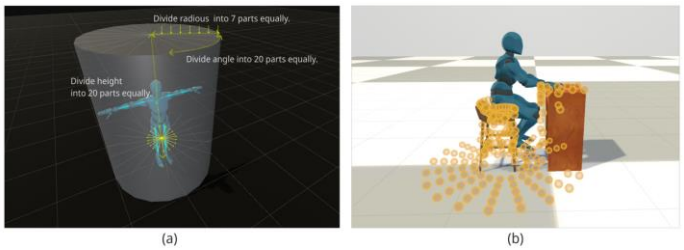


그림1: 540개의 광선을 통한 환경 정보 탐색

2.2. 모션 리타게팅 프레임워크

본 연구는 소스 모션을 타겟 환경에 맞게 조정하여 모션을 리타게팅하는 새로운 방식의 프레임워크를 제안한다. 이 프레임워크는 크게 3가지 구성 요소로 이루어져 있다: 두 개의 생성 모델과 하나의 최적화 모델이다. 본 연구는 생성 모델을 통한 환경 적응에 중점을 두고 연구하며, 이에 대해 서술한다. 최적화 모델은 소스의 상호작용을 정교하게 유지하도록 한다.

두 생성 모델인 ImportanceNet과 MotionNet은 모두 조건부 가변 자동 인코더(cVAE) [7] 구조를 사용한다. 먼저 ImportanceNet을 훈련하고 이를 고정한 후, MotionNet을 훈련하는 두 단계의 과정을 거친다. 소스에서 과거 10프레임의 루트(골반의 위치를 지면으로 투영)의 속도와 각 신체 부위와 가구 사이의 현재 접촉 정보를 추출하고, 타겟에서 환경의 현재 위치와 라벨 정보를 추출한다. 입력 데이터에는 이전 프레임의 주요 관절(joint=15) 포즈 정보도 포함된다. ImportanceNet은 입력 데이터를 기반으로 환경의 중요도를 예측한다. MotionNet은 입력 데이터와 예측된 환경 중요도를 기반으로 리타게팅된 주요 관절 포즈를 예측한다. 이 리타게팅된 포즈는 다음 프레임을 예측하기 위한 포즈 정보로 자동 회귀적으로 사용된다.

두 생성 모델을 통해 생성된 환경에 적응한 주요 관절 모션은 최적화 모델의 초기 값으로 사용되며, 최적화 모델은 전신(joint=25) 모션을 계산한다. 이때, 최적화 모델은 하반신에 대해서는 생성된 모션을 따르되, 상반신에서는 그림 3과 같이 주어진 가중치에 따라 소스 모션이나 생성된 모션을 따르도록 한다.

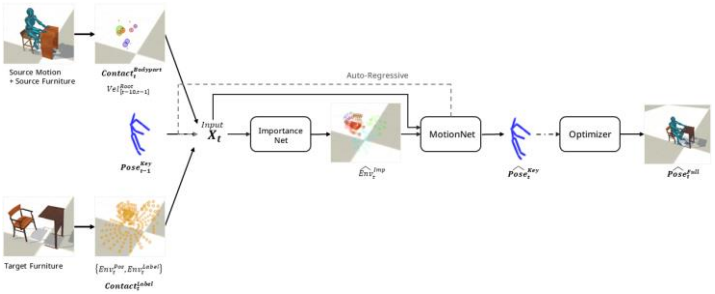


그림 2: 제안된 프레임워크를 통한 모션 리타게팅 과정 개요

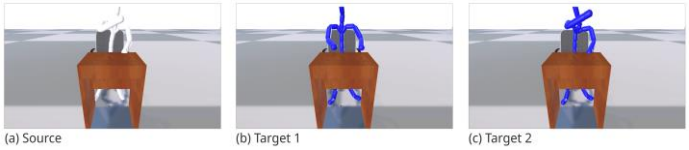


그림 3: 최적화 모델의 가중치에 따라 (b) 생성된 모션의 상반신, (c) 소스 모션의 상반신을 따르는 결과

3. 결과

그림 4-6에서 흰색 캐릭터는 소스 모션, 파란색 캐릭터는 리타게팅된 모션을 나타낸다. 그림 3은 높은 의자, 높은 책상과 상호작용하는 하나의 모션을 다양한 환경으로 리타게팅한 결과를 보여준다. 가구의 일부 또는 전부를 제거하거나 책상과 의자의 높이를 변경하는 등 다양한 가구 구성에 적응한다. 그림 4는 가구의 라벨이 달라지는 경우, 이에 적응하는 결과를 보여준다. 그림 5는 학습 데이터에 사용되지 않은 새로운 가구에 대해서도 적응할 수 있음을 보여준다. 이러한 결과는 본 프레임워크가 환경의 변화에 대해 강인하게 적응하는 모션 리타게팅 프레임워크임을 입증한다.

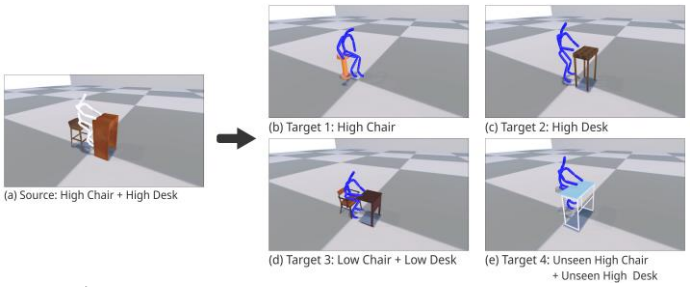


그림 4: 다양한 가구 구성에 따른 모션 리타게팅 결과

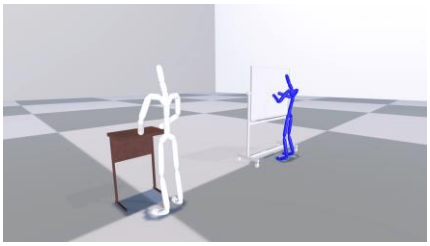


그림 5: 책상에서 칠판으로 모션 리타게팅 한 결과

그림 6: 높이가 다른, 처음 보는 가구에 모션 리타게팅 한 결과

참고문헌

[1] S. Tonneau, N. Mansard, A. Herzog, J. Vaillant, M. Botsch, and J. Pettré, Character contact re-positioning under large environment deformation, *Computer Graphics Forum*, vol. 35, no. 2, pp. 455-463, 2016.
[2] S. Choi, D. Lee, H. Shin, and S.-H. Lee, Online Avatar Motion Adaptation to Morphologically-similar Spaces, *Computer Graphics Forum*, vol. 42, no. 2, pp. 153-164, 2023.
[3] Y. Kim, J. Lee, S. Ha, and J. Kim, Retargeting human-object interaction to virtual avatars, *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2405-2412, 2016.
[4] M. Hassan, J. Chibane, A. Zanfir, C. Tang, and A. Black, Stochastic scene-aware motion prediction, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11124-11133, 2021.
[5] S. Starke, H. T. Holden, and J. Komura, Neural state machine for character-scene interactions, *ACM Transactions on Graphics*, vol. 38, no. 6, pp. 1-14, 2019.
[6] T. Jin and S.-H. Lee, DAFNet: Generating Diverse Actions for Furniture Interaction by Learning Conditional Pose Distribution, *Computer Graphics Forum*, vol. 42, no. 7, pp. 25-35, 2023.
[7] K. Sohn, H. Lee, and X. Yan, Learning structured output representation using deep conditional generative models, *Advances in Neural Information Processing Systems*, vol. 28, pp. 3483-3491, 2015.

유체-옷감의 상호작용에 의한 파동 난류, 확산, 주름을 표현하는 GPU 프레임워크*

이주용[0,1], 박은수[0,1], 김종현[2], 박인규[1]
[1]인하대학교 정보통신공학과, [2]인하대학교 디자인테크놀로지학과
{juyong9864, qkrdmstn1014}@inha.edu, {jonghyunkim, pik}@inha.ac.kr

A GPU Framework for Representing of Wave Turbulence, Diffusion and Wrinkling in Fluid-Cloth Interaction

Juyong Lee[0,1], Eun Su Park[0,1], Jong-Hyun Kim[2], In Kyu Park[1]
[1]Dept. of Information and Communication Engineering, Inha University,
[2]College of Software and Convergence (Dept. of Design Technology), Inha University

요약

본 논문에서는 액체-옷감 상호작용 시 난류, 확산, 주름 등의 현상을 표현하는 GPU 프레임워크를 제안한다. 기존의 방법은 운동량 보존 법칙에만 의존하여 상호작용을 표현하기 때문에 세부적인 시각적 특징을 놓치는 경우가 있다. 따라서, 본 논문에서는 난류, 확산, 주름 등의 시각적 특징을 표현하는 것에 초점을 맞추고 있다. 이러한 특징들은 연산이 많아 CPU에서 처리하기 어렵기 때문에, GPU 기반의 그래프 컬러링과 메모리 최적화를 통해 유체/옷감 시뮬레이션 및 상호작용을 병렬 처리하여 성능을 개선한다.

1. 서론

액체-옷감의 상호작용을 표현하기 위한 다양한 물리 기반 시뮬레이션 기법들이 있다[4]. 그러나 대부분 액체-옷감 사이의 운동량만을 집중적으로 다루고 있기 때문에 시각적 특징이 제대로 표현되지 못하는 경우가 있다. 본 논문에서는 상호작용을 통해 얻어지는 추가적인 특징인 물방울에 의한 파동 난류, 중력에 의한 확산, 젖은 옷감의 기하학 특징을 고려한 주름 등을 다루며, 이 과정에서 증가되는 연산을 해결하기 위해 GPU 기반의 새로운 프레임 워크를 제안한다.

2. 제안하는 프레임워크

본 연구에서는 액체-옷감 상호작용을 다루기 위해 다음과 같은 과정을 따른다(그림 1 참조).

액체를 표현하기 위한 기반 시뮬레이션은 FLIP (Fluid-Implicit Particle)과 Mercier 등이 제안한 난류(turbulence)를 활용하며[1] 옷감의 경우 PBD (Position-Based Dynamics) 기법과 Patkar 등이 제안한 젖은 옷감을 활용한다[3].

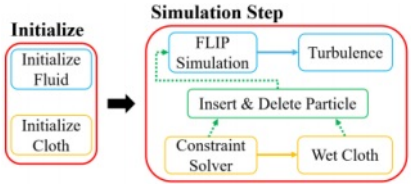


그림 1: 유체-옷감 시뮬레이션 과정

2.1. 파동 난류가 포함된 액체 시뮬레이션 GPU 병렬화

본 논문에서 초점을 두고 있는 파동 난류의 병렬 처리 부분은 크게 두 가지로 나뉜다 (그림 2 참조).

난류 입자의 초기화 및 표면 유지 과정을 병렬화 하기 위해 GPU 기반 메모리 최적화 과정을 거쳐 매 프레임 개수가 달라지는 난류 입자의 삽입/삭제 과정을 효율적으로 설계한다 (그림 2(a) 참조).

파동 시뮬레이션은 각 난류 입자마다 스레드를 할당하여 파동의 높이를 계산하고, 최종적으로 시각적으로 확인 가능한 새로운 입자를 생성한다 (그림 2(b) 참조).

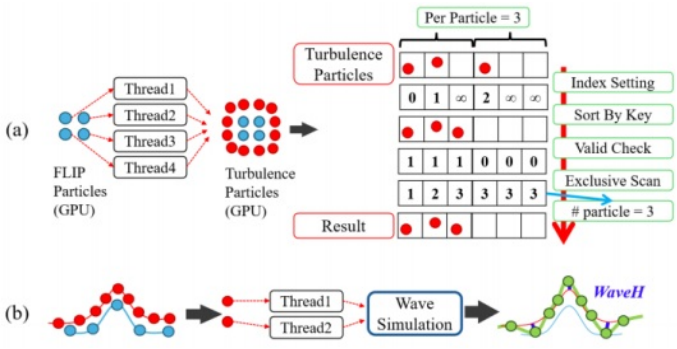


그림 2: 액체의 파동 난류 병렬화 과정

* 구두 발표논문
* 학부생 주저자 논문임.
* 이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 한국연구재단과 정보통신기획평가원의 지원을 받아 수행한 연구임(No. RS-2023-00254695, 2022R1F1A1063180, No.RS-2022-00155915, 인공지능융합혁신 인재양성사업(인하대학교)).

2.2. 수분이 포함된 옷감 시뮬레이션 GPU 병렬화

본 연구의 PBD 병렬화 과정에서 제약(constraint)은 각 스프링(spring)마다 스레드를 할당하여 병렬화 하며 경합 조건(race condition)을 피하기 위해 [2]에서 제안하는 그래프 컬러링을 활용한다 (그림 3(a) 참조).

유체-옷감 상호작용의 흡수, 확산, 물방울 떨어짐 현상은 [3]에서 제안한 모델을 따른다. [3]에서 면(face)에 포화도를 누적하는 형식의 모델을 제안하기 때문에 면마다 스레드를 할당한다 (그림 3(b) 참조).

젖은 옷감의 주름 표현은 다양한 방식으로 구현될 수 있다. 본 연구에서는 인접 정점(vertex)과의 곡률을 고려하여 주름을 부풀듯이 과장하여 표현한다. 이때 곡률을 고려하기 위해 에지(edge)마다 스레드를 할당하여 이면각(dihedral angle)을 계산하고 이를 가중치로 하여 정점마다 스레드를 할당해 주름 표현을 위한 속도를 계산한다 (그림 3(c) 참조).

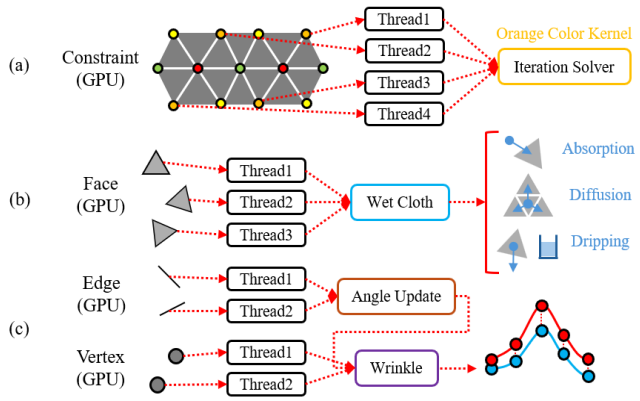


그림 3: 수분이 포함된 옷감 시뮬레이션의 병렬화 과정

3. 실험 결과

CPU와 GPU에서 frame/sec로 측정하여 비교해 보았을 때 CPU 대비 유체 시뮬레이션의 경우 약 30~70배 (표 1 참조), 옷감 시뮬레이션의 경우 약 2.5~70배 향상되었다 (표 2 참조). 두 시뮬레이션 모두 장면의 복잡도가 증가할수록 높은 성능 향상을 보였으며, 유체-옷감 상호작용 장면의 성능 측정 결과 약 1.15 frame/sec로 측정되었다 (표 3 참조). 결과적으로, GPU를 활용한 최적화를 통하여 각 시뮬레이션의 성능 향상을 이뤄낼 수 있었다.

Scene		*Water Drop*	*Dam Break*	*Double Water Drop*
GPU	*# FLIP Particle*	20k	24k	135k
	# Turbulence Particle	11k	16k	15k
	frame/sec	2.72	1.76	0.83
CPU	*# FLIP Particle*	20k	24k	47k
	# Turbulence Particle	8k	15k	14k
	frame/sec	0.09	0.035	0.012

표 1: 유체 시뮬레이션 성능 비교(GPU : RTX 6000 Ada / CPU : AMD Ryzen Threadripper PRO 7975WX 32-Core)

Model	*Dragon*		*Bunny*	
# Vertex/Face	50k	100k	4.8k	2.4k
# Structural/Bend	150k	150k	7.2k	7.2k
CPU/GPU (frame/sec)	0.98	68.3	36.5	92.4

표 2: 옷감 시뮬레이션 성능 비교(GPU : RTX 3080 / CPU : Intel 13700K)

Scene		*Dripping*	
Fluid	*# FLIP Particle*	135k	
	# Turbulence Particle	197k	
Cloth	*# Vertex/Face*	18k	36k
	# Structural/Bend	54k	54k
Result	*frame/sec*	1.15	

표 3: 유체-옷감 상호작용 성능 측정(GPU : RTX 6000 Ada / CPU : AMD Ryzen Threadripper PRO 7975WX 32-Core)

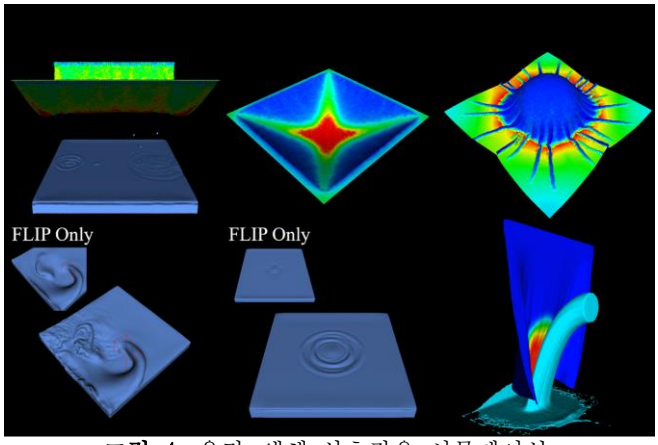


그림 4: 옷감-액체 상호작용 시뮬레이션.

4. 결론

본 논문에서는 유체/옷감 시뮬레이션에서 CPU 대비 수십 배의 성능 향상을 이루었다. 이를 통해 기존 연구에서 성능적인 제약으로 인해 표현하기 어려웠던 파동 난류, 확산, 주름 등의 세부적인 현상을 표현하는 장면을 제작할 수 있었다 (그림 4 참조).

향후, 실시간 시뮬레이션을 위해 다중 GPU를 활용한 프레임워크 제작을 진행할 예정이다.

참고문헌

[1] O. Mercier, C. Beauchemin, N. Thuerey, T. Kim, and D. Nowrouzezahrai, “Surface turbulence for particle-based liquid simulations,” *ACM Trans. on Graphics*, 34:6, 1-10, 2015.

[2] M. Fratarcangelin, F. Pellacini, “A GPU-based implementation of position based dynamics for interactive deformable bodies,” *Journal of Graphics Tools*, 17:3, 59-66, 2013.

[3] S. Patkar, P. Chaudhuri, “Wetting of porous solids,” *IEEE Trans. on Visualization and Computer Graphics*, 19:9, 1592-1604, 2013.

[4] Y. Fei, C. Batty, E. Grinspun, and C. Zheng, “A Multi-scale model for simulating liquid-fabric interactions,” *ACM Trans. on Graphics*, 37:4, 1-16, 2018.

물리기반 캐릭터의 축구 드리블 스킬 학습*

김민수 [0], 이윤상
한양대학교 인공지능학과, 한양대학교 컴퓨터 소프트웨어학과
igotaspot426@gmail.com, yoonsanglee@hanyang.ac.kr

Learning Football Dribble Skills for Physics Based Character

Minsu Kim[0], Yoonsang Lee
Dept. of Artificial Intelligence, Hanyang University, Dept. of Computer Science, Hanyang University

요약
우리는 물리 시뮬레이션 되는 휴머노이드 캐릭터가 축구공으로 드리블할 수 있도록 하는 정책을 학습하는 방법을 제안한다. 다양한 축구 모션 캡처 데이터를 잠재공간에 임베딩하고, 상황에 따라 적절한 잠재변수를 활용해 공과 복잡한 상호작용하는 드리블 정책을 학습시켰다.

1. 서론
드리블은 축구선수에게 기본적이면서 중요한 스킬이다. EA SPORTS FC, eFootball와 같은 축구 게임에서도 축구선수 캐릭터가 드리블하는 것이 게임 플레이의 주요 부분 중 하나이지만, 완전한 물리 시뮬레이션에 기반하지 않다 보니 선수의 동작과 공의 움직임이 실제 축구와는 차이가 있다. 완전한 물리 시뮬레이션을 통해 계산된 캐릭터의 동작과 축구공과의 움직임을 통해 드리블 스킬을 구현하게 되면 이러한 축구 게임의 사실성 향상에 크게 도움이 될 수 있을 뿐만 아니라, 물리 시뮬레이션에 기반한 축구선수의 퍼포먼스 분석 및 예측을 가능하게 하는 등 스포츠 사이언스 관점에서의 기여도 가능할 것이다. 이전에도 물리기반 캐릭터가 축구공을 드리블하는 것을 보여준 연구들이 있었지만, 걸으면서 공을 다루거나 [1] 동작이 부자연스럽거나 [2] 축구공과 선수 사이의 거리를 가깝게 유지하지 못하는 [3] 등의 한계점들이 있었다. 이에 우리는 심층강화학습을 통해 물리기반 캐릭터가 기민하게 드리블을 할 수 있도록 하는 정책을 학습하는 방법을 제안한다. 특히 게임과 같은 인터랙티브 애플리케이션에 적용가능 하도록 우리는 사용자의 실시간 입력에 기반하여 조종되는 캐릭터가 공을 드리블할 수 있도록 하는 것에 초점을 맞추었다.

* 구두 발표논문, 요약논문 (Extended Abstract)
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 과 정보통신기획평가원의 지원(No.RS-2020-II201373,인공지능대학원지원(한양대학교)) 과 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

2. 물리기반 모션 임베딩 모델
물리기반 모션 임베딩 모델은 잠재공간에 모션 시퀀스를 임베딩하는 모델이다. 모션 시퀀스는 주로 모션 캡처 데이터에서 추출되고 각 시퀀스는 잠재공간의 특정 포인트에 의해 표현된다. 우리는 물리기반 모션 임베딩 모델 학습을 하기 위해서 저수준 정책과 인코더, 판별자를 학습하는 CALM [4]의 방법을 사용하였다. 인코더는 고차원의 모션 시퀀스를 입력으로 받아 저차원의 잠재변수로 변환하여 출력하는 역할을 한다. 저수준 정책은 인코더로부터 만들어진 잠재변수와 캐릭터의 각 관절의 위치, 각도, 속도, 각속로 이루어진 캐릭터의 상태가 합쳐진 벡터를 정책의 상태로서 입력으로 받고 시뮬레이션에 필요한 관절의 목표 각도를 정책의 액션으로서 출력한다. 판별자는 시뮬레이션 된 모션과 원본 모션 시퀀스와 차이를 구분하지 못하면 높은 보상을 얻게 되는데, 이 보상 신호를 이용해 인코더와 저수준 정책, 판별자가 한꺼번에 end-to-end로 학습된다. 물리기반 모션 임베딩 모델의 학습이 완료되면 모션 시퀀스를 잠재변수로 변환하는 인코더의 역할은 더 이상 필요하지 않으므로 런타임에는 인코더를 제거해도 되며, 저수준 정책 모델에는 캐릭터의 상태와 학습된 잠재변수만을 입력으로 넣어서 캐릭터를 컨트롤 할 수 있다.

3. 드리블 정책 학습
학습에 사용된 축구 스킬 모션 캡처 데이터는 공과 상호작용하는 부분의 정보가 없는 데이터다. 따라서 모션이 임베딩 되어도 캐릭터가 축구공과 물리적으로 상호작용하고 컨트롤하는 능력은 추가로 학습을 해야 한다. 그렇기 때문에 모델을 계층적으로 구성하고 드리블하는 정책을 학습시켰다.

고수준 정책 모델은 입력으로 저수준 정책에서 사용하는 캐릭터의 상태에 더해 추가로 유저 입력 (목표 속력, 목표 방향)을 정책의 상태로서 입력받고 잠재변수를 정책의 액션으로서 출력한다. 이 고수준 정책 모델은 미리 학습된 저수준 정책모델에 연결되어 계층적인 모델 구조를 갖는다. 이 두 모델은 함께 시뮬레이션 되며 학습하지만 저수준 정책의 파라미터는 고정되어 고수준 정책 모델만 학습이 된다.

고수준 정책을 학습하는데 사용된 보상 함수는 다음과 같다.

$$r_t = 0.3 \exp\left(-10\left(\mathbf{x}_t^{root} - \mathbf{x}_t^{ball}\right)^2 \right) + 0.3 \exp\left(-\left(\mathbf{v}_t^{tar} - \mathbf{v}_t^{*ball}\right)^2 \right) + 0.4\, \mathbf{d}_t^{tar} \cdot \mathbf{d}_t^{ball}$$

$\mathbf{x}_t^{root}$와 $\mathbf{x}_t^{ball}$ 는 각각 캐릭터와 공의 위치이며 $\mathbf{v}_t^{tar}$ 와 $\mathbf{v}_t^{*ball}$ 는 각각 목표 드리블 속력과 공의 목표 드리블 방향으로의 속력이다. $\mathbf{d}_t^{tar}$ 와 $\mathbf{d}_t^{ball}$는 각각 목표 드리블 방향, 공의 이동 방향을 의미한다.

4. 실험

학습환경은 Isaac Gym을 사용했고 빠른 학습을 위해 4096 개의 환경을 병렬적으로 실행시켜 데이터를 수집하고 policy 를 학습했다. 학습은 에피소드 단위로 진행되는데 에피소드 길이는 30 fps 시뮬레이션 환경에서 300 프레임으로 고정한다. 목표 속력과 목표 방향을 랜덤으로 설정한다. 150 ~ 200 프레임 시점에 목표 방향과 목표 속력을 한 번 재설정하는데, 이를 통해 방향 및 속력 변화에 자연스럽게 대응하는 능력을 획득하게 하였다.

자연스러운 축구공 컨트롤을 위해 실제와 비슷한 공의 물리량을 설정하였다. 축구공의 크기는 공식 경기에 사용되는 크기인 22cm이고, 질량은 200g으로 설정하였으며, 반발계수는 0.8 로 설정하였다.

드리블을 잘 하기 위해서 발의 모양을 적절한 메쉬로 설정할 필요가 있었다. 저수준 정책 모델을 학습할 때 사용한 휴머노이드의 발 메쉬 부분을 축구화 메쉬로 대체하였다. 축구화 메쉬가 아닌 단순한 박스 형태의 발 메쉬로 학습하면 드리블을 전혀 학습하지 못하는 것을 확인했다.

모션 저수준 정책을 학습할 때 사용한 데이터셋은 총 89 개 모션, 약 6 분 분량의 축구 스킬 모션 캡처 데이터로, 걷기, 뛰기, 슈팅, 드리블 등이 포함되어 있다.

우리는 보상 설계의 차이에 따라 드리블 능력을 실험했다. 최대 목표 드리블 속력을 5m/s, 7.5m/s, 10m/s 로 설정하고 학습한 후 비교하였다. 원본 모션 캡처 데이터에 달리는 모션의 최대 속력이 초속 약 5 m/s 라서 학습할 때도 최대 속력을 5 m/s로 기본값으로 설정하였다. 학습할 때 최대 속력을 7.5 m/s 로 설정하면 드리블 능력을 유지하는 선에서 최대 약 6 m/s까지 드리블하는 모습을 보여준다. 하지만 학습할 때 최대 속력을 10 m/s 로 설정하면 드리블을 전혀 하지 못하는 것을 확인했다.
학습한 후에는 게임패드를 연결하고 고수준 정책에 유저 입력 (목표 방향, 목표 속력)으로 입력해줘서 직접 컨트롤 할 수 있다. 그림 1 은 유저 입력에 의해 드리블 스킬이 시뮬레이션 되고 있는 모습을 보여주고 있다.

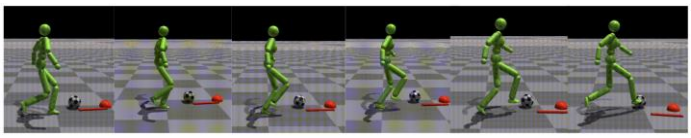


그림 1: 런타임에 시뮬레이션 되고 있는 드리블 스킬

5. 결론

우리는 심층강화학습을 통해 물리기반 캐릭터의 사실적이고 민첩한 축구 드리블 정책을 학습하는 방법을 제안했다. 적절한 보상설계와 학습 환경 구성으로 유저 입력에 민첩하게 반응하는 드리블 스킬을 배우게 할 수 있음을 보였다. 추후에는 더 다양한 축구 스킬들을 학습하고 물리환경에서 플레이할 수 있는 축구게임을 위한 기반 연구를 고려할 수 있다.

참고문헌

[1] Xue Bin Peng, Glen Berseth, Kangkang Yin, and Michiel Van De Panne, DeepLoco: dynamic locomotion skills using hierarchical deep reinforcement learning. ACM Trans. Graph. 36, 4, Article 41 (August 2017), 13 pages, 2017
[2] S. Liu, G. Lever, Z. Wang, J. Merel, S. M. A. Eslami, D. Hennes, W. M. Czarnecki, Y. Tassa, S. Omidshafiei, A. Abdolmaleki, N. Y. Siegel, L. Hasenclever, L. Marris, S. Tunyasuvunakool, H. F. Song, M. Wulfmeier, P. Muller, T. Haarnoja, B. D. Tracey, K. Tuyls, T. Graepel, and N. Heess, From motor control to team play in simulated humanoid football, Science Robotics, vol. 7, no. 69, 2022
[3] Seokpyo Hong, Daseong Han, Kyungmin Cho, Joseph S. Shin, and Junyong Noh, Physics-based full-body soccer motion control for dribbling and shooting. ACM Trans. Graph. 38, 4, Article 74 (August 2019), 12 pages, 2019
[4] Chen Tessler, Yoni Kasten, Yunrong Guo, Shie Mannor, Gal Chechik, and Xue Bin Pen, CALM: Conditional Adversarial Latent Models for Directable Virtual Characters. In ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23). Association for Computing Machinery, Article 37, 1–9, 2023

논문 발표

가상/증강현실1

확장현실 협업 환경을 위한 장면 합성 프레임워크 설계*

나기리[0,1], 김진모[1,2]
한성대학교 일반대학원 컴퓨터공학과[0,1], 한성대학교 컴퓨터공학부[2]
{nagiri0209, jinmo.kim}@hansung.ac.kr

Design of Scene Synthesis Framework for Collaborative XR Environment

Giri Na[0,1], Jinmo Kim[1,2]
Department of Computer Engineering, Graduate School, Hansung University[0,1],
Division of Computer Engineering, Hansung University[2]

요약

본 연구는 현실과 가상이 동일한 좌표를 기준으로 완벽히 대응되는 확장현실 협업 환경을 구축하기 위해 현실 기반의 가상 장면을 생성하는 장면 합성 프레임워크를 제안한다. 프레임워크는 현실 공간에 대한 3차원 정보를 파악하기 위한 기하학적 분석과 현실 공간을 이루는 객체를 정의하기 위한 의미론적 해석 과정을 통해 현실 공간을 탐색한다. 그리고 탐색 된 현실 공간 정보를 토대로 가상 장면을 생성한다. 이를 통해, 혼합현실 사용자와 가상현실 사용자가 유기적으로 협업하는 확장현실 체험 환경을 제작하고 가상현실 사용자를 대상으로 설문 실험을 진행하여 제작한 체험 환경의 유의미성을 확인한다. 실험을 통해 현실을 기반으로 만들어진 가상 장면과 유기적인 상호작용을 지원하는 체험 환경이 사용자들에게 긍정적인 영향을 미친다는 것을 확인하였다.

1. 서론

혼합현실과 가상현실 사용자가 함께 협업하는 확장현실 환경은 디지털 트윈 등의 방향으로 다양한 산업에 응용되어 여러 이점을 제공한다[1]. 이러한 환경에 필요한 가상 장면을 직접 제작하는 것은 시간 소모가 크기 때문에 알고리즘을 활용하여 효율적으로 가상 장면을 생성하는 장면 합성에 관한 연구가 진행되었다[2]. 하지만, 이는 생성된 장면 결과에만 주목하고 있고, 현실 정보를 고려하지 않는다. 이와 달리, 현실 공간 정보를 기반으로 만들어진 가상 환경을 증강현실 사용자 콘텐츠에 통합하는 시스템을 제안하는 연구도 있었다[3].

본 연구는 현실 세계에 존재하는 혼합현실 사용자와 가상 세계에서 행동하는 가상현실 사용자가 서로 정확하고 유기적으로 협업할 수 있는 확장현실 환경을 제작하기 위한 장면 합성 프레임워크를 설계한다. 이는 서로 다른 공간(현실과 가상)에 있다 하더라도 같은 장면을 공유함으로써 같은 공간에 있는 듯한 느낌을 주고, 사용자의 몰입도와 흥미를 높이는 등의 이점을 제공한다.

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출 예정.
* 교신저자: 김진모(jinmo.kim@hansung.ac.kr)

2. 장면 합성 프레임워크 설계

장면 합성 프레임워크는 원활한 상호작용이 가능한 확장현실 환경을 지원하기 위해 현실과 가상의 각기 다른 기준으로 해석되던 공간을 동일한 좌표계를 기준으로 통일시켜서 하나의 협업 공간으로 구성하는 과정이다. 현실 공간 탐색을 위해 마이크로소프트 홀로렌즈(Microsoft HoloLens) 2와 혼합현실 개발 도구인 MRTK(Mixed Reality Toolkit) 2를 사용한다. 탐색한 현실 공간 정보는 유니티(Unity) 3D 엔진으로 통합하여 가상 공간을 계산하고 가상 객체를 자동으로 배치하는 데 사용된다. 그림 1은 제안하는 장면 합성 프레임워크의 전체적인 구조를 보여준다.

2.1. 현실 공간의 기하학적 분석

현실 공간의 좌표를 기준으로 가상 공간의 기하학적인 정보를 계산하기 위해서 Scene Understanding SDK를 통해 현실 세계의 3차원 기하학적 공간 정보를 추론한다. 홀로렌즈 2에 통합된 작업 환경을 실행하면 Scene Understanding의 장면 스캔이 동작하고 사용자가 위치한 좌표를 기준으로 그 주위를 이루는 현실 공간의 기하학적 정보(벽, 바닥, 객체 등)를 탐색하여 메쉬와 SceneObject 라벨 정보를 반환한다. 해당 정보를 통해 가상 장면의 전체적인 구조와 범위를 계산할 수 있다.

2.2. 객체의 의미론적 해석

현실 공간의 기하학적 분석으로 반환된 정보에는 현실 공간을 이루는 객체의 세부적인 유형이 포함되어 있지 않기 때문에 각 객체를 분석 및 분류하는 의미론적 해석 과정을 수행한다. 객체 분석을 위해, 본 연구는 Azure에서 제공하는 AI 서비스인 커스텀 비전(Custom Vision)을 사용해서 이미지 분류 모델을 생성하였다. 모델과 홀로렌즈 2를 연동하고, 객체 인식이 필요한 현실 장면 이미지 분석을 요청하면 모델이 장면 이미지 안에 인식된 객체 정보를 반환한다. 다음으로, 분석된 객체 결과를 토대로 가상 공간에 배치할 현실 객체를 분류하여 객체 이름으로 된 태그 라벨을 배치하는 라벨링 과정을 수행한다. 분류된 객체 태그 라벨에 대한 정보와 라벨이 위치한 3차원 변환 정보는 가상 공간에 가상 객체를 자동으로 배치하는 데 활용된다.

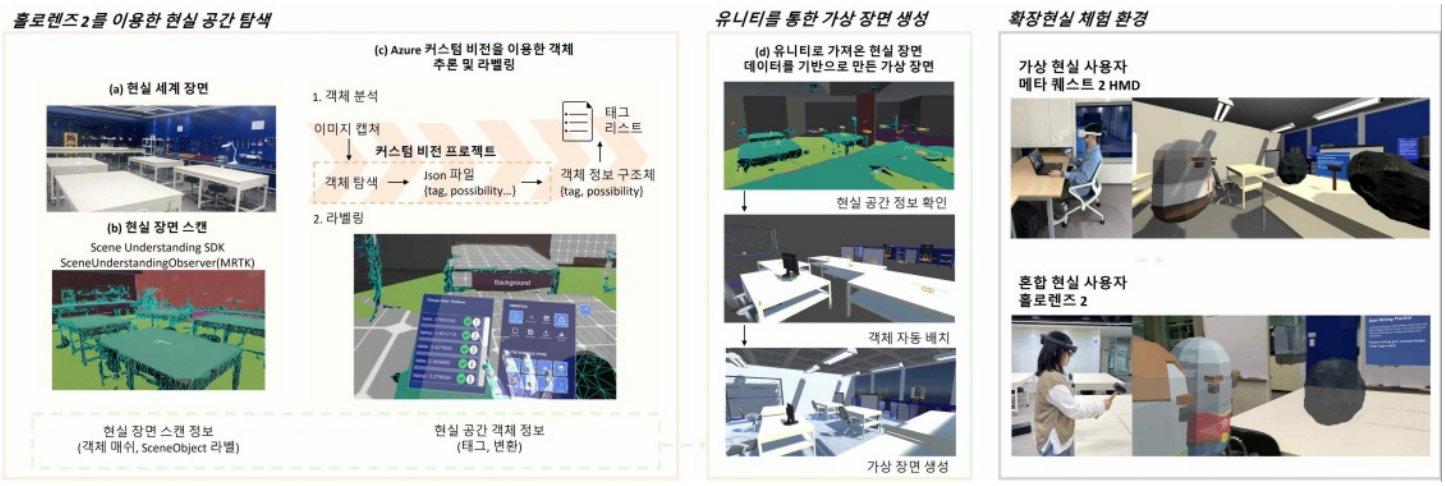


그림 1: 제안하는 장면 합성 프레임워크의 전체적인 구조

2.3. 객체 자동 배치

현실 공간의 기하학적 분석을 통해 얻은 현실 장면 정보(메쉬, SceneObject 라벨)와 객체의 의미론적 해석을 통해 추출한 객체 정보(태그, 3차원 변환 값(위치))를 토대로 가상 장면을 구성한다. 먼저, 객체 정보를 바탕으로 가상 객체를 자동으로 배치한다. 이때 가상 객체는 위치만을 고려하기 때문에 객체의 크기 및 방향 값은 장면 정보를 기반으로 재조정해준다. 마지막으로 완성도를 위해 조명 및 추가 객체를 배치해서 현실 장면 기반의 가상 장면을 생성한다.

3. 확장현실 협업 환경 구축

본 연구는 제안하는 프레임워크를 통해 현실과 대응되는 가상 공간을 제작하여 확장현실 체험 환경을 구축하는 것이 궁극적인 목표이다. 이를 위해, 3D 프린터 실습교육과 광물 채석 교육을 주제로 하는 협업 환경에서의 교육 콘텐츠를 제작한다. 제안하는 프레임워크를 통해 분석한 현실 공간은 크리에이티브 스튜디오로, 교육을 진행하기 어려운 환경에 있는 학생들이 현실 기반의 몰입도 높은 교육 활동을 수행할 수 있는 새로운 응용 방향을 제시하기 위하여 선택하였다. 사용자는 각각 홀로렌즈 2와 오큘러스 퀘스트 2 HMD를 통해 참여하고 사용자가 상호작용하는데 사용된 객체는 대화형 객체로 분류되어 동기화된다. 혼합현실 사용자는 콘텐츠에 교육자로 역할로, 대화형 객체로 정의한 현실 세계 도구를 직접 사용하여 교육을 주도한다. 가상현실 사용자는 피교육자로 참여해서 혼합현실 사용자가 보여주는 교육 활동에 집중하고, 현실 도구와 동일시되어 움직이는 가상 객체를 통해 현실감 있는 교육 체험이 가능하다.

4. 실험

제안하는 장면 합성 프레임워크와 제작한 확장현실 체험 환경을, 실험을 통해 분석하였다. 우선, 정량적 실험으로 프레임워크를 통해 가상 장면을 생성하기까지 걸린 시간을 기록했을 때, 8m x 10m 면적과 55개의 객체로 이루어진 현실 공간을 탐색하는데 걸린 시간은 9분 소요되었다. 다음으로, 가상 장면 생성 시간은 14분 소요되었다. 또한, 제작한 확장현실 체험 환경에 참여한 가상현실 사용자들을 대상으로 체험 환경이 현실 기반으로 제작된 것이라는 것을 알리기 전과 후로 나누어 인식 차이에 대한 유의미성을 알아보기 위해 비교 설문 실험을 진행하였다. 그 결과 현실과 대응된 가상 공간에서의 활동이 사용자들의 몰입감과 교육 활동에 대한 적응력을 높일 수 있다는 것을 확인하였다.

5. 결론

본 연구는 혼합현실 사용자와 가상현실 사용자가 함께 협업하는 확장현실 환경에서 사용자의 유기적인 상호작용을 지원하기 위해 현실 좌표를 기준으로 가상 공간을 해석하여 가상 장면을 효율적으로 생성하는 프레임워크를 설계하고, 현실과 가상이 대응되는 체험 환경을 제작하여 실험을 통해 분석해보았다. 혼합현실 사용자가 가상현실 사용자에게 사실적인 정보 전달이 가능하다는 점에서 협업 확장현실 환경의 발전 가능성과 효율성을 검증하였다. 향후, 새로운 형태의 디지털 경험을 제공하기 위한 연구를 지속할 예정이다.

참고문헌

[1] L.Qppermann and F.Buchholz and Y.Uzun, Industrial Metaverse: Supporting remote maintenance with avatars and digital twins in collaborative XR environments, *CHI EA '23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, Article 178, pp. 1-5, 2023.

[2] J.M.Sun and J.Yang and K.Mo and Y.K.Lai and L.Guibas and L.Gao, Haisor: Human-aware Indoor Scene Optimization via Deep Reinforcement Learning, *ACM Transactions on Graphics*, Article 15, pp. 1-17, 2024.

[3] X.Qian and F.He and X.Hu and T.Wang and A.Ipsita and K.Ramani, ScalAR: Authoring Semantically Adaptive Augmented Reality Experiences in Virtual Reality, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, Article 65, pp. 1-18, 2022.

동적 환경 적용을 위한 방향전환보행 전략의 선택적 전환

이호정[0,1], 전상빈[1], 조용훈[2], 이인권[1]
연세대학교 컴퓨터과학과[1], 고려대학교 디지털 체험 연구실[2]
dearshawn@yonsei.ac.kr, ludens0508@yonsei.ac.kr, maxburst88@gmail.com, iklee@yonsei.ac.kr

Redirection Strategy Switching: Selective Redirection Controller for Dynamic Environment Adaption

Ho Jung Lee[1], Sang-Bin Jeon[2], Yong-Hun Cho[3], In-Kwon Lee[4]
Dept. of Computer Science, Yonsei University[1,2,4], Digital eXPerience Lab, Korea University[3]

요약

본 연구에서는 가상현실을 체험하는 사용자의 주변 현실 및 가상 환경에 따라 최적의 방향전환 컨트롤러 (Redirection Controller: RC)를 선택하는 Selective Redirection Controller (SRC)를 소개한다. SRC는 강화학습을 통해 사용자 주변의 상황에 따라 네가지 RC (Steer-to-Center (S2C), Thomas et al.'s Artificial Potential Function (TAPF), Alignment-based Redirection Controller (ARC), Steering algorithm via Reinforcement Learning (SRL))들을 동적으로 전환하여 각 RC의 장점을 극대화하도록 학습되었다. SRC는 다양한 현실 및 가상 환경에서 시뮬레이션 및 사용자 실험을 통해 기존 RC들에 비해 더 적은 리셋 횟수를 보이며 우수한 성능을 보였다. SRC는 사용자의 주변 환경에 따라 RC들을 교차 적용하며 각 기술들의 장점을 극대화하여 사용자에게 우수한 경험을 제공한다.

1. 서론

방향전환보행 (Redirected Walking: RDW)은 가상현실을 체험하는 사용자의 화면을 미세하게 조정하여 한정된 현실 공간에서 가상 공간을 무한히 보행할 수 있게 한다 [1]. 하지만, 미세조정만으로는 현실공간의 벽이나 장애물과의 충돌을 피할 수 없을 때, 회전 이득 (rotation gain)을 활용하는 리셋 (reset)을 통해 사용자의 보행 방향을 재조정한다. 빈번한 리셋의 사용은 사용자의 몰입감을 저하시키기 때문에, 여러 전략들을 바탕으로 사용자를 최적의 방향으로 미세조정하기 위한 RC (S2C, TAPF[2], ARC[3], SRL[4])들이 연구되었다. 기존에 제안된 RC들이 좋은 성능을 보이는 물리 및 가상 환경은 각각 다르다. 본 연구에서는 강화학습을 통해 다

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract)으로서, 원본 논문은 IEEE VR 2024에 발표 및 IEEE Transactions on Visualization and Computer Graphics 30(5). 2024에 게재되었음.
* 이 연구는 정부 (과학기술정보통신부)의 재원으로 한국연구재단 (No. RS-2024-00348094) 및 한국전파진흥협회 (No. RNIX20230200)의 지원으로 수행되었음.

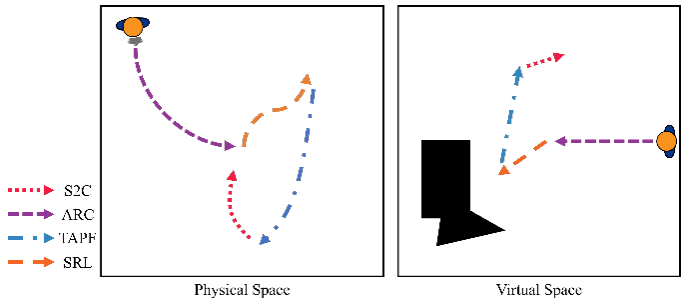


그림 1: SRC: 환경에 따라 최적의 RC를 적용

양한 환경에서 RDW를 체험하는 사용자의 주변 상황에 따라 S2C, TAPF, ARC, SRL 중 최적의 RC를 선택하도록 학습하는 SRC를 제안한다 (그림 1).

2. 방법

SRC 모델은 강화 학습의 Soft Actor-Critic 알고리즘을 통해 최적의 RC를 선택하여 사용자에게 적용하도록 설계되었다. 우리는 에이전트 (agent)를 아래와 같이 정의된 모델을 기반으로 다양한 환경에서 학습시켰다.

2.1. 상태 (State)

Agent가 시간 t에서 가지는 사용자 주변 정보를 o_t라 할 때, o_t는 현실공간의 정보 (o_t^p)와 가상공간의 정보 (o_t^v)로 다음과 같이 구성한다:

$$o_t = [o_t^p, o_t^v] = \left[\bigcup_{i=1}^{8}\{d_{i,t}^p\}, \bigcup_{i=1}^{8}\{d_{i,t}^v\}\right],$$

여기서 $d_{i,t}^p$와 $d_{i,t}^v$는 각각 현실 및 가상 공간의 사용자 위치에서부터 보행중인 방향을 기준으로 시계 방향으로 $\pi/4$ 각도 간격으로 측정된 보행 가능 경계선 또는 장애물까지의 최소 거리이다. i가 1일 때는 사용자가 걷고 있는 방향을 나타내고, i 값이 커질수록 시계 방향으로 $\pi/4$만큼 회전한 방향을 나타낸다 (그림 2). SRC는 64회에 걸쳐 관측된 o_t 들을 Long Short-Term Memory (LSTM) 레이어에 누적하여 상태 s_t를 구성한다:

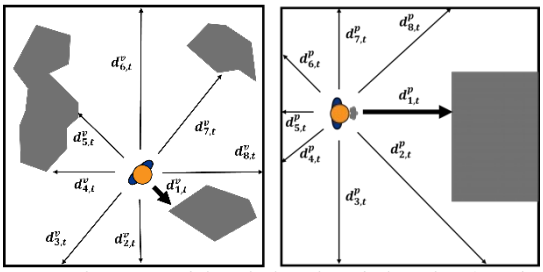


그림 2: SRC의 state 관측 과정 (좌: 현실공간, 우: 가상공간)

$$s_t = (o_{t-63}, o_{t-62}, \dots, o_{t-1}, o_t).$$

2.2. 행동 (Action)

SRC의 학습 목표는 사용자가 RDW를 수행하는 동안 구성된 s_t를 기반으로 사용자 주변 현실 및 가상환경의 상황에 따라 최적의 RC를 선택하여 사용자에게 적용하는 것이다. 이를 위하여 시간 t에 SRC agent가 RC 후보 중 하나를 결정하는 이산 (discrete) action을 수행하도록 a_t를 정의한다. 본 연구에서는 SRC가 S2C, TAPF, ARC, SRL 중 하나를 선택하도록 학습시켰다.

2.3. 보상함수 (Reward Function)

우리는 RDW를 경험하는 사용자가 현실 공간의 열린 공간을 탐색하면서 리셋을 최소화하는 정책을 SRC가 학습하는 것을 목표로 하므로, 보상함수는 다음과 같다:

$$R(s_t, a_t) = w_r R_r + w_a R_a + R_0,$$

여기서 R은 하위보상함수 R_r과 R_a, 기본 보상 $R_0\,(10^{-5})$, 가중치 w로 구성된다. R_r은 사용자가 리셋을 수행할 때 발생하는 벌점 (penalty)을 나타내며, -1로 설정된다. R_a은 사용자가 보행중일때 매시간 t마다 아래와 같이 계산된다:

$$R_a = d^p_{1,t} + \min\left[d^p_{3,t}, d^p_{7,t}\right],$$

여기서 사용자가 보행중인 방향으로 사용자 위치에서 장애물까지의 거리 ($d^p_{1,t}$)에 사용자의 왼쪽 ($d^p_{3,t}$)과 오른쪽 ($d^p_{7,t}$) 거리 중 짧은 것을 더하여 보상을 계산함으로써 사용자를 열린 공간으로 이동시키도록 학습된다.

3. 실험 및 결과

3.1. 시뮬레이션 실험

우리는 세 종류의 현실 환경과 네 종류의 가상 환경에서 시뮬레이션을 통해 SRC와 기존 RC (S2C, TAPF, ARC, SRL)를 각각 시뮬레이션 사용자에게 적용하였다. 시뮬레이션 사용자가 각 공간에서 500m를 보행할 때 발생한 리셋 횟수를 측정하였다. 수집된 데이터가 등분산과 정규성 검정을 통과하지 못하여 비모수검정인

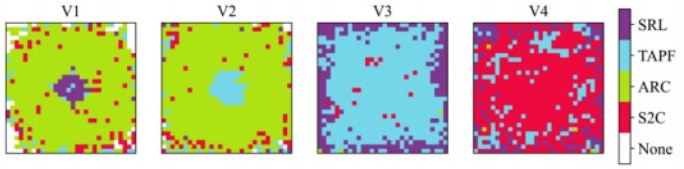


그림 3: 가장 많이 선택된 RC를 나타내는 히트맵

Kruskal-Wallis test와 Dunn's test를 통한 사후 검정을 통해 통계 분석을 수행했다. SRC는 모든 환경에서 유의미하게 기존 기법들보다 리셋 횟수를 감소시켰다. 우리는 각각의 환경별로 SRC가 현실 공간의 특정 위치에서 가장 많이 선택한 RC를 히트맵을 통해 시각화하였고 그 패턴을 분석하였다 (그림 3). SRC를 통해 상황에 따라 RC들을 교차 적용하는 것이 단일 RC를 사용하는 것보다 더욱 좋은 성능을 보임을 확인하였다.

3.2. 사용자 테스트

우리는 $6 \times 6m^2$의 장애물이 존재하지 않는 현실 공간과 여러 장애물들이 배치된 $50 \times 50m^2$의 유적지처럼 꾸며진 형태의 가상공간에서 실제 사용자 테스트를 수행하였다. 실험 참가자들은 SRL과 SRC를 각각 적용하여 가상 공간에서 80m를 보행하였다. 실험에서 측정된 데이터가 사전검정을 통과하지 못하여 비모수검정인 Wilcoxon signed-rank test로 확인한 결과, SRC가 더 적은 리셋 ($p < .001$), 더 긴 리셋간 평균 이동 거리 ($p < .001$), SSQ 설문 결과 더 적은 멀미 ($p < .01$)를 보였다.

4 토의 및 결론

우리는 가상 공간에서 자유롭게 보행하는 사용자에게 RC를 상황에 따라 교차 적용하여 여러 RC들의 장점을 최대화시키는 SRC를 제안한다. SRC는 RDW 체험 중 하나의 RC만을 사용하는 것보다 여러가지 환경에서 유의미하게 리셋 횟수를 감소시켰다. 추후 새로운 전략을 바탕으로 설계된 새로운 RC가 제안되었을 때 해당 알고리즘을 SRC의 선택에 추가함으로써 SRC는 자신의 성능을 더욱 향상시킬 잠재적 가능성을 지니고 있다.

참고문헌

[1] RAZZAQUE, Sharif. Redirected walking. The University of North Carolina at Chapel Hill, 2005.
[2] THOMAS, Jerald; ROSENBERG, Evan Suma. A general reactive algorithm for redirected walking using artificial potential functions. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019. p. 56-62.
[3] WILLIAMS, Niall L.; BERA, Aniket; MANOCHA, Dinesh. Arc: Alignment-based redirection controller for redirected walking in complex environments. IEEE Transactions on Visualization and Computer Graphics, 2021, 27.5: 2535-2544.
[4] STRAUSS, Ryan R., et al. A steering algorithm for redirected walking using reinforcement learning. IEEE transactions on visualization and computer graphics, 2020, 26.5: 1955-1963.

색채 심리를 활용한 VR 환경에서의 스트레스 완화 프로그램 개발*

김은지[0], 이채현, 권영남, 천애지, 방기석, 김선정§
한림대학교 융합소프트웨어학과[0], 한림대학교 소프트웨어학부
ej961@naver.com, cogusp1@gmail.com, {duddndgkt, aeji4123}@naver.com, {mysaver, sunkim}@hallym.ac.kr

Development of a Stress Relief Program Using Color Psychology in a VR Environment

Eun-Ji Kim[0], Chae-Hyne Lee, Young-Nam Kwon, Ae-Ji Cheon, Ki-Seok Bang, Sun-Jeong Kim,
Dept. of Convergence Software, Hallym University, Major of Contents IT, Hallym University

요약

현대인들은 일상생활에서 마주하는 다양한 상황으로 인해 스트레스와 불안감, 긴장감을 경험한다. 이러한 이유로 심리 치료를 받는 사람들이 점점 증가하고 있다. 심리 치료 프로그램에는 대면, 비대면을 포함한 다양한 방법이 있으며, 본 논문에서는 색채 치료(컬러 테라피)를 활용하여 마음의 안정을 추구하였다. 특히, VR 환경에서 노랑과 초록 색상의 테마를 구현하여 사용자가 집과 같은 편안한 환경에서 반복적으로 프로그램을 진행할 수 있도록 하여, 편리하고 안전하게 스트레스를 완화할 수 있도록 하였다.

1. 서론

1.1. 연구 배경과 목적

일상생활에서 겪게 되는 긴장감과 불안, 스트레스는 현대인의 삶에 부정적인 영향을 미칠 수 있는 요소로 여겨진다. 우울증이나 대인 기피증 같은 증상도 흔히 발생하는 병으로 자리 잡게 되었다[1]. 이를 완화시키기 위한 다양한 심리 상담 및 치유 프로그램이 많이 개발되고 있다. 전통적인 스트레스 관리법에는 꾸준한 운동과 대화 치료, 행동 요법을 포함한 심리 치료, 규칙적인 수면, 식습관 관리, 여가 활동과 명상 등이 있지만 적지 않은 시간과 비용을 요구하거나 개인의 환경에 따라 실천하기 어렵다는 문제점이 있다. VR(가상현실)을 활용한다면 사용자에게 현실적인 맞춤형 환경을 제공할 수 있으며, 개인에게 필요한 경험으로 환경을 조절할 수 있다. 본 연구는 HMD를 활용하여 색채 심리를 활용한 VR 환경을 구현하였다.

1.2. 색채 심리

색채 심리[2]에서는 색이 가지는 의미를 통해 심리 상태를 분석한다. 본 연구에서는 두 가지 색상 테마를 사용하여 콘텐츠를 구현하였다. 노란색은 상쾌, 찬란, 따뜻함, 즐거움 등의 이미지를 가지고 명랑하고 생동감 있는 느낌을 준다. 초록색은 스트레스와 격한 감정을 차분히 하도록 도와주고 심리적으로 자극을 거의 미치지 않는 색상으로 두 색을 선택하였다.

* 구두 발표논문
* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00254695).

2. 색채 심리 기반의 VR 환경 구현

2.1. 색채 심리와 컬러 테라피

컬러 테라피(Color Therapy)란 컬러와 테라피의 합성어로 색채를 통해 심리적인 문제를 치료하는 진단 방법이다[3]. 색상에 따라 의미를 해석하는 여러 방법이 있다[4]. 대표적으로 노란색은 긍정적이고 기쁜 마음이 생기게 한다. 초록색은 긴장을 완화시키고 스트레스를 해소하는 등의 효과가 있다는 것을 여러 연구에서 밝혀낸 바 있다. 본 논문에서는 색채 심리를 이용하여 스트레스를 완화하기 위한 심리적 안정 프로그램을 개발하고자 한다.

2.2. VR 환경 구현

본 연구에서는 HMD를 활용한 VR 환경에서 노랑, 초록 색상을 주제로 한 자연환경을 체험하게 된다. 사용자는 처음 VR 콘텐츠 체험 시 개개인에게 부여된 ID를 입력하게 된다. 총 3회의 실험을 진행하는 데, 사용자가 첫 번째 회기의 실험을 시작하기 전에 튜토리얼을 진행하도록 설계하였다(그림 1 참조).

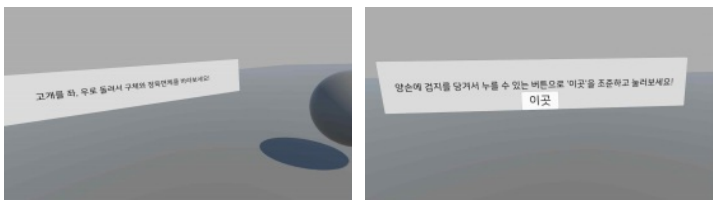


그림 1: 튜토리얼 진행 화면

사용자는 HMD와 컨트롤러 사용 방법을 숙지하기 위해 튜토리얼 단계를 거치게 된다. HMD를 착용한 상태로 시선을 돌려 주변을 바라보고 컨트롤러를 활용하여 이동하거나 물체를 터치하는 방법을 익히게 된다. 튜토리얼을 마치면 자연환경으로 이동된다. 사용자는 자연환경에서 주어지는 미션을 수행해야 한다(그림 2 참조). 미션은 총 세 가지로 소리에 집중하기, 심호흡하기, 사슴 찾기로 구성하였다. 미션은 회기마다 랜덤 한 순서로 제시되며, 하나의 미션을 완료한 뒤에 다음 미션으로 넘어갈 수 있다.

그림 2: 초록 색상 VR 콘텐츠 체험 화면

미션 사이에는 휴식을 위한 시간 간격이 있는데, 이때 사용자는 자유롭게 주변을 돌아다니거나 가만히 멈춰서 휴식을 취할 수 있다. 노란 색상에서도 동일한 미션이 제공된다(그림 3 참조). 미션 하나 당 약 30초에서 1분 정도의 시간을 가지고 진행되며, 멀미를 예방하기 위해 너무 길지 않은 시간으로 설정하였다.

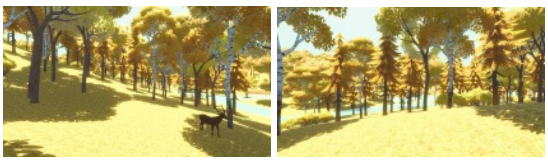


그림 3: 노란 색상 VR 콘텐츠 체험 화면

3. 연구 방법

3.1. 연구 설계 및 절차

본 연구는 만 18세 이상의 시각적인 능력에 이상이 없는 성인 38명을 대상으로 진행 예정이다. 초록 색상과 노란 색상에 각각 19명씩 나누어 진행하며, 서로 다른 날짜에 하루 한 번씩 3회 참여하게 된다. 색채 심리 기반의 VR 콘텐츠이므로, 시각적인 이상이 있는지 확인하기 위해 첫 번째 실험 전 색맹 테스트를 진행한다. 색맹 테스트 이상 유무와 상관없이 실험은 진행된다. 사전, 사후 설문으로 벡 불안척도[5] 검사를 시행하는데 본 검사의 점수 합계를 통해 불안 정도를 수치로 확인할 수 있다. 실험 참여 전, 후의 불안 정도를 확인하기 위해 실시하였다.

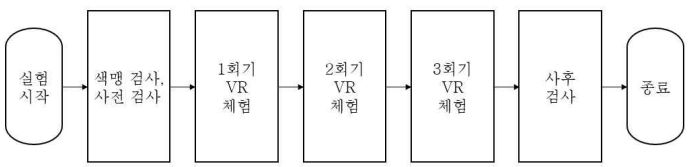


그림 4: 실험 절차

3.2. 측정 도구

색맹 판단을 위해 이시하라 색맹 테스트를 진행하였다. 이시하라 색맹 검사는 1917년 제작된 검사로, 약 100년 동안 꾸준히 사용되는 검사이다. 실험자들은 동일한 태블릿(Galaxy Tab A8)을 사용하여 색맹 테스트를 진행하였다. 사전, 사후에 실시하는 벡 불안척도(BAI, Beck Anxiety Inventory)란 불안에 대한 정도를 평가하기 위해 개발된 검사로 21개의 질문으로 구성되어 있다. 검사 문항에 대해 예, 아니요, 잘 모르겠다 등으로 간단히 답하는 자기 보고식 검사법이며, 각 질문은 0~3점 사이로 선택할 수 있다. 21개 질문의 점수를 더해 총점이 0~7점 사이이면 정상, 8~15점 사이면 가벼운 불안, 16~25 사이이면 중증도 불안, 26~63점 사이라면 심한 불안으로 평가한다.

4. 결론

4.1. 데이터 분석

본 실험에 앞서 10명의 시각적 이상이 없는 성인을 대상으로 파일럿 실험을 진행하였다. 색맹 테스트에서 1명은 색채 이상을 가지고 있는 것으로 확인되었다. 각기 다른 날짜에 3회 실험에 참여하여 진행하였다. 사전, 사후에 벡 불안척도 검사를 진행하고 수치를 비교한 결과 대부분은 불안이 감소하거나 유지되었지만, 수치가 소폭 상승한 경우도 있었다(표 1 참조).

표 1: 파일럿 테스트 결과

ID	1	2	3	4	5	6	7	8	9	10
색맹	정상	이상*	정상	정상	정상	정상	정상	정상	정상	정상
사전	0	18	11	16	0	6	4	14	4	0
사후	0	22	5	14	0	9	1	7	1	0

* 정답률 11% (심각한 색채 이상)

4.2. 결과 및 논의

파일럿 테스트에서 확인한 결과를 살펴본 뒤, 사전, 사후 결과만으로 심리 상태를 파악하기에는 무리가 있다는 것을 인지했다. 본 실험에서는 간단한 인터뷰를 통해 심리 상태에 영향을 줄 만한 중대한 사건이 있었는지 등의 정보를 물어볼 예정이다. 또, 불안 증상뿐 아니라 우울 척도의 검사도 함께 진행한다면 조금 더 세밀한 분석이 가능할 것이라는 예상이 된다. 추후 다양한 색상의 테마를 추가해 사용자에게 적합한 색상의 프로그램을 선택해 진행할 수 있도록 발전시켜 볼 예정이다.

참고문헌

[1] 이양민, 박조안 외, 불안 측정 앱 및 공황장애 노출 훈련용 VR 시스템 개발, 한국정보과학회, 제24권 제5호, pp.227-233, 2018.
[2] 김나원, 화상환자의 PTSD 증상 완화를 위한 컬러테라피 연구, 석사학위논문, 2013.
[3] 홍기윤, 이언석, 색채심리치료를 위한 VR색채영상의 영향, 한국콘텐츠학회, *Volume 17 Issue 10*, pp.376-384, 2017.
[4] 최세영, 김수진 외, 컬러테라피를 활용한 VR 힐링 콘텐츠, '노르니르' 개발, 한국컴퓨터그래픽스학회, *Vol. 26, No. 3*, pp.143-153, 2020.
[5] Thomas Fydrich, Reliability and validity of the Beck Anxiety Inventory, Journal of Anxiety Disorders, Vol. 6. pp.55-61, 1992.

논문 발표

가상/증강현실2

가상 환경에서의 연속적인 포인팅 타겟 예측*

정충호[0], 이성희
문화기술대학원 (한국과학기술원)
{thegenuine, sunghee.lee}@kaist.ac.kr

Continuous Prediction of Pointing Target in Virtual Reality

Choongho Chung[0], Sung-Hee, Lee
GSCT, Korea Advanced Institute of Science and Technology

Abstract

We present a study on continuously predicting the direction to a pointing target in virtual environments using motion and eye-tracker data throughout the pointing process. We first collect and process time series data for user motion and eye-tracker in a cursorless, single-target pointing. Following this, we train a recurrent neural network that feeds on sliding window inputs containing temporal history of input features to predict the direction of the target at any given time during pointing. Our results suggest that using features from both the eye-tracker and joint motion contributes to higher prediction performance, as well as faster stabilization of output values at the starting phase of pointing.

1. Introduction

User behavior before pointing fixation can be affected by individual differences in motion and body dimensions. This can pose challenges in formulating an adaptive structure for a pointing prediction model. We address this problem by leveraging a deep neural network structure to continuously predict a pointing target throughout the course of the pointing action.

To develop a framework that can provide accurate predictions for various pointing states and individuals, we first focus on identifying useful input features for pointing targets, derived from motion[2] and gaze[3] data. To gain insights into how a continuously operating model should function for this application and to come up with appropriate input features for our prediction model, we analyze the characteristics of these features not only at times of fixation but also during the course of pointing.

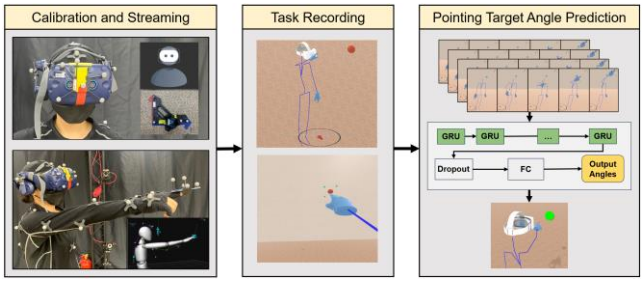


Figure 1: Overview of the Pointing Prediction System

2. Data Collection

We gather time-series data from human subjects performing single-target pointing tasks (gender: male=13, female=11, age: mean=26.4, standard deviation=2.5). The participants were divided into two groups to record data in two target placement scenarios (Fig.2): 1) a spherical grid target setup, and 2) a randomized target placement setup.

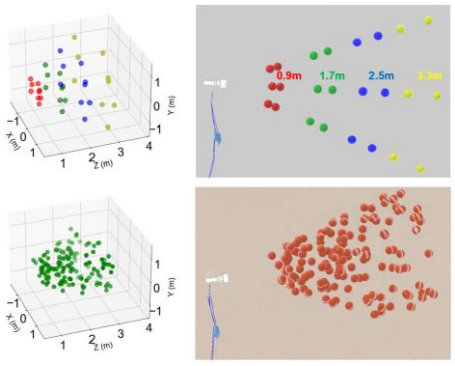


Figure 2 Data Collection Scenarios for Spherical(top), and Randomized (down) Target Placements

* Oral Presentation Paper
* This work is an extended abstract, originally submitted and published at IEEE ACCESS 31(3). 2024.
* This work was supported by NRF, South Korea, under Grant No.2022R1A4A5033689.

3. System

Category	Features	Form
Motion Features	HMD linear, angular velocity Dominant hand transform Dominant hand linear, angular velocity Index fingertip position, linear velocity	3x3 matrix 3x1 vector
Ray-Casting Features	Head-Finger-RayCast (HFRC) Eye-Finger-RayCast (EFRC)	3x1 vector
Eye-Tracker Features	Gaze vector (left, right) Gaze vector delta (left, right)	3x1 vector

Table 1 Input Window Features for our system

The randomized target placement data were utilized later for preventing model overfitting to uniform target positions in the later stages.
Participants wearing Vive Pro Eye headset, motion capture suit and manually constructed finger markers, were instructed to point at a single appearing target upon audible beep signal in a 3D virtual environment (Fig.1). The target direction, or final output for our model is represented as azimuth and altitude angles, with respect to VR headset position.
Based on the previous collected motion and eyetracking data, we build a target angle prediction model using a recurrent neural network architecture (Fig.4), feeding on input window consisting of multiple frames of motion features, eye-tracker features, and ray-casting features (Tab.1).

