

KOREA  
COMPUTER  
GRAPHICS  
SOCIETY  
한국컴퓨터그래픽스학회

# PAINTING THE METAVERSE

[cg-korea.org](http://cg-korea.org)

KCGS 2021  
학술대회논문집

# 한국컴퓨터그래픽스학회

# 2021 학술대회 논문집

2021년 7월 6일 ~ 2021년 7월 8일

온라인 학술대회

주관기관: 한국컴퓨터그래픽스학회

# KCGS 2021 환영사

두 번째 온라인 학술대회에 오신 것을 환영합니다. 작년에는 처음이다 보니 하나부터 열까지 서툴렀다면, 올해는 조금 익숙해진 느낌입니다. 준비 과정에서 여러 선택지 중에서 고민하는 시간은 최소화하였고, 대신 빠르게 현실적인 방향을 선택한 후 내실을 갖추고자 했습니다. 올해 학회의 가장 큰 변화 중 하나는 등록비를 완전히 없앤 것입니다. 온라인 학회의 비용 절감 덕분이기도 하지만, 무엇보다 기꺼이 후원해주신 기업들의 도움이 컸습니다. 후원 기업인 엔씨소프트, 바로AI, 클로 여러분에게 감사의 마음을 전합니다. 무료 등록을 통해 우리 학회의 저변이 한층 확대되기를 기대해봅니다. 작년과 달라진 또 하나의 변화로 “여름학교”의 복귀를 들 수 있습니다. 소중한 학습의 장이 되어준 여름학교를 작년 한 해 쉬었던 부분이 적잖게 아쉬웠습니다. 다행히 올해에는 어려운 여건에도 기꺼이 여름학교를 맡아주신 교수님들 덕분에 Computational Photography라는 흥미로운 연구 분야를 일목요연하게 배워볼 수 있는 기회가 생겼습니다.

학술대회의 중심에는 언제나 그렇듯 최신 연구결과를 교류하는 논문 발표가 자리합니다. 학술대회 기간 동안 6개 분과로 나누어서 총 33편의 논문이 발표될 예정입니다. 작년과 마찬가지로 실시간 온라인 회의 플랫폼을 통해 미리 녹화된 발표 영상을 순차적으로 공개하고, 이후 좌장의 주재 하에 질의/응답을 진행합니다. 많은 질문을 던진 참가자 분들 중 “질문 왕”을 선정하여 마지막 날에 경품을 시상할 예정이니, 학술대회 홈페이지에서 경품 내역과 선정 기준을 한 번쯤 확인해 둘 것을 권해 드립니다. 중간 휴식시간에는 이번 학술대회에서 특별히 준비한 Gather.town에 둘러보시기 바랍니다. 기존 학술대회처럼 발표자와 청중들이 자연스럽게 어우러져 대화를 나눌 수 있는 메타버스 공간을 마련해 두었습니다. 아쉽게도 올해에는 포스터 발표는 진행하지 않습니다. 작년의 경험에 비추어 볼 때 낮은 참여율과 실질적인 운영의 한계를 극복하기 어려울 것으로 예상하여 내린 결정이니 양해를 부탁드립니다.

초청강연과 교수/박사연구원급 발표는 우리 학술대회에서 놓치지 말아야 할 주요 행사입니다. 한국기술교육대학교의 김상연 교수님은 “From Metaverse to Haptaverse”라는 주제로 최근 각광 받고 있는 메타버스의 몰입감을 극대화 할 수 있는 햅틱 기술에 대해 소개해 주실 예정입니다. 또한 Victoria University of Wellington의 이태현 교수님은 “Cinematic XR – Teleport into the Video”라는 주제로 미래형 미디어 융복합 기술 및 연구 내용을 공유하는 자리를 마련해 주실 예정입니다. 지구 반대편 뉴질랜드에서 실시간으로 초청강연을 진행해 주시는 것 자체가 이미 눈앞에 다가온 미래를 보여주는 사례가 아닐까 싶습니다. 교수/박사연구원급 발표 세션에서는 세 분의 교수님이 최근 국제적으로 인정받은 연구 성과를 발표해 주실 예정입니다. 컴퓨터그래픽스 분야의 최신 연구 흐름을 한 눈에 확인할 수 있는 자리가 될 것으로 기대합니다.

이번 학술대회의 여러 발표와 이벤트에서 반복적으로 발견하게 될 “메타버스”라는 키워드는 코로나 시대 혹은 그 이후의 사회 변화를 단적으로 나타내는 용어로 자리 잡는 듯합니다. 작년에 처음 온라인 학술대회를 준비할 때에는 컴퓨터그래픽스 연구 분야가 포스트 코로나 시대의 핵심 기술로 부상할 것이라는 조심스러운 예측을 한 적이 있는데 이제는 어느덧 명백한 현실이 된 것 같습니다. 올해 학술대회 슬로건인 “Painting the Metaverse” 역시

바로 그러한 변화의 시기에 우리 컴퓨터그래픽스 분야 연구자들이 적극적으로 역량을 발휘하여, 코로나 이후 사회에서 중요한 역할을 감당하기를 기대하는 마음이 반영된 것입니다. 모쪼록 이번 학술대회가 국내 컴퓨터그래픽스 분야 연구에 새로운 활기를 불어넣고 연구자들 간의 교류와 화합을 도모할 수 있기를 기대합니다. 모든 참가자 여러분들의 건강과 행복을 기원합니다.

감사합니다.

2021년 6월 25일  
KCGS 2021 학술대회 조직위원장 이강훈, 황재인

# 학술대회 조직위원회

대회장 이제희 (서울대학교)

조직위원장 이강훈 (광운대학교), 황재인 (KIST)

프로그램위원장 윤성의 (KAIST), 이성길 (성균관대학교)

여름학교 준비위원장 신현준 (아주대학교)

수상위원회 위원장 김영준 (이화여자대학교)

조직위원 김승원 (전남대), 성민혁 (KAIST), 이명호 (부산대), 이윤상 (한양대),  
이지은 (한성대), 조성현 (포항공대), 최명걸 (가톨릭대)

프로그램위원 홍헬렌 (서울여자대학교), 김구진 (경북대학교), 손봉수 (중앙대학교),  
경민호(아주대학교), 이정진 (숭실대학교), 한정현 (고려대학교),  
오경수 (숭실대학교), 허재필 (성균관대학교), 김덕수 (한국기술교육대학교),  
하종성 (우석대학교), 김영민 (서울대학교), 이영호 (목포대학교),  
권구주 (배화여자대학교), 김선정 (한림대학교), 최장환 (이화여자대학교),  
정원기 (고려대학교), 노준용 (KAIST), 이종원 (세종대학교),  
박경주 (중앙대학교), 최명걸 (가톨릭대학교), 이성희 (KAIST),  
정문열 (서강대학교), 이정 (고려대학교), 김종민 (강원대학교),  
권태수 (한양대학교), 김민혁 (KAIST), 김정현 (고려대학교),  
김준호 (국민대학교), 김익재 (KIST), 김진모 (한성대학교),  
조성현 (POSTECH), 백승환 (Princeton University), 고성안 (UNIST),  
이주행 (ETRI), 김형석 (동의대학교), 김영봉 (부경대학교),  
이인권 (연세대학교), 장윤 (세종대학교), 남승우 (ETRI),  
송창근 (한림대학교), 이환용 (아주대학교), 송오영 (세종대학교),  
최정주 (아주대학교), 안상철 (KIST), 박진호 (숭실대학교),  
최민규 (광운대학교), 이윤진 (아주대학교), 이창하 (중앙대학교),  
이화자 (공주대학교), 한다성 (한동대학교), 윤승현 (동국대학교),  
박진아 (KAIST), 백낙훈 (경북대학교), 김형석 (건국대학교)

# 목 차

---

## 초청 강연

From Metaverse to Haptaverse 김상연 (한국기술교육대학교) .....	2021.7.6(화) 14:00-14:50 1
Cinematic XR - Teleport into the Video 이태현 (Victoria University of Wellington, New Zealand) .....	2021.7.7(수) 11:00-11:50 2

---

## 석사논문상 후보자 발표

장원종 (POSTECH) .....	2021.7.7(수) 14:00-14:30 3
---------------------	------------------------------

---

## 특별 논문 발표: 교수/박사연구원급 발표

Learning to Deform: 3D Content Creation via Mesh Deformation .....	2021.7.7(수) 12:50-13:50 4 성민혁(KAIST)
Mixed Reality Tabletop Gameplay: Social Interaction with a Virtual Human Capable of Physical Influence .....	5 이명호(부산대학교), Nahal Norouzi, Gerd Bruder, Pamela Wisniewski, Greg Welch(University of Central Florida)
Deep Combiner for Independent and Correlated Pixel Estimates .....	6 백종희(광주과학기술원), Binh-Son Hua(VinAI Research, VinUniversity), Toshiya Hachisuka(The University of Tokyo), 문보창(광주과학기술원)

---

## 논문 발표 1: 이미지/비디오/렌더링

딥러닝기반 실내와 실외 환경에서의 광원 추출 .....	2021.7.6(화) 15:05-16:35 특별호 이지원, 서광균, 이하늬, 유정은, 노준용(KAIST)
초박형 라이트필드 카메라의 실시간 분해능 향상 알고리즘 개발 .....	특별호 김동건(한국과학기술원), 유재관, 조용진(LIG 넥스원), 김민혁(한국과학기술원)
기하적 변형이 포함된 이미지의 GAN Inversion .....	7 강경국, 김성태, 조성현(POSTECH)
광선추적 GPU를 이용한 확률적 볼륨매핑 가속화 .....	9 민혜정, 한경민, 김영준(이화여대)
음성 입력으로부터 사실적인 말하는 영상 생성 .....	11 주광진, 장원종, 이승용(포항공과대학교)
단노출 플래시 스마트폰 영상에서 저속 동조 영상 생성 .....	특별호 이종협, 조성현, 이승용(포항공과대학교)

---

**논문 발표 2: 애니메이션/시뮬레이션 I**

2021.7.6(화) 16:50-18:05

물리 시뮬레이션 및 모션캡처 데이터의 물체 및 관절 사이의 관계 인식 .....	13
이도해, 오영진, 이인권(연세대학교)	
PointNet을 이용한 연체 동역학의 실시간 예측 .....	15
박진형, 이도해, 이인권(연세대학교)	
보정-보간 기반 연기 시뮬레이션 .....	17
이윤지, 오영진, 이도해, 이인권(연세대학교)	
심층 강화학습을 활용한 단일 강체 캐릭터의 모션 생성 .....	특별호
안제원, 구태홍, 권태수(한양대학교)	
저가형 모션 캡처 장비를 이용한 실시간 상호작용 애니메이션 시스템 .....	19
김정호, 강다운, 권태수(한양대학교)	

---

**논문 발표 3: 애니메이션/시뮬레이션 II**

2021.7.7(수) 09:30-10:45

자연스러운 피격 반응 생성 .....	21
박수환, 이제희(서울대학교)	
하나의 모션 클립으로부터 모션 집단 학습 .....	23
이세영, 이선민, 이용우, 이제희(서울대학교)	
사람의 운동을 이용한 사족 보행 로봇 제어기 .....	25
김선우(서울대학교), 막심 소로킨, 하세훈(조지아공과대학교), 이제희(서울대학교)	
병적 보행 환자의 모션 캡처 데이터로부터 근골격 모델 추정 .....	27
박범수, 민세희, 박정남, 이제희(서울대학교)	
마커를 이용한 관절 연체 모델 변형의 IK 계산 가속화 .....	29
박현아, 강다운, 권태수(한양대학교)	

---

**논문 발표 4: 가상/증강현실**

2021.7.7(수) 14:45-16:15

가상 실내 환경을 위한 변화맹 기반 무한공간탐험 .....	31
황준영, 조용훈, 권순욱, 이인권(연세대학교)	
고정된 위치에서의 Reset과 블록 이어 붙이기를 이용한 무한 가상 공간 탐험 .....	34
권순욱, 황준영, 조용훈, 이인권(연세대학교)	
무한공간탐험을 위한 옆걸음 및 뒷걸음의 인지 한계점 측정 .....	36
조용훈, 이인권(연세대학교)	
모바일 가상현실에서의 이동 인터페이스에 관한 연구 .....	특별호
홍승현, 나기리, 조윤식, 김진모(한성대학교)	
Unity MARS를 이용한 증강현실 기반 볼 특수효과 콘텐츠 개발 .....	38
김보성, 송오영(세종대학교)	

Interactive Wind System을 이용한 VR 사이버 멀미 개선 연구 ..... 특별호  
임도전, 이예원, 조예솔, 류태동, 한다성(한동대학교)

---

**논문 발표 5: 모델링/그래픽스응용**

2021.7.8(목) 09:30-11:00

환자 특이적 수술 내비게이션 개발 .....	40
한예진, 홍기윤, 김성재, 성낙준((주)휴툰)	
모바일 매니플레이터 로봇을 이용한 TSP 펜아트 .....	42
임은정, 박지윤, 송다은, 김영준(이화여자대학교)	
사람의 EOS X-ray 스캔으로부터 표면 메시 추정 .....	44
장필식, 이제희(서울대학교)	
맞춤형 의료기기 제작을 위한 클라우드기반 협업서비스 개발 .....	46
정영진, 김태한, 서안나((주)시안솔루션)	
GPU를 이용한 삼각형 집합의 근사적 외부 경계 계산 .....	48
최규진, 경민호(아주대학교)	
3차원 형상 복원을 위한 점진적 점유 예측 네트워크 .....	특별호
김용규, 김덕수(한국기술교육대학교)	

---

**논문 발표 6: 애니메이션 III**

2021.7.8(목) 11:15-12:30

물리적인 상호작용을 위한 유연한 휴머노이드 제어와 시뮬레이션 .....	50
이승환, 이제희(서울대학교)	
상태 표현 방식에 따른 심층 강화 학습 기반 캐릭터 제어기의 학습 성능 비교 .....	52
손채준, 이윤상(한양대학교)	
단일 보행 패턴으로부터 신체 조건 유추 .....	57
민세희, 이제희(서울대학교)	
2D 이동 경로 지정을 통한 모션 매칭 기반 캐릭터 모션 생성 .....	59
이정민, 이윤상(한양대학교)	
Constrained Projective Dynamics: 에너지와 운동량을 보존하는 실시간 물리 기반 변형체 시뮬레이션 .....	61
기민형(고려대학교), 엄기원(텔레콤파리대학), 정우석, 한정현(고려대학교)	



초청강연

## From Metaverse to Haptaverse

김상연 교수

한국기술교육대학교 컴퓨터공학부

<http://lilab.koreatech.ac.kr>

### 강연 내용

최근들어 다시 주목 받고 있는 메타버스 (Metaverse) 플랫폼/시스템에서는 주로 사람의 오감 중 주로 시·청각 정보만을 제공하는 형태로 개발되고 있어 메타버스 환경이 제공하는 다양한 상황 및 감각을 사용자에게 온전히 전달하기 어렵다. 메타버스 환경에서 사용자의 몰입감을 증대시키기 위하여 햅틱컨트롤러를 사용해 사용자에게 오감 중 촉감을 제공하고자 하는 연구가 진행되고 있으나, 기 개발된 햅틱컨트롤러의 경우 단순한 진동 감각만을 전달하거나 물체의 강성 혹은 재질감 등 한 가지 감각만을 전달할 수 있어 사용자에게 충분한 몰입감을 제공하기에는 부족하다. 또한 하나의 컨트롤러를 이용하여 다양한 콘텐츠를 제어할 때 메타버스환경에서는 사용자가 잡고 있는 툴이 한 개가 아니고 상황에 따라서 달라지게 되므로, 컨트롤러의 형태 및 부피가 바뀌고, 컨트롤러의 무게 중심 등이 바뀌어야 한다. 본 강연에서는 메타버스환경, 가상/증강현실의 상황에 따라 컨트롤러의 형태 및 부피가 자동으로 변화되어 도구 종류에 관계없이 사용자가 가상의 도구 (탁구채, 테니스라켓, 야구방망이, 골프채 등)를 잡는 느낌을 그대로 모사 할 수 있으며, 콘텐츠 상황에 따라서 컨트롤러의 무게 중심이 변화되어 마치 실제 환경에서 사용자가 도구를 잡고 움직일 때 발생하는 운동감을 가상 환경에서 실감나게 표현할 수 있는 신개념의 "형상변형이 가능한 햅틱컨트롤러"를 소개한다.