4. Analysis and Discussion

We demonstrate the accuracy of our model in three aspects: 1) an ablation study of input features (Fig.3), 2) performance comparisons for different neural network structures (Fig.5), and 3) an analysis for different input window lengths.

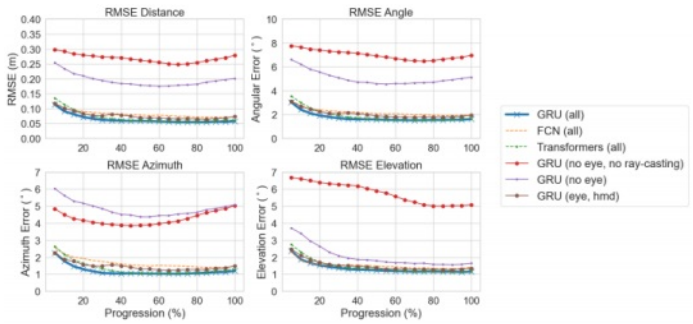
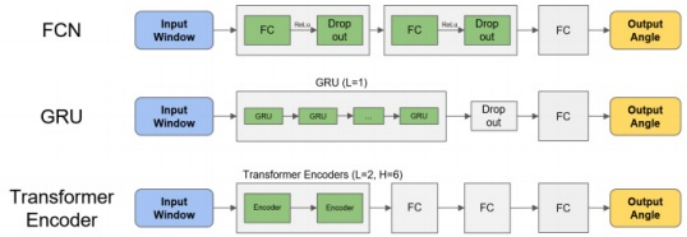


Figure 3 Ablation and Network Comparison results

To visualize the model performance, we draw a progression error plot representation by averaging the temporal RMSE series for all test data sequences. Our initial baseline excluded any eye-tracker derived data, achieving an overall deviance of 6.16◦ (SD=3.45◦) or minimum 5.0◦at 80% progression.

The second baseline assumed an HMD equipped with eyetracking, without controller or motion tracker support. This model delivered a significant leap in performance, achieving an RMSE of 1.71◦ (SD=1.16◦). The third baseline excluded only the gaze vectors, and attained an angular deviance of 4.20◦ (SD=2.86◦). The fully featured mode achieves the minimum angular error of 1.48◦ (SD=0.99◦), and a faster decrease in error values over the initial pointing period.



For comparing performance for different models, we compared a) a simple double-layered Fully-Connected-Network (FCN) with dropouts and b) a double-stacked Transformer encoder model. Our GRU neural network outperformed the other two models, achieving the lowest overall RMSE value.
In analyzing the impact of window length on performance, we tested window lengths of 0.67s, 0.41s, 0.33s, and 0.17s, each corresponding to 20, 15, 10, and 5 frames at a 30 FPS input sampling rate. The model with the 20-frame input window achieved the highest performance and fastest saturation.

5. Conclusion

We present a prediction study on pointing target direction during pointing progression, employing motion and eye-tracker features. Our model shows improved accuracy by incorporating eye-tracking data along with motion and ray-casting features. Addition of motion-based features is also found to reduce errors at start of pointing and maximizes the overall accuracy.

References

[1] Koehler, M Pruzinec, T Feldmann, and A Wörner, Automatic human model parametrization from 3d marker data for motion recognition. In *Proceedings of WSCG*, 2008.
[2] Tor-Salve Dalsgaard, Jarrod Knibbe, and Joanna Bergström, Modeling pointing for 3d target selection in vr. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, pages 1–10, 2021.
[3] Katrin Plaumann, MatthiasWeing, ChristianWinkler, Michael Müller, and Enrico Rukzio, Towards accurate cursorless pointing: the effects of ocular dominance and handedness. *Personal and Ubiquitous Computing*, 22:633–646, 2018.

안구 운동을 위한 DDA 유도 VR 게임이 몰입 및 안구 피로에 미치는 영향

김현진[1], 김영조[0,1], 이영성[1], 김서원[1], 임재호[2], 한다성[1]
[1]한동대학교 ICT창업학부, [2](주)덱스터스튜디오
{22000205, 22000128, 21800539, 22000085}@handong.ac.kr, jaeho.im@dexterstudios.com., dshan@handong.edu

Effect of a DDA-Guided VR Game for Ocular Exercises on Immersiveness and Visual Fatigue

Hyeonjin Kim[1], Yeongjo Kim[0,1], Youngsung Lee[1], Seowon Kim[1], Jaeho Im[2], Daseong Han[1]
[1]School of Global Entrepreneurship and ICT, Handong Global University, [2]Dexter Studios., Ltd

요약

본 연구는 동적 난이도 조절(DDA) 메커니즘을 적용한 VR 안구 운동 게임이 사용자 몰입과 안구 피로에 미치는 영향을 분석한다. 이를 위하여 two-step 강화학습에 기반한 DDA 에이전트 학습 기법을 제안한다. 실험적으로, DDA 메커니즘이 몰입을 증가시키는 데 도움이 되었지만, 안구 피로를 완화하지는 못함을 확인하였다.

1. 서론

디지털 화면 사용 증가로 인해 눈의 불편함과 시력 문제가 증가하는 가운데, 기존 연구에서는 간단한 안구 운동이 안구 피로를 완화할 수 있음을 입증하였다[1]. 그러나 이러한 안구 운동은 단순 반복으로 인해 지루함을 유발한다[2].

적극적인 참여와 몰입이 중요한 재활의 특성에 따라, 기존 연구들은 사용자의 수행 능력에 따라 동적으로 난이도를 조절하는 동적 난이도 조절(Dynamic Difficulty Adjustment, DDA) 메커니즘을 재활에 사용하였다[3]. 그러나 이는 손과 팔에 재활치료가 요구되는 뇌졸중, 척수 손상 등의 상지 재활 치료에 초점이 맞추어져 있으며, 꾸준한 참여가 중요한 안구 운동에 DDA 가 미치는 영향을 연구한 사례는 드물다. 본 연구에서는 DDA 를 적용한 안구 운동 VR 게임에서 사용자의 몰입과 안구 피로의 상관관계를 분석하고, 안구 운동에 적절한 구성 방식을 제안한다.

* 구두 발표논문
* 학부생 주저자 논문임.
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술지에 제출 준비 중임.
* 이 연구는 문화체육관광부 - 한국콘텐츠진흥원(KOCCA)에서 지원하는 연구비(No. R2022020123, 가상공연 참여를 위한 감정 표현이 가능한 소통형 관객 아바타 생성 기술 개발)로 수행됨.

2. 시스템 개요

본 논문에서 제안하는 시스템은 HMD 의 시선 추적 기능을 활용한 게임과 DDA 메커니즘으로 구성된다. 본 연구에서는 별도의 장비 없이 기기 내의 카메라를 통해 사용자의 시선을 추적할 수 있는 Meta Quest Pro 를 활용하였으며, Unity 의 Oculus Integration SDK 와 Meta Movement SDK 를 사용하여 환경을 구현하였다.

자연스러운 안구 운동을 유도하기 위해, 대표적인 안구 운동인 원활추종운동과 빠른안구운동에 기반하여 [4] 시선으로 장애물을 조준하고 컨트롤러로 장애물을 파괴하며, 흥미 증진을 위해 게이미피케이션 요소들을 사용하였다(그림 1).

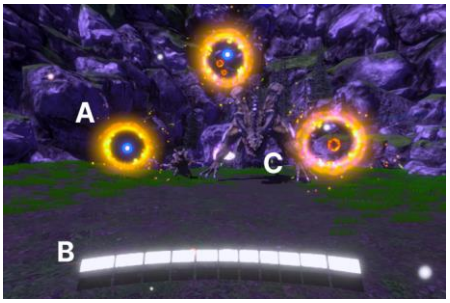


그림 1: 게임 플레이 화면. (A)장애물 생성 스포너, (B)사용자의 체력, (C)사용자의 공격 대상

3. DDA 에이전트 학습 기법

사용자와 게임의 상호작용을 시뮬레이션하여 난이도를 적절하게 조절하는 DDA 에이전트를 생성하기 위해 본 논문에서는 two-step 강화학습 기법을 제안한다.

1단계: 플레이어 에이전트를 위한 강화학습
이 단계에서는 사용자의 게임 수행 행동을 흉내 내는 플레이어 에이전트를 정의하고, 강화학습을 통해 초급, 중급, 상급 수준의 플레이어 에이전트를 학습시킨다.

플레이어 에이전트의 observation 벡터는 공격 대상과 사용자의 체력, 가상 눈의 회전 값, 장애물의 정보로 구성되어 있다. Action 벡터는 장애물을 구분해서 공격하는 2 개의 이산 변수와 최우선 파괴대상을 결정하는 1 개의 이산 변수로 구성된다.

2단계: DDA 에이전트를 위한 강화학습

두 번째 단계에서는 게임 난이도를 조절하는 DDA 에이전트를 정의하고, 학습된 플레이어 에이전트들에 대해 적절한 난이도를 유지하도록 학습시킨다. DDA 에이전트의 observation 벡터는 사용자의 체력 변화량, 파괴된 장애물과 사용자의 평균 거리, 난이도 조절 변수, 사용자의 체감 난이도로 구성되어 있다. Action 벡터는 난이도의 상승, 하락, 유지를 선택하는 이산 변수 1 개로 구성된다.

체감 난이도가 적절한 범위 안으로 들어오면 보상을 주고, 그 범위를 벗어나면 처벌을 주는 보상함수를 통해 DDA 에이전트가 난이도를 적절하게 유지하도록 학습하였고, 그 과정에서 앞서 학습된 플레이어 에이전트들과 번갈아 게임을 진행하도록 하여 각각의 사용자들에 대해 적절하게 반응하도록 하였다.

4. 실험 설계

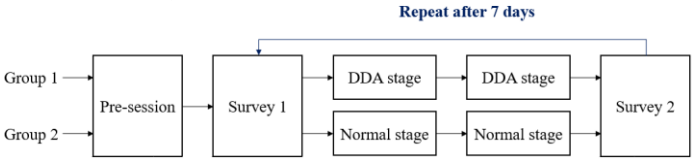


그림 2: 실험 설계 개요

본 연구의 실험 설계는 그림 2 와 같다. 열 명씩 DDA 메커니즘을 경험하는 그룹 1 과 기존 메커니즘을 경험하는 그룹 2 로 무작위로 나누었다. 피험자들은 상호작용에 익숙해지도록 사전 세션을 진행한 후, 총 두 번의 설문과 게임을 진행하였다. 설문은 게임을 경험 전후의 안구 피로 차이를 확인하기 위한 목적으로, 모든 게임이 끝난 후에는 몰입에 관한 설문을 진행하였다.

5. 실험 결과

표 1: 각 그룹의 몰입도, 증가율, 감소율 평균

	몰입도 평균	증가율 평균	감소율 평균
Group 1	37.95점	10.18%	-7.76%
Group 2	33.3점	5.54%	-13.65%

표 2: 각 그룹의 안구 피로도 증감별 인원 수

		증가	유지	감소
Group 1	Day 1	5	1	4
	Day 2	4	1	5
Group 2	Day 1	6	0	4
	Day 2	2	0	8

표 1은 몰입도와 관련된 문항을 분석한 결과이며, 표 2는 안구 피로와 관련된 문항을 분석한 결과이다. 몰입도 전체 평균에서는 그룹 1의 몰입도가 그룹 2보다 높은 것으로 나타났다. 또한, 그룹 1의 경우 총 4번의 게임동안 피험자의 50%가 몰입도 증가를 경험하였으며, 그룹 2의 경우 피험자의 30%가 몰입도 증가를, 50%가 감소를 경험하였다. 몰입도가 감소한 피험자의 비율을 동일하지만, 몰입도가 증가한 피험자를 비교했을 때, 그룹 1의 몰입도 증가율 평균이 그룹 2보다 2배 높았다. 몰입도가 감소한 피험자를 비교했을 때, 몰입도 감소율 평균 또한 그룹 1이 더 작았다.

안구 피로도의 경우, 그룹 1 에서는 총 4 번의 게임 동안 안구 피로도가 증가한 피험자의 비율과 감소한 피험자의 비율이 동일한 반면, 그룹 2 의 경우 Day 1 에서는 안구 피로도가 증가하였다고 응답한 피험자의 75%가 Day 2 에서 안구 피로도의 감소를 경험했다고 응답하였다. 결론적으로, DDA 메커니즘을 경험했을 때 몰입은 증가하였지만, 기존 메커니즘을 꾸준히 경험했을 때 안구 피로가 감소하였다는 것을 확인할 수 있었다.

6. 결론 및 향후 연구

본 연구는 DDA 를 활용한 안구 운동 VR 게임을 통해 몰입과 안구 피로의 상관관계를 분석하였다. 본 논문에서는 DDA 메커니즘이 몰입을 증가시키는 데 도움을 주지만, DDA 가 적용되지 않은 기존 게임이 안구 피로를 감소시키는 데 더 효과적인 것을 확인할 수 있었다. 이는 꾸준한 참여를 위해서 사용자의 몰입을 증가시키는 DDA 메커니즘이 긍정적일 수 있으나, 안구 피로를 완화하지는 못하기 때문에 실제 콘텐츠로 구성할 때 몰입과 안구 피로 사이의 균형을 맞출 수 있는 적절한 DDA 메커니즘이 필요함을 제시한다.

그러나 본 연구는 피험자 수가 적어 효과를 통계적으로 검증하기에는 한계가 존재한다. 향후 연구에서는 더 많은 실험군을 모집하여 반영할 것이며, DDA 를 적용한 안구 운동 VR 게임을 여러 번 경험했을 때, 높은 몰입을 유지하며 안구 피로를 완화하는 요소를 중점으로 연구할 계획이다.

참고문헌

[1] H.S.Shin, S.C.Park and W.C.Maples, Effectiveness of vision therapy for convergence dysfunctions and long-term stability after vision therapy, *Ophthalmic Physiol Opt*, 31(2):180-189, 2011.

[2] O.Hurd, S. Kurniawan and M. Teodorescu, Virtual Reality Video Game Paired with Physical Monocular Blurring as Accessible Therapy for Amblyopia, *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp.492-499, Osaka, Japan, IEEE.

[3] N.Hocine, A.Gouaich and S.A.Cerri, Dynamic Difficulty Adaptation in Serious Games for Motor Rehabilitation, *Games for Training, Education, Health and Sports*, pp.115-128, Springer, Cham.

[4] D.W. Lawrence, Concussion Vision Therapy and Oculomotor Rehab, *David L.Macintosh Sport Medicine Clinic*, 2022.

논문 발표

이미지/비디오

레퍼런스 착용 스타일을 반영하는 가상 착용 이미지 합성*

박순찬[0], 박진아
한국전자통신연구원, 한국과학기술원
parksc@etri.re.kr, jinahpark@kaist.ac.kr

Image-based Virtual Try-on Reflecting Wearing Style of Reference Image

Soonchan Park[0], Jinah Park
ETRI, KAIST

요약

본 연구는 타겟 모델 이미지와 타겟 옷 이미지를 입력으로 타겟 모델이 타겟 옷을 착용한 것과 같은 이미지를 생성적 신경망을 통해 합성하는 착용 이미지 합성 기술에 관한 것으로, 합성하는 과정에서 레퍼런스 모델 이미지의 착용 스타일을 반영하기 위한 방법론을 제안한다. 착용 스타일을 모사하고자 하는 레퍼런스 모델의 정보와 합성하고자 하는 타겟 모델의 상이한 정보(예. 체형, 자세 등)를 극복하기 위해 본 연구에서는 착용 스타일 이전 모듈(Wearing Style Transfer Module)을 제안하여 레퍼런스 모델의 착용 스타일을 타겟 모델에 맞게 매칭시킨 다음 이미지 합성을 수행한다. 본 연구에서 수행한 정성적 결과에 따르면 제안한 WSTM은 임의의 레퍼런스 모델을 타겟 모델의 체형과 자세에 맞게 변형하여 레퍼런스 착용 스타일을 효과적으로 전달하고 결과적으로 다양한 착용 스타일을 가지는 이미지가 합성될 수 있음을 보였다.

1. 서론

가상 착용 (virtual try-on) 이미지 합성은 옷과 모델의 이미지를 입력으로 하여 입력된 모델이 입력된 옷을 실제 착용한 듯한 모델 이미지를 합성하는 방법으로 최근 다양한 형태의 방법들이 소개되어왔다. 최근 상, 하의를 동시에 입력하여 전신 착용 이미지를 합성하는 연구에서는 복수의 상품을 활용할 때, 다양한 착용 스타일을 반영해야 한다는 문제가 제기되었다. 이에 추가적인 마스크 입력을 활용해 상의의 길이를 조절하여 착용 스타일을 조절하는 방법이 제안되었으나 다양한 모델 이미지들에 일반적으로 적용되기에는 한계가 있다[1].

본 논문에서는 모델 및 상, 하의 상품을 동시에 입력받아 전신 착용 이미지를 합성하는 과정에서 추가적으로 입력되는 레퍼런스 모델 이미지의 착용 스타일을 적용하도록 유도하는 조건부 이미지 합성 방법을 소개한다. 보다 자세하게는 입력된 레퍼런스 모델을 타겟 모델의 체형과 자세에 맞게 일치시킨 뒤 착용 스타일을 이전시키는 착용 스타일 이전 모듈 (Wearing Style Transfer Module)을 제안한다. 이렇게 이전된 착용 스타일을 바탕으로 타겟 모델과 옷에 적합한 착용 이미지 합성을 수행케 함으로써 조건부 착용 이미지 합성 과정이 이루어진다. 본 논문에서 Fashion-TB 데이터세트를 활용하여 수행한 정성적 평가 결과에 따르면, 제안한 방법은 임의의 레퍼런스 모델 이미지의 착용 스타일을 타겟 모델의 체형과 자세에 맞도록 전이할 수 있음을 확인하였으며, 결과적으로 다양한 착용 방법을 모사하는 가상 착용 이미지 합성을 수행할 수 있음을 확인하였다.

2. 착용 스타일을 반영한 착용 이미지 합성

스타일을 반영한 착용 이미지 합성은 레퍼런스 모델 이미지의 착용 스타일을 추출하고 추출된 스타일과 유사한 스타일을 가지는 모델 이미지가 결과로 합성되게끔 유도하는 방법이다. 본 연구에서는 레퍼런스의 착용 스타일을 분석하고 전달하는 방법을 제안한다. 본 장에서는 착용 스타일의 정의와 레퍼런스 모델의 착용 스타일을 체형과 자세가 다를 수 있는 타겟 모델로 이전하는 방법에 대해서 상세히 설명하고, 이를 활용하여 가상 착용 이미지 합성을 수행하는 과정을 설명한다.

2.1. 착용 스타일의 정의

본 논문에서 정의한 상, 하의 착용 과정에서의 착용 스타일은 모델의 신체 정보와 상의, 하의의 위치적 관계로 정의된다. 예컨데 상, 하의 경계가 허리 위쪽으로 높게 위치하는 하이웨이스트 (high-waist), 상의를 내어 입는 루즈핏, 상의를 바지에 넣어 입는 턱인(tuck-in) 스타일이 그 예가 될 수 있다. 본 논문에서는 사전 실험을 통하여 사용자 신체를 잘 표현하는 Densepose[2] 추론 결과(D)와 의미론적 영역 정보(S)를 이용하여, Densepose와 하의의 골반 영역을 연결하여 3채널로 만듦으로써 하의 골반 영역의 크기와 신체상에서의 위치로 모델의 착용 스타일을 정의하였다 (그림 2.(a)). 이를 수식으로 정리하면 $T_{(r,style)} = (S_{(r,hip)}, D_r)$과 같이 정의될 수 있다.

2.2. 착용 스타일 전이 모듈 및 학습방법

앞선 장에서 설명한 착용 스타일 $T_{(r,style)}$은 레퍼런스 모델의 착용 스타일이므로 이를 주어진 타겟 모델의 크기와 자세에 맞게 전달하는 것이 중요하다. 이를 위해 본 방법에서는 두 입력 모델의 신체를 일치시키기 위한

* 구두발표논문
* 본 연구는 2023년도 정부(과학기술정보통신부)의 재원으로 국가과학기술연구회 창의형 융합연구사업(No. CAP23081-000)의 지원을 받아 수행되었습니다.

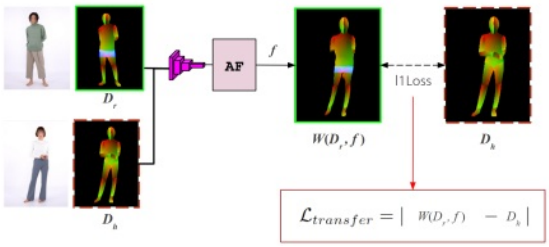


그림 1: 레퍼런스 모델과 타겟 모델의 체형, 자세를 일치시키는 착용 스타일 전이 모듈의 및 학습방법

Appearance Flow[3]를 추론하는 착용 스타일 전이 모듈(Weairng Style Transfer Module, WSTM)을 제안한다. 보다 자세하게는 그림 1와 같이 WSTM은 레퍼런스 모델의 정보(D_r)와 타겟 모델의 정보(D_h)를 입력으로 받아 D_r을 D_h에 일치시킬 수 있는 2차원 와핑 오프셋을 출력하도록 학습된다. 이상적으로 학습이 이루어지면 우리는 임의의 레퍼런스 모델이 표현하는 착용 스타일($T_{(r,style)}$)역시 학습된 와핑을 통해 임의의 타겟 모델의 자세와 체형에 맞게 전달할 수 있게 된다.

2.3. 착용 스타일을 반영한 착용 이미지 합성

본 연구에서 제안한 착용 스타일 이전 모듈의 효용성을 검증하기 위해서 본 연구에서는 기존 연구에서 고안한 단일 스테이지 기반 상, 하의 착용 이미지 합성 네트워크 WG-VITON[1]을 바탕으로 착용 스타일 전이 모듈(WSTM)을 적용하였다. WG-VITON은 입력된 상, 하의 옷 이미지를 타겟 모델에 맞게 와핑하는 모듈과 와핑된 결과를 활용하여 모델 이미지를 합성하는 모듈이 통합되어 종단간학습을 통해 학습된다. 본 연구에서는 WSTM의 와핑 함수 W를 통해 이전된 레퍼런스 모델의 착용 스타일($W(T_{(r,style)})$)을 목표로 WG-VITON의 타겟 옷 와핑이 이루어지게 함으로써 최종적으로는 레퍼런스 착용 스타일과 유사한 스타일을 가지는 착용 이미지가 합성된다.

3. 실험

본 연구에서 제안한 착용 스타일 전이 모듈의 효용성을 검증하기 위하여 Fashion-TB[1] 데이터세트를 활용하여 정성적 평가를 진행하였다. 그림 2는 WSTM이 레퍼런스 모델의 정보를 와핑한 결과를 보여주고 있다. 1행과 2행은 상의를 내려 입거나 하의 안으로 넣어 입은 레퍼런스 이미지를 보여주고 있으며(그림 2.(a)), 레퍼런스의 착용 스타일은 WSTM에 의해 타겟모델(그림 2.(b))에 맞춰 와핑된다(그림 2.(c)). 와핑 결과를 확인할 때 신체가 완벽하게 일치되도록 와핑할 순 없지만 하늘색 가시화가 된 바지 골반 영역으로 대표되는 착용 스타일이 타겟 모델의 신체와 자세에 맞게 전달되는 것을 확인할 수 있다.

그림 3은 스타일을 반영한 착용 이미지 합성의 전체 과정을 보여주고 있다. 같은 타겟 모델과 옷 입력(그림 3.(a))에도 두 다른 레퍼런스 이미지를 스타일로 제시(그림 3.(b))함으로써 스타일이 타겟 모델의 신체에 맞게 변형된 결과(그림3.(c))에 맞게 타겟 옷 와핑이 이루

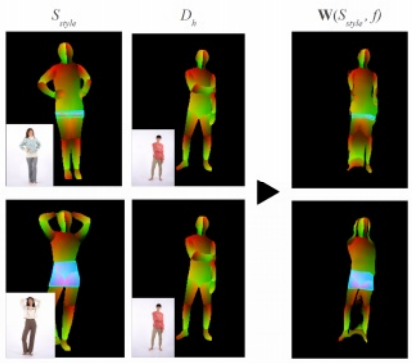


그림 2: (a) 다른 스타일의 레퍼런스 Densepose 하늘색 가시화는 하의의 골반 부분, (b) 타겟 모델의 Densepose, (c) 레퍼런스가 타겟에 맞게 와핑된 결과.

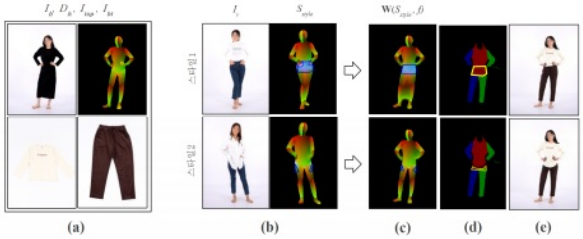


그림 3: (a) 입력된 타겟 모델 및 옷, (b) 다른 착용 스타일의 레퍼런스 모델 이미지, (c) WSTM에 의해 와핑된 레퍼런스 모델 정보, (d) 스타일에 맞게 입력옷이 와핑 된 결과, (e) 스타일을 반영한 가상 착용 이미지 합성 결과

어지고 (그림 3.(d)), 그에 따라 다른 착용 스타일을 가지는 착용 이미지(그림3.(e))가 합성되는 것을 확인할 수 있다.

4. 결론

본 연구에서는 복수의 옷을 활용하는 가상 착용 이미지 합성에 있어서, 추가적 입력으로 주어지는 레퍼런스 모델 이미지의 착용 스타일을 적용하여 가상 착용 이미지 합성을 수행하는 방법에 대해서 소개하였다. 본 연구에 따르면 레퍼런스 모델과 타겟 모델의 자세와 체형이 다른 경우에도 착용 스타일 전이 모듈이 레퍼런스 스타일을 타겟 모델에 맞게 잘 전달하여 조건부 이미지 합성이 수행되게 할 수 있음을 확인하였다.

참고문헌

[1] S. Park and J. Park, “Single-stage virtual try-on for top and bottom clothes with wearing style control,” Available at SSRN 4379142, 2023.

[2] R. A. G ¨uler, N. Neverova, and I. Kokkinos, “Densepose: Dense human pose estimation in the wild,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7297–7306.

[3] Zhou, Tinghui, et al. "View synthesis by appearance flow." Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016.

RNA: ROI 기반 아틀라스 신경망을 사용한 동영상 편집*

이재경[0,1], 김건웅[1], 조성현[1,2]
포항공과대학교 {인공지능대학원[1], 컴퓨터공학과[2]}
{jaekyeong, k2woong92, s.cho}@postech.ac.kr

RNA: Video Editing with ROI-based Neural Atlas

Jae-Kyeong Lee[0,1], Geon-Ung Kim[1], Sung-Hyun Cho[1,2]
Graduate School of Artificial Intelligence[1], POSTECH
Department of Computer Science[2], POSTECH

요약

최근 시간 축으로 일관된 동영상 편집을 위해서 아틀라스 기반 동영상 표현 방법이 제안되었다. 하지만 이러한 방법은 사용자에게 전경 객체 선택 작업을 요구하고, 복잡한 전경 객체들을 정확히 다루지 못하며, 작은 편집 영역에서도 비효율적인 학습을 요구한다. 본 연구는 편집하기를 원하는 영역을 지정했을 때, 선택된 영역만 아틀라스를 추정하는 방식으로 다양한 동영상에서 효율적으로 학습하면서 편집하는 방법을 제안한다.

1. 서론

최근 시간 축으로 일관된 동영상 편집을 위해서 LNA [1], Hashing NVD [2]와 같은 아틀라스 기반 동영상 표현 방법이 제안되었다. 이러한 방법들은 각 동영상 프레임이 배경과 전경 객체를 나타내는 아틀라스들로 구성되어 있고, 아틀라스 외형은 일정하지만, 위치와 형태는 움직임에 따라 변한다는 가정으로 동작한다. 이러한 가정 아래 아틀라스 기반 방법은 각 레이어에 아틀라스와 움직임을 추정 후, 아틀라스를 편집한 다음 복원해서 시간 축으로 일관된 편집을 할 수 있다. 하지만 이러한 방식은 전경 객체 분할을 위해 사용자에게 전경 객체 선택 작업을 요구하고, 전경 객체가 복잡한 움직임이 있을 때 정확한 아틀라스를 추정하는 것이 어렵다. 또한 모든 전경 객체의 아틀라스를 모델링함으로 전경 객체 수에 따라 학습 시간이 증가하게 된다.

본 연구는 참조 프레임(reference frame)에 사용자가 편집하기를 원하는 영역에 ROI를 지정하면, ROI 영역만 아틀라스를 추정하는 ROI 기반 아틀라스 신경망을 제안한다. 이러한 방식은 전경 객체와 배경을 분리하는 작업과 전경 객체의 아틀라스 모델링 과정을 제거해서, 전경 객체의 복잡한 움직임을 가진 경우에도 자연스러운 편집을 가능하게 한다. 또한 전경 객체의 수와 상관없이 일정한 계산 자원을 사용하여 학습할 수 있다.

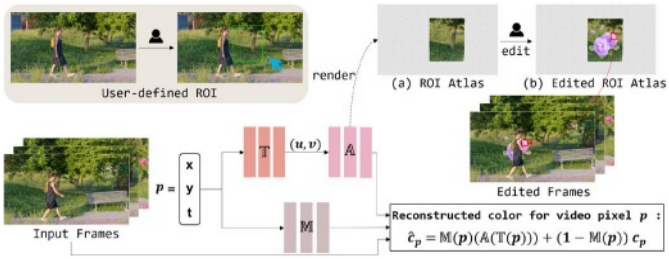


그림 1: ROI 기반 아틀라스 신경망 모델

2. ROI 기반 아틀라스 신경망

본 연구에서 ROI 기반 아틀라스 신경망 모델을 아래와 같이 제안한다. (그림 1 참조)

$$\hat{c}_p = \mathbb{M}(p)\left(\mathbb{A}(\mathbb{T}(p))\right) + (1 - \mathbb{M}(p))c_p$$

p는 픽셀 위치를 나타내고, c는 입력 동영상과 $\hat{c}$는 복원된 동영상의 픽셀 값이다. $\mathbb{T}: \mathbb{R}^3 \rightarrow \mathbb{R}^2$는 픽셀 좌표와 아틀라스 좌표의 매핑 관계를 학습하고, $\mathbb{A}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$는 아틀라스에 RGB 값을 학습한다. $\mathbb{M}: \mathbb{R}^3 \rightarrow \mathbb{R}^1$은 픽셀 좌표에서 마스크 값이고, ROI 영역에 속하면 $\mathbb{M}(p) = 1$, 속하지 않으면 $\mathbb{M}(p) = 0$으로 학습한다. 이러한 마스크 값은 사용자가 ROI 아틀라스에 편집 후 복원하였을 때, 전경 객체로부터 가림(occlusion)이 발생했을 때, 자연스러운 편집 결과를 만들어 준다.

2.1. 아틀라스 추정

아틀라스 추정 손실함수는 아래와 같다.

$$\mathcal{L} = \mathcal{L}_{recon} + \mathcal{L}_{rigid} + \mathcal{L}_{pos} + \mathcal{L}_{corr} + \mathcal{L}_{mask}$$

$\mathcal{L}_{recon}$은 복원된 픽셀 값 $\hat{c}_p$가 입력 동영상의 픽셀 값인 c_p와 가까워지도록 한다. $\mathcal{L}_{rigid}$는 $\mathbb{T}$의 국소 변환이 강체(rigid)에 가깝게 유도한다. $\mathcal{L}_{pos}$는 그림 1(a)와 같이 아틀라스 좌표가 참조 프레임에 ROI 좌표와 비슷하게 나오도록 만든다. $\mathcal{L}_{corr}$는 마스크가 1인 부분에서 픽셀 위치 p에 대한 기존 광학 흐름 모델(optical flow)로 미리

* 구두발표논문
* 현재 타 학술대회(논문지)에 제출 중.
* 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2019-0-01906, 인공지능대학원지원(포항공과대학교))과 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(RS-2024-00395401, 생성 AI 기반 특수효과 자동 생성 및 합성 기술 개발)을 받아 수행된 연구임.

구해 놓은 참조 프레임 대응점과 인접한 프레임 대응점이 같은 아틀라스 좌표가 나오도록 만든다. $\mathcal{L}_{mask}$는 참조 프레임에서 ROI 영역은 마스크 값을 1, 그 외 영역은 0으로 만든다. 그리고 픽셀 위치 p에 대한 참조 프레임과 인접한 프레임의 대응점에서 같은 마스크 값을 가지도록 학습한다.

하지만 이러한 아틀라스 추정 손실 함수는 정확한 마스크 값을 구하는 것이 어렵다. 광학 흐름 모델의 신뢰도가 좋지 않은 전경 객체 경계 부분에서 $\mathcal{L}_{recon}$이 마스크를 0으로 만들기 때문이다. 이러한 마스크로 편집했을 때, 그림 2(a)에 빨간색 화살표처럼 원본 색상이 보이는 자연스럽지 못한 편집 결과를 만든다.

그림 2: ROI 기반 아틀라스 신경망의 단계별 편집 결과

2.2. 마스크 정제

마스크 정제 단계는 마스크를 개선하기 위해 아래 손실 함수를 사용하여 $\mathbb{M}$ 신경망을 추가적으로 학습한다.

$$\mathcal{L}_{refine} = \mathcal{L}_{recon} + \mathcal{L}_{no}$$

$\mathcal{L}_{no}$는 2.1장에서 추정한 아틀라스를 렌더링한 픽셀 값 $\mathbb{A}(\mathbb{T}(p))$가 입력 동영상의 픽셀 값 c_p와 비슷한 특성을 활용하여 비가림 픽셀(non-occluded pixel)을 식별하고, 이러한 픽셀에서 마스크 값을 1로 유도한다.

$\mathcal{L}_{refine}$을 사용하여 정제된 마스크는 그림 2(b)와 같은 개선된 편집 결과를 보여준다. 하지만 이러한 아틀라스 신경망 모델은 마스크가 0 또는 1로 학습되는 경향이 있다. 이러한 마스크는 그림 2(b)에 주황색 화살표처럼 여전히 자연스럽지 않은 편집 결과를 보여준다.

2.3. 소프트 아틀라스 신경망 모델을 사용한 복원

더 자연스러운 편집 결과를 얻기 위해 아틀라스와 가림을 만드는 전경 객체를 소프트 마스크로 혼합(blending)하는 소프트 아틀라스 신경망 모델을 제안한다.

$$c_p = \hat{\mathbb{M}}(p)\left(\mathbb{A}(\mathbb{T}(p))\right) + (1 - \hat{\mathbb{M}}(p))c_p^{oc}$$

먼저 입력 프레임과 2.2장에서 구한 마스크를 사용하여 기존 매팅 네트워크로부터 소프트 마스크 $\hat{\mathbb{M}}$을 구한다. 그러고 나면 제안된 모델로부터 c_p^{oc}를 유도할 수 있다.

최종적으로 편집된 픽셀 값 c_p^{edit}은 $\mathbb{A}$를 편집한 아틀라스 $\mathbb{A}_{edit}$로 대체하고, $\hat{\mathbb{M}}(p)$와 c_p^{oc}를 사용하여 구할 수 있다. 이러한 수식으로 전체 픽셀을 복원하였을 때, 그림 2(c)와 같이 자연스러운 편집 결과를 얻을 수 있다.

3. 실험 결과

3.1. 정성적 비교

그림 3은 정성적 비교 결과를 보여준다. 기존 방법은 전경 객체들의 복잡한 관계 또는 전경 객체의 복잡한 움직임으로 전경 객체 아틀라스 모델링이 실패하여 자연스럽지 못한 결과를 보인다. 반면 본 연구는 훨씬 더 자연스러운 편집 결과를 보여준다.

그림 3: 정성적 비교 결과

3.2. 정량적 비교

표 1은 DAVIS 데이터 셋에서 동영상의 전경 객체 수가 증가함에 따라 학습 시간과 ROI영역의 PSNR을 비교한 결과이다. 기존 방법은 전경 객체 수가 증가할수록 학습 시간이 증가하는 반면, 본 연구는 일정 시간동안 학습하는 것이 가능하다.

표 1: 학습 시간 및 ROI 영역의 PSNR

데이터 셋 (전경 객체 수)	Hashing NVD [2]		Ours	
	학습 시간	PSNR	학습 시간	PSNR
lucia (1)	72분	31.0	40분	31.5
dogs-jump (3)	132분	33.2	40분	33.5
dog-gooses (5)	192분	27.3	40분	27.7

4. 결론

본 논문은 사용자가 편집하기를 원하는 영역에 ROI를 지정하는 방식으로 동영상을 편집할 수 있게 한다. 이러한 ROI 기반 방식은 사용자에게 모든 전경 객체를 선택해야 하는 작업을 요구하지 않고, 효율적으로 마스크와 아틀라스를 추정할 수 있다. 또한 우리는 자연스러운 편집 결과를 얻기 위해 마스크 정제와 소프트 아틀라스 모델 방법을 제안한다. 결과적으로 이러한 방법은 다양한 동영상에서 기존 방법보다 더 효율적이면서 직관적으로 편집하는 것을 가능하게 한다.

참고문헌

[1] Kasten, Yoni, et al. "Layered neural atlases for consistent video editing." *ACM Transactions on Graphics (TOG)* 40.6 (2021): 1-12.

[2] Chan, Cheng-Hung, et al. "Hashing Neural Video Decomposition with Multiplicative Residuals in Space-Time." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

다단계 미세구조 구역화 네트워크를 활용한 3D 의료영상 구역화*

김자연[0], 김지온, 신병석
인하대학교 전기컴퓨터공학과
{Kimjayeon1, jionkim}@inha.edu, bsshin@inha.ac.kr

3D Medical Image Segmentation Using a Multi-Stage Microstructure Segmentation Network

Ja-Yeon Kim[0], Ji-On Kim, Byeong-Seok Shin
Department of Electrical and Computer Engineering, Inha University

요약

인체는 다양한 미세구조들로 이루어져 있으므로 의료용 3D 이미지 내의 미세구조를 구역화 하는 작업은 중요하다. 기존 구역화 방법은 구역화 할 미세구조가 포함된 영역이 전체 이미지의 작은 부분만 차지하는 경우에도 이미지 전체를 대상으로 하여 계산을 수행하므로 GPU 메모리가 낭비된다. 이를 해결하기 위해 먼저 구역화 할 ROI 영역을 결정한 뒤 해당 영역만을 집중적으로 구역화 하는 다단계 네트워크를 제시한다. 이를 통해 이미지 전체를 계산하지 않고도 미세구조를 효과적으로 구역화 할 수 있다. 제안하는 방법은 Dice, IoU, 그리고 HD95 평가지표를 통해 구역화 정확도를 평가하였고 GPU 메모리 사용량을 측정하였다. 실험 결과는 제안하는 방법이 비교모델과 유사하거나 더 좋은 성능을 보이는 것을 보여준다.

1. 서론

의료분야에서의 이미지 구역화는 병변이나 해부조직을 분리하여 그 경계를 파악하기 위해 사용한다. 인체내 미세구조들은 대상 병변이나 조직이 이미지 내에서 차지하는 영역이 작기 때문에 구역화가 어렵다. 또한 미세구조는 영상의 작은 영역을 차지하기 때문에 전체 영역에서 구역화 작업을 수행하면 불필요한 연산을 하게 되고 많은 GPU 메모리가 낭비된다. 기존에는 GPU 메모리 문제를 해결하기 위해 네트워크 경량화를 주된 전략으로 삼았다. 그중의 하나의 방법으로 입력 이미지의 크기를 줄이는 방법이 있다. 입력 이미지를 줄이는 방법은 간단하지만 정보가 손실되어 구역화 정확도가 저하될 수 있다. 특히 2D 이미지가 아니고 CT, MRI등 단층촬영 장치에서 생성되는 3D 이미지의 경우는 이러한 비효율성이 더 심각해지며 심지어 처리가 불가능한 경우도 생길 수 있다. 이 연구는 미세구조가 존재하는 이미지에서 ROI를 사용하여 GPU 메모리 사용량을 줄이고자 한다. 기존에도 ROI를 사용하는 기법[1]이 존재하였지만 정확도 향상만을 목적으로 하며 GPU 메모리 절감이 고려되지 않았다. 이와 달리, 본 논문에서는 미세구조가 존재하는 ROI(region of interest)만을 구역화하는 구조를 통해 정확도를 유지하며 GPU 메모리 사용량을 절감하는 네트워크를 제안한다.

본 연구에서는 ROI 영역을 탐지하는 label map localization 단계와 구역화를 수행하는 ROI segmentation 단계로 이루어진 다단계 네트워크를 제안한다. 이를 통해 1단계에서 탐지된 ROI 영역에 대해서만 구역화를 함으로써 정확도를 유지하면서 계산량과 메모리 사용량을 크게 줄일 수 있다. 실험 과정에서 Dice, IoU(intersection over union), 그리고 HD95 평가지표를 통해 구역화 정확도를 측정하였다. 또한 GPU 메모리 사용량을 측정하여 제안하는 네트워크의 성능을 평가하였다.

2. 본문

제안하는 네트워크는 그림 1과 같이 label map localization 단계 ROI segmentation 단계, 두 단계로 구성된 transformer 기반 네트워크이다. 제안하는 네트워크는 ROI 영역에 대해서만 구역화를 수행하여 낮은 GPU 메모리 사용량으로 이미지 내 미세구조의 영역을 추출할 수 있도록 한다.

Label map localization 단계는 다운샘플링된 3D 이미지를 사용하여 최종 예측 라벨의 근사 영역을 얻는 단계이다. 라벨이 존재하는 영역을 얻기 위해 우선 다운샘플링된 이미지로부터 구역화 하는 coarse segmentation 모델을 통해 근사 라벨을 구한다. 얻은 라벨의 너비, 높이, 깊이 인덱스들의 최대값과 최소값으로부터 경계상자(bounding box) 좌표를 추출한다.

ROI segmentation 단계는 label map localization 단계에서 얻은 예측 경계상자 좌표와 ground truth 경계상자 좌표를 사용해 고해상도 이미지를 잘라내어 ROI 영역에 해당하는 이미지를 얻는다. 잘라낸 이미지는 고

* 구두 발표논문
* 이 연구는 주식회사 삼성메디슨의 지원에 의해 수행되었음.
* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2022-00155915, 인공지능융합 혁신인재양성(인하대학교)).

해상도 라벨을 구역화하는 fine segmentation 모델의 입력으로 사용하여 구역화를 수행한다. 이때 잘라낸 이미지는 다양한 크기를 갖으나 모델의 입력으로 사용하는 크기는 고정되어 있으므로 입력 이미지는 크기를 조정(resize)하여 사용한다. Fine segmentation 모델을 통해 추출된 라벨은 입력 전 이미지 크기와 동일하도록 다시 원래크기로 복원한다.

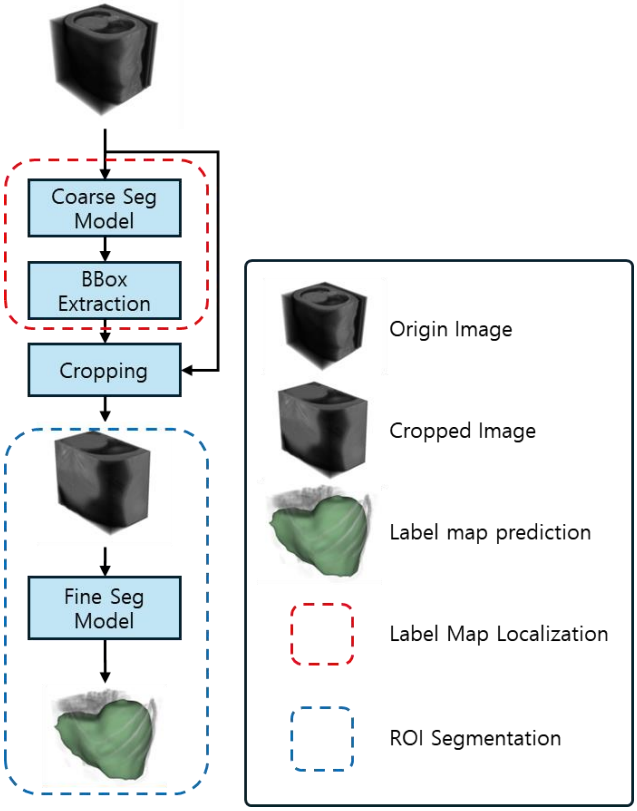


그림 1: 제안하는 다단계 네트워크 구조

3. 실험결과

AMOS[2] 데이터셋의 간, 전립선, 그리고 자궁 라벨이 포함된 CT 이미지를 사용하여 실험하였다. 실험 과정에서 4의 배치크기와 0.002의 학습률을 적용하였고 네트워크를 1000회의 epoch으로 학습하였다. 실험과정에서 제안한 기법과 Slim UNETR[3] 모델 간의 결과 비교를 진행하였다. 제안하는 네트워크와 비교모델을 Dice, IoU, 그리고 HD95 지표로 평가하였다. Dice와 IoU는 두 픽셀 집합 간의 유사도를 비교하는 지표이다. HD95는 상위 95% 백분위수에 해당하는 이미지 외곽선 픽셀 간의 거리를 계산하는 지표이다. 지표를 사용한 결과는 표 1과 같다. 제안하는 기법은 기존 기법과 비교하여 간, 전립선/자궁 클래스에서 각각 Dice 지표가 0.16%p, 1.37%p 향상되었고, IoU 지표가 0.02%p, 3.09%p 향상되었다. HD95 지표는 각각 2.8834mm, 2.9059mm 개선되었다. 이는 제안하는 기법이 구역화해야 하는 대상 영역을 ROI로 한정시킨 것이 이유로 보인다. 따라서, 제안하는 기법은 ROI를 효과적으로 활용하여 간, 전립선/자궁 CT 이미지의 구역화 성능을 향상시켰음을 알 수 있다.

제안하는 네트워크와 비교모델의 GPU 메모리 사용량은 표 2에 나타나 있다. 제안하는 네트워크는 훈련과정에서 간, 전립선/자궁 클래스에서 각각 74.36%, 74.74%의 GPU 메모리 사용량을 줄였고 추론과정에서는 각각 16.37%, 10.99%가량 줄였다. 이러한 결과를 통해, 제안하는 네트워크는 입력 이미지 전체를 사용하는 것과 비교하여 모델을 경량화 함으로써 GPU 메모리 사용량을 줄이는 것을 확인할 수 있다.

표 1: 제안하는 네트워크와 비교 모델의 평가지표 점수

Approach	Class	Dice	IoU	HD95
Slim UNETR	간	0.9247	0.8633	5.6943
	전립선/자궁	0.5902	0.4507	9.0886
ours	간	0.9263	0.8635	2.8109
	전립선/자궁	0.6039	0.4816	6.1827

표 2 : 제안한 네트워크와 비교 모델의 GPU 메모리 사용량

Approach	Class	GPU Memory (MiB)	
		학습단계	추론단계
Slim UNETR	간	4,314	782
	전립선/자궁	4,450	746
ours	간	1,106	654
	전립선/자궁	1,124	664

4. 결론

본 연구는 미세구조의 구역화를 위해 ROI 영역을 탐지한 후 해당 영역에서 구역화를 수행하는 다단계 네트워크를 제안한다. 제안하는 네트워크는 3D 이미지 구역화 과정에서 더 적은 GPU 메모리를 소모하기 때문에 더 많은 의료 분야에서 3D 이미지 구역화를 수행할 수 있는 환경을 갖추는데 기여할 것으로 보인다. 특히, 메모리 사용량이 제한적인 다양한 임상 환경에서 제안하는 기법을 적용할 수 있을 것으로 기대된다.

참고문헌

[1] H. Chen *et al.*, “Harnessing 2D Networks and 3D Features for Automated Pancreas Segmentation from Volumetric CT Images,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, vol. 11769, in Lecture Notes in Computer Science, vol. 11769., Cham: Springer International Publishing, 2019, pp. 339–347.
[2] Y. Ji *et al.*, “AMOS: A Large-Scale Abdominal Multi-Organ Benchmark for Versatile Medical Image Segmentation”, in *Advances in Neural Information Processing Systems*, pp. 36722–36732, 2022.
[3] Y. Pang et al., “Slim UNETR: Scale Hybrid Trans-formers to Efficient 3D Medical Image Segmentation Under Limited Computational Resources”, IEEE Trans. Med. Imaging, vol. 43, no. 3, pp. 994–1005, Mar. 2024.

엣지 기반 초해상도 이미지 혼합 기법*

김동윤[0], 나재호†
상명대학교 컴퓨터과학과
edw0728@gmail.com, jaeho.nah@smu.ac.kr

Edge-Based Super-Resolution Image Blending Method

Dong-Yun Kim[0], Jae-Ho Nah†
Dept. of Computer Science, Sangmyung University

요약

현재 다양한 특성을 가지는 여러 종류의 SISR(single image super resolution) 모델들이 나와있지만 각 모델은 각기 다른 장단점을 가지고 있다. 본 연구에서는 SwinIR의 classical 및 real-world image SR 모델의 결과물을 엣지(edge) 기준으로 결합하는 방법을 제안한다. 이는 각각의 SR 결과물에서 나타나는 아티팩트(artifact)를 완화하고 이미지 품질을 향상시킬 수 있다.

1. 서론

단일 이미지 초해상도(SISR, single image super resolution) 기술은 저해상도의 단일 이미지 입력을 고해상도 이미지로 변환해 주는 기술로, 현재 의료, 위성, 감시 등 다양한 분야에서 활용되고 있다. 과거 오랜 기간동안 CNN을 기반으로 한 SISR 모델들이 연구되어 왔으나, 최근에는 GAN, Transformer 등 제한된 입력에서 새로운 데이터를 생성해 내는 모델을 이용한 기법들(BSRGAN[1], SwinIR[2], Real-ESRGAN[3] 등)이 점차 주목받고 있다. 하지만 단일 모델 결과만 사용해서는 모든 상황에서 최적의 결과를 얻기는 어렵다.

이러한 측면에서, 본 연구는 SwinIR의 두 가지 모델로부터 얻어진 결과물들을 결합하여 보다 개선된 초해상도(SR) 이미지를 생성하는 방법을 제안한다. 먼저 classical image SR의 결과 이미지로부터 엣지 마스크(edge mask)를 생성하고, 이 마스크를 이용하여 real-world image SR의 베이스 이미지(base image)에 classical image SR로부터 얻어진 엣지 영역을 결합한다.

* 구두발표논문, 학부생 주저자 논문
* 본 논문은 요약논문(Extended Abstract)으로, 현재 연구 진행중인(work in progress) 결과를 포함
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음 (2019-0-018801)
† 교신저자

제안하는 방법의 검증을 위해 Urban 100 데이터셋에서 실험을 수행하였다. 그 결과, 제안하는 방법은 SwinIR의 classical 및 real-world image SR 모델이 각각 제공하는 형태 보존과 선명함의 장점을 동시에 얻어내는 것으로 나타났다.

2. SwinIR

SwinIR[2]은 Swin Transformer를 사용한 대표적인 이미지 복구(restoration) 기법이다. 이 방법은 SR 뿐 아니라 디노이징(denoising) 및 JPEG 압축 아티팩트 감소에도 사용 가능하다.

SwinIR은 SR을 위해 classical, lightweight, real-world image SR의 세 가지 모델을 제공하는데, 앞의 두 모델은 높은 PSNR을 가지도록 훈련시키되 모델의 파라미터 수를 달리한 것으로, 발표 당시 최고 수준(SOTA, state-of-the-art)의 성능을 성취했다. 하지만, 여전히 업스케일링(upscaling) 배수가 높아질수록 CNN 기반 SR과 유사하게 다소 흐릿한 결과물이 도출될 수 있다. 후자는 BSRGAN[1]의 열화 모델(degradation model)을 통해 훈련시켜 이러한 블러링(blurring) 아티팩트를 감소시키지만, 대신 원본의 형태가 왜곡되는 GAN 아티팩트가 나타날 수 있다. 이 모델도 이 분야 SOTA에 해당하는 Real-ESRGAN[3]과 대등한 수준을 보여주었다.

3. 제안하는 초해상도 이미지 혼합 기법

본 연구는 SwinIR의 classical과 real-world image SR의 장점을 취하기 위한 픽셀 단위 혼합(blending) 방법을 제안한다. 이 방법의 목적은 classical image SR의 매끄러운 엣지와 real-world image SR의 선명한 이미지를 픽셀 단위로 취사 선택하는 것이다. 이를 통해 전역적인 단일 혼합 비율을 사용하는 기존 방식[4]과 달리 국부 단위의 화질 향상을 이루어 내고자 한다.

본 연구의 전체적 처리 과정은 [그림 1]과 같다. 본 연구에서는 픽셀별 혼합 선택 기준으로 엣지를 이용하기 때문에 가장 먼저 픽셀의 밝기가 급격하게 변하는 지점을 찾아 이를 엣지로 판별해 낼 필요가 있다. 이를 위해 본 연구에서는 먼저 PSNR 기반 모델(classical image SR)에서 얻은 이미지에 고주파 통과 필터(HPF, high-pass filter)인 Sobel filter를 적용한 후, 추가적으로 팽창(dilate) 기법을 적용하여 최종 결과물이 보다 자연스러워질 수 있도록 한다. 이와 같이 만들어진 엣지 마스크를 이용하여 두 가지 다른 SR 모델(classical 및 real-world image SR)로부터 만들어진 결과물들을 픽셀 단위 보간을 통해 혼합하면 최종 이미지가 완성된다.

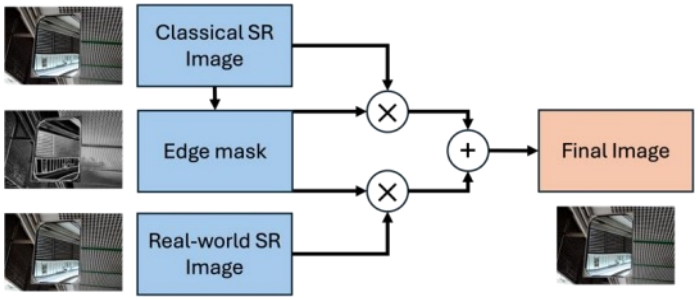


그림 1: 제안하는 엣지 기반 SR 이미지 혼합 과정

4. 실험 결과

본 연구의 실험을 위해 SwinIR의 두 가지 SR 모델(classicalSRx4 및 realSRx4)을 Urban100 데이터셋에 적용하여 4배 업스케일링을 수행하였고, 그 결과는 [표 1]과 같다. 평균 PSNR 및 SSIM 수치는 classicalSRx4가 가장 좋지만, SR 이미지에서는 이러한 메트릭(metric)들의 수치가 실제 사람의 눈이 인지하는 품질과 항상 비례하지는 않는다[4]. 그러므로, 두 SR 모델의 중간 정도로 나타난 본 연구의 결과는 이 두 모델의 특성이 적절히 섞였음을 의미하며, 눈으로 볼 때에는 대체로 realSRx4에 훨씬 가깝다.

표 1: Urban100 데이터셋에서 4배 SR 과정을 수행한 후의 평균 PSNR 및 SSIM 수치

사용 방법	PSNR (dB)	SSIM
classicalSRx4	25.91	0.8062
realSRx4	21.46	0.6735
본 연구	23.75	0.7234

[그림 2]에서는 세 가지 대표적인 결과 이미지별로 품질 비교를 수행한다. 3번 이미지에서 classicalSRx4과 realSRx4는 풀의 질감 표현시 각각 과도한 블러링 및 샤프닝(sharpening) 현상을 보여줬지만, 본 연구는 그 중간 수준의 자연스러운 결과물을 산출하였다. 4번 이미지에서 classicalSRx4와 본 연구는 원형의 구멍을 잘 표현한 반면, realSRx4는 이를 사각형으로 왜곡하였다. 28번 이미지에서 classicalSRx4는 전체적으로 블러링이 매우 심하고, realSRx4는 결과가 선명하지만 타일 형태가 왜곡되었고 발 주위에 링잉(ringing) 아티팩트가 나타났는데, 본 연구는 타일 형태 왜곡까지는 보정하지 못했으나 블러링 및 링잉 현상을 크게 감소시켰다.

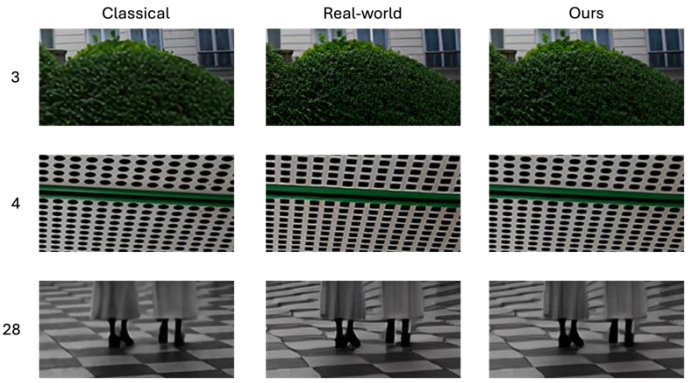


그림 2: Urban100 내 3개 이미지에서의 SR 품질 비교

5. 결론 및 향후 연구

본 연구는 SwinIR의 두 가지 모델을 통해 얻은 이미지들을 엣지 기반으로 혼합하여 개선된 SR 결과물을 얻을 수 있는 방법을 제안하였다. SwinIR은 classical 및 real-world image SR 분야 모두에서 SOTA를 달성하였었기 때문에, 이 두 모델이 보여주는 각기 다른 단점을 동시에 개선한 본 연구는 화질 개선 측면에서 의미가 있다고 생각한다. 본 연구의 혼합 과정은 SwinIR의 훈련 및 추론과는 독립적으로 수행되므로 다른 SR 모델에도 쉽게 확장 적용이 가능하다. 향후 연구에서는 보다 높은 범용성을 위해 여러가지 SR 이미지의 특성을 다양하게 추출하고, 앙상블 기법으로 다수의 결과물을 혼합할 수 있도록 발전시킬 예정이다.

참고문헌

[1] K. Zhang, J. Liang, L. V. Gool, and R. Timofte, “Designing a practical degradation model for deep blind image super-resolution,” in *Proc. of ICCV*, 2021, pp. 4791–4800.
[2] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “SwinIR: Image restoration using Swin Transformer,” in *Proc. of ICCV Workshops*, 2021, pp. 1833–1844.
[3] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,” in *Proc. of ICCV Workshops*, 2021, pp. 1905–1914.
[4] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *Proc. of ECCV Workshops*, 2018, pp. 63–79.

논문 발표

그래픽스 응용

동적 계획법을 통한 빠르고 직관적인 실내 장면 합성*

조윤식[1], 김진모[1,2,*]
한성대학교 일반대학원 정보컴퓨터공학과[1], 한성대학교 컴퓨터공학부[2]
yunsik.cho@hansung.ac.kr, jinmo.kim@hansung.ac.kr

Fast and Intuitive Indoor Scene Synthesis via Dynamic Programming

Yunsik Cho[1], Jinmo Kim[2,*]
Department of Information and Computer Engineering, Graduate School, Hansung University[1],
Division of Computer Engineering, Hansung University[2]

요약

본 연구는 동적 계획법을 이용해 실내 장면 내 객체를 빠르고 직관적으로 재배치하는 새로운 방법을 제안한다. 최적의 객체 배치 접근 방식은 장면의 전체 구조 및 샘플 장면에서 도출된 관계를 기반으로 하는 Top-Down 방식과 객체 간의 하위 관계를 분석하는 Bottom-Up 방식을 혼합하여 사용한다. 각 방법에 대한 비용함수를 활용해 선택, 탐색, 수렴 및 행동의 객체 배치 과정을 통해 빠르고 정확한 장면 합성을 달성한다.

1. 서론

가상현실, 증강현실, 혼합현실을 포함한 확장현실을 기반으로 한 디지털 트윈 기술이 발전하고 있다. 이와 더불어 현실을 디지털 기반의 가상세계로 확장한 가상 공간에서 다양한 활동을 수행할 수 있게 만드는 메타버스는 경제, 사회, 문화 등 여러 분야에서 활용되고 있다. 그리고, 가상 장면 생성은 디지털화된 가상환경에서 시각적 몰입을 높이고 다양한 경험을 제공하는 핵심 기술이다. Yu et al.[1]은 샘플 장면을 토대로 객체 배치의 주요 정보를 추출하고, 마르코프 연쇄 몬테카를로(Marcov Chain Monte Carlo)를 활용한 확률적 최적해 알고리즘인 모의 담금질 기법(Simulated Annealing)을 활용하여 최적화된 장면재배치를 연구하였다. 최근에는 딥러닝을 활용한 장면 합성연구[2]도 수행되었지만, 이는 훈련, 추론시간을 고려하였을 때 최적해 기법보다 느리고 장면 배치 객체 수가 제한적이다.

본 연구는 동적 계획법을 응용한 새로운 실내 장면 합성을 제안한다. 동적 계획법은 복잡한 문제를 간단한 여러 개의 문제로 나누어 푸는 방법이다. 장면을 구성하는 객체 배치 과정에서의 해석을 겹치는 부분 문제와 최적 부분구조 조건에 대응하여 빠르고 직관적인 구조로 장면 합성결과를 도출하는 새로운 방법을 설계한다. 이를 위하여 Top-Down과 Bottom-Up 문제 해결방법으로 구분되는 효율적인 장면 해석과 배치 과정을 설계한다.

2. 동적 계획법 기반 장면 합성

동적 계획법[3]은 복잡한 문제를 더 작은 하위 문제로 분해하여 각 문제를 해결하고 이를 기반으로 원래 문제의 해결책을 도출하는 알고리즘 설계 기법이다. 자연스럽고 최적화된 실내 장면 합성을 위해서 Yu et al.[1]이 제안한 객체 표현, 관계를 활용하여 객체의 기본 속성(위치, 방향)과 샘플로부터 객체와 레이아웃과의 관계, 객체와 객체 간의 관계, 그리고 객체들 사이의 중첩 등을 정의하고, 이를 기반으로 동적 계획법 기반 장면 합성을 수행한다. 그림 1은 본 연구가 제안하는 빠르고 직관적인 실내 장면 합성을 위한 동적 계획법 기반의 장면 합성방법의 전체 구조를 요약한 것이다.

2.1 Top-Down 문제 해결 방법

Top-Down 문제 해결 방법은 큰 문제를 작은 문제로 나누어 해결하는 방식이다. 이는 샘플(Top)로부터 계산된 전체 장면을 구성하는 객체 대한 구조, 정보를 토대로 객체(Down)를 해석한다(그림 1(a)). 식 (1)은 Top-Down 방법에서 사용되는 사전 관계(C_d^{pr}, C_θ^{pr}), 쌍별 관계($C_d^{pair}, C_\theta^{pair}$)에 관한 비용값을 나타낸다.

$$C_d^{pr}(\text{or } C_d^{pair}) = \sum_i \|d_i - \overline{d_i}\|$$
$$C_\theta^{pr}(\text{or } C_\theta^{pair}) = \sum_i \|\theta_i - \overline{\theta_i}\| \quad (1)$$

계산된 비용값을 토대로 제안하는 Top-Down 비용함수는 식 (2)와 같다.

$$C_d^{td} = w_d^{pr} C_d^{pr} + w_d^{pair} C_d^{pair},$$
$$C_\theta^{td} = w_\theta^{pr} C_\theta^{pr} + w_\theta^{pair} C_\theta^{pair} + w_\phi^{pair} C_\phi^{pair} \quad (2)$$

w는 각 비용에 대한 가중치로 0.1~0.01로 정의한다.

2.2 Bottom-Up 문제 해결 방법

Bottom-Up 문제 해결 방법은 작은 하위 문제들로부터 시작하여 점진적으로 큰 문제의 해를 구하는 방식이다. 이는 객체 단위로 해석이 가능한 중첩을 해석하는 과정이다(그림 1(b)). 장면을 구성하는 모든 객체에 대하여 현재 계산 중인 객체의 위치, 방향 값이 최종적으로 결

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출 중.
* 교신저자: 김진모(jinmo.kim@hansung.ac.kr)

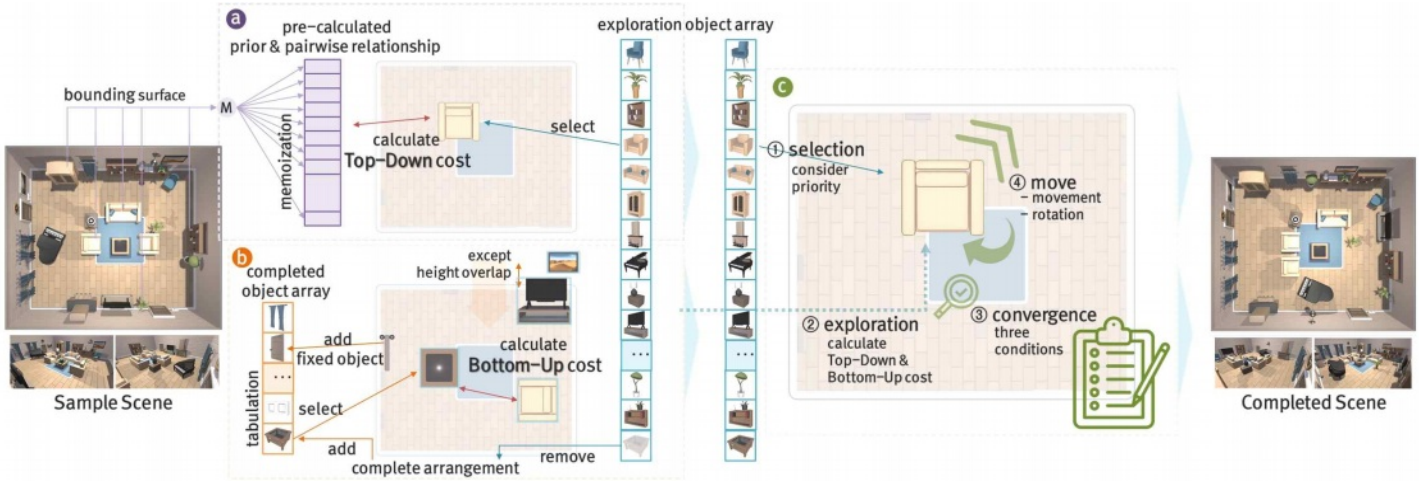


그림 1: 제안하는 동적 계획법 기반의 객체 배치 구조, (a) Top-Down 문제 해결, (b) Bottom-Up 문제 해결, (c) 객체 배치과정

정되면 완료된 객체정보를 배열에 하나씩 채워나간다. 식 (3)은 중첩비용 값과 Bottom-Up 비용함수를 계산하는 과정이다.

$$C_{ac} = \sum_i \sum_j \sum_k \max\left[0, 1 - \frac{\|p_i - a_{jk}\|}{b_i + ad_{jk}}\right], (i \neq j)$$
$$C_a^{bu} = w_{ac} C_{ac} \quad (3)$$

C_{ac}는 현재 객체(i)를 기준으로 배치가 완료된 모든 객체(j)와의 중첩비용(객체의 4방향 볼륨(k)과의 중첩을 계산)을 계산한 결과이며, w_{ac}는 객체의 접근성에 관한 Bottom-Up 비용함수의 가중치(0.05~0.1)이다.

2.3 Process

제안하는 객체 배치과정은 선택, 탐색, 수렴, 그리고 행동의 4단계 흐름으로 하나의 객체씩 순차적으로 진행된다(그림 1(c), 알고리즘 1). 그림 2는 제안하는 방법을 통해 배치된 다양한 배경의 실내 장면 합성결과이다.

그림 2: 다양한 배경의 장면 합성결과

알고리즘 1. Object Arrangement Process

1. **procedure** OBJECT_ARRANGEMENT
2. **exploration object array** ← store the entire objects that make up the scene.
3. **completed object array** ← add objects whose arrangement has been completed.
4. fixed objects are added to the completed object array.
5. **select** objects based on three conditions defined in the exploration object array.
6. **explore** for current object.
7. calculate of C_d^{td}, C_r^{td}, C_a^{bu}.
8. **convergence** judgment based on three defined conditions.
9. **if** convergence **then**
10. add the current object to the completed object array and remove it from the exploration object array.
11. **else**
12. **move** on current object.
13. **end if**
14. repeat steps 5-13 until the exploration object array is empty.
15. **end procedure**

3. 결론

본 연구는 빠르고 직관적인 실내 장면 합성을 위해 장면 합성을 크게 Top-Down, Bottom-Up 방식의 문제 해결방법으로 나누어 동적 계획법 기반 새로운 실내 장면 합성방법을 제안하였다.

참고문헌

[1] L.F. Yu, S.K. Yeung, C.K. Tang, D. Terzopoulos, T.F. Chan, and S.J. Osher, "Make it home: automatic optimization of furniture arrangement," *ACM Trans. Graph*, vol. 30, issue. 4, 86:1-12, 2011.

[2] J. Sun, J. Yang, K. Mo, Y. Lai, L. Guibas, and L. Gao, "Haisor: Human-aware Indoor Scene Optimization via Deep Reinforcement Learning," *ACM Trans. Graph*, vol. 43, issue. 2, 15:1-17, 2024.

[3] R. Bellman, Dynamic Programming, *Dover Publications,* 2013

퍼지 추론 시스템을 이용한 2차원 보행 영상 기반의 발 접촉 판별 네트워크 학습*

이연수[01], 신정환[2], 윤기범[3], 유리[4]
[1]아주대학교 국방디지털융합학과 [2]서울대학교병원 신경과 [3]디딤 [4]아주대학교 소프트웨어학과
dldustnajoo@ajou.ac.kr, neo2003@snu.ac.kr, kibeom.youn@gmail.com, riyu@ajou.ac.kr

Training a Foot Contact Detection Network From 2D Gait Video Using Fuzzy Inference System

Yeonsu Lee[01], Jung Hwan Shin[2], Kibeom Youn[3], Ri Yu[4]
[1]Dept. of Military Digital Convergence, Ajou University
[2]Dept. of Neurology, Seoul National University College of Medicine
[3]Didim
[4]Dept. of Software and Computer Engineering, Ajou University

요약

지면과 발의 접촉을 판별하기 위해 지도학습으로 모델을 학습하는 연구[3]에서는 라벨링을 위해 접촉 여부를 결정하는 기준을 경험적으로 정한다. 이는 대규모 데이터를 처리하기에도 적절하지 않을뿐더러 모델 성능에 큰 영향을 미친다. 이를 해결하기 위해 본 연구에서는 퍼지 추론 시스템(Fuzzy Inference System, 이하 FIS)을 사용한 발 접촉 판단 방법을 제안하고 이를 바탕으로 2차원 보행 영상 기반의 발 접촉 판별 네트워크를 학습한다. FIS는 모호성을 처리하는 데 강점을 가지며, 다양한 보행 종류에 대해 매번 새로운 기준을 설정할 필요 없이 안정적인 성능을 제공할 수 있다. 본 연구의 효과를 입증하기 위해 기존 연구 방식과의 정확도를 비교하였고, 본 연구의 정확도가 더 높음을 확인하였다.

1. 서론

보행 분석은 파킨슨병, 뇌졸중과 같은 질환을 앓고 있는 환자들의 상태를 평가하고 치료 효과를 모니터링하는 데 유용하게 활용될 수 있다. 기존의 센서 기반 보행 분석 방법은 장비가 비싸고 장비 설치 및 센서 착용 등에 많은 시간이 소요되며, 특정 환경에서만 분석이 가능하다는 단점이 있다. 이에 반해 2차원 영상 기반 보행 분석 방법은 고가의 센서 장비 없이 다양한 환경에서 촬영된 보행 영상만으로도 보행 분석이 가능하다는 장점이 있다. 보행 분석에서 보행 주기를 파악하는 것은 중요하며, 이는 발 관절과 지면 간의 접촉 여부를 판별하여 얻을 수 있다.

본 연구에서는 발과 지면 간의 접촉을 판별하기 위해 지도학습 방식으로 네트워크를 학습하기로 한다. 학습을 위해서는 접촉 라벨링이 필요하다. 이에 발-지면 접촉 라벨링을 위해 모션 캡쳐 데이터(.trc)의 3차원 관절 위치 정보를 바탕으로 속력과 가속력을 계산하였다. 그러나 속력과 가속력만으로 접촉 여부를 판단하기 위한 명확한 기준을 설정하는 것은 쉽지 않으며, 모호한 부분이 존재한다. 예를 들어 정상 보행과 병적 보행의 접촉 여부 판별하는 가속력 값은 다르게 설정해야만 한다. 이에 본 연구에서는 모호성에 강한 FIS를 설계하여 네트워크 학습에 필요한 접촉 여부 라벨을 얻어내었다.

* 구두 발표논문
* 학부생 주저자 논문임
* "본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2022-0-01077)

2. 방법

본 연구는 Rempe et al.[1]의 연구를 바탕으로, 보행에 특화된 발 관절 접촉 판별 네트워크를 학습하였다. 네트워크 학습에 다음과 같은 네 종류의 데이터가 사용되었다. 1) 보행 영상으로부터 추출한 프레임별 이미지(.png), 2) 2차원 관절 위치 정보(.json), 3) 프레임별 발 접촉 데이터(.npy), 4) 모션 데이터(.bvh)

2.1. 보행 영상에서 이미지 및 2차원 관절 위치 추출

본 연구에 활용된 보행 영상은 신경과 전문의에 의해 시연된 5가지 보행 데이터(normal walking, parkinsonian walking, freezing of gait, ataxic gait, asymmetric gait)이다. 영상은 정면과 측면에서 촬영되었고, 길이는 1분에서 3분 30초 사이이다. 총 10개의 보행 영상에 대해 Openpose[2]를 활용하여 매 프레임마다 2차원 관절 위치를 얻어내었다.

2.2. 모션 캡쳐 데이터로부터 속력, 가속력 계산

발 관절(발 뒤꿈치, 발가락)과 지면 간의 접촉 여부 라벨을 얻어내기 위해 먼저 모션 캡쳐 데이터(.trc)로부터 속력과 가속력을 계산하였다. 속력은 각 발 관절에 대해 현재 프레임과 이전, 다음 프레임 간의 유클리드 거리의 평균을 구한 후 모션 캡쳐 데이터의 fps를 곱하여 구하였다. 가속력은 연속되는 두 속도 측정값의 차이를 시간 간격으로 나눈 후 절댓값을 취하여 계산하였다.

2.3. 퍼지 추론 시스템을 통한 접촉 여부 라벨 생성

FIS란 불확실성을 처리하는 데 사용되는 컴퓨팅 프레임워크이다. 본 연구에서는 MATLAB을 활용하여 FIS를

설계하였으며 다음과 같이 작동한다.

먼저, 실제 수치 값을 입력받아 이를 언어적으로 모호한 퍼지 집합으로 변환한다. 입력값은 프레임별 속력과 가속력을 각각 해당 보행 영상의 최대 속력과 최대 가속력으로 나눈 0과 1사이의 값이다. 퍼지 집합의 경우 '매우 작음'을 의미하는 SS부터 '매우 큼'을 의미하는 BB까지 크기 순으로 6가지 범주(SS, S, NS, NB, B, BB)로 분류하였다. 이때, 퍼지 집합의 각 요소에 대한 소속도를 나타내는 멤버십 함수를 이용하여 각 요소에 대한 입력값의 소속도를 할당한다. 그림 1은 속력에 대한 멤버십 함수를 나타낸 것이고, 가속력에도 동일한 멤버십 함수가 사용되었다. 멤버십 함수는 불확실성에 강한 Interval Type-2 function으로 설정하여 상한과 하한을 두었다. 예를 들어, 속력 0.65, 가속력 0이 입력되었다고 하자. 그림 1의 멤버십 함수에 따르면, 속력 0.65는 'B' 범주에서 상한 멤버십 값이 0.25, 하한 멤버십 값이 0.0625이며 'NB' 범주에서 상한 멤버십 값이 0.6875, 하한 멤버십 값이 0.75이다. 가속력 0은 'SS' 범주에서 멤버십 값이 1, 'S' 범주에서 멤버십 값이 0이다 .

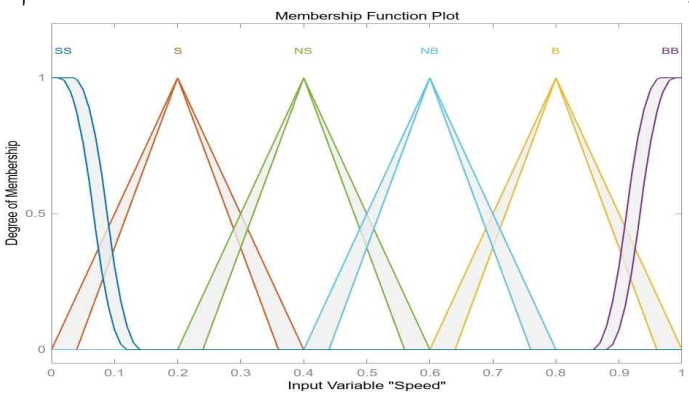


그림 1 속력에 대한 멤버십 함수

이후, 입력과 출력간의 관계를 나타내는 퍼지 규칙을 퍼지 집합에 적용하여 퍼지 출력을 생성한다. 표 1은 본 연구에서 사용한 퍼지 추론 시스템의 퍼지 규칙을 나타낸 것이다.

표1 : 본 연구의 퍼지 추론 시스템에서 사용한 퍼지 규칙

		Acceleration					
		SS	S	NS	NB	B	BB
Speed	SS	SC	SC	SC	WC	WC	WC
	S	WNC	WNC	WNC	WNC	WNC	WNC
	NS	WNC	WNC	WNC	WNC	WNC	WNC
	NB	WNC	WNC	SNC	SNC	SNC	SNC
	B	SNC	SNC	SNC	SNC	SNC	SNC
	BB	SNC	SNC	SNC	SNC	SNC	SNC

출력값의 경우 '강한 접촉'을 의미하는 SC부터 '강한 미접촉'을 의미하는 SNC까지 4가지 범주(SC, WC, WNC, SNC)로 분류하였고, 각각 1, 0.7, 0.3, 0의 상수값이다. 속력 0.65은 'NB', 'B' 범주 모두에 해당하고, 가속력 0은 'SS' 범주에만 해당한다. 표 1에 의하여 Speed = NB and Acc = SS then output = WNC, Speed = B and Acc = SS then output = SNC 의 퍼지 규칙이 활성화된다. 이러한 퍼지 규칙은 앞서 할당된 멤버십 값을 바탕으로 강도를 가지게 되는데, 이 경우 output = WNC가 되는 퍼지 규칙의 강도가 output = SNC가 되는 퍼지 규칙의 강도보다 강하다. 본 연구에서는 가중평균 방법을 이용하여, 규칙의 강도와 해당 규칙의 출력값(WNC는 0.3, SNC는 0)을 곱한 값들의 합을 모든 활성화된 규칙의 강도 합으로 나누어 최종 출력값을 산출하였다. 최종 출력값은 0과 1사이의 값으로 접촉 정도를 나타내며 이 값이 0.5보다 크면 접촉 0.5 이하이면 미접촉이라 판단하였다. 접촉 여부는 매 프레임마다 .npy 파일로 저장하였다.

3. 결과

본 연구의 유효성을 입증하기 위해 기존 연구와의 정확도 비교를 수행하였다. 기존연구[3]는 보행 종류에 따라 속력과 가속력의 threshold를 달리하여 접촉 유무를 결정한 데이터셋을 사용하여 네트워크를 학습하였다. 반면, 본 연구에서는 기존 연구[3]과 동일한 보행 데이터를 사용하되 퍼지 추론 시스템으로 생성한 라벨을 이용하여 네트워크를 학습하였다.

표 2는 학습한 데이터셋과 다른 환경, 다른 사람의 보행 영상과 압력 센서(pressure sensor mat)를 통해 얻은 ground-truth를 바탕으로 연구[3]과 본 연구의 정확도를 비교한 것을 나타낸 것이다. 왼쪽 발, 오른쪽 발에 대해 본 연구에서 제안한 네트워크가 더 높은 정확성을 보였다.

표2 : 정확도 비교 (단위 : %)

	left foot	right foot
연구[3]	74.99	74.63
본 연구	**81.14**	**84.17**

4. 결론 및 후속 연구 방향

본 논문에서는 접촉 여부 라벨을 얻기 위해 FIS를 사용하는 방식을 제안하였다. FIS는 한 번 설계하면 다양한 보행 패턴에 대해 재조정할 필요가 없어 효율적이다. FIS를 사용하여 얻은 데이터로 학습된 네트워크는 보행 종류에 따라 경험적으로 접촉 판단 threshold를 설정하는 방법과 거의 유사한 정확도를 보였다. 또한, 새로운 데이터에도 강건함을 보였다.

본 연구는 한 사람의 보행 데이터를 기반으로 네트워크를 학습하여 데이터 규모가 작다. 이때 다양한 사람의 보행 데이터를 포함한다면 네트워크의 정확성과 강건함을 더욱 높일 수 있을 것이다.

참고문헌

[1] Davis Rempe, Leonidas J Guibas, Aaron Hertzmann, Bryan Russel, Ruben Villegas, and Jimei Yang, Contact and Human Dynamics from Monocular Video, in European Conference on Computer Vision(ECCV), 2020

[2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)

[3] 이연수, 유리, 2차원 보행영상으로부터 보행주기를 판별하는 네트워크 학습, 한국 컴퓨터 그래픽스 하계 학술대회 논문지, 2023.7, pp.95-96

논문 발표

강화학습

IB-LBM 시뮬레이션을 이용한 물고기 동작 정책 학습

최완욱 [0], 이윤상
한양대학교 기계기술연구소, 한양대학교 컴퓨터소프트웨어학부
{dhksdnr2003, yoonsanglee}@hanyang.ac.kr

Fish Behavior Policy Learning Using Immersed Boundary-Lattice Boltzmann Methods

Wanuk Choi[0], Yoonsang Lee
Mechanical Engineering and Technology Research Institute, Hanyang University
Dept. of Computer Science, Hanyang University

요약

본 연구에서는 유체와의 상호작용이 중요시되는 물체의 움직임을 구현하는 방법으로 강화학습 환경에 IB-LBM 을 결합하는 시스템을 제안하고, 위 환경에서 사실적인 시뮬레이션이 되는지를 물 속 물고기 객체를 움직이는 동작정책 강화학습 결과를 통해 보인다.

1. 서론

유체 내 물체의 사실적인 행동 시뮬레이션에 대한 수요는 다양한 산업 분야에서 지속적으로 존재해 왔다. 예를 들어, 영화 산업에서는 '니모를 찾아서'에서 만타가오리, 블루탱, 백상아리 등 각 생물의 생체특성에 맞는 시뮬레이션이 사용되었고, 수산업에서는 컴퓨터 비전처리를 활용한 양식장 관리를 위해, 물고기 군집의 시뮬레이션을 학습에 활용하는 연구가 수행되었다[1].
위 수요에는 물속 생물의 움직임을 시뮬레이션하는 데 있어 객체별 움직임을 손쉽게 생성할 수 있게 됨으로써 개선될 수 있는 점들이 존재한다. 예를 들어, 앞선 예시에서는 종별 행동 특성을 연구하고 모션 동작을 조정하는 과정을 생략할 수 있으며, 후자의 경우 시뮬레이션에 객체의 감가속에 맞는 움직임을 추가하여 더 사실적인 결과를 얻을 수 있다. 이에 본 논문에서는 물고기의 형상정보인 뼈의 구조 및 물고기의 외형정보를 바탕으로 two-way coupling 된 강화학습환경에서 행동정책을 학습함으로서 사실적인 시뮬레이션을 쉽게 획득할 수 있는 방법을 제안한다.

* 구두 발표논문, 요약논문 (Extended Abstract)
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비 중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 및 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

2. 방법

2.1. 격자볼츠만 방법(LBM)

3 차원 공간상에서 속도분포를 19 개의 방향으로 이산화한 단위격자 내의 속도 별 입자분포함수 f_i 를 식(1)의 이산화된 격자볼츠만방정식을 통해 풀이하였다[2]

$$f_i(x + c_i\Delta t, t + \Delta t) - f_i(x,t) = \Omega_i(x,t) + F_i(x,t)\Delta t \quad (1)$$

수식에서 c_i는 이산화된 입자의 속도를, Ω_i는 충돌 연산자를, F_i 는 외력에 의한 입자분포함수의 변화량을 나타낸다. 충돌연산 모델로 식(2)의 BGK 충돌모델[3]을 적용하였다:

$$\Omega_i(x,t) = -\frac{f_i - f_{eq}}{\tau}\Delta t\,. \quad (2)$$

식에서 f_{eq}는 평형분포함수를, τ는 완화계수를 나타낸다. 외력에 의한 입자분포함수의 변화량에 관한 식은 비압축성을 가정하여 아래의 수식을 적용하였다:

$$F_i(x,t) = w_i\frac{c_{i\alpha}}{{c_s}^2}F_\alpha(x,t)\,. \quad (3)$$

식에서 c_s 는 입자의 속력를, $c_{i\alpha}$ 는 입자속도의 α 방향 성분을, w_i 격자 내 이산화 가중치를 나타낸다. 유체의 밀도 ρ, 속도 u, 압력 p 는 아래 식을 통해 계산되었다:

$$\rho = \sum_i f_i\,, \qquad u = \frac{1}{\rho}\sum_i c_i f_i\,, \qquad p = c_s^2\rho\,. \quad (4)$$

2.2. 가상경계-격자볼츠만 방법(IB-LBM)

물체를 경계면 위에 여러 점들로 대체하여 연산을 진행하는 IB-LBM 방법에서, 물체의 j 번째 경계점 $r_j(t)$와 x 위치의

유체격자 사이에 연산을 수행하기 위한 이산화 디락-델타함수 $\Delta(r_j(t),x)$를 아래와 같이 정의하였다:

$$\Delta(r_j(t),x)=\frac{1}{\Delta x^3}\prod_{\alpha=1}^{3}\emptyset\left(\left(r_j(t)-x\right)_\alpha\right). \quad (7)$$

$\emptyset(z)$는 Peskin[4]이 정의한 커널함수를 사용하였다.
물체의 j 번째 경계점에서의 유체속도 $\dot{r}_j$은 이산화 디락-델타함수를 사용하여 식(8)을 통해 구했다:

$$\dot{r}_j(t)=\sum_x \Delta x^3 u(x,t)\Delta(r_j(t),x). \quad (8)$$

해당지점에서의 유체와 경계점의 상대속도에 비례하는 힘 $-f_j(t)$를 강체에 작용, 유체엔 아래 수식을 통해 구한 $F(x,t)$를 식(3)에 대입함으로 상호작용을 구현하였다:

$$F(x,t)=\sum_j f_j(t)\Delta(r_j(t),x). \quad (9)$$

2.3. 물체의 경계점 생성 및 변형

IB-LBM 연산을 위해 물체의 경계면에 분포되는 점들은 Poisson Disk Sampling[5] 방법을 통해 생성하였다. 경계점의 속도와 변위는 Linear Blend Skinning 방법을 통해 뼈의 운동학 정보를 바탕으로 계산하였다.

2.4. 강화 학습

PPO[6] 알고리즘을 통해 학습을 진행했으며, 에이전트의 관측값으로 목표지점과 머리의 상대각도, 머리의 속도 및 각속도, 뼈 관절점에서의 자유도 위치, 자유도 속도, 자유도 가속도, 유체로 받는 힘과 토크를 사용하였다. 그리고 한스텝동안 목표방향으로 단축시킨 거리에 타임스텝을 나눈 값을 보상으로 주도록 설정하였다.

3. 실험 및 결과

원점을 기준으로 x, y, z 방향으로 각각 100,60,60 만큼의 격자를 채운 유체 도메인에 (20, 30, 30) 지점에서 시작해 목표위치인 (80,30,30) 으로 향하도록 움직이는 정책을 학습하였다. 학습된 정책을 수행하는 물고기 주변의 압력분포를 그림 1 로 도식화하였다 (좌상단을 기준으로 우하단까지 시열 순). 이를 통해 절반의 주기를 움직이는 동안 주변 유동장으로부터 압력을 받아가며 앞으로 나아가려 함을 확인할 수 있었다.

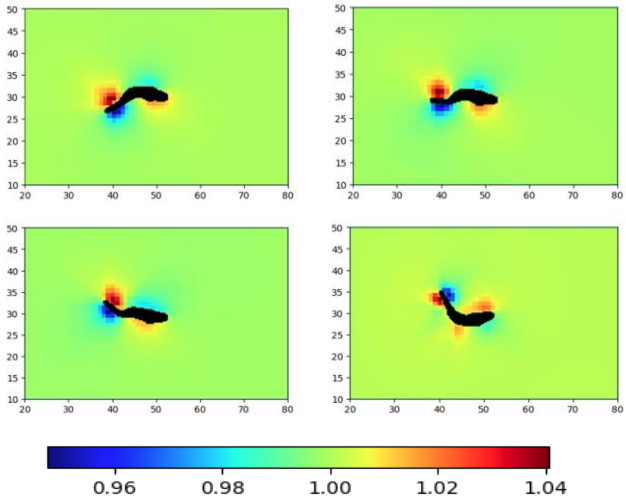


그림 1: 학습된 정책을 통해 움직이는 물고기 주위의 압력분포

4. 결론

강체와 유체가 상호작용할 수 있는 강화학습환경을 IB-LBM 을 통해 구현함으로서, 뼈대와 외형을 바탕으로 물속에서의 사실적인 움직임을 강화학습을 통해 시뮬레이션 할 수 있음을 보였다. 위 과정에서 유체 연산이 GPU 연산장치의 계산 성능에 큰 영향을 받기에, Multi-GPU 를 통한 연산이 가능하도록하여 더 넓은 유체도메인에서 다양한 정책을 학습하는 방향으로 개선연구를 진행할 것이다.

참고문헌

[1] Ishiwaka, Y., Zeng, X. S., Ogawa, S., Westwater, D., Tone, T., and Nakada, M. DeepFoids: Adaptive Bio-Inspired Fish Simulation with Deep Reinforcement Learning, *NeurIPS*, 2022.
[2] Timm, K., Kusumaatmaja, H., Kuzmin, A., Shardt, O., Silva, G., and Viggen, E. *The lattice Boltzmann method: principles and practice.* Cham, Switzerland: Springer International Publishing AG, 2016
[3] Bhatnagar, P. L., Gross, E. P., and Krook, M. A model for collision processes in gases, *Physical Review,* vol. 94, No. 3 pp. 511-525, 1954
[4] Peskin, C. S. The immersed boundary method, *Acta Numerica*, vol. 11, pp. 479-517, 2002
[5] Yuksel, C. Sample Elimination for Generating Poisson Disk Sample Sets, *Computer Graphics Forum,* vol. 34, No. 2, pp. 25-32, 2015
[6] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017

물리 환경 속 가상의 탄도적 스턴트 연기 동작 생성*

김민석[0], 서원정, 이성희, 원정담
서울대학교 컴퓨터공학과, KAIST 문화기술대학원
{minseok, wjeong.west }@imo.snu.ac.kr, sunghee.lee@kaist.ac.kr, jungdam@imo.snu.ac.kr

ViSA: Physics-based Virtual Stunt Actors for Ballistic Stunts

Minseok Kim[0], Wonjeong Seo, Sung-Hee Lee, Jungdam Won
Dept. of Computer Science and Engineering, Seoul National University
Graduate School of Culture Technology, KAIST

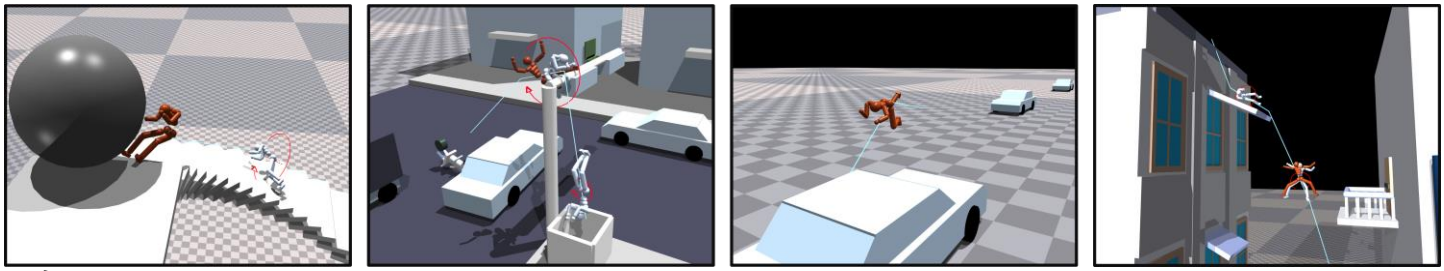


그림 1: ViSA를 통해 생성한 탄도적 스턴트 연기 동작 생성 예시. 왼쪽부터 나선형 계단 추락, 오토바이 사고, 다중 교통사고, 건물 추락 시나리오에서 사용자가 입력한 궤적과 자세에 맞춰 스턴트 동작을 수행하는 모습.

요약

본 논문에서는 가상 환경에서 탄도적 스턴트 연기 동작을 생성할 수 있는 시스템 ViSA를 제안한다. 영화 제작 및 TV 프로덕션에서는 사람이 차에 치여 날아간다든가, 높은 빌딩에서 떨어진다든가 하는 위험하고 극적인 탄도적 스턴트 동작들을 자주 볼 수 있다. 우리의 시스템은 스턴트 장면에 유저가 원하는 공간 제약을 설정하면, 주어진 제약을 충족하는 물리적으로 타당한 동작을 생성한다. 본 논문에서는 새로운 축소된 상태 및 행동 공간에 더해 직관적이고 효과적인 보상을 통합한 심층 강화 학습 시스템을 제안한다. 이 시스템을 통해 사용자는 몇 분 내에 빠른 응답을 받을 수 있으며 상호작용 방식으로 전체의 긴 장면을 계속 만들어낼 수 있게 된다.

1. 서론

스턴트 배우는 화재/가스에 의한 폭발, 차량 충돌, 높은 곳에서의 추락과 같이 대담하고 위험한 행동을 수행하는 배우이다. 극적인 연출을 위해 배우들이 위험에 처하는 상황은 자주 발생하면서도 스턴트 장면들은 영화의 인기 요소로 자리잡게 되었고, 스턴트 장면이 없는 액션 영화를 상상하기 어려워 졌다. 현대 영화 제작에서는 디지털로 스턴트 장면을 만드는 경우가 많아지긴 했지만, 스턴트 장면을 제작하는 것은 상당히 도전적인 문제일 수 밖에 없다. 첫째, 물리학에 대한 광범위한 지식을 요구하며 두번째로는 애니메이션 곡선이 시각적 및 물리적 현실성을 충족하면서 감독이 제시하는 특정 시공간 제약에 맞춰야 하기 때문이다. 이러한 제약들 때문에 감독들은 여전히 실제 스턴트 장면을 촬영하는 것이 디지털 접근 방식보다 궁극적으로는 더 효율적이라고 여기고 있다.

본 논문에서는 물리적으로 시뮬레이션된 가상 스턴트 배우를 사용하여 탄도적 스턴트 연기 동작을 생성하는, 유저 상호작용이 가능한 애니메이션 시스템 ViSA를 소개한다. 물리 시뮬레이션 환경에서 특정 스턴트 장면이 주어지면, 사용자는 직관적인 인터페이스를 통해 스턴트 배우의 이동 궤적과 그 사이의 자세를 설정할 수 있다. 그 뒤 우리 시스템은 심층 강화 학습을 통해 사용자의 입력을 만족하는 물리적으로 타당한 애니메이션을 효율적으로 생성한다. 물리 환경 속에서 주어진 궤적을 따라가는 기존 연구들은 몇 시간의 학습 시간, 혹은 사전 학습 시간이 필요했던 반면[1,2], 우리 시스템의 새로운 강화 학습 환경 구성과 스턴트에 특화된 가속 기술을 통해 모션의 생성 시간이 단 몇 분 이내에 이루어지며, 사용자는 상호작용 방식으로 장면들을 수정하거나 계속해서 이어 붙여 나갈 수 있다.

2. 시스템 개요

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출중.

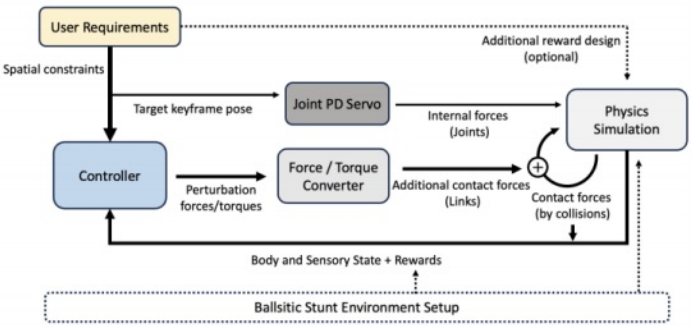


그림 2: 시스템 개요

우리 시스템이 제공하는 인터페이스를 사용하여 사용자가 생성한 공간적 제약의 예를 그림 3 에서 볼 수 있다. 스턴트 배우와 주변 객체를 포함한 물리적으로 시뮬레이션 되는 환경이 주어지면, 사용자는 시작 위치와 종료 위치로 정의된 3D 선형 구간을 설정한다. 이 때에 필요하다면 여러 개의 선형 구간을 이어 긴 곡선 경로를 생성할 수 도 있다. 각 선형 구간에 대해 사용자는 특정 자세, 몸(root)의 방향, 몸의 각속도, 최대 무게중심 선형 속도를 삽입할 수 있다. 사용자가 위의 설정을 완료하고 나면, 우리 시스템은 심층 강화 학습을 통해 시뮬레이션된 스턴트 배우의 제어기를 학습하여 물리적으로 타당한 스턴트 애니메이션을 생성한다.

3. 학습 환경 및 방법

가상의 스턴트 배우는 각 관절에 PD모터가 장착된 관절형 강체 시스템으로 모델링 되어있다. 기존의 물리 기반 캐릭터 컨트롤러와의 주요 차이점 중 하나는, 우리 시스템에서 캐릭터는 접촉 힘의 섭동(perturbation)으로 제어된다는 것이다. 기존 방법들은 PD 모터에 가하는 목표 자세를 통해 캐릭터를 제어한다. 하지만 우리 시스템에서는 사용자가 미리 지정한 키프레임 애니메이션을 추적하기 위해서만 PD 모터를 사용하고, 무게중심의 궤적, 몸의 방향 등을 사용자의 제약에 맞추는 작업은 접촉 힘의 섭동을 제어함으로써 이루어진다.

접촉 힘의 섭동을 제어한다는 아이디어는 *다중세계 탐색*(Many-worlds browsing)에서 사용되는 개념에서 영감을 받았다. *다중세계 탐색* 은 물체들 간의 충돌에 의해 발생한 접촉 힘에 낮은 비율로 왜곡을 가해도 사람은 그것에 왜곡이 있었는지 잘 인지하지 못한다는 점을 이용한다. 이전의 연구에서는 동일한 다수의 환경에서 섭동을 임의로 가하여 서로 다르게 시뮬레이션 한 뒤에, 유저가 원하는 결과물들을 직접 골라냈다[3]. 우리는 이 섭동을 직접적으로 강화학습에서 제어한다. 우리의 강화학습 시스템에서 행동은 섭동 힘, 섭동 돌림 힘 총 6개의 차원으로 구성되어있다.

3.1 섭동 힘, 돌림 힘 변환

행동으로 주어지는 섭동 힘, 돌림 힘은 몸의 무게중심에

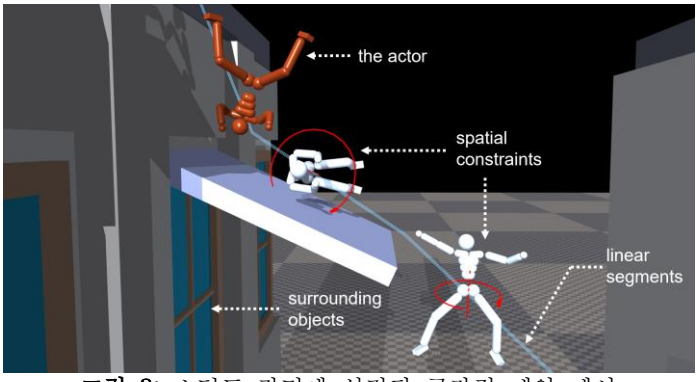


그림 3: 스턴트 장면에 설정된 공간적 제약 예시

가해지는데 무게중심은 가상의 작용점이므로, 실제로는 몸에 직접 가해주어야 한다. 캐릭터는 자연스러운 동작을 위해 낮은 PD 계수가 설정되어있기 때문에(기존의 5%정도) 몸의 한 부위에만 이 힘들을 적용한다면 몸이 와이어에 매달린 것처럼 끌려 다니는 모션이 나온다. 우리 시스템에서는 이 힘들을 몸에 고루 나누어 주기 위해 선형시스템을 풀어 몸의 각 부위에 나누어 주었다.

4. 실험 및 결과

본 실험은 IsaacGym 물리 시뮬레이터에서 이루어졌으며 제어기는 Proximal Policy Optimization (PPO) 알고리즘을 통해 학습되었다. 우리는 그림 1 에서처럼 (1) 거대한 나선형 계단을 따라 굴러 내려가거나 (2) 빠른 속도로 달리는 오토바이가 차에 부딪혀 사람이 지형지물에 걸쳐지거나 (3) 반대편에서 달려오는 차들에 순차적으로 부딪히거나 (4) 빌딩 옥상부터 건물 난간에 차례로 부딪혀 바닥까지 떨어지는 시나리오 등을 통해 우리의 시스템의 활용 가능성에 대해 보였다.

5. 결론

본 시스템을 활용하여 상업 영화 수준의 고품질 스턴트 동작을 복잡하고 위험한 시나리오에서도 수 십 분 내에 생성할 수 있음을 확인했다. 또한 본 논문에서 제안한 적절한 환경 설계와 힘의 섭동을 통한 캐릭터 제어 기법이 결합되었을 때, 자연스러운 동작을 구현할 수 있음을 입증했다.

참고문헌

[1] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. 2021. Amp: Adversarial motion priors for stylized physics-based character control. ACM Transactions on Graphics (TOG) 40, 4 (2021). 1-20

[2] Jungdam Won, Deepark Gopinath, and Jessica Hodgins. 2022. Physics-based character controllers using conditional vaes. ACM Transactions on Graphics (TOG) 41, 4(2022),1-12

[3] Christopher D Twigg and Doug L James. 2007. Many-worlds browsing for control of multibody dynamics. In ACM SIGGRAPH 2007 papers. 14-es.

강화학습을 이용한 나비의 비행 동작 생성*

[1]정은호[0], [2]장이권, [1]이윤상
[1]한양대학교 컴퓨터·소프트웨어학과, [2]이화여자대학교 에코과학부
[1]{jho6394, yoonsanglee}@hanyang.ac.kr, [2]jangy@ewha.ac.kr

Butterfly Flight Generation Using Deep Reinforcement Learning

[1]Eunho Jung[0], [2]Yikweon Jang, [1]Yoonsang Lee
[1]Dept. of Computer Software Engineering, Hanyang University,
[2]Division of EcoScience, Ewha Womans University

요약

나비의 비행은 다른 곤충의 비행과 비교하여 훨씬 역동적이면서 독특한 모습을 보인다. 이러한 나비의 비행을 재현하기 위해서는 생물학적인 특성을 고려하는 것이 필수적이다. 본 연구에서는 나비의 생물학적인 특성을 반영한 나비 모델과 심층강화학습 설계를 통해 사실적인 나비의 비행 동작을 재현할 수 있음을 보이고자 한다.

1. 서론

인간 외의 동물의 움직임을 컴퓨터로 재현하려는 시도는 상당히 오래전부터 이어져왔다. 근래 들어서는 모션 캡처 데이터의 확보가 상대적으로 용이한 네발 동물의 움직임을 생성하는 연구 결과들이 발표된 바 있다 [1, 2]. 이에 비해 곤충의 움직임은 상대적으로 많이 다루어지지 않았는데, 특히 비행 곤충인 나비는 불규칙적이고 역동적인 움직임을 보이지만 크기가 작아 모션 캡처에 어려움이 있다는 특성이 있다. Chen [3] 등은 미리 정의된 규칙에 의해 나비의 날갯짓 커브의 주기와 진폭을 나비의 속력에 따라 조절하고, 목표 경로를 따르도록 가상의 힘을 나비에 가하는 방식으로 실시간으로 나비 비행 동작을 생성하는 연구 결과를 발표한 바 있다.

본 연구는 비교적 단순화된 공기역학을 포함한 물리 시뮬레이션과 생물학적인 근거에 기반하여 설계된 강화학습 모델을 이용하여 실시간 나비 애니메이션을 생성할 수 있는 시스템을 제안한다. 제안하고자 하는 방법론에는 인위적인 힘과 가속도가 필요치 않고, 강화학습을 사용해 변화를 즉각적으로 반영할 수 있다는 점에서 기존 연구와 차별점이 있다. 또한 생물학 연구들에 근거한 나비의 특성을 학습에 반영함으로써 더욱 자연스러운 움직임을 생성해 낼 수 있었기에, 생물학적으로도 큰 의미를 지닌다.

2. 나비의 역학적 특성 및 생물학적 특성 반영

2.1. 공기역학적 특성

공기역학을 계산할 때는 Chen [3]등의 연구와 같이 아래의 단순화된 공기역학 식을 사용하였다.

$$F_{lift} = \frac{1}{2}\rho A_i |V_i|^2 C_l(\alpha), \qquad F_{drag} = \frac{1}{2}\rho A_i |V_i|^2 C_d(\alpha)$$

ρ는 공기의 밀도, A_i는 나비의 날개 단면을 구성하는 삼각형들 중 i번째 삼각형, V_i는 i번째 삼각형의 속도, $C_l(\alpha)$과 $C_d(\alpha)$는 α(angle of attack)에 따른 양력 계수와 항력 계수이다. 조금 더 정확한 값을 활용하기 위해, 나비와 같은 크기의 모형으로 양력 계수와 항력 계수를 측정한 Ancel [4]등의 연구를 참고하여 그림 1과 같이 함수식을 구성하여 사용하였다.

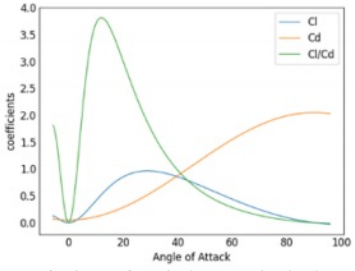


그림 1: 시뮬레이션에 사용된 나비의 공기역학계수

또한, 나비가 공기 역학으로 인해 발생하는 돌림힘을 활용해 방향전환을 학습할 수 있도록, 계산된 양력과 항력이 가해진 위치를 활용해 각 관절부를 기준으로 돌림힘을 계산하여 시뮬레이션에 적용하였다.

---

* 구두 발표논문, 요약논문 (Extended Abstract)
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 및 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

2.2. 생물학적 특성

곤충의 날개에는 앞날개와 뒷날개의 움직임을 연동시키는 다양한 커플링 메커니즘이 존재한다. 그 중 나비의 커플링 메커니즘은 포착형 결합(amplexiform coupling)으로, 앞날개와 뒷날개의 일부분이 겹쳐짐으로써 힘이 전달되는 메커니즘을 가지고 있다. Ma [5]등은 포착형 결합이 사라지면 비행은 가능하지만 날개 간 연동성이 확연하게 떨어진다는 실험 결과를 보여, 자연스러운 날갯짓에 포착형 결합의 영향이 크다는 것을 입증하였다. 포착형 결합을 재현하기 위해, 시뮬레이션에서 메시 간 충돌을 활성화하여 앞날개와 뒷날개가 서로에게 힘을 전달할 수 있도록 하였다.

나비는 상승 비행에서 가장 격렬한 날갯짓을 보이는데, 상승 비행시의 날개 관절의 움직임에 대한 Roy [6] 등의 연구를 참고하여 각 관절의 가동 범위를 설정하였다. 물리 시뮬레이션에 큰 영향을 주는 각 부위의 무게 또한 말린 표본을 기준으로 무게를 측정한 Andrew [7] 등의 연구에 기반하여 설정하였다. 다만 살아 있을 때와 말린 표본 간의 무게 차이가 체액의 유무로 인해 다를 수 있다는 점, 측정되지 않은 부위들이 있다는 점을 고려하여, 실험적으로 일부 조정하여 사용하였다.

3. 강화학습 설계

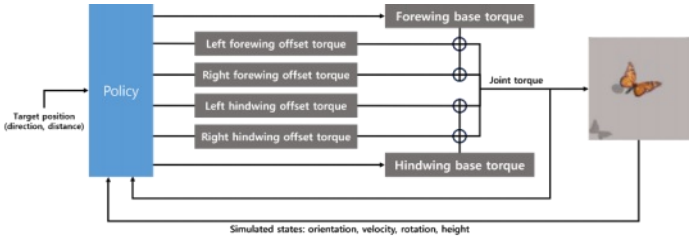


그림 2: Base Action Model

나비의 날갯짓은 대칭을 이루고, 약간의 날개 각도 조절과 배 관절의 움직임 만으로 방향을 바꿀 수 있다는 점에서 착안하여, 그림 2와 같은 Base Action Model을 제안한다. 해당 모델은 Policy의 출력을 그대로 사용할 때에 비해 보상 함수 설계를 단순화할 수 있게 하였고, 학습 파라미터 튜닝 또한 쉽게 할 수 있었다.

Policy는 입력으로 목표 지점과의 상대적인 거리와 방향벡터, 현재 높이, 공기역학으로 인한 힘과 돌림힘, 이전 프레임에서의 action과 6D rotation representation을 받는다. 입력으로 받는 값들은 나비의 local 좌표계를 xy평면에 사영한 좌표계로 표현하였다.

학습에 이용되는 보상함수는 목표 지점과의 거리에 따른 거리 보상, 속도가 정면 방향이도록 하는 비행 방향 보상, 각 관절에 가하는 힘이 클수록 페널티를 주는 에너지 페널티, 떨림을 방지하는 규제 보상으로 구성하였다.

4. 실험결과

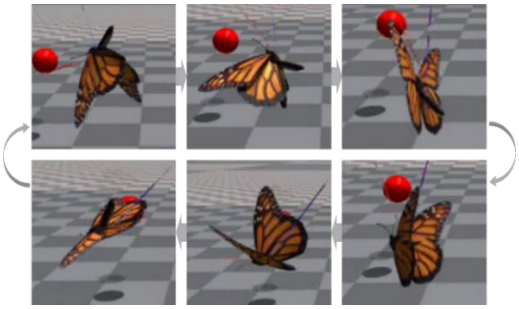


그림 3: 학습된 Model로 재현한 나비의 날갯짓

그림3과 같이 나비의 비행 간 생물학적 특징인 날개와 배의 움직임의 주기성, 그리고 앞날개와 뒷날개 간의 포착형 결합을 잘 보이는 나비 비행 정책 학습 결과를 얻을 수 있었다.

참고문헌

[1] Zhang, He, Sebastian Starke, Taku Komura and Jun Saito. “Mode-adaptive Neural Networks for Quadruped Motion Control”. *ACM Trans. Graph.* 37, 호 4: 145:1-145:11. 2018.

[2] Taesoo Kwon, Hoimin Kim, Yoonsang Lee. Control of an Iguana Character Using Soft-Body Simulation. *IEEE Access*, Volume 6, 77931-77939, December 2018.

[3] Qiang Chen, Tingsong Lu, Yang Tong, Guoliang Luo, Xiaogang Jin, and Zhigang Deng. A practical model for realistic butterfly flight simulation. *ACM Trans. Graph*., 41(3), mar 2022.

[4] Ortega Ancel, A., Eastwood, R., Vogt, D., Ithier, C., Smith, M., Wood, R., & Kovač, M. Aerodynamic evaluation of wing shape and wing orientation in four butterfly species using numerical simulations and a low-speed wind tunnel, and its implications for the design of flying micro-robots. *Interface focus*, *7*(1), 20160087. 2017.

[5] Ma, Y., Zhao, H., Ma, T., Ning, J., & Gorb, S. Wing coupling mechanism in the butterfly Pieris rapae (Lepidoptera, Pieridae) and its role in taking off. *Journal of insect physiology*, *131*, 104212. 2021.

[6] Camille Le Roy, Nicolas Silva, Ramiro Godoy-Diana, Vincent Debat, Violaine Llaurens, Florian Titus Muijres; Divergence of climbing escape flight performance in Morpho butterflies living in different microhabitats. *J Exp Biol*, 225 (15), August 2022.

[7] Andrew K. Davis, Michael T. Holden, Measuring Intraspecific Variation in Flight-Related Morphology of Monarch Butterflies (Danaus plexippus): Which Sex Has the Best Flying Gear? *Journal of Insects*, vol. 2015, Article ID 591705, 6 pages, 2015.

포스터발표

XR환경에서 사용자-아바타 상호작용을 위한 음성 기반 감정인식 방법*

이민정[0,1], 신영훈[1], 이미란[†,2]
대구대학교 일반대학원 IT융합공학과[1], 대구대학교 컴퓨터정보공학부(컴퓨터소프트웨어전공)[2]
{minjeong.hci.du, younghoon.hci.du}@gmail.com, miran@daegu.ac.kr([†]교신저자)

Speech-based Emotion Recognition Method for User-Avatar Interaction in XR Environment

Minjeong Lee[0,1], Younghoon Shin[1], and Miran Lee[†,2]
Graduate School of IT Convergence Engineering, Daegu University[1]
Department of Computer and Information Engineering, Daegu University[2]

요약

증강(AR)·가상(VR)·혼합(XR) 현실을 활용한 심리치료, 재활 등의 다양한 콘텐츠가 등장하면서, 이러한 가상현실 기술들은 노년층의 우울증 등의 정서적 문제를 해결하는 새로운 대안으로 주목받고 있다. 본 논문에서는 VR기기에 내장된 마이크를 통해 사용자의 음성을 녹음하고, ResNet 모델을 이용하여 4종의 감정을 분류하는 방법에 대해 제안한다. 음성을 통해 사용자 내면의 감정을 인식하고, 이를 AI 아바타 비서가 적절한 표정으로 표현함으로써 감정적 공감을 받을 수 있는 환경을 제공하며, 향후, 실시간으로 건강 상태 모니터링을 통해 간호사 보조를 할 수 있는 용도로도 활용될 예정이다.

1. 서론

OECD(Organization for Economic Co-operation and Development, 경제협력개발기구) Health Statistics에 따르면, 2022년 기준 임상 간호 인력 수는 인구 1,000명당 8.4명으로 OECD 평균 9.7명보다 1.3명 적게 나타났다[1]. 정부는 이를 해결하기 위해 부족한 간호 인력을 충원하여 담당 환자의 수를 5명으로 낮추고자 노력하는 중이다. 그러나, 2022년 통계청 자료에 따르면 65세 이상 인구는 9,018,412명이며, 매년 증가하는 추세이다[2]. 따라서 간호 인력을 충원하더라도 65세 이상 인구가 증가하는 추세를 따라잡기엔 어려우며, 이로 인해 간호 인력은 여전히 부족할 것으로 보인다.

노인 환자의 경우 만성질환과 더불어 우울은 흔히 나타나는 증후군 중 하나이며, 입원한 노인 환자의 경우 40%가 우울증 증상이 있다고 보고된다[3]. 이에 대한 노인의 전반적인 건강과 삶의 질에 영향을 미치는 우울증 증상에 대한 적절한 관리가 필요하다[3]. 그러나, 우울증 증상을 가진 환자들의 집중적인 관리를 위해서는 간호사와 환자 간에 1:1 돌봄 서비스를 제공하기에는 인력이 부족하다. 따라서 환자의 우울증 증상을 예방할 수 있으면서 간호사를 보조할 수 있는 AI 아바타 비서를 개발하고자 하며, 가상현실의 콘텐츠를 바탕으로 구성하고자 한다.

증강(AR)·가상(VR)·확장(XR)현실의 콘텐츠는 게임, 교육, 운동뿐만 아니라 심리치료, 재활 등의 치료목적으로도 활용이 된다. 예를 들면, 교통사고로 인한 사고현장의 기억에 따른 정신적 스트레스[4], 가상현실을 활용한 청소년 주의력결핍/과잉행동장애(PTSD)와 품행장애(CD) 심리치료 방법[5] 등 가상현실로 심리치료를 하는 방법에 대해서도 많이 개발되고 있다.

본 연구의 궁극적인 목적은 가상현실의 콘텐츠를 통해 우울증 예방, 의료종사자의 부담을 덜기 위한 1:1 돌봄 서비스를 제공하는 것이다. 이러한 연구목적 달성을 위해, 본 논문에서는 음성을 통해 감정을 인식할 수 있는 ResNet모델을 개발하고 성능을 검증하고자 한다 (그림 1).

2. 제안하는 방법

2.1. 음성 감정 데이터베이스

한국지능정보사회진흥원에서 운영하는 AI통합 플랫폼인 AI-Hub 사이트에서 제공하는 음성 데이터베이스 중 '감정 분류를 위한 대화 음성 데이터셋'을 데이터베이스로 사용한다. 해당 음성 데이터베이스에서 4가지(Happiness, Angry, Sadness, Neutral)의 감정을 추출하여 학습 모델의 데이터베이스로 활용하고자 한다. 각 감정의 음성 데이터 수는 균일하게 적용하기 위해 가장 적은 개수를 기준으로 2,798개로 한다. 총 11,192개의 음성 데이터를 통해 검증을 위한 학습모델을 구현한다.

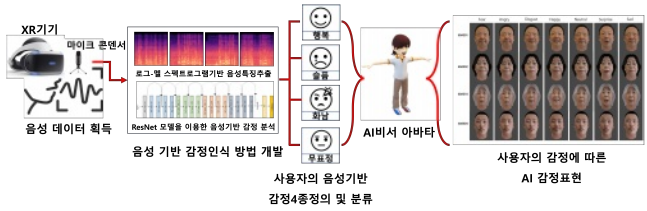


그림 1: 전체 시스템 구성도

---

* 포스터 발표논문

* 본 성과물은 2024년도 대구광역시 지역혁신중심 대학연구활동 지원사업의 지원과 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (P0012724, 2024년 산업혁신인재성장지원사업)

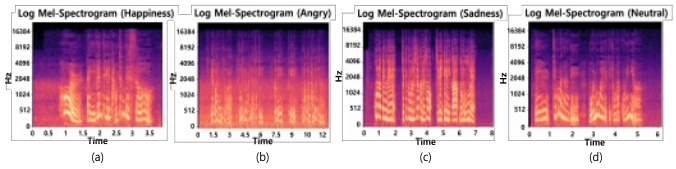


그림 2: 로그 멜-스펙트로그램 (a) 행복 (b) 화남 (c) 슬픔 (d) 무표정

이는 VR기기로부터 피험자의 음성을 녹음하여 전달받았을 때, 해당 음성의 감정 상태와 검증 모델을 통해 판단한 감정 상태가 일치하는지를 확인하기 위해 사용한다.

HCI-VEE(Voice-based Emotion Expression) 데이터베이스는 대구대학교 HCI연구실에서 구축된 자체 데이터베이스로, 5명의 피험자를 대상으로 음성 신호를 수집한 DB이다. 음성을 녹음할 피험자에게 미리 대본과 지표가 될 음성을 전달하고, 감정에 대한 사전 연습을 준비하도록 요청하였다[6]. 녹음 시에는 피험자가 감정 이입을 도와줄 수 있는 보조 지문을 삽입하여 감정을 유도하였다[6]. 본 연구에서는 VR기기 내에 탑재된 마이크를 활용하여 음성 녹음을 진행하였으며, 대사는 드라마, 영화, 일상대화를 통해 각 감정당 10개로 총 40개를 선정하였다. 최종적으로, 피험자 5명을 대상으로 200개(피험자 5명*10개 대사*감정상태 4종)의 음성을 취득하였다(그림 2).

2.2. 음성 특징 추출 방법

수집한 원시 데이터는 BPF(Band-Pass Filter)를 통해 가청주파수인 20~20,000Hz 대역의 주파수로 제한하여 잡음 신호, 저주파 등을 제거한다[7]. 잡음이 제거된 음성 데이터에 STFT(Short-Time Fourier Transform)를 적용하여 시간적 정보를 유지하며, Spectrogram을 추출한다. 그 후, Mel Bank Filter와 log 변환을 적용하여 주파수 대역을 인간 청각에 가깝게 표현한 것을 로그 멜-스펙트로그램이라고 한다. 이를 그림 3과 같이 이미지로 변환하여 사용자의 감정 정보를 4가지 종류(행복, 화남, 슬픔, 무표정)로 나타낸 그림이다.

2.3. ResNet모델 기반 음성 감정 분류

ResNet 모델은 layer가 깊어질수록 기울기 소실 문제가 발생하는 특정 모델들의 문제점을 보완하기 위해 고안된 것으로, Skip Connection을 적용한 모델이다[8]. 해당 신경망은 224*224의 이미지 크기를 입력받기 때문에[9] 수집한 이미지의 크기를 동일한 크기로 변경하였다. 본 논문에서는 Layer 개수가 18개인 ResNet을 구축하여 로그 멜-스펙트로그램 이미지를 행복, 슬픔, 화남, 무감정 각각의 감정 라벨을 적용하여 학습을 진행하였다.

3. 결과

본 논문에서는 ResNet 모델을 이용하여 4가지의 감정으로 분류하는 실험을 진행하였으며, 실험은 모두 5번의 교차 검증 결과를 평균과 편차로 그래프로 나타냈다. 그림 4(a)는 ResNet18 모델을 이용하여 AIHub 데이터

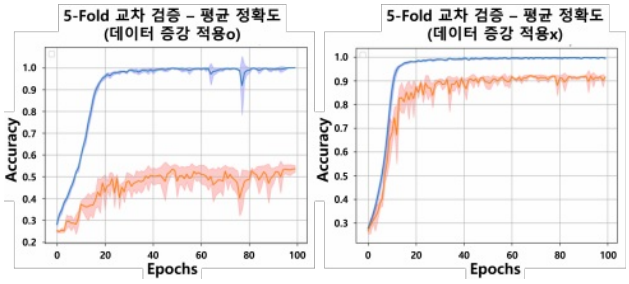


그림 3: k-Fold 교초 검증 결과 (k=5) (a) 데이터 증강 적용한 결과 (b) 데이터 증강을 적용하지 않은 결과

베이스로 학습을 진행하였으며, 평균 정확도는 54.02% (±0.013)로 과적합 문제가 발생하였다. 이를 해결하기 위해 Data Augmentation 기법[10] 중 이미지의 색 saturation(채도)를 변경하여 적용했으며, 22,384개의 이미지를 추가로 획득하였다. 따라서 그림 4(b)와 같이 AIHub 데이터베이스에 색 채도를 적용한 이미지 추가하여 ResNet18 모델의 학습을 진행하였고, 평균 정확도가 91.68% (±0.003)이다. 이는 Data Augmentation 기법을 적용하기 전보다 정확도가 37.66% 증가하였으며, 이를 통해 과적합 문제를 해결하였다. 그러나 학습된 모델로 음성 데이터를 예측한 결과는 내부 데이터(AIHub)가 45%, 외부 데이터(HCI-VEE)가 33%로 전체적으로 낮은 결과를 보였다. 따라서, HCI-VEE 데이터베이스의 음성 데이터를 추가로 수집하여 실험을 확장할 필요성이 있다.

4. 결론

본 논문에서는 음성을 통해 사용자 내면의 감정을 인식하기 위하여 로그 멜-스펙트로그램을 통해 음성 특성을 추출하였으며, ResNet모델을 이용하여 감정을 4종으로 분류하는 방법을 제안하였다. 향후, VR기기 대신 XR 기기에 적용함으로써 실제 사용자가 착용하였을 때 음성에 내재된 감정에 따라 AI 아바타 비서의 표정을 변화시켜 감정적 공감을 받을 수 있도록 개발할 예정이다. 또한, 의료기기와 연동하여 환자의 건강 상태를 실시간 모니터링하여 간호사의 보조 역할을 할 수 있는 기능을 추가할 예정이다.

참고문헌

[1] OECD Health Statistics, *Ministry of health and welfare*, 2022.
[2] Population Projection 2020, *Statistics Korea*, 2022.
[3] E.H. Cho, J.H. Chun, S.W. Lee, A Model for the Factors Affecting Depression in the Elderly with Chronic Disease, *Journal of the Korean Geriatrics Society*, 4(3):148-163, 2000.
[4] J.W. Park, S.H. Oh, A Study on the Development of VR-Based Life Care Contents for Post-Traumatic Stress Disorder(PTSD) for Traffic Accident. *The Journal of Korean Institute of Next Generation Computing*, 14(3):56-65, 2018.

[5] C.H. Ryu, Implications of VR-Based Psychotherapeutic Effects for ADHD and CD among Adolescents. *Korean Association of Addiction Crime Review*, 12(1):25-55, 2022.

[6] I.C. Jang, M.K. Park, T.S. Kim, and M.W. Park, Study of Emotion in Speech, *Proceedings of th Korean Society for Precision Engineering*, pp.25-28, 2004.

[7] C.H. Lee, S.K. Hong, A study on the hearing characteristic based equalizer design for the elderly, *Journal of Digital Contents Society*, 19(4):779-787, 2018.

[8] S.H. Kim, K.J. Lee, Developing a ResNet-Based Forest Fire Detection System Using Training Data from Unmanned Aerial Vehicle, *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, pp.384-385, 2023.

[9] A.H. Jo, K.C. Kwak, Emotion Recognition Using Deep Learning Based on Transfer Learning from Speech Emotion Signals, *Proceedings of the Korean Institute of Electrical Engineers (KIEE) Annual Conference*, pp.196-198, 2021.

[10] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 25, 2012.

사용자 악력 근활성이 반영된 VR인터페이스의 개선*

신영훈[0,1], 이민정[1], 이미란[†,2]
대구대학교 일반대학원 IT융합공학과[1], 대구대학교 컴퓨터정보공학부[†,2]
{younghoon.hci.du, minjeong.hci.du}@gmail.com, miran@daegu.ac.kr([†]교신저자)

Improvement of VR Interface that Reflects User's Grip Strength

Younghoon Shin[0,1], Minjeong Lee[1], and Miran Lee[†,2]
Graduate School of IT Convergence Engineering, Daegu University[1]
Department of Computer and Information Engineering, Daegu University[2]

요약

가상현실에서 몰입감을 극대화 시킬 수 있는 요인은 직관적이고 감각적인 인터랙션 방법을 개발하는 것이다. 기존의 가상현실 컨트롤러는 사용자의 동작의도 제어를 위해 사용되고 있지만, 손의 악력이나 세밀한 근력 등을 반영하지 못한다는 문제점이 있다. 따라서, 본 논문에서는 사용자의 악력, 즉, 심지굴근의 수축을 가상현실 콘텐츠에 반영하여, 기존의 방법보다 감각적인 가상현실 환경을 구축하고자 한다. 실험에서는 6명의 피험자로부터 표면근전도를 기반으로 심지굴근의 근활성도를 취득하고, 가상현실 내에서 강도(intensity)가 다른 4가지의 객체를 생성하였으며, 근 활성도에 따라 객체를 파괴할 수 있도록 제작하였다. 결과적으로, 보다 감각적으로 상호작용할 수 있다는 결과를 얻었다.

1. 서론

가상현실(Virtual Reality, VR)은 교육, 의료, 예술, 비즈니스 등 다양한 분야와 접목되어 새로운 인터페이스로 자리 잡고 있으며, 가상현실 컨텐츠의 이용자 수는 늘어나는 추세이다[1]. 특히 재활, 학습, 모의 훈련에 이용될 때, VR은 특유의 몰입감으로 업무의 적응이나 능률 향상을 통해 효과적인 훈련 수단으로 기대되는데[2-5], 이는 VR의 특성인 실재감(presence)이 몰입(flow)을 유발하여 학습효과에 영향을 미치기 때문이다[3]. 즉, VR 환경에서 몰입감을 높이는 것은 학습효과를 높이는 것으로 이어진다.
한편, 감각적인 몰입감을 향상시키기 위해 생체신호를 VR 환경에 활용하고자 하는 연구도 진행되었으며[6], SEMG(Surface Electromyography, SEMG)는 인간의 모션 의도를 인식할 수 있는 신뢰할 수 있는 방법 중에 하나이다[7]. 이는 사용자가 능동적으로 동작 의도를 제어할 수 있고, 별다른 외과적 수술 없이 바로 피부에 부착할 수 있기 때문이다.
이에 본 연구는 VR 콘텐츠 내에 오브젝트와 상호작용하는데 있어서 SEMG를 활용했을 때에 몰입감에 대해 서술하였다. 핸드트래킹을 기반으로 무구속적인 환경을 구축하였다. 그림 1 과 같이 VR 객체를 손으로 쥐는 동안, 측정되는 EMG 값을 기반으로 객체가 반응하도록 하며, 이를 시각적으로 확인할 수 있도록 Unity엔진을 사용하여 시스템을 개발하였고, 설문조사를 통해 제안하는 방법의 효과성을 검증하였다.

2. 관련 연구

VR은 실재감, 몰입, 상호작용성이 있어야 한다고 정의된다[9]. 권종산 외 [3]는 VR의 presence가 몰입을 유발하며, 몰입이 학습효과의 향상과 직접적인 관련이 있

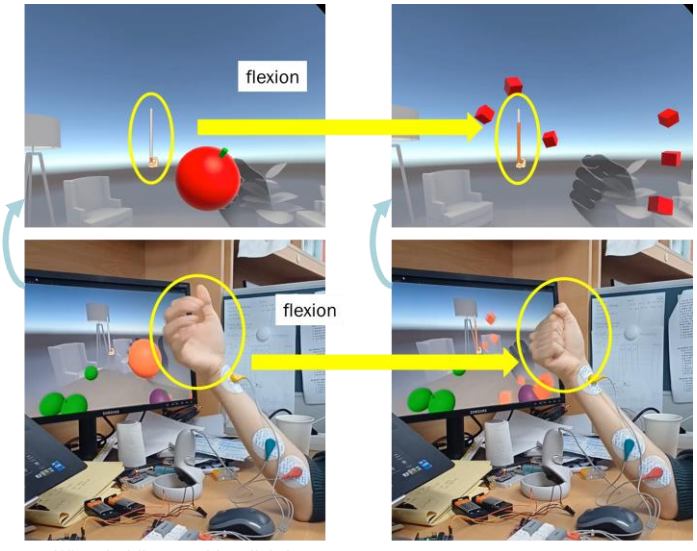


그림 1: 근육 수축 시 객체의 이벤트가 발생하는 예시

* 포스터 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 Journal of Korea Multimedia Society 27(3). 2024에 게재되었음.
* 본 성과물은 2024년도 대구광역시 지역혁신중심 대학연구활동 지원사업의 지원과 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (P0012724, 2024년 산업혁신인재성장지원사업).

어 상호작용의 향상이 학습 효과의 향상을 기대할 수 있으리라 보았으며, HMD 기반 몰입형 학습에서 방해요인과 몰입, 학습효과의 관계를 분석하였다. 이러한 관련 연구는 공통적으로 VR에서의 몰입을 중요시하여 감각 정보를 이용한 상호작용의 중요성을 강조하고 있다.

EMG는 사용자가 주체적으로 동작의도에 대해 제어가 가능하고, 근활동과 해석까지의 시간적 간격이 짧아 강력한 도구로 사용된다. Anany Dwivedi 외 [8]의 연구는, VR/AR 시스템에서 휴대용 컨트롤러를 사용하는 것은 손을 쓰는 것에 비해 덜 직관적이기 때문에, 데이터 글로브와 EMG를 활용한 손 제스처 추적을 제시하여, 근육-컴퓨터 인터페이스가 몰입적인 상호 작용을 촉진할 수 있음을 시사한다. 김상수 외 [9]은, 다채널 EMG 모듈을 이용하여 4 가지 방향의 인터페이스를 개발하였다.

상기 연구들은 인터페이스를 개발할 때 EMG의 활용이 현실적이고 유용함을 내포한다. 다만, [8]은 여전히 손이 하드웨어와의 접촉에서 자유롭지 못한 한계점이 있다. 또한, [9]에서 제시된 시스템은 채널 당 명령 가짓수가 제한된다는 한계점이 있다. 이러한 문제점들을 개선하기 위해, 본 연구에서는 HMD 내장 카메라를 사용하여 맨손의 위치와 방향, 손가락의 관절 각도를 감지하는 핸드트래킹을 적용하여 불편함을 완화한 시스템을 제안한다.

3. 제안한 방법

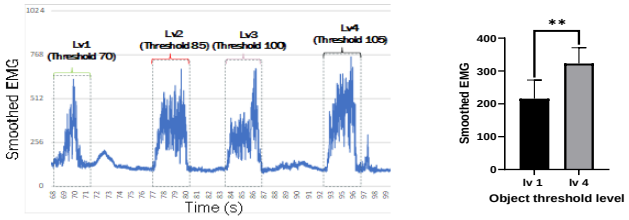
3.1. SEMG 신호처리

FDP에 Ag-AgCl 전극 및 스냅전극을 부착하여 EMG 모듈 KIT0012 을 통해 아두이노 UNO에 연결하였다. EMG 신호는 500Hz의 주요 주파수 대역을 가진다고 알려져 있으며[10], 나이퀴스트 이론에 따라 1kHz 이상의 주기로 샘플링함으로써 신호의 왜곡(aliasing)을 방지하였다.

실험의 조정 단계에서 참가자에게 최대한의 노력으로 FDP를 수축하도록 요청하여 최대자발적수축 MVC(Maximum Voluntary Contraction) 값을 구하고, MVC 값과 실시간으로 입력받는 smoothed EMG 값을 기반으로 Min-Max 정규화를 통해, Normalized EMG 값을 산출하였다[11]. 데이터 전처리 후, 실시간 입력의 Normalized EMG 값을 토대로 객체들은 악력에 따른 이벤트를 발생할 때 활용된다.

3.2. 시뮬레이션 환경 구축

객체의 Intensity를 다르게 설정하기 위해 Normalized EMG 기반으로 임계치가 각기 다르게 설정된 4 개의 객체를 생성하고, 강도에 따라 초록, 빨강, 보라, 회색 색상으로 디자인하였다. 파괴 가능한 객체들은 Grabbed 상태이면서 Normalized EMG가 객체별 임계값을 넘을 때 파편 파티클을 발생시키고 제거된다. 본 연구를 위해 제안된 시스템의 구동 영상은 다음 링크를 통해 제공한다(https://www.youtube.com/watch?v=Xp1-zRuINwg).



(a) (b)

그림 2: 실험참가자가 각 객체를 파괴할 때 기록한 EMG 값의 결과. (a) 각 단계의 객체를 부술 때의 EMG 그래프 (b) 1단계와 4단계 객체 간 t-test 결과($p<0.001$).