### 강연자 이력

2006.03 ~ 현재 : 한국기술교육대학교 컴퓨터공학부 교수

~2006년 2월 삼성종합기술원 책임연구원

2004년 2월 : KAIST 박사졸업

1997년 2월 ~ 1999년 2월: 통신연구소 근무

1997년 2월 : KAIST 석사졸업

1995년 2월 : 고려대학교 학사졸업

초청강연

# Cinematic XR - Teleport into the Video

이태현 교수

Director, Computational Media Innovation Centre

Victoria University of Wellington, New Zealand

## 강연 내용

청정 자연을 배경으로 한 특수 촬영, 영화 및 광고 특수효과 기술 등으로 잘 알려진 뉴질랜드의 수도 웰링턴에서 벌어지고 있는 컴퓨터 그래픽스 관련 산학연 인프라 구축 사례를 소개하고, 이를 기반으로 진행된 미래형 미디어 융복합 기술 및 연구 내용을 공유하는 자리를 마련하고자 합니다. 특히 Immersive (몰입형), Interactive (대화형), Intelligent (지능형) 미디어 기술 및 발전 동향을 정리하고, 이들의 융복합을 통해 구축 가능한 신개념 Cinematic XR 기술 및 플랫폼을 소개합니다. 이를 활용하여 제작된 응용 사례, 그리고 비디오 속 사물과 상호 작용이 가능한 몰입형, 대화형 콘텐츠를 소개합니다. 나아가, 비대면 환경에 적용 가능한 가상 순간이동 (Augmented Virtual Teleportation), 혼합 현실을 이용한 원격 협업 시스템으로의 확장 방안을 제시하고, 미래의 영화 및 미디어 산업과의 연계 방안, 활용 가능성 등을 논의하는 자리를 갖고자 합니다.

## 강연자 이력

이태현 교수는 뉴질랜드 수도 웰링턴 소재 빅토리아 대학교 (Victoria University of Wellington), 엔지니어링 학부 교수로 재직 중이며, 2018년 설립한 산학 연구 센터인 Computational Media Innovation Centre (CMIC) 센터장을 역임하고 있다. 2012년 Weta Digital과 함께 빅토리아 대학 컴퓨터 그래픽스 과정을 신설하였고, 2014년 빅토리아 그래픽스 연구실을 설립하였다. 2017년, 혼합 현실 스타트업 DreamFlux를 창업, 뉴질랜드 정부에서 연구자에게 수여하는 창업가 상을 수상하였다.

미국 서던 캘리포니아 대학 (University of Southern California)에서 컴퓨터 그래픽스 박사학위를 취득했으며, 미래형 대화형 미디어 기술 및 서비스 플랫폼; 영화/게임 등에 적용 가능한 실시간 조명, 실사 렌더링, 자동 합성 기술; Cinematic XR; 혼합 현실을 이용한 몰입형 비대면 화상 통화 및 협업 시스템; 사람과 디지털 콘텐츠 간 상호 작용 관련 연구 과제들을 수행 중이다.

Pacific Graphics 2020, 2021의 Conference Chair, Asia Graphics Association의 Executive Committee로 활동 중이며, 뉴질랜드 이주 전에는, 삼성 종합 기술원 수석연구원 (2008-12), 삼성전자 책임연구원/연구원 (1996-2003)으로 재직하였다.

# 2021 석사논문상 후보

**장원종 (Wonjong Jang)** wonjong@postech.ac.kr

장원종은 2019년 2월에 포항공과대학교 컴퓨터공학과 대학원에 석사과정으로 입학했으며 이승용 교수의 지도하에 석사학위를 취득하고 현재 박사과정 1학기를 이수하였다.

이 학생은 얼굴 영상의 생성 및 조작에 관한 연구에 관심을 가지고 있다. 학부때부터 이승용 교수의 지도 아래에서 기하적인 얼굴 미화 방법과 실제 세계의 얼굴 영상 초해상도 기법에 관한 연구를 수행하였으며, 석사과정 기간에는 2차원 및 3차원 캐리커처의 생성에 관한 연구에 참여하였다.



그 중 2차원 캐리커처 생성에 관한 연구는 기존 캐리커처 생성 연구들이 2차원 이미지 와핑으로 얼굴의 과장을 모델링하던 통념에서 벗어나 사실적인 캐리커처 생성을 위해 StyleGAN의 coarse-level feature map을 modulation하는 방법을 제안하였다. 이 연구 결과는 SIGGRAPH 2021에서 "StyleCariGAN: Caricature Generation via StyleGAN Feature Map Modulation"이라는 제목으로 발표될 예정이다.



그림 1: StyleCariGAN의 캐리커처 생성 예시

교수/박사연구원급 발표

# Learning to Deform: 3D Content Creation via Mesh Deformation

성민혁

KAIST

본 발표는 SIGGRAPH ASIA 2020, CVPR 2021, ECCV 2020에 발표되었음

## **DeformSyncNet: Deformation Transfer via Synchronized Shape Deformation Spaces**

Minhyuk Sung<sup>1\*</sup>, Zhenyu Jiang<sup>2\*</sup>, Panos Achlioptas<sup>3</sup>, Niloy Mitra<sup>4</sup>, and Leonidas Guibas<sup>3</sup> (\*equal contribution)

<sup>1</sup>KAIST, <sup>2</sup>University of Texas at Austin, <sup>3</sup>Stanford University, <sup>4</sup>UCL/Adobe Research  
SIGGRAPH Asia 2020

## **DeepMetaHandles: Learning Deformation Meta-Handles of 3D Meshes with Biharmonic Coordinates**

Minghua Liu<sup>1</sup>, Minhyuk Sung<sup>2</sup>, Radomír Měch<sup>3</sup>, and Hao Su<sup>1</sup>

<sup>1</sup>UCSD, <sup>2</sup>KAIST, <sup>3</sup>Adobe Research, <sup>4</sup>UCSD  
CVPR 2021

## **Deformation-Aware 3D Shape Embedding and Retrieval**

Mikaela Angelina Uy<sup>1</sup>, Jingwei Huang<sup>1</sup>, Minhyuk Sung<sup>2</sup>, Tolga Birdal<sup>1</sup>, and Leonidas Guibas<sup>1</sup>

<sup>1</sup>Stanford University, <sup>2</sup>Adobe Research  
ECCV 2020

교수/박사연구원급 발표

# **Mixed Reality Tabletop Gameplay: Social Interaction with a Virtual Human Capable of Physical Influence**

이명호<sup>0,1</sup>, Nahal Norouzi<sup>2</sup>, Gerd Bruder<sup>2</sup>, Pamela Wisniewski<sup>2</sup>, Greg Welch<sup>2</sup>

<sup>1</sup>부산대학교, <sup>2</sup>University of Central Florida

본 논문은 IEEE Transactions on Visualization and Computer Graphics 게재 예정임 (Early Access).

교수/박사연구원급 발표

# Deep Combiner for Independent and Correlated Pixel Estimates

백종희<sup>1</sup>, Binh-Son Hua<sup>2,3</sup>, Toshiya Hachisuka<sup>4</sup>, 문보창<sup>1</sup>

<sup>1</sup>광주과학기술원, <sup>2</sup>VinAI Research, <sup>3</sup>VinUniversity, <sup>4</sup>The University of Tokyo

본 논문은 ACM SIGGRAPH ASIA 2020에 발표되었음.

# 기하적 변형이 포함된 이미지의 GAN Inversion\*

강경국<sup>0,1</sup>, 김성태<sup>2</sup>, 조성현<sup>1,2</sup>  
 포항공과대학교 <sup>1</sup>컴퓨터공학과, <sup>2</sup>인공지능대학원  
 {kkang831, seongtae0205, s.cho}@postech.ac.kr

## GAN Inversion for Out-of-Range Images with Geometric Transformations

Kyoungkook Kang<sup>0,1</sup>, Seongtae Kim<sup>2</sup>, Sunghyun Cho<sup>1,2</sup>  
 {<sup>1</sup>Dept. of Computer Science and Engineering, <sup>2</sup>Graduate School of Artificial Intelligence} POSTECH

### Abstract

In this paper, we propose a novel GAN inversion approach to semantic manipulation of out-of-range images that are geometrically unaligned with the training images of a GAN model. To find a latent code that is semantically editable, our approach inverts an input out-of-range image into an alternative latent space than the original latent space. We also propose a regularized inversion method to find a proper solution that supports semantic manipulation in the alternative space. Our experiments show that our approach effectively supports semantic manipulation of out-of-range images with geometric transformations.

### 1. Introduction

Recently, it has been shown that rich semantic information is encoded in the latent space of GANs, and furthermore, that images can be effectively manipulated in a semantically meaningful way by modifying latent code. To enable such semantic manipulation for real images, GAN inversion has recently attracted much attention [3, 4, 5, 6]. GAN inversion maps a real image into the latent space of a pre-trained GAN model. As shown in [3], for successful semantic manipulation of real images, it is critical to find an in-domain latent code of a pre-trained GAN model.

Unfortunately, such in-domain latent codes can be found only for a small fraction of real images that align with the training images of a pre-trained GAN model.

For example, most GAN models use geometrically aligned face images as their training data for ease of training. As a result, images with a small amount of translation or other geometric transforms are out of their ranges, i.e. images without the align method used when creating the training set, and the previous GAN inversion methods cannot find in-domain latent codes for such out-of-range images. This severely limits the applicability of semantic editing of real images using GAN inversion.

In this paper, we propose a novel GAN inversion approach to semantic manipulation of out-of-range images. Specifically, Our approach inverts an image that is geometrically unaligned with the training images for the StyleGAN [1] and StyleGAN2 [2] frameworks. To this end, we propose to invert an image into another space  $F/W^+$  space. To find a in-domain latent code in the  $F/W^+$  space that faithfully reconstructs the input image and supports semantic manipulation, we also propose a regularization approach for the  $F/W^+$  space based on an encoder network.

### 2. Our Approach

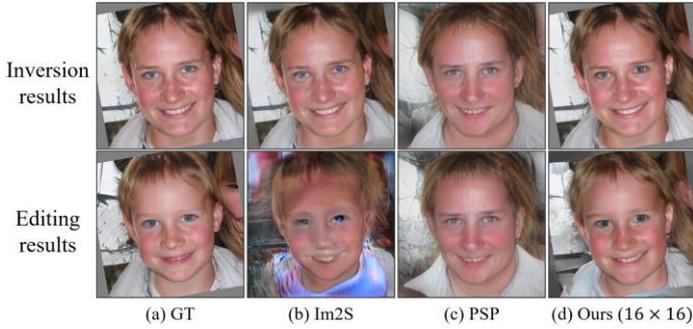
StyleGAN frameworks [1, 2] have a unique intermediate latent space, and  $w \in W$  is fed to multiple layers of different scales of the generator to control the style of each scale. For StyleGAN inversion, more extended latent space  $W^+$  is generally used to enhance the reconstruction accuracy [4]. A  $w^+ \in W^+$  is a set of  $w$  vectors  $\{w_1, w_2, \dots, w_N\}$ , where each  $w$  is the input of each layer.

Nonetheless, GAN inversion to the extended latent space  $W^+$  still fails to find an in-domain latent code for out-of-range images as discussed in Sec. 1. To overcome this limitation, we propose another latent space  $F/W^+$  where each element  $w^*$  in the space is defined as  $w^* = (f, w_{M+})$ .  $w_{M+}$  is a subset of  $w^+$ , which is the inputs of the layers larger than a specific

\* 구두발표논문

\* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회에 제출 중

\* 본 연구는 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2019-0-01906, 인공지능대학원 지원(포항공과대학교))과 한국연구재단의 지원(No. 2020R1C1C1014863)을 받아 수행되었음.



**Figure 1.** Qualitative comparison of the reconstruction quality and editing quality of different methods on geometrically transformed images. We sample 50 images from the CelebA-HQ dataset and applied different degrees of translation, rotation, and scaling. In this example, an input image is rotated by 20 degrees, and aging editing is applied.

$M$ th layer for the fine scales of the generator.  $f$  is a coarse-scale feature map of the generator before the  $M$ th layer. In our experiments, we test two scales,  $8 \times 8$  and  $16 \times 16$ , for  $f$ .

The  $F/W^+$  space provides a couple of nice properties that enable semantic editing of out-of-range images. First, compared to  $\{w_1, \dots, w_{M-1}\}$ ,  $f$  can represent a wider range of images including images with geometric transformations, as  $f$  has a greater degree of freedom. Second,  $w_{M+}$  is invariant to translations of images, as it is the input of the spatially global operation of the StyleGAN frameworks.

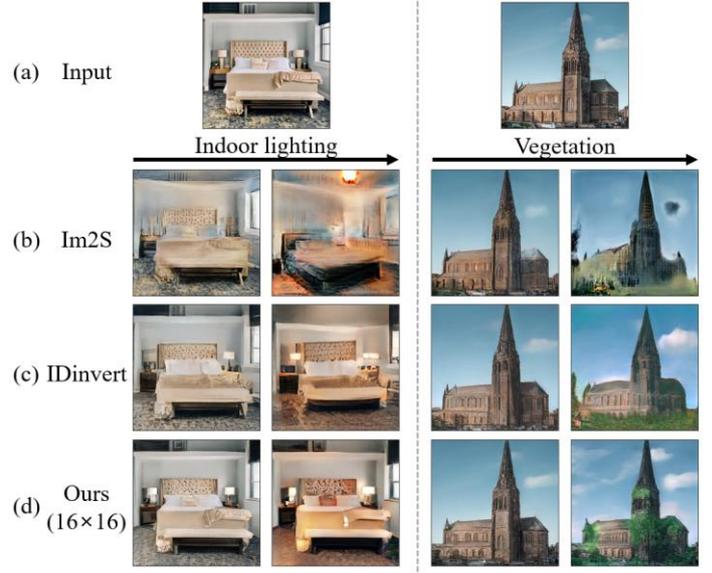
To enable semantic manipulation for out-of-range images, both  $f$  and  $w_{M+}$  must be in proper domains. To this end, we adopt a regularized optimization scheme both on  $f$  and  $w_{M+}$ . For  $w_{M+}$ , we adopt the  $P - norm^+$  space-based regularization proposed by Zhu et al. [5]. For  $f$ , we first find an initial latent code  $f^o$  that lies in the extended domain of  $f$  using an encoder  $E$  and find a latent code  $f$  that is close to  $f^o$ .

For the training of the encoder, we sample latent codes  $(f^{gt}, w_{M+}^{gt})$  and its image  $I$  and train our encoder  $E$  with a loss function defined as:

$$L_{enc} = \|G(E(I), w_{M+}^{gt}) - I\|^2 + \lambda \|F(G(E(I), w_{M+}^{gt})) - F(I)\|^2$$

where  $\lambda$  is a weight to balance two terms, and  $F$  is a LPIPS [7] network to calculate perceptual similarity. The encoder has a VGG-like architecture consisting of 11 convolution blocks and three pooling layers. Our training procedure does not use geometrically transformed images. Nevertheless, our encoder still performs effectively for geometrically transformed images thanks to the spatially-invariant property of CNNs.

### 3. Experimental Results



**Figure 2.** Qualitative comparison of the reconstruction quality and editing quality of different methods on natural images. The input images on the top row are collected from the internet. We use StyleGAN [1] models pre-trained on the LSUN bedroom and tower datasets.

We compare our method with recent state-of-the-art approaches: IDinvert[3], Im2S [4], and PSP [6]. Figure 1 shows a reconstruction and editing quality comparison. The figure shows that our  $16 \times 16$  version can reconstruct input image and successfully edit inversion result. Only Im2S shows high-quality reconstruction results. However, due to the lacks in-domain constraints, Im2S tends to produce out-of-domain latent codes that are not semantically editable.