4. 결과 및 토의

4.1. 실험 환경

제안하는 악력반영 시스템의 만족도 조사 및 시스템 검증을 위해 남성 3 명과 여성 3 명으로 이루어진 6 명의 대상자(22.17 세, ±2.13)를 모집하였다. 피험자들은 대구대학교 소속의 학생으로, FDP에 어떠한 근골격계 질환을 가지지 않는 건강한 사람을 대상으로 진행하였다. 또한, 해당 실험에 대해 사전에 충분히 전달하고 실험 참가 동의를 받았다.

4.2. 객체 강도별 EMG의 결과

Raw EMG 및 Smoothing 처리된 신호를 포함한 정량적인 데이터를 확보하였으며, 그림 2 (a)는 실험대상자 한 명의 각 객체를 부술 때의 Onset과 Cessation 사이의 Smoothed EMG 값을 기록하여 그래프로 나타낸 것이다.

그림 2 (b)는 실험대상자들이 각 객체를 부수려고 시도하는 동안의 Smoothed EMG 값의 평균을 t-test 결과로 나타낸 것이며, 두 레벨 간 유의미한 차이를 확인할 수 있었다 ($p<0.001$).

5. 결론

본 논문에서는 감각적인 VR 시스템을 위해 FDP으로부터 EMG와 핸드트래킹을 활용하였으며, MVC를 토대로 한 EMG신호 Normalization을 통해 VR에 악력을 반영시킴으로써 사용자가 VR기기와 콘텐츠를 사용할 때 인터랙션 측면에서의 만족감을 향상시키는 방법을 제안하였다. 제안하는 방법의 평가를 위해 실험대상자를 모집하여 정량적 평가를 진행하였다. 그 결과, 사용자들이 상호작용하는 객체의 강도가 다를 때 근활성도의 유의미한 차이를 확인할 수 있었다. 이를 통해 EMG를 기반으로 악력을 설정하고 가상 객체가 서로 다른 악력 임계치를 가질 때, 사용자로부터 가상 객체와 상호작용하기 위해 능동적이고 구분되는 EMG 신호를 취득할 수 있음을 확인하였다.

참고문헌

[1] T. Kim and S.Kim, “Metaverse App Market and Leisure: Analysis on Oculus Apps,” *Knowledge Management Research*, Vol. 23, No. 2, pp. 37-60, 2022.

[2] M. Lee, J. Youn, and E. Kim, “Evaluation and Analysis of VR Content Dementia Prevention Training Based on Musculoskeletal Motion Tracking,” *Journal of Korea Multimedia Society*, Vol. 23, No. 1, pp. 15-23, 2020.

[3] C.S. Kwon, “A Study on the Relationship of Distraction Factors, Presence, Flow, and Learning Effects in HMD - Based Immersed VR Learning,” *Journal of Korea Multimedia Society*, Vol. 21, No. 8, pp. 1002-1020, 2018.

[4] J. Shin, D. Choi, S. Kim, and K. Jin, “A Study on Efficiency of the Experience Oriented Self-directed Learning in the VR Vocational Training Contents,” *Journal of Knowledge Information Technology and Systems*, Vol. 14, No. 1, pp. 71-80, 2019.

[5] J. Jo, “Analysis of Visual Attention of Students with Developmental Disabilities in Virtual Reality Based Training Contents,” *Journal of Korea Multimedia Society*, Vol. 24, No. 2, pp. 328-225, 2021.

[6] C. Kim, J. Youn, I. Kang, and B. Kim, “Development and Assessment of Multi-Sensory Effector System to Improve the Realistic of Virtual Underwater Simulation,” *Journal of Korea Multimedia Society*, Vol. 17, No. 1, pp. 104-112, 2014.

[7] K. Li, J. Zhang. L. Wang, M. Zhang, J. Li, and S. Bao, “A Review of the Key Technologies for sEMG-Based Human-Robot Interaction Systems,” *Biomedical Signal Processing and Control*, Vol. 62, September 2020, pp. 102074, 2020.

[8] A. Dwivedi, Y.J. Kwon, and M. Lioarokapis, “EMG-Based Decoding of Manipulation Motions in Virtual Reality: Towards Immersive Interfaces,” *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3296-3303, 2020.

[9] K. Kim, Y. Han, W. Jung, Y. Lee, J. Kang, H. Choi, et al., “Technical Development of Interactive Game Interface Using Multi-Channel EMG Signal,” *Journal of Korea Game Society*, Vol. 10, No. 5, pp. 65-73, 2010.

[10] M. Lee, J. Lee, and D. Kim, “Gender Recognition Using Optimal Gait Feature Based on Recursive Feature Elimination in Normal Walking,” *Expert Systems with Applications*, Vol 189, March 2022, pp. 116040, 2022.

[11] M.J. Islam, S. Ahmad, F. Haque, M.B.I. Reaz, M.A.S. Bhuiyan, and M.R. Islam, “Application of Min-Max Normalization on Subject- Invariant EMG Pattern Recognition,” *IEEE Transactions on Instrumentation and Measurement*, Vol. 71, pp. 1-12, 2022.

깊이 감독 Dynamic NeRF: 제한된 카메라를 이용한 동적 장면 재구성

김규민[0], 박동하[0], 이채은[1]
인하대학교 정보통신공학과[0], 한양대학교 융합전자공학부[1]
{kkm8121, donghaha12}@naver.com, chree@hanyang.ac.kr

Depth Supervised Dynamic NeRF: Dynamic Scene Reconstruction with Limited Cameras

Kyu Min Kim[0], Dong Ha Park[0], Chae Eun Rhee[1]
Inha University Information and Communication Engineering[0], Hanyang University Electronic Engineering[1]

요약

본 논문은 고품질 3D 장면을 합성하는 Neural radiance fields (NeRF)[1] 기반 연구 중 다중 시점 비디오 입력으로 동적 장면을 효율적으로 학습하는 MixVoxels[2]에서 학습에 너무 많은 카메라가 사용되는 문제를 해결하는 방법을 제시한다. 본 논문에서는 먼저 한 대의 카메라를 통해 얻은 데이터를 이용하여 정적 모델을 학습한다. 그리고 정적 모델에서 얻은 깊이 정보를 증류하여 Fine모델을 학습한다. 최종적으로 얻은 모델은 학습에 적은 수의 카메라를 사용했음에도 정적인 영역과 동적인 영역이 높은 품질로 학습된 모델이다.

1. 서론

다중 시점 비디오 입력으로 동적 장면을 재구성하는 것은 큰 주목을 받는 주제이다. 이는 virtual reality/augmented reality (VR/AR), 게임 및 영화 산업, 3-dimensional (3D) 콘텐츠 등 많은 응용 분야에 사용된다. MixVoxels[2]는 다중 시점 비디오 합성에 심층 신경망을 이용하여 새로운 시점의 장면을 높은 성능으로 합성하는 NeRF[1]를 접목하여 동적 장면을 매우 사실적으로 재구성하고 있다. MixVoxels[2]은 높은 품질로 새로운 시점의 장면을 재구성하지만 학습에 매우 많은 카메라가 사용되고 그로 인해 많은 이미지가 학습에 입력으로 주어진다는 단점이 있다. 또한 학습에 사용되는 카메라의 개수를 줄였을 경우에 매우 큰 성능 하락이 발생한다. 본 연구는 학습에 사용되는 카메라의 개수를 줄이더라도 동등하거나 높은 품질의 장면 재구성을 달성하는 것을 목표로 한다. 이를 위해서 깊이 정보를 얻을 수 있는 정적 모델 학습과, 깊이 정보를 이용하여 적은 숫자의 카메라로도 사실적인 동적 장면을 재구성하는 Fine 모델 학습으로 나누어 학습 파이프라인을 구성한다. 또한 정적 모델로 얻은 깊이 정보를 이용하여 3차원에 존재하는 실제 물체의 위치를 파악하고 이 위치의 주변에 더 많은 샘플링을 진행하는 깊이 감독 샘플링을 제안한다.

2. 깊이 감독 Dynamic NeRF

그림 1은 본 논문의 학습 파이프라인을 의미한다. 우선 정적인 영역의 깊이 정보를 추출할 수 있는 정적 모델을 학습한다. 정적 모델로부터 장면의 정적인 영역의 깊이 정보를 추출할 수 있다. 그리고 정적 모델로부터 얻은 깊이 맵(depth map)을 추가의 입력으로 사용하여 Fine model을 학습한다. NeRF[1]에서는 ground truth(GT) RGB와 모델이 추론한 RGB 사이의 평균제곱오차를 계산하여 손실 함수를 정의한다. 본 논문에서는 NeRF의 RGB 손실함수에 깊이 손실 항을 추가한다. 깊이 손실 항은 Fine 모델이 추론한 장면의 깊이 맵과 정적 모델로 얻은 깊이 맵 사이의 평균제곱오차를 계산하여 얻은 값이다.

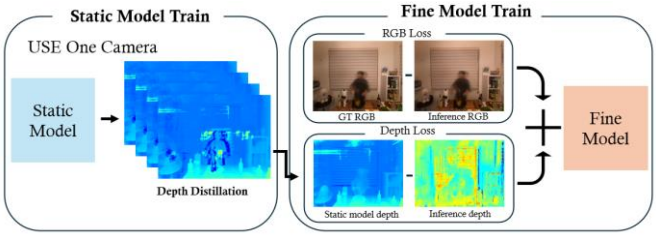


그림 1. 깊이 감독 Dynamic NeRF 전체 파이프라인

2.1. 정적 모델 학습

한 대의 카메라를 통해 3차원 장면의 정적인 영역을 학습하여 깊이 정보를 추출한다. 학습하는 장면이 정적이라고 가정한다면, 카메라 한 대를 사용하여 시점을 변화시켜가며 얻은 RGB 이미지를 통해 정적 모델을 학습할 수 있다. 3D 장면의 대부분은 정적인 영역이기 때문에 정적 모델을 통해 얻은 깊이 정보는 신뢰할 수 있는 정보이다.

---

* [1]교신저자
* 포스터 발표논문, 학부생 주저자 논문
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT연구센터육성지원사업의 연구결과로 수행되었음. (IITP-2022-RS-2022-00156225)

2.3. 깊이 감독 샘플링

그림 2은 MixVoxels[2]와 본 논문의 샘플링 방식 차이를 설명한다. MixVoxels[2]은 그림2(a)에서처럼 광선을 따라 3차원의 점의 위치를 균일하게 샘플링한다. 따라서 카메라 사용이 적은 상황에서 학습에 사용되는 정보가 부족한 위치와 충분한 위치의 차이를 고려하지 못한다. 따라서 재구성된 장면에서 두 부분의 합성 품질은 큰 차이가 발생하게 된다. 이를 해결하기 위해 본 논문에서는 그림2(b)의 깊이 감독 샘플링을 제안한다. 정적 모델에서 얻은 깊이 정보를 이용하여 3차원 상의 물체의 위치를 추정할 수 있고 이 위치가 샘플링 확률 분포의 평균 지점이 되도도록 가우시안(Gaussian) 샘플링 한다. 이를 통해 학습에 사용되는 정보가 부족한 부분은 상대적으로 많이 샘플링하고 정보가 충분한 부분은 적게 샘플링 한다. 이는 재구성된 장면이 균일한 합성 품질을 달성할 수 있게 한다.

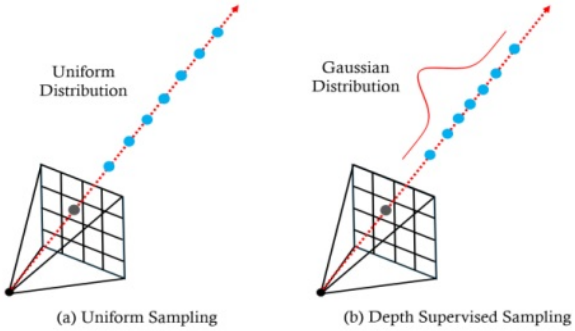


그림 2. 깊이 감독 샘플링

3. 실험 결과

본 논문은 20개의 서로 다른 시점의 비디오 데이터셋인 Plenoptic Video Dataset을 사용하여 성능을 검증하였다. 표 1은 MixVoxels[2]에서 학습에 사용되는 카메라 수에 따른 재구성 품질 차이를 나타낸 표이다. 카메라 수가 감소함에 따라 재구성되는 장면에서 큰 성능 하락이 발생했다. 표 2는 본 논문 성능 검증을 위하여 카메라 2대를 사용하여 진행한 학습을 기준으로 본 논문과 MixVoxels[2]의 재구성 품질 차이를 나타낸 표이다. 본 논문의 방법에서 동일한 카메라 개수로 얻은 이미지에 깊이 맵 입력 추가로 정적인 영역 PSNR 6.19db, 평균 PSNR 4.05db의 성능 향상을 달성했다. 그림 3은 표1, 표2의 학습 결과이다. 그림3(a)는 카메라로 취득한 이미지, 그림3(b)는 MixVoxels[2]에서 카메라 20대로 학습한 결과, 그림3(c)는 MixVoxels[2]에서 카메라 2대로 학습한 결과, 그림3(d)는 본 논문의 방법으로 카메라 2대로 학습한 결과이다. 그림3(b)와 그림3(c)를 비교하면 학습에 사용된 카메라의 개수가 줄어듦에 따라 큰 성능 하락이 발생했다. 그림3(c)와 그림3(d)를 비교하면 본 논문의 방법을 통해 학습에 사용되는 정보의 부족을 깊이 정보로 보완하여 재구성 품질을 향상시켰다.

카메라 수	이미지 수	PSNR(dB)	PSNR_STA(dB)
20	6000	32.55	32.76
5	1500	23.73	24.45
2	600	20.96	21.76

표 1. MixVoxels[2] 학습에 사용되는 카메라 수에 따른 성능 차이, PSNR_STA는 정적인 영역의 PSNR

모델	이미지 수	PSNR(dB)	PSNR_STA(dB)
MixVoxels	600	20.96	21.76
Ours	620	25.01	27.95

표 2. 카메라 두 대를 사용한 학습을 기준으로 성능 비교

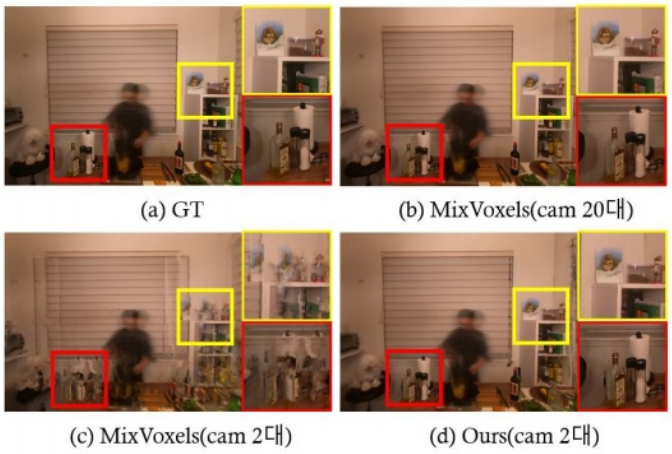


그림 3. 새로운 시점의 장면 추론 결과

4. 결론

본 논문은 깊이 감독을 사용하여 동적 장면을 효율적으로 학습하는 방법을 제안하며, 정적 모델과 Fine 모델을 분리하여 학습하는 방식을 통해 적은 수의 카메라로도 높은 품질의 동적 장면 재구성을 가능하게 한다. 본 논문의 방법으로 학습된 모델은 정적인 영역에서 상당한 성능 향상을 달성했지만 동적인 영역은 정적인 영역에 비해 상대적으로 적은 성능 향상을 달성했다. 향후 연구에서는 정적인 영역과 동적인 영역이 균일한 품질로 학습을 진행하는 방법을 연구해보고자 한다.

참고문헌

[1] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” In European Conference on Computer Vision (ECCV), pp. 405-421, 2020.

[2] Feng Wang, Sinan Tan, Xinghang Li, Zeyue Tian, Yafei Song, and Huaping Lui, “Mixed Neural Voxels for Fast Multi-view Video Synthesis,” In IEEE International Conference on Computer Vision (ICCV), pp. 19706-19716, 2023.

[3] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramana, “Depth-supervised NeRF: Fewer Views and Faster Training for Free,” In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12882-12891, 2022.

이벤트 카메라를 이용한 동적 장면에 대한 편광 반사 특성 계측*

前田涼汰[0], 백승환
POSTECH 컴퓨터공학과
{maeda, shwbaek}@postech.ac.kr

Event-based Polarimetric Reflectance Imaging for Dynamic Scene

Ryota Maeda[0], Seung-Hwan Baek
Dept. of Computer Science and Engineering, POSTECH

Abstract

Polarization delivers crucial physical properties for understanding real-world environments such as shape, material, and light transport. Polarimetric reflectance imaging aims to acquire comprehensive polarimetric info that ties both a light source and a camera. However, this imaging requires multiple shots by rotating the polarizers, thus requiring a long acquisition time. In this paper, we introduce event-based polarimetric reflectance imaging, which allows for fast acquisition and capability for dynamic scenes. Our proposed imaging system comprises an event camera and continuously rotating quarter-wave plates. Continuous rotation enables fast polarization modulation, and the event camera can detect these changes as events. We introduce an image formulation and Mueller matrix reconstruction method that is robust to outlier events caused by sensor noise and object movements.

1. Introduction

Polarimetric reflectance imaging aims to acquire the comprehensive polarimetric relationship between the light source and the camera [1]. This rich information enables detailed and reliable polarization-based image analysis such as 3D shape estimation, material actuation, light transport analysis. Unfortunately, this imaging technique requires capture multiple images by mechanically rotating the polarizers, leading to a long acquisition time.

* 본 논문은 요약논문 (Extended Abstract)으로서 현재 수행중인 연구임.
* 본 연구는 한국연구재단 MSIT의 연구비 지원 (RS-2023-00211658)으로 수행되었음.

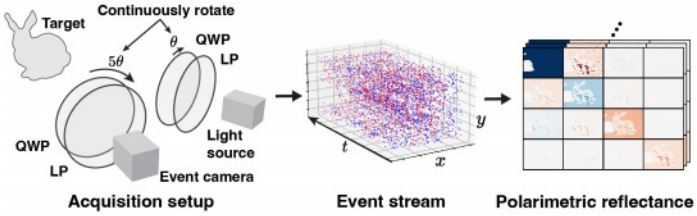


Figure 1 Overview of our event-based polarimetric reflectance imaging.

Event cameras asynchronously measure per-pixel brightness changes with high temporal resolution and have been widely used for capturing dynamic movements. In contrast to movement analysis, recent works utilized high-temporal resolution of event cameras for photometric analysis by modulating light intensity. The most relevant work is shown by Muglikar et al. [2]. They measured the polarization state by rotating the linear polarizer in front of the event camera. Our work expands this concept to polarimetric reflectance imaging by rotating the two quarter-wave plates (QWPs) in front of the light source and the event camera.

In this paper, we introduce event-based polarimetric reflectance imaging for dynamic scenes. As shown in the overview in Figure1, our method captures the polarimetric reflectance by rotating two QWPs and reconstructing the Mueller matrix from event streams. This reconstruction method is strong against outlier events caused by unexpected noise and movements of dynamic scenes.

2. Image Formation

2.1. Polarization Measurement

In our approach, we modulate the brightness of light by rotating two quarter-wave plates (QWPs). This

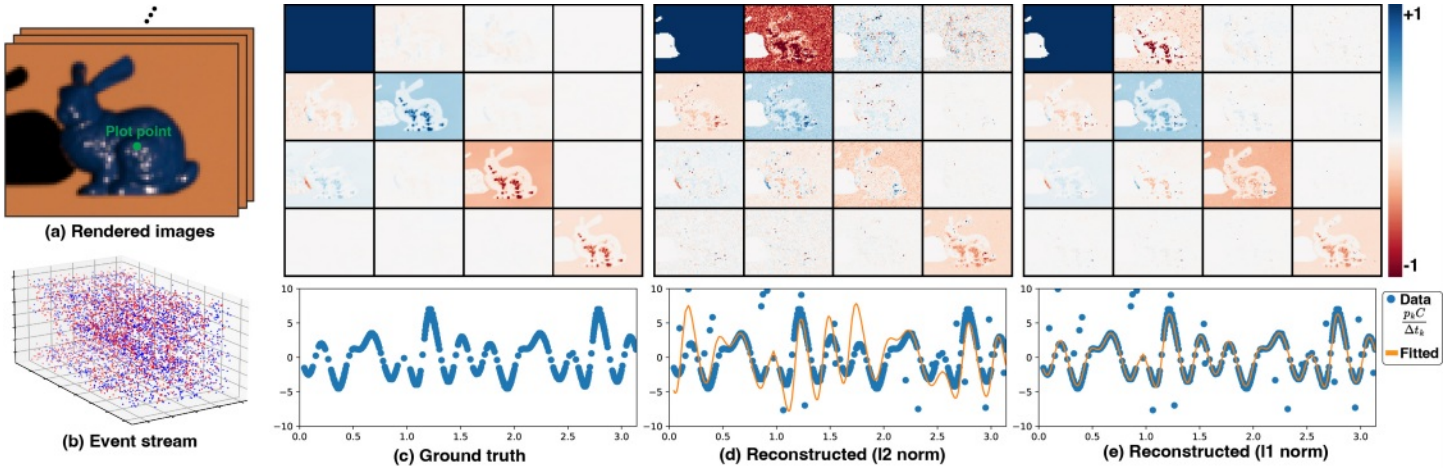


Figure 2 Reconstructed Mueller matrix image. (a) The sequence of rendered images. (b) Generated event str eam from rendered images. (c)-(e) Mueller matrix and plot of generated event and fitted line. (c) Ground truth. (d) Reconstructed result with l2 norm. (e) Reconstructed result with l1 norm using IRLS algorithm. Note that all the Mueller matrix images are normalized by m00 components.

technique is known as the dual-rotating-retarder method, and the modulated intensity I is formulated as follows:

$$I(\theta) = [L(0)Q(5\theta)M_{\text{meas}}Q(\theta)L(0)]_{00},$$

where θ is the angle of QWP. Here, $\boldsymbol{L}$ and $\boldsymbol{Q}$ denote the Mueller matrix of a linear polarizer (LP) and QWP, respectively, and $\boldsymbol{M}_{\text{meas}}$ is the Mueller matrix of the measurement target. As indicated in the equation, the QWP on the camera side rotated five times faster than the one on the light source side. This ratio is well conditioned for Mueller matrix reconstruction, and we incorporate this ratio in our method.

2.2. Polarization Measurement with Event Sensor

We now formulate the observation of modulated light in an event camera. Inside the event sensor, the radiance of incident light is converted to log photocurrent $log\,(I)$. When the change of log photocurrent reaches a threshold C, an event $e = (x, t, p)$ is recorded. This event is consisted with pixel location $x = (x, y)$, timestamp t, and polarity $p \in \{+1, -1\}$. When the time difference $\Delta t_k = t_k - t_{k-1}$ (where k is an index of the event at pixel x) of adjacent two events is small enough, the derivative of log photocurrent can be approximated according to Taylor's expansion as follows:

$$\frac{\partial \log(I)}{\partial t}(t_k) \approx \frac{p_k C}{\Delta t_k}.$$

2.3. Mueller Matrix Reconstruction

We formulate an objective function to reconstruct the Mueller matrix from the event stream as follows:

$$\text{minimize} \sum_{k=1}^{K} \rho\left(\frac{\partial \log(I)}{\partial t}(t_k) - \frac{p_k C}{\Delta t_k}\right),$$

where ρ is the loss function. We employ L1 absolute loss to be robust for outlier events. To solve the minimization equation, we convert the above equation to linear matrix form and use singular value decomposition (SVD) to solve it. To solve the equation with arbitrary loss function ρ, we use an iteratively reweighted least squares (IRLS) algorithm.

3. Result

We evaluate our method using synthetic data. We rendered the image sequence for each interval at $\theta = 0.01°$ using Mitsuba3 and pBRDF dataset [1]. The rendered images are converted to an event stream using the DVS-Voltmeter. To evaluate the robustness of outlier noise, we randomly replaced 5% of events with the randomly generated values. Figure2 shows the generated synthetic data and reconstructed Mueller matrix from events. Our method robustly reconstructs the Mueller matrix under the outlier events.

References

[1] Seung-Hwan Baek, Tizian Zeltner, Hyunjin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H Kim. Imagebased acquisition and modeling of polarimetric reflectance. ACM Trans. Graph., 2020.
[2] Manasi Muglikar, Leonard Bauersfeld, Diederik Paul Moeys, and Davide Scaramuzza. Event-based shape from polarization. Proc. of CVPR, 2023.

개인별 속도차를 반영한 물리 기반 군중 탈출 시뮬레이션*

김가을[0], 나경민, 이성호, 김나영, 최명걸
가톨릭대학교 미디어기술콘텐츠학과
{rkdmf11103, nkm3266, dltjdgh115, nykim0508, mgchoi}@catholic.ac.kr

Physics-Based Crowd Escape Simulation Reflecting Individual Speed Differences

Gaeul Kim[0], Kyungmin Na, Seongho Lee, Nayeong Kim, Myung Geol Choi
Dept. of Media Technology & Media Contents, The Catholic University of Korea

요약

인파 사고 분석 및 예측을 위해 정확한 시뮬레이션이 중요하다. 일반적인 군중 시뮬레이션에서는 군중 전체 움직임을 재현하는 데 계산 자원을 집중하며, 각각의 에이전트는 최대한 단순하게 모델링된다. 많은 경우, 에이전트는 원으로 표현되고 속도와 방향만으로 제어된다. 이러한 단순화는 인체의 신체 구조의 복잡성을 무시하므로, 인체와 인체 사이 또는 인체와 환경 사이가 매우 밀접한 상황에서는 정확한 시뮬레이션을 제공하지 못한다. 본 연구에서는 인파 밀도가 높고 탈출 경로가 매우 좁은 상황에서의 탈출 시뮬레이션에 물리 기반 캐릭터 애니메이션을 적용한 사실적인 탈출 시뮬레이션을 제시한다. 또한, 각 에이전트이 신체 조건에 따라 최대 이동 속도가 다를 수 있음을 고려하고, 이러한 요소가 탈출 시뮬레이션에 미치는 영향을 분석하고자 한다.

1. 서론

인파 사고는 많은 인명 피해를 초래할 수 있으며, 이를 방지하기 위해서는 정확한 시뮬레이션을 통한 예측과 분석이 필요하다. 기존의 군중 시뮬레이션 연구는 주로 군중의 전체적인 움직임을 모델링하는 데 중점을 두었으며, 각각의 에이전트는 단순한 형태로 표현되어 왔다. 기존의 군중 시뮬레이션은 주로 2D 평면상에서 단순한 원 형태의 에이전트를 사용하여, 속도와 방향만을 고려한 모델링이 일반적이었다. 이러한 접근법은 계산 효율성 면에서는 유리하나, 실제 인체의 복잡한 상호작용을 재현하기에는 부족하다. 특히 고밀도 인파 상황에서의 주변 사물과 상호작용 또는 사람과 사람 간의 상호작용을 정확히 재현하지 못한다는 한계를 가진다. 최근 연구에서는 물리 기반 캐릭터 애니메이션을 적용하여, 에이전트 간의 물리적 상호작용을 보다 정밀하게 모델링하는 시도가 이루어졌다. 특히, 돌발적인 위험 상황에서의 대피 시뮬레이션에 있어, 주변 사물과의 충돌 및 상호작용을 고려한 연구가 제안되었다 [1]. 다만 이 연구에서는 비교적 대피 공간이 넓고 공간 내의 사람 밀도는 낮은 상황에 대해서만 실험을 진행하여 사람과 사람 사이의 충돌이나 길 막힘의 경우는 고려되지 못하였다.

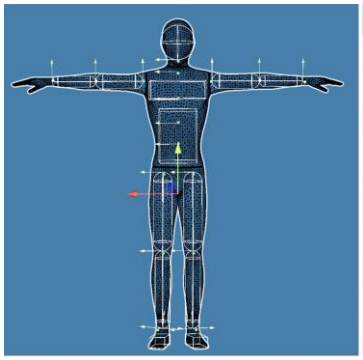
본 연구에서는 인파 밀도가 높고 통로가 좁은 경우 탈출 시뮬레이션을 위해 물리 기반 캐릭터 애니메이션을 적용한다. 각 에이전트가 신체 조건에 따라 최대 이동 속도가 다를 수 있다고 가정하여 실제 상황에서의 복잡한 상호작용을 보다 정확하게 재현하고자 하였다. 구체적으로 매우 밀집된 인파가 좁은 통로를 지나가는 상황을 설정하고, 물리 기반으로 제어되는 각 인체 모델이 서로 다른 최대 이동 속도로 움직이는 상황을 시뮬레이션 하였다. 최대 속도의차이와 비율에 따라 탈출 속도의 변화를 관찰하였으며, 실험 결과 의미 있는 차이가 있음을 발견하였고, 모두가 최대 속도로 탈출하는 패닉 상황에서는 느린 사람이 빠른 사람이 먼저 탈출하도록 기다리는 것이 전체 탈출 속도에 도움이 됨을 확인하였다.

2. 물리 기반 인체 모델 및 제어

물리 기반 인체 모델은 그림 1과 같이 18개 신체 부위로 구성된 관절체로 만들었다. 오른쪽 표는 각 신체 부위 이름과 무게를 나타낸다. 관절의 회전값 제어를 위해 PD 제어기를 사용하였다. 자연스러운 걷는 동작 제어를 위해 한 번의 사이클에 해당하는 앞으로 걸어가는 모션 캡쳐 데이터를 레퍼런스 데이터로 사용하였다. 걷는 속도에 변화를 주기 위해 레퍼런스 모션 데이터의 프레임 레이트를 조정하였다. 걷는 방향을 바꾸어야 하는 경우 루트 노드의 방향에 회전을 적용하여 조정하였다.

인체가 걸어가면서 넘어지지 않고 균형을 유지하려면 많은 노력이 필요하다 예를 들어, [1]의 연구에서는 PD

* 포스터 발표논문
* 학부생 주저자 논문
* 본 연구는 경찰청 과학기술기반 군중밀집관리 기술 개발 연구 사업(RS-2024-004051 00)의 지원으로 수행되었음



Body Part	Weight (kg)	Body Part	Weight (kg)
L. UpLeg	6.56	R. UpLeg	6.56
L. Leg	6.56	R. Leg	6.56
L. Foot	1.00	R. Foot	1.00
L. Arm	4.37	R. Arm	4.37
L. ForeArm	4.37	R. ForeArm	4.37
L. Hand	1.00	R. Hand	1.00
Spine	4.31	Neck	1.00
Spine1	4.31	Head	4.37
Spine2	4.31	Hip	10.93

그림 2 물리 법칙이 적용된 캐릭터 모델. 총 18개의 신체 부분으로 구성. 오른쪽 표는 신체 이름과 무게

제어가 균형 유지까지 고려하도록 하기 위해 강화학습을 적용하 학습하는 과정을 거쳤다. 본 연구에서는 각 캐릭터가 탈출하는 중 넘어지는 일이 없다고 가정하여 이 과정을 단순화하였다. 균형을 잡는 방법을 따로 학습하는 대신, 캐릭터 중심에 넘어지는 방향과 반대 방향으로 인공적인 힘을 가하여 넘어지지 않게 하였다.

3. 탈출 시뮬레이션 실험

100명의 사람이 탈출하는 동일한 상황을 설정하고, 서로 다른 최대 속도를 가지는 캐릭터들이 탈출하는 실험을 수행하였다. 각 사람의 최대 속도는 빠른 걸음(시속 5.35 km), 보통 걸음(시속 4.15 km), 느린 걸음(시속 2.7 km) 중 하나로 설정하였다.

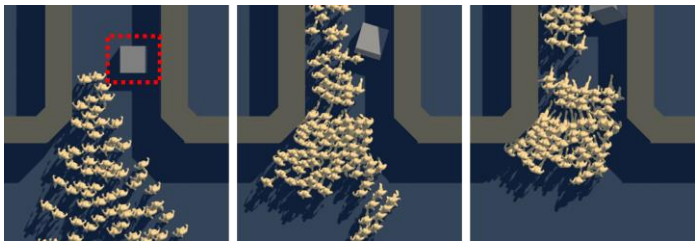


그림 1 빠른 속도로 100명이 탈출하는 실험 결과의 캡쳐 이미지 (왼쪽부터 시간 순으로)

그림 2는 기준이 되는 실험으로, 100명 모두 빠른 걸음으로 탈출하는 경우이다. 빨간색 점선으로 표시된 박스는 장애물을 나타낸다. 처음에는 사람들이 장애물을 비켜가지만, 인파 밀도가 높아지면 장애물을 밀치고 가는 것을 볼 수 있다. 이는 일반 에이전트 기반 시뮬레이션과의 차이를 나타내며, 물리 시뮬레이션을 통해 사물과 상호작용하는 것을 확인할 수 있다. 이 실험에서 모든 사람이 탈출하는 데 걸린 시간은 11.26초였다.

그림 3은 서로 다른 최대 속도를 가지는 사람들을 혼합하여 탈출 시뮬레이션을 실행한 두 가지 결과이다. 빠른 걸음 34명, 보통 걸음 33명, 느린 걸음 33명으로 구성하였고, 그림에서 각각 노랑, 초록, 빨강으로 표시되었다. 첫 번째 실험(그림 3의 상단)은 속도가 다른 사람들 사이의 출발 위치를 무작위로 배치한 것이고, 두 번째 실험(그림 3의 하단)은 출발 시 빠른 걸음 사람들을 가장 앞에, 보통 걸음 사람을 중간, 느린 걸음 사람을 가장 뒤에 배치한 후 시뮬레이션을 시작한 것이다. 모든 군중이 탈출하는 데 걸린 시간은 각각 21.53초와 18.31초로 나타났다. 빠른 걸음 사람이 먼저 출발한 것이 모두 탈출하는 시간을 줄이는 것으로 나타났다. 그림에서 보듯이 속도가 다른 사람이 혼합되어 있으면 밀도가 높아졌을 때 느린 걸음 사람들이 다른 사람들을 막아 정체가 심해지는 것을 확인할 수 있다. 또한 두 번째 실험에서는 빠른 걸음 사람들이 장애물을 먼저 제거하여 느린 걸음 사람들이 따라오는 것을 쉽게 하는 경향이 발견된다.

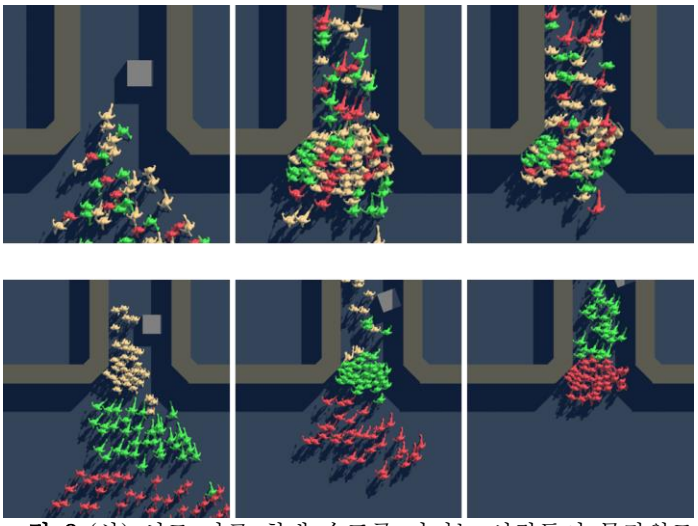


그림 3 (상) 서로 다른 최대 속도를 가지는 사람들이 무작위로 배치되어 출발한 결과 (하) 빠른 걸음 사람들이 앞에 서고 느린 걸음 사람들이 뒤에서 출발한 결과

4. 결론

물리 기반 시뮬레이션을 통해 사람과 사물 간의 상호작용을 보다 현실적으로 재현하는 구중 시뮬레이션 결과를 도출하였으며. 이를 통해 서로 다른 속도를 가지는 사람들이 혼합된 경우, 빠른 사람들이 먼저 출발하는 것이 전체 탈출 시간을 줄이는 데 효과적임을 확인하였다. 향후 인파 사고 예방 및 군중 관리 시스템의 설계와 개선에 유용하게 활용될 수 있을 것이다. 에이전트를 원통 등으로 단순화하는 기존 연구와의 직접적인 정량적 비교가 없는 점이 한계로 남아있다. 향후 이러한 비교 분석을 수행하고 더불어 개인의 나이, 성별, 장애 등에서 오는 신체적 차이를 반영하여 물리 시뮬레이션 적용하여 사실성을 높여갈 예정이다.

참고문헌

[1] Ye, J., Liu, Z., Liu, T., Wu, Y., & Wang, Y. (2024). Crowd evacuation simulation based on hierarchical agent model and physics-based character control. *Computer Animation and Virtual Worlds*, 35(3), e2263.

GAIA: 문제 해결 전략과 감정 조절 전략 제시를 통합한 게임 AI 어시스턴트 서비스 프레임워크*

엄가람[1,†], 박은별[2,†], 채지훈[1,†], 도영임[1,2]
한국과학기술원 문화기술대학원[1], 메타버스대학원[2], Games and Life Lab[1,2]
{karam.eum, silverstar, chaejihun, yydoh}@kaist.ac.kr

GAIA: A Game AI Assistant Service Framework Integrating Problem-Solving and Emotion Regulation Strategies

Karam Eum[1], Eunbyul Park[2], Jihun Chae[1], Young Yim Doh[1,2]
KAIST Graduate School of Culture and Technology[1], Graduate School of Metaverse[2], Games and Life Lab[1,2]

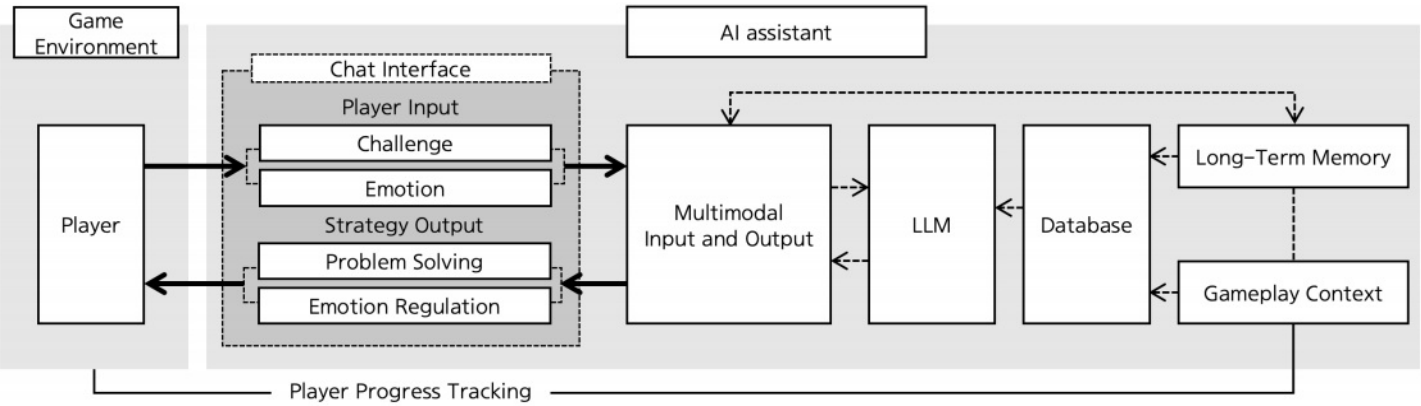


그림 1. GAIA 서비스 프레임워크

요약

본 논문은 플레이어가 게임 플레이 중 겪는 어려움을 해결하기 위해 활용할 수 있는 문제 해결 및 감정 조절 전략을 제공하는 하이브리드 게임 AI 어시스턴트 GAIA(Game AI Assistant) 서비스 프레임워크를 제시한다. GAIA는 플레이어의 어려움을 파악하여 적절한 문제 해결 또는 감정 조절 전략을 제공함으로써 플레이어의 게임 지속 경험을 개선하는데 기여할 수 있다.

1. 서론

최근 AI의 급격한 발전으로 게임 산업에서도 생성형 AI를 활용한 게임 AI 어시스턴트가 주목받고 있다. 예를 들어, Nvidia의 Project G-Assist는 플레이어와 실시간으로 대화하며 게임 전략을 추천하고 조작법을 안내한다[1]. 그러나 현재 게임 AI 어시스턴트는 정보 제공 등 인지적 지원에 치중하여 감정적 어려움에 적절한 지원을 제공하지 못하는 한계가 있다. 선행연구에서는 감정적 지원을 해주는 AI 개발을 위해 협동게임에서 플레이어가 감정을 조절하는 방법을 다룬 데이터 프레임워크를 제시한 바 있다[2]. 여기에서 나아가, 본 연구는 플레이어의 어려움에 맞춰 적절한 문제 해결 및 감정 조절 전략을 지원하는 하이브리드 게임 AI 어시스턴트 프레임워크인 GAIA를 제안한다.

2. GAIA 서비스 프레임워크

GAIA는 플레이어가 게임 중 겪는 어려움을 전달하고 필요한 지원을 받을 수 있는 게임 외부 AI 서비스로, 채팅 인터페이스와 게임 맥락(Gameplay Context), 대화 정보 등을 저장하고 처리하는 AI 어시스턴트로 구성되어 있다(그림 1 참고). GAIA는 플레이어의 어려움을 정보 및 기술 습득의 어려움과 감정 조절 어려움으로 분류한다(표 1 참고). 정보 및 기술 습득의 어려움은 정보 접근 문제나 게임 문법에 익숙하지 않은 경우를 포함하며, 감정 조절 어려움은 좌절감을 인식하고 적절히 표현하는 데 어려움을 겪는 경우로 세분화된다.

2.1. 채팅 인터페이스

플레이어와 AI 어시스턴트는 채팅 인터페이스를 매개하여 상호작용한다. 채팅 인터페이스는 게임 화면과 오버레이되어 플레이어가 글이나 음성 형태로 정보를 입력하거나 AI 어시스턴트가 출력한 정보를 확인할 수 있다.

2.2. AI 어시스턴트

게임 환경(Game Environment) 안에서 플레이어가 겪는

어려움은 게임 요소와의 상호작용 맥락 안에 위치해 있다. AI 어시스턴트는 해당 어려움이 위치한 맥락을 파악하기 위해 플레이어의 행동, 선택, 반응 등 게임 진행 상황을 실시간으로 트래킹한다.

	분류	종류
정보 및 기술 습득의 어려움	정보 부족	환경설정
		공략 정보
		게임 진행 정보
	숙련도 부족	조작 정보
		위기 대처
		실력 장벽
감정 조절의 어려움	분류	
	감정 인식 : 플레이어가 느끼는 감정의 종류를 구분하기 어려움	
	감정 표현 : 플레이어가 느끼는 감정을 적절히 공유하기 어려움	
	감정 강도 조절 : 플레이어가 적절한 감정 표현 수준을 조절하기 어려움	

표 1. 플레이어가 겪는 어려움 분류

플레이어가 채팅에 자신이 겪는 어려움을 입력하면, 어시스턴트 내부의 LLM은 게임 진행 상황 정보를 불러오고, 플레이어가 입력한 정보를 종합해 어려움의 종류를 세부적으로 분류한다. 정보 및 기술 습득의 어려움일 경우 데이터베이스에서 적합한 환경 설정이나 전략 정보를 불러와 제공하고, 감정 조절의 어려움일 경우 감정 전략 데이터베이스에서 감정 구체화, 감정 표현 전략 등 적절한 감정 조절 전략을 불러와 제시한다. 감정 수치가 일정 수준 이상인 대화 정보는 장기 기억(Long-Term Memory)으로 저장되어 추후 플레이어가 비슷한 종류의 어려움을 호소할 때 적절한 문제 해결 전략을 판단하는 데에 사용된다.

3. UX 시나리오

GAIA 서비스의 작동 결과를 보여줄 수 있는 사용자 경험(UX) 시나리오를 제작하였다(그림 2 참고). 플레이어가 문제 상황을 입력하면 GAIA는 플레이어의 입력과 게임 맥락을 분석하여 알맞은 전략을 제시한다. 예를 들어, 자막 크기 조정이나 보스 전투 공략 등 정보 및 기술 습득의 어려움일 경우 환경 설정 방법이나 적합한 게임 전략을 알려준다. 한편 다른 플레이어와의 갈등 상황 등에서 감정 조절의 어려움을 겪을 경우, 감정 표현 방법과 강도에 대한 조언을 제공한다.

4. 결론

GAIA는 플레이어가 게임 플레이 중 겪는 어려움에 대처하는 데 활용할 수 있는 문제 해결 전략과 감정 조절 전략을 제공하는 AI 어시스턴트 서비스 프레임워크이다. 후속 연구로는 실제로 GAIA를 구축해 이용자 조사를 진행하고 평가하는 실증 연구가 필요하다. 또한 게임 진행 맥락에 맞는 감정 조절 전략을 학습, 제공하기 위한 감정 전략 데이터베이스 구축이 필요하다.

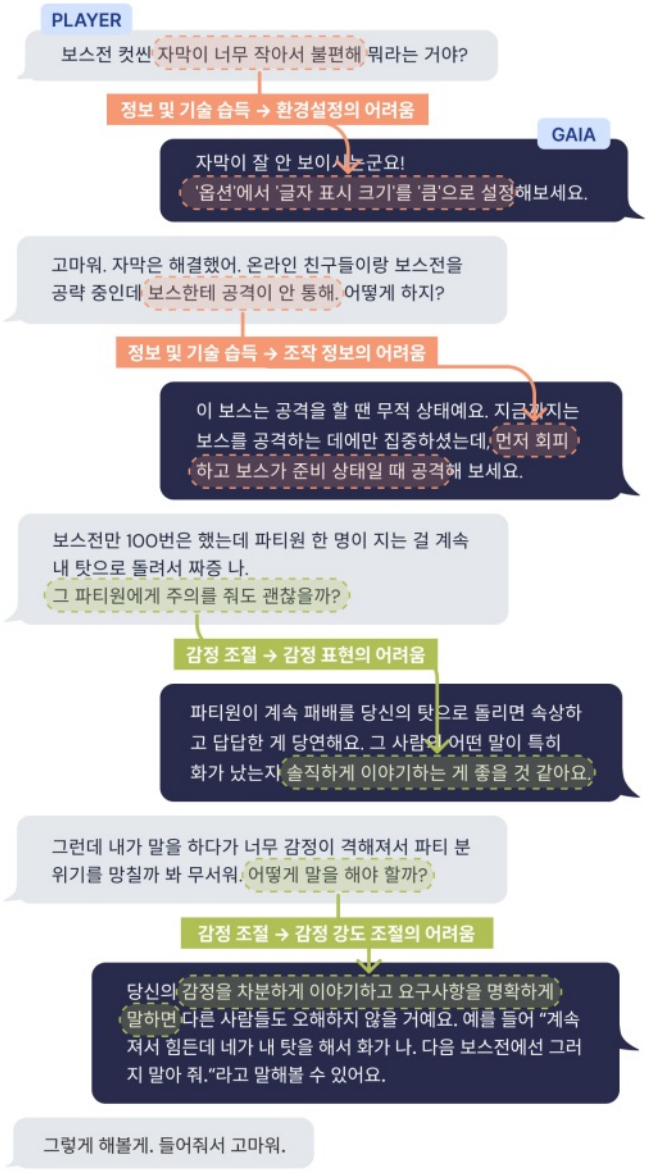


그림 2. UX 시나리오

* 구두(포스터) 발표논문
* † 동일하게 기여함
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2024-00355534).

참고문헌

[1] Nvidia. (2024.06.02). Project G-Assist 소개: 게임 및 앱 경험을 향상시키는 AI 어시스턴트 미리보기. https://www.nvidia.com/ko-kr/geforce/news/g-assist-ai-assistant/

[2] Habibi, R., Pfau, J., Holmes, J., & El-Nasr, M. S. (2023). Empathetic AI for empowering resilience in games. arXiv preprint arXiv:2302.09070.

Bullet Sign: 익스트랙션 슈터 멀티게임 개발

윤세진[0], 김준호, 조인호, 최예찬, 이강윤, 김선정
한림대학교 소프트웨어학부 콘텐츠IT 전공
{yunsejin1028, junhokim0916official, zjxps2007, dpcks7282, dkdpdh9}@gmail.com
sunkim@hallym.ac.kr

Bullet Sign: Development of Extraction Shooter Multi-Game

Se-Jin Yun[0], Jun-Ho Kim, In-Ho Jo, Ye-Chan Choi, Kang-Yun Lee, Sun-Jeong Kim
Major of Contents IT, Hallym University

요약

유니티 3D 엔진과 PUN2(Photon Unity Network 2) 네트워크 엔진을 사용하여 새로운 재미 요소가 포함된 익스트랙션 슈터 장르의 게임 Bullet Sign을 개발하였다. 기존 게임의 기능을 자체적으로 구현하면서도 새로운 화물 시스템을 추가해 전투와 파밍에서 신규 재미 요소를 추가했다.

1. 서론

익스트랙션 슈터 장르는 “Escape From Tarkov”[1]의 출시로 새롭게 정립된 장르로, 한 세션에 진입하여 NPC (Non Player Character)나 플레이어를 제거하거나 회피하고, 파밍, 루팅하여 무기와 아이템 등의 재화를 수집하며 해당 세션을 탈출하여 생존하는 게임 장르이다.

기존 게임은 극사실주의 전투를 기반으로 하여 높은 집중력을 요구하므로 피로도가 높다는 단점이 있었기 때문에 장르는 유지하면서 새로운 시스템을 도입하여 피로도를 낮추고 새로운 재미 요소를 추가하고자 Bullet Sign을 제작하게 되었다.

게임의 주요 특징으로는 기존 익스트랙션 슈터 장르에서는 볼 수 없던 화물 시스템 도입을 통해 기존과는 다른 전투 및 게임 플레이의 다양성을 확보했다는 점이다. 개발에 사용된 게임 엔진과 네트워크 솔루션은 유니티 3D 엔진과 포톤 PUN2 엔진[2]으로 이미 다양한 게임에 사용되고 있어 활용도가 높고 낮은 접근성 및 학습의 용의성이 높아 개발이 용이하기 때문에 선택하게 되었다.

2. 개발 과정

본 게임을 제작하기에 앞서 먼저 기획, 디자인, 개발 및 테스트의 세 단계의 과정을 거치도록 설계하였다. 기획 단계에서는 게임의 장르와 스토리, 디자인 단계에서는 캐릭터 디자인, 배경 그래픽, UI/UX 디자인 등을 결정하였다. 마지막으로 개발 및 테스트 단계에서는 멀티플레이 시스템, 캐릭터, UI/UX, 배경 맵 및 아이템으로 나누어 개발하였다.

멀티플레이 시스템은 닉네임 입력, 서버 접속, 로비 연결, 게임 참가 등의 기능과 방에 모집할 수 있는 최대 인원이 전부 참여하거나 일정 시간이 경과하면 자동으로 게임이 시작되고 게임이 끝나면 플레이어는 자동으로 게임에서 로비로 이동하는 기능을 개발하였다.

캐릭터는 다시 세분화하여 기본적인 이동과 무기를 관리하는 기본기능, 소지한 물건이나 화물과 상호작용하는 인벤토리, 게임의 핵심 특징이 되는 화물 시스템 개발로 나누었다.

캐릭터 기본기능은 구매한 에셋[3]의 이동, 무기 조준 및 발사 기능을 사용하면서도, 화물의 장착과 사용이 가능하도록 변경하였다. 또한 사격 시 탄환 프리팹 발사와 필드 아이템을 습득 등 게임에 필요한 동기화 기능도 구현하였다. 기본기능을 구현하며 발생한 가장 주요한 문제는 캐릭터 애니메이션과 관련된 IK(Inverse Kinematics) 문제였다. 이는 캐릭터가 무기를 장착하여 활용할 때, 무기가 원하는 위치에 부착되지 않는 문제로 이를 해결하기 위해 사용하는 무기마다 캐릭터의 자세 IK를 일일이 수정하여 각자 위치에 맞는 수치값을 미리 저장해둔 뒤, 불러오는 방식으로 해결하였다.

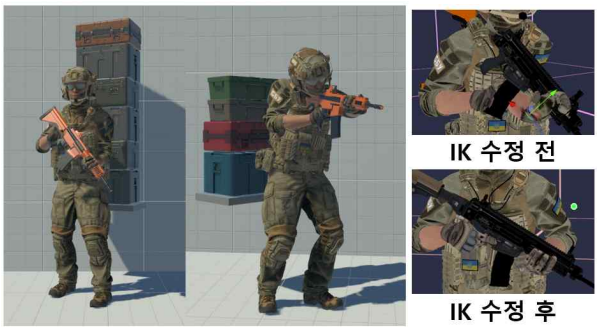


그림 1 구현된 캐릭터 및 IK 오류 수정

인벤토리 개발은 구매한 에셋[4]이 싱글 플레이를 기반으로 만들어져 있었기 때문에 멀티플레이용으로 변경하는 과정이 필요했다. 대표적으로 플레이어 사망 또는 가방, 지갑 등 내부에 아이템을 보관할 수 있는 아이템을 드롭하는 경우 아이템 커스텀 클래스와 트리 구조를 PUN2의 RPC(Remote Procedure Call) 함수가 동기화할 수 없는 문제가 있었다.

따라서 아이템 클래스와 트리 구조를 JSON을 이용해 직렬화한 다음 JSON 문자열을 RPC 함수를 이용해 동기화한 뒤, JSON 문자열을 역직렬화하여 아이템 트리 구조를 복원하는 방법으로 해결하였다.

또한, JSON 트리 구조의 경우 재귀적 직렬화, 역직렬

* 포스터 발표논문
* 학부생 주저자 논문임
* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00254695).

화가 필요하지만 유니티의 경우, 무한 루프의 가능성을 우려하여 재귀의 깊이를 10으로 제한하고 경고가 출력하는 문제도 있었다. 이는 직렬화를 위한 타입을 따로 정의하여 직렬화 클래스를 구성함으로써 해결하였다.

그림 2 그리드 스타일 인벤토리

화물 시스템 개발은 독자적인 시스템으로 개발되었다. 각각의 화물은 폭발성, 유독성, 파손 주의 등의 특성을 가진다. 외부의 공격을 받아 제거되는 경우, 폭발성 화물은 즉시 폭발하여 주변의 모든 캐릭터에게 피해를 입히고, 유독성 화물은 광역 지속 피해를 입히는 지역을 생성한다. 이처럼 전투에 도움 되는 화물을 사용하여 다양하고 특색 있는 전투를 진행할 수도 있는 반면, 화물을 각 세션 도중 물자창고에 배달하여 한 세션에서 파밍 할 수 있는 재화를 극대화하는 방법도 존재한다.

화물 시스템의 개발은 기존 게임의 아이템 파밍 및 전투 시스템 양측에 새로운 재미 요소를 추가할 수 있었지만, 각 화물에 필요한 데이터가 많아지고 각 데이터마다 상호작용 요소가 증가함에 따라 데이터 관리의 복잡성이 늘어나는 문제가 존재했다. 이를 해결하기 위해 데이터 테이블, Scriptable Object 기법을 사용하여 모든 화물이 생성될 때, 원본 값의 메모리를 참조하는 얕은 복사를 수행하도록 하여 할당 메모리를 감소시키는 방식으로 최적화를 진행하였으며 데이터 일괄 관리의 용의성을 얻을 수 있었다. 추후 화물 종류가 증가하면 SO로 모든 정보를 다루기 어렵기 때문에 JSON을 이용한 직렬화, 역직렬화를 사용한 관리 방법도 추가하였다.

그림 3 화물 종류와 제거될 때 발생하는 이펙트

UI/UX 개발은 다양한 게임의 UI 배치와 기능을 참고하여 제작되었다. 메인화면 UI는 현재 게임에 접속 중인 인원을 알 수 있는 텍스트와 게임에 참여할 수 있는 Match Start 버튼, 캐릭터를 한눈에 볼 수 있는 동적 바탕화면으로 구성하였다.

재화로 다양한 아이템을 구매 및 판매할 수 있는 상점 화면과 아이템을 장비할 수 있는 인벤토리 화면도 개발하였다. 임무를 수행하여 게임 진행 후 보상을 얻을수 있는 퀘스트 화면과 업적 화면도 추가하였다.

인게임 UI는 현재 진행 중인 게임에 최대한 방해가 되지 않도록 좌측 하단에 현재 HP와 달리기, 점프 등에 쓰이는 스태미너가 표시된다. 우측 하단에는 현재 무기와 장탄 수가 표기되며 화면의 정중앙에는 정확한 조준을 돕는 크로스 헤어가 표기된다.

로비 UI 인게임 UI

그림 4 로비 UI와 인게임 UI

배경 맵 및 아이템 개발은 주로 기존의 에셋에 추가하는 방향으로 진행되었다. 에셋에서 기본적으로 제공하는 데모 맵을 동일한 스타일로 확장하는 작업이 이루어졌다. 플레이어가 생성 및 탈출 지점을 추가하고 아이템은 주변의 사물 크기와 동일하도록 조정하는 작업을 진행하였다. 또한 맵에 아이템 생성 지점을 추가하여 게임 시작 전에 아이템이 정해진 위치에 생성되도록 하였다.

그림 5 게임 내 배경

3. 결론 및 향후 계획

익스트랙션 슈터 장르에 화물 요소를 추가하여 재미 요소를 추가한 게임을 제작하였다. Bullet Sign은 PUN2 네트워크 엔진을 사용하여 만들었기 때문에 지연시간이 최소화된 기존 게임과 달리 지연시간이 약간 더 필요하다는 단점이 있다. 같은 포톤 사의 신규 네트워크 엔진 FUSION2로 변경하여 지연시간과 관련된 문제들을 개선할 예정이다. 향후 온라인 배급사에 정규 출시를 진행하거나 배틀 패스와 같은 BM(Business Model)을 적용하여 수익을 창출할 수 있을 것으로 기대한다.

참고문헌

[1] Escape From Tarkov, Homepage, https://www.escapefromtarkov.com/

[2] Photon PUN2, Homepage, https://www.photonengine.com/ko-kr/pun

[3] Unity Asset, https://assetstore.unity.com/packages/3d/characters/humanoids/humans/modular-soldier-pack-222861

[4] Unity Asset, https://assetstore.unity.com/packages/tools/game-toolkits/ultimate-grid-inventory-215256

탄성체 시뮬레이션을 위한 모멘텀 보존 요소 반전 완화 기법*

정희조[0,a], 김승욱[b], 이재현[a], 엄기원[c], 기민형[a], 한정현[a]
고려대학교 컴퓨터학과[a], 한국외국어대학교 컴퓨터공학부[b], 텔레콤파리대학 이미지, 데이터, 신호학과[c]
jhj1629343@korea.ac.kr, seungwook.kim@hufs.ac.kr, dlwogus1223@korea.ac.kr,
kiwon.um@telecom-paris.fr, minh702@korea.ac.kr, jhan@korea.ac.kr

Momentum-preserving Inversion Alleviation for Elastic Material Simulation

Heejo Jeong[0,a], Seung-wook Kim[b], JaeHyun Lee[a], Kiwon Um[c], Min Hyung Kee[a], JungHyun Han[a]
Korea University [a], Hankuk University of Foreign Studies [b], LTCI, Telecom Paris, IP Paris [c]

Abstract

This paper proposes a novel method that enhances the optimization-based elastic body solver. The proposed method alleviates element inversions in the prediction stage of the prediction-projection approach, thus enhancing stability and efficiency in the subsequent projection stage. We introduce a velocity decomposition method and adjust only the non-rigid motion while preserving both linear and angular momenta in prediction stage. Experimental results demonstrate improvements in stability of the simulation, particularly when large deformations hinder the solver.

1. Introduction

In the computer graphics field, elastic body simulations are exploited in various applications including virtual surgery and clothing design. For such purposes, optimization-based solvers(i.e., Projective Dynamics(PD)[1]) are widely adopted thanks to their stability and efficiency. In PD, time integration is performed in two stages: prediction and projection. In the prediction stage, the next physical state is initialized with inertial motion. Then, in the projection stage, the initialized state is projected onto the optimal state, searching for the compromise between the prediction and material properties.

However, incorrect predictions result in instability and inefficiency in the projection stage, and this problem may arise from *element inversions* problem. To address this problem, Lee et al. [2] alleviates the inversion at the prediction stage. However, this method leads to losses in angular momentum and energy.

* 포스터 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 CASA 2024에 채택되었음
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2024- 2020-0-01460)

In this paper, we propose a new method to alleviate element inversions in the prediction stage. We introduce velocity decomposition to adjust only the non-rigid velocity, thereby preserving both linear and angular momentum. Consequently, it effectively represents the motion of elastic bodies and enhances simulation stability.

2. Method

Given the current position $\mathbf{x}^n$ and velocity $\mathbf{v}^n$, the problem of determining the next time step position $\mathbf{x}^{n+1}$ using Backward Euler is reformulated as follows:

$$\underset{x^{n+1}}{\mathbf{argmin}} \frac{\mathbf{1}}{\mathbf{2}h^2}||\mathbf{x}^{n+1} - \mathbf{y}||_{\mathbf{M}}^2 + W(\mathbf{x}^{n+1}) \quad \cdots \quad (1)$$

Here, h is the time step size, $\mathbf{y}$ is the predicted state, $\mathbf{M}$ is the mass matrix, $\mathbf{f}_{\text{ext}}$ is the external force, and W is elastic potential. In the prediction stage, $\mathbf{y}$ is predicted as $\mathbf{y} = \mathbf{x}^n + h\mathbf{v}^n + h^2\mathbf{M}^{-1}\mathbf{f}_{\text{ext}}$. After prediction stage, the iterative projection algorithm is performed to satisfy Eq.(1) and obtain $\mathbf{x}^{n+1}$. For the inverted elements arising in the prediction stage, our method modifies $\mathbf{y}$. We first present how to resolve a single element inversion and then move on to the iterative method for resolving multiple inversions.

2.1 Single inversion

Given an inverted tetrahedral element, its vertex indices are represented as $i \in \{1,2,3,4\}$.

Let us define the "predicted" velocity $\tilde{\mathbf{v}}_i$ as follows:

$$\tilde{\mathbf{v}}_i = \frac{y_i - x_i}{h} = v_i + \frac{h}{m_i} f_{ext,i} \quad \cdots \quad (2)$$

Our method decomposes the predicted velocity into a *rigid* part, which has no influence on deformation, and a *non-rigid* part, which deforms the element. We adjust only the *non-rigid* part to preserve linear and angular momenta.

Let $\tilde{\mathbf{v}}_{\text{rigid}}$ denote the *rigid* velocity. It involves the linear velocity, $\tilde{\mathbf{v}}_l$, and the angular velocity, $\tilde{\omega}$. The linear velocity, $\tilde{\mathbf{v}}_l$, is computed as the average velocity of the four vertices, i.e., $\tilde{\mathbf{v}}_l = \sum_i \mathrm{m}_i \tilde{\mathbf{v}}_i / \sum_i \mathrm{m}_i$. To obtain the angular velocity, $\tilde{\omega}$, we calculate angular momentum, L, and the inertia tensor, $\mathbf{I}$ as follows:

$$L = \sum_i m_i [\mathbf{x}_i - x_c]_\times \tilde{\mathbf{v}}_l \quad \cdots \quad (3)$$

$$I = \sum_i m_i [\mathbf{x}_i - x_c]_\times ([\mathbf{x}_i - x_c]_\times)^T \quad \cdots \quad (4)$$

Where $[\cdot]_\times$ represents the cross product matrix and x_c is the center of mass, that is, $x_c = \frac{\sum_i m_i x_i}{\sum_i m_i}$. Finally, we can evaluate the angular velocity, $\tilde{\omega} = \mathrm{I}^{-1}\mathrm{L}$.

With the linear and angular velocities, we can define the *rigid* velocity of each vertex as $\tilde{\mathbf{v}}_{\text{rigid},i} = \tilde{\mathbf{v}}_l + [\tilde{\omega}]_\times [\mathbf{x}_i - x_c]$. Using this, we can simply define the *non-rigid* velocity as $\tilde{\mathbf{v}}_{\text{non-rigid},i} = \tilde{\mathbf{v}}_i - \tilde{\mathbf{v}}_{\text{rigid},i}$.

Now we redefine predicted velocity of each vertex $\tilde{\mathbf{v}}_i$ as follows:

$$\hat{\mathbf{v}}_i = \tilde{\mathbf{v}}_{rigid,i} + \alpha \tilde{\mathbf{v}}_{non-rigid,i} \quad \cdots \quad (5)$$

where $\alpha (\in [0,1])$ represents the *alleviation factor*.

To find α that ensure the predicted velocity prevent inversion, we use the signed volume J of the tetrahedron, defined in terms of the predicted position of each vertex $\hat{\mathbf{y}}_i = \mathbf{x}_i + h\hat{\mathbf{v}}_i$, as follows:

$$J = \det\left(F(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_3, \hat{\mathbf{y}}_4)\right) \quad \cdots \quad (6)$$

where F is the deformation gradient operator and $\det$ is the determinant operator. Note that in Eq.(6) J is a cubic function of α. By solving the cubic polynomial equation, $J(\alpha) = 0$, we can find α. However, we approximate α as follows:

$$\alpha = \max\left(0, \min\left(1, \frac{\mathrm{J}(0)}{\mathrm{J}(0) - \mathrm{J}(1)}\right)\right). \quad \cdots \quad (7)$$

Note that α in Eq.(7) is the solution clamped between 0 and 1 of $J(\alpha) = 0$, when J is approximated as a linear function of α. Using α the predicted state $\hat{\mathbf{y}}$ is updated. Finally, the objective function of the optimization problem in Eq.(1) is updated with $\hat{\mathbf{y}}$.

2.2 Multiple inversion

In general, alleviating an element may invert its neighbors. To address this problem, we employ a Jacobi-style iterative algorithm. The algorithm continues until no inversion remains or the iteration count surpasses a predefined threshold. Also, we monitor inversion severity using the signed volume of tetrahedra. If the sum of absolute signed volumes of inverted tetrahedra increases, we halt the algorithm.

3. Results

Given PD as an underlying solver, we compare our method with the previous inversion alleviation method, Linear Inversion Alleviation(LIA)[2], which reverts the inverted elements preserving their linear motions[2].

Fig 1: The elastic cubes swing in an elliptical motion.

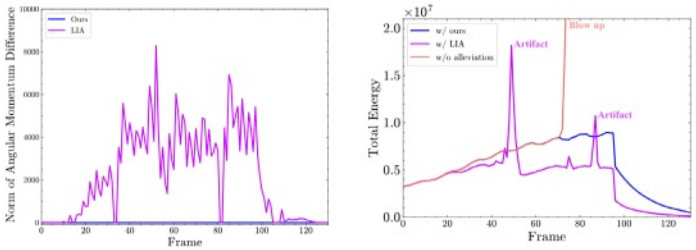


Fig 2: The left graph shows the norm of angular momentum difference, while the right graph shows the total energy.

Fig 1 compares the simulation results generated with the following three strategies; PD (in red), PD+ LIA (in purple), and PD+ Ours (in blue). As shown in the snapshots of Fig 1, "PD+ Ours" successfully stabilizes the simulation, while PD fails and PD+ LIA shows dissipative motion. The graphs in Fig 2 shows that our method preserves angular momentum, resulting in less energy loss.

4. Conclusion

This paper proposes a new method to alleviate the *element inversion* in *optimization-based* simulations. We decompose the velocity and adjusts only the *non-rigid* components, preserving angular momentum and minimizing the impact on the original motion, thereby improving the stability of simulation and maintaining vivid motion. Our method is performed independently of the projection stage, making it applicable to other method included in the *prediction-projection* approach. Applying our method to different solvers and examining its benefits is our future work.