Due to the large diversity of natural images, it is difficult to accurately reconstruct and edit a natural image using previous GAN inversion approaches. On the other hand, thanks to the high degree-of-freedom of the  $F/W^+$  space, our approach is especially effective in handling such natural images. Figure 2 shows reconstruction and editing quality comparisons.

### Reference

- [1] Karras, T et al. A style-based generator architecture for generative adversarial networks. In CVPR, 2019.
- [2] Karras, T et al. Analyzing and improving the image quality of stylegan. In CVPR, 2020.
- [3] Zhu, J et al. In-domain gan inversion for real image editing. In ECCV, 2020.
- [4] Abdal, R et al. Image2stylegan: How to embed images into the stylegan latent space? In CVPR, 2019.
- [5] Zhu, P et al. Improved stylegan embedding: Where are the good latents? arXiv preprint arXiv:2012.09036, 2020.
- [6] Richardson, E et al. Encoding in style: a stylegan encoder for image-to-image translation. In CVPR, 2021.
- [7] Zhang, R et al. The unreasonable effectiveness of deep features as a perceptual metric. In CVPR, 2018.

# 광선추적 GPU를 이용한 확률적 볼륨매핑 가속화\*

민혜정<sup>0</sup>, 한경민, 김영준  
이화여자대학교 컴퓨터공학과  
hjmin@ewhain.net, {hankm, kimy}@ewha.ac.kr

## Accelerating Probabilistic Volumetric Mapping using Ray-Tracing Graphics Hardware

Heajung Min<sup>0</sup>, Kyung Min Han, Young J. Kim  
Dept. of Computer Science and Engineering, Ewha Womans University

### Abstract

Probabilistic volumetric mapping (PVM) represents a 3D environmental map for an autonomous robotic navigational task. A popular implementation such as Octomap relies on an octree to represent a PVM and its main bottleneck lies in massive ray-shooting to determine the occupancy of the underlying volumetric voxel grids. In this paper, we propose GPU-based ray shooting to drastically improve the ray shooting performance in Octomap. Our main idea is based on the use of the recent ray-tracing RTX GPU. Our ray-shooting first maps leaf-level voxels in the given octree to a set of axis-aligned bounding boxes (AABBs) and employ massively parallel ray shooting on them using GPUs to find free and occupied voxels. These are fed back into the CPU to update the voxel occupancy and restructure the octree. In our experiments, we have observed more than three-orders-of-magnitude performance improvement in terms of ray shooting using ray-tracing GPU over a state-of-the-art Octomap CPU implementation, where the benchmarking environments consist of more than 77K points and 25K~34K voxel grids.

### 1. Introduction

3D mapping is an essential component for autonomous navigational tasks since the accuracy of 3D mapping significantly affects estimating the surroundings where the robot is deployed. PVM is a popular strategy for representing such 3D maps. As a choice for volumetric reconstruction, an octree is a *de facto* standard. Octomap[1] is a state-of-the-

art implementation for PVM using the octree representation that consists of the following steps to build a PVM: 1) Scan and generate a point cloud for the environment. 2) Shoot rays toward each point in the point cloud. 3) Find and identify free or occupied voxels in the observed space. 4) Update the octree with updated occupancy. Often, the second and third steps, i.e., the ray-shooting step, are the most time-consuming operation. In this paper, we perform GPU-based ray shooting to drastically improve the ray shooting performance in Octomap, which is the main bottleneck in octree-based PVM. In our experiments, we have observed more than three-orders-of-magnitude performance improvement in terms of ray shooting using ray-tracing RTX GPU over an Octomap CPU implementation.

### 2. GPU-Accelerated Ray Shooting

#### 2.1. Fast and Massively-parallel Ray Shooting on GPU

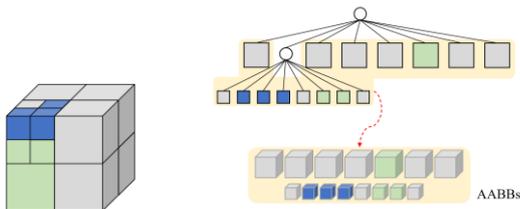
Since ray-tracing is a computationally intensive technique, dedicated hardware support is highly beneficial. Timely, NVIDIA introduced and started to mass-produce ray tracing hardware, known as RTX. Many RT cores installed in the RTX consist of two units, one responsible for the bounding volume hierarchy (BVH) traversal and the other for ray-triangle or ray-AABB intersection test. Also, shader programs in the RTX pipeline are executed in a massively-parallel fashion for all rays[2].

#### 2.2. Mapping an Octree on CPU to AABBs on GPU

In order to leverage GPU-accelerated ray shooting for an octree, the voxel elements of the octree must be converted to geometric primitives that the ray-tracing GPU such as the RTX can process, which should be a set of triangles or AABBs in case of RTX. In our problem, we opt for AABBs as target

\* 구두 발표논문, 본 논문은 요약논문(Extended Abstract)으로서, 본 논문의 원본 논문은 IEEE International Conference on Robotics and Automation 2021에 발표되었음.

\* 본 연구는 ITRC/IITP 프로그램(IITP-2021-2020-0-01460)과 연구재단 중견연구자지원사업(2017R1A2B3012701)의 지원으로 수행되었음.



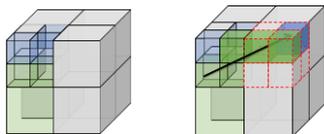
**Figure 1:** (left) spatial subdivision with different labels to represent occupied (blue), free (green), or unknown (gray) voxel space, (right) an octree representation on CPU with leaf-level nodes highlighted in yellow, and a set of AABBs on GPU that correspond to the leaf-level voxel nodes on CPU.

primitives as their geometries are close to the shape of a voxel. Specifically, we convert all leaf-level voxels in an octree to individual AABBs.

As illustrated in Fig. 1, a leaf-level voxel is mapped to an AABB by matching their geometries including the position and size. After all leaf-level voxels are converted to a set of AABBs, they are uploaded to the GPU, and the RTX GPU builds a bounding volume hierarchy (BVH) of AABBs.

### 2.3. Massively-Parallel Ray Shooting

Once the BVH of AABB is computed on GPUs, we set up multiple rays in the ray generation shader and shoot them. The direction of each ray is defined by the sensor origin (ray start) and the position of each point (ray end) in the cloud. All AABBs intersected with a given ray are found within the ray's extent. When ray-AABB intersection occurs, determining the occupancy of the voxel is performed in the intersection shader.



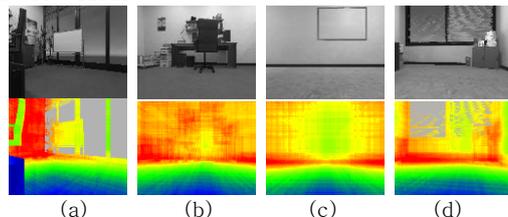
**Figure 2:** (left) leaf-level voxels mapped to a set of AABBs, (right) subdivision to the finest resolution for occupancy labeling.

We use CPU for octree and occupancy update using Octomap, and the Octomap expects voxels of the finest resolution in the octree as a result of ray shooting. In order to meet this interfacing requirement, if the size of an AABB intersected with a ray is greater than that of a finest-resolution voxel, we subdivide the subspace that the ray traverses to a set of sub-voxels in the finest resolution (typically 16) using DDA. These voxels are labeled as free except that the voxel containing the endpoint of the ray is labeled as occupied. Fig. 2 illustrates this situation. If the intersection of ray and AABB no longer occurs, the miss shader is executed to simply terminate the ray shooting. After a set of voxels hit by rays are found, octree

and occupancy update are executed on the CPU by sending the voxel data back to the CPU.

## 3. Experiments and Results

GPU-based ray shooting for map building was implemented on NVIDIA's RTX 2080 GPU. The acquired point cloud data is built into an octree with a maximum depth of 16, corresponding to 25K~34K leaf-level voxels.



**Figure 3:** (top) the target scene, (bottom) the corresponding hit count of rays with voxels in the space; as the color changes from blue to red, more voxels are intersected with rays.

Fig. 3 illustrates five views from different viewpoints. Here, we shoot 320x240 rays (76,800 rays) per view to collect and identify octree cells. Table 1 is the result of evaluating ray shooting time for each view shown in Fig. 3. GPU-based ray shooting can be performed three-orders-of-magnitude faster than CPU-based ray shooting on average excluding GPU-CPU readback. Even though GPU-CPU readback time is included, the performance improvement is still two-orders-of-magnitude faster than the CPU version.

**Table 1:** Comparisons of Ray Shooting Performance on Octomap (CPU) and Ours (GPU), and Timing Breakdown

Benchmarking Views		(a)	(b)	(c)	(d)
# of Free, Occupied or Unknown Voxels in Octree		25,515	33,806	34,412	21,832
Octomap (CPU)	Ray Shooting (ms)	886.83	1,446.52	1,349.87	1,294.31
Ours (GPU)	Build BVH (ms)	0.59	0.63	0.58	0.55
	Ray Shooting (ms)	1.42	1.67	2.07	1.14
	Readback from GPU to CPU (ms)	14.19	14.63	14.39	14.23

## 4. Conclusions

In this paper, we propose GPU-based ray shooting to improve the ray shooting performance, which is the main bottleneck, in the state-of-the-art PVM algorithm such as Octomap. The limitation is that the octree itself is not maintained on GPUs but CPU. However, this can be addressed in the future using various GPU-based octree maintenance techniques.

## References

[1] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, OctoMap: An efficient probabilistic 3D mapping framework based on octrees, *Autonomous Robots*, 34(3):189-206, 2013.

[2] M. Stich, Introduction to NVIDIA RTX and DirectX Ray Tracing, <https://devblogs.nvidia.com/introductionnvidia-rtx-directx-ray-tracing/>, 2018

# 음성 입력으로부터 사실적인 말하는 영상 생성\*

주광진, 장원중, 이승용  
 포항공과대학교 인공지능대학원  
 {gwangjin, wonjong, leesy}@postech.ac.kr

## Realistic Talking Head Generation from Audio Input

Gwangjin Ju, Wonjong Jang, Seungyong Lee  
 Graduate School of Artificial Intelligence, POSTECH

### 요약

본 논문에서는 하나의 입력 사진과 입력 음성으로부터 사실적인 말하는 얼굴을 생성하는 모델을 소개한다. 이 모델은 얼굴 특징점을 중간 표현으로 사용하여 음성과 얼굴 생성 사이의 연결한다. 이를 통해 효율적인 encoder 및 decoder를 구성할 수 있다. 또한 여기서 제안한 StyleGAN에 기반한 구조인 StyleUNet은 입력 정보의 skip-connection을 통해 더 사실적인 영상을 생성할 수 있다.

### 1. 서론

본 논문은 음성에서 말하는 얼굴을 생성하는 것을 목적으로 하며, 얼굴 특징점을 중간 표현으로 사용하여 음성에서 고차원인 얼굴의 움직임으로 직접 매핑하는 것을 피하고, encoder와 decoder의 구성이 효율적이 되도록 한다. 또한 StyleGAN[1]을 기반으로 한 StyleUNet을 제안하여 입력 사진 하나만으로 자연스러운 얼굴을 생성한다. 이를 통해 제안된 모델은 audio-to-lip 모델, 얼굴 움직임 생성 모델, 얼굴 영상 생성 모델로 총 3가지 부분으로 이루어진다.

Audio-to-lip 모델은 음성으로부터 입의 움직임을 예측하여 얼굴 특징점을 생성하며, 얼굴 움직임 생성 모델은 자연스러운 움직임을 생성하여 얼굴 영상 생성 모델에 전달한다. 얼굴 영상 생성 모델은 전달받은 정보와 입력 영상을 이용해 입력 음성 신호에 맞는 얼굴 영상을 생성한다.

실험 결과를 통해 본 논문에서 제안한 모델이 사실적이고 자연스러운 얼굴 영상을 만들어낸다는 것을 볼 수 있다. 특히 이 모델은 원 샷 학습을 진행하였기 때문에

학습 때 관측되지 않은 입력 영상이 주어지더라도 실패하지 않는 것을 확인할 수 있다.

### 2. 음성을 이용한 얼굴 영상 생성 모델

본 논문에서 제안한 모델의 전체적인 구조는 아래 그림과 같다.

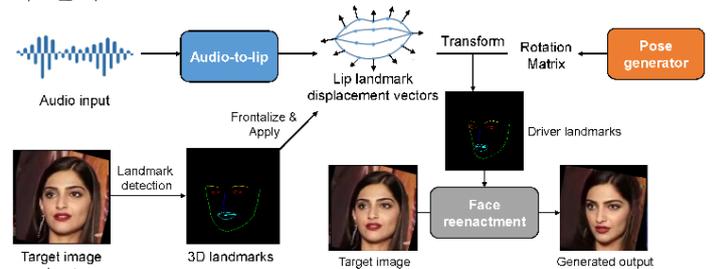


그림 1 제안된 모델 구조

#### 2.1. Audio-to-lip 모델

Audio-to-lip 모델은 입력 음성 신호를 이용하여 입의 모양을 예측하며, 이 때 음성 신호의 연속된 상관성을 잘 학습할 수 있도록 LSTM 모델을 이용한다.

음성 신호는 50ms의 범위로 각 40ms마다 추출한 MFCC 특성 값을 추출하여 학습의 입력 데이터로 사용한다. 학습의 정답으로 사용되는 입의 특징점은 모두 정면화한 뒤 주성분 분석 방법을 적용하여 입의 움직임과 관련된 정보만 갖도록 한 뒤 학습의 정답 데이터로 사용한다.

또한 모델이 음성과 입의 모양의 관계를 학습할 수 있도록 음운을 분석하여 음운에 해당하는 표준적인 입 모양을 제약으로 준다. 이와 함께 audio-to-lip 모델이 음성에 해당하는 음운을 예측하는 문제를 추가적으로 해결하도록 한다.

최종적으로 학습을 위한 목적 함수는 아래와 같다.

$$L_{lip} = E \left[ \|L_{predict} - L_{GT}\|_2 + \|L_{predict} - L_{prior}\|_2 + \text{crossentropy}(P_{pred}, P_{GT}) \right]$$

$L_{predict}$  는 예측한 입의 움직임,  $L_{GT}$  는 정답 입의 움직임

\* 구두발표논문

\* 본 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (IITP-2020-0-01649, 마이크로소프트연구소아시아와 공동연구협력 프로젝트; IITP-2015-0-00174, SW 스타랩)

이다.  $L_{prior}$  는 해당 음운에 해당하는 표준적인 입의 움직임이며  $P_{pred}$  는 예측한 음운,  $P_{GT}$  는 정답 음운이다.

### 2.2. 얼굴 움직임 생성 모델

사람이 말을 할 때 생기는 자연스러운 움직임을 생성하기 위해 긴 시간의 연관성을 잘 고려하는 모델인 Transformer[2]의 구조를 이용하여 모델을 구성한다. 학습 데이터는 입력 영상 데이터에서 얼굴의 정면 방향을 추출하여 사용한다.

### 2.3. 얼굴 영상 생성 모델

이 모델은 target encoder, driver encoder, 그리고 StyleGAN을 decoder로 이용한 encoder-decoder 구조를 이용하여 자연스러운 영상을 생성한다.

얼굴 영상 생성 모델의 target encoder는 입력 영상을, driver encoder는 생성된 얼굴 특징점 영상을 encode하여 낮은 단계 feature map과 압축된 은닉 벡터를 얻는다. Decoder인 StyleGAN은 두 encoder에서 받은 은닉 벡터로부터 영상을 생성한다. 이 때 은닉 벡터로 압축되는 과정에서 정보가 손실되어 정확도가 낮아지므로 encoder로부터 낮은 단계의 feature map을 skip connection으로 전달받아 세밀한 내용을 복원한다.

얼굴 영상 생성 모델의 학습은 3단계로 구성된다. 첫 번째 단계에서는 FFHQ dataset로 미리 학습된 StyleGAN을 대상 dataset에 fine-tune하여 decoder를 학습한다. 두 번째 단계에서는 target encoder와 decoder만을 이용해 auto-encoder의 목적 함수로 학습한다. 세 번째 단계에서는 driver encoder까지 포함하여 전체 모델을 학습한다. 이 단계의 학습에서는 같은 사람의 다른 영상 2개를 이용한 self-face reenactment의 형태로 학습하므로 정답 영상이 존재하며, 이를 이용한 목적 함수는 아래와 같다.

$$L_{face} = L_{GAN} + \lambda_{pixel}L_{pixel} + \lambda_{percept}L_{percept} + \lambda_{identity}L_{identity} + \lambda_{landmark}L_{landmark}$$

$L_{GAN}$ 은 adversarial loss,  $L_{pixel}$ 은 pixelwise L1 loss이다.  $L_{percept}$ 은 perceptual loss이며,  $L_{identity}$ 은 얼굴 인식 모델을 이용한 identity preservation loss,  $L_{landmark}$ 은 정답 영상과의 특징점 차이를 이용한 손실 함수이다.  $\lambda$ 는 각각의 가중치이다.

## 3. 결과

본 논문에서 제안된 모델은 VoxCeleb1[4]의 얼굴 영상과 얼굴 영상에서 추출한 특징점, 그리고 음성을 학습 데이터로 사용한다.

### 3.1. Audio-to-lip 모델

이 모델의 결과를 이용하여 mean squared error(MSE)를 측정할 경우 0.1266의 결과를 보인다.

### 3.2. 얼굴 영상 생성 모델

동일 인물의 다른 영상을 생성한 영상의 peak signal-to-noise ratio(PSNR)을 측정할 경우 본 논문에서 제안한 모델은 23.642, 기존 face reenactment 연구인 MarioNETte[3]는 23.244의 결과를 보이며, 결과는 아래 그림과 같다.



그림 2 얼굴 영상 생성 모델 결과 비교

## 4. 결론

본 논문에서는 LSTM 구조를 이용해 음성 신호로부터 입의 움직임을 예측하였으며 더 사실적인 얼굴 영상을 합성하기 위해 StyleGAN을 기반으로 한 StyleUNet을 제안하였다. 제안된 모델은 입력 영상과 같은 사람이면 서로 사실적인 얼굴을 생성할 수 있다.

## 참고문헌

- [1] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. *In Proc. CVPR*, 2020.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *In Proc. NIPS*, 2017.
- [3] Sungjoo Ha, Martin Kersner, Beomsu Kim, Seokjun Seo, and Dongyoung Kim. MarioNETte: Few-shot face reenactment preserving identity of unseen targets. *In Proc. AAAI*, 2020.
- [4] Arsha Nagrani, Joon Son Chung, and Andrew Senior. Voxceleb: a large-scale speaker identification dataset. *In Proc. INTERSPEECH*, 2017.

# 물리 시뮬레이션 및 모션캡처 데이터의 물체 및 관절 사이의 관계 인식\*

이도해<sup>0</sup>, 오영진, 이인권  
연세대학교 컴퓨터과학과

dlehgo1414@gmail.com, skrcjstk@yonsei.ac.kr, iklee@yonsei.ac.kr

## Reasoning Relationship between Entities for Physics Simulation and Mocap Data

Dohae Lee<sup>0</sup>, Young Jin Oh, In-Kwon Lee  
Dept. of Computer Science, Yonsei University

### 요약

개체들이 서로 특정한 관계를 가지고 상호작용하는 시스템 속 개체 사이의 관계를 추론하기 위한 graph neural network 기반의 방법들이 활발히 연구되고 있다. 이러한 방법들을 이용하여 물체 시뮬레이션 데이터 및 캐릭터 애니메이션 데이터 등의 상호작용 시스템에서 개체 간 관계를 추론하고 미래 움직임을 예측하는 것이 가능하였다. 하지만, 이전 연구들에서 제안된 방법들은 시스템 속 관계의 수를 모르거나, 관계의 수가 많은 데이터에 적용하기 어려웠다. 본 논문에서는 개체 간 관계를 잠재 공간에 임베딩하여 추론할 수 있는 방법을 제안하고, 기존 방법들과의 성능 비교 실험을 수행하였다. 가상 물리 시뮬레이션 데이터와 모션 캡처 데이터를 이용한 성능 비교 실험 결과, 본 논문에서 제안한 방법은 기존 방법들보다 상호작용 시스템 속 개체 간 관계를 더 정확하게 추론할 수 있었다.

### 1. 서론

상호작용 시스템은 개체들이 서로 특정한 관계를 맺고 상호작용하는 시스템을 의미한다. 물리 시뮬레이션 데이터와 캐릭터 애니메이션 데이터 등이 상호작용 시스템에 해당하는데, 이들은 시간이 흐르면서 개체의 특성이 변화하는 동적 그래프로 나타낼 수 있다. 이러한 상호작용 시스템을 관찰하여 숨겨져 있는 관계를 추론할 수 있다면, 시스템을 이해하고 개체의 상태변화를 예측하는데 도움이 될 수 있다. 예를 들어, 물체들이 서로 다양한 물리적 관계를 맺고 상호작용할 때, 물체 간 관계를 파악함으로써 전체 시스템을 이해하는데 도움이 될 수 있고, 물체의 미래 움직임을 더 정확히 예측할 수

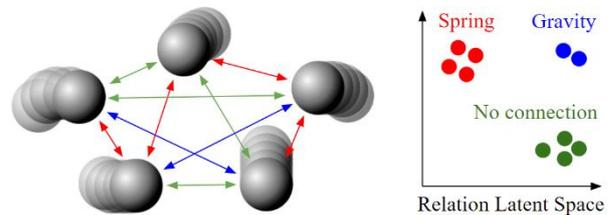


그림 1: 물리 시뮬레이션 데이터 속 물체 사이의 관계를 잠재 공간 위에 표현

있다. 또는, 캐릭터의 움직임을 관찰하여 관절 간의 관계를 추론함으로써, 캐릭터의 미래 움직임을 더 정확히 예측할 수도 있다.

최근 들어, 동적 그래프에서 개체의 상태 변화를 관찰하여 개체 사이의 관계를 유추하고, 이를 기반으로 개체의 미래 상태를 예측하는 방법들이 활발히 연구되고 있다 [2,3]. 그러나, 이전 방법들은 각각의 관계를 별개의 네트워크를 이용하여 모델링하기 때문에 관계의 수를 모르거나 관계가 복잡할 경우 적용이 어려웠다. 반면, 본 논문에서 제안하는 방법은 관계를 잠재 공간 속 연속적인 벡터로 임베딩하여 표현함으로써 관계의 수를 모르거나 관계가 복잡한 시스템에도 잘 적용될 수 있다 (그림 1).

### 2. 방법

#### 2.1. 모델 구조

본 논문에서 제안하는 모델은 두개의 graph neural network [1] 기반의 세부 네트워크로 구성되어 있다: 관계 인코더, 관계 디코더.

##### 2.1.1. 관계 인코더

관계 인코더는 일정 시간동안 상호작용 시스템 속 개체의 상태를 관찰하여 개체 사이의 관계를 추론하는 네트워크이다. 먼저, 관계 인코더는 edge의 상태를 추론한다.  $n_t^i$ 가  $t$  타임스텝에서의  $i$ 번째 개체의 상태이고,  $F$ 가 edge 상태를 업데이트 하는 학습 가능한 함수일

\* 구두발표논문

\* 본 논문은 요약논문(Extended Abstract)으로서, 현재 다른 publication에 제출 중임

\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2021-2018-0-01419\*).

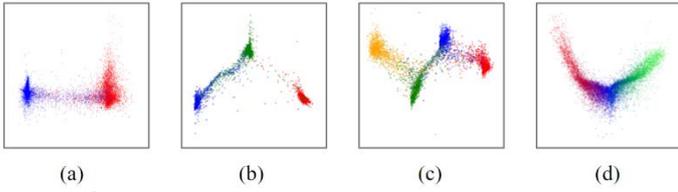


그림 2: 관계 잠재 벡터를 2차원 평면상에 표현

때,  $t$  타임스텝에서의  $i, j$  번째 개체 사이의 edge 상태  $s_{ij}^t = F(n_i^t, n_j^t, s_{ij}^{t-1})$ 이다. 충분한 시간동안 그래프를 관찰하여 edge 상태를 업데이트 한 후, edge 상태로부터 개체 사이 관계를 추론한다.  $G$ 가 edge 상태로부터 관계의 분포를 출력하는 함수이고,  $\sim$ 가 reparameterization trick을 이용한 샘플링일 때,  $i, j$  번째 개체 사이의 관계  $r_{ij} \sim G(s_{ij}^t)$ 이다.

### 2.1.2 관계 디코더

관계 디코더는 관계 인코더에서 추론한 개체 사이의 관계와 특정 시간에서의 개체들의 상태를 입력 받아, 한 타임스텝이 흐른 후의 개체 상태를 예측하는 네트워크이다.  $K$ 가 개체 사이의 영향력을 계산하는 함수일 때,  $j$  번째 개체가  $i$  번째 개체에게 전달하는 영향력  $f_{ij} = K(n_i^t, n_j^t, r_{ij})$ 이다.  $J$ 가 개체의 상태변화량을 계산하는 함수일 때,  $i$  번째 개체의 다음 상태  $n_i^{t+1} = n_i^t + J(n_i^t, \sum_j f_{ij}^t)$ 이다.

### 2.2. 랜덤 샘플링 기법

우리는 관계 인코더가 개체 사이의 관계를 올바르게 표현할 수 있도록 훈련하기 위해 랜덤 샘플링 기법을 사용한다. 랜덤 샘플링 기법은 관계 인코더와 관계 디코더에 입력되는 그래프의 시간을 매번 임의로 샘플링하는 훈련 방식을 말한다. 랜덤 샘플링 기법을 사용함으로써, 관계 인코더가 단순히 입력되는 그래프의 상태를 압축하도록 훈련되는 것을 방지할 수 있다.

### 2.3. 손실 함수

본 논문에서 제안하는 모델은 3가지 손실함수를 이용하여 훈련한다: 개체 상태 예측 loss, KL divergence loss, 그리고 관계 표준편차 loss. 개체 상태 예측 loss는 관계 디코더가 예측한 개체의 미래 상태와 ground truth 상태 간의 제곱 평균 오차이다. KL divergence loss는 관계 벡터의 각 원소들의 분포가 정규분포를 따르도록 유도한다. 마지막으로, 관계 표준편차 loss는 하나의 그래프에서 여러 시간에 걸쳐 추론한 관계 벡터들의 표준편차를 구하는 함수로, 같은 관계를 표현하는 관계 벡터 사이의 차이가 줄어들도록 유도한다.

## 3. 실험 및 결과

### 3.1. 데이터셋

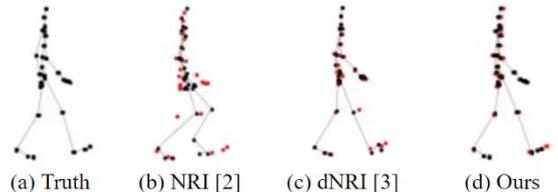


그림 3: 관절의 미래 움직임을 예측

본 논문에서는 물리 시뮬레이션 데이터와 모션캡처 데이터 [4]를 활용하여 성능 비교 실험을 수행하였다.

### 3.2. 물리 시뮬레이션 데이터

제안한 모델이 올바르게 물체 사이 관계를 추론하는지 확인하기 위하여, 관계 벡터를 차원 축소하여 2차원 평면상에 표현하였다 (그림 2). 그림 2의 a,b,c, 그리고 d는 각각 시스템 속 관계의 수가 2,3,4, 그리고 100개 일 때 모델이 인식한 관계 벡터들을 표현한 그림이다. 각 점들의 색은 관계의 종류를 의미한다. 그림 2에서 점들이 색깔 별로 잘 구분이 되어있는 것은 제안한 모델이 관계의 수에 상관없이 관계의 잠재 벡터를 잘 임베딩하고 있음을 의미한다.

### 3.3. 모션 캡처 데이터

제안한 모델로 모션 캡처 데이터에서 관절의 미래 움직임을 예측하는 실험을 수행하였다 (그림 3). 그림 3은 비교모델 [2,3]과 제안한 모델이 예측한 49 타임 스텝 후의 관절의 위치를 보여주는데, b,c, 그리고 d에서 빨간 점은 실제 관절의 위치이고, 검은 점은 각 모델이 예측한 관절의 위치이다. 제안한 모델이 가장 정확하게 관절의 미래 움직임을 예측할 수 있었다.

## 4. 토의 및 결론

본 논문에서 상호작용 시스템에서 개체 간 관계를 비감독 방식으로 추론할 수 있는 방법을 제안하였다. 물리 시뮬레이션 데이터와 모션 캡처 데이터를 이용한 성능 비교 실험 결과, 제안한 모델은 이전 방법들보다 정확하게 관계를 추론하고 미래 움직임을 예측할 수 있었다.

### 참고문헌

[1] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009

[2] Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. Neural relational inference for interacting systems. In *International Conference on Machine Learning*, pages 2688–2697, 2018.

[3] Colin Graber and Alexander G Schwing. Dynamic neural relational inference. In *Computer Vision and Pattern Recognition*, pages 8513–8522, 2020.

[4] CMU. Carnegie-mellon motion capture database. 2003. URL <http://mocap.cs.cmu.edu>.

# PointNet을 이용한 연체 동역학의 실시간 예측\*

박진형<sup>0</sup>, 이도해, 이인권  
연세대학교 컴퓨터과학과

jh9604@yonsei.ac.kr, dleho1414@gmail.com, iklee@yonsei.ac.kr

## Predicting Softbody Dynamics Using PointNet in Real Time

Jin-hyung Park<sup>0</sup>, Do-hae Lee, In-Kwon Lee  
Dept. of Computer Science, Yonsei University

### 요약

기계학습을 이용하여 물리 현상을 예측하고자 한 연구들은 많이 존재하지만, 주로 CG 데이터와 같은 가상 데이터를 사용하여 모델을 구축하여 왔다. 본 연구에서는 실제 세계에서 연체의 이동경로와 형태변화를 실시간으로 예측하기 위해 가상 데이터로 모델을 훈련시킨 후 실제 데이터로 미세조정 하는 방법을 제안한다. 제안한 방법은 기본 순환 신경망 모델에 비해 더 정확하고, 실시간 영상 예측이 가능하기 때문에 실시간으로 연체의 미래 상태 예측이 필요한 스포츠 분야 등에서 폭넓게 활용될 수 있을 것으로 기대된다.

### 1. 서론

가상 세계에서 강체나 연체, 그리고 유동체를 입자 (particle)들로 정의하여 물리 현상을 예측하는 방법들은 활발히 연구되고 있다[1-3]. HRN[1]과 GNS[2]는 입자들의 위치와 속력 정보를 이용하여 다양한 물체들의 이동 경로를 예측할 수 있었고, NRI[3]는 입자들의 이동경로를 기반으로 입자 사이의 관계를 추론하고, 물체의 미래 이동경로를 예측할 수 있었다. 하지만, 가상 데이터가 아닌 실제 데이터로 연체(softbody)의 이동경로와 형태변화를 예측하는 방법은 제안된 바 없다.