Reference

[1]Sofien Bouaziz, Sebanstian Martin, Tiantian Liu, Ladislav Kavan, and Mark Pauly. Projective dynamics: Fusing constraint projections for fast simulation. *ACM transaction on graphics (TOG), 33(4):1-11,2014*

[2] JaeHyun Lee, Seung-wook Kim, Kiwon Um, Min Hyung Kee, and JungHyun Han. Inversion alleviation for stable elastic body simulation. *Computer Animation and Virtual Worlds, 34(3-4):e2183, 2023.*

스토리텔링 기반 장면을 이해하는 다중 인간 동작 생성

임동근[0], 배진석, 황인우, 김영민
서울대학교
{rms2836, capoo95, inusu0818, youngmin.kim}@snu.ac.kr

Storytelling-Driven Multi-Human Motion Synthesis with Scene Understanding

Donggeun Lim[0], Jinseok Bae, Inwoo Hwang, Young Min Kim
Seoul National University

요약

본 연구에서는 주어진 3차원 공간과 스토리텔링을 기반으로, 공간의 맥락을 이해하며 주변 환경 및 사람들과 자연스럽게 상호작용하는 다수의 가상 인간의 움직임을 계층적으로 생성하는 프레임워크를 제안한다. 제안된 프레임워크는 거대 언어 모델을 활용하여 개별 사람의 행동 계획을 결정하는 상위 수준의 행동 계획 모듈과 모션 매칭 기법을 활용해 3차원 공간에서 실제 움직임을 구현하는 하위 수준의 움직성 생성 모듈로 구분된다. 우리는 거대 언어 모델을 활용해 주어진 스토리텔링에 따라 다수의 가상 인간의 의사결정을 효과적으로 상호조율하고, 주변 3차원 공간의 맥락을 이해하며 행동을 계획할 수 있는 방안을 제시한다. 또한, 인간의 움직임을 자연어로 기술하기 위해 개발된 라반 움직임 분석(Laban Movement Analysis)을 활용하여 저수준의 인체 동작을 생성하는 과정에서도 거대 언어 모델을 활용할 수 있는 방안을 탐구한다.

1. 서론

컴퓨터 애니메이션 분야에서 주어진 3차원 공간 내 다수의 가상 인간들이 주변 환경 및 사람들과 상호작용하며 제각기 활동하고 있는 장면을 연출하는 것은 많은 시간과 노력이 요구되는 작업이다. 이를 위해서는 주변 상황 및 연출 의도에 맞게 개별 사람의 행동 계획을 적절하게 수립하는 것과 더불어 이를 자연스러운 3차원 움직임으로 구현하는 종합적인 과정이 필요하다. 최근 거대 언어 모델의 비약적인 발전에 힘입어, 거대 언어 모델을 활용해 로봇 혹은 가상 인간의 행동 계획을 수립하는 연구들이 다수 진행되었다. 하지만, 이러한 기존의 연구들은 많은 경우 실제 움직임을 구현하는 단계까지는 포함하지 않으며, 특히, 오직 단일 개체를 대상으로 하거나, 3차원 공간에 기반한 행동 계획을 수립하지는 못하는 등의 한계점들이 존재한다.

본 논문에서는 이러한 한계점들을 극복하여, 거대 언어 모델을 활용해 주어진 3차원 공간과 스토리텔링을 종합적으로 고려하면서 여러 가상 인간의 행동을 계획하고, 이를 올바른 상호작용을 포함한 3차원 움직임으로까지 구현하는 프레임워크를 제안한다. 우리의 프레임워크는 개별 사람의 행동 계획을 결정하는 상위 수준의 행동 계획 모듈과 3차원 공간에 실제 움직임을 구현하는 하위 수준의 움직임 생성 모듈로 계층적으로 구성된다. 우리의 연구는 거대 언어 모델을 파인튜닝하거나, 움직임 생성을 위한 새로운 모델을 학습하지 않고도 앞서 설명된 도전적인 상황에서의 행동 계획을 수행하고 그에 따른 움직임을 생성할 수 있음을 보인다.

2. 방법

2.1. 상위 수준의 행동 계획 모듈

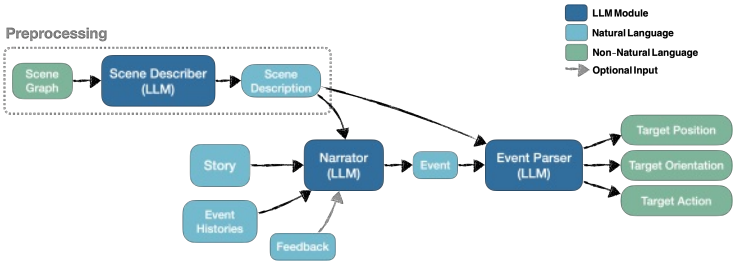


그림 1. 행동 계획 모듈 구조도

상위 수준의 행동 계획 모듈은 공간의 맥락 및 개체간의 상호작용을 고려하며 여러 개체의 행동 계획을 수립한다. 우리는 이러한 복잡한 문제를 여러 작은 단위의 문제들로 분할함으로써 전체 문제를 보다 효율적으로 해결한다 (그림 1).

거대 언어 모델에게 3차원 공간에 대한 정보를 전달하기 위해, 우리는 3차원 공간으로부터 장면 그래프를 추출하고 이로부터 물체들과 물체들간의 연결관계에 대한 정보를 얻어 공간에 대한 묘사를 생성한다. 이 때, 주어진 공간을 의미론적 문맥에 따라 계층적으로 파악하여 묘사하도록 주문함으로써, 거대 언어 모델이 마치 사람이 공간을 인식하는 것과 유사하게 공간을 효율적으로 이해하고 표현할 수 있게 한다 (그림 2).

주어진 스토리텔링에 따라 여러 사람들의 행동들을 올바르게 계획하기 위해서는 거대 언어 모델은 현재 장면의 전체적인 상황을 종합적으로 파악할 수 있어야 한다. 이를 위해, 우리는 이벤트 기반의 (event-driven) 방식으로 가상 인간들의 행동 계획 및 장면의 상태를 표현함으로써, 거대 언어 모델이 이벤트 기록을 통해 장면의

* 구두(포스터) 발표논문

현재 상황을 손쉽게 파악하고, 여러 가상 인간들의 상호작용을 보다 효과적으로 조율할 수 있게끔하였다.

최종적으로, 각각의 생성된 이벤트를 가상 인간의 움직임으로 구현하기 위해서는 고수준의 자연어로 표현된 이벤트를 목표 위치, 목표 방향, 목표 행동 라벨 등과 같은 저수준의 정보들로 변환해야한다. 하지만, 일반적으로 거대 언어 모델은 3차원 좌표 등과 같은 저수준의 정보를 직접적으로 처리하는 데에는 많은 어려움을 겪기 때문에, 우리는 거대 언어 모델이 자연어로서 위치 등에 대한 조건들을 출력하게 하고 이를 기반으로 비용 함수를 정의하고 MCMC 샘플링을 통해 목표하는 저수준 정보를 획득한다 (그림 3).

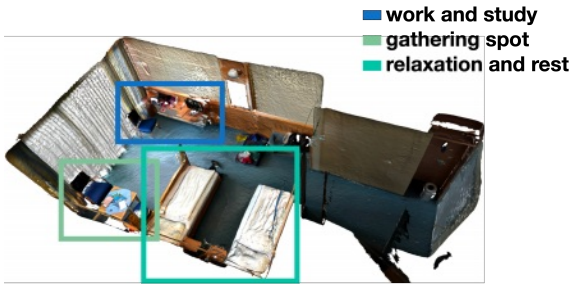


그림 2. 장면 그래프 및 거대 언어 모델을 활용해 추출한 공간의 의미적 문맥 예시

그림 3. 자언어로 표현된 위치 조건 및 그에 따른 위치 샘플링 결과 예시

2.1. 하위 수준의 움직임 생성 모듈

하위 수준의 움직임 생성 모듈에서는 텍스트 기반 모션 검색 기법 [1]과 텍스트 기반 모션 매칭 기법 [2]을 활용하여 각각의 이벤트를 3차원 공간 상에서 실제 이루어지는 가상 인간들의 자연스러운 움직임으로 구현한다 (그림 4). 더불어, 우리는 인간의 움직임을 자연어를 통해 체계적으로 표현하기 위해 널리 쓰이는 라반 움직임 분석 (Laban Movement Analysis)을 활용하여 거대 언어 모델이 인체 동작의 질적 측면 (강도, 직행성 등)에 대해서도 표현할 수 있도록 하였다. 그리고, 라반 움직임 분석의 요소들을 모션 검색 시에 반영하여 [3] 같은 행동 내에서도 여러 가지 다른 스타일의 모션들을 검색할 수 있도록 하였다 (그림 5).

3. 결과 및 결론

본 논문에서는 다수의 가상 인간들이 서로 간의 상호작용을 수반하며 동일한 3차원 공간에 함께 존재하는 상황에서, 주어진 스토리텔링에 따라 여러 사람들의 행동을 적절하게 계획하고 이를 자연스러운 움직임으로까지 구현할 수 있는 프레임워크를 제안하였다. 우리는 추가적인 학습 없이도, 여러 가상 인간들의 의사결정을 효과적으로 상호조율하고, 주변 3차원 공간의 맥락을 이해하며 행동을 계획할 수 있는 방안을 제시하였으며, 라반 움직임 분석을 이용해 거대 언어 모델이 저수준의 움직임 생성에 응용될 수 있는 방안을 탐구하였다.

그림 4. 움직임 생성 모듈을 통해 구현된 가상 인간의 움직임 예시. (좌) Event: [man_1, man_2] are handshaking. (우) Event: [man_3] is drinking coffee in front of the coffee machine.

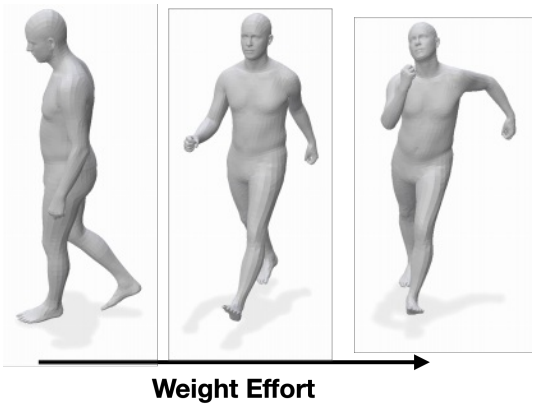


그림 5. 라반 움직임 분석의 Weight Effort 세기에 따른 '걷기' 움직임 검색 결과 예시

참고문헌

[1] Petrovich, Mathis, Michael J. Black, and Gül Varol. "TMR: Text-to-motion retrieval using contrastive 3D human motion synthesis."Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[2] Qing, Zhongfei, et al. "Story-to-motion: Synthesizing infinite and controllable character animation from long text."SIGGRAPH Asia 2023 Technical Communications. 2023. 1-4.

[3] Kapadia, Mubbasir, et al. "Efficient motion retrieval in large motion databases."Proceedings of the acm siggraph symposium on interactive 3D graphics and games. 2013.

캡쳐된 전신 동작에 대한 손가락 동작 생성 기술*

이지성[0,1], 김영빈[1], 김윤섭[1], 이지나[1], 임재호[2], 한다성[1]
[1]한동대학교 ICT창업학부
[2](주)덱스터스튜디오
{zero579, 22000127, 21900170, 22100571}@handong.ac.kr, jaeho.im@dexterstudios.com, dshan@handong.edu

Finger Motion Generation for Captured Whole Body Motion

Jiseong Lee[0,1], Young Bin Kim[1], Yunsub Kim[1], Jeana Lee[1], Jaeho Im[2], Daseong Han[1]
[1]School of Global Entrepreneurship and Information Communication Technology, Handong Global University
[2]Dexter Studios., Ltd

요약

본 논문은 주어진 전신 동작에 어울리는 손가락 동작을 생성하여 더욱 자연스러운 캐릭터 애니메이션을 합성하는 기법을 제안한다. 이를 위해 데이터 수집 단계에서는 전신 동작과 손가락 동작을 동시에 캡쳐하고 학습 단계에서는 입력으로 주어진 전신 동작으로부터 손가락 동작을 생성하는 VAE(Variational Auto-Encoder)를 학습시킨다. 실험 결과, 제안된 방법이 캐릭터의 움직임을 더욱 자연스럽게 하는 데 효과적임을 확인할 수 있었다.

1. 서론

최근, 전신 동작과 더불어 다양한 신체 부위의 동작을 생성하는 연구가 활발히 진행되고 있다[1, 2]. 본 연구는 손가락 동작을 효율적으로 생성하기 위해 VAE를 활용하여 전신 동작에 어울리는 자연스러운 손가락 동작을 생성하는 방법을 제안한다. 이를 위해 모션 캡쳐 방법을 사용하여 전신 동작과 손가락 동작 데이터를 동시에 수집하였고 전처리 과정을 거쳐 모델 학습에 사용하였다. 본 논문에서는 VAE 기반 방법론을 통해 전신 동작이 주어졌을 때 손가락 동작을 생성하는 과정을 설명하고, 이에 대한 실험 결과를 제시한다.

2. VAE 기반의 손가락 동작 생성 기법

2.1. 데이터 수집 및 전처리

본 연구에서는 자연스러운 움직임 데이터를 획득하기 위해 전신 및 손가락 데이터를 동시에 수집하였다. 전신 동작을 수집하는 방법으로는 광학식 모션 캡쳐 시스템(OptiTrack Prime 17W 12대)을 사용하였고 손가락 동작을 수집하는 방법으로는 Quantum Mocap Metagloves 장비를 사용하였다. 수집된 동작 데이터셋은 약 117개의 다양한 동작으로 구성되고(걷는 동작: 56개, 서 있는 동작: 61개), 각 동작 데이터는 24 fps의 속도로 약 8초간 캡쳐되었다. 이 데이터셋을 바탕으로 매 프레임마다 각 관절의 로컬 위치 및 로컬 회전 정보를 그 상위 관절을 기준으로 표현하여 학습 데이터셋을 구축하였다.

2.2. 학습 및 손가락 동작 생성 과정

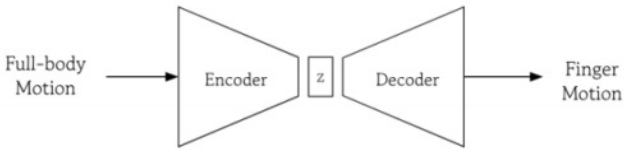


그림 1: 모델 아키텍처

본 연구에서는 전신 동작에 적합한 손가락 동작을 생성하기 위해 VAE를 사용한다(그림 1). VAE는 입력 데이터를 잠재 공간으로 압축하고 이를 재구성하여 새로운 데이터를 생성하는 확률적 생성 모델이다. VAE는의 인코더는 전신 동작 데이터를 입력받아 이를 평균과 로그 분산으로 구성된 잠재 공간으로 압축한다. 잠재 공간은 고차원 입력 데이터를 저차원 벡터로 변환하는 역할을 하며, 전신 동작 데이터는 잠재 벡터로 표현된다. 이후, 잠재 공간에서 재파라미터화(reparametrization)를 통해 잠재 변수를 샘플링하고, 디코더는 이 잠재 변수를 이용해 손가락 동작을 생성한다.
생성된 손가락 동작의 정확성을 평가하기 위해, 출력과

---

* 포스터 발표논문
* 학부생 주저자 논문
* 이 연구는 문화체육관광부 - 한국콘텐츠진흥원(KOCCA)에서 지원하는 연구비(No. R2022020123, 가상공연 참여를 위한 감정 표현이 가능한 소통형 관객 아바타 생성 기술 개발)로 수행되었습니다.

실제 손가락 동작 데이터(ground truth) 간의 손실을 계산한다. 손실 함수는 세 가지 요소로 구성된다(그림 2). 재구성 손실은 예측값과 실제 값 간의 평균 제곱 오차(Mean Squared Error, MSE)를 측정하여 생성된 출력이 실제 값에 가까운지 평가한다. KLD(Kullback–Leibler divergence)은 학습된 잠재 변수 분포와 표준 정규 분포 간의 차이를 측정하여 잠재 공간을 정규화한다. β는 MSE와 KLD 항목 간의 균형을 조절하는 가중치이다. 이러한 요소들을 통해 손실을 계산하고 모델의 성능을 평가한다.

$$\text{Loss} = \text{MSE}(output_finger, ground_finger) + \beta \times \text{KLD}(\mu, \log(\sigma^2))$$

그림 2: 손실 함수 (output_finger: 디코더의 출력 손가락 동작, ground_finger: 실제 손가락 동작)

3. 결과

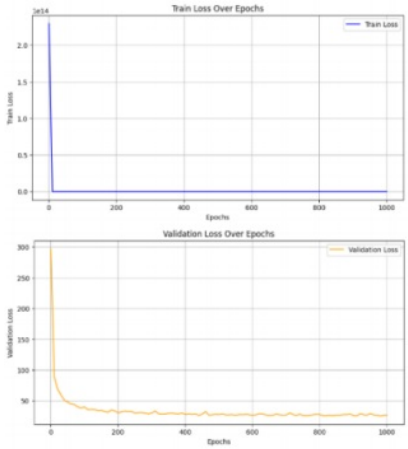


그림 3: (상) 학습 로스 그래프, (하) 검증 로스 그래프

표 1: 학습 환경 및 실험 결과

beta	hidden size	latent size	learning rate
1.0	256	16	0.001
validation loss		test loss	
24.47		24.54	

모델을 학습시키기 위해 데이터를 학습, 검증, 테스트 데이터셋으로 6:2:2 비율로 분할하였다. 모델의 성능을 높이는 데 최적화된 하이퍼파라미터(beta, hidden size, latent size, learning rate) 값을 찾기 위해, 하이퍼파라미터마다 VAE 모델에 통상적으로 많이 사용되는 후보 값들을 사전에 정의한 후, 가능한 모든 조합에 대하여 모델의 성능을 평가하는 grid search 방식을 사용하였다. 가장 높은 성능을 보이는 하이퍼파라미터 조합을 표 1에 기재하였다. 해당 하이퍼파라미터 조합으로 모델을 1000 epochs 동안 학습시켰다. 그 결과, 그림 3에 나와 있듯이 학습 손실과 검증 손실이 유사한 패턴으로 감소하고 있으며, 표 1에 나온 것처럼 검증 손실과 테스트 손실이 유사한 값을 가짐을 알 수 있다. 이를 통해 모델이 적절하게 학습되었으며, 새로운 데이터에서도 잘 작동함을 알 수 있다. 마지막으로, 생성된 손가락 동작을 전신 동작에 추가하여 시각적으로 살펴본 결과, 그림 4에 보이는 것처럼 실제 손가락 동작과 매우 유사함을 확인할 수 있었다.

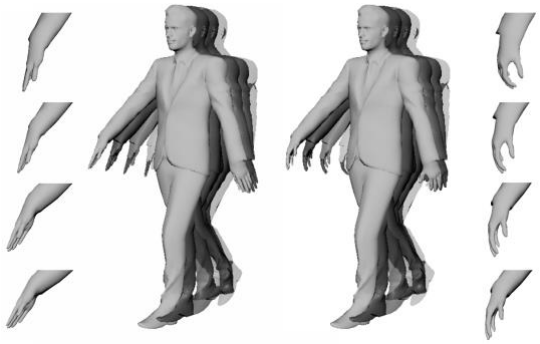


그림 4: (좌) 손가락 동작이 포함되지 않은 전신 동작, (우) 생성된 손가락 동작이 포함된 전신 동작

4. 결론 및 향후 연구

본 연구에서는 VAE를 활용하여 전신 동작을 넣었을 때 이에 어울리는 손가락 동작을 생성하는 방법을 제안한다. 이를 위해 모션 캡쳐를 통해 전신 동작과 손가락 동작 데이터를 수집하였고, 이렇게 수집된 데이터셋에 기반하여 VAE를 학습하였다.

본 연구의 한계점으로는 프레임 사이에서 손가락 동작의 연속성을 고려하지 않은 점이다. 또한 데이터셋이 제한적이어서 다양한 동작을 생성하기에는 한계가 있다. 향후 연구에서는 입력으로 전신 동작뿐 아니라, 이전 프레임의 손가락 동작을 넣어 제안된 기법을 개선할 예정이다. 또한, 모션 캡쳐 이외 영상에서 동작을 추출하는 방법을 통해 다양한 학습 데이터를 수집할 계획이다.

참고문헌

[1] Mathis Petrovich, Michael J. Black and Gül Varol, Action-Conditioned 3D Human Motion Synthesis with Transformer VAE, International Conference on Computer Vision (ICCV), 2021, pp.10985-10995

[2] S. Lu, L.-H. Chen, A. Zeng, J. Lin, R. Zhang, L. Zhang, and H.-Y. Shum, HumanTOMATO: Text-aligned Whole-body Motion Generation, arXiv, 2023, vol. 2310.12978

레이-트레이싱 코어를 활용한 3차원 표면 재구성 가속화*

김영우[0], 김덕수
한국기술교육대학교
{aister9, bluekds}@koreatech.ac.kr

Acceleration of 3D Surface Reconstruction using Hardware accelerated Ray-Tracing

YoungWoo Kim[0], Duksu Kim
KOREATECH (Korea University of Technology and Education)

요약

본 연구는 최신 GPU의 레이-트레이싱 코어를 활용한 3차원 표면 재구성 알고리즘을 제안한다. 제안하는 알고리즘은 사면체 그래프를 삼각형으로 분해하여 가속 구조를 생성하고, 광선 추적 기반의 가시성 가중치 계산을 수행한다. 또한, 광선 추적 도중 각 사면체에 대한 중복 연산 제거를 위해 후면 컬링을 사용한다. 본 연구는 제안하는 알고리즘을 다양한 세대의 RTX GPU에서 검증하였다. 그 결과, CPU 기반 가시성 가중치 계산 알고리즘 대비 최대 52.11배 높은 성능을 보여주었다.

1. 서론

3차원 재구성은 과학적 가시화, 가상 증강 현실, 그리고 컴퓨터 지원 설계 등 다양한 분야에 사용되는 기술로, 컴퓨터 비전 및 그래픽스의 중요한 연구 분야로 입력 이미지들로부터 3차원 형상을 구성한다 [1]. 이러한 3차원 재구성 기술은 입력 데이터의 잡음과 이상치로부터 큰 영향을 받으며, 이를 최소화하여 고품질의 결과를 얻기 위한 노력이 계속되어왔다 [2]. 가시성 기반 3차원 표면 재구성 기술은 이러한 3차원 재구성에 대한 강건하고 효율적인 알고리즘 중 하나이다 [2]. 최근 이러한 가시성 기반 3차원 표면 재구성을 가속하기 위한 다양한 병렬처리 알고리즘들이 제안되고 있다. 특히 GPU가 활발히 활용되고 있다 [1]. 또한, 이러한 가시성 기반 3차원 표면 재구성 기술은 고품질 3차원 재구성 기술을 위한 미분 가능한 렌더링의 전처리 단계로 활용되고 있다 [3].

본 연구는 가시성 기반 3차원 표면 재구성 기술을 가속화하기 위해 레이-트레이싱(RT)-코어를 활용한 가시성 가중치 계산 알고리즘을 제안한다. 제안하는 알고리즘은 RTX 플랫폼 활용을 위해 사면체를 삼각형으로 분리하고, 분리된 삼각형에 광선을 투사하는 것으로, 각 사면체에 대한 가시성 가중치를 계산한다. 그 결과, CPU 알고리즘 대비 최대 52.11배 높은 성능을 보여주었다.

* 포스터 발표논문
* 본 연구는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받은 기초연구사업(2021R1I1A3048263, 50%)과 지자체-대학 협력기반 지역혁신 사업(2021RIS-004, 50%)의 결과임

2. 레이-트레이싱 코어 기반 가시성 가중치 계산

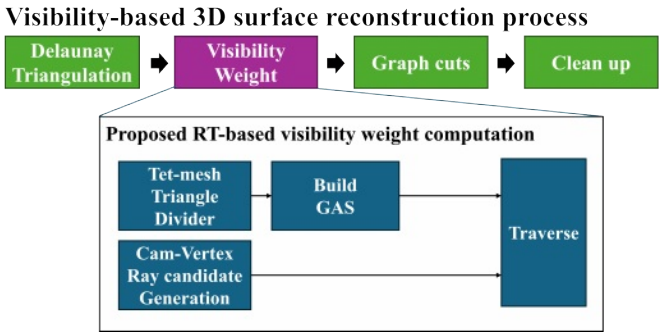


그림 1: 가시성 기반 3차원 표면 재구성 프로세스와 제안하는 RT-코어 기반 가시성 가중치 계산 방법의 개요

그림 1은 본 논문이 제안하는 알고리즘의 개요를 보여준다. 3차원 표면 재구성은 일반적으로 들로네 삼각 분할, 가시성 가중치 계산, 그래프 절단, 그리고 메쉬 클린업 단계로 구성되며 [4], 본 연구는 가시성 가중치 계산 단계를 RT-코어를 활용하여 가속한다.

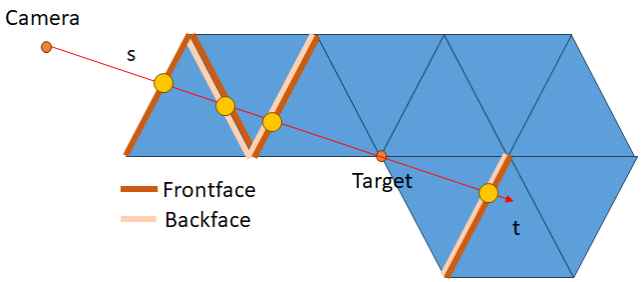
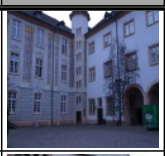


그림 2: 가시성 가중치 계산 방법

가시성 가중치 계산 알고리즘(그림 2)은 우선 카메라로부터 정점을 향해 길게 뻗은 선(line-of-sight)를 만든다. 그리고 사면체의 표면과 선이 교차가 일어날 때마다, 사면체에 목표 정점과 충돌 지점 사이 거리만큼의 가중치를 계산한다.

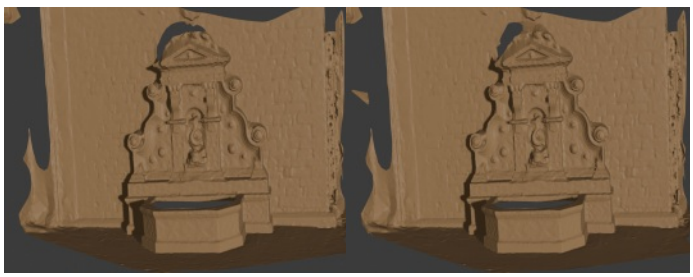
RTX 플랫폼에서는 CPU 알고리즘에서 사용하는 사면체의 그래프 기반 구조를 바로 사용할 수 없다. 따라서, 본 연구는 사면체를 삼각형으로 분해하고, 분해된 삼각형을 기반으로 RTX 플랫폼의 가속 구조인 (Geometry Acceleration Structure, GAS)를 생성한다. 또한, 각 카메라와 정점 사이의 선을 카메라로부터 정점을 향한 광선으로 표현한다. 그리고 가속 구조를 기반으로 광선

표 1: 가시성 가중치 계산 알고리즘 수행 시간 (단위: 초)

	Image	Points	Rays	CPU	RTX 2080	RTX 3080	RTX 4080
Castle-P30		636K	2.162M	13.987	0.673	0.383	0.320
Herz-Jesu-P8		1.285M	5.633M	35.140	2.275	1.026	0.794
Fountain-P11		2.225M	11.154M	102.284	6.013	2.510	1.963
Herz-Jesu-P25		4.819M	23.118M	185.006	12.888	5.533	4.096

이 교차하는 모든 삼각형에 대한 거리를 계산하고, 해당 삼각형이 속한 사면체를 추적하여 그래프 가중치를 업데이트한다. 이러한 그래프 가중치는 목표 정점으로부터 멀어질수록 높은 가중치가 부여되며, 결과 표면이 아닌 삼각형에 대해 높은 가중치를 부여한다. 각각의 사면체는 네 개의 삼각형으로 분해되며, 두 사면체가 접한 부분에 중복된 삼각형이 생성될 수 있다. 따라서 광선 순회과정에 후면 컬링(backface culling)을 적용하여 중복 연산을 제거하였다.

3. 실험 및 결과



(a) OpenMVS (CPU) (b) 제안하는 알고리즘 (RT-코어)

그림 3: OpenMVS와 RT-기반 방법을 이용하여 각각 가시성 가중치를 계산하고, 표면을 재구성한 결과(Fountain-P11).

제안하는 알고리즘의 성능을 평가하기 위해 Strecha 등이 제안한 비전 데이터 세트 [5]를 사용하였다. OpenMVG[6]와 OpenMVS[4]를 통해 이미지로부터 밀집된 점 구름을 구성하였다. 제안하는 알고리즘은 서로 다른 세 종류의 RTX 플랫폼에서 CUDA 12.3와 OptiX 7.4.0을 사용하여 구현되었으며, OpenMVS가 제공하는 CPU 기반 가시성 검사 알고리즘과 성능을 비교하였다.

표 1은 제안하는 알고리즘과 CPU 알고리즘의 성능 비교한다. 제안하는 알고리즘은 RTX 2080에서 CPU 대비 최대 20.78배 성능 향상을 얻었으며. RTX 3080에서 최대 40.75배, RTX 4080에서 최대 52.11배 성능 향상을 달성했다. 그림 3은 CPU 알고리즘과 제안하는 알고리즘의 표면을 재구성 결과를 보여주며, 두 알고리즘이 동등한 품질의 결과를 생성함을 볼 수 있다.

이러한 결과는 제안하는 RT-코어 기반 가시성 가중치 계산 알고리즘의 우수성을 보여주며, RT-코어를 3차원 재구성 알고리즘에 사용하는 것에 대한 가능성을 보여준다.

참고문헌

[1] H.Pan, T.Guan, K.Luo, and J.Yu, A visibility-based surface reconstruction method on the gpu, *Computer Aided Geometrice Design*, 84:101956, 2021

[2] P.Labatut, J.-P.Pons, and R.Keriven Robust and efficient surface reconstruction from range data, *Computer graphcis forum*, 28(8):2275-2290, 2009

[3] J.Zhang, Z.Wan, and J.Liao, Adaptive joint optimization for 3d reconstruction with differentiable rendering, *IEEE Transactions on Visualization and Computer Grpahics*, 29(6): 3039-3051, 2022

[4] D.Cernea, OpenMVS: Multi-view stereo reconstruction library, *[Online] Available:https://cdcseacave.github.io/openMVS*, 2020

[5] C.Strecha, W.Von Hansen, L. Van Gool, P.Gua, and U. Thoennessen, On benchmarking camera calibration and multi-view stereo for high resoultion imagery, *in 2008 IEEE conference on computer vision and pattern recognition*, pp. 1-8, 2008

[6] P.Moulon, P.Monasse, R.Perrot, and R.Marlet, OpenMVG: Open multiple view geometry, *in International Workshop on Reproducible Research in Pattern Recognition*. pp. 60-67, 2016

Image-to-3D 기술을 활용한 반려동물 아바타 생성*

황수미[0], 최준서, 송오영[1]
세종대학교 컴퓨터공학과, 컴퓨터공학과, 소프트웨어학과
hwangsm614@naver.com, wnstj0126@naver.com, oysong@sejong.ac.kr

Generating Pet Avatars using Image-to-3D Techniques

Su-mi Hwang[0], Jun-seo Choi, Oh-yong Song[1]
Dept. of Computer Science, Dept. of Computer Science, Dept. of Software, Sejong University

요약

본 연구는 메타버스 환경에서 반려동물을 3D 아바타로 구현해 가상공간에서의 생성을 목표로 한다. 단일 이미지의 Matting 전처리 후, Meshy API를 이용해 3D 모델을 생성하고, Anything World API를 통해 Rigging 및 애니메이션을 적용한다. 최종적으로 게임 엔진 등에서 애니메이션 및 조작이 가능한 아바타로 만들어 사용자들이 다양한 활동을 즐길 수 있게 한다. 이는 반려동물과의 정서적 교감을 디지털 경험으로 확장하며, 메타버스 내 새로운 응용 가능성을 제시한다.

1. 서론

현대의 메타버스 환경에서는 사용자들이 주로 자신을 아바타로 표현해 왔으며, 사용자들은 아바타를 통해 자신의 외모, 성격, 라이프 스타일 등을 나타내기도 한다. 본 연구에서는 이 개념을 확장하여 사용자들이 자기 자신뿐만 아니라 자신의 반려동물을 가상공간에서 함께 즐길 수 있는 경험이 가능하도록 AI를 활용하여 동물을 3D 아바타로 구현하는 연구를 진행하였다. 이는 사용자에게 동물과의 정서적 교감과 풍부한 메타버스 경험을 제공할 것이다.

사람의 경우, 기존 연구와 상용화된 서비스가 많아 아바타를 만드는 데 있어 다양한 선택지가 존재한다. 반면, 동물 아바타에 대한 데이터와 서비스는 상대적으로 미비하지만, 사람에 비해 옷, 악세서리, 소품 등으로 인한 복잡성이 적어 단순한 텍스처를 가지고 있기 때문에 비교적 사실적인 표현이 가능하다. 본 연구에서는 동물의 옆모습 이미지를 입력 받아 Matting 전처리 후 Image-to-3D 변환을 거쳐 3D 모델로 생성한다. 이 후 생성된 모델에 뼈대를 붙이는 Rigging 과정을 거치고, 최종적으로 애니메이션 및 조작이 가능한 아바타를 생성하여 게임 엔진 등에 적용한다.

[1]교신저자
* 포스터발표논문, 학부생 주저자 논문임
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업(IITP-2024-RS-2022-00156354), 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원)사업(IITP-2024-RS-2023-00254529), 2024년 과학기술정통신부 및 정보통신기획평가원의 SW중심대학사업(2024-0-00037)의 연구결과로 수행되었음

2. 본론

본 연구는 3D 아바타 생성을 위한 모델링 및 텍스처링 기술이 없는 비전문가도 사진 한 장으로 3D 모델을 생성할 수 있도록 AI 기술을 활용하며, [그림 1]과 같은 파이프라인으로 3D 아바타를 생성한다.

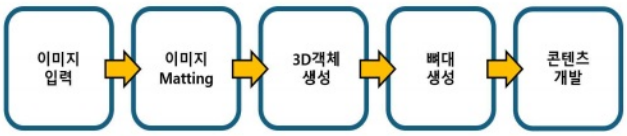


그림 1 : 전체 파이프라인

2.1. Matting 전처리

Image-to-3D란 2D 이미지에서 대상 객체를 3D 모델로 재구성하는 작업으로, 최근 컴퓨터 비전 및 3D 모델링 분야에서 주목받고 있다[1]. 특히, 복수가 아닌 단일 이미지에 대한 3D 모델 생성 연구가 활발히 진행되고 있다[2, 3]. 하지만 단일 이미지만을 사용한 3D 모델 생성은 배경과 대상 객체의 구분이 명확한 이미지인 경우 3D 모델 생성의 정확도가 높은 반면, 배경과 대상 객체 사이의 구분이 명확하지 않거나 배경이 복잡한 경우 정확도가 낮아진다는 한계를 가진다. 이러한 문제를 해결하기 위해서는 Image-to-3D 수행 전 배경과 대상 객체의 정교한 분리를 위한 Matting 전처리가 필수적이다. 이를 위해 본 연구에서는 널리 알려진 딥러닝 기반 파이썬 라이브러리 rembg와 딥러닝 기반의 Paddleseg의 Matting 기술을 이용하여 전처리 테스트 및 비교를 수행하였다[4, 5]. rembg는 U2-Net기반 구조[4]를 가지고, Paddleseg는 U-Net기반에 Attention 기능 등을 추가한 구조를 가진다[5]. Matting 결과 이미지를 픽셀 간 알파 매트값 평균의 차이를 나타내는 MSE, MAE와 구조적 유사성을 나타내는 SSIM 등의 지표로 비교 분석한 결과, [표 1]과 같이 Paddleseg의 성능이 2~5배 정도 더 우수한 것으로 나타났다. MSE와 MAE는 낮을수록, SSIM은 1에 가까울수록 성능이 우수하다. 또한 [그림 2]를 육안으로 확인했을 때, rembg는 객체의 외곽에 배경색이 많이 남아 있는 반면, Paddleseg는 원본의 배경색을 거의 유추할 수 없을 정도로 개선된 것을 볼 수 있다. 따라서, 더 정확한 결과를 산출하는 Paddle-seg의 Matting 기술을 전처리 단계에 사용하였다.

지표 / DNN모델	MSE	MAE	SSIM
rembg	0.739	0.758	0.213
PaddleSeg	0.159	0.328	0.831

표 1 : 정량적 분석 결과

그림 2 : rembg(좌), Paddleseg(우) Matting 결과

2.2. 단일 이미지 기반 3D 모델 생성

텍스처가 복잡한 경우 Image-to-3D 변환 모델[3]의 결과는 품질이 현저히 낮았다. 따라서 대량의 데이터셋으로 충분히 학습시켜 더욱 견고하고 일반화된 모델이 필요했고, 이를 만족하는 단일 이미지로부터 고품질의 3D 모델을 빠르게 생성할 수 있는 도구인 Meshy API를 전처리로 사용하였다.

3D 모델 생성 결과, [그림 3]과 같이 Matting 전처리를 적용하지 않은 이미지로 생성한 모델은 배경 색상이 모델 가장자리 부분에 포함되거나 배경과 객체를 정확히 구분하지 못하여 꼬리가 2개로 나타나는 등의 한계가 존재했다. 하지만 Matting 전처리를 적용한 후 이러한 한계가 눈에 띄게 개선되는 것을 확인할 수 있었다.

그림 3 : Matting 전처리 적용 전(좌), 적용 후(우)

2.3. 리깅 및 애니메이션 생성

3D 모델에 애니메이션을 적용하기 위해서 skeleton이라는 객체의 뼈대를 생성해야 하며, 이를 Rigging이라고 한다. 3D 모델 생성과 같이 Rigging 또한 자동화하기 위해 딥러닝 모델인 Anything World API를 선택하였다. Anything World API는 Image-to-3D 변환으로 생성된 3D 모델을 입력하면 자동으로 Rigging을 적용하여 skeleton을 생성한다. 또한, 다양한 애니메이션 시퀀스를 제공하므로 별도의 추가적인 작업 없이 아바타의 동작을 생성할 수 있다.

3. 평가 및 한계

단일 이미지를 통한 3D 모델의 생성을 위해 3D 모델 변환 전 입력 이미지에 Matting 전처리를 적용함으로써 보다 정확한 3D 모델 생성이 가능함을 확인하였으나, 여전히 몇 가지 제약 사항이 존재한다. 동물이 정확하게 정자세로 서 있는 옆모습에 대한 이미지가 아니라면, 리깅 과정에서 동물의 머리 부분을 앞다리로 잘못 인식하는 등의 오류가 발생하며 비현실적인 애니메이션이 적용되는 것으로 이어진다. 또한, 동물 아바타가 걷거나 달리는 모션을 할 때 진행 방향을 정면으로 바라보지 못하고 옆만 바라보고 달리는 한계가 존재한다. 3D 모델이 정자세로 구현되었어도 skeleton의 관절 좌표가 정확한 위치에 생성되지 않으면 시각적으로 불안정하거나 자연스럽지 않은 움직임이 초래된다. 향후 연구에서는 skeleton 구조의 정밀도를 개선하고, 보다 정확한 Rigging을 통해 모델의 움직임을 개선할 필요가 있다.

4. 결론

본 연구는 Image-to-3D 기반 동물 아바타 생성을 위해 반려동물의 옆모습 사진을 입력받아 Matting 전처리를 통해 배경과 객체를 정확히 분리하고, Meshy API를 활용하여 고품질의 3D 모델을 생성한다. 이어 Anything World API를 통해 Rigging 및 애니메이션을 적용함으로써 동물 아바타가 실제 동물처럼 자연스러운 움직임을 구현할 수 있도록 한다. 최종적으로 게임 엔진 등에서 애니메이션 및 조작이 가능한 형태로 구현하여, 사용자들이 자신의 반려동물 아바타와 다양한 활동을 즐길 수 있다.

본 연구는 반려동물과의 정서적 교감을 중요시하는 현대인들에게 메타버스 환경에서의 새로운 형태의 디지털 교류를 제공함으로써, 한층 더 풍부한 사용자 경험을 선사한다. 향후 Matting 기술과 3D 모델링 기술의 고도화를 통해 다양한 반려동물의 특성을 보다 사실적으로 반영하는 아바타를 생성할 수 있을 것이며, 본 연구에서 동물 아바타 생성과 활용에 대한 새로운 가능성을 제시함으로써, 향후 관련 기술 발전과 응용 분야 확대에 기여할 수 있을 것으로 전망된다.

참고문헌

[1] LI, Xiaoyu, et al. Advances in 3D Generation: A Survey. arXiv preprint arXiv:2401.17807, 2024.

[2] LONG, Xiaoxiao, et al. Wonder3d: Single image to 3d using cross-domain diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024. p. 9970-9980.

[3] LIU, Minghua, et al. One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization. Advances in Neural Information Processing Systems, 2024, 36.

[4] QIN, Xuebin, et al. U2-Net: Going deeper with nested U-structure for salient object detection. Pattern recognition, 2020, 106: 107404.

[5] LIU, Yi, et al. Paddleseg: A high-efficient development toolkit for image segmentation. arXiv preprint arXiv:2101.06175, 2021.

실 수업내 협력 학습 연구를 위한 카메라 기반 모션 데이터 추출 방법론에 관한 예비 연구

조준형 [0,1], 조영욱 [2], 배정은 [2], 신윤희 [3] 김광욱*[,2]
한양대학교 헬스케어디지털공학과 [1],한양대학교 컴퓨터·소프트웨어학과 [2], 한양대학교 교육공학과 [3]
{anywhere13, ooik9823, je_0922}@naver.com, {kenny, yoonheeshin}@hanyang.ac.kr

A Pilot Study on Camera-based Motion Data Extraction for Real Classroom Collaborative Learning

Junhyeong Cho[0,1], Youngwug Cho[2], JeongEun Bae[2], Yoonhee Shin[3] Kwanguk (Kenny) Kim*[,2],
Dept. of HealthCare Digital Engineering, Hanyang University[1], Dept. of Computer Software, Hanyang University[2], Dept. of Education Engineering, Hanyang University[3]

요약

협력 학습이란 주어진 목표를 그룹원들과 협력하여 학습을 증진시키는 교육 방식이다. 관련 선행 연구는 설문지 연구를 통해 협력 학습이 중요하고 긍정적인 영향을 준다는 사실에 대해서만 알려졌을 뿐, 객관적인 데이터를 통한 협력 학습 연구는 매우 제한되었다. 본 연구에선 실제 수업 환경에서 협력 학습을 진행할 때 발생하는 학습자의 몸 움직임, 얼굴 회전량, 표정 변화 데이터 값을 카메라 기반으로 추출할 수 있는 방법을 제안하고자 한다. 설문지를 통해 측정된 고/저 성과 그룹 차이를 정량적으로 측정하였다. 이를 통해 잠재적으로 몸 움직임, 얼굴 회전량, 표정 변화 데이터가 협력 학습 연구에 활용될 수 있는지 예비 연구하고자 한다.

1. 서론

협력 학습이란 학습자들이 함께 문제를 해결하거나 지식을 구축하는 과정으로[1], 높은 수준의 사고력, 말하기 능력, 학생과 교수 간 상호작용 증가, 학생들에게 책임감과 관련 있는 것으로 알려졌다. 설문지 연구에 따르면 상호작용이 협력학습의 성공적 요인으로 보았고, 그룹 내 공동 목표를 달성하기 위해 그룹원의 행동과 감정을 조절하는 과정인 Socially-Shared Regulation (SSR)를 강조했다[2]. SSR 이 협력 학습에 미치는 영향을 확인하기 위한 방법으로 그룹원들의 감정과 상호작용을 파악하기 위한 self-report questionnaire, 그룹의 성과 및 협력학습 결과를 평가하기 위한 observation-report 가 사용되었다[1]. 그룹 응집력을 측정하는 Group Environment Questionnaire (GEQ) [3] 설문지, Observation-report 방법에는 대표적으로 Creative Problem Solving (CPS) 모델[4] 기반 평가지 등이 사용되었으며, 이를 통해 그룹원의 상호작용 및 감정과 그룹 성과 간의 연관성을 확인하였다.

최근 연구에선 컴퓨터 비전 기술을 활용하여 수업 영상에서 얼굴 감정을 추출함으로써 주관적 보고에 의존하지 않는 정서 측정 방법론이 가능함을 제안하였다.[1] 이를 기초로 실제 수업에 적용되기 어렵고 객관적인 분석이 힘든 기존 self-report questionnaire 방식을 보완할 수 있는, 카메라 기반 모션 데이터 추출 방법론을 제안하고자 한다. 실제 교실 내에서 카메라에서 얻을 수 있는 몸 움직임, 얼굴 회전량, 표정 변화를 통한 감정 인식등의 데이터를 활용하여 고성과 그룹과 저성과 그룹에서 차이가 있는지 확인하고자 한다. 이를 통해 컴퓨터 비전 기반 측정 방법론이 협력 학습 연구에 활용될 수 있는지 예비 연구하고자 한다.

2. 방법

2.1. 참가자

피험자의 경우, 총 8 명의 학생 (Mean age = 22.3; S.D. = 2.9047; 5 females)의 수업 데이터를 분석하였다. 총 4 주에 걸쳐 실제 한양대학교 수업을 진행하며 생겨나는 데이터를 이용하여 분석하였으며, 각 주차 동안 45 분의 협력 학습을 진행하였다. 협력 학습 결과 고/저 성과자 그룹 2 그룹을 선정하였으며, 각 그룹은 4 명으로 이루어져 서로 마주하여 위치하였다. (그림 1)

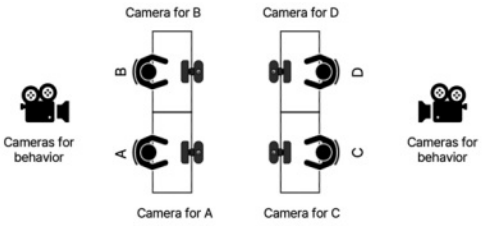


그림 1: 수업 환경 세팅 구조도

2.2. 시스템 구성

각 그룹원의 노트북 앞에 웹 캠 및 zoom(Zoom Video Communication, 미국) 녹화 프로그램을 사용하여 수업 중에 발생하는 얼굴 움직임을 기록하였다.

* 포스터 발표논문
* 본 논문은 요약논문 (Extended Abstract)으로서, 본 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2024-00355411).
* Correspondence to K. Kim (kenny@hanyang.ac.kr)

표 1 컴퓨터 비전 결과(A,B 그룹)
*몸 움직임 : 프레임 당 전체 몸 움직임량, 얼굴 회전 량 : 프레임 당 전체 얼굴 회전량

	고성과자(A 그룹): CPS 점수 44.7				저성과자(B 그룹): CPS 점수 35.3			
	1 주차	2 주차	3 주차	4 주차	1 주차	2 주차	3 주차	4 주차
앞으로 기울인 횟수	452 회	544 회	267 회	332 회	191 회	167 회	646 회	172 회
감정 표현 비율	90%	88%	85%	89%	82%	26%	45%	27%
긍정 감정 비율	5.6%	4.5%	10.6%	10.1%	7.3%	7.7%	8.9%	3.7%
총 몸 움직임 량	534.3 cm/s	470.7 cm/s	679.8 cm/s	593.7 cm/s	375.3 cm/s	384.3 cm/s	450.9 cm/s	339.3 cm/s
총 얼굴 회전 량	55.0 °/s	39.0 °/s	37.75 °/s	29.0 °/s	25.25 °/s	22.5 °/s	29.0 °/s	25.0 °/s

수업 중 발생하는 그룹원의 몸 움직임을 분석하기 위한 카메라의 경우 각 그룹원들 뒤에 설치하여(1080p, 29.97fps) 기록하였다.

2.3. 실험 절차

실험의 경우 총 4 주에 걸쳐 진행되었으며, 모든 주차가 종료된 후에는 교수자가 Creative Problem Solving (CPS) 기반 성과 측정 설문지[4]를 통해 각 그룹원의 성과를 평가하였다.

2.4. 컴퓨터 비전 기반 측정 데이터

MediaPipe 의 Face Detection 을 사용하여 매 프레임별 얼굴 주위에 Bounding Box 를 생성하였다. 생성된 Box 의 너비와 높이 값을 활용해 Size 를 계산하였고, 매 프레임 간 Size 값의 변화를 계산하여 이전 프레임 대비 5% 이상 Box Size 가 커질 시, 앞으로 기울인 횟수로 측정했다. 그리고 협력 학습 내 얼마나 상호작용이 일어났는지 확인하고자, 4 명 중 3 명 이상이 동시에 앞으로 얼굴 기울임이 발생한 횟수를 Time-stamp 데이터를 활용하여 계산하였다. 감정 측정의 경우, 웹 캠 영상에서 Haarcascade 기반 딥러닝 모델을 통해 얼굴을 탐지하였고, 탐지된 얼굴에서 Xception 기반 딥러닝 모델[5]을 통해 3 가지의 감정(긍정, 부정, 중립)을 추출하고, 감정의 유무, 긍/부정에 따라 분류하였다(표 1). 몸 움직임의 경우, MediaPipe 의 body pose landmark 를 사용하여 매 프레임별 몸 관절들 좌표 값을 추출하였다(그림 2). Occlusion 이 발생한 프레임을 제외하고, 매 프레임마다 모든 관절 좌표의 변위를 총합하고, 이를 시간(프레임)에 대해 평균 내어 전체 움직임 값을 계산하였다. 몸 얼굴 회전의 경우, 6DRepNet 모델을 활용하여 웹 캠 영상에서 추출한 얼굴 회전(Yaw, Pitch, Roll)값을 합한 뒤 시간에 대해 평균 내어 얼굴 회전 량을 계산하였다.

그림 2 MediaPipe 몸 포즈 데이터 추출

2.5. 평가 요소

2.5.1. Observation-report 기반 그룹 성과

Creative Problem Solving (CPS): 창의적 문제 해결 능력 수준을 평가하기 위한 방식으로 Fact Finding, Problem Finding, Idea Finding, Solution Finding 에 대해 관찰자가 평가를 진행한다.

3. 결과

3.1. 그룹 성과 평가 결과

CPS 총점 계산결과, A 그룹은 44.7점으로 고성과자로, B 그룹은 35.3점으로 저성과자로 교수자에 의해 지정되었다.

3.2. 컴퓨터 비전 기반 측정 결과

앞으로 기울인 횟수, 감정 표현 비율, 긍정 감정 비율, 몸 움직임 량, 얼굴 회전 량에서 모두 고성과자 그룹이 저성과자 그룹에 비해 높게 나타났다. 부정 감정 비율은 고성과자 그룹이 저성과자 그룹에 비해 높게 나타났다.

4. 논의

본 논문은 실제 수업에서 협력 학습을 객관적으로 측정하기 위해 카메라 기반 모션 데이터 추출 방법론을 예비 연구하였다. 예비 연구 결과 고/저 성과자 사이에서 컴퓨터 비전 기반 측정 결과가 분명한 차이가 있음을 확인하였다. 이를 통해 실 교실 내에서 협력 학습 연구를 위해 카메라 기반 모션 데이터 추출 방법론이 활용될 수 있음을 확인하였고, 향후 대단위의 본 연구를 통해 관련 연구를 진행할 예정이다.

참고문헌

[1] Li, Yante, et al. "Exploring Interactions and Regulations in Collaborative Learning: An Interdisciplinary Multimodal Dataset." arXiv preprint arXiv:2210.05419 (2022).
[2] Allyson Hadwin, Sanna Järvelä, and Mariel Miller. 2018. Self-regulation, co-regulation, and shared regulation in collaborative learning environments. (2018). *psychology* 36.1 (2011):
[3] Whitton, Sarah M., and Richard B. Fletcher. "The Group Environment Questionnaire: A multilevel confirmatory factor analysis." Small Groupesearch 45.1 (2014): 68-88.
[4] Treffinger, D. J. (1995). Creative problem solving: Overview and educational implications. Educational psychology review, 7, 301-312.
[5] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

CLIP의 Value Embedding을 이용한 개방 어휘 환경 3차원 세분화*

김상민[01], 이준호[1], 이용현[2], 김영민[1]
서울대학교[1], 고등과학원[2]
{tkdals9082, twjhlee, youngmin.kim}@snu.ac.kr[1], ylee@kias.re.kr[2]

3D Open-Vocabulary Segmentation using Value Embedding of CLIP

Sang Min Kim[01], Junho Lee[1], Yonghyeon Lee[2], Young Min Kim[1]
Seoul National University[1], Korea Institute for Advanced Study[2]

요약

3차원 공간의 의미를 이해하는 것은 컴퓨터 그래픽스와 로보틱스 분야에서 다양한 작업을 수행하는 데에 필수적이다. 최근의 방법론들은 CLIP과 NeRF를 결합하여 의미론적 feature field를 만들어 낼 수 있음을 보였다. 이때, NeRF에서 사용하는 렌더링 손실 함수를 CLIP에 적용하기 위해서는, 조밀한 feature map을 뽑아내는 방법이 필요하다. 본 연구에서는 CLIP의 마지막 attention layer에서 추출한 value embedding이 기존 방법에 비해 해당 위치의 전역적인 의미를 잘 유지함을 보인다. 또한, 이러한 임베딩을 강화하기 위해 다른 비전 파운데이션 모델 (VFM)들의 정보를 통합하는 잠재 공간을 제안한다. Mutual Information Neural Estimation (MINE)를 활용하여 잠재 공간이 VFM들의 유용한 정보를 보존하도록 했다. 마지막으로, 향상된 CLIP feature를 이용해 3차원 feature field를 재구성하고, 해당 field가 전역적인 정보를 잘 유지하고 있음을 보였다.

1. 서론

CLIP[1]을 NeRF[2]와 결합하여 만든 feature field를 통해 우리는 3차원의 의미를 텍스트로 해석할 수 있다. 이러한 feature field를 만들기 위해서는 조밀한 feature map이 필요하다. 기존 연구에서는 이를 위해 이미지를 여러 개의 작은 이미지 영역으로 나누어 CLIP에 통과시키거나, 마스크 세분화 모듈 (SAM[3])을 이용해 잘라낸 마스크를 CLIP에 통과시켰다. 하지만, 그림 1에서 볼 수 있듯, 이렇게 잘라낸 이미지는 주변 환경에 대한 정보를 잃게 되어 그 의미를 제대로 추정하기 어렵다. 본 논문에서는 CLIP의 Value Embedding을 사용하는 것이 기존 방법에 대한 좋은 대안이 될 수 있음을 보인다. 추가적으로, 다른 비전 파운데이션 모델 (VFM)과 상호 정보량을 높이는 것으로 CLIP의 Value Embedding을 향상시킬 수 있음을 보인다.

* 포스터 발표논문

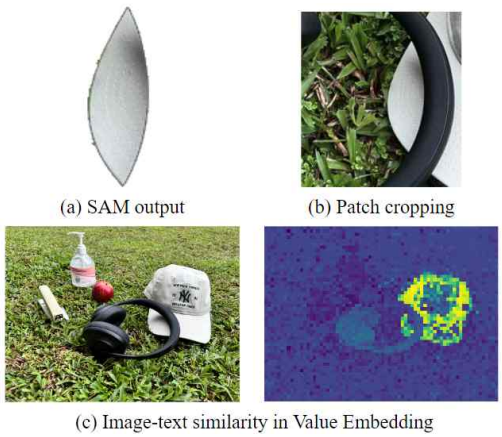


그림 1: 조밀한 Feature map을 뽑기 위한 기존 방법 (a), (b)과 Value Embedding (c)의 차이. 기존 방법론은 주변 환경에 대한 정보가 없어 의미를 정확히 파악하기 어렵다. Value embedding은 충분히 작은 patch에 대한 의미를 추출하며 전역적인 정보를 종합한다. "모자"라는 입력이 주어졌을 때, 임베딩과 텍스트와의 유사도를 (c)의 오른쪽처럼 시각화하면, 헤드폰에 의해 나누어진 모자의 작은 부분도 모자로 인식한다.

2. 방법

2.1. 상호 정보량 최대화

상호 정보량 최대화 파이프라인은 그림 2와 같다. 실제 정보의 교환은 잠재 공간 (Latent Image)에서 이루어진다. 해당 잠재 공간은 Coordinate MLP (f_{xy})를 이용해 구현했다. 임의의 2D 픽셀 좌표에 대해 잠재 벡터를 얻고 해당 벡터가 정보 제공자와 상호 정보량이 높아지게 최적화를 진행했다. 또한, MLP로 구현된 decoder (f_{CLIP})를 추가로 두어 해당 잠재 벡터가 CLIP의 feature 공간으로 decoding 될 수 있도록 했다.

상호 정보량의 계산은 [4]의 논문을 따랐다. 상호 정보량을 높이는 대상이 되는 VFM인 정보 제공자는 꼭 VFM일 필요 없이 수행하고자 하는 작업에 따라 설정할 수 있다. 본 연구에서는 feature field가 다음과 같은 두 가지 직관을 만족하도록 VFM들을 선택했다. 하나의 객체는 비슷한 의미를 가진다 (DINO), 공간 상에서 비슷한 위치에 있으면 비슷한 의미를 가진다 (MiDaS).

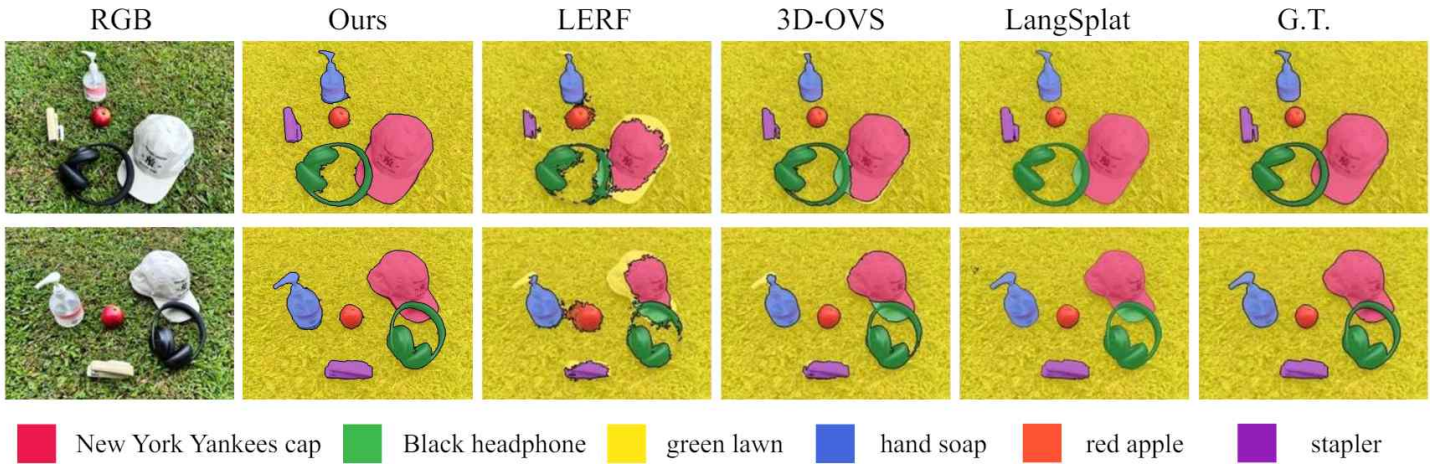


그림 3. 개방 어휘 환경에서 3차원 세분화 결과 시각화. 세분화를 위해 사용한 개방 어휘와 대응되는 마스크의 색이 표시되어 있다. 본 연구에서 제안한 방법이 헤드폰에 의해 잘린 모자 부분을 모자로 잘 세분화하는 것을 확인할 수 있다.

최적화를 위해 두가지 손실함수를 정의하였다. L_{MI}는 [4]를 따라 정보 제공자와 Latent Image의 상호 정보량을 높인다. L_{CLIP}는 Latent Image가 CLIP 공간으로 복호화되도록 유지하며, MSE로 구현되었다.

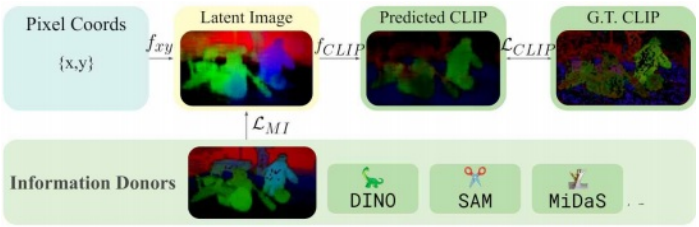


그림 2. 상호 정보량 최대화를 통한 CLIP 향상 파이프라인.

2.2. 3차원 Feature Field 재건

촬영한 위치를 알고 있는 이미지의 집합이 주어졌을 때, NeRF[2]를 이용해 3차원 Field를 재건할 수 있다. 본 연구에서는 상호 정보량 최대화를 통해 향상된 2차원 CLIP feature를 이미지처럼 사용하여 3차원 Feature Field를 재건한다. 자세한 구현은 [9]를 따랐다. [9]에서는 Field가 3차원 색상과 해당 좌표의 밀도를 출력하는데, 이에 더해 CLIP과 동일한 차원의 feature vector를 출력하도록 했다. 해당 feature vector와 입력으로 주어진 2차원 CLIP feature가 같아지도록 MSE 손실함수를 최소화하여 학습을 진행했다.

3. 결과

3D-OVS[7] dataset의 *lawn*에 대해 실험을 진행했다. 학습은 1대의 RTX 4090에서 총 4000 iter 만큼 진행했고 약 2분이 소요되었다. 3차원 feature field의 재건이 잘 진행되었음을 확인하기 위해 해당 feature field를 특정 시점에서 렌더링하였다. 렌더링된 2차원 feature에 대해 텍스트와 유사도를 판별하여 이미지를 주어진 텍스트와 가장 유사한 집합으로 세분화하였다.

개방 어휘 환경 3차원 세분화 결과는 그림 3과 같다. 다른 방법론의 결과 그림은 LangSplat[8]에서 가져왔다. 본 연구에서 제안한 방법론 (Ours)만이 모자의 작게 잘린 부분을 모자로 정확하게 세분화하는 것을 확인할 수 있다.

참고문헌

[1] Radford, Alec, et al. Learning transferable visual models from natural language supervision. *International conference on machine learning. PMLR*, 2021.

[2] Mildenhall, Ben, et al. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* 65.1 (2021): 99-106.

[3] Kirillov, Alexander, et al. Segment anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

[4] Belghazi, Mohamed Ishmael, et al. Mutual information neural estimation. *International conference on machine learning. PMLR*, 2018.

[5] Caron, Mathilde, et al. Emerging properties in self-supervised vision transformers. *Proceedings of the IEEE/CVF international conference on computer vision.* 2021.

[6] Ranftl, René, et al. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence* 44.3 (2020): 1623-1637.

[7] Liu, Kunhao, et al. Weakly supervised 3d open-vocabulary segmentation. *Advances in Neural Information Processing Systems* 36 (2023): 53433-53456.

[8] Qin, Minghan, et al. Langsplat: 3d language gaussian splatting. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2024.

[9] Tancik, Matthew, et al. "Nerfstudio: A modular framework for neural radiance field development." ACM SIGGRAPH 2023 Conference Proceedings. 2023.

부정적 감정 해소를 위한 야외활동 유도 게임 애플리케이션 개발 연구*

전아름 [1], 권민성 [2]
고려대학교 창의융합디자인협동과정 박사 과정, 고려대학교 산업디자인학과 교수
dkfma1233@korea.ac.kr, kwonminsung@korea.ac.kr

Development of a Game Application to Encourage Outdoor Activities for the Alleviation of Negative Emotions

Jeon A Rum[1], Kwon Min Sung[2]
Korea university Creative Convergence Design course, Korea university Department of Industrial Design

요약

집 안에서 생활하는 사람들이 늘어나면서 부족한 신체활동으로 인해 우울, 무기력, 권태 등 부정적인 감정을 겪는 사람들도 늘어났다. 본 논문에서는 신체활동이 부족한 사람들이 겪는 부정적인 감정을 해소하기 위한 야외 활동을 유도하는 모바일 애플리케이션을 제안하였다. 이를 구현하기 위해 지속적으로 하루에 한 개씩 미션을 준다는 컨셉을 중축으로 진행되었으며, 몰입감을 향상시키기 위해 비밀 조직의 요원이라는 스토리 텔링을 도입하였다. 또한, 지속적으로 미션을 완수를 유도하는 세계관 설정을 통해 컨셉 스케치 구상과 기능 지원 방식을 분석하였고, Smart maker 시스템을 활용하여 애플리케이션의 디자인 청사진을 탐구하였다.

1. 서론

빠른 배달 시스템과 다양한 콘텐츠가 있는 OTT 서비스로 집 안에서 모든 것을 해결하는 홈 루덴스(Home ludens) 족, 홈 족인 사람들이 늘어났다. 집 밖을 나가지 않아도 모든 것을 집 안에서 해결할 수 있는 장점이 생겼지만 집 안에서 생활하다 보니 신체활동의 부족으로 우울, 무기력, 권태 등 부정적인 감정을 겪게 되는 단점도 생겨났다. 신체활동은 모든 연령층에서 우울과 불안 등의 심리상태와 높은 상관관계[1]가 있기 때문에 발생하는 문제점으로 부정적인 감정 해소를 위해 집 밖에서의 야외 활동을 통한 신체활동을 높여줘야 한다.
하여 본 논문에서는 집 안에서 우울, 무기력, 권태 등 부정적인 감정을 겪는 사람들을 집 밖에서 야외 활동을 통해 부정적인 감정에서 벗어날 수 있게 지원하는 모바일 애플리케이션을 제시하고자 한다.

2. 앱 게임 개발을 위한 컨셉 스케치

2.1. 사용자 대상 선정

- 야외 활동 시 쉽고 편리하게 사용할 수 있도록 모바일 어플리케이션(이하 앱)으로 게임 개발을 진행하였다. 사용자는 앱에서 제공하는 미션을 하루에 한 개만 수행하게끔 제안하기 때문에 부담감이 적고, 매일 미션이 주어지므로 야외 활동의 동기 부여를 제공해준다.
본 연구의 앱 사용 대상자는 혼자 생활하고 있는 20대 후반에서 30대 중반의 사람이며, 게임 및 OTT시청을 취미로 하는 사람으로 진행하였다. 앱 사용 대상자 인터뷰를 진행하기 전에 페르소나 프로필을 작성하여 인터뷰 대상자를 분석하였다(그림 1 참조).

이름	유지희(가명)
나이/성별	26세/여성
MBTI	ISFP
가족관계	부모님, 자기 자신(외동)
거주지	인천 출생. 부모님과 떨어져 혼자 원룸 거주
학력/직업	대학교를 졸업하고 취업을 하지 못해 알바를 하는중
취미/여가활동	디즈니 플러스로 마블 시리즈 정주행 하기, 보드게임하기, 일기쓰기, 카페가기
특징	다양한 색으로 머리를 염색 하는 걸 좋아함
목표	디즈니 랜드 놀러가서 마블 캐릭터와 사진찍기
불만사항	돈도 없고 재미도 없고 설렘도 없고 낭만도 없다.

그림 1: 페르소나 프로필

2.2. 몰입감을 위한 세계관 설정 및 스토리 텔링 구상

- 게임의 몰입감은 게임의 참여도와 관련되어[2] 있기 때문에 체계적인 세계관으로 몰입감을 높여 사용자의 참여를 유도하였다. 하루에 한 개의 미션을 받는다는 컨셉에 맞춰 *비밀 조직 일원으로 활동하며 어둠의 조직을 소탕하기 위해 미션을 완수 한다*라는 스토리 텔링을 구상하여 세계관을 설정하였다. 직업은 사용자의 선택권을 높이기 위해 4개의 직업으로 해커, 연구원, 요원, 특수부대로 설정하였고, 각 직업에 따른 상징색을 설정해 색상이 주는 의미와 연결하였다. 또한 앱 사용자의 행동에 정당성을 부여하기 위해 최종 레벨 단계에서 어둠의 범죄 조직 소탕 미션을 달성하면 앱 내 명예의 전당에 닉네임을 새길 수 있게 하였다. 미션 내용은 세계관 내용과 관련되었으며, 해당 미션을 완료할 때마다 게임 머니 및 미션을 완료한 모습의 이미지를 받는다. 완료한 미션은 다시 볼 수 있으며, 수행한 후기를 음성으로 기록할 수 있어 사용자가 세계관안에서 감정이입이 될 수 있게 유도한다. 미션을 완료하고 받은 게임 머니를 통해 상점에서 각 직업과 관련된 게임 아이템을 구매할 수 있고 정해진 게임 아이템을 모두 수집하면 레벨이 올라가는 방식을 지원하여 사용자의 참여도를 높였다(표 1 참조).