가상 세계에서 연체는 외곽선을 불연속적인 입자들로 나타낼 수 있다. 따라서 가상 데이터는 입자별로 고유번호를 부여하여 구성할 수 있고, 각 프레임별로 입자들의 대응관계를 쉽게 파악하여 입자의 움직임을 예측할 수 있다. 그러나, 가상 데이터와 다르게 실제 세계에서 연체 데이터는 영상으로 입력되기 때문에,

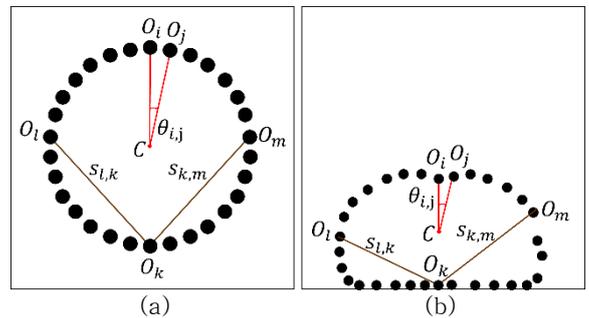


그림 1: (a)는 연체의 초기/정지 상태, (b)는 변형된 상태

외곽선을 인식하여 최대한 균등하게 입자를 직접 추출하여 데이터를 구성해야 한다. 이 때, 연체는 매 시각 회전하고 입자들의 간격이 좁기 때문에 시각  $t$ 에서 추출한 입자  $O_i$ 가 시각  $t+1$ 에서 추출된 입자들 중 어떤 입자에 대응되는지 확인할 수 없다. 즉, 각 프레임별로 입자들의 대응관계를 확인할 수 없기 때문에, 입자의 움직임을 예측하기 어렵다.

본 연구는 각 프레임별로 입자들의 순서가 무작위로 섞여 있어도 연체의 미래 상태를 예측할 수 있는 모델을 제안함으로써 이러한 어려움을 해결하였다.

### 2. 방법

#### 2.1 데이터 생성

**2.1.1 가상 데이터 생성** 본 논문에서는 가상 연체 데이터를 생성하기 위해 Box2D 시뮬레이터를 사용했다. 연체의 표면을  $n$ 개의 2차원 입자들의 집합  $\{O_i | 1 \leq i \leq n\}$ 으로 정의했으며,  $|i - j| < \frac{2}{5}n$  인 입자  $O_i$ 와  $O_j$ 는 복원력이 있는 용수철  $s_{i,j}$ 로 연결되어 있다 (그림 1).

**2.1.2 실제세계 데이터 생성** 실제세계 데이터를 생성하기 위해 여러 각도와 세기로 던져지는 연체를 초고속 카메라로 촬영하고, 연체의 외곽선을 구성하는  $n$ 개의 입자를 추출했다. 입자를 최대한 균등하게 추출하기 위해 인접한 두 입자가 연체의 중심점  $C$ 와 이루는 각도  $\theta_{i,i+1}$ 가 최대한 일정하도록 추출했다 (그림 1).

\* 포스터발표 논문, 학부학생 주 저자 논문  
\* 본 논문은 요약논문(Extended Abstract)으로서, 초기 연구 결과를 담고 있음  
\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2021-2018-0-01419\*).

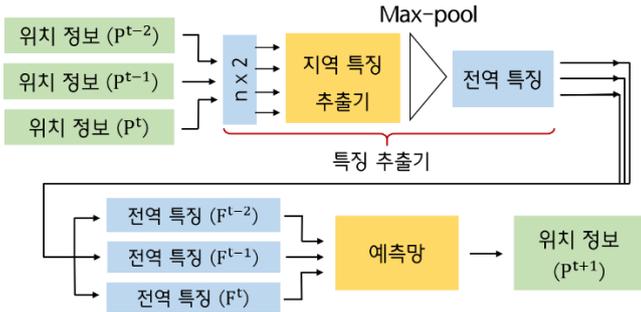


그림 2: PointNet을 활용한 우리의 시계열 예측 모델 구조

## 2.2 모델 구조

본 연구에서 제안하는 방법은 모델에 입력하는 입자들의 순서에 관계없이 동일한 값을 예측하는 PointNet[4] 기반의 모델을 이용한다 (그림 2). 우리 모델은 시각  $t-2$ ,  $t-1$ ,  $t$  일 때의 입자들의 위치  $p^{t-2}$ ,  $p^{t-1}$ ,  $p^t$ 를 입력 받아서  $t+1$ 일 때의 입자들의 위치  $p^{t+1}$ 를 예측한다. 이 때, 각 시각 별로 입자들의 순서가 서로 다르게 섞여 있더라도, 우리 모델은 미래 상태를 높은 정확도로 예측할 수 있다. PointNet의 구조를 활용한 특징 추출기에서는 MLP를 통해 입자별로 지역 특징을 추출한 후, 입력 값의 순서에 관계없이 동일한 연산을 하는 대칭함수(max pooling)를 거쳐 전역 특징  $F^{t-2}$ ,  $F^{t-1}$ ,  $F^t$ 를 추출한다. 각 시각별로 추출된 전역 특징들은 3개의 LSTM layer와 Fully-Connected layer로 구성된 예측망에 입력되어  $t+1$ 일 때의 입자들의 위치  $p^{t+1}$ 을 예측한다.

## 2.3 모델 훈련

먼저, 우리는 가상 데이터로 모델을 사전 훈련시킨 후 실제 데이터로 모델을 미세 조정(fine-tuning)했다. 미세 조정 시 특징 추출기는 고정시킨 상태에서 예측망만 재학습했다. 사전 훈련 및 미세 조정 시에는 입자의 실제 위치와 모델이 예측한 위치 간의 Chamfer Distance를 비용함수로 사용하여 모델을 훈련시켰다.

## 3. 실험 및 결과

우리는 연체를 초고속 카메라로 촬영한 영상의 첫 3개 프레임은 제안된 모델에 입력한 후, 미래 프레임을 예측하도록 실험했다. 또한, 입력하는 입자들의 순서에 영향 받지 않는 우리 모델의 성능을 입증하기 위해 입자들을 특정 방식( $x, y$  좌표 기준)으로 정렬해서 입력 받는 기본 순환 신경망의 예측결과와 비교했다.

그림 3. (a)는 우리 모델과 기본 순환 신경망이 예측한 프레임의 개수에 따른 누적 위치 오차를, 그림 3. (b)는 누적 형태 오차를 비교한 그래프이다. 위치 오차는 연체의 외곽선을 구성하는 실제 입자들의 위치와 모델이 예측한 외곽선 입자들 사이의 Chamfer Distance를 의미하고, 형태 오차는 예측한 입자들과

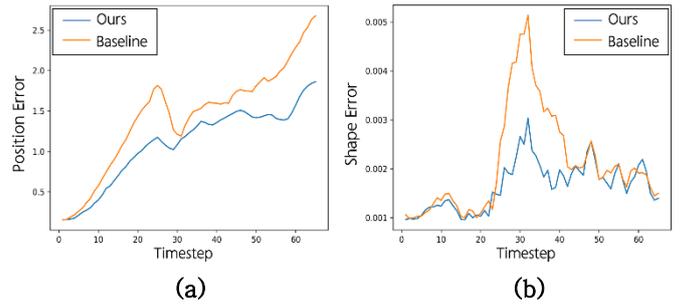


그림 3: (a)는 각 모델이 예측한 프레임 개수에 따른 누적 위치 오차, (b)는 누적 형태 오차

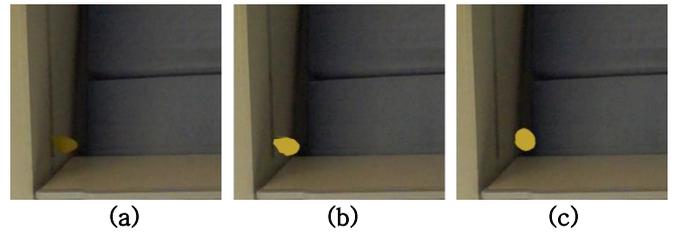


그림 4: (a)는 실제 영상, (b)는 제안된 모델의 예측결과, (c)는 기본 순환 신경망 모델의 예측결과

실제 입자들의 중심점이 각각 원점이 되도록 입자들을 평행 이동시킨 후, 입자들을 따라서 외곽선을 각각 그렸을 때 생기는 영역들의 차집합을 의미한다. 두 지표에 따르면 제안된 모델은 기본 순환 신경망 모델에 비해 연체의 위치와 형태를 더 정확하게 예측했다.

그림 4. (a)는 실제 영상, 그리고 (b)와 (c)는 제안된 모델과 기본 순환 신경망 모델이 예측한 34번째 프레임인데, 제안된 모델이 기본 순환 신경망 모델에 비해 연체의 미래 프레임을 더 정확하게 예측했다.

## 4. 결론

본 연구는 실제 세계에서의 연체 동역학을 학습하여 연체의 미래 상태를 실시간으로 예측하는 방법을 제안했다. 이를 통해 각 프레임별로 연체를 구성하는 입자들의 순서가 비결정적이어도 충돌 전후의 연체의 이동 경로와 형태변화를 모두 정확하게 예측할 수 있었다. 그러나, 다양한 연체와 상황에 대한 모델의 적용 가능성을 검증하기 위해 추가적인 실험이 필요하다.

## 참고문헌

[1] D.Mrowca, C. Zhuang, E.Wang, N.Haber, L.Fei-Fei, J.B.Tenenbaum, D.L.K.Yamins, Flexible Neural Representation for Physics Prediction, *NIPS*, 2018  
 [2] A.S.Gonzalez, J.Godwin, T.Pfaff, R.Ying, J.Leskovec, P.W.Battaglia, Learning to simulate complex physics with graph networks, *ICML*, 2020  
 [3] T.Kipf, E.Fetaya, K.Wang, M.Welling, R.Zemel, Neural relational inference for interacting systems, *ICML*, 2018  
 [4] C.R.Qi, H.Su, K.Mo, L.J.Guibas, PointNet: Deep Learning on Point Sets for 3D Classification, *CVPR*, 2017.

# 보정-보간 기반 연기 시뮬레이션\*

이윤지<sup>0</sup>, 오영진, 이도해, 이인권  
 연세대학교 컴퓨터과학과  
 {yoong15, skrcjstk, dlehgo1414, iklee}@yonsei.ac.kr

## Correction-Interpolation based Smoke Simulation

Yunjee Lee<sup>0</sup>, Young Jin Oh, Dohae Lee, In-Kwon Lee  
 Dept. of Computer Science, Yonsei University

### 요약

본 논문에서는 딥러닝을 이용하여 큰 타임스텝 시뮬레이션을 같은 시간 동안의 작은 타임스텝 시뮬레이션 결과와 가까워지도록 보정하는 방법을 제안한다. 제안한 방법을 통해 기존 연기 시뮬레이션 보간의 한계를 극복할 수 있고, 기존의 딥러닝 기반 연기 시뮬레이션 방법보다 더 빠르고 정확한 생성이 가능하다.

### 1. 서론

전통적인 물리 기반 시뮬레이션은 연산량이 많기 때문에, 정교한 시뮬레이션 및 렌더링을 위해서는 많은 자원과 시간이 필요하다. 최근 들어, 딥러닝을 이용하여 유체 시뮬레이션의 연산량을 줄이고 속도를 개선하려는 방법들이 활발히 연구되고 있다[1,2]. DeepFluids[1]는 딥러닝을 이용하여 매 스텝마다 속도장을 예측함으로써 연산 속도를 개선하였다. 전통적인 방법으로 큰 타임스텝 시뮬레이션을 생성한 후 사이의 프레임들을 딥러닝 방법을 사용하여 보간함으로써 연산 속도를 개선하려는 시도도 있었다[2]. 하지만, 큰 타임스텝 시뮬레이션(그림 1 a)은 작은 타임스텝 시뮬레이션(그림 1 b)보다 연산 과정에서의 오류가 크기 때문에 같은 시간 동안을 계산하더라도 오차가 발생한다(그림 1 t=4.0). 따라서 이 사이를 보간하면 실제 작은 타임스텝 시뮬레이션 결과와 차이가 있다. 따라서 실제 시뮬레이션에 적용하기 어렵다. 본 연구에서는 시뮬레이션 결과를 딥러닝으로 보정 및 보간하는 방식으로 이러한 한계를 극복하고, 기존 딥러닝 기반 시뮬레이션 방식보다 더 빠르고 정확한 방법을 제시한다.

### 2. 방법

본 연구에서 제시하는 방법은 그림 2와 같이 두 단계로 이루어진다.

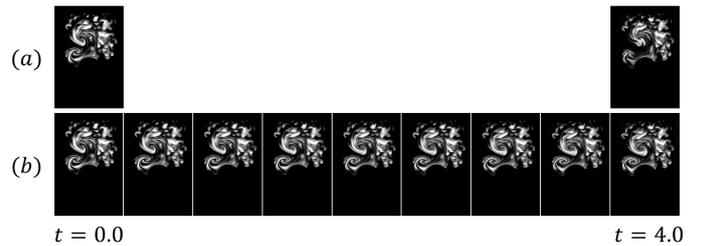


그림 1: 같은 시간 동안의 큰 타임스텝 시뮬레이션 결과와 작은 타임스텝 시뮬레이션 결과 차이

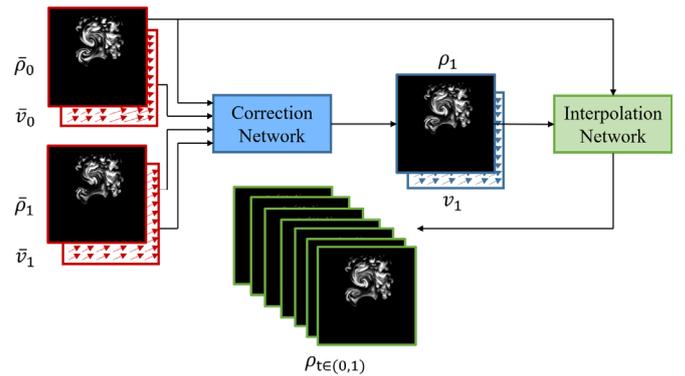


그림 2: 제안 방법 개요

먼저, 큰 타임스텝 시뮬레이션으로 계산한 결과를 보정 네트워크를 통해 작은 타임스텝 시뮬레이션의 결과와 유사해지도록 보정한다. 다음으로, 보정된 프레임과 직전 프레임 사이를 보간 네트워크를 통해 보간함으로써 프레임율(frame rate)을 높인다.

#### 2.1. 보정 네트워크 (Correction Network)

보정 네트워크는 큰 타임스텝 시뮬레이션의 결과의 밀도  $\bar{\rho}_1$ 와 속도  $\bar{v}_1$ 를 같은 시간 동안 작은 타임스텝 시뮬레이션으로 계산된 결과와 비슷하게 보정한다. 그림 3은 보정 네트워크의 구조를 나타낸다. 밀도는 시뮬레이션 생성 결과의 질에 가장 중요한 요인이다. 따라서 모델이 예측한 밀도  $\hat{\rho}_1$ 를 광학 흐름(optical flow)을 통해 추가적으로 보정함으로써 더욱 정교하게 하였다[3]; 먼저, 타겟 밀도로의 광학 흐름  $F$ 를 예측한 후, 이를 바탕으로 소스 밀도  $\bar{\rho}_1$ 를 와핑(warping)한다.

\* 구두발표논문  
 \* 본 논문은 요약논문(Extended Abstract)으로서, 초기 연구 결과임  
 \* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구 결과로 수행되었음 (IITP-2021-2018-0-01419\*)

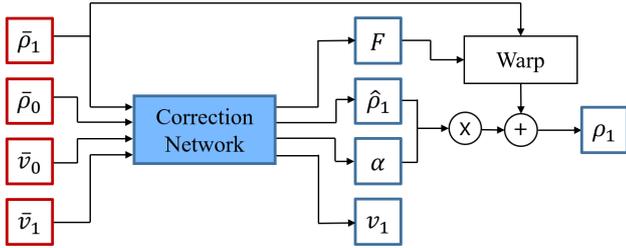


그림 3: 보정 네트워크 구조 개요

그 다음, 와핑한 결과와  $\hat{\rho}_1$ 를 적절한 학습 가능한 가중치  $\alpha$ 를 이용하여 가중합함으로써 최종적으로 보정된  $\rho_1$ 을 얻는다. 보정 네트워크는 CNN 기반 인코더-디코더의 각 층이 skip-connection으로 연결된 U-Net으로 구성된다. 손실 함수  $L$ 은 다음과 같이 정의하였다:

$$L = \lambda_{rD}L_{rD} + \lambda_{pD}L_{pD} + \lambda_{rV}L_V + \lambda_{pV}L_{pV} + \lambda_{gV}L_{gV} + \lambda_I L_I,$$

$L_{rD}$ 는 ground truth(gt) 밀도와와의 L1 손실(L1 loss),  $L_{pD}$ 는 gt 밀도와와의 지각 손실(perceptual loss)이다.  $L_{rV}$ 는 gt 속도와의 L1 손실,  $L_{pV}$ 는 gt 속도와의 지각 손실이다.  $L_{gV}$ 는 gt 속도 그라디언트와의 L1 손실이다.  $L_I$ 는  $\bar{\rho}_0$ 와  $\rho_1$  사이를 보간 네트워크를 통해 보간한 결과와 gt의 L1 손실이다.  $\lambda_r, \dots$ 은 각 손실함수의 각 항의 가중치이다.

### 2.2. 보간 네트워크 (Interpolation Network)

보간 네트워크는 연속한 두 프레임 사이의 밀도를 U-Net 기반 인공신경망을 통해 보간하는 모델[2]을 사용하였다.

## 3. 실험 및 결과

MantaFlow[4]를 이용하여 학습 및 실험 데이터를 생성하였다. 작은 타임스텝은 0.5, 큰 타임스텝은 4.0으로 설정하였다. 학습 데이터는 시작 조건을 동일하게 한 후 서로 다른 타임스텝으로 시뮬레이션한 데이터 1000개로 구성된다. 실험 데이터는 총 200개로, 데이터 하나당 128프레임으로 구성하였다. 실험 단계에는 큰 타임스텝으로 한 스텝 계산한 후, 제시한 모델로 보정 및 보간하고, 보정된 결과로부터 다시 큰 타임스텝으로 계산을 이어나가는 것을 반복하였다. 이 결과를 물리 기반 방식에서 작은 타임스텝으로 계산한 결과와 정확도 및 속도를 비교하였다.

### 3.1. 정확도 향상

그림 4는 보정 전의 큰 타임스텝 시뮬레이션 결과와 같은 시간 동안의 gt 작은 타임스텝 시뮬레이션 결과, 보정 후의 결과를 비교한 사진이다. 제시한 모델이 입력값을 gt와 근접하게 보정하였음을 확인할 수 있다. gt와의 평균 L1 Loss가 보정 전 0.05에서 보정 후 0.02로 50% 이상 감소하였다.

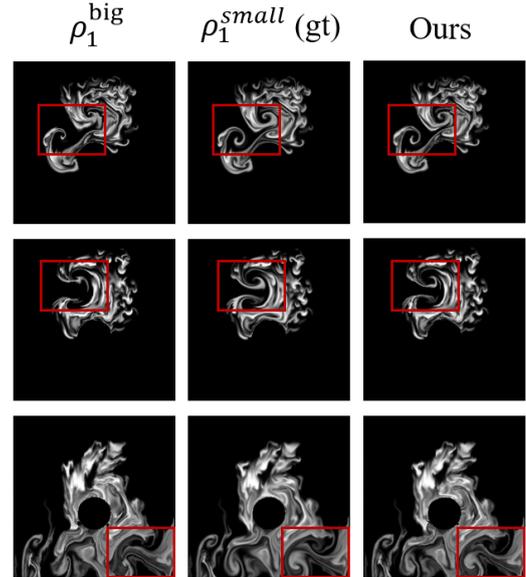


그림 4: 제안 방법의 정성적 평가

표 1: 128프레임 생성에 걸린 총 시간 평균

	Physics-Based	DeepFluids	ours
Time (s)	6.12	5.36	4.70

### 3.2. 속도 향상

표 1은 본 모델을 이용한 프레임 생성의 총 시간을 물리 기반 시뮬레이션 및 DeepFluids[1]와 비교한 결과이다. 총 시간은 모델의 인퍼런스(inference) 시간과 시뮬레이션 시간을 포함한다. 각각에 대해 약 30%, 14%의 속도 향상이 있었다.

## 4. 결론

본 논문에서는 딥러닝을 이용한 프레임의 보정 및 보간을 통한 새로운 연기 시뮬레이션 방법을 제시하였다. 제시한 방법은 기존의 보간 기반 방식[2]보다 정확하며, 기존의 딥러닝 기반 생성 방식[1]보다도 빠르다.

## 참고문헌

- [1] Kim, Byungsoo, et al, Deep fluids: A generative network for parameterized fluid simulations, *Computer Graphics Forum*, Vol. 38, No. 2, 2019.
- [2] Oh, Young Jin, and In-Kwon Lee, Two-step Temporal Interpolation Network Using Forward Advection for Efficient Smoke Simulation, *Computer Graphics Forum*, Vol. 40, No. 2, 2021.
- [3] Wang, Ting-Chun, et al, "Video-to-video synthesis." *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018.
- [4] THUEREY N., PFAFF T.: MantaFlow, URL: <http://mantaflow.com>, 2018

# 저가형 모션 캡처 장비를 이용한 실시간 상호작용 애니메이션 시스템\*

김정호<sup>0</sup>, 강다은, 권태수  
 한양대학교 일반대학원 컴퓨터소프트웨어학과  
 {cpls93, teddysiah, taesoo}@hanyang.ac.kr

## Real-time Interactive Animation System for Low-Priced Motion Capture Sensors

Jeongho Kim<sup>0</sup>, Daeun Kang, Taesoo Kwon  
 Department of Computer Science, Hanyang University Graduate School, South Korea

### 요약

본 논문은 입력 정확도가 낮은 저가형(보급형) 모션 캡처 장비를 이용하여, 사용자 아바타와 가상의 아바타가 실시간으로 상호작용하는 애니메이션 시스템을 연구하였다. 전처리 과정에서는 예제 동작 데이터베이스 내에 있는 동작을 가상 환경의 아바타에 적용하기 위한 리타게팅 및 동작 맵핑을 수행한다. 실시간 과정에서는 사용자 입력 동작에 적합한 주 캐릭터 동작 및 이에 대응하는 부 캐릭터 동작을 선별하여 사용자 및 가상 아바타에 합성한 후, 가상 환경에 적합하게 최종적으로 보정한다. 실험 결과 사용자 아바타의 인사, 악수, 터치 등 다양한 동작 시도에 대하여 가상 아바타 또한 적절한 반응 동작을 취하며, 비슷한 유형의 동작이라도 신체부위의 높낮이나 각도와 같은 다양한 변수에도 적절히 대응한다.

### 1. 서론

최근 가상현실(virtual reality, VR) 및 증강현실(augmented reality, AR) 기술을 활용한 콘텐츠의 수요가 꾸준히 증가하고 있으며, 이에 발맞추어 관련 하드웨어 및 소프트웨어 기술 또한 지속적인 연구와 발전을 거듭하고 있다. 특히 사용자가 직접 가상의 환경에 참여하는 상호작용 형태의 콘텐츠 또한 관심을 이끌고 있다.

본 논문은 입력 정확도가 낮은 저가형(보급형) 모션 캡처 장비를 이용하여, 사용자 아바타와 가상의 아바타가 실시간으로 상호작용하는 애니메이션 시스템을 연구하였다. Kinect™[1]을 이용하여 사용자가 실시간으로 동작(motion)을 입력하는데 이를 그대로 사용하면 노이즈로 인해 부자연스러운 동작이 생성된다. 이것을 보완하기 위해 기존의 연구들에서는 정교한 모션 캡처를 이용하여 생성한 동작을 예제 동작으로 사용한다.[2] 본 논문에서도 이를 활용하여 Kinect 센서로 받은 입력을 기반으로 사용자 아바타는 예제 동작과 비슷한 동작을 가져와 보정한 후 이것을 수행한다. 디지털 아바타는 이에 대응하는 반응 동작을 실시간으로 출력한다.

### 2. 실시간 상호작용 애니메이션 시스템

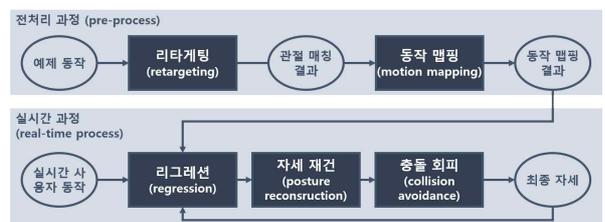


그림 1 상호작용 애니메이션 시스템 흐름도

그림 1은 해당 시스템이 작동하는 과정의 흐름을 묘사한다. 사전에 입력되는 예제 동작 데이터베이스는 주 캐릭터(main character)와 부 캐릭터(sub character)가 상호작용하는 동작을 포함하고 있다. 전처리 과정에서는 예제 동작 데이터베이스 내에 있는 동작을 가상 환경의 아바타에 적용하기 위한 리타게팅 및 동작 맵핑을 수행한다. 실시간 과정에서는 사용자 입력 동작에 적합한 주 캐릭터 동작 및 이에 대응하는 부 캐릭터 동작을 선별하여 사용자 및 가상 아바타에 합성한 후, 가상 환경에 적합하게 최종적으로 보정한다.

#### 2.1. 전처리 과정 (pre-process)

● **리타게팅**: 일반적으로 예제 동작 데이터베이스의 캐릭터 모델과 실제 가상환경에서 사용하는 캐릭터 모델의 골격구조가 다르므로, 서로 다른 두 캐릭터 간의 골격구조 및 비율(길이)를 비교하여 각 관절부위를 대응시키는 작업이 필요하다. 본 연구에서는 사용자가 기본자세(T-pose)를 취한 후 아바타와 길이를 비교하여 각 관절부위를 매칭시키는 반자동 인터페이스를 이용함으로써 사용자가 손쉽게 리타게팅을 진행할 수 있도록 설계하였다.[3]

● **동작 맵핑**: 주어진 예제 동작 데이터베이스로부터 주 캐릭터와 부 캐릭터의 각 프레임 별 상대적 위치 정보를 계산하여 저장한다. 이는 추후 실시간 과정에서 예제 동작을 효율적으로 활용하기 위하여, 동작의 정보를 처리하기 쉬운 형태로 재가공하는 전처리 과정이다. 예제 캐릭터의 동작 변수들은 표 1과 같이 최소화된 변수들을 갖는 단순화된 물리 모델(simplified physics models) 캐릭터에 대응하여 재계산되는데, k 근접 역거리 가중치 기법(k-nearest neighbor inverse distance weighting, KNN-IDW)을 사용하여 보간된 변수값을 도출하였다. KNN-IDW는 다차원 검색 트리(k-dimensional tree) 형태의 데이터에서 빠르게 작동하므로 비선형 리그레션에 유리하게 활용할 수 있다.

\* 구두(포스터) 발표논문  
 \* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출 중.  
 \* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행되었음 (NRF-2020R1A2C1012847).

표 1 단순화된 물리 모델의 상태 및 조절 변수

상태	제어	의미
$\alpha_y$	$\dot{\alpha}_y$	캐릭터가 횡평면에 대해 이루는 각도/각속도
$\delta$	$\dot{\delta}$	골반 설정의 오프셋/해당 미분값
$p$	$\dot{p}$	각 관절의 현재 각도/해당 미분값

2.2. 실시간 과정 (real-time process)

● **리그래션**: 단순화된 모델의 다음 이동 경로를 계산하고, 실시간 사용자 입력 및 주변 환경의 영향을 반영하여 목표 자세를 결정한다. 매 프레임마다 단순화된 모델을 모델링하고, 현재의 특징 벡터(feature vectors)를 토대로 KNN-IDW를 이용하여 제어 변수 및 맵핑 변수를 도출한다. 이 때 제어 변수는 단순화된 캐릭터 모델의 자세를 변화시키는 역할을 하고, 맵핑 변수는 단순화된 모델과 실제 사용하는 캐릭터 모델 간의 맵핑 관계를 표시하는 역할을 한다. 예제 데이터베이스를 토대로 다음 프레임에 예상되는 발의 상태에 따라 KNN-IDW를 여러 번 사용하고, 그 중에서 현재 상태에 가장 알맞은 KNN-IDW 모델을 하나 선택한다.

● **자세 재건**: 사용자 및 가상 캐릭터가 현재 자세로부터 리그래션을 통해 계산된 다음 목표 자세로 관절을 움직이도록 캐릭터 모델을 제어한다. 동작 맵핑 과정에서 추정된 다음 자세와 현재 관절 속도의 적분값을 선형 보간함으로써 연속적인 다음 자세를 결정하고 이에 따라 캐릭터 자세를 제어하였다[4].

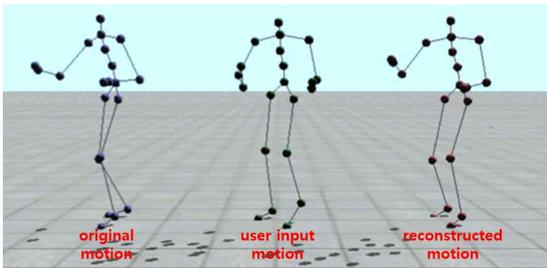


그림 2 예제 동작과 실제 사용자 입력 동작, 최종 출력 동작의 비교

그림 2는 예제 동작과 실제 사용자 입력 동작, 최종 출력 동작을 비교한 것으로, 사용자가 시스템에 내제된 동작(예제 동작)과 정확히 일치하지 않는 동작을 취하더라도 예제 데이터베이스로부터 가장 유사한 동작을 검색하여 수정을 거쳐 출력하는 것을 확인할 수 있다. 이 과정에서 모션 캡처 장비의 노이즈와 트래킹 오류 또한 개선될 수 있다.

● **충돌 회피**: 리그래션을 통해 도출된 각각의 캐릭터의 자세는 자연스러우나, 보간 오류 및 오차 누적 등으로 인하여 두 캐릭터의 상호작용이 부자연스러운 경우가 발생한다(예: 각자 악수하는 포즈를 취하고 있으나 서로의 손이 맞닿지 않음). 이러한 문제점을 개선하기 위하여, 상호작용이 발생하는 관절 부위를 엔드 이펙터(end effectors)로서 별도 지정하고, 적합한 상호작용을 취할 수 있도록 자세를 수정한다. 엔드 이펙터의 현재 상태를 감지하고, 부자연스러운 상태 감지 시 및 목표 위치와 방향, 깊이 값을 측정하여 자세를 수정한다. 목표 자세는 엔드 이펙터 및 그에 가까운 관절의 순서대로, 각 위치를 목표 위치 방향으로 끌어내는 방식으로 역기구학

(inverse kinematics, IK) 풀이를 반복수행하여 계산한다. 이 때, 자연스러운 자세 유지를 위하여 각 관절의 위치 수정은 1cm 이내로 제한하였다.

3. 실험결과 및 결론

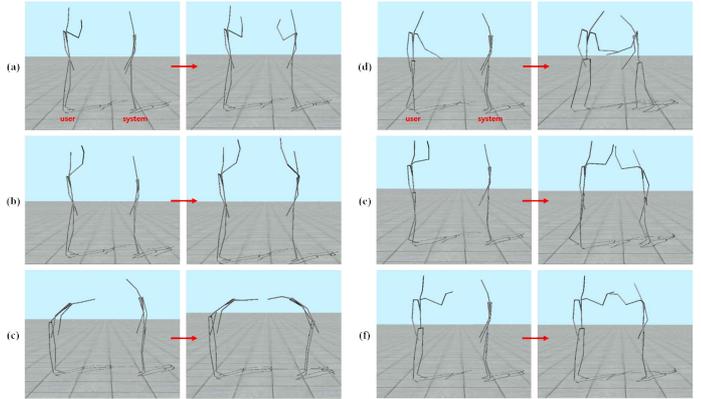


그림 3 사용자의 실시간 동작 입력에 따른 아바타 간 상호작용 실험 결과: (a) 손을 흔들어 인사, (b) 손을 높이 흔들어 인사 (c) 허리를 굽혀 인사, (d) 악수, (e) 손바닥 맞대기, (f) 주먹 맞대기

그림 3은 실시간 사용자 동작 입력에 대하여 사용자 아바타와 가상 아바타가 실시간으로 상호작용하는 것을 실험한 결과이다. 사용자 아바타의 인사, 악수, 터치 등 다양한 동작 시도에 대하여 가상 아바타 또한 적절한 반응 동작을 취하며, 비슷한 유형의 동작이라도 신체부위의 높낮이나 각도와 같은 다양한 변수(예: 손을 흔드는 높낮이)에도 적절히 대응한다.