비밀 조직 이름: 바크로	
게임 배경	비밀 조직 바크로는 어둠의 범죄 조직인 리시키를 소탕하기 위해 각 직업에 따라 미션을 수행한다.

주저자 [1] 교신저자 [2]
* 포스터 발표 논문
* 본 연구는 2024년도 산업통상자원부의 재원으로 “한국산업기술진흥원” 지원을 받아 “신기술분야융합디자인 전문인력양성사업” 의 일환으로 수행되었다. (P00012725)

직업 종류	해커	연구원	요원	특수부대
상징 색상	초록색	파란색	검은색	노란색
미션 수행 장소	유희 위주: 노래방, 영화관, PC 방, 놀이터	교육 위주: 도서관, 미술관, 박물관, 서점	일반 위주: 카페, 편의점, 음식점, 지하철	운동 위주: 운동장, 산, 공원, 정원
미션 수행 내용 예시	노래방에서 ㅂ,ㄹ,ㅋ가 포함된 노래 찾아 부르기	도서관에서 ISBN 9791162 143605 책 찾기	편의점에서 ㄹ,ㅅ,ㅋ가 들어간 제품 찾아 사진찍기	공원에서 세 잎 클로버를 찾아서 사진찍기
미션 완료 보상	외장하드, 태블릿, 스마트 손목시계	비커, 실린더, 온도계, 스포이트	멀티 툴, 가스 총, 마취 총, 3단 봉	군용 칼, 망원경, 도끼, 수류탄

표 1: 세계관 설정

3. 세계관 속 사용자의 욕구를 유도하는 기능 지원

사용자는 몰입된 세계관 속에서 자신의 욕구를 반영하고자 한다. 하여, 사용자의 욕구를 유도하는 알림, GPS, 카메라, 음성 인식 등 4가지 기능 지원을 통해 사용자가 자신의 욕구를 반영할 수 있게 하였다(표 2 참조).

기능 유형	지원 방식
알림 기능	미션을 하고 싶은 인지적 욕구 유도.
GPS 기능	위치 파악을 통한 안전의 욕구 유도.
카메라 기능	미션 완료 인증을 통한 자존의 욕구 유도.
음성 기능	미션 평가 기록을 통한 소속의 욕구 유도

표 2: 세계관 설정

4. 디자인 스케치 단계

사용자가 앱을 실행하는 단계를 3단계로 분류하여 디자인 스케치를 진행하였다. 첫 번째, 앱 시작 단계. 사용자는 앱을 실행하고 직업을 선택한 뒤 소속증을 발급받는다(그림 2 참조).

그림 2: 앱 시작 단계

두 번째, 미션 수행 단계. 원하는 미션 수행 장소를 선택하고 1개의 미션을 받아서 미션을 수행한다(그림 3 참조).

그림 3: 미션 수행 단계

세 번째, 보상 획득 및 기록 단계. 미션을 완료하고 보상으로 게임 머니와 미션을 완료한 모습의 이미지를 받아 수집하고 완수한 미션에 대한 평가를 기록한다(그림 4 참조).

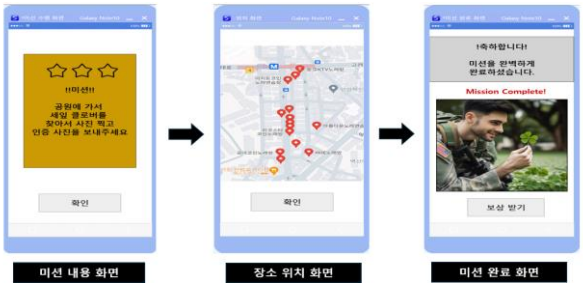


그림 4: 보상 획득 단계

5. 기대 효과 및 결론

본 연구는 부족한 야외 활동으로 인해 발생하는 우울, 무기력, 권태 등의 부정적인 감정을 해소하고, 사용자의 신체 활동을 제안 및 유도하는 앱을 개발하였다.
해당 앱의 기대 효과는 3가지로 안정감, 성취감, 만족감이다. 사용자는 첫 번째로 조직의 일원이 되었다는 소속감을 통해 안정감을 얻을 수 있다. 두 번째로 매일 미션을 완수하면서 성취감과 자신감을 얻을 수 있다[3]. 세 번째로 사용자의 인지적, 안전, 자존, 소속의 욕구를 일으키게 유도하는 기능들을 통해 사용자는 깊은 몰입감을 얻고, 유도된 욕구에 반응하며 만족감과 해방감을 얻을 수 있다. 앱을 통해 긍정적인 감정을 얻게 된 사용자가 지속적으로 야외 활동을 유지하길 기대하는 바이다.
향후 앱 개발에 있어 디자인 스케치를 보완하여 사용자 테스터를 진행할 예정이다. 이러한 개발 연구 과정이 사용자의 부정적인 감정 해소에 기여하기를 예상해 본다.

참고문헌

[1] Trost, S. G., Pate, R. R., Dowda, M., Saunders, R., Ward, D. S., & Felton, G. (1996). Gender differences in physical activity and determinants of physical activity in rural fifth grade children. Journal of School Health April, 66(4), 145-50.
[2] Emily Brown and Paul Cairns, A Grounded Investigation of Game Immersion, CHI 2004 Conference on Human Factors in Computing Systems, pp.1297-1300, 2004.04.
[3] **윤준. 나는 매일 작은 성공을 합니다. 서울: 인플루엔셜. 2021.12.20**

생성형 AI를 활용한 공연예술 씬 디자인 시뮬레이션 도구*

김명은†,1, 박은별†,1, 윤세린†,2, 안재홍1,2
한국과학기술원 메타버스대학원1, 문화기술대학원2
{meongeun, silverstar, serinyoon, ahnjh}@kaist.ac.kr

Simulation Tool for Theatrical Scene Design Using Generative AI

Meongeun Kim†,1, Eunbyul Park†,1, Serin Yoon†,2, Jaehong Ahn1,2
Graduate School of Metaverse1, Graduate School of Culture Technology2
Korea Advanced Institute of Science and Technology

Abstract

In this paper, we propose a simulation tool that uses Generative AI to complete musical scenes based on scenarios, character and prop placement, and user requirements. This tool enables users to simulate stage direction and design with Generative AI, fostering innovative and creative execution. It improves collaboration efficiency between directors and stage designers in the performing arts production process, while also enhancing user creativity.

1. Introduction

Performing arts is expanding artistic expression through the use of various digital technologies during stage production. Traditionally, sound and lighting have led to new creative possibilities in the performing arts[1][2]. Recently, Generative AI technologies have been introduced into stage production[3][4], showing significant potential in enhancing creative processes and necessitating research on its effective application.

This study explores enhancing stage production efficiency using AI as a simulation tool. Stage production relies on collaboration between directors and designers, where directors provide general ambience and symbolic cues that designers translate into concrete designs[5]. For example, blocking (movement), a key aspect of directing[6], may change in stage composition. Generative AI can enhance this process by offering pre-implementation for stage design and blocking. Currently, no tools allow simultaneous visualization of stage design, blocking, and music. By using Generative AI simulation and visualization, this research aims to improve creative idea development and collaboration efficiency between directors and stage designers.

2. Concept

In the performing arts, simulation tools like SketchUp and set design transform visual concepts into actionable plans for the stage. However, these tools often limit free conceptualization and struggle with dynamic coordination between blocking and stage elements, making it challenging to visualize changes effectively.

To address these challenges and enhance both efficiency and creativity, we propose the introduction of Generative AI in the stage design and direction. This technology enables flexible creation and adjustment of stage designs through advanced simulation capabilities.

3. Prototype

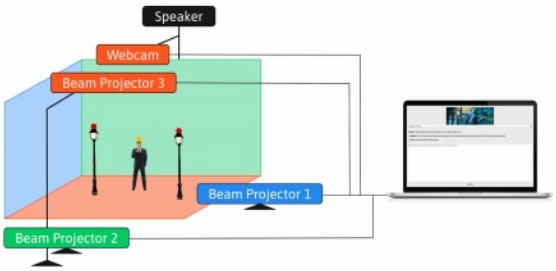


Figure 1: Prototype Structure Diagram

We propose a tool that uses Generative AI to facilitate integrated stage design simulations. Our prototype features character blocking recommendations, background scene generation, and background music generation. As shown in Fig 1, it uses three beam projectors to display images onto the floor and two walls. A webcam tracks the

* 구두(포스터) 발표논문
† 동일하게 기여함

positions of characters and props. The generated music is played through speakers, and the process starts with a button click on the user interface.

The system allows users to place characters and props on an empty stage and input the musical's scenario and requirements. OpenCV detects the positions of characters and props on the floor, while ChatGPT analyzes the provided script to determine how characters should interact with the props and recommends their blocking paths.

Figure 2: Generating process of the scene from the musical "Singing in the Rain"

For instance, as shown in Fig 2, the script of the musical "Singing in the Rain" is input into the tool via the user interface. The user places the character and streetlamps on an empty stage. The output is a completed set that includes the character's blocking path, background images, and background music.

For the blocking path, ChatGPT is provided with the script and the coordinates of the character and props to first understand the mood of the scene and the character's situation. Based on this, it recommends blocking that involves interactions with the props. The response includes explanations based on the coordinates where the character will move. For example, '1. (300, 200): Start from the center of the stage.' Then, these blocking paths are projected onto the floor.

The Generative AI employed state-of-the-art models to create the background and floor images using Stability AI's SDXL model [7], and to generate the music with Meta's MusicGen[8]. We implemented a dual structure of large language models (LLMs) for generation, as the script could not be directly inputted. To interpret the script, we utilized Meta's Llama-2-70b-chat[9], a prominent LLM, to initially extract the mood and atmosphere of the scene along with potential visual elements. This information was then provided as input to each Generative AI to produce the images and music. The images were displayed through the respective beam projectors, and the music was played through the speakers. This setup provides the director with an audio-visual preview of the final theatre set design, enabling real-time adjustments as necessary.

4. Conclusion

Our prototype was successfully demonstrated, allowing simultaneous simulation of stage design, blocking, and music, with the capability to repeatedly simulate various scenes. This enables directors or designers to input scripts and directions, visually confirming various ideas.

Although current prototype does not fully address all aspects of stage design, further development could make it an invaluable tool for performing arts production, reflecting various directing and design intentions to provide a comprehensive simulation and visualization experience.

The first enhancement focuses on fine-tuning LLM blocking optimization, tailoring it to theater directing to enable realistic and detailed stage simulations. The second enhancement involves prompt engineering to reduce extraneous noise in automated content generation, ensuring image and music align with the script. The third enhancement addresses system optimization to overcome performance limitations from integrating multiple APIs for image processing, music synchronization, and motion tracking, thereby reducing processing times and enhancing efficiency. These enhancements will significantly improve the dynamism and executability of the simulation tool combining Generative AI with performing arts.

References

[1] "How Has Technology Changed Theatre? 4 Ways." *TheatreArtLife*. Feb 5, 2018.

[2] 한아림, and 홍일태. "연출을 통해 나타나는 공연무대의 효과에 관한 연구: 국내 공연된 뮤지컬 무대 중심으로." *한국실내디자인학회 논문집*. 21(5): 328-335, 2012.

[3] AI And Theatre: Playwriting, Stage Design, And Ticketing. *Arts Management&Technology Laboratory*. Jun 13.

[4] Practical Artificial Intelligence for Stage Design. *HowlRound*. Oct 06, 2022.

[5] Di Benedetto, Stephen. *An introduction to theatre design*. Routledge, 2013.

[6] Crook, Paul B. *The art and practice of directing for theatre*. Routledge, 2016.

[7] Podell, Dustin, et al. "Sdxl: Improving latent diffusion models for high-resolution image synthesis." *arXiv preprint arXiv*:2307.01952, 2023.

[8] Copet, Jade, et al. "Simple and controllable music generation." *Advances in Neural Information Processing Systems* 36, 2024.

[9] Touvron, Hugo, et al. "Llama 2: Open foundation and fine-tuned chat models." *arXiv preprint arXiv*:2307.09288, 2023.

깊이를 색상으로써 활용한 NeRF 렌더링*

임석현[01], 김민성[2], 이성길[1]
[1]성균관대학교 소프트웨어학과, [2]성균관대학교 실감미디어공학과
{ish990730, leon0106}@g.skku.edu, sungkil@skku.edu

Depth-as-Color NeRF Rendering

SeokHyun Lim[01], Minseong Kim[2], Sungkil Lee[1]
[1]Dept. of Computer Science and Engineering, Sungkyunkwan University
[2]Dept. of Immersive Media Engineering, Sungkyunkwan University

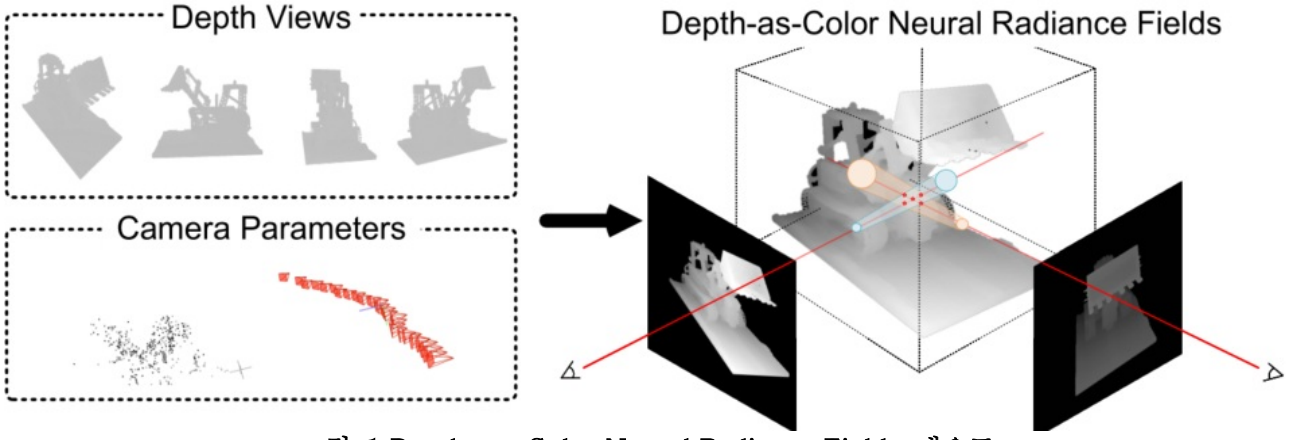


그림 1 Depth-as-Color Neural Radiance Fields 개요도

요약

본 연구는 컴퓨터 그래픽스 분야에서 Neural Radiance Fields(NeRFs)의 새로운 응용을 탐구한다. 기존 NeRF 모델의 RGB 이미지 학습을 통한 3D scene 재구성의 방식에서 벗어나, 깊이 맵을 학습 데이터로 사용하는 새로운 모델을 제안한다. 제안된 모델은 novel view에서의 깊이 맵을 높은 정확도로 추정하는 것을 목표로 하며, 결과적으로 기존 MiDaS나 RGB NeRF 기반의 깊이 추정보다 높은 정확도(PSNR)를 달성하였다.

1. 서론

Neural Radiance Fields(NeRFs) 모델은 3D 장면의 복잡한 구조를 정확하게 재구성하는 강력한 도구로 자리잡았다. 그러나 기존 NeRF 모델은 픽셀의 색상을 예측하는 데에 초점을 맞추고 있어 깊이 정보 추정에는 한계가 있다. 본 논문은 이러한 한계를 극복하고자, 깊이 맵 만을 학습하여 novel view에서의 깊이를 추론하는 Depth-as-Color NeRF 모델을 제안하여 보다 정확한 깊이 추정이 가능함을 보인다.

2. 관련 연구

Barron et al. [1]은 NeRF에서 카메라와 대상과의 거리에 따라 렌더링 시에 발생하는 얼라이어싱 문제를 해결하기 위한 mipmapping 방법을 제안하였다. 따라서 mipmap의 개념을 응용하여 ray가 아닌 cone방식으로의 새로운 인코딩 방식을 제안한다. Barron et al. [2]는 경계가 없는 장면의 사실적인 렌더링을 위한 360도 scene에서 Mip-NeRF의 확장을 제시한다.

Ranftl et al. [3]은 Midas 모델을 통해 서로 다른 bias를 가진 여러 데이터셋을 함께 학습시킴으로써 다양한 환경에서 촬영된 한번도 보지 못한 scene의 single image에 대한 깊이 추정을 수행할 수 있게 하였다.

3. Depth-as-Color NeRF

3.1. 실험 방법

* 구두(포스터) 발표논문
* 본 연구는 과학기술정보통신부 한국연구재단 중견연구자 지원사업(RS-2024-00339681), 정보통신기획평가원 메타버스 융합대학원(IITP-2024-RS-2023-00254129)의 지원으로 수행된 연구임.

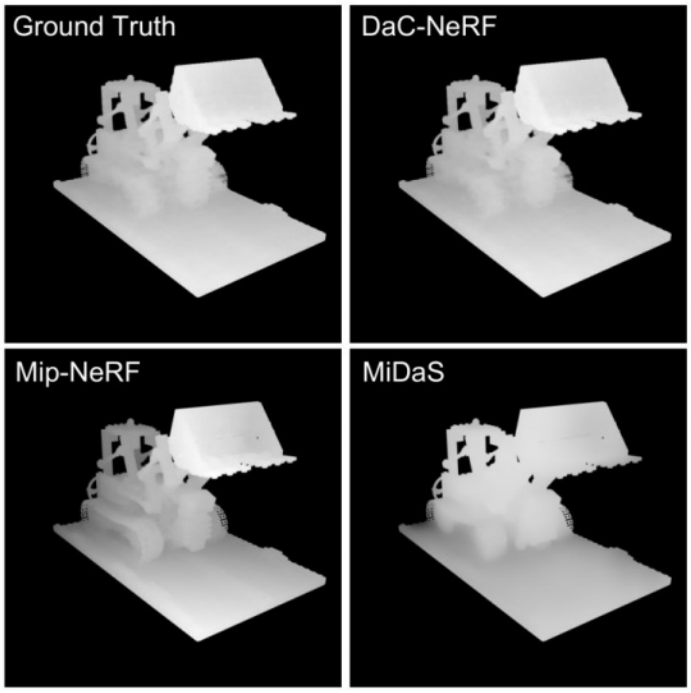


그림 2 Qualitative Comparison

	Mip-NeRF	MiDaS	DaC-NeRF
PSNR ↑	23.556	23.907	**29.363**
SSIM ↑	0.979	0.971	**0.981**

표 1 Evaluation

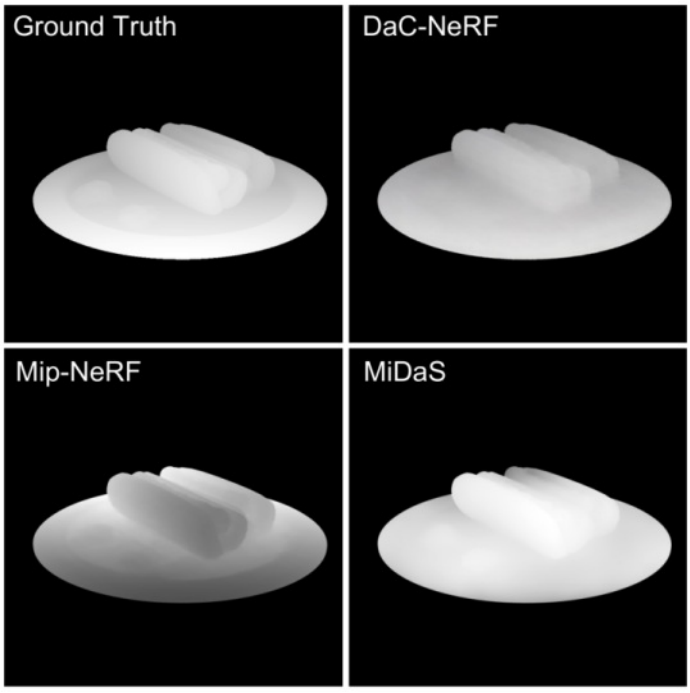


그림 3 Qualitative Comparison

	Mip-NeRF	MiDaS	DaC-NeRF
PSNR ↑	15.026	25.255	**26.663**
SSIM ↑	0.874	0.976	**0.991**

표 2 Evaluation

기존 NeRF와 다르게, 렌더링 데이터셋의 깊이 맵을 RGBA 4개의 채널로 변환하여 각 채널의 깊이 정보를 색상 데이터로 매핑한다. 이를 통해 NeRF 모델은 깊이 정보를 단일 채널이 아닌 4채널 색상 데이터로 인식하며 scene을 학습한다. 기준이 정해져 있지 않은 깊이의 특성 상 최소, 최대값을 기준으로 정규화 또한 필요하다.

4. 결과

본 기법은 두 개의 Intel(R) Core(TM) i9-10900 2.8GHz, GeForce RTX 3090 위에서 학습, 구현되었으며, 800×800의 해상도를 사용한다. Train, test, validation set은 각각 140, 30, 30장으로 구성하였으며 총 iteration은 25000 번으로 설정하여 학습 시간은 약 85분이 소요되었다.

4.1. Depth Evaluation

그림 2, 3은 깊이 정보만을 활용하여 학습한 DaC-NeRF와 다른 모델의 깊이 추정 실험 결과이다. Mip-NeRF에서는 RGB 데이터셋을 활용하여 학습한 후 렌더링 시 깊이 맵을 시각화 하였으며 대표적인 깊이 추정 모델인 MiDaS에서도 RGB를 통해 깊이를 추출하였다.

표 1, 2에서 볼 수 있듯, MiDaS, Mip-NeRF의 방법과 비교하였을 때, 렌더링 데이터 셋에서 뛰어난 품질의 깊이 맵을 볼 수 있다.

5. 결론 및 한계점

본 논문은 깊이 맵을 통해 모델을 학습하여 새로운 시각의 깊이를 예측하는 NeRF 모델을 제안하였으며, 이를 통해 기존에는 고려되지 않았던 색상정보가 없는 환경에서 높은 품질의 3D 재구성과 시각화를 가능케 한다. 허나, 이미지의 깊이 추정을 위해 이미지의 특징이 애초에 명확해야만 한다는 한계점이 존재한다. 또한, 동일한 장면이라도 서로 다른 시점에서 볼 때, 각 시점 별로 상이한 깊이 맵이 도출될 수 있기에 이러한 차이를 효과적으로 처리하고 일관된 깊이 정보를 학습하도록 하기 위해 추후 정규화나, multi-view를 사용하여 해결한다.

참고문헌

[1] Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., & Srinivasan, P. P. (2021). MIP-NERF: a multiscale representation for Anti-Aliasing Neural Radiance fields. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*.
[2] Barron, Jonathan T., et al. "Mip-nerf 360: Unbounded anti-aliased neural radiance fields." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[3] Rene Ranftl, Katrin Lasinger, David Hafner, Konrad ´ Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1623–1637, 2020.

다중 FPGA 병렬 렌더링을 활용한 실시간 레이 트레이싱 가속화*

Real-Time Ray Tracing Acceleration Using Multi-FPGA Parallel Rendering

Yunho Han[0], Jiyoung Kim, Woochan Park
Dept. of Computer Science, Sejong University
{yhhan, jykim}@rayman.sejong.ac.kr
pwchan@sejong.ac.kr

Abstract

In this work, we implemented a multi-FPGA (chip) hardware architecture for accelerating ray tracing on multi-FPGA boards. To achieve this, we introduce a parallel acceleration structure that utilizes inter-chip frame partitioning and intra-chip screen partitioning. This provides the advantage of accelerating in parallel and enables scalable performance improvement. Experimental results show that the proposed multi-FPGA architecture improves FPS performance by an average of 3.67x compared to the single-FPGA structure.

1. Introduction

Ray tracing is an advanced rendering technique used to generate realistic graphics. However, it requires massive amounts of computation, so real-time processing is only possible on dedicated hardware.

In this study, we built a multi-FPGA-based ray tracing system suitable for real-time processing and implemented multi-core hardware based on parallel frame rendering. For this purpose, we designed a frame allocation-based rendering HW for multiple FPGAs and verified its effectiveness through performance measurements on various scenes.

2. Construction of a Multi-FPGA Ray Tracing Rendering System based on Frame Partitioning

Due to the limited resources available on a single FPGA board, there are limitations in improving rendering performance. To solve this problem, we studied a multicore ray tracing rendering architecture that can be implemented on multiple FPGAs, and built a multichip architecture based on a parallel frame partitioning method suitable for real-time ray tracing [1]. This paper extends the system in [1] to improve its performance.

The traditional screen partitioning-based multi-chip architecture needs to transfer the updated acceleration structure data from the dynamic scene to each chip. In the proposed frame partitioning-based parallel rendering scheme, frames are divided and allocated according to the number of chips in advance, and each frame performs tree building and ray tracing operations on the same chip. This method enables the performance of the entire system to be scalable in proportion to the number of FPGAs and enables parallel accelerated rendering using the screen partitioning method inside the chip.

3. Experimental Environment

The implemented system consisted of an Intel Core i9-9920X (3.5 GHz) CPU and four Alveo U250 boards. Twelve ray tracing cores were deployed on each chip to achieve real-time ray tracing. Table 1 shows the main components of the scene.

Components / Scenes Name	Cornell Box	Coke	Gear	Jewel	Veget-able	Wood
Total Primitives	34	12,228	7,430	8,970	2,720	20,160
# of Lights	1	3	3	1	3	2
Total Frames	1,000	1,000	1,000	851	601	1,001

표 1: scenes major components

* This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency(KOCCA) grant funded by the Ministry of Culture, Sports and Tourism(MCST) in 2024(Project Name:Development of AI-based content production technology to advance volumetric studio and virtual production, Project Number:RS-2024-00351961, Contribution Rate: 100%)

Figure 1 shows the multi-FPGA system we built. Figure 2 shows the configuration of the proposed multi-FPGA system. Lastly, Figure 3 shows the scene used in the benchmark.

그림 1: Multi-FPGA system

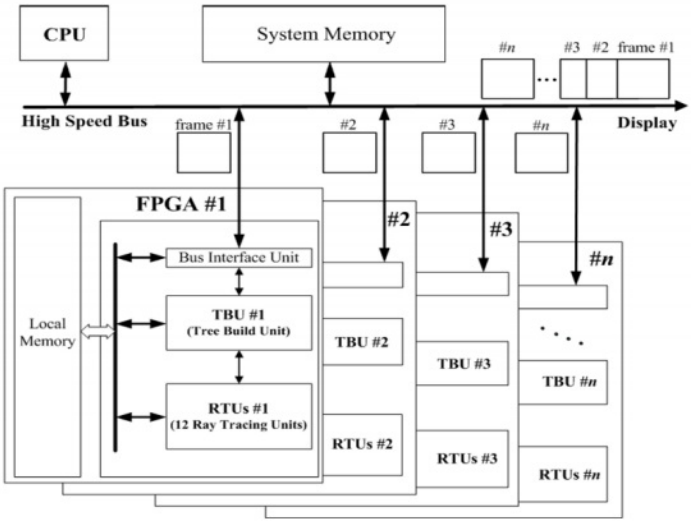


그림 2: Proposed Multi-FPGA system configurations

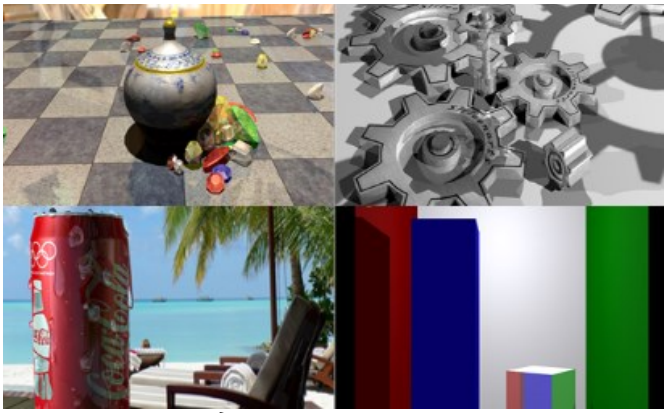


그림 3: Benchmark scenes

4. Experimental Results

To ensure the accuracy of the experiment, it was repeated three times and averaged. The scene used in the experiment was run at 1920 * 1080 resolution. In Table 2, we see an average improvement of 3.67x compared to single-board performance.

Compare to /Performance	Cornell Box	Coke	Gear	Jewel	Veget-able	Wood
Single-Core (FPS)	10.2	6.36	7.77	5.2	9.78	8.6
Multi-Core (FPS)	37.1	24.6	29.0	18.3	37.3	29.6
Performance Improvement(%)	362	387	373.3	351.9	381.7	344.1

표 2: performance comparison

References

[1] J. Lee, J.-H. Nah, W. Chung, T.-H. Lee, and W.-C. Park, "An Implementation of Multi-Chip Architecture for Real-Time Ray Tracing Based on Parallel Frame Rendering." *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers (IEEE), pp. 118968–118976, 2021.

XR 서비스를 위한 대용량 3D 도면 전처리 기술*

서명배[0], 박형진[1], 신승엽[23]
한국건설기술연구원 미래스마트건설연구본부[0,1], PMRC[2], KI-ITC ARRC[3]
{smb, parkhyungjin}@kict.re.kr[0,1], youb1649@kaist.ac.kr[23]

Large-sized 3D drawing preprocessing technology for XR services

MyoungBae Seo[0], HyungJin Park[1], SeungYoub Shin[23]
Korea Institute of Civil Engineering and Building Technology[0,1], PMRC[2], KI-ITC ARRC[3]

요약

건설분야 XR 서비스를 위해서는 3D 도면 모델인 BIM이 필요하나 대부분 용량이 커서 XR로 서비스할 때 속도지연, 성능저하 등이 발생한다. 이에 본 논문에서는 대용량 BIM 데이터를 기반으로 XR로 서비스가 가능한 데이터 변환 및 경량화 기술을 개발하고 테스트를 진행하였다.

1. 서론

건설산업에서 파라메트릭 설계(Parametirc Design)를 가능하게 하는 BIM (Building Information Modelling)과 몰입형 가상설계 등을 가능하게 하는 XR(Extended Reality) 은 건설 전과정 시각화를 통해 다양한 사회현상을 해결할 수 있는 핵심 기술로 최근 부각되고 있다. 이에 최근 BIM과 XR 기술을 결합하기 위한 다양한 시도가 건설분야에서 진행중에 있다[1]. 하지만 BIM 데이터를 XR로 서비스하기 위해서는 다양한 XR 장비, 대용량 BIM 데이터 지원, 네트워크 속도에 따른 전송속도 지연, 이종 OS 지원 등 다양한 한계가 존재한다. 이를 해결하기 위해서는 BIM 데이터를 XR로 서비스하기 위해서는 데이터 경량화, XR 기반 중립파일 포맷 지원, BIM to XR 변환 시간 단축, 데이터 재활용 등 선결해야 하는 문제들이 존재하며 본 논문에서는 이를 해결하기 위한 방법을 제시한다.

2. BIM to XR 핵심기술 개발

2.1. 기존 문제점 분석

BIM to XR은 대용량 BIM 데이터를 XR 환경에서 서비스가 가능한 형태로 변환 및 분리하는 작업을 이야기한다. 현재는 3D 모델링→렌더링 S/W(경량화)→게임엔진(Unity, Unreal 등) 등을 거쳐 XR로 서비스가 되며 단순한 XR 환경 도면뷰잉는 기존 S/W를 활용하면 가능하나 시뮬레이션 등을 수행하기 위한 BIM to XR 변환 등에는 많은 시간이 소요된다(그림 1 참조). 특히 XR로 서비스하기 위해서는 대용량 BIM 데이터 처리방식, 재활용이 가능하기 위해 XR에서 쓸 수 있는 포맷 지원, 모바일 및 저속 네트워트 환경에서 지원 가능한 BIM to XR 변환기술은 매우 중요하다.

그림 1. 일반적인 XR 시각화 제작 프로세스

2.2. BIM to XR 데이터 변환, 저장 및 압축기술

이러한 문제를 해결하기 위해 대용량 BIM 데이터를 형상정보와 속성정보로 분리하여 클라우스 서버에 저장한 후 XR 서비스 환경에 맞춰 데이터를 선별하여 로딩하는 방식을 개발하였다. BIM은 표준포맷인 IFC를 활용하였으며 IFC를 XR 환경에 서비스하기 위해 Obj, Dae, Step, GLB 등 다양한 포맷을 대상으로 변환 테스트를 진행하였다. 모바일 환경에 적합하고 가시화가 빠르며 WebGL 호환이 용이한 GLB(GLTF) 파일을 XR 서비스를 위한 형상 중립파일 포맷으로 선정하였고 속성은 XML로 별도로 분리하여 최종적으로는 IFC to GLB, IFC to XML 형태로 데이터가 분리 가능하도록 개발하였으며, 또한 네트워크 환경이 열악할 경우를 대비하여 구글의 Draco를 활용하여 데이터 압축율을 높일 수 있도록 개발하였다.변환과정은 총 3단계에 걸쳐서 진행된다. IFC파일을 Converter를 통해서 형상정보인 GLB와 속성정보인 XML로 분류하고 추출된 3D 형상정보는 구글의 Draco 알고리즘을 통해 경량화 작업이 진행된다. 경량화 작업이 진행된 파일은 객체 수만큼 *.drc가 생성되며 리패키징을 통해 1개의 파일로 재 구성된다. GLB와 XML의 객체속성 정보 연동을 위해 GUID를 활용하였으며 XML 데이터는 RDB 형태로 저장하여 향후 로딩시 성능향상이 가능하도록 하였다. 이를 정리하면 그림 2와 같다.

* 포스터 발표논문
* 본 연구는 2023년 정부(과학기술정보통신부)의 재원으로 국가과학기술연구회 오픈XR(OXR) 융합연구사업(No. CRC21011)의 지원을 받아 수행된 연구임.

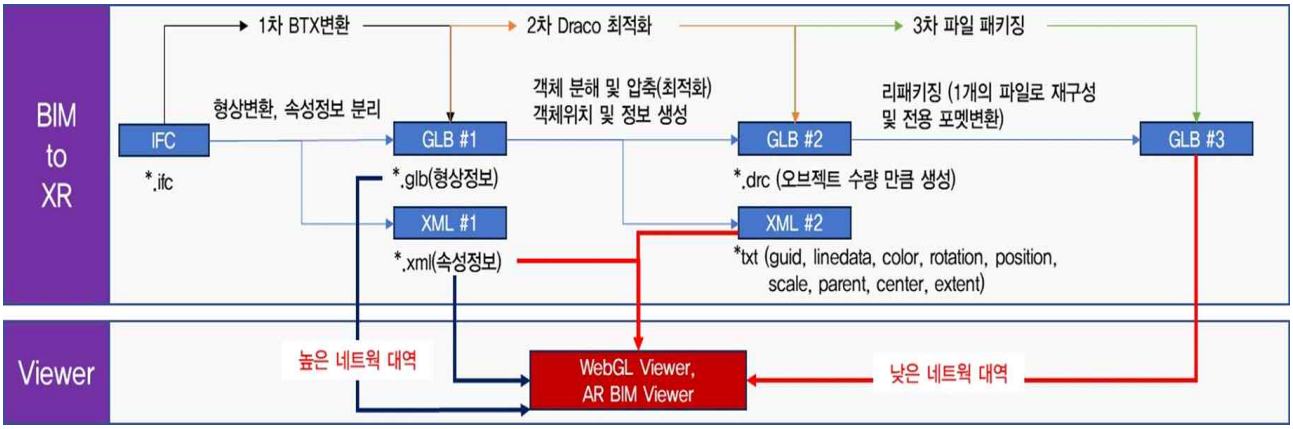
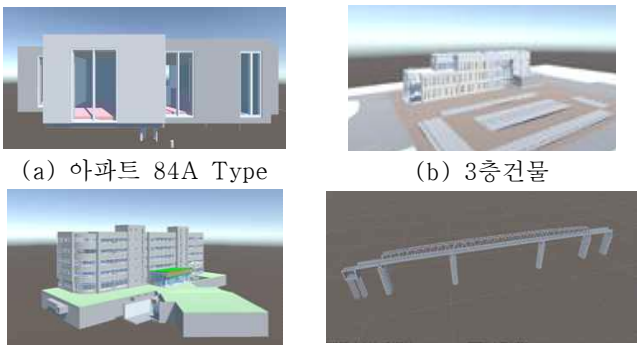


그림 2. 데이터 변환 및 압축이 포함된 BIM to XR 프로세스

2.3. 테스트

BIM to XR 변환 테스트를 위해 아파트, 건물, 교량 등 다양한 형태의 IFC 데이터를 기반으로 변환 테스트를 진행하였다. 테스트를 위해 아파트, 3층 건물, 4층 건물, 속성정보가 포함된 모델, 교량 등 다양한 형태의 모델을 준비하였다. IFC 파일 용량은 작게는 32MB부터 큰 것은 2.8GB까지 다양하게 준비하였으며 테스트한 모델은 그림 3과 같다.



(a) 아파트 84A Type (b) 3층건물

(c)4층 건물 (d)교량

그림 3. 테스트에 활용된 3D 모델(건축, 토목)

4가지 모델에 대해 BIM to XR 변환 테스트를 진행하였다. 테스트 결과 평균 85% 이상의 압축률을 확인할 수 있었다(표 1 참조). 복잡한 폴리곤이 많은 형상의 경우 압축율이 상대적으로 높았다. 또한 변환 단계에서 속성과 형상을 분리하였기 때문에 향후 뷰어에서 다양한 속성정보도 확인할 수 있어 향후 XR 서비스 단계에서 활용도는 매우 높을 것으로 판단된다.

표 1. BIM to XR 및 Draco를 기반으로 변환한 결과

모델명	IFC 용량 (MB)	GLB#1 (MB)	GLB #3 -Draco	압축률
아파트 84A Type	101.00	50.00	5.49	89.02%
3층 건물	65.80	70.60	8.07	88.57%
4층 건물	32.90	26.70	6.30	76.40%
교량	2,820.00	1,379.00	198.00	85.64%

3. 결론 및 향후 연구

향후 BIM to XR 변환 자동화, 변환시 데이터의 손실율, 다양한 XR 장비 호환성 테스트, 도시모델 지원 및 경량화를 위한 Cesium 및 3D Tiles 연동, WebGPU 지원 등을 검토할 예정이다.

참고문헌

[1] Sepehr Alizadehsalehi, Ahmad Hadavi, Joseph Chuenhuei Huang, “From BIM to extended reality in AEC industry”, Automation in Construction, Volume 116, 2020.
[2] 박혜진, 추승연. “BIM 기반 VR융합 기술을 적용한 다차원 시각화 설계검토 프로세스에 관한 연구”, 『대한건축학회 학술발표대회 논문집』 제38권 제1호, 2018. pp.287.

스케치와 점유 지점을 활용한 2차원 솔리드 경계표현법 전환 알고리즘*

이민기[1] , 장동수[2], 김영민[1] [2]
[1]서울대학교 전기정보공학부
[2]서울대학교 뉴미디어통신 연구소
{mingi1019, 96lives, youngmin.kim}@snu.ac.kr

Sketch and occupancy to 2D solid boundary representation conversion algorithm

Mingi Lee[1], Dongsu Zhang[2], Young Min Kim[1] [2]
[1]Dept. of Electrical and Computer Engineering, Seoul National University
[2]Institute of New Media and Communications, Seoul National University

요약

솔리드 경계표현법(boundary representation)은 기하를 정확하게 표현할 수 있으면서 면적을 구성하는 폐곡선에 대한 정보를 담고 있기 때문에 CAD 모델링에서 중요한 역할을 한다. 본 논문에서는 이차원 솔리드 경계표현법을 parameteric한 곡선으로 구성된 이차원 스케치와 이차원 공간의 점에 대해 내/외부를 분간하는 점유 지점을 바탕으로 추론하는 알고리즘을 제시하고 구현한다. 단순하지만 다양한 이차원 형상에 제시한 알고리즘을 적용함으로써 알고리즘이 다양한 형상에 대해 효과적으로 작동하는 것을 질적으로 보인다.

1. 서론

CAD 모델링에서 경계 표현법(boundary representation)은 기하를 표현하는데 있어서 핵심적인 역할을 담당한다. 경계표현법은 primitive들로 구성된 parameteric surface와 그 표면의 경계에 대한 정보로 기하를 표현하기 때문에 mesh와 point cloud와 같은 다른 표현법과 달리 기하를 정확하게 저장할 수 있는 특성을 지닌다. 이 때문에 CAD 사용자들은 CAD 프로그램에서 내부 기하 모델을 경계표현법으로 바꾼 후, 경계표현법을 출력하거나, 혹은 편의에 맞게 mesh, point cloud, voxel과 같은 표현법으로 근사하여 기하를 출력한다.

본 연구에서는 경계표현법 중에서 유효한 면적을 지니는 2차원 솔리드한 경계표현법의 생성에 대해 다룬다. 솔리드한 경계표현법은 3차원으로 확장될 시 유효한 공간(부피)을 지니는 물체에 대한 표현법으로 확장이 되기 때문에 실제 물체를 만드는 computational fabrication 분야에서 핵심적인 역할을 한다. 2차원에서 일반적인 경계표현법은 기하를 선과 선의 경계(시작과 끝)으로 표현하지만, 솔리드한 경계표현법은 선들로 구성된 폐곡선으로 인해 나뉜 공간에 대한 내/외부의 판단이 추가적으로 요구된다. 본 논문에서는 선분들의 집합으로 표현된 스케치와 한 점의 경계면 내/외부에 있는지에 대한 정보인 점유 지점을 바탕으로 폐곡선을 찾아 솔리드한 경계표현법으로 표현법을 전환하는 알고리즘을 제시한다.

2. 방법론

본 논문에서는 내부와 외부가 구분되지 않은 매개변수 곡선들로만 이루어진 스케치와 점유 지점에 대한 정보를 이용하여 닫힌 루프를 형성하는 곡선들을 식별하고, 이를 바탕으로 유효한 2차원 모양의 경계 표현법을 생성하는 방법을 제안한다. 이를 위해 컴퓨터 그래픽스 분야에서 널리 사용되는 flood fill [1] 알고리즘을 활용하였다. flood fill 알고리즘이 이산적으로 작동하는 반면 스케치의 매개변수 곡선들을 연속적인 실수이기 때문에, 입력받은 점유 지점들을 0.01 단위로 반올림하고 한 반복마다 0.01만큼씩 탐색하도록 하였다. 곡선에 대한 교차 여부 역시 0.01 이내의 오차를 허락하였다.

구현된 알고리즘은 다음과 같은 단계로 구성된다.
-초기 설정 및 데이터 입력: 스케치의 곡선 데이터를 입력 받고, 점유 지점을 초기 설정한다.
- 영역 채우기: 설정된 시작점에서부터 인접한 점들을 탐색하며 영역을 채운다. 탐색 과정에서 각 점이 곡선 위에 있는지를 확인하여, 해당 점이 곡선 위에 있는 경우 해당 곡선을 추적하고, 아닌 경우 탐색을 계속한다.
- 곡선 식별: 채워진 영역을 둘러싸는 곡선들을 식별한다.
- 영역 저장 및 반복: 최종적으로, 채워진 영역과 이를

둘러싸는 곡선들을 저장한다. 여러 점유 지점들로부터 과정을 반복한 뒤 저장된 영역들을 비교하여 일정 개수 이상 반복된 영역을 채택한다.
- 2차원 경계표현형으로 변환: Python OCC(Open Cascade)[2]를 사용하여 위에서 채택한 모양을 경계표현형으로 변환한다.

다음은 알고리즘을 구성하는 주요 부분에 대한 설명이다. 곡선 추적을 위하여 다양한 형태의 매개변수 곡선(직선, 원, 호)에 대해 점이 곡선 위에 있는지를 확인한다. 각 곡선 타입에 따라 다른 검사를 수행하며, 직선의 경우 교차 곱을 이용해 점이 직선 위에 있는지, 원의 경우 원의 방정식을 이용해 점이 원 위에 있는지를 확인한다. 호의 경우, 점의 좌표를 극좌표로 변환하여 호의 시작 각도와 끝 각도 사이에 있는지를 검사한다. 이를 통해 호 내부에 점이 위치하는지 확인할 수 있다. 또한, 두 영역이 동일한지 확인하기 위해서는, 각 영역을 구성하는 점들을 집합으로 변환하여 비교한다. 두 영역의 점들이 일정 허용 오차 내에서 일치하는지를 검사하여, 동일한 영역인지 확인한다.

이 방법론을 통해 스케치 상의 매개변수 곡선들로부터 닫힌 루프를 식별하고, 이를 기반으로 유효한 2차원 경계를 효과적으로 생성할 수 있다.

3. 결과

제안된 알고리즘을 실험하기 위해서 Skexgen [3] 스케치 데이터셋에서 실험을 하였다. 그림 1에 입력 스케치, 점유 지점, 그리고 출력 2차원 경계표현법을 나타내었다. 이를 통해 제안한 알고리즘이 효과적으로 2차원 경계표현법을 생성할 수 있음을 질적으로 검증하였다.

4. 결론

본 연구에서는 스케치와 점유지점에 대한 표현법을 2차원 솔리드 경계표현법으로 전환하는 알고리즘에 대해 제시하였다. 알고리즘은 flood fill을 바탕으로 작동하며, 간단하지만 다양한 2차원 형상에 대해 효과적으로 작동하는 것을 질적으로 확인하였다. 본 알고리즘은 2차원에서 작동하는 한계를 지니지만, 유사한 방법론으로 3차원에서 솔리드 경계표현법으로 나타난 기하로의 전환도 가능할 것으로 보인다.

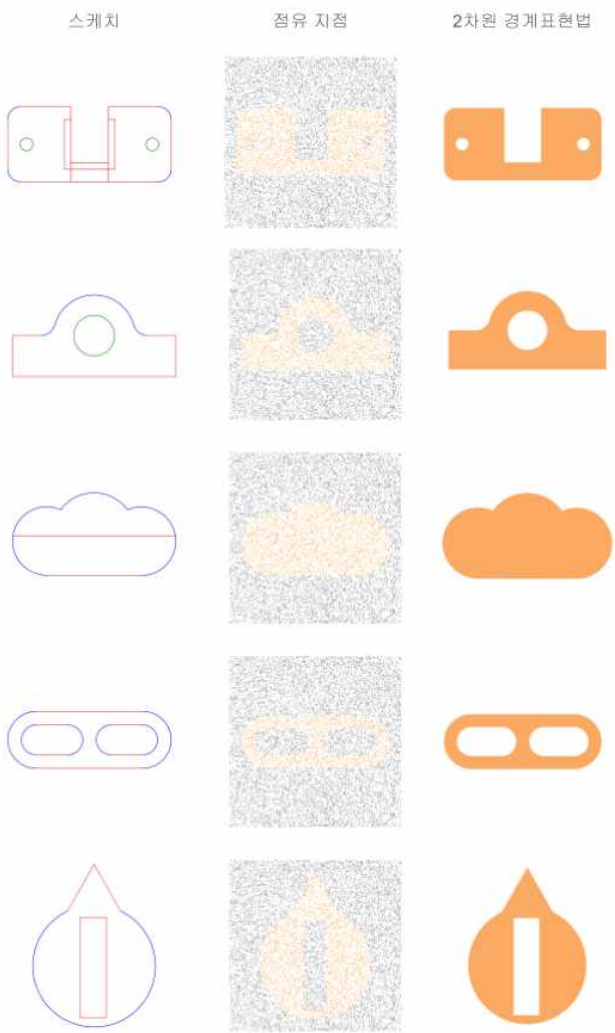


그림 1. 왼쪽부터 순서대로 입력 스케치, 점유 지점과 출력 2차원 경계표현법. 스케치와 점유지점이 주어졌을 때, 제시된 방법론을 통하여 2차원 솔리드 경계표현법으로 형상을 표현할 수 있다.

참고문헌

[1] Burtsev, S.V. and Kuzmin, Y.P. An efficient flood-filling algorithm. Computers & graphics, 1993
[2] PythonOCC, https://github.com/tpaviot/pythonocc-core
[3] X. Xu and K. Willis and J. Lambourne and C. Cheng and P. Jayaraman and Y. Furukawa. SkexGen: Autoregressive Generation of CAD Construction Sequences with Distentangled Codebooks. ICML, 2022

특수 목적의 웹 CAD 시스템을 위한 데이터 변환 도구의 구현*

권초염[0], 김규환, 이환용
아주대학교 소프트웨어학과
{shikuawo, rbgur05, hwan}@ajou.ac.kr

Implementation of Data Format Conversion for Special Purpose Web CAD

Chuyan Quan[0], Gyuhwan Kim, Hwanyong Lee
Dept. of Software Engineering and Computer Engineering, Ajou University

요약

특수한 목적으로 개발된 CAD 시스템은 자주 사용되는 핵심 기능만을 효과적으로 구현하여 활용할 수 있는 반면, 일반 상용 CAD 시스템이 갖고 있는 방대한 기능을 사용할 수 없다는 단점이 있다. 본 연구에서는 알루미늄 프로파일 설계에 특화된 웹 CAD 시스템에서 생성된 데이터를 일반 CAD 시스템에서 사용할 수 있도록 DXF 포맷으로 변환하는 도구를 구현하였다. 이를 통해 서로 다른 목적으로 제작된 CAD 시스템 간 데이터를 전달함으로써, 작업의 효율성과 융통성을 높일 수 있었다. 구현에는 Python 언어 기반의 ezdxf 인터페이스를 사용하여 웹 CAD에서 생성된 JSON 데이터를 DXF 형식으로 변환하는 도구를 구현하였으며, 구현 결과 데이터의 정확성과 효율성을 확인하였다.

1. 서론

특수 목적의 CAD 시스템은 특정한 설계 요소와 규격을 포함하고 있어 효율적인 반면, 기능이 제한적이다. 이를 해결하기 위해 범용 CAD 시스템과 상호 보완적으로 사용하는 것이 일반적이다. [그림1]은 EFSOFT 사의 EFCAD 시스템으로 알루미늄 프로파일 기구 설계에 특화된 웹 CAD 시스템이다[1]. 프로파일의 기구의 특성상 세그먼트가 직교하는 방식으로 체결되며, 프로파일 규격에 따라 체결 방식, 부품이 자동으로 결정된다.

[그림1] EFSOFT의 EFCAD 화면

이러한 제한 조건을 반영하여 효율적인 CAD 시스템을 구현할 수 있다. 그러나, 아주 낮은 빈도로 발생하는 특수 가공에 대처하지 못하는 경우가 있다. 이를 해결하는 방법으로 고려할 수 있는 것이, 특수 목적의 CAD 기능을 범용 시스템의 플러그인 형태로 개발하는 방법이 있다. 이 경우에는 사용자가 특정 범용 CAD를 보유하고, 학습해야 한다는 단점이 있다. 본 연구는 다양한 범용 CAD와 상호 운용성을 제공하기 위하여 데이터 변환을 통해 이를 해결하고자 하였다. 본 연구에서는 DXF 포맷을 변환 대상으로 선정하였다. 소스가 공개된 캐드 포맷으로 사실상의 산업 표준으로 사용되고 있기 때문이다.

2. 시스템 설계 및 구현

2.1 시스템 요구사항 분석

본 연구에서 제안하는 시스템은 다음과 같은 요구사항을 만족해야 한다:

- 정확성: 데이터 변환 과정에서 정보 손실이 없어야 하며, 설계 의도가 정확히 전달되어야 한다.
- 효율성: 변환 과정이 빠르고 웹 인터페이스에서 실시간에 가까운 입출력 처리가 가능해야 한다.
- 이식성: 웹 CAD 시스템에 간단하게 이식하여 사용할 수 있어야 한다.

위와 같은 요구사항을 만족하기 위해 다음과 같은 절차를 갖는 변환 도구를 구현하였다.

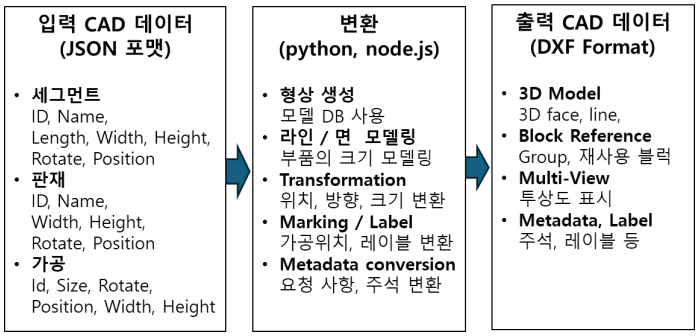
2.2. 프리미티브 JSON 데이터 추출 프로세스

먼저 CAD엔진은 사용자의 요청에 의해 JSON[2]으로 정의된 기구 설계 정보를 생성한다. 이 JSON 데이터 안에는 프로파일 세그먼트, 체결, 가공, 판재 등의 모델, 위치, 크기 등의 형상정보와, 여러 메타데이터가 포함되어 있다. 이 정보를 구조화된 데이터로 변환한다. 이 단계에서는 프로파일 모델과 가공의 형상은 반영하지 않고, 단순한 프리미티브 데이터로 전달한다. 즉 세그먼트, 조인트는 직육면체로, 가공은 위치와 가공 종류 정보만을 갖고 있다. 이 JSON 형식 데이터는 웹 CAD 시스템의 렌더링을 위한 Scene Graph와 동일한 형식이다.

* 포스터 발표논문
* 본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2022-0-01077)

2.3. 부품 모델 변환 단계

추출된 데이터를 기반으로 ezdxf[3] 인터페이스를 활용하여 DXF 포맷으로 알루미늄 프로파일의 부품 모델을 구현한다. 각 부품을 하나의 블록으로 캡슐화하여 모델 간의 결합과 각 부품의 세부적인 3차원 형상을 생성하는 데 사용된다. 부품의 길이, 너비, 높이 및 회전 정보를 이용하여 ezdxf의 3D 면구성 기능을 이용해 각 면을 정의하고 조립하여 하나의 기구로 구성한다. 패널 및 가공 형상 생성에도 동일한 접근 방식이 적용되며, 필요에 따라 적절한 DXF 속성과 함께 블록 참조를 모델 공간에 배치하는 역할을 한다. [그림2]



[그림2] 변환 전후의 데이터

2.4. 위치/회전/크기 변환 단계

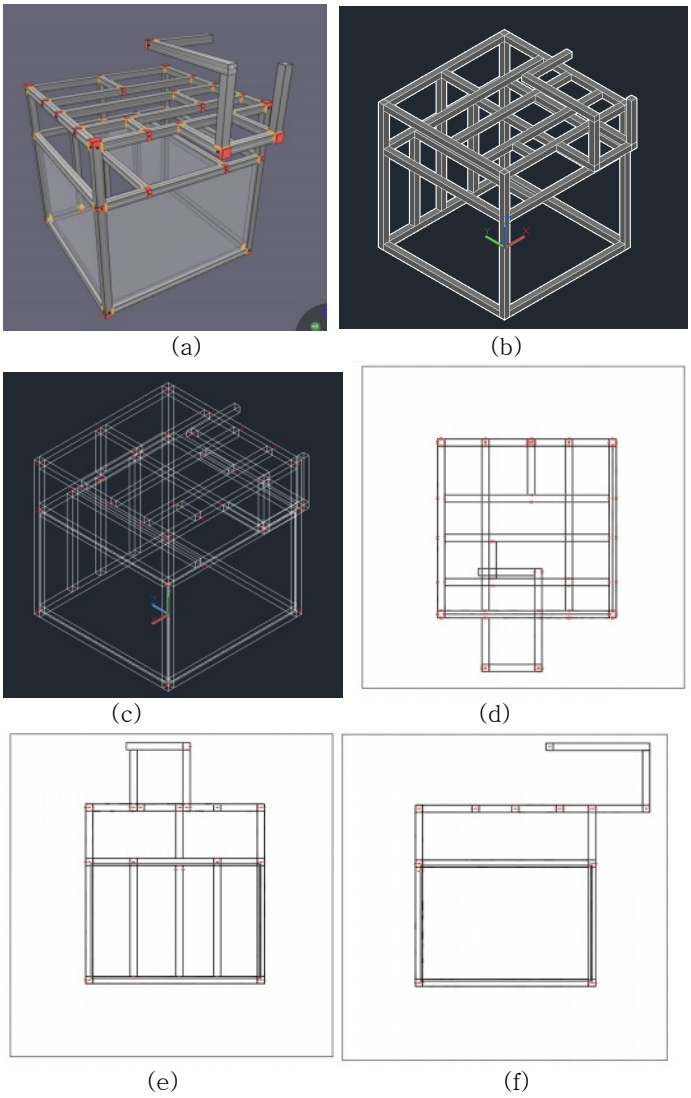
각 부품이 크기와 방향에 따라 변환하고, 필요한 경우, 회전 조건에 따라, 부품의 시작 위치 좌표는 변환되고 모델이 어떤 방향으로 회전하든 간에, 구조적 정합성과 시각적 정확성이 유지되도록 하여 변환 데이터를 만든다.

2.5 생성된 DXF의 입출력 단계

변환된 데이터는 ezdxf 인터페이스를 통해 DXF 파일로 변경된다. ezdxf는 Python 언어로 제작되어 있어, 웹 시스템에 연동하기 위해서는 Node.js를 이용하여 웹 서비스에 포팅된다. 생성된 DXF 파일을 사용자가 자신의 컴퓨터에 저장하는 웹 인터페이스가 제공하였다.

3. 결과 및 분석

구현된 도구는 전용 CAD 프로그램에서 생성된 기구 데이터를 변환하여 DXF 파일로 변환하여, AutoCAD로 추가적인 작업을 할 수 있었다. [그림3] 그림에서 (d), (e), (f)는 ezdxf로 직접 생성된 투상도 이며, (b), (c)는 AutoCAD에서 데이터 가져오기를 통해 DXF 파일을 가져와 표시한 결과이다. 이 과정에서 웹 CAD에서 작성한 각종 메타데이터가 전달됨을 확인하였다. 범용 웹 서버 환경에서 [그림3]의 프로파일 데이터의 변환에 50msec 이내에 생성이 가능함을 확인하였다.



[그림3] 변환 결과 예
(a) EFCAD 화면 캡처 (b),(c) AutoCAD 화면 캡처, 와이어프레임 (d),(e),(f) ezdxf로 생성한 투상도

전용 CAD가 제공하는 효과적인 설계 방식을 통해 설계 작업의 생산성을 높이고, 전용 CAD가 제공하지 못하는 특수 가공에 대한 표시, 구조분석, 조립 검증, 투상도 출력, 단면도 출력, 등의 기능을 범용 CAD로부터 제공받을 수 있어 전체적으로 설계 작업의 생산성을 높일 수 있을 것으로 기대된다.

참고문헌

[1] EFSOFT, “알루미늄 프로파일 캐드 시스템 - EFCAD System”, Available: https://efsoft.kr [액세스: 5 2024].
[2] ISO, ISO/IEC 21778:2017 - The JSON data interchange syntax, ISO, 2017.
[3] M. Moitzi, “documentation for ezdxf release 1.3.1,” Available: https://ezdxf.readthedocs.io/ [액세스: 5 2024].

자유형상 곡선에 기반한 3D 텍스트 메쉬의 변형

권성현[01], 정현석[2], 윤승현[*1]
동국대학교 컴퓨터·AI학과[1] 동국대학교 멀티미디어공학과[2]
kshyunee@dgu.ac.kr, fiijk1255@dongguk.edu, shyun@dongguk.edu

Deformation of 3D Text Meshes Based on Freeform Curve

Seong-Hyeon Kweon[01], Hyun Seok Jung[2], Seung-Hyun Yoon[*1]
Dept. of Computer-AI, Dongguk University[1] Dept. of Multimedia Engineering, Dongguk University[2]

요약

3D 텍스트 모델은 다양한 산업에서 시각적 효과를 극대화하고, 정보를 효과적으로 전달하는 데 중요한 도구로 사용된다. 본 논문은 3D 텍스트 메쉬를 생성하고 자유로운 형상으로 변형하는 기술을 제안한다. 먼저 트루타입 글꼴 기반으로 문자의 윤곽선을 추출한 후, 제약 Delaunay 삼각화를 통해 3D 텍스트 메쉬를 생성한다. 이후 자유형상 곡선을 사용하여 텍스트 메쉬의 형상을 변형한다. 제안된 기법은 다양한 시각적 효과를 부여할 수 있으며, 3D 모델의 형상을 반영한 워터 마크 기술에 활용될 수 있다.

1. 서론

2D 텍스트를 3차원으로 표현한 3D 텍스트 모델은 온라인 광고나 영화, VR/AR 등 미디어 분야 뿐만 아니라 보안 및 저작권 보호 분야에서도 활용되고 있다.

Dhiman 등[1]은 윤곽선 샘플링과 Delaunay 삼각화를 통해 3D 텍스트 메쉬를 생성하고, 자기 교차와 홀(hole) 처리 및 부드러운 모서리 생성 방법에 대해 다루었다. 하지만, 생성한 텍스트 모델의 응용 기술에 대해서는 논의하지 않았다. Li 등[2]은 트루타입 글꼴에서 텍스트의 윤곽선을 샘플링하고, Delaunay 삼각화를 적용하여 3D 텍스트 메쉬를 생성한 후, 3D Boolean 연산을 통해 다른 모델에 워터 마크로 삽입하는 방법을 제안했다. 하지만 이 기법은 곡률이 거의 없는 편평한 영역만 워터마크 삽입이 가능하며, 단조로운 형상의 텍스트 메쉬를 생성하는 한계를 지닌다.

본 논문에서는 문자의 윤곽선을 샘플링한 점들로 제약 Delaunay 삼각화를 수행하여 3D 텍스트 메쉬를 구축하고 자유형상 곡선 및 대상 메쉬의 형상에 따라 변형하는 기법을 제안한다. 또한, 이 기법의 다양한 응용 사례를 소개한다.

* 포스터 발표논문
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음 (IITP-2024-RS-2023-00254592)

2. 3D 텍스트 메쉬 생성

2.1. 문자 윤곽선 추출

트루타입 글꼴에 기반한 문자는 베지에(Bezier) 곡선으로 표현된 윤곽선을 가진다. 본 연구에서는 FreeType 라이브러리[3]를 사용하여 문자의 윤곽선을 나타내는 선분과 베지에 곡선의 제어점 정보를 얻는다. 이 곡선들을 일정 간격으로 샘플링하고 윤곽선의 샘플링 점을 선분으로 연결함으로써, Figure 1(a)과 같이 닫힌 폴리라인들을 얻을 수 있다.

2.2. 3차원 텍스트 메쉬 생성

3D 텍스트 메쉬의 상단 표면을 얻기 위해 CDT 라이브러리[4]를 사용하여 제약 Delaunay 삼각화를 수행한다[5]. Figure 1(b)에서 볼 수 있듯이, 생성된 삼각형은 윤곽선 샘플링 점과 이를 연결하는 선분을 포함하며, 닫힌 폴리라인들로 둘러싸인 영역 내에 존재한다.

상단 표면의 삼각형들을 모두 복제한 후 법선을 뒤집어 하단 표면을 생성한다. 그 다음, 상단 표면과 하단 표면을 z방향으로 각각 $\frac{d}{2}, -\frac{d}{2}$만큼 이동하고 동일한 x, y 좌표를 가진 정점들을 연결하여 측면 표면을 구성한다. 여기서 d $(d > 0)$는 사용자가 지정한 텍스트 메쉬의 깊이 값이다. 중복 정점과 에지를 제거하고 연결 정보를 갱신하면 Figure 1(c)와 같은 3D 텍스트 메쉬를 얻을 수 있다. 마지막으로 문자열 내 각 문자의 너비와 베어링, 문자 사이 공백을 고려하여 스케일 및 이동 변환을 수행하면 최종적인 3D 텍스트 메쉬를 얻을 수 있다.

(a) (b) (c)

Figure 1: 3D 텍스트 메쉬 생성 과정: (a) 윤곽선 샘플링, (b) 제약 Delaunay 삼각화 (c) 3D 텍스트 메쉬.

3. 자유형상 곡선 기반 텍스트 메쉬 변형

본 논문에서는 3차원 자유형상 곡선 $C(t)$를 기반으로 텍스트 메쉬의 형상을 변형한다. 텍스트 메쉬의 상단 및 하단 표면 상의 임의의 정점을 각각 $\mathbf{v}_f = (x_f, y_f, z_f)$, $\mathbf{v}_b = (x_b, y_b, z_b)$라고 할 때, 다음과 같은 두 가지 변형 방법을 제안한다.

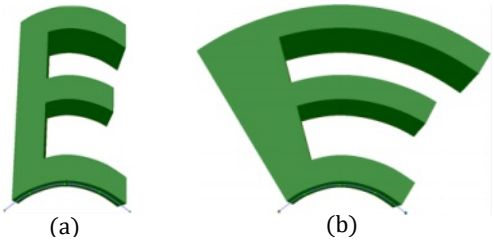
첫 번째, 고정 매핑 방법은 곡선 위 하나의 기준 점 $C(t_0)$에서의 프레네 틀장 $\{\mathbf{T}(t_0), \mathbf{N}(t_0), \mathbf{B}(t_0)\}$을 이용하여 정점 $\mathbf{v}_f$와 $\mathbf{v}_b$를 다음과 같이 $\mathbf{v}_f'$와 $\mathbf{v}_b'$으로 매핑한다:

$$\mathbf{v}_f' = C(x_\mathrm{f}) + y_\mathrm{f} \cdot \mathbf{N}(t_0) + z_\mathrm{f} \cdot \mathbf{B}(t_0),$$
$$\mathbf{v}_b' = C(x_\mathrm{b}) + y_\mathrm{b} \cdot \mathbf{N}(t_0) - z_\mathrm{b} \cdot \mathbf{B}(t_0).$$

두 번째, 동적 매핑 방법은 곡선 위 임의의 점 $C(t)$에서의 프레네 틀장이 $\{T(t), N(t), B(t)\}$일 때, 각 정점의 x 값을 t에 대입한 프레네 틀장을 이용하여 $\mathbf{v}_f$와 $\mathbf{v}_b$를 다음과 같이 변환한다:

$$\mathbf{v}_f' = C(x_f) + y_f \cdot \mathbf{N}(x_f) + z_f \cdot \mathbf{B}(x_f),$$
$$\mathbf{v}_b' = C(x_b) + y_b \cdot \mathbf{N}(x_f) - z_b \cdot \mathbf{B}(x_b).$$

결과적으로, 벡터 N과 B가 모든 정점에서 일정한 방향(고정 매핑)이나 각 정점에 따라 다른 방향(동적 매핑)을 가지도록 텍스트 메쉬를 변형할 수 있다(Figure 2).



(a) (b)

Figure 2: 자유형 곡선으로 텍스트 메쉬를 변형한 결과: (a) 고정 매핑 방법, (b) 동적 매핑 방법.

4. 실험 결과 및 응용

제안된 기술은 3D 텍스트 메쉬를 복잡한 자유형상 곡선에 매핑하여 다양한 시각적 효과를 생성할 수 있다. Figure 3은 여러 곡선으로 변형된 텍스트 메쉬의 예시를 보여준다. 또한, Figure 4는 변형 과정에서 시작 파라미터 값의 조절을 통해 텍스트 메쉬가 실시간으로 곡선을 따라 이동하고 변형될 수 있음을 보여준다. 이 기술은 3D 타이포그래피와 키네틱 타이포그래피에 활용될 수 있으며, 변형된 3D 텍스트를 사용하여 컴퓨터와 사용자를 구분하는 CAPTCHA 기술로의 확장이 가능하다.