본 연구는 입력 정확도가 낮은 저가형 모션 캡처 장비를 이용하여 사용자의 입력 동작을 사용자 아바타에 적용하고, 가상의 아바타가 상호작용 동작으로 응대하는 실시간 애니메이션 시스템을 개발하였다. 예제에 없는 동작을 입력하면 모호한 동작이 나올 수 있는 것, 예제 동작에서와 비슷한 골격 구조를 가져야 시스템이 작동한다는 것이 본 연구의 한계점이다. 현재 딥러닝의 활용이 높아짐에 따라 향후 연구에는 본 논문에서 사용된 KNN-IDW와 딥러닝의 성능을 비교할 예정이다.

참고문헌

[1] Z. Zhang, Microsoft kinect sensor and its effect, *IEEE multimedia*, 19(2), 4-10, 2012.  
 [2] Victor Brian Zordan and Jessica K Hodgins. "Motion capture-driven simulations that hit and react". In: *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*. ACM. 2002,pp. 89-96.  
 [3] Hwang, J., Kim, K., Suh, I. H., & Kwon, T. (2017). Performance-based animation using constraints for virtual object manipulation. *IEEE computer graphics and applications*, 37(4), 95-102.  
 [4] Y. Lee, K. Wampler, G. Bernstein, J. Popović, & Z. Popović, Motion fields for interactive character locomotion. In *ACM SIGGRAPH Asia 2010 papers* (pp. 1-8), 2010.

# 자연스러운 피격 반응 생성\*

박수환<sup>0</sup>, 이제희  
 서울대학교 컴퓨터공학부  
 {shpark, jehee}@mrl.snu.ac.kr

## Generating Natural Hit Reaction

Soohwan Park<sup>0</sup>, Jehee Lee  
 Dept. of Computer Science and Engineering, Seoul National University

### 요약

본 연구는 다양한 종류의 피격에 대해 자연스럽게 반응할 수 있는 가상의 캐릭터를 생성하는 것을 목표로 한다. 피격 반응은 외부와의 상호작용으로 인해 발생하는 동작으로 다양한 조건(세기, 방향, 부위, 캐릭터의 자세)에 영향을 받는다. 물리적 상호작용에 큰 영향을 받는 동작이기 때문에 물리적으로 부자연스러운 동작이 생성 경우 그 결과가 매우 어색해 보이게 된다. 우리는 물리 시뮬레이션과 모션 매칭을 기반으로 하여 피격 조건에 맞는 자연스러운 반응 동작을 생성할 수 있는 시스템을 제안한다. 또한, 위의 시스템을 mixture of experts 기반 네트워크에 학습시킴으로써 런타임에 실시간으로 피격 반응 동작을 생성할 수 있도록 한다. 우리는 이 시스템이 다양한 종류의 피격 상황에서 적절히 동작하는 여러 예제를 보일 것이다.

### 1. 서론

다양한 종류의 게임에서 피격 반응은 매우 빈번하게 요구된다. 사용자 캐릭터의 공격에 대해 상대 캐릭터가 적절한 피격 반응을 보이도록 하는 것은 공격의 타격감에 매우 중요한 영향을 미치며 게임의 퀄리티에도 큰 영향을 미친다. 최근 AAA급 게임들은 게임의 퀄리티를 높이기 위해 많은 노력을 기울이고 있으며, 이에 따라 자연스러운 피격 반응을 생성하기 위한 방법 역시 많은 연구가 이루어지고 있다. 피격 반응은 다른 동작들과 달리 외부와의 물리적 상호작용에 의해 생성되는 동작이다. 따라서 자연스러운 반응 동작을 위해선 얼마나 사람처럼 자연스럽게 행동하는지 뿐 아니라 외부의 충격에 대한 물리적 적합성 역시 중요하다. 모션 캡처와 기존의 시공간 기반 모션 에디팅 방식으로는 물리적 충격에 대

해 적절한 동작을 합성하는 것이 어렵다. 물리 시뮬레이션을 기반으로 심층 강화학습을 통해 동작을 생성하는 방식들[1]의 경우, 가상의 물리 환경에서 피격 상황을 직접 시뮬레이션해서 동작을 생성하는 것이 가능하지만, 심층 강화학습과 PD컨트롤러의 복원력으로 인해 매우 뻣뻣한 부자연스러운 피격 반응이 생성된다.

우리는 최근 게임업계에서 널리 사용되는 모션 매칭[2] 방식과 물리 시뮬레이션을 결합하여 충격에 대해 물리적으로 자연스럽게 반응하면서 동시에 사람답게 반응하는 자연스러운 피격 반응 생성 시스템을 제안한다. 또한, 이 시스템을 mixture of experts 기반 네트워크[3]에 학습시킴으로써 런타임의 연산 속도와 메모리 사용량을 개선할 것이다. 시스템을 통해 합성될 수 있는 피격 데이터는 수백~수천 시간에 달하는데, 이 데이터를 적절히 샘플링해서 네트워크에 학습시킬 수 있도록 하는 알고리즘을 제안한다.

### 2. 물리 기반 피격 반응 생성 시스템

충격에 대해 물리적으로 자연스러우면서 동시에 사람처럼 자연스러운 동작을 생성하기 위해, 우리는 실제 사람의 피격 반응 방식을 기반으로 하여 피격 반응 생성 시스템을 설계했다. 사람은 예상치 못한 갑작스러운 충격이 발생할 경우 이를 인지하고 반응하기 까지 약 100~250ms의 지연이 발생한다. 즉 피격 직후의 짧은 시간 동안은 사람의 의지가 아닌 물리법칙에 크게 영향을 받는 동작이 생성되며, 그 이후부터 사람의 의지에 의한 동작이 나타나게 된다. 우리는 이런 사람의 반응 특성을 기반으로 순간 피격 반응과 후속 피격 반응의 두 단계로 구성되는 시스템을 설계하였다.

순간 피격 반응은 피격 직후의 짧은 순간 동안의 피격 반응으로, 사람이 아직 피격을 인지하지 못한 상황에서 생성되는 반응 동작을 의미한다. 이 구간에서는 충격에 대한 물리적 반응이 동작 생성에 지배적인 영향을 미친다. 우리는 물리 시뮬레이션을 통해 이 구간에서의 동작을 생성하였다. 피격이 발생하면 그 순간의 캐릭터 자세에 주어진 피격 조건에 따른 힘을 가해서 물리 시뮬레

\* 구두발표논문

\* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00878, SW컴퓨팅산업원천기술개발사업(SW스타랩)).

이션을 진행하여 동작을 생성한다. 순간 피격 반응 구간은 약 167ms로서 30hz로 정의되는 모션에서 5프레임으로 정의된다. 물리 시뮬레이션이 진행되는 동안 캐릭터는 PD 컨트롤을 이용해 기존의 자세를 유지하려고 하는 약한 힘을 내게 된다.

후속 피격 반응은 순간 피격 반응 이후에 이어지는, 사람의 의지에 의해 동작이 생성되는 구간이다. 사람과 같은 자연스러운 동작을 생성하기 위하여 우리는 피격 동작이 포함된 모션 데이터베이스를 기반으로 모션 매칭 방식을 사용하여 동작을 생성한다. 순간 피격 구간이 종료되면 종료 순간의 캐릭터 자세와 피격 조건을 바탕으로 모션 데이터베이스에서 가장 적합한 피격 동작을 탐색한다. 모션의 적합성을 비교하기 위한 피쳐는 캐릭터의 자세와 속도, 엔드 이펙터의 위치와 속도, 피격 조건이 사용된다. 데이터베이스에서 가장 적합한 모션을 찾아내면 캐릭터는 해당 모션으로 블렌딩되며 동작을 생성하게 된다.

### 3. 네트워크 학습

Mixture of experts 네트워크 모델은 최근 여러 동작 생성 연구에서 좋은 성능을 보여주고 있는 네트워크 모델이다. 여러 개의 expert 네트워크와 하나의 gating 네트워크로 구성되며, gating 네트워크의 출력을 기반으로 experts 네트워크의 출력을 블렌딩 해서 최종 출력을 결정한다. 이 모델은 다양한 동작을 하나의 네트워크에 효과적으로 학습할 수 있다. 우리의 연구 역시 수많은 피격 동작들을 하나의 네트워크에 학습하는 것을 목표로 하며 따라서 이 모델을 기반으로 네트워크를 디자인했다.

$$p_t = \text{MOE}(p_{t-1}, c_{t-1}, \dots, p_{t-n}, c_{t-n})$$

네트워크는 지난 n프레임의 캐릭터 자세와 피격 조건을 입력으로 받아서 다음 프레임의 자세를 생성한다. 캐릭터의 자세는 캐릭터 각 조인트의 로테이션 매트릭스로 표현되며, 피격 조건은 방향, 부위, 세기로 구성된다.

#### 3.1. 학습 데이터 생성

물리 기반 피격 반응 생성 시스템은 주어진 피격 조건과 캐릭터의 상태로부터 피격 반응 동작을 생성한다. 물리 시뮬레이션과 블렌딩에 의해 기존의 모션데이터에서 벗어난 동작들이 생성되며 이는 피격 조건에 따라 무수히 늘어날 수 있다. 약 한 시간 분량의 모션 데이터에 주어진 경우, 그로부터 생성될 수 있는 피격 반응 데이터의 양은 수백~수천 시간에 달한다. 이 피격 반응들을 모두 생성해서 학습하는 것은 현실적으로 불가능하기 때문에, 우리는 이를 적절히 샘플링해서 하나의 네트워크에 학습시킬 수 있는 알고리즘을 디자인했다.

시스템에서 생성되는 동작은 물리 시뮬레이션에 의해 생성되는 순간 피격 반응 구간과 그 후 모션데이터로 블렌딩 되는 구간을 제외하면 기존의 모션데이터를 그대로 재현하게 된다. 따라서 일반적인 데이터 생성 방식처럼 랜덤한 피격 조건으로 데이터를 생성하게 될 경우 중복되는 동작이 데이터의 대부분을 차지하고 정작 시스템에 의해 다양화되는 부분은 많이 포함되지 못하게 된다. 우리의 알고리즘은 학습용 피격 반응 동작을 생성할 때 중복되는 모션으로 수렴할 경우 더 이상 동작을 생성하지 않고 멈추고 새로운 피격 반응 동작을 생성하도록 함으로써, 같은 메모리로 더 다양한 피격 반응들을 포함 할 수 있도록 한다.

### 4. 결론

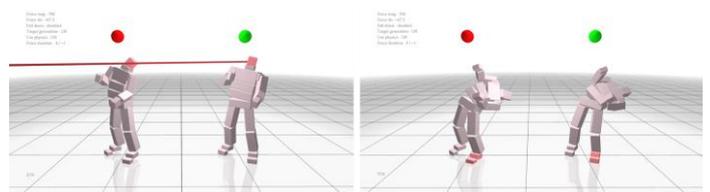


그림 1: 피격 시뮬레이션, 붉은색: 물리 기반 피격 반응 생성 시스템, 초록색: MOE 네트워크

본 논문에서 우리는 주어진 모션데이터로부터 다양한 피격조건에 대해 자연스러운 피격 반응을 생성 할 수 있는 시스템을 학습하는 방법을 제안했다. 이를 통해 약 한 시간 분량의 총에 의해 피격 당하는 데이터로부터 다양한 충격에 대해 적절히 반응하는 시스템을 학습 할 수 있었다. 시스템은 다양한 방향, 세기, 부위에 가해지는 충격에 대해 적절히 반응 할 수 있으며, 연속적으로 가해지는 충격에도 적합한 피격 반응을 생성 할 수 있었다.

### 참고문헌

[1] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne, Deepmimic: Example-guided deep reinforcement learning of physics-based character skills, *ACM Trans. Graph* 37(4): Article 143, 2018.  
 [2] Simon Clavet. 2016. Motion Matching and The Road to Next-Gen Animation. In *GDC 2016*  
 [3] Jungdam Won, Deepak Gopinath, and Jessica Hodgins, A Scalable Approach to Control Diverse Behaviors for Physically Simulated Characters, *ACM Trans. Graph*, 39(4): Article 33, 2020.

# 하나의 모션 클립으로부터 모션 집단 학습\*

이세영<sup>0</sup>, 이선민, 이용우, 이제희  
 서울대학교 컴퓨터공학부

{seyounglee, sunmin.lee, yongwoo.lee, jehee}@mrl.snu.ac.kr

## Learning a Family of Motor Skills from a Single Clip

Seyoung Lee<sup>0</sup>, Sunmin Lee, Yongwoo Lee, Jehee Lee  
 Seoul National University



그림 1: 매개변수화된 장애물 점프, 점프, 백플립, 카트휠, 킥 모션

### 요약

우리는 단일 모션 클립에서 매개 변수화된 모션 집단을 학습하는 새로운 알고리즘을 제시한다. 각 동작은 심층 정책 네트워크로 표현되며, 연속적인 동작 공간을 탐색하여 사용자 입력 및 환경 상호 작용에 대응하여 물리 시뮬레이션 상에서 모션을 생성한다. 모션 매개 변수화와 동작 추적의 동시 학습은 학습된 모터 기술의 성능과 시각적 품질을 크게 향상시킨다. 우리는 매우 역동적인 동작들을 작업 목표, 신체 비율, 물리적 수치 및 환경 조건에 대해 매개 변수화하여서 알고리즘의 유연성을 입증했다.

### 1. 서론

사람은 주어진 동작을 다른 동작으로 손쉽게 변형해낼 수 있다. 예를 들어서, 펀치하는 동작이 주어졌을 때 팔을 뒤로 더 당기고 허리의 회전을 이용해서 더 세게 펀치하는 동작을 생성해낼 수 있다. 마찬가지로, 동작 데이터를 일반화하기 위해서는 물리적 및 생물학적 원칙에 따라 운동 기술이 어떻게 변형될지를 알아야 한다. 이전의 접근 방식은 종종 동작을 일반화하기 위해 설명

하기 위해 많은 양의 동작 데이터를 이용했다. 이러한 대규모 데이터를 수집하려면 상당한 시간과 노력이 필요하다.

본 논문에서는 단일 모션 클립에서 매개 변수화된 모션 집단을 구성하는 새로운 접근 방식을 제시한다. 우리의 알고리즘은 기존 동작을 일반화하여 새로운 조건과 목표를 충족하는 다양한 동작을 생성한다. 매개 변수화된 동작은 심층 정책 네트워크로 표현되며, 지속적인 작업 공간을 탐색하여 사용자 입력 및 환경 상호 작용에 대응하여 물리 시뮬레이션 상에서 모션을 생성한다.

구체적으로, 두 가지 하위 문제를 해결해야 한다. 하나는 입력 동작을 모방하는 동작 기술을 배우는 것이고, 다른 하나는 연속적인 작업 공간을 형성하는 다양한 조건과 작업에 대해 물리적으로 유효한 방식으로 적응된 동작을 구하는 것이다. 모션 매개 변수화 네트워크를 명시적으로 구성하고 심층 정책 네트워크의 학습 중에 발생하는 시뮬레이션 튜플들을 활용하여 두 네트워크를 학습하는 방식으로 두 가지 문제를 동시에 해결할 수 있다. 우리의 새로운 알고리즘은 학습 튜플을 효과적으로 재사용하고 작업 공간 탐색을 계획함으로써 현실적인 시간 내에 넓은 범위의 동작을 생성할 수 있다.

우리는 매우 역동적인 동작들을 매개 변수화하여서 알고리즘의 유연성을 입증할 것이다. 우리의 캐릭터는 실시간 물리 시뮬레이션에서 다양한 운동 기술을 수행하기 위해 인터랙티브하게 시뮬레이션 및 제어된다.