Figure 3: 자유형상 곡선으로 변형된 다양한 3D 텍스트 메쉬

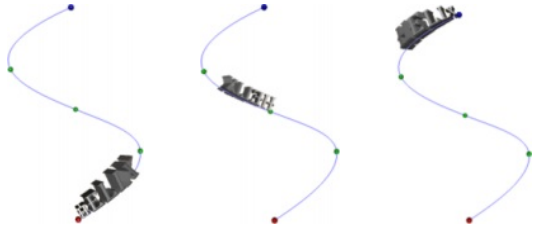


Figure 4: 곡선을 따라 이동 및 변형하는 3D 텍스트 메쉬

또한, 입력 메쉬 위의 측지 곡선[6]을 기반으로 텍스트 메쉬를 변형하고 Boolean 연산을 수행하여 해당 영역의 형상을 반영한 워터 마크를 생성할 수 있다. Figure 5는 측지 곡선으로 변형된 텍스트 메쉬와 입력 메쉬의 union 연산을 통해 양각의 워터 마크를, difference 연산을 통해 음각의 워터 마크를 생성한 결과를 보여준다.

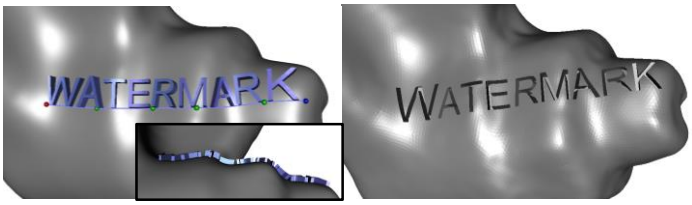


Figure 5: 입력 메쉬의 형상을 반영한 양각, 음각 워터마크.

5. 결론

본 논문은 트루타입 폰트 기반 3D 텍스트 메쉬 생성 및 자유형상 곡선 매핑을 통한 다양한 시각적 효과 제공 방식을 제안하였다. 문자의 윤곽선을 일정 길이로 샘플링하고 제약 Delaunay 삼각화를 수행하여 3D 텍스트 메쉬를 생성하였다. 또한, 자유형상 곡선을 사용한 고정 및 동적 매핑 방법을 소개하고 다양한 변형 사례를 통해 본 기술의 활용 가능성을 입증한다.

제안 기술은 텍스트 메쉬 변형 시 자기 교차 또는 곡선 상의 여러 텍스트 메쉬 간 교차가 발생할 수 있다는 한계점이 존재한다. 향후 연구에서는 텍스트 메쉬의 교차 제거 기법과 다국어 메쉬 생성 기술을 연구하고, 기존 연구와의 비교 분석을 수행할 예정이다.

참고문헌

[1] A. Dhiman, P. Agrawal, S. K. Bose, and B. S. Vandrotti, Textgen3d: A real-time 3d-mesh generation with intersecting contours for text, *Proceedings of the Satellite Workshops of ICVGIP 2021,* pp.251-264, 2022.
[2] A.-B. Li, H. Chen, and X.-L. Xie, Visible watermarking for 3d models based on 3d Boolean operation, *Egyptian Informatics Journal,* vol.25, p. 100436, 2024.
[3] David Turner, Robert Wilhelm, Werner Lemberg, Freetype 2, 2000. URL: www.freetype.org
[4] A. Amirkhanov and K. Åkerblom, artem-ogre/CDT: Constrained Delaunay Triangulation (C++), 2019. URL: https://github.com/artem-ogre/CDT
[5] Borut Žalik and Ivana Kolingerová, An incremental construction algorithm for Delaunay triangulation using the nearest-point paradigm, *International Journal of Geographical Information Science*, Volume 17, Issue 2, Pages 119-138, 2003.
[6] Ha Y, Park J-H, Yoon S-H. Geodesic Hermite Spline Curve on Triangular Meshes. *Symmetry*, Volume 13. 2021

사람의 얼굴을 이용한 동물 3D 모델링 생성 서비스

이종혁[01], 소병욱[02], 조현태[1], 이준수[2], 백지웅[1], 최명걸[2*]
[1]가톨릭대학교 컴퓨터정보공학부
[2]가톨릭대학교 미디어기술콘텐츠학과
{ljh20011, thquddnr123, hyuntae9912, land8746, jw6133, mgchoi}@catholic.ac.kr

Service for Generating Animal 3D Objects Using Human Faces

Jong-Hyuk Lee[01], SO BYUNGUK[02], JO HYUNTAE[1], LEE JUNSOO[2], BACK JIWOONG[1], Myung Geol Choi[2*]
[1]School of Computer Science and Information Engineering, The Catholic University of Korea
[2]Dept. of Media Technology and Media Contents, The Catholic University of Korea

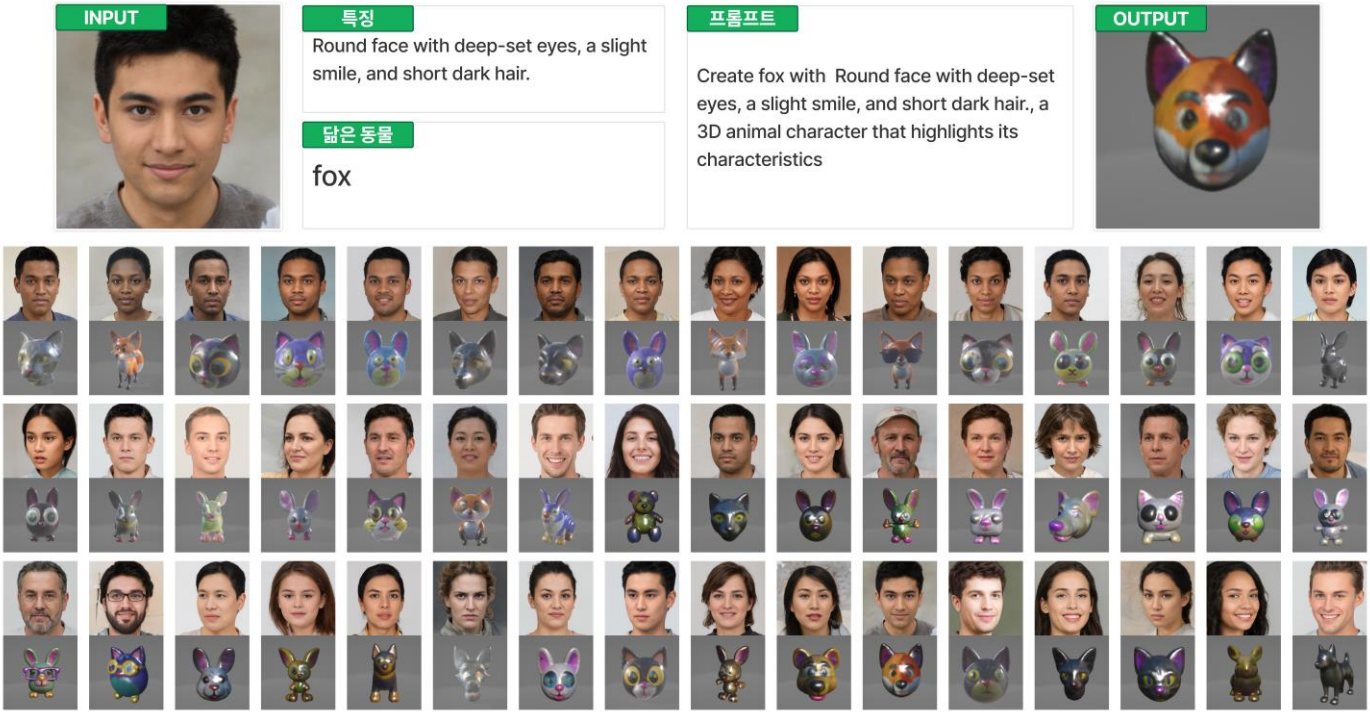


그림 1 : 사람의 얼굴 이미지를 사용하여 생성된 3D 모델링

요약

최근 일어난 생성형 AI의 발전은 컴퓨터 그래픽스 분야에서, 텍스트만으로도 3D 모델링을 생성할 수 있게 되었다. 본 연구에서는 사람의 얼굴 이미지로부터 추출한 특징들을 프롬프트로 사용하여 3D 동물 모델링을 생성하는 서비스를 제안한다. 유사한 동물을 추론하기만 하는 기존의 서비스와 달리, 인물의 특징과 닮은 동물의 정보를 포함한 프롬프트를 생성하여 보다 개성 있는 3D 동물 모델링을 생성한다. 이를 위해 ResNet34 모델을 사용하여 유사한 동물을 추론하고, ChatGPT API를 통해 인물의 특징을 추출하였다. 이렇게 추출된 정보들은 하나의 프롬프트로 연결되어, meshy AI의 API를 통해 3D 모델링으로 변환되었다. 실험 결과, 전문가 평가를 통해 생성된 3D 모델의 질적 우수성을 확인하였다. 본 연구는 얼굴 이미지 한 장만으로도 특징이 잘 드러나는 3D 동물 모델링을 제공하며, 향후 다양한 동물에 대한 실험을 통해 사용자 경험을 극대화할 계획이다.

1. 서론

텍스트로부터 이미지를 생성하는 모델인 Stable Diffusion[1]과 같은 Diffusion 모델의 성공으로 인해 텍스트-이미지로 구성된 엄청난 양의 학습 데이터가 확보되면서, 텍스트를 사용하여 고품질의 2D 이미지 생성이 가능하게 되었다. 이는 2D 이미지를 기반으로 3D 모델링을 생성하는 과정에서 요구되는 비용을 감소시켰고, DreamFusion[2]과 같이 텍스트만으로도 고품질의 3D 모델링을 생성하는 모델의 발전에 기여하였다. 동시에,

* 포스터 발표논문
* 학부생 주저자 논문임.

LLaVA[3], CLIP[4]와 같이 이미지와 텍스트를 동시에 처리할 수 있는 VLM (Vision-Language Model)의 등장은 하나의 모델을 사용하여 이미지에 대한 정보를 처리할 수 있게 되었다.
우리는 이러한 기술들을 활용하여 사람의 얼굴 이미지로부터 추출한 특징을 사용하여 3D 동물 모델링을 생성하는 서비스를 제안한다. 기존에 분류 모델을 사용하여 유사한 동물을 추론하는 서비스를 제공한 경우는 있었지만, 추론 결과로 나온 동물에 대한 정보만을 사용자에게 전달하였다. 본 연구에서는 이에서 더 나아가, 단순히 동물을 추론하고 결과를 반환하는 것이 아니라 분류 모델이 반환한 인물과 닮은 동물에 VLM이 반환하는 이미지 속 인물의 특징을 프롬프트에 첨가하여 3D 모델링을 생성하는 방법을 통해 보다 현실감 있고 개성 있는 동물 캐릭터를 만들어내는 서비스를 제안한다.

2. 방법론

사람의 사진을 통해 유사한 동물을 추론하기 위해 ResNet34[5]를 사용하였다. 동물의 이미지로 학습된 ResNet 모델은 사용자가 입력한 사람 이미지에 대해 가장 확률적으로 유사한 동물의 클래스를 텍스트의 형태로 반환한다. 동시에, ChatGPT API[6]를 이용하여 사람의 이미지에 대한 특징을 추출한다. 이렇게 만들어진 동물 키워드와 인물의 특징을 사용하여 프롬프트를 제작한 뒤, 제작된 프롬프트를 사용하여 3D 모델링을 생성하였다. 3D 모델링 생성에는 meshy AI[7]의 API를 활용하였으며, 이는 고품질의 3D 모델을 효과적으로 생성하는 데 사용되었다.
모든 과정을 원활하게 연결하기 위해 Spring Boot와 Flask를 사용하여 시스템을 구축하였다. Spring Boot는 백 엔드 서버를 구축하고 프론트 엔드와의 연동에 사용되었으며, Flask는 모델들의 API를 담당하였다. 이 시스템을 통해 사용자는 웹사이트에 이미지를 업로드하면 3D 모델링이 생성되고, 결과를 확인할 수 있다.

3. 실험 결과

실험에 사용한 ResNet34는 Colab환경에서 16GB NVIDIA Tesla T4를 통하여 학습하였고, 데이터셋은 저작권이 없는 동물 사진 600장을 사용하였다. 동물은 개, 고양이, 사슴, 곰, 토끼, 여우의 6종류로 구분하였다. ChatGPT API의 경우 이미지와 텍스트를 모두 처리할 수 있어야 하므로 GPT-4o를 사용하였다. 그림 1을 통해 얼굴과 유사한 동물과, 얼굴의 특징을 포함하고 있는 프롬프트를 통해 3D 모델링을 생성할 수 있음을 확인할 수 있다. 우리의 결과에 대해 사용자 평가를 수행하지 않았지만, 전문가를 대상으로 3D 모델링의 품질에 대한 질적 평가를 수행하였다. 질적 평가를 위해 다양한 실험 결과를 생성할 필요가 있었고, 이때 사용된 사람의 얼굴 이미지는 생성형 AI를 이용하여 만들어진 사람의 얼굴 이미지[8]를 사용하였다. 전문가 평가 결과 사람과 생성된 동물 캐릭터 얼굴 간 표정의 유사성을 인정받았다.

4. 결론

본 연구에서는 사람의 얼굴 이미지로부터 특징과 닮은 동물을 추정하여 3D 모델링을 만드는 방법을 제안하였다. 본 연구에서 제시하는 방법의 모든 절차는 자동화가 가능하여 일반 사용자가 한 장의 사진으로 본인의 개성을 반영한 3D 동물 모델링을 생성하는 서비스를 제공할 수 있을 것으로 기대된다.
본 연구의 한계점 중 하나는 결과에 대한 정량 평가가 이루지지 못했다는 점이다. 향후 다수의 일반인을 대상으로 체계적인 사용자 평가를 진행하여 결과를 수치화하고, 성능을 개선해 나갈 계획이다. 전문가의 자문을 받아 진행된 질적 평가에서는 제시된 방법의 유용성에 긍정적 평가를 받았으나, 두 가지 한계점에 대한 피드백도 있었다. 첫 번째 한계점은 결과로 출력되는 동물 외형의 스타일이 모두 유사하게 정형화되어 나온다는 것이고, 두 번째 한계점은 동물 얼굴의 표정이 풍부하지 못하고 세부 표현이 약하다는 점이다. 이는 연구에 사용된 Meshy AI가 일반적인 3D 모델을 제작하기 위해 만들어진 서비스로, 동물의 얼굴이나 표정 생성에 특화되어 있지 않기 때문에 발생하는 한계이다. 이러한 한계점을 해결하기 위해, 향후 동물 모델과 동물 표정 생성에 특화된 인공신경망을 개발하는 것을 목표로 하고 있다.

참고 문헌

[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, **“High-resolution image synthesis with latent diffusion models,”** in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10 684–10 695
[2] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, **“Dreamfusion: Text-to-3d using 2d diffusion,”** *arXiv preprint arXiv:2209.14988*, 2022
[3] H. Liu, C. Li, Q. Wu, and Y. J. Lee, **“Visual instruction tuning,”** Advances in neural information processing systems, vol. 36, 2024
[4] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., **“Learning transferable visual models from natural language supervision,”** in International conference on machine learn- ing. PMLR, 2021, pp. 8748–8763.
[5] K. He, X. Zhang, S. Ren, and J. Sun, **“Deep residual learning for image recognition,”** in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770– 778.
[6] OpenAI. (2024) ChatGPT [Large language model]. https://www.chatgpt.com
[7] Meshy LLC. (2024) Meshy. ai. [Web service]. https://www.meshy.ai
[8] Generated Photos Inc. (2024) Generated Photos [Web service]. https://www.generated.photos

감사의 글

한 학기 동안 지도해주신 서효중 교수님과,
오수민 멘토님께 감사를 표합니다.

3차원 공간 속 로봇 에이전트 표현을 위한 혼합 신경장*

민철희[01], 문성빈[2], 김영민[1]
서울대학교[1], 고려대학교[2]
{mch5048, youngmin.kim}@snu.ac.kr[1], brian0429@korea.ac.kr[2]

A Mixture Neural Field for the Robot Agent Representation in a 3D Space

Cheol-Hui Min[01], Sung Bin Mun[2], and Young Min Kim[1]
Seoul National University [1], Korea University [2]

요약

로봇 에이전트가 3차원 공간 속에서 시각 정보를 바탕으로 활동하기 위해서는, 로봇에 부착된 카메라 정보에서 취득하는 영상 활용이 필수적이다. 이때, 모바일 매니퓰레이터와 같은 플랫폼에서 취득하는 ego-centric 시각 정보에는 로봇 팔 (robot arm)이 필연적으로 포함되어 있다. 이러한 시각 정보로부터 3차원 신경장 (3D neural field)을 표현할 때는, 카메라의 3차원 자세 (pose)에 따른 광선 (ray)에서 학습된다. 하지만, ego-centric 시각 정보에서 로봇 팔은 카메라 자세에 관계 없이 영상에 존재하므로, 3차원 공간에 임베딩 (embedding)하기가 어렵다. 본 연구에서는 이러한 문제를 해결하기 위해 로봇과 그 이외 배경 및 물체를 독립된 신경장에 임베딩 하는 방법론을 제시한다. 실험 결과, 본 연구는 독립된 신경장 모델링을 통해, 카메라 자세와 분리되어 관측되는 로봇 팔을 성공적으로 모델링하여 향후 비전 기반 모바일 매니퓰레이션 인공 지능 개발에 활용 가능성을 보인다.

1. 서론

객체 지향 표현 (object-centric representation; OCR) [5]은 에이전트가 관측하는 시각 정보 속에 존재하는 물체 및 다양한 정보들을 독립된 잠재 변수 (latent variable)들로 분리하여 임베딩하는 학습법이다. 기존 연구 [2]에서는 고정된 카메라에서 관측된 로봇 에이전트가 물체들과 상호작용하는 상황을 OCR로 표현할 수 있었다. 이 경우, camera pose가 상관 없기 때문에, auto-encoder 형태로 손쉽게 OCR로 장면 속 객체들을 분리하여 임베딩 할 수 있다.

최근에는 3차원 신경장과 통합하여 OCR을 3차원 공간에서 얻는 연구가 많이 수행되고 있다 [3, 4]. 이는 인코딩 과정에서 CNN 및 트랜스포머를 통해 피쳐 (feature)를 얻고 OCR 공간으로 정보를 mapping 한 후 [3], 디코딩 과정에서 카메라가 바라보는 시점 (viewpoint)을 query로 하고, OCR을 key 및 value값으로 하는 self-attention을 수행한다. 이로써, 3차원 카메라 자세에 condition된 디코딩을 수행하여 OCR을 3차원 공간에 임베딩 할 수 있게 된다.

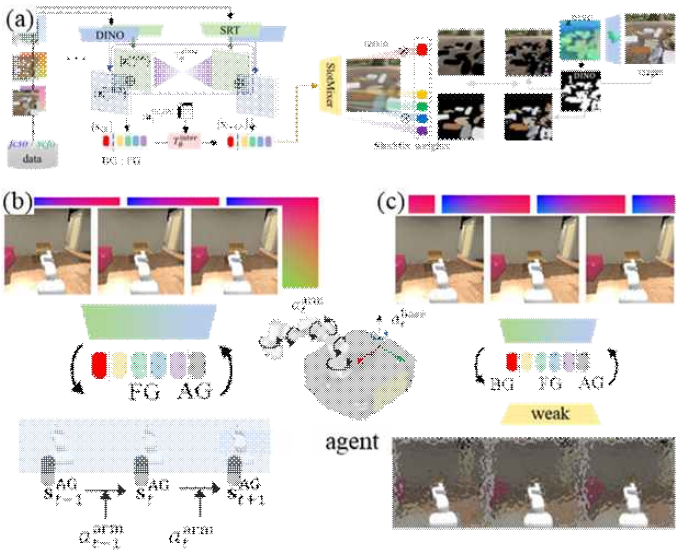


그림 1. (a): TRITON을 통한 3차원 OCR 모델링 (b): 에이전트 동작 (action) a_t을 활용한 에이전트 종속 OCR의 취득. (c): autoencoder 형태의 agent 신경장 학습

하지만 이 방법론은 관측되는 객체가 3차원 공간에 정적 (static)으로 존재함을 가정한다. TRITON[3](**그림 1a**) 에서는 동적인 물체를 모델링하지만, 카메라 모션과 분리 (decouple) 된 객체들을 모델링한다. 따라서 **그림 1b**의 모바일 매니퓰레이터 [1]의 ego-centric 관측정보와 같이 카메라 자세와 couple 된 로봇 팔의 경우, 기존의 방법론으로는 OCR로써 3차원 공간에 임베딩 할 수 없다. 이에, 본 연구에서는 에이전트에 종속된 (dedicated) 별개의 신경장을 모델링하여, 에이전트 신경장과 3차원 배경 및 공간 속 객체들의 신경장을 통합하는 혼합 신경장 (mixture neural field)를 제안한다.

2. 방법

2.1. 에이전트 종속 OCR 취득

에이전트 종속 신경장을 학습하기 위해서는 우선 **그림 1a, b**과 같이 OCR의 집합 원소 중에서 에이전트에

* 포스터 발표논문. 추후 국제 학술대회 투고 예정 연구내용임.

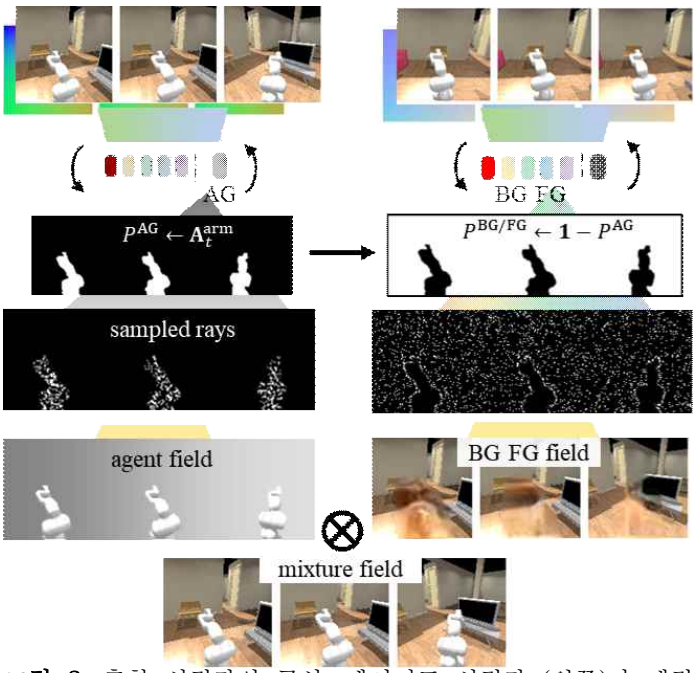


그림 2. 혼합 신경장의 구성. 에이전트 신경장 (왼쪽)과 배경+물체 신경장 (오른쪽)의 Gaussian mixture 형태로 구성됨.

종속된 OCR을 정해야 한다. 본 연구에서는 학습 데이터에서 에이전트의 동작 (action) a_t 정보를 활용할 수 있다는 점을 이용한다. 이는 간단히 $\arg\max_\theta I(a_t; s_t^{arm}, s_{t+1}^{arm})$와 같이 에이전트 OCR s_t^{arm}, s_{t+1}^{arm}과 a_t간의 mutual information 최대화함으로써, s_t^{arm}, s_{t+1}^{arm} 가 로봇 팔 에이전트에 종속된 정보만 표현하면서 학습할 수 있다.

2.2. 에이전트/배경+물체 독립 광선과 혼합 신경장

3차원 신경망을 디코딩하기 위해서는 카메라 기준 매 픽셀 마다 광선을 샘플링하여 렌더링을 수행한다. 이를 위해서는 디코딩 과정 이전에 광선을 샘플링할 수 있는 공간적 확률 분포를 얻어야 한다. 본 연구에서는 **그림 1c**와 같이 autoencoder 형태로 에이전트 신경장을 사전 학습하여 attention map A_t^{arm} (**그림2**)을 얻는다. 이 영역을 바탕으로 광선을 샘플링 하면 **그림2**와 같이 에이전트와 배경+물체 영역에 대해 mutually exclusive 한 광선으로 학습 시그널을 얻을 수 있다. 다음으로, 학습된 에이전트 신경장과 기존 TRITON을 배경과 물체 영역에 대한 광선 분포 $p^{BG/FG}$에서 학습을 수행하면, 최종적으로 혼합 신경장을 구성할 수 있다.

3. 결과

실험 결과 **그림3**과 같이 모바일 매니퓰레이터 에이전트가 3차원 공간속에서 동작하면서 수집한 여러 에피소드의 ego-centric GT이미지 obs에 대해 비지도학습을 통해 에이전트 샘플링 영역 A_t^{arm} 을 얻은후, **그림 2** 의 도식과 같이 구한 에이전트 신경장 agent 및 배경+물체 신경장 BG/FG를 성공적으로 얻을 수 있었다. 최종적으로 두 신경장을 Gaussian mixture 형태로 구성하여 혼합 신경장 mix를 얻었으며, 관측 이미지와 비교했을 때 충실히 장면 속 에이전트, 배경, 그리고 물체들을 혼합 신경장에 임베딩 했음을 확인할 수 있었다.

4. 결론

본 연구에서는 OCR을 통해 3차원 신경장을 표현할 때, 카메라의 자세와 couple 된 로봇 에이전트가 관측되는 경우, 기존 방법을 통해 3차원 공간에 정보를 임베딩 할수 없었던 한계점을 극복하였다. 이는 에이전트와 이외 영역을 서로 다른 신경장으로 모델링하여 혼합하는 방식을 활용하였으며, 에이전트 영역의 광선 샘플링분포를 OCR을 통해 얻음으로써 독립된 광선 샘플링을 통해 카메라 동작 및 에이전트 정보 사이의 ambiguity를 해소함으로써 구현되었다.

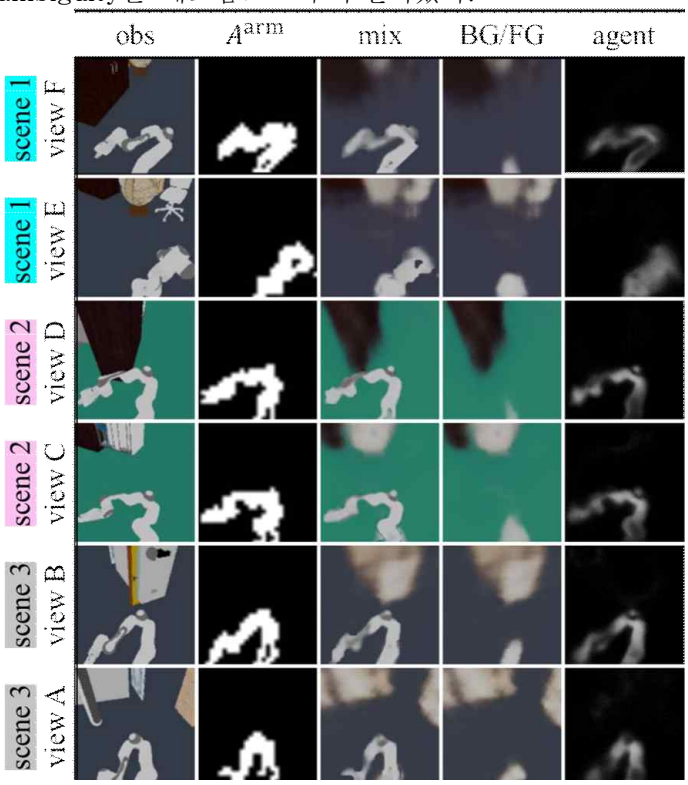


그림 3. 실험 결과. ego-centric GT 이미지 obs에 대한 에이전트 샘플링 분포 영역 A_t^{arm}, 혼합 신경장 mix, 배경+물체 신경장 BG/FG, 그리고 에이전트 신경장 (agent)을 표현.

참고문헌

[1] Siegwart, R., & Nieto, J. (2020). Whole-body control of a mobile manipulator using end-to-end reinforcement learning. arXiv preprint arXiv:2003.02637.

[2] Min, C. H., & Kim, Y. M. (2024). ASIMO: Agent-centric scene representation in multi-object manipulation. The International Journal of Robotics Research, 02783649241257537.

[3] Min, C.-H., & Kim, Y. M. (*in submission*). TRITON: TRansforming Image-level object interactions TO factorized dynamic Neural fields.

[4] Sajjadi, M. S., Duckworth, D., Mahendran, A., Van Steenkiste, S., Pavetic, F., Lucic, M., ... & Kipf, T. (2022). Object scene representation transformer. Advances in Neural Information Processing Systems, 35, 9512-9524.

[5] Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., ... & Kipf, T. (2020). Object-centric learning with slot attention. Advances in neural information processing systems, 33, 11525-11538.

의류의 겉감과 안감의 3D 가우시안 스플래팅 복원 방법

정문경[0], 이도해, 이인권
연세대학교 컴퓨터과학과
{jungk0419, dlehgo1414, iklee}@yonsei.ac.kr

3D Gaussian Splatting for Shells and Linings of Clothing

Mun-Kyung Jung[0], Dohae Lee, In-Kwon Lee
Dept of Computer Science, Yonsei University

요약

본 논문은 의류의 안감과 겉감을 동시에 3D로 복원하기 위한 Gaussian Splatting 기반의 방법을 제안한다. 의류를 3D로 복원하기 위해 마네킹이나 사람에 의류를 착용시킨 사진을 활용하면 의류의 겉감만 복원되며, 보이지 않는 안감 부분을 복원하기 어렵다. 본 논문에서는 의류의 양면을 모두 복원하기 위한 새로운 데이터셋 구성 방법을 제안하고, 3D Gaussian 포인트들의 법선 벡터 방향과 view direction에 따라 각 포인트의 렌더링 여부를 결정하는 알고리즘을 제안한다. 가상 데이터를 활용한 실험 결과, 제안한 방법은 기존 방법에 비해 의류의 양면 3D 복원 퀄리티를 정량적, 정성적으로 상승시켰음을 확인했다. 우리의 방법은 3D 가상피팅 등 다양한 분야에서 응용될 것으로 기대된다.

1. 서론

현실의 3차원 물체를 컴퓨터 프로그램 상에 실제와 같이 복원하는 것은 영화, 게임 등 콘텐츠 제작에 활용될 수 있어 활발히 연구되고 있다. 전통적으로, Structure from Motion (SfM)기법을 활용하여 정지된 물체를 여러 방향에서 촬영한 이미지들로 물체의 3D 정보를 추출하는 방법이 연구되어 왔다. 그러나 SfM 기반의 방법은 Non-Lambertian 표면의 복원이 어렵고, photorealistic한 렌더링 결과를 얻기 어렵다는 한계가 있다. 최근 Neural Radiance Field(NeRF)[1]기술을 활용하여 유한한 시점에서의 객체 이미지를 통해 무한한 시점에서의 렌더링을 학습하는 기술들이 연구되었으나, 학습과 렌더링 시간이 오래 걸린다는 단점이 있다. 이후 발표된 3D Gaussian Splatting(3DGS)[2]기법은 3D 가우시안 포인트들을 2D래스터화 하여 photorealistic한 결과를 실시간으로 렌더링 할 수 있기에 큰 관심을 받고 있다. 3DGS 기법을 활용하여 다양한 물체를 높은 퀄리티로 3D 렌더링 하는 것이 가능하나, 의류를 독립적인 객체로 복원하는 것에는 어려움이 있다.

* 포스터 발표논문
* 본 논문은 요약논문 (Extended Abstract)임.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단 (No. RS-2024-00348094) 및 한국전파진흥협회 (No. RNIX 2023200)의 지원으로 수행되었음.

그림 1. 데이터셋의 구성

의류는 open-surface로 안감과 겉감으로 구성되어 있는데, 마네킹이나 사람이 의류를 착용하였을 때, 한쪽 면(겉감)만 보여, 보이지 않는 면(안감)의 복원이 어렵기 때문이다.

본 논문에서는 의류의 양면을 모두 3D로 복원하기 위한 데이터셋 구성 방법을 제안한다. 또한, 3D Gaussian의 법선벡터와 view direction을 기준으로 의류의 안감과 겉감을 구분하고, 의류의 양면을 동시에 복원할 수 있는 새로운 알고리즘을 제안한다. 가상 데이터를 이용한 실험으로, 제안하는 방법의 복원 결과는 기존 방법에 비해 의류의 양면 복원 퀄리티가 정량적, 정성적으로 상승했음을 확인했다.

2. 방법

2.1. 데이터셋 구축

제안하는 방법은 다음의 단계를 거쳐 의류의 양면을 동시에 3D로 복원하기 위한 데이터를 생성한다. 먼저, 마네킹에 의류의 겉감이 바깥을 향하도록 착용시킨 후 여러 방향에서 사진을 촬영한다(그림1(a)). 다음으로, 의류의 안감이 바깥을 향하도록 옷을 뒤집어서 착용시킨 후 같은 방식으로 촬영한다(그림1(c)). 모든 촬영 이미지에서 배경을 제거한 후 학습에 사용한다 (그림1(b),(d)). 일단, 본 연구에서는 실제 촬영 데이터가 아닌 “Blender”로 생성한 가상 의류 데이터를 사용하였고, 안감과 겉감의 geometry는 동일하고, texture는 다른 가상 데이터가 생성된다.

2.2. 3D 복원 및 렌더링

겉감과 안감의 데이터셋을 SuGaR[3]모델을 통해 각각 학습한다(그림2(b)). 대상 의류가 아닌 불필요한 포인트들(마네킹 등)을 제거한다. 이는 3DGS segmentation방법[4]을 활용하여 수행할 수 있다(그림2(c)). SuGaR모델은 3D Gaussian과 메쉬를 함께 복원

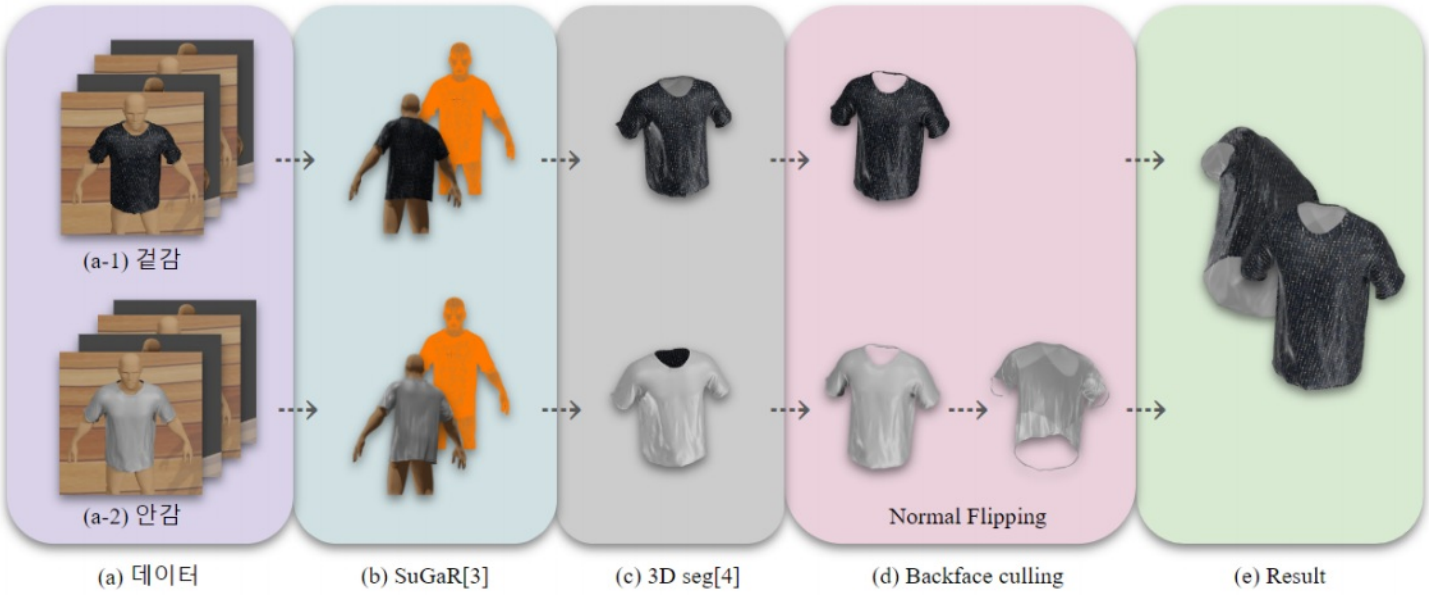


그림 2. 제안 방법의 모식도

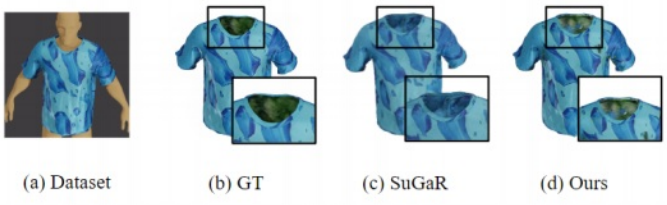


그림 3. 실험 결과

Algorithm	PSNR	SSIM
3DGS	37.38494	0.94520
SuGaR	37.86382	0.94712
Ours	39.13630	0.96289

표1. 정량평가

하며, 모든 포인트는 메쉬의 surface위에 정렬된다. 각 포인트의 법선 벡터는 그 포인트가 속한 surface의 법선 벡터로 정할 수 있다. 이 때, 안감에 해당하는 포인트의 법선 벡터는 옷의 내부를 향할 수 있도록 반대 방향으로 flipping한다(그림2(d))하단). 다음 스텝에서, 독립적으로 복원 과정을 거친 겉감과 안감의 법선 벡터 정보를 포함한 3DGS 결과물을 합친다(그림2(e)). 이 과정을 통해 안감과 겉감이 동시에 복원된 의류의 최종 포인트 클라우드가 생성되며, 모든 포인트들은 법선 벡터 정보를 포함한다. 렌더링 시에는 법선 벡터의 방향과 view direction이 이루는 각도가 90도보다 클 경우에만 포인트를 렌더링한다(그림2(e)). 이를 통해 특정 view direction에서 보여야 할 의류의 안감과 겉감을 구분하여 렌더링한다.

3. 실험 및 결과

그림3의 (a)는 학습에 사용한 겉감 촬영 이미지이다. 그림3(b)는 옷의 안감과 겉감이 동시에 복원되는 ground truth로, 마네킹 혹은 사람을 필요로 하는 현실에서는 촬영이 불가능하다. 그림3(c)은 데이터셋(그림3(a))을 활용한 SuGaR[3]모델의 학습 결과이다. 해당 모델로는 의류의 안감과 겉감을 동시에 복원할 수 없다. 그림3(d)는 제안하는 기법으로, 겉감과 안감을 동시에 복원한 이미지이다. 확대된 목 부분에서(의류의 안감과 겉감이 교차하는 지점) 확인할 수 있듯이, 의류의 안감과 겉감이 구분되는 것과 동시에 3D로 복원된다. 표1은 ground truth와 복원 결과의 렌더링된 이미지 간의 차이를 PSNR과 SSIM으로 보여주고 있으며, 제안한 모델이 기존 모델에 비해 정량적으로 우수함을 보이고 있다.

4. 결론

본 연구에서는, 의류의 안감과 겉감을 동시에 3D Gaussian 포인트로 복원하는 기법을 제안한다. 본 기술은 기존 방법으로는 해결하지 못한 3D 의류 복원 기술의 특정 케이스를 가능하게 한다. 하지만, 티셔츠보다 더 복잡한 형태의 의류에서는 아직 실험이 더 필요하며, 가상 데이터가 아닌 실제 데이터로의 검증이 필요하다.

향후 실제 데이터를 사용하여 제안한 기법을 검증하고, 복원의 정확성과 효율성 등을 개선해야 할 것이다. 이를 통해 가상 피팅 분야에서 실제와 같은 응용에 활용할 수 있을 것으로 기대된다.

참고문헌

[1] Mildenhall et al. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In ECCV 2020
[2] Kerbl et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. In ACM Transactions on Graphics. 2023
[3] Guedon et al. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High Quality Mesh Rendering. arXiv 2024
[4] Lyu et al. Gaga: Group Any Gaussians via 3D-aware Memory Bank. arXiv 2024

Quality Distribution Function for 3D Gaussian Splatting*

Minseo Choi[0], Dohae Lee, In-Kwon Lee
Dept. of Computer Science, Yonsei University

3D 가우시안 스플래팅을 위한 퀄리티 분포 함수*

최민서[0], 이도해, 이인권
연세대학교 컴퓨터과학과
{cmins1011, dlehgo1414, iklee}@yonsei.ac.kr

Abstract

3D reconstruction algorithms that use partial datasets suffer from the problem that the output quality depends on factors such as the number of input images and the arrangement of adjacent input images, which requires re-acquisition of images to obtain satisfactory reconstruction results. To overcome these issues, we propose a quality distribution function (QDF) based on consecutive camera intervals to quantitatively analyze the performance of the 3D Gaussian Splatting (3DGS) algorithm, which can provide guidelines for the acquisition of input images. The efficiency value determined by the QDF enables the selection of an appropriate 3DGS algorithm based on the reconstruction purpose.

1. Introduction

In contrast to neural radiance field-based image synthesis, which requires significant rendering time, a recent advancement is 3D Gaussian splatting (3DGS), which uses ellipsoid-shaped 3D Gaussians [1]. 3DGS achieves high temporal rendering performance due to its explicit representation.

Prior to 3DGS, a deep learning solution called local light field fusion (LLFF) [2] was introduced, which uses a multi-plane image (MPI) representation to capture and render novel views of real-world scenes.

* 포스터발표논문
* 학부생 주저자 논문임.
* 본 논문은 extended abstract 로 연구의 초기결과임.
* 이 연구는 정부 (과학기술정보통신부)의 재원으로 한국연구재단 (No. RS-2024-00348094) 및 한국전파진흥협회 (No. RNIX20230200)의 지원으로 수행되었음.

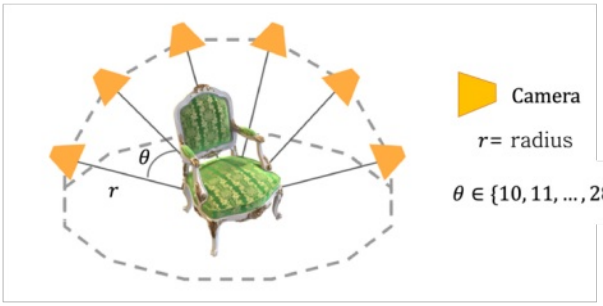


Figure 1: Dataset Construction

They provided guidelines for sampling locations and spacing to help users effectively capture input images. However, their method has the limitation that it is not generally applicable to methods in other domains that are not derived from LLFF [2].

In this paper, we present a general image acquisition guideline for 3D reconstruction techniques that have been proposed so far. Our quality distribution function (QDF) allows users to verify the quality distribution of 3D reconstruction algorithms as a function of the number of images and use it as a guideline for building an input image dataset. Furthermore, the efficiency value, which indicates the quality robustness of a 3D reconstruction algorithm with decreasing number of images, can be used to select an appropriate algorithm based on the given constraints.

2. Methods

2.1. Dataset Construction

Based on a spherical coordinate system, the target to be 3D reconstructed is located at the origin, and cameras are uniformly distributed at a distance r from the target (Figure 1). All camera orientations are towards the origin. Images are taken at camera spacings ranging from 10° to 29°, with 289 images at 10 degrees and 25 images at 29 degrees.

Algorithm	PSNR	SSIM	LPIPS
3DGS	4.38	0.05	0.04
C3DGS	2.91	0.06	0.03
GES	4.24	0.07	0.05

Table 1: RMSE of Regression Fitted with Quadratic Functions

2.2. Quality Distribution Function (QDF)

The changes in image quality metrics, such as PSNR, SSIM, LPIPS values, of the output of 3DGS-based algorithms are fitted by a quadratic function, QDF, through polynomial regression analysis.

The general quadratic form of the QDF we have assumed is:

$$f(x) = a\,(x+b)^2 + c, \quad (a < 0),$$

where x is the angle between the cameras. Smaller $|a|$ makes the increment of $f(x)$ smaller as x increases, making the parabola wider, and conversely, the larger $|a|$, the narrower the parabola. In our case, a is negative, so the value $-1/a$ is proportional to the width of the parabola.

Let $f(x)$ be the PSNR. As x increases, fewer cameras can be used. Since $f(x)$ is a better value the larger it is, the model is more efficient if $f(x)$ decreases less as x increases, i.e., the wider the parabola. In other words, $-1/a$ represents the efficiency of the 3DGS algorithm given the PSNR value. Also, c represents the maximum PSNR of the 3DGS algorithm. In the maximum $f(x)$, x represents the optimal angle between the cameras.

3. Result

In our experiments, the input images were captured using Blender to utilize NeRF [3] based synthesis data. Polynomial regression analysis was performed on a dataset rendered uniformly with the angle spacing between cameras, $\theta = \{10, 11, \ldots, 29\}$. It was assumed that the synthetic data is able to initialize points with COLMAP. The target algorithms for analysis in this experiment are 3DGS [1], GES [4], and Compact-3DGS (C3DGS) [5]. The accuracy of the fitted function was assessed using the Root Mean Square Error (RMSE), with the RMSE values presented in Table 1.

Theoretically, it is feasible to predict the quadratic function with only three data points. The graph in Figure 2 represents the function extracted from the

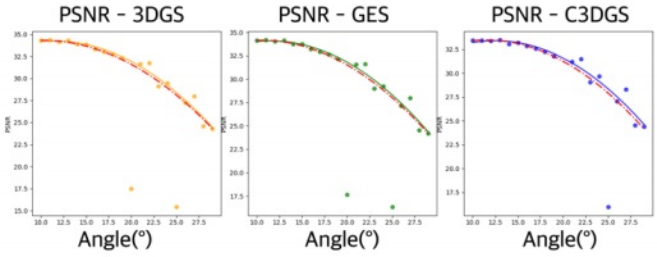


Figure 2: Result of Regression Analysis (solid line: sample prediction, dashed line: population prediction).

Algorithm	Efficiency Value↑	Max Accuracy↑
3DGS	**31.3**	**34.13**
C3DGS	30.9	33.40
GES	**31.3**	34.11

Table 2: Efficiency Analysis

sampled data excluding outliers. Figure 2 also shows the estimated population quadratic function for each of the different 3DGS algorithms.

From the result of the efficiency analysis shown in Table 2, it is clear that the most efficient method is 3DGS[1], which achieves maximum accuracy when the dataset is sampled with an interval of 11.46°.

4. Conclusion

In this work, we contribute in the following ways. First, Quality Distribution Function was developed to quantitatively assess the performance of the 3DGS algorithm. Secondly, efficiency metric was proposed to aid in algorithm selection based on the reconstruction purpose, which helps reduce dataset construction costs. Finally, general guidelines for capturing input images were provided.

Reference

[1] Kerbl, Bernhard, et al., "3D Gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, 42(4):1-14, 2023.
[2] Mildenhall, Ben, et al., "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics*, 38(4): 1-14, 2019.
[3] Mildenhall, Ben, et al., "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM* 65(1):99-106, 2021.
[4] Hamdi, Abdullah, et al. "Ges: Generalized exponential splatting for efficient radiance field rendering," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.
[5] Lee, Joo Chan, et al., "Compact 3D Gaussian representation for radiance field," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

방향전환보행 사용자를 위한 가상환경의 실시간 리셋 예측

전수림[0], 이호정, 이인권
연세대학교 컴퓨터과학과
slchun@yonsei.ac.kr, dearshawn@yonsei.ac.kr, iklee@yonsei.ac.kr

Real-time Reset Prediction in Virtual Environments for Redirected Walking Users

Sulim Chun, Ho Jung Lee, In-Kwon Lee
Dept. of Computer Science, Yonsei University

요약

방향전환보행(Redirected Walking: RDW) 사용자들이 가상환경을 체험할 때 경험하는 리셋(reset)은 몰입감을 저하시키는 중요한 요인이지만, 시뮬레이션이나 실제 사용자 실험을 거쳐야만 파악할 수 있다. 그러나 이러한 방식은 상당한 시간이 소요되기 때문에 가상환경의 다양한 물체와 그 배치에 대해 리셋을 파악하기가 어려운 면이 있다. 따라서 본 연구에서는 방향전환보행 사용자들이 체험할 가상환경의 물체 배치에 따라 발생될 리셋 횟수를 예측하는 딥러닝 모델을 제안한다. 시뮬레이션을 통해 RDW를 체험할 가상환경 물체들을 임의로 배치하며, 이에 따라 발생되는 리셋 횟수와 물체의 top-view 데이터를 수집하였다. 우리는 수집된 데이터를 바탕으로 CNN 기반의 인공신경망을 학습시켰다. 우리의 모델은 임의로 배치된 가상환경에서 발생될 평균 리셋 횟수를 MAE 0.33, MSE 0.21로 예측하였다. 본 연구에서 제안하는 예측 모델은 가상 환경에 대한 평가 뿐만 아니라 최적의 가상환경 배치에도 사용될 수 있다.

1. 서론

가상공간을 직접 보행하며 가상현실을 체험하는 것은 사용자에게 높은 몰입감을 제공한다. 하지만, 현실공간과 가상공간의 물리적인 차이가 존재하기 때문에, 충돌 없이 지속적으로 보행하기 위해서는 사용자의 움직임을 조정해야 한다. RDW는 사용자의 화면에 미세 조정을 가함으로써 현실과 가상의 보행경로에 차이를 만들고, 가상 공간을 무한히 보행할 수 있게 한다 [1].

하지만, RDW를 이용한 미세 조정에도 불구하고 현실공간과의 충돌을 회피하지 못할 경우에는 회전 이득(rotation gain)을 활용하여 사용자의 보행 방향을 재조정하는데, 이를 리셋이라고 한다. 빈번한 리셋의 사용은 사용자의 몰입감을 저하시킨다. 이에 따라 가상현실을 체험하는 사용자가 더욱 적은 리셋을 수행하고 몰입감 높은 콘텐츠를 체험할 수 있게 하기 위해 다양한 RDW 알고리즘들이 제안되어 왔다. 하지만 알고리즘의 개선에도 불구하고 사용자가 체험하는 가상환경의 배치에 따라 리셋 발생 빈도는 달라진다. ENI [2]는 현실과 가상환경의 유사도를 분석하는 지표로서, 리셋의 발생횟수와 비례 관계를 보인다고 연구된 바 있다. 하지만 ENI는 복잡한 환경에서는 실시간 계산이 불가능하며, 리셋 횟수를 정확하게 예측하지 못한다. 우리는 가상환경에 따라 사용자에게 발생할 리셋의 횟수를 예측해주는 CNN 기반의 모델을 제안한다. 우리의 모델은 시뮬레이션을 통해 수집한 데이터로 학습되었으며 여러 종류의 장애물에 따라 MAE 0.33, MSE 0.21로 리셋을 예측하였다.

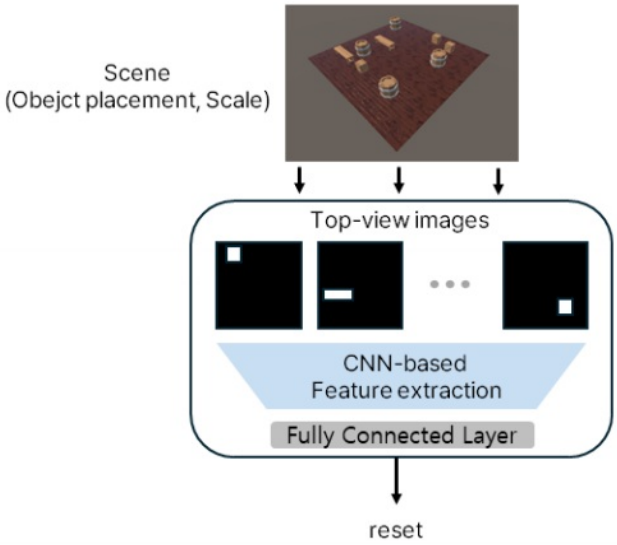


그림 1: 가상환경 리셋 예측 모델 구조

2. 방법

본 연구에서는 CNN(Convolution Neural Network)을 기반으로 물체가 배치된 가상 환경의 상측 이미지에 따른 리셋 횟수를 예측하는 모델을 학습 및 평가한다 (그림 1).

모델은 가상공간 물체의 위치 및 Box Collider 크기 정보를 입력 받아 한 물체당 하나의 상측 이미지를 생성한다. 물체 각각에서부터 전체적 특징을 학습하여 추출할 수 있도록 각 물체의 top-view 이미지들의 총 10개

* 포스터발표논문
* 본 논문은 extended abstract 로 연구의 초기결과임.
* 이 연구는 정부 (과학기술정보통신부)의 재원으로 한국연구재단 (No. RS-2024-00348094) 및 한국전파진흥협회 (No. RNIX20230200)의 지원으로 수행되었음.

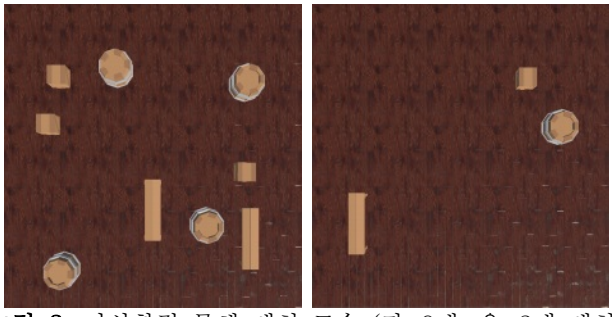


그림 2: 가상환경 물체 배치 모습 (좌: 9개, 우: 3개 배치)

채널로 구성하며 물체가 없는 채널은 이미지가 0으로 채워진다. 각각의 이미지는 너비와 높이가 64 픽셀이며 물체의 위치 및 크기 정보는 이미지의 크기에 맞게 조정되었다.
모델은 채널을 128개까지 연속적으로 확장하는 4개의 합성곱층 및 ReLU 활성화 함수가 MaxPooling층과 Flatten층을 거치고, 그 출력은 초기 입력 데이터를 이미지 별로 평균한 벡터와 결합되어 Fully connected층으로 전달되는 구조이다. 3개의 Fully connected층을 거친 모델은 평균 리셋을 예측한다.

3. 실험 및 결과

3.1. 데이터 수집

OpenRDW [3]를 기반으로 하여 다양한 물체를 배치하고 가상 아바타 시뮬레이션을 통해 리셋을 측정할 수 있는 시뮬레이션을 구축해서 데이터를 수집하였다. 본 실험에는 $5 \times 5m^2$의 장애물이 없는 현실공간과 $10 \times 10m^2$의 가상 공간이 사용되었다. 가상 공간 물체의 종류는 4종류로 하나의 데이터마다 임의로 1~10개가 선택되어 공간의 경계를 벗어나지 않는 임의의 위치에 배치된다 (그림2). 가상 아바타는 직전 목표점을 기준으로 2~6m의 거리에 생성되는 목표점을 향해 걸어가며, 연속된 20개의 목표점이 각기 다른 물체 배치마다 50가지 생성된다. 이러한 50번의 리셋 횟수를 평균하여 그 데이터의 정답 리셋으로 사용하였다. 물체는 모두 장애물로 인식하여 아바타가 부딪히지 않도록 목표점을 생성하였다. 미세조정 기법과 방향전환 컨트롤러로는 TAPF [4]와 Reset-to-Gradient(R2G)를 사용하였다. 총 350개의 데이터를 수집하였고, 그 중 무작위로 선택된 70개의 데이터를 학습하지 않고 검증에 사용하였다.

3.2 검증 결과

우리의 모델과 비교하는 Baseline model은 학습 데이터의 전체 평균으로 예측하는 것으로 선정하였다. 모델 평가 결과 우리의 모델은 MAE(Mean Absolute Error) 가 0.33, MSE(Mean Squared Error)은 0.21이었다. Baseline model은 MAE가 0.51, MSE가 0.39로, 우리의 모델이 더 예측 오차가 적었다. 그림3에 우리 모델의 예측 값과 실제 값의 분포를 산점도로 나타냈다.

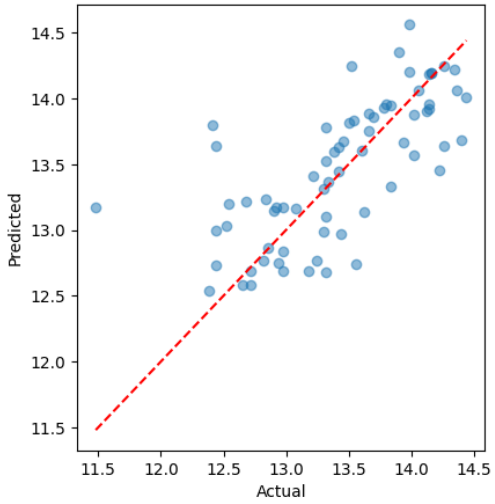


그림 3: 검증 데이터셋에 대한 모델의 평균 리셋 예측 결과

4 토의 및 결론

우리의 모델은 수집된 검증 데이터에 대하여 적은 예측 오차를 보이나, 일정한 크기의 가상환경과 현실공간을 가정하고 있다는 한계가 있다. 향후 연구에서 여러 크기를 고려하고 장애물의 다양성 및 개수를 더욱 확장하여 일반화된 응용을 가능케 할 것이다.
본 연구에서는 가상환경의 물체 배치에 따라 발생될 리셋 횟수를 예측하는 딥러닝 모델을 제안했다. 다양한 종류와 개수의 물체 배치에도 우리의 모델은 Baseline과 비교하여 경향성을 잘 파악하여 적은 오차로 평균 리셋 횟수를 예측하였다. 본 연구에서 제안된 모델의 설계를 개선하고 더욱 다양한 환경에서 데이터를 수집하여 일반화된 환경에서도 사용이 가능하도록 시도하고 있다. 또한, 리셋뿐만 아니라 물체 간의 연관성도 고려하여 최적의 가상환경 배치를 찾는 방법도 향후에 연구될 수 있다.

참고문헌

[1] Razzaque, Sharif. Redirected walking. The University of North Carolina at Chapel Hill, 2005.
[2] Williams, Niall L., Aniket Bera, and Dinesh Manocha. ENI: Quantifying environment compatibility for natural walking in virtual reality. IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2022.
[3] Li, Yi-Jun, et al. Openrdw: A redirected walking library and benchmark with multi-user, learning-based functionalities and state-of-the-art algorithms. IEEE International symposium on mixed and augmented reality (ISMAR), 2021.
[4] Thomas, Jerald, and Eva Suma Rosenberg. A general reactive algorithm for redirected walking using artificial potential functions. In IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 2019.

GEMINI 연동 가상 에이전트를 이용한 요리 교육 VR 프로그램*

이혜은[0], 이바다, 이채현, 김소희, 김선정[§]
한림대학교 콘텐츠IT전공, 융합소프트웨어학과
{hyesilver781, leebd188, cogusp1, kimsohee12360}@gmail.com, sunkim@hallym.ac.kr

Cooking Education VR Program Utilizing GEMINI-Integrated Virtual Agents

Hye-Eun Lee[0], Ba-Da Lee, Chae-Hyun Lee, So-Hui Kim, Sun-Jeong Kim[§]
Dept. of Content IT, Convergence Software, Hallym University

요약

본 논문에서는 가상 에이전트를 활용한 요리 교육 VR 프로그램을 제안한다. 동양인 모델 기반 가상 에이전트는 시각화 기술을 통해 가상환경의 적응을 돕고, 시각적 단서를 제공하여 사용자의 몰입과 이해를 돕는다. 또한, Python을 이용해 Google GEMINI와 연동된 가상 에이전트는 Google TTS 및 STT 기술을 통해 음성 질의응답 기능을 지원한다. 사용자는 가상 에이전트의 구두 설명에 따라 간단한 요리를 진행하고, 버튼을 눌러 질문하면 가상 에이전트가 적절한 답변을 제공한다. 이를 통해 사용자는 요리와 관련된 기본 지식에 대한 습득이 가능하며, 프로그램에 적극적으로 참여할 수 있고, 몰입감을 높일 수 있다.

1. 서론

서울시 먹거리 통계조사 보고서(2022)에 따르면, 서울시민의 성별 먹거리 이해력 점수는 여성 65.1점, 남성 58.7점으로, 남성이 6.4점 낮은 것으로 나타났다[1]. 먹거리 이해력이란 개인의 생명 유지 및 건강한 생활을 위한 먹거리 관련 역량을 포괄적으로 정의하는 개념이다. 이에 본 논문에서는 먹거리 이해력 향상을 목표로 가상 에이전트를 활용한 요리 교육 VR 프로그램을 제안한다. 가상 에이전트는 시각화 기술을 통해 구현되어 사용자가 가상환경에 더 쉽게 적응하도록 돕고, 표정과 몸짓 등의 시각적 단서를 제공하여 이해도를 높인다[2, 3, 4]. 이러한 접근 방식은 가상환경을 활용하여 요리를 포함한 다양한 일상생활 능력 교육에서 큰 도움이 될 것이며 추후 다양한 교육에도 활용될 수 있을 것으로 기대된다.

* 포스터 발표 논문
* 학부생 주저자 논문임.
* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00254695).

2. 본론

본 논문에서는 Unity3D와 Meta Quest 3를 사용하여 가상환경을 제작했으며, 사용자의 편의성을 극대화하기 위해 Hand Tracking 기술을 도입했다. 또한, 실제 물체를 스캔함으로써 현실과 가상환경의 이질감을 감소시키기 위하여 3D 스캐너인 Revopoint MIRACO를 사용했다.

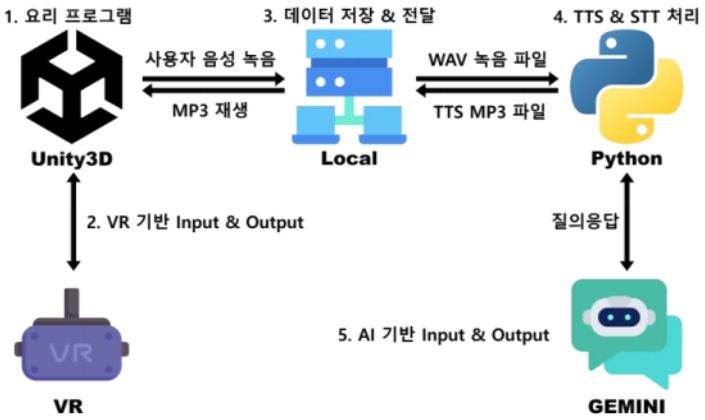


그림 1: 시스템 개요도

2.1. 가상 에이전트

사용자가 가상환경에 더 잘 몰입할 수 있도록 동양인 모델을 사용했다. Blender를 활용하여 가상 에이전트 모델의 손가락에 뼈대를 삽입하고, 얼굴에는 Blend Shape 기술을 적용하여 Unity에서 손가락 및 립싱크 애니메이션을 제작했다. 또한, 사용자가 가상환경 내에서 더 높은 사회적 실재감(Social Presence)을 느낄 수 있도록 Python을 통해 Google GEMINI를 연동하여 가상 에이전트와 상호작용이 가능하게 했고, 가상 에이전트와의 상호작용에서 사용자의 전반적 만족감에 긍정적인 영향을 효과적으로 미치기 위해 Google TTS(Text-to-Speech) 및 STT(Speech-to-Text) 기술을 이용해 가상 에이전트를 구현했다[5].

그림 2: 칼 잡는 법을 가르쳐주는 가상 에이전트

Unity3D에서는 사용자의 음성을 WAV 파일로 녹음하여 로컬 서버로 전달하고, 해당 파일을 Python에서 Google STT를 통해 텍스트로 변환한 후 GEMINI에게 질문을 전송한다. GEMINI의 답변은 Google TTS를 통해 MP3 파일로 변환되어 로컬 서버를 통해 Unity3D로 전달되고 재생된다. 이러한 과정은 사용자와 가상 에이전트 간의 상호작용을 원활하게 하여 사용자 경험을 향상시키고, 교육 효과를 극대화하는 데 기여한다[6].

2.2. 요리 교육 VR 프로그램

멀미를 줄이기 위해 움직임을 최소화할 수 있는 환경을 제공했다. 가상 에이전트는 사용자의 움직임에 따라 고정된 자리에서 몸의 방향을 돌려 사용자를 바라보도록 설계했다. 또한, 사용자의 몰입을 높이기 위해 3D 스캐너를 사용해 햇반을 스캔하고, 오브젝트로 만들어 활용했다.

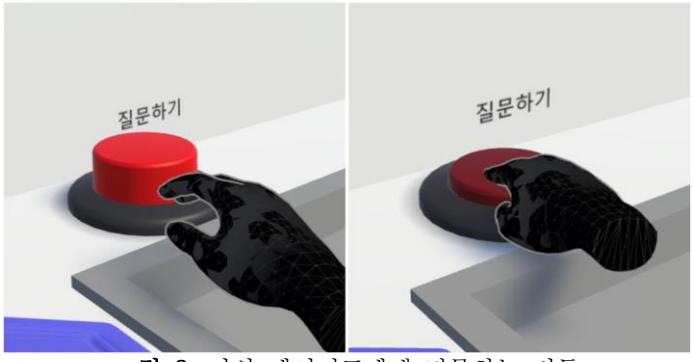


그림 3: 가상 에이전트에게 질문하는 버튼

사용자는 가상 에이전트의 구두 설명을 듣고 절차에 따라 간단한 요리를 진행한다. 전기레인지를 사용하여 계란프라이를 하거나 전자레인지에 햇반을 데우고, 야채를 씻고 써는 등의 기본적인 요리 지식을 습득할 수 있다. 또한, 사용자가 앞에 위치한 버튼을 누른 뒤 말을 하면 GEMINI와 연동된 가상 에이전트가 적절한 답변을 제공한다. 이로써 사용자는 가상 에이전트와의 질의응답을 매개체로 프로그램에 적극적인 참여가 가능하며, 이를 통해 몰입감을 높일 수 있다.

그림 4: 가상 에이전트의 설명에 따라 요리 진행

3. 결론

본 논문에서는 먹거리 이해력 향상을 목표로 가상 에이전트를 활용한 요리 교육 VR 프로그램을 제안한다. 가상 에이전트의 실체화를 통해 사용자들이 가상환경에 더 잘 적응할 수 있도록 하고, 표정이나 몸짓 등을 통해 시각적 단서를 제공하여 칼질 등의 기본적인 요리 방법에 대한 이해를 돕는다. 이를 토대로, 가상 에이전트를 활용한 VR 교육 방법이 향후 요리를 포함한 일상생활 능력 향상을 돕는 다양한 교육에도 적용될 수 있음을 제시한다.

향후 칼질과 같은 다양한 상호작용을 추가하여 본 프로그램이 사용자의 요리 능력 향상에 어떤 영향을 미치는지 검증하고, 음성과 텍스트 비교 및 가상 에이전트의 실체화 유무에 따른 교육 효과 등 가상 에이전트에 대한 연구를 계획하고 있다[7]. 이를 통해 가상 에이전트 기술이 교육 분야에서의 활용 가능성을 더욱 확대시킬 수 있을 것으로 기대된다.

참고문헌

[1] 2022 서울시 먹거리 통계조사 보고서, 서울특별시 식생활종합지원센터, pp.676, 2022
[2] 천은영, 가상현실 기반의 인공지능 영어회화 시스템, 한국융합학회논문지, 10(11):55-61, 2019
[3] Manuel Guimarães, Rui Prada, Pedro A. Santos, João Dias, Arnav Jhala and Samuel Mascarenhas, The Impact of Virtual Reality in the Social Presence of a Virtual Agent, IVA 2020, 23:1–8, 2020
[4] Issac Wang, Jesse Smith and Jaime Ruiz, Exploring Virtual Agents for Augmented Reality, CHI 2019, 281:1-12, 2019
[5] 이은지, 인간과 인공지능의 상호작용: 사용자, 기기, 환경적 측면의 상호작용 요인 중심으로, 2020
[6] Jiawen Chu, Recipe Bot: The Application of Conversational AI in Home Cooking Assistant, 2021 2nd ICBASE, pp.696-700, 2021
[7] Ivan Gris, Diego A. Rivera, Alex Rayon, Adriana Camacho and David Novick, Young Merlin: an embodied conversational agent in virtual reality, ICMI '16: Proceedings of the 18th ACM International Conference on Multimodal Interaction, pp.425-426, 2016

가상현실을 활용한 동물 해부 교육 콘텐츠 및 커리큘럼에 관한 연구*

장문수[0,1], 오병택[1], 백종훈[1], 변다솔[1], 김진모[1,*]
한성대학교 컴퓨터공학부[0,1]
{joey300, gnswhd0415, dennis99510, ekthf2013}@naver.com, jinmo.kim@hansung.ac.kr

A Study on Animal Anatomy Education Content and Curriculum using Virtual Reality

Moonsu Jang[0,1], Jonghun Baek[1], Byungtaek Oh[1], Dasol Byun[1], Jinmo Kim[1,*]
Division of Computer Engineering, Hansung University[1]

요약

본 연구는 동물보호법 시행규칙 개정으로 인하여 해부 실습이 불가능한 교육 환경의 문제를 해결하기 위한 하나의 대안으로 가상현실을 활용한 동물 해부 교육 콘텐츠를 제작한다. 그리고 동물 해부와 관련된 윤리문제를 안내하는 사전 교육, 가상현실을 활용한 실습 콘텐츠 교육, 마지막으로 퀴즈와 보고서로 구성된 사후 교육으로 구성된 커리큘럼을 새롭게 제안한다. 여기에, 가상현실 환경에서 교육을 지원하는 인공지능 에이전트를 함께 구현하여 교육 효과 향상에 도움을 줄 수 있도록 한다.

1. 서론

가상현실은 Meta Quest, HTC Vive를 비롯한 다양한 HMD(Head Mounted Display) 장비의 보급과 더불어 몰입감 높은 시각정보를 통해 현실과 유사한 경험을 제공하는 다양한 콘텐츠 제작에 활용되고 있다. 사용자는 현실 세계와 완전히 차단된 시각정보와 함께, 청각, 촉각 등의 여러 감각을 기반으로 가상의 환경을 체험한다. 이러한 기술의 발전과 더불어 가상환경 또는 객체와 몰입감 높은 상호작용은 물론 햅틱, 모션 플랫폼 등의 핵심 기술 연구와 다양한 응용 콘텐츠 제작 연구들이 다양한 관점에서 진행되고 있다. 이와 관련하여, 국내 전래동화 콘텐츠에 가상현실 기술을 적용하여 시각적 몰입감을 만족하게 하고 전래동화의 이야기를 전개하는 과정을 보다 동적으로 전달하기 위한 연구를 진행하였다[1]. 또한, 물리학의 전기장, 자기장의 개념을 3차원 가상 공간으로 확장하여 가상현실 기술을 적용한 교육 환경 및 학습 콘텐츠를 개발하고 학습 효과를 분석하는 연구를 수행하였다[2].

하지만, 본 연구는 전체 교육 과정을 가상현실 환경으로 제공하는 것은 가상현실 기술 경험이 부족한 사용자나 저연령의 학습자에게는 조작 또는 장비의 불편함, 시각적 어지러움, 낯선 교육방식 등의 이유로 가상현실 멀미를 유발하거나 교육 효과를 저해할 수 있다고 판단하였다.

따라서, 가상현실을 활용한 교육 콘텐츠를 전체 커리큘럼의 한 요소로 활용하여 교육 효과를 높일 수 있는 동물 해부 교육 콘텐츠 및 커리큘럼을 제안한다.

2. 동물 해부 교육 커리큘럼

2020년 3월 동물보호법 시행규칙 개정안 제24조의 2로 인하여 미성년자들의 해부실습이 금지되었다[3]. 하지만, 동물 해부실습을 통해 동물의 장기나 뼈, 근육 등을 직접 관찰하면서 생물학에 관한 지식을 배울 수 있으므로 필요하다고 판단하기도 한다. 이러한 이유로, 시청각 자료, 모형 등으로 수업을 진행하지만, 현실성이 떨어져 교육 효과는 감소하는 문제가 존재한다. 또한, 생명을 보호하고 권리를 존중해야 하는 윤리적 관점에서 해부실습에 대하여 부정적인 견해를 보이는 경우들도 다수 존재한다. 따라서, 본 연구는 두 가지 관점을 최대한 고려한 방식의 교육 커리큘럼을 제안하고자 한다.

2.1. 커리큘럼 구성

제안하는 커리큘럼은 사전 교육, 본 교육, 사후 교육의 과정으로 구성한다. 본 연구의 핵심은 가상현실 교육 콘텐츠만으로는 해부실습 교육이 가지는 문제를 해결하기 어렵다고 판단하기 때문에 교육자의 직접적 참여를 고려하는 커리큘럼을 구성한다.

우선, 동물보호법 시행규칙에 관한 내용과 윤리적 문제에 관한 주제를 해부실습에 관한 본 교육을 시작하기에 앞서 사전 교육으로 진행한다. 이는 가상현실에서 아바타를 통해 제공되는 것보다는 교육자(선생님 등)가 직접 전달하는 것이 정확하고, 명확할 수 있다. 따라서, 교육자가 동물 해부실습에 관한 윤리적 내용과 해부실습 외에 다른 목적(흥미, 재미 등)으로 사용하는 것을 금지한다는 내용을 사전 교육을 통해 진행한다. 본 교육은 해부실습 교육으로 현장감을 높여 과학적 지식을 학습하는 교육 효과를 높이기 위하여 가상현실을 활용한 교육 콘텐츠로 제작한다. 마지막은 사후 교육으로 가상현실을 통해 학습한 내용이 제공하는 교육적 효과에 대해서 정확히 검증하기 어려우므로 퀴즈와 보고서 등을 통해 확인할 수 있도록 한다. 또한, 단순히 가상현실 콘텐츠를 통해 신기하고 재미있는 게임적 요소로 활용되지 않도록 교육자가 주도하는 복습을 통해 수업을 완성할 수 있도록 커리큘럼을 제안한다.

---

* 포스터발표논문
* 학부생 주저자 논문임
* 본 논문은 요약논문 (Extended Abstract) 으로서, 한성대학교 컴퓨터공학부 캡스톤디자인에 발표된 콘텐츠 제작 결과를 토대로 정리한 연구 결과임
* 교신저자: 김진모(jinmo.kim@hansung.ac.kr)

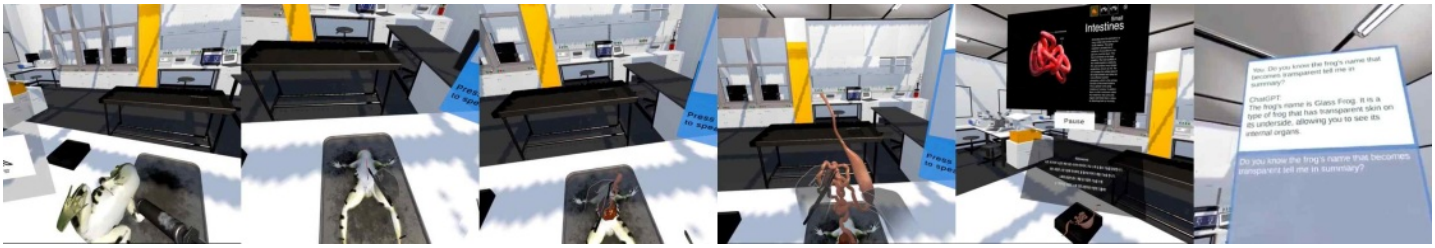


그림 1: 제안하는 가상현실을 활용한 동물 해부 교육 콘텐츠 제작 결과의 예

2.2 가상현실을 활용한 해부실습 콘텐츠

본 연구는 메타 퀘스트 2를 기반으로 가상현실에서 해부실습 교육을 진행하는 교육 콘텐츠를 제작한다. 제안하는 콘텐츠는 개구리 해부실습을 주제로 교육을 진행한다. 손 추적 상호작용 기능을 활용하여 손을 직접 사용하고, 주어진 도구(칼, 핀셋, 주사기 등)를 활용하여 해부실습을 단계적으로 수행한다. 사용자의 해부실습 상호작용은 크게 3가지로 구성된다. 다음은 제작된 개구리 해부실습의 과정을 정리한 것이다.

1단계) **해부**
a. 주사기를 사용하여 개구리를 마취 b. 고정핀을 사용하여 개구리를 고정 c. 칼을 사용하여 배를 보조선을 따라 해부
2단계) **관찰**
a. 핀셋을 사용하여 개구리의 장기를 관찰 b. 장기를 조작(확대 등)하여 다양하게 탐색
3단계) **정보**
a. 설명 도구를 통해 각 장기에 대한 교육자료를 글, 이미지, 동영상 등을 통해 제공

칼 또는 장기를 잡고 수행하는 행동을 포함한 모든 상호작용은 컨트롤러가 아닌 손을 직접 활용한다. 마지막으로, 본 교육은 가상현실에서 진행되는 환경이기 때문에 교육자가 필요한 정보를 직접 전달하기 어려울 수 있어 ChatGPT를 활용한 강사 에이전트를 함께 제공하여, 필요한 궁금증을 실시간으로 해결할 수 있도록 지원한다. 본 연구는 STT(Speech to Text) 기능을 통해 학습자는 음성으로 질의하고, 답변은 텍스트로 GUI를 통해 확인한다. 그림 1은 본 연구에서 제작한 해부실습 콘텐츠의 진행 과정을 나타낸 것이다.

2.3 학습

제안하는 동물 해부 교육에 관한 커리큘럼은 가상현실을 활용하여 몰입감 높은 해부실습을 제공하는 것뿐만 아니라, 교육의 질을 높이는 하나의 교육 과정을 제공하는 것이 핵심 목적이다. 따라서, 콘텐츠를 통해 학습한 내용에 대한 퀴즈 수행평가와 실험 보고서를 작성하는 사후 교육을 함께 제시한다. 그림 2는 본 연구에서 제시하는 사후 교육 프로그램의 일부를 보여주고 있다.

3. 결론

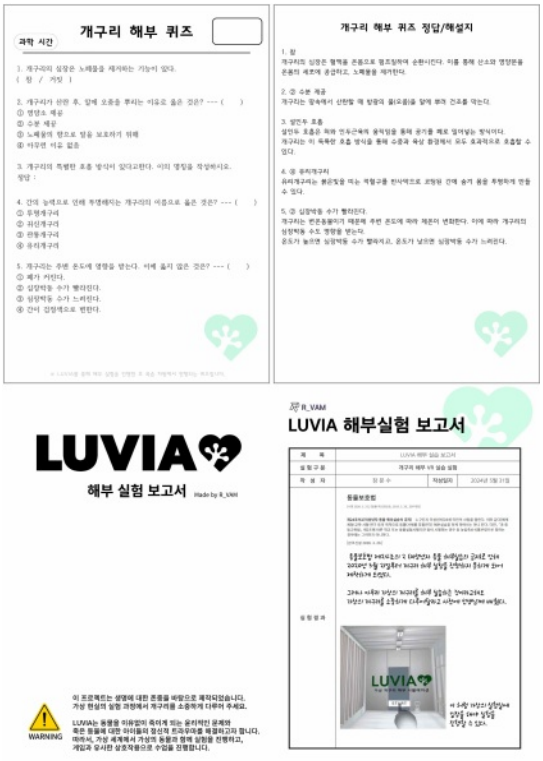


그림 2: 사후 교육을 위해 제작한 퀴즈 및 실험 보고서의 예

본 연구는 개정된 동물보호법에 따라 해부실습과 관련된 제한된 교육 환경을 개선함을 목적으로 가상현실을 활용한 동물 해부 교육 콘텐츠를 제작한다. 또한, 윤리적 문제의식, 해부실습에 관한 과학적 지식을 학습하는 교육 과정의 질적 향상을 모두 고려하기 위하여 사전 교육, 본 교육, 그리고 사후 교육으로 구성된 커리큘럼을 제안하였다.

참고문헌

[1] K. Jeong, S. Han, D. Lee and J. Kim, “A Study on Virtual Reality Techniques for Immersive Traditional Fairy Tale Contents Production,” *Journal of the Korea Computer Graphics Society*, vol. 22. no. 3 pp. 43-52, 2016.

[2] J. Lee and T. Lee, “Development of contents based on virtual environment of basic physics education,” *Journal of the Korea Computer Graphics Society*, vol. 29. no. 3 pp. 149-158, 2023.

[3] 동물보호법, [Internet] Available: https://casenote.kr/법령/동물보호법/제24조의2

VR 사용자 경험 강화를 위한 장착형 후각 장치 개선*

조인호[0], 박시연, 김선정
한림대학교 소프트웨어학부 콘텐츠IT 전공, 한림대학교 융합소프트웨어학과
{ zjxps2007, M23051, sunkim }@ hallym.ac.kr

Improvement of a wearable olfactory device to enhance VR user experience

Inho Jo[0], Siyeon Bak, Sun-Jeong Kim
Department of Convergence Software, Hallym University

요약

본 논문은 VR 환경에서 몰입감 높은 사용자 경험을 제공하기 위해 Meta3용 향기 장치를 제안한다. 기존의 Meta2용 향기 장치와 비교할 때, VR 환경의 사용자 경험 측면에서 정확도와 선택 시간면에서는 큰 차이가 없었으나, 다음과 같이 개선되었다. 첫째, 향기 장치의 무게를 크게 줄여 사용자의 착용감을 개선하였다. 둘째, 향기를 발생시키기 위해 돌아가는 팬(Fan)으로부터 느껴지던 바람을 약하게 하여, 바람의 방향을 인지하지 못하게 하였다.

1. 서론

기술의 급속한 발전과 함께 가상현실(VR) 분야에서도 혁신적인 기기들이 지속적으로 등장하고 있다. 햅틱 장치와 향기 장치[1]가 대표적으로, 이러한 발전은 사용자들로 하여금 더욱 높은 수준의 몰입도와 실감 나는 경험을 요구하게 만든다. 기존의 향기 장치는 정밀한 향기 제어와 배포를 통해 사용자에게 높은 몰입도와 실감 나는 경험을 제공해왔다. Meta3의 등장은 사용자들에게 향상된 경험을 제공할 수 있는 가능성을 열어주었다. 사용자 몰입도를 극대화하기 위해서는 Meta3에 최적화된 향기 장치의 개발이 필요하다. 본 논문에서는 Meta3 기반의 향기 장치를 제안하고자 한다.

2. 설계 및 구현

2.1. Meta3 기반의 향기 장치

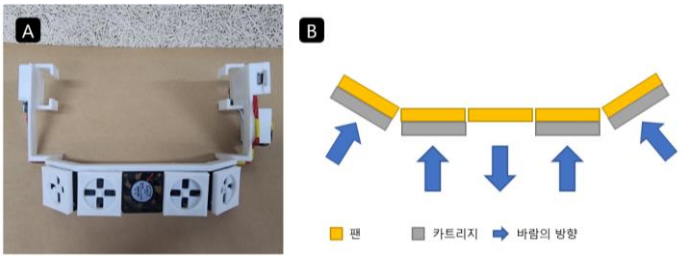


그림 1. Meta3용 향기 장치
(A - Meta3에 장착된 향기 장치 실물, B- 향기 장치의 구조)

본 연구에서 새롭게 제안하는 향기 장치는 기존 향기장치와 마찬가지로 HMD에 직접적으로 장착하며, 아두이노와 5V 팬과 3D프린터를 사용하여 제작된 몸체로 구성되어 있다. 최대 4개의 향기를 사용이 가능하도록 설계되었으며, 정면의 팬은 반대 방향으로 장착하여 향기 분출 후 남을 수 있는 잔향을 해결하였다[2]. 또한 무게 및 바람방향 인지문제를 해결하고자 30mm * 30mm의 팬을 사용하였다.

2.2. 향기 선택

본 연구에서 사용된 향기는 기존의 향기 장치와 개선된 향기 장치 간의 성능 비교를 위해 기존 연구[2]에서 사용한 향기인 사과, 포도, 수박, 오렌지향을 사용하였다.

3. Meta2용 향기 장치과의 비교

본 논문에서 사용한 실험은 기존의 향기 장치와 개선된 향기 장치 간의 성능 비교를 위해 기존 논문에서 사용했던 실험과 동일하게 진행하였다[2]. 실험 참가자들은 Simulator Sickness Questionnaire (SSQ)[3]와 System Usability Scale (SUS)[4] 및 User Experience Questionnaire (UEQ)[5] 설문을 진행하여 멀미와 사용자의 경험과 평가를 진행하였다. 실험 전 아래와 같은 가설을 준비하였다.

(가설 1) 기존 장치에 비해 가벼워 사용자의 느끼는 부담이 적을 것이다.

(가설 2) 팬의 크기가 작아졌지만 코와의 거리가 가까워져 사용자 경험에는 차이가 없을 것이다.

실험은 가설 검증 및 Meta3용 향기 장치의 유효성을 검증하기위해 실행하였다.

3.1. 실험

실험 환경은 창문이 없는 4.3m*2.9m*3m 크기의 방에서 성인 남녀 10명(남성=5, 여성=5명, age M=24.2, SD=3.12)을 대상으로 Meta2용 향기장치와 Meta3용 향기 장치로 진행하였다. 실험은 컨베이어 벨트에서 중앙으로 상자가 나오면 참가자가 박스를 터치했을 때 향기가 분출하여 참가자가 컨트롤러로 어떤 향기인지 맞추는 회기를 10회 시행하였다.

3.2 결과

실험결과 아래와 같은 결과를 얻었다.

---

* 포스터 발표논문
* 학부생 주저자 논문임

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00254695).

표 1 SUS 결과

Types	Device	M	SD	P	η_p^2
System Satisfaction	Meta2	12	2.9	.013	.131
	Meta3	16	1.9		
System Usability	Meta2	14.25	2.6	.177	.045
	Meta3	16.5	2.6		
System Effectiveness	Meta2	23	2.5	.671	.003
	Meta3	23.75	2		
System Attractiveness	Meta2	7.25	2.5	.638	.012
	Meta3	7.75	2.2		
System Reliability	Meta2	13.25	2.3	<.001	.186
	Meta3	17.25	1.3		

표 2 정답률, 소요시간 및 바람방향인지(WDP)

System	Duration (Second)	Accuracy (%)	WDP[1] (%)
Meta2 scent VR	11.33	51	70
Meta3 scent VR	12.97	56	56

표 3 가벼운 향기 장치 선택 및 향기 장치의 무게

System	Select (%)	Weight (g)
Meta2 scent VR	0	250
Meta3 scent VR	80	102
Both	20	N/A

표1은 향기 장치 각각의 사용자 평가결과를 나타낸다. 신뢰성과 만족감을 제외하고 기존 시스템과 유의미한 차이를 보이지 않는다. 하지만 사용자 평가 평균이 전체적으로 높아진 것과 그림2의 Novelty을 제외하고 평균이 올랐다는 것으로 보아 개선이 되었다는 것을 알 수 있다. 표2는 정답률과 소요시간, 바람을 인지했는지를 보여준다. Meta2 scent VR은 70%의 사용자가 바람을 인지하였으며, Meta3 scent VR은 56%의 사용자가 바람의 방향을 인지하였다. 정확도와 선택시간에는 큰 차이가 없었으며, 그림2의 UEQ설문 결과도 큰 차이가 없었다. 이는 기존 시스템과 성능과 사용자 경험에 차이가 없다는 것을 의미한다. 표3은 향기장치의 무게에 대한 사용자의 평가 및 장치의 무게를 보여준다. 80%의 사용자가 Meta3 scent VR을 선택했으며, 20%의 사용자가 같다고 응답했다. 이는 본 논문에서 제안하는 Meta3용 향기 장치가 실제로 가벼우며, 사용자가 사용하고 있을 때도 가볍게 느껴졌다는 의미이다. SSQ 설문은 사용자 모두 큰 차이를 보이지 않았다.

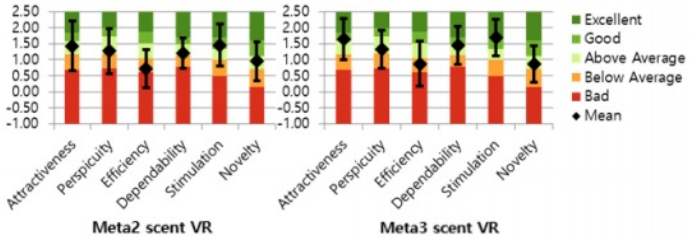


그림 2. UEQ 설문결과

[1] Wind direction perception

실험용 Map

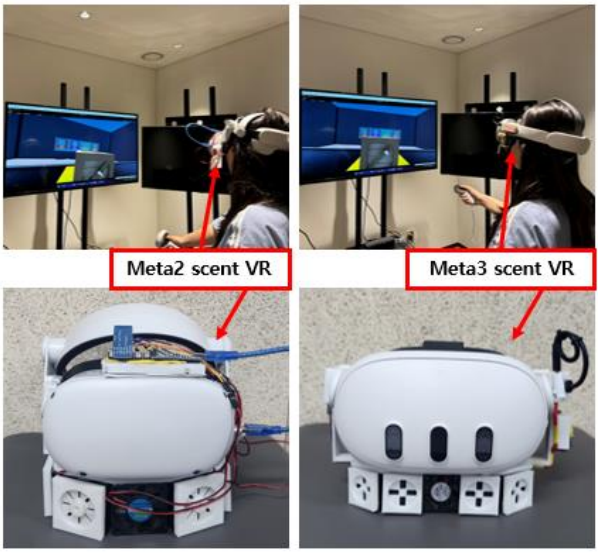


그림 3. 실험 환경

5. 결론 및 향후 연구

본 논문에서는 기존의 Mate2용 향기 장치를 개선한 Meta3용 향기 장치를 제안하였고 전반적으로 사용자의 경험을 향상시켰으나, 기존 장치와 비교했을 때 정확도와 선택 시간 면에서 큰 차이가 없음을 알 수 있었다.
추후 연구 방향으로 제안한 향기 장치를 사용하여 VR 환경에서 향기가 특정 환경에 대한 인지와 감정 변화에 미치는 영향을 평가할 예정이다.

참고문헌

[1]Ranasinghe, Nimesha, et al. "Season traveller: Multisensory narration for enhancing the virtual reality experience." Proceedings of the 2018 CHI conference on human factors in computing systems. 2018.
[2] Myung, Je-Won, et al. "Enhancing User Experience in VR using Wearable Olfactory System." Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology. 2023.
[3] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology, 3(3), 203-220.
[4] Brooke, J. (1996). SUS: A 'quick and dirty'usability scale. Thomas, BA We erdmeester and AL McClelland.
[5] Laugwitz, B., Schrepp, M. & Held, T. (2008). Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (Ed.): USAB 2008, LNCS 5298, pp. 63-76. Show on Research Gate

XR 환경 특수효과 제어를 위한 멀티모달 인터페이스

최종민[0], 김민채, 송오영[1]
세종대학교 소프트웨어학과, 고려대학교 컴퓨터학과, 세종대학교 소프트웨어학과
103solomon@sju.ac.kr, yjk199900@korea.ac.kr, oysong@sejong.edu

Multimodal Interface for Controlling Visual Effects in XR Environment

Jong-min Choi[0], Min-chae Kim, Oh-young Song[1]
Dept. of Software, Sejong University, Dept. of Computer Science and Engineering, Korea University,
Dept. of Software, Sejong University

요약

최근 XR(확장 현실)에 대한 관심이 높아지면서 관련 기술도 확대되고 있는데, XR은 기존의 마우스 및 키보드와는 다른 형식인 제스처와 같은 직관적 상호작용 방식을 요구한다. 본 연구에서는 XR 환경에서 다양한 특수효과를 효과적으로 제어하기 위해 제스처와 스피치를 결합한 멀티모달 인터페이스를 개발하였다. 제스처 인터페이스는 사용자의 손동작을 인식하여 직관적이고 시각적인 상호작용을 제공하고, 스피치 인터페이스는 음성 명령을 통한 제어를 가능하게 하는데, 본 연구에서는 두 인터페이스를 통합함으로써 XR 환경에서 직관적으로 특수효과의 세부 사항을 제어할 수 있도록 한다. 본 시스템은 Unity 엔진, Meta Quest Pro HMD, Oculus Integration SDK 등을 활용하였다.

1. 서론

최근 XR기술의 발전과 함께 기존의 컴퓨터에서 사용되었던 키보드, 마우스 등과는 다른 형식의 인터페이스 대한 필요성이 제기되고 있다. 이를 위해 Zhang, Fan, et al., Arora, Rahul, et al.[1,2] 등의 연구에서와 같이 제스처 인터페이스에 대한 연구가 진행되어 왔다. 특히, Arora, Rahul, et al.[2]에서는 특수효과 제어를 위한 제스처 활용법을 제시하였으나, 제스처 만으로는 특수효과의 세부적인 부분을 고려할 수 없었다. 본 연구에서는 XR 환경에서 특수효과를 직관적이고 효과적으로 제어할 수 있는 방법을 제안한다. 이를 위해 제스처와 스피치를 결합한 멀티모달 인터페이스를 개발하였다.

[1] 교신저자
* 포스터발표논문, 학부생 주저자 논문임
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT 연구센터육성지원사업(IITP-2024-RS-2022-00156354), 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원)사업(IITP-2024-RS-2023-00254529), 2024년 과학기술정통신부 및 정보통신기획평가원의 SW중심대학사업(2024-0-00037)의 연구결과로 수행되었음

2. 멀티모달 인터페이스 설계

2.1. 제스처 인터페이스

제스처 인터페이스는 사용자의 손동작을 실시간으로 인식하여 제스처에 해당하는 기능을 수행함으로써 XR 환경의 객체 및 효과를 제어한다. 본 시스템에서는 한 손당 손목을 포함한 손가락 관절 24개 벡터 정보를 수집하여 필요한 제스처를 사전에 저장할 수 있도록 하였다. 이 때, 각 손가락 관절의 위치 정보는 손목을 기준으로 한 상대 위치로 기록된다. 이 후, 녹화된 제스처와 사용자 입력 제스처에서 대응되는 관절 사이의 위치 차이를 오차로 하여 오차의 합산치가 임계값보다 낮을 시 동일한 제스처로 판단한다.

2.2. 스피치 인터페이스

스피치 인터페이스는 사용자의 음성 명령을 인식하여 텍스트로 변환한 후, 미리 정의된 프롬프트에 따라 적절한 작업을 실행한다. 본 시스템에서는 OpenAI의 Whisper API[3]를 사용하여 음성을 텍스트로 변환(STT)하고, ChatGPT API를 활용하여 명령을 해석하였다.

2.3. 멀티모달 통합

제스처는 특수효과의 선택, 크기, 방향과 같은 공통적인 속성에 대하여 직관적이고 시각적인 상호작용을 제공한다. 그러나 비의 세기, 연기의 밀도 등과 같은 고유 속성 조절을 위해 각 속성마다 다른 제스처를 부여하는 것은 상호작용의 복잡도를 증가시킬 수 있다. 스피치는 즉각적 피드백은 부족하나, 자연어를 통하여 특수효과의 고유 속성에 대한 자유롭고 세부적인 접근과 제어가 가능하다. 따라서 두 인터페이스를 통합함으로써 제스처의 직관성을 유지하면서 스피치를 통해 세부적인 조절을 가능하게 할 수 있다. 또한 제스처와 스피치 인터페이스의 충돌을 피하기 위해, 특정 제스

처를 취했을 때만 스피치 인식이 활성화되도록 하였다.

기능	제스처/스피치
생성/증강	스피치
삭제	제스처/스피치
선택	제스처
회전	제스처
크기	제스처
고유 속성	스피치
모션	제스처

표 1 제스처, 스피치 역할 구분

3. 시스템 구현 및 결과

3.1. 하드웨어 및 소프트웨어 환경

본 시스템은 Unity 엔진을 기반으로 개발되었으며, Meta Quest Pro HMD와 Oculus Integration SDK를 활용하였다.

3.2. 제스처 인식

사용자는 왼손 검지로 가리키는 제스처를 취함으로써 광선을 생성하여 해당 광선이 가리키는 객체/특수효과를 선택할 수 있다. 해당 객체/특수효과가 선택된 상태에서 사용자의 오른손 제스처에 따라 선택된 객체/특수효과의 속성(회전, 크기 등)을 변경할 수 있다. 예를 들어 그림 1과 같이 오른손을 편 상태에서 손을 회전시키면 비가 내리는 방향이 변한다.

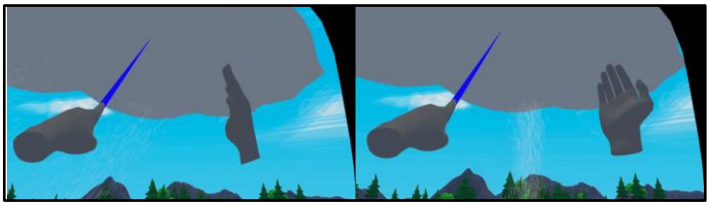


그림 1: 제스처를 활용하여 비의 방향 회전 수행

또한, 그림 2와 같이 오른손 엄지 방향과 함께 손이 움직이는 속도를 계산하여 객체의 움직임을 제어할 수 있다. 엄지 방향이 공의 방향을 결정하고, 속력이 공의 속력을 결정한다.

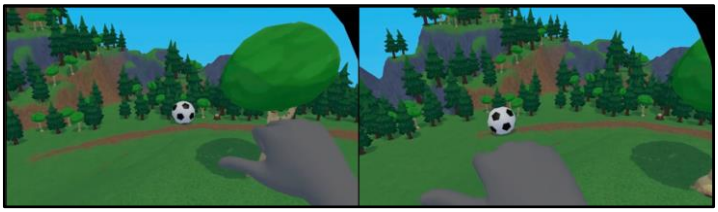


그림 2: 제스처의 속력에 비례하여 공의 물리량 변화

3.3. 스피치 명령 해석

녹음 시작 제스처를 취하면 HMD 마이크를 통해 음성 명령이 입력된다. 입력된 음성은 Whisper API를 거쳐 텍스트로 변환되고, ChatGPT API로 전송되어 미리 정의된 프롬프트에 따라 해석된다. 예를 들어 “Start fire”라는 명령으로 불 특수효과를 증강시킬 수 있으며, "Heavy rain is falling"과 같은 명령에 따라 비의 파티클 수를 증가시킬 수 있다.

그림 3: 스피치를 이용한 불 특수효과의 증강

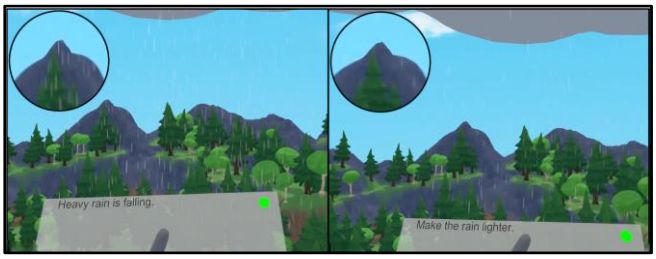


그림 4: 스피치를 이용한 비의 세기 조절

4. 결론

본 연구에서는 XR 환경에서 특수 효과를 직관적이고 효과적으로 제어할 수 있는 멀티모달 인터페이스에 대한 연구를 진행하였으며, 제스처와 스피치 인터페이스를 통합함으로써 사용자는 가장 적절한 방식으로 XR 환경을 직관적이고 효율적으로 제어할 수 있다. 향후 연구에서는 제스처의 다양성을 확대하고 특수효과 별 추가적인 음성 명령을 지원하는 데 중점을 둘 것이다. 또한, 제스처와 음성 인식의 정확도를 향상시킴으로써 XR 환경에서 특수효과 제어에 대한 편리성을 확보할 예정이다. 이를 통해 XR 환경에서 보다 원활한 상호작용을 제공할 수 있을 것으로 기대된다.

참고문헌

[1] Arora, Rahul, et al. "Magicalhands: Mid-air hand gestures for animating in vr." *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 2019.
[2] Zhang, Fan, et al. "Double hand-gesture interaction for walk-through in VR environment." *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. IEEE, 2017.Society Press.
[3] https://whisperapi.com/

F-RDW: 미래 위치 예측을 통한 방향 전환 보행*

전상빈[0], 정재호, 박진형, 이인권
연세대학교 컴퓨터과학과
{ludens0508, gkgkgkwogh, jh9604, iklee}@yonsei.ac.kr

F-RDW: Redirected Walking with Forecasting Future Position

Sang-Bin Jeon[0], Jaeho Jung, Jinhyung Park, In-Kwon Lee
Dept. of Computer Science, Yonsei University

요약

기존의 예측형 방향전환보행 방법들은 미래 정보를 활용하여 reset의 수를 줄인다. 그러나 이러한 방법들은 가상 환경의 레이아웃이나 사용자의 보행 방향에 대한 전제를 반드시 요구하여 보편적인 적용에 제약을 준다. 이 문제를 해결하기 위해, 우리는 두 단계로 구성된 머신러닝 기반 예측형 방향전환보행 메커니즘을 제안한다: (1) 사용자의 공간 및 시선 데이터로 학습된 예측 모델을 사용하여 가상 공간에서 사용자의 미래 정보를 어떠한 가정 없이 예측하고, (2) 이 정보를 기존 RDW 알고리즘에 조합하는 것이다. 시뮬레이션 테스트와 사용자 연구 결과, 작은 물리적 공간이나 복잡한 환경에서 RDW를 사용할 때 미래 정보가 더욱 중요함을 확인하였다. 우리는 제안된 메커니즘이 reset의 수를 크게 줄이고 reset 발생 사이의 이동 거리를 증가시켜, 이 연구에서 탐구한 모든 RDW 알고리즘의 성능을 향상시킴을 증명하였다.

1. 서론

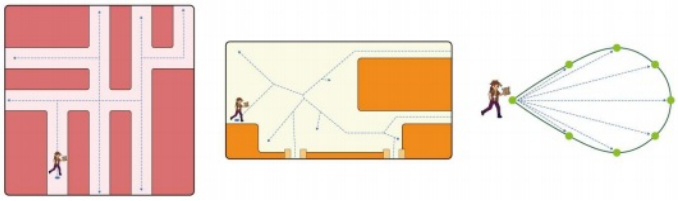


그림 1: 기존 연구의 한계 (보행 영역과 방향에 대한 임의의 가정들)

방향전환보행 (Redirected Walking: RDW) 기술은 동시에 걷기로 유한한 물리 공간에서 가상 공간을 무한히 탐험하게 해주는 기술이다 [1]. RDW는 teleportation, joystick 등의 다른 가상공간 이동 방법들보다 사용자에게 높은 현존감과 몰입감을 제공한다 [2]. 한편, 미래 정보 예측 기반 RDW 방법들은 사용자의 미래 정보를 바탕으로 방향전환을 조절하여, 몰입감을 떨어뜨리는 reset 사용 횟수를 줄이는 것을 목표로 한다.

* 포스터 발표논문
* 본 논문은 요약논문 (Extended Abstract)으로서, 본 논문의 원 논문은 IEEE Transactions on Visualization and Computer Graphics (2024) 에 게재되었음.
* 이 연구는 정부 (과학기술정보통신부)의 재원으로 한국연구재단 (No. RS-2024-00348094) 및 한국전파진흥협회 (No. RNIX20230200)의 지원을 받아 수행된 연구임.

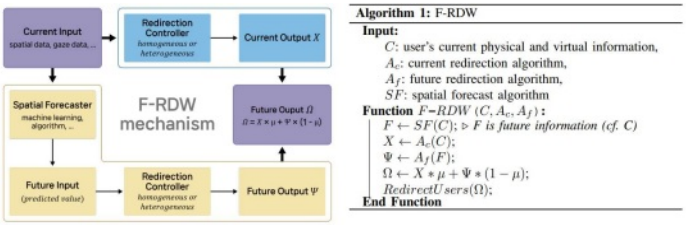


그림 2: 제안하는 메커니즘의 오버뷰 및 알고리즘

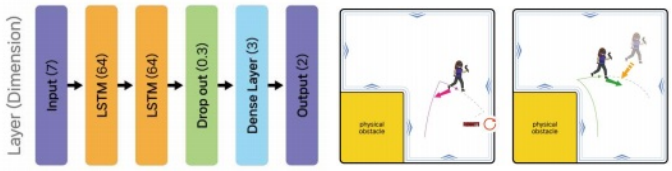


그림 3: 미래 정보 예측에 활용된 네트워크 구성과 미래 가상 정보가 반영되어 방향전환보행 물리 경로가 변경되는 예시

이러한 기존 접근법들은 그림 1과 같이 미리 정의된 가상 보행 경로 그래프를 사용하나, 사용자의 보행 방향에 대한 선제적 가정을 필요로 한다. 이것은 무한한 가상공간의 사용성 및 응용력을 떨어뜨리고 미래 예측 정확도를 보장하지 못하여 RDW 성능에 부정적 요인이 된다.

2. 방법

우리는 기존 연구의 한계를 해소하기 위해, 두 단계로 구성된 머신러닝을 이용한 예측 기반 방향전환보행 메커니즘을 제안한다 (그림 2).

2.1. 미래 가상 보행 위치 예측

첫 번째 단계로, 우리는 매 trial마다 미로와 열린 공간으로 구성되는 20x20m의 가상 공간에서 가상 타겟을 수집하는 콘텐츠를 구성하였다. 이 환경에서 10명의 실험참가자로부터 수집된 spatial locomotion data와 eye tracking data를 이용하여, 사용자의 n초 뒤 (ex. 1초 뒤) 미래 가상 위치를 예측를 예측하는 네트워크 (그림 3)를 학습시킨다. 이때, 관련 연구 [3]를 기반으로 하여 구현한 결과, 1초 뒤 가상 위치 예측 정확도는 77%이다.

2.2. 미래 정보와 기존 RDW 알고리즘의 조화

두 번째 단계로, 예측된 미래 가상 위치를 활용할 수 있

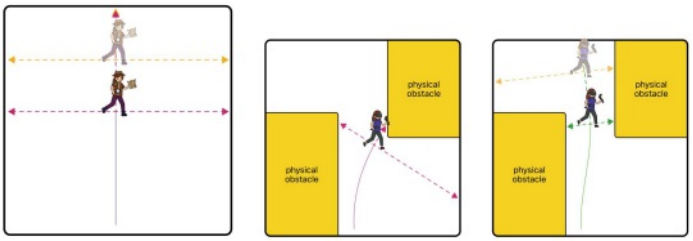


그림 4: 제안된 메커니즘이 적용된 F-ARC (b)는 미래 가상 위치를 반영하여 spatial align을 계산, 사용자를 방향전환보행 시킨다.

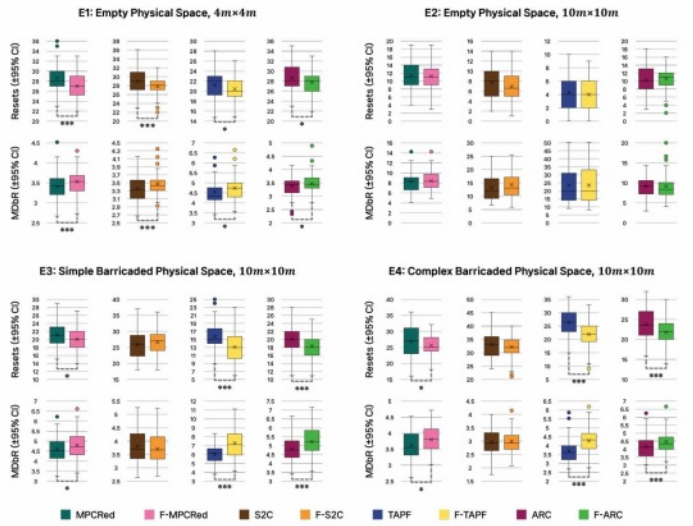


그림 5 각 환경 조건 별 바닐라 RDW 알고리즘과 제안된 메커니즘이 적용된 F-RDW의 시뮬레이션 실험 결과

도록 기존에 제안된 반응형 RDW 알고리즘 (S2C, APF, ARC) 혹은 예측형 RDW 알고리즘을 변형하고 이를 적용한다. 즉 현재 사용자의 위치와 미래 사용자 위치 각각에 적합한 해당 RDW 알고리즘의 결과들을 weighted sum하여 최종적으로 방향전환 값을 산출한다 (그림 2).

3. 실험 및 결과

우리는 F-RDW 메커니즘의 성능을 평가하기 위해 주어진 개별 크기 공간의 빈 공간 혹은 장애물이 존재하는 물리 공간을 셋팅하고, F-RDW 메커니즘의 적용 유무를 독립변인으로, reset의 수와 reset 발생 사이의 이동거리를 종속변인으로 하여 시뮬레이션 테스트를 수행하였다. 통계분석 결과, 그림 5와 같이 F-RDW 메커니즘이 적용되어 유의미한 방향전환 효과을 얻을 수 있음을 확인하였다. 결과에서 얻을 수 있는 인사이트로는 비교적 작은 물리적 공간이나 복잡한 환경에서 RDW를 사용할 때 미래 정보가 더욱 중요하다는 것이다. 이것은 여유 물리 공간이 없는 상황에서 미래 정보의 활용이 큰 영향을 미친다는 것을 의미한다.

한편, 우리는 사용자 가상 위치 예측 단계에서 머신러닝 기반의 예측이 임의의 보행 방향 가정보다 얼마나 우세한지, 몇 초 뒤를 예측할지, 현재/미래에 대한 F-RDW의 가중치는 각 알고리즘마다 어떻게 다른지에 대한 시뮬레이션 테스트를 수행하였다 (그림 6).

또한, 우리는 시뮬레이션 테스트를 확장하여 28명의 실험참가자와 함께 사용자 실험을 수행하였고, 실적용에서도 F-RDW가 유의미한 RDW 퍼포먼스 향상을 얻을

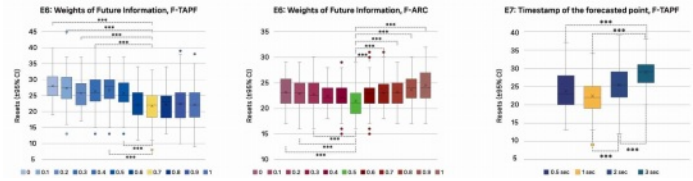


그림 6 제안된 알고리즘에 대한 ablation study 결과

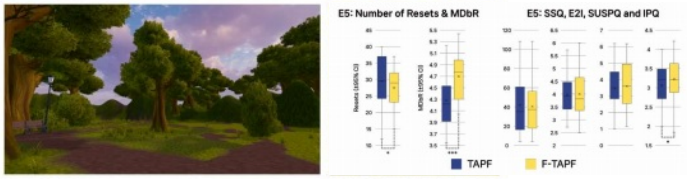


그림 7 제안하는 메커니즘에 대한 사용자 실험에 사용된 가상환경과 실험 결과 그래프

수 있음을 그림 7과 같이 증명하였다.

4. 결론

사용자의 미래 정보를 활용하는 예측형 RDW 방법은 사용자의 몰입감과 현존감을 저하시키는 reset 발생 횟수를 줄여준다. 그러나 미래 정보의 예측 정확도가 낮으면 방향전환 성능이 떨어지며, 가능한 미래 공간 정보의 거의 무한한 수를 고려하는 데 드는 막대한 계산 비용 때문에 실시간 제공이 불가능하다. 이러한 문제를 우회하기 위해, 이전 연구들은 사전 정의된 가상 보행 경로 그래프나 사용자의 보행 방향을 가정하는 추가적인 제약을 두게 되었고, 이는 예측형 RDW 방법의 일반화를 저해시켰다. 이러한 문제를 보다 근본적으로 해결하기 위해, 우리는 기존과 같은 가정이나 제약 없는 예측형 방향전환 메커니즘 (F-RDW)을 제안한다. F-RDW는 공간 및 시선 데이터 기반의 머신러닝 모델을 통해 사용자의 가상 공간 미래 위치를 예측함으로써 어떤 제약이나 가정에 독립적이며, 예측된 값을 기존에 연구된 RDW 방법들에 반영하면서 그들의 내부 방향전환 아이덴티티를 존중한다. 우리의 시뮬레이션 테스트와 사용자 연구 결과, F-RDW가 작은 물리적 공간이나 장애물이 있는 복잡한 환경에서 기존 RDW 방법의 방향전환 성능을 향상시킨다는 것을 증명하였다. 한계점이나 미래 연구로서, 가상 위치 예측을 위해 더 다양한 실험참가자로부터 다양한 학습 데이터 수집하여 분석하고 예측 모델을 보강할 수 있다. 또한, 제안하는 메커니즘이 작동하기 어려운 reset 상황에 대해 개선 방법을 고안할 필요가 있다.

참고문헌

[1] Razzaque, S., et al. Redirected walking. Eurographics 2001 - Short Presentations. 2001.
[2] Nilsson, N., et al. 15 years of research on redirected walking in immersive virtual environments. IEEE computer graphics and applications 38(2):44-56, 2018.
[3] Bremer, G., et al. Predicting future position from natural walking and eye movements with machine learning," in 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). 2021.

무요소법을 활용한 KSTAR 중성입자빔가열 유발 플라즈마 대면 요소 열속 해석 및 가시화*

문태욱[0], 윤의성
울산과학기술원 원자력공학과
{esyoon}@unist.ac.kr

Analysis and Visualization of Heat Flux on Plasma-Facing Components Induced by Neutral Beam Heating in KSTAR Using the Meshfree Method

Taeuk Moon[0], Eisung Yoon
Dept. of Nuclear Engineering, Ulsan National Institute of Science and Technology

요약

한국형 핵융합 에너지 연구 개발 장치인 KSTAR (Korea Superconducting Tokamak Advanced Research)의 운전 중 조사된 중성이온빔 입자 소실(shine through)와 조사에 의해 발생하는 고에너지의 하전 입자의 소실(fast ion loss)에 의한 장치 내벽 열속의 분포를 분석 및 가시화하는 접근법에 대해 소개한다. 기존에는 KSTAR 장치 내벽의 컴퓨터 지원 설계 (computer-aided design) 모델을 기반으로 만든 삼각형 격자(mesh) 요소(element)별로 열속의 평균을 구했다면, 본 연구에서는 삼각형의 꼭짓점 개체(entity)만을 활용하여 무요소법(meshfree method)으로 해석과 시각화를 수행한다.

1. 서론

한국형 핵융합 에너지 연구 개발 장치인 KSTAR (Korea Superconducting Tokamak Advanced Research)는 운전 중 플라즈마 가열을 위해 중성이온빔을 장치 내부로 조사한다. 해당 빔은 플라즈마와 반응하여 고에너지의 하전 입자를 만드는데, 중성이온빔 입자가 플라즈마와 반응하지 않고 장치 내벽에 부딪히거나 고에너지 하전 입자가 KSTAR 장치 내부에 가두어지지 않고 이탈하여 내벽에 충돌하면 해당 부분에 열속이 가해진다. 이러한 열속이 장치 내부 취약한 부분에 국소적으로 집중되면 손상이 발생하여 장치 수명을 단축시킬 수 있다.

* 포스터 발표논문
* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국과학기술정보연구원의 지원을 받아 수행된 연구임(No. KSC-2021-CRE-0504).

이를 방지하고자 열속의 분포를 추산하기 위한 다양한 시도가 있었으며 [1, 2, 3], 본 연구에서는 NuBDeC 코드[1]의 열속 시뮬레이션 결과를 분석 및 가시화하는 접근법을 소개한다. 기존에는 KSTAR 장치 내벽의 CAD 모델을 기반으로 삼각형 격자 요소별로 열속의 평균을 구했으나, 본 연구에서는 삼각형의 꼭짓점 개체만을 활용한 무요소법으로 해석과 시각화를 수행한다.

2. 무요소법 활용 중성이온빔 유발 입자 소실 분석 및 가시화

2.1. CAD 모델 기반 KSTAR 내벽 격자 생성

KSTAR 장치의 CAD 설계 데이터를 활용하여 삼각형 격자를 생성하고, 해당 격자와 NuBDeC 코드에서 모사된 입자의 경로 사이의 교차 지점을 3차원 공간상에서 계산하여 충돌 지점을 파악한다 [4].

2.2. 무요소법 활용 하전입자 소실 열속 분석

무요소법에서는 수치해석을 위해 격자 요소를 필요로 하지 않는다. 요소를 따로 생성하지 않고 노드만 배치하며, 노드들 사이의 공간은 노드에서의 값으로부터 내삽을 통해 값을 얻는다. 기존의 격자를 활용한 방법에서는 격자 요소별로 충돌한 하전입자의 에너지를 총합하여 격자 요소의 넓이로 나누어 요소별 평균 입자속을 계산했으나, 무요소법을 활용한 방법에서는 KSTAR 내벽 격자 위에서 충돌 지점으로부터 사용자가 지정한 영향 영역(influence domain) 반경 내에 있는 격자 노드들에 대해, 격자 표면 위에서 최단거리를 계산하여 최단거리의 길이만큼 지수적으로 감쇄하는 열속값을 기록한다.

2.3. 입자 소실 열속 가시화

격자의 노드별로 열속 계산이 완료되면, 이들 사이

표면에서의 열속값은 노드에서의 값을 선형적으로 내삽하여 가시화한다(Figure 1).

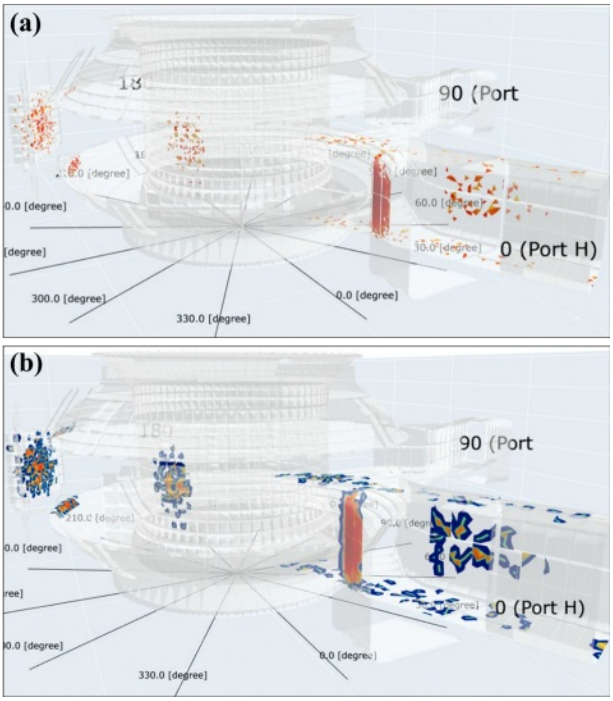


Figure 1. (a) 삼각형 격자 요소별 열속 분포 가시화, (b) 무요소법 활용 열속 분포 가시화 (24,000개 표본 활용 시뮬레이션 결과)

3. 무요소법 활용 시 입자 소실 분석 및 가시화 효과

NuBDeC 코드 시뮬레이션 결과로 계산된 입자 소실 분포는 몬테카를로 방법을 활용한 입자 샘플링으로 계산된 것이다. 그리고 해당 소실 분포 및 그 열속값은 표본 입자의 개수가 증가할수록 통계적 오차가 줄어 특정 분포 및 값으로 수렴하는 경향이 있다. 기존 삼각형 요소별 열속 계산 방식에서는 열속값이 요소의 넓이에 대한 소실된 입자 개수의 평균값으로 계산되므로, 요소의 넓이가 지나치게 작을 경우 통계적 오차가 증폭될 가능성이 있다. 그래서 일부 매우 작은 요소들의 경우, 열속 계산 시 집계된 오차가 해당 요소에서 큰 열속이 발생한 것처럼 보일 수 있다. 따라서 국소적인 열속 집중 부분의 위치를 파악하기 위해 시뮬레이션을 반복할 경우, 충분히 많은 표본 입자를 동원하지 않는 이상 일관되지 않은 결과가 도출되는 경우가 많다. 무요소법은 요소의 넓이가 아닌 충돌 지점으로부터 떨어진 거리에 따라 열속이 결정되므로, 기존 방법에서 오차가 증폭되는 문제를 완화할 수 있다.

각 분석 방법별로 얼마나 통계적 오차를 줄일 수 있는지 확인하기 위해 표본 입자가 12,000개, 24,000개일 경우를 비교하였다. 각 경우가 서로 얼마나 비슷한 열속 분포를 계산하였는지, 즉 얼마나 시뮬레이션 결과가 수렴에 가까워졌는지 비교하였다. 삼각형 요소별 평균 열속을 비교한 결과를 보면($\sqrt{\frac{\sum_i(\Gamma_{i,24k}-\Gamma_{i,12k})^2}{\sum_i \Gamma_{i,24k}{}^2}} \cong 0.411$) 삼각형 꼭짓점별로 무요소법으로 계산된 열속들($\sqrt{\frac{\sum_i(\sum_j \Gamma_{ij,24k}-\sum_j \Gamma_{ij,12k})^2}{\sum_i(\sum_j \Gamma_{ij,24k})^2}} \cong 0.112$)보다 표본 수가 2배로 증가하였을 때 값이 크게 변화하는 경향을 보였다.

Figure 3에서는, 분석 방법별 상대오차의 제곱 즉, 삼각형 요소별 평균은 $\left(\frac{\Gamma_{i,12k}-\Gamma_{i,24k}}{\Gamma_{i,24k}}\right)^2$, 꼭짓점별 무요소법은 $\left(\frac{\Gamma_{ij,12k}-\Gamma_{ij,24k}}{\Gamma_{ij,24k}}\right)^2$ 을, Γ_i 와 Γ_{ij} 의 값이 큰 순서대로 왼쪽부터 오른쪽으로 정렬하였다. 이는 무요소법을 사용할 때 최댓값의 변화가 낮아짐, 즉 수렴에 가까워진 모습을 보인다.

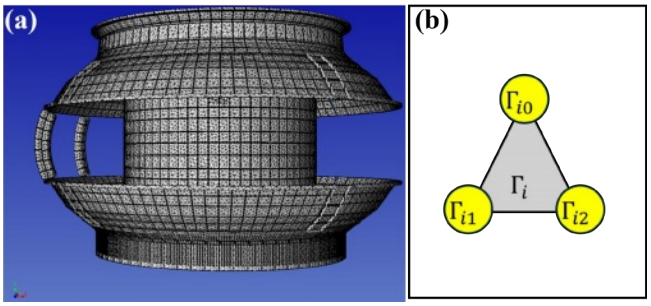


Figure 2. (a) NuBDeC 코드에 활용된 삼각형 격자, (b) i번째 삼각형 격자 요소에 대한 분석 방법별 열속 분포 (Γ_i: 삼각형 격자 요소별 평균 열속값, Γ_{ij}: 해당 삼각형의 j번째 꼭짓점에서 무요소법으로 계산된 열속값)

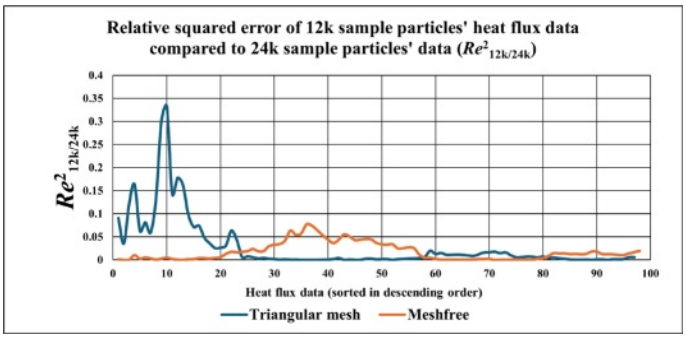


Figure 3. 24,000개 표본 입자 활용 열속 데이터에 대한 12,000개 표본 입자 활용 열속 데이터의 상대오차 제곱 (Re^2) 분포

4. 결론

요소별 평균이라는 해석 방법에서 작은 삼각형 요소의 통계적 오차 증폭 문제를 해결하기 위해 무요소법을 활용한 해석 및 가시화 방법론을 제안한다. 이 방법론은 요소의 넓이에 구애받지 않고 중성입자빔 유발 고에너지 하전입자의 KSTAR 장치 내벽 상 열속 분포에서 효과적으로 국소적 열속 집중 지점을 판별하는 데 도움을 준다. 이 기술은 KSTAR 장치 내

중성이온빔의 설계 및 운영에 기여할 수 있다.

참고문헌

[1] T.Rhee, K.Shinohara, H.S.Kim, J.M.Kwon, Y.M.Jeon, J.Kim, ... and B.Na, *Numerical study of neutral beam injection prompt loss at KSTAR poloidal limiters*. *Physics of Plasmas*, 26(11), 2019

[2] T.Rhee, J.Kim, K.Kim, J.Kang, M.Podestà, Y.S.Park, and B.H.Park, *Simulation study of fast ion losses associated with the rotating n= 1 resonant magnetic perturbations in KSTAR*. *Nuclear Fusion*, 62(6), 066028, 2022

[3] K.Kim, H.Jhang, J.Kim, and T.Rhee, *Enhanced fast ion prompt loss due to resonant magnetic perturbations in KSTAR*. *Physics of Plasmas*, 25(12), 2018

[4] T.Moon, T.Rhee, J.M.Kwon, and E.Yoon, *Development of Novel Collision Detection Algorithms for the Estimation of Fast Ion Losses in Tokamak Fusion Device*. Available at SSRN 4537156.

ScanToSafeGuard: 공간 객체 인식 AI 기술을 활용한 건설 현장의 안전 감시 시스템

신승엽[12○] 박형진[4] 서명배[4] 우운택[123*]

[1]PMRC, [2]KI-ITC ARRC, [3]UVR Lab KAIST
[4]한국건설기술연구원

{youb1649, wwoo}@kaist.ac.kr, {parkhyungjin, smb}@kict.re.kr

ScanToSafeGuard: A Construction Site Safety Monitoring System Using Spatial Object Recognition AI Technology

Seungyoub Ssin[12○] Hyungjin Park[4] Myoungbae Seo[4] Woontack Woo[123*]

[1]PMRC, [2]KI-ITC ARRC, [3]UVR Lab KAIST
[4]KOREA INSTITUTE of CIVIL ENGINEERING and BUILDING TECHNOLOGY

요 약

건설 현장은 다양한 위험 요소를 내포하고 있으며, 기존의 수작업 안전 관리 방법으로는 이러한 복잡한 위험 요소를 효과적으로 다루기 어렵다. 본 연구에서는 증강 현실(AR) 기술과 실시간 객체 인식 기술을 결합한 ScanToSafeGuard 시스템을 제안한다. 이 시스템은 HoloLens 2, ScanToCAD, Omni3D, Cube R-CNN을 사용하여 실시간 데이터 캡처와 정밀한 객체 탐지를 통해 작업자의 안전을 향상시키고 사고율을 줄이는 것을 목표로 한다. ScanToSafeGuard는 건설 현장의 객체 인식 기반 디지털 트윈을 생성하여 지속적인 모니터링과 즉각적인 위험 요소 식별을 가능하게 하며, 기존의 수작업 안전 관리 시스템의 한계를 극복할 수 있다.

Abstract

Construction sites inherently contain various hazards, and traditional manual safety management methods struggle to effectively address these complex risks. This study proposes the ScanToSafeGuard system, which combines augmented reality (AR) technology with real-time object recognition. This system utilizes HoloLens 2, ScanToCAD, Omni3D, and Cube R-CNN to capture real-time data and perform precise object detection, aiming to enhance worker safety and reduce accident rates. ScanToSafeGuard creates a digital twin of the construction site based on object recognition, enabling continuous monitoring and immediate identification of hazards, thereby overcoming the limitations of traditional manual safety management systems.

키워드: 공간 객체 인식, 실시간 모니터링, 3D 객체 탐지, 건설 현장 안전, 증강 현실
Keywords: Spatial Object Recognition, Real-Time Monitoring, 3D Object Detection, Construction Site Safety, Augmented Reality

1 서론

건설 현장은 본질적으로 다양한 위험 요소를 내포하며, 기존의 수작업 안전 관리 방법은 이러한 복잡한 위험 요소들을 효과적으로 다루기에는 한계가 있다. 비약적으로 발전하고 있는 실시간 객체 인식 기술은 증강현실(AR) 기술과 결합하여 이러한 문제를 개선함으로써 더 정확하고 신속한 위험 감지와 대응이 가능해질 수 있다(1). 본 연구에서는 AR과 AI를 결합한 ScanToSafeGuard 시스템을 제안한다. 이 시스템은 작업자의 안전을 향상시키고 사

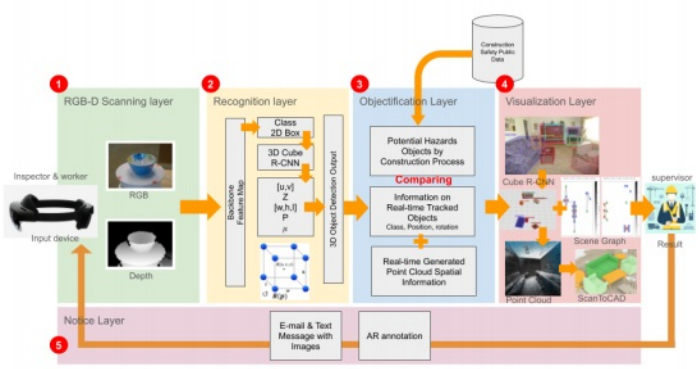


Figure 1: 1번에서 5번까지 시스템 내부의 여러 레이어들의 관계 및 흐름도

고율을 줄이는 것을 목표로 한다. ScanToSafeGuard는 HoloLens 2, ScanToCAD(2), Omni3D(3)을 사용하여 실시간 데이터 캡처와 정밀한 객체 탐지 기능을 이용한다. 기존의 수작업 안전 관리 시스템과 달리, 이를 통해 건설 현장의 객체 인식 기반 디지털 트윈을 생성하여 지속적인 모니터링과 즉각적인 위험 요소 식별을 가능하게 한다.

2 시스템 레이어 구성

본 시스템은 관리 영역별로 팀별 작업을 예상하고 작업 레이어의 개념으로 구성된다.

- **1번 레이어(RGB-D Scanning Layer)**: 첫 번째 레이어는 실내 공간을 스캐닝하여 3D 포인트 클라우드를 생성하는 초기 작업을 수행한다. RGB-D로 각종 객체들을 스캐닝하는 작업이 포함되며, 이는 ScanToCAD에서 ScanNet의 역할을 담당한다.
- **2번 레이어(Recognition Layer)**: 두 번째 레이어는 객체를 인식하는 작업을 수행하며, RGB-D로 생성된 포인트 클라우드나 RGB 이미지를 사용한다. ScanToCAD 방식에서는 ShapeNet의 역할을 하며, Omni3D에서는 RGB 이미지 셋을 사용하여 객체를 인식한다. 3차원 인식 AI Convolution Neural Network(CNN)을 통해 객체를 분류하고, 최종적으로 인식된 객체의 종류와 9Dof(position, rotation, scaling) 정보를 얻게 된다.
- **3번 레이어(Objectification Layer)**: 세 번째 레이어는 공공 데이터(예: 건설공사 안전관리 정보망)와 최종 인식된 객체 리스트를 비교하여 현재 작업 과정에서 있어서는 안 되는 객체들이나 가까운 거리에 서로 존재하면 위험할 수 있는 객체들을 인식한다. 이 데이터들은 시간의 흐름에 따라 순서대로 관리되며, 과거의 상황을 언제라도 다시 볼 수 있는 시간 단위 기반의 디지털 트윈 시스템으로 제공된다.
- **4번 레이어(Visualization Layer)**: 네 번째 레이어는 Objectification Layer에서 얻은 정보를 시각화하는 역할을 한다. 관리자가 필터링된 데이터를 통해 신속한 판단을 할 수 있도록 돕는 정보 가시화 부분이다. 포인트 클라우드 상의 인식된 객체를 출력하는 뷰, 실제 인식된 객체를 CAD 모델로 변환하여 보여주는 뷰, 그리고 상관관계를 동시에 보여주는 씬 그래프 뷰 등이 포함된다.
- **5번 레이어(Notice Layer)**: 마지막 레이어는 관리자가 확인한 사항을 현장 작업자에게 전달하는 과정을 담당한다. 본 시스템에서는 AR 방식으로 문자나 이미지를 메시지 형식으로 전달하는 방식을 제안한다. 작업 현장의 스캔된 상황 및 위험 요소에 해당하는 이미지를 보고 현장 작업자는 빠른 대처를 할 수 있다.

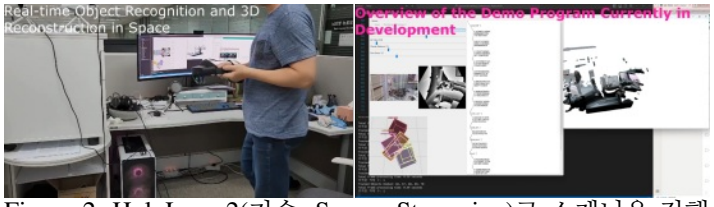


Figure 2: HoloLens 2(기술: Sensor Streaming)로 스케닝을 진행하는 모습(좌)과 실시간 객체 인식을 통해 객체 정보가 생성되고, 함께 표시되는 포인트 클라우드 생성 모습(우)

3 결론

본 논문에서는 건설 현장의 안전 관리를 위해 실시간 객체 인식과 증강 현실(AR) 기술을 검토였고, 제안한 ScanToSafeGuard 시스템은 HoloLens 2, ScanToCAD, Omni3D, Cube R-CNN을 통합하여 실시간 데이터 캡처와 정밀한 객체 탐지로 작업자의 안전을 향상시키는 것을 목표로 한다.

감사의 글

이 논문은 2023년 정부(과학기술정보통신부)의 재원으로 국가과학기술연구회 오픈XR(OXR) 융합연구사업(No. CRC21011)의 지원을 받아 수행된 연구임

References

[1] H. Nassereddine, A. S. Hanna, D. Veeramani, and W. Lotfallah, "Augmented reality in the construction industry: Use-cases, benefits, obstacles, and future trends," *Frontiers in Built Environment*, vol. 8, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbuil.2022.730094

[2] A. Avetisyan, M. Dahnert, A. Dai, M. Savva, A. X. Chang, and M. Nießner, "Scan2cad: Learning cad model alignment in rgb-d scans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2614–2623.

[3] G. Brazil, A. Kumar, J. Straub, N. Ravi, J. Johnson, and G. Gkioxari, "Omni3d: A large benchmark and model for 3d object detection in the wild," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 13 154–13 164.

한국컴퓨터그래픽스학회 2024 학술대회 논문집

발행인 최수미

편집인 이성희

발행처 사단법인 한국컴퓨터그래픽스학회

주소 05006 서울특별시 광진구 능동로 209 (군자동) 세종대학교 광개토관 1011호

전화 02-6935-2539

URL http://www.cg-korea.org/

KCGS 2024
Generative Imagination

CLO

OSSTEM IMPLANT

BARO
Bring AI Right ON

OpenXR
플랫폼 융합연구단

세종대학교 초실감 XR 연구센터
EXTENDED REALITY RESERCH CENTER

숭실대학교 메타버스융합대학원
Graduate School of Metaverse

IITP 정보통신기획평가원
Institute of Information & Communications Technology Planning & Evaluation

Simulated Reality
Ewha ITRC

KOREA UNIVERSITY 고려대학교 ICT명품인재양성사업단
Korea University ICT Creative Consilience Foundation

BK21 FOUR 세종대학교 BK21 지능형드론 교육연구단

세종대학교 메타버스 융합대학원

MOVIN

VIZINF creative technology

박찬모 교수
평양과기대 명예총장