### 2. 매개변수화 네트워크 및 동작 추적기의 동시 학습

시스템은 기본이 되는 모션 클립과 매개 변수화된 작업 공간을 입력으로 받는다 (그림 2 참조). 작업 공간은 사

\* 구두발표논문

\* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 ACM Transactions on Graphics (SIGGRAPH 2021) 에 게재 예정이다.

\* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00878, SW컴퓨팅산업원천기술개발사업(SW스타랩))

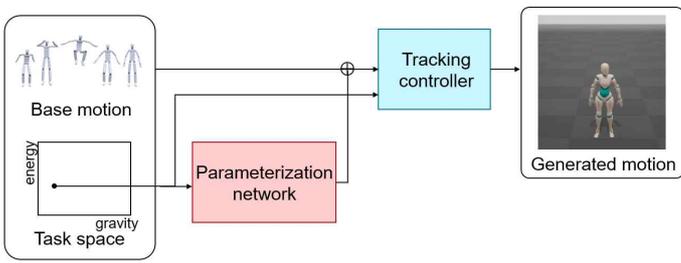


그림 2: 시스템 개요

용자가 변형으로 충족하고자 하는 작업의 조합이다. 작업 목표는 키, 몸무게 또는 관절의 최대 힘과 같은 캐릭터의 신체 조건이나 중력과 같은 환경 조건 또는 점프 높이 또는 펀치 위치와 같은 시공간적인 목표 등 다양하게 설정할 수 있다.

매개 변수화 네트워크는 작업 매개 변수에 해당하는 적절한 기준 모션을 생성하는 매핑 함수이다. 네트워크의 입력은 매개 변수이며, 네트워크의 출력은 기본 모션과의 변위 매핑이다. 기본 모션과 매개 변수화 네트워크의 출력을 추가하면 작업 매개 변수에 대한 새 모션이 생성된다.

위의 과정을 통해 생성된 모션이 추적 컨트롤러의 목표가 되는 모션이 된다. 추적 컨트롤러의 출력 모션은 물리 시뮬레이션 상에서 주어진 목표를 수행하는 동시에 레퍼런스 모션을 모방하는 모션이다. 우리의 네트워크는 심층 강화 학습 알고리즘인 proximal policy optimization [1] 을 통해서 학습되었으며, Deepmimic [2] 컨트롤러를 기반으로 작동한다. 심층 정책 네트워크는 학습된 분포로부터 동작을 확률적으로 출력하므로, 동일한 상태가 주어지더라도 다른 동작이 나온다. 그 결과, 하나의 레퍼런스 모션으로부터 다양한 동작이 생성될 수 있다.

매개 변수화 네트워크를 최적화하려면 매개 변수와 동작 쌍이 필요하다. 각 데이터는 추적 정책을 교육하는 과정에서 생성된 동작으로부터 얻을 수 있다. 우리는 hindsight experience replay [3]의 개념을 사용하여 시뮬레이션 롤아웃을 재사용한다. 추적 정책을 훈련하는 동안 생성된 대부분의 동작은 정확히 의도된 목표를 달성하지 못한다. 그것들은 의도된 목표에 대한 최적의 움직임 찾기를 찾는 데는 유용하지 않지만, 시뮬레이션된 움직임과 실제로 달성한 작업 사이의 상관 관계를 제공할 수 있다. 가능한 최고 품질의 동작을 매개 변수화 네트워크의 학습 데이터로 주기 위해서, 우리는 각 롤아웃이 처음에 주어진 모션과 얼마나 유사한지를 평가하여 가장 좋은 품질의 모션들만 학습 데이터로 제공한다.

### 3. 결과 및 결론

우리의 시뮬레이션 시스템은 C++로 작성되었으며 DART[4]를 기반으로 한다. 강화 학습 및 회귀 네트워크는 Python으로 작성되며 텐서플로 라이브러리를 기반으로 한다. 학습 계산 시간은 작업 공간의 크기와 범위에 따라 다르다. AMD Ryzen 93950x를 사용하는 단일 PC에서는 12시간에서 72시간이 걸린다. 우리는 밀기, 펀치, 킥, 백플립, 카트휠, 점프, 장애물 점프 등 다양한

동작을 가지고 실험을 했다 (그림 3 참조). 학습된 매개 변수화된 동작은 사진과 같다. 이를 통해 우리의 알고리즘이 역동적이고 변형이 큰 동작을 포함한 다양한 매개 변수와 동작에 대해서 높은 퀄리티의 동작을 생성해낼 수 있다는 것을 확인할 수 있다.

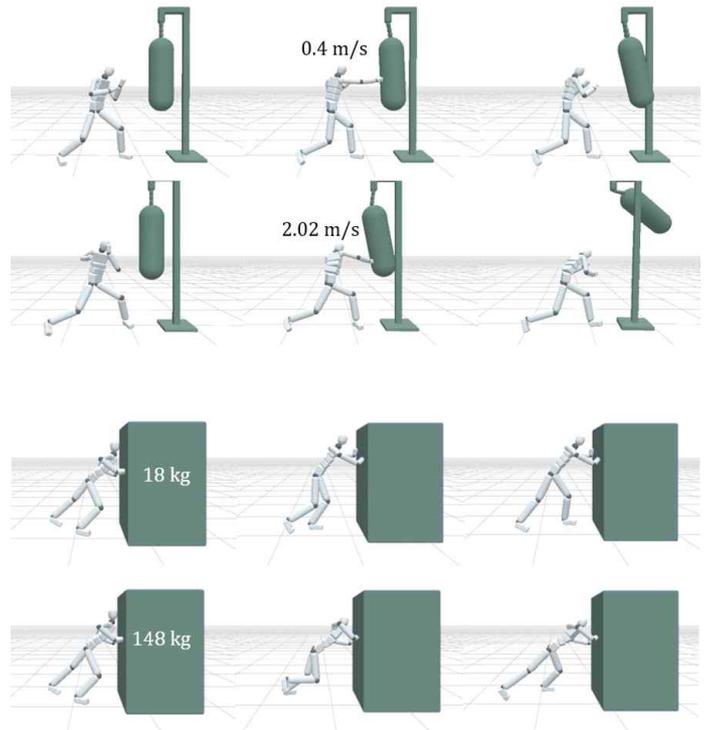


그림 3: 세기에 대해 매개 변수화된 펀치 동작과 물체의 무게에 대해 매개 변수화된 미는 동작

### 참고문헌

[1] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov, 2017, Proximal Policy Optimization Algorithms, CoRR abs/1707.06347.  
 [2] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne, 2018, Deepmimic: Example-guided deep reinforcement learning of physics-based character skills, ACM Transactions on Graphics 37, 4, Article 143.  
 [3] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba, 2017, Hindsight Experience Replay. In Advances in Neural Information Processing Systems, 5048–5058.  
 [4] Jeongseok Lee, Michael X Grey, Sehoon Ha, Tobias Kunz, Sumit Jain, Yuting Ye, Siddhartha S Srinivasa, Mike Stilman, and C Karen Liu, 2018, DART: Dynamic animation and robotics toolkit, The Journal of Open Source Software 3, 22.

# 사람의 운동을 이용한 사족 보행 로봇 제어기\*

김선우<sup>0</sup>, 막심 소로킨, 하세훈, 이제희  
 서울대학교 컴퓨터 공학과  
 조지아 공과대학교 인터랙티브 컴퓨팅스쿨  
 {sunwoo, jehee}@mrl.snu.ac.kr  
 {maks, sehoonha}@gatech.edu

## Human Embodied Control of Legged Robots using Deep Reinforcement Learning

Sunwoo Kim<sup>0</sup>, Maksim Sorokin, Sehoon Ha, Jehee Lee  
 Dept of Computer Science, Seoul National University  
 School of Interactive Computing, Georgia Institute of Technology

### 요약

본 연구에서는 사람의 운동을 이용하여 사족 보행 로봇을 제어하는 새로운 시스템을 제시한다. 사람의 운동을 통한 제어는 사족 보행 로봇에게 한정되지 않은 다양한 제어 목표를 갖고 즉각적으로 반응하여 움직일 수 있게끔 한다. 이와 같은 시스템을 만들기 위해서 본 연구에서는 기존의 직접적인 방식의 자세 매핑 방법이 아닌 인공지능망을 이용한 매핑 방식을 제안하였다. 또한, 강화학습을 통해 로봇의 운동 정책을 수립하였다. 마지막으로 새로운 제어 시스템 모델을 제시함으로써, 사람의 자유분방한 움직임을 통해 사족 보행 로봇 제어에 제한을 없애고 다양한 움직임을 생성해냈다. 본 연구에서는 보행, 뛰기, 팔 뻗기, 몸체 회전 및 굽히기 등 다양한 사람의 운동을 통해 물리 시뮬레이션 상의 사족 보행 로봇을 제어할 수 있음을 입증하였다.

### 1. 서론

사족 보행 로봇은 다양한 상황에서 사람을 대체할 가능성이 있다. 특히, 재난 현장 등 안전이 위협받는 상황에서 사람 대신 투입되어 여러 임무를 수행할 수 있는 능력을 보유하고 있다. 지난 수십 년간, 로봇공학자들은 고수준 명령을 통한 로봇을 제어하는 방법들을 제안하였다.[1] 하지만, 이런 고수준 명령들은 즉각적인 반응이 어려우며, 명령 내에서만 수행 가능한 행동이 제한되는 단점이 있다. 실제 세상의 임무들은 예측 불가능하며, 때로는 로봇이 비구조화된 조종 제어를 해야 하는 상황에 놓이기도 한다. 그렇기에, 미리 개발된 제어기가 아닌 다양한 상황에서 제어 가능하며 즉각적인 반응을 보이는 시스템이 필요하다.

본 연구에서는 사족 보행 로봇의 새로운 제어 방법을 제안한다. 사람의 운동 동작을 이용하여 로봇에게 제어 목표를 수립하도록 한다. 이 방법을 통해 기존의 고수준 제어 명령에 따른 한정적인 로봇의 움직임이 아닌, 다양한 움직임의 제어가 가능해진다. 이 제어 시스템을 구축할 때 인공지능망을 통한 자세 매핑 방법을 새로이 고안하여 제안하였다.

### 2. 시스템 개요

사람의 움직임을 이용하여 사족 보행 로봇의 제어를 수행하기 위한 전체의 시스템은 다음과 같다. 우선 입력으로 Microsoft Azure Kinect를 이용하여 사람의 움직임을 감지한다. 사람 각 관절에 해당하는 세계좌표를 관측할 수 있다. 관측한 사람의 각 관절의 좌표를 정규화된 사람의 좌표 공간으로 치환한다. 정규화된 사람의 자세에 대응하는 사족보행 로봇의 자세를 매핑 인공지능망을 통하여 생성한다. 이전에 생성된 로봇의 자세 기록을 이용하여 현재 생성된 로봇의 자세를 부드러운 참고 자세가 될 수 있도록 조절한다. 생성된 자세를 제어 목표로 하는 강화학습 정책을 이용하여 최종적으로 로봇의 모터의 토크를 제어하여 사족 보행의 움직임을 생성함이 전체 시스템의 구조이다. 이 시스템을 수립하기 위해 두 가지 부분을 딥러닝을 통하여 미리 학습시켜 두어야 한다. 첫째, 사람의 자세에 대응하는 사족 보행 로봇의 자세를 학습해야 한다. 이 부분은 운동학적인 자세를 매핑해주는 역할을 하며 최종적으로 사람의 자세를 사족 보행 로봇의 자세에 대응시켜 사족 보행 로봇이 달성해야 할 자세의 목표를 수립하게 만든다. 두 번째 부분은 로봇의 운동 정책을 학습시키는 부분이다. 물리 시뮬레이션 상에서 로봇이 균형을 잃지 않으면서 목표 자세를 따라갈 수 있게끔 제어 정책을 수립한다.

### 3. 운동학적 자세 매핑

한 물체의 운동학적 자세를 다른 물체로 대응시키는 방법은 여러 분야에서 중요하다. 특히 사람의 자세를 로봇에 매핑시켜 적용하는 방안은 여러 로봇공학 분야와 컴퓨터 애니메이션 분야에서 연구되었던 주제이다. 한 자세를 다른 자세로 매핑하는 전통적인 방법으로는 직접 매핑의 법칙을 기술하는 방법이 있다. 이 방법은 주로

\* 구두 발표논문

\* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00878, SW컴퓨팅산업원천기술개발사업(SW스타랩))

최적화 목표를 설계하여 주어진 제한 조건 내에서 목표를 가장 만족시키는 결과를 사용하는 방법이다. 예를 들어 팔을 뻗은 길이를 보존시키거나, 팔목의 각도를 유지하는 등을 최적화 목표로 주어, 한 자세로부터 다른 자세를 생성해 낼 수 있다. 이 방법의 장점은 정확한 목표를 만족시키는 자세들의 매핑을 얻을 수 있다는 점이다. 하지만, 사용자가 정확한 최적화 목표를 주어야 한다는 한계점이 있다.

본 연구에서는 사람의 자세로부터 사족 보행 로봇의 자세를 매핑한다. 여기서, 사람과 사족 보행 로봇의 구조가 다르므로 자세를 매핑하는 최적화 목표를 정의하기 모호하다. 또한, 사람의 자세에 해당하는 의미를 사족 보행 로봇의 자세에 해당하는 의미로 바꾸는 최적화 목표를 정의하고 푸는데에 많은 시간이 소요된다. 본 연구에서는 위의 문제를 해결하기 위해 인공신경망을 이용하여 매핑을 생성하였다. 이는 사용자가 선택한 몇 개의 사람의 자세와 그에 해당하는 사족 보행 로봇의 자세 쌍을 이용하여 사람 자세 공간에 해당하는 전체 자세에 대응하는 로봇의 자세를 생성한다. 이를 통해 사용자가 선택한 자세 쌍의 의도를 전체 매핑에 반영시킬 수 있다. 그뿐만 아니라 적은 양의 자세 쌍만을 사용자가 선택한다면 나머지 사람의 자세들에 해당하는 대응 자세를 쉽게 얻을 수 있다. 본 연구에서는 다층 퍼셉트론을 학습시켜 사람의 자세로부터 사족 보행 로봇의 자세를 표현할 수 있게끔 하였다.

#### 4. 강화학습을 통한 물리 환경에서의 운동 정책

##### 4.1. 물리 환경에서의 운동 모사 학습

매핑 된 사족 보행 로봇의 자세는 물리적인 환경을 고려하지 않은 운동학적 자세일 뿐이다. 위의 매핑에서 얻은 자세를 곧바로 PD(Proportional-derivative) 제어를 통하여 로봇의 운동을 생성하면 균형을 잃는 등, 원하는 제어 목표를 달성할 수 없다. 기존의 연구들에서 이와 같은 문제를 해결하기 위하여 강화학습 방법, 특히 운동 모사 방법을 사용하였다.[2] 이는 주어진 참고 자세를 최대한 모사하면서 균형을 잃지 않는 제어 정책을 수립하는 강화학습 방법이다. 본 시스템에서는 사람의 행동에 제한을 두고 있지 않기 때문에, 이에 따라 생성된 사족 보행 로봇의 운동도 넓은 범위에 분포한다. 넓은 범위의 운동을 모두 모사할 수 있으면서도 균형을 잃지 않는 제어 정책을 학습하기 위해서는 다량의 참조 자세 배열을 학습해야 한다. 본 연구에서는 사람의 자세들을 임의로 대량으로 생성한 뒤, 매핑을 통해 참고 자세를 대량으로 생성하였다. 강화학습 단계의 시뮬레이션마다 생성된 참고 자세 중 임의로 몇 개의 자세들을 선택한다. 선택된 자세들을 보간하여 운동 배열을 생성하여 이를 모사하는 강화학습 정책을 학습하였다. 이 과정을 통하여 다양한 자세로부터 생성될 수 있는 넓은 범위의 운동을 강화학습 정책이 경험할 수 있도록 만들었다.

##### 4.2. 기민한 움직임에 대한 학습

참고 자세 학습 시, 참고 자세가 기민하게 바뀌는 환경에서 종종 제어 정책 수립에 실패한다. 이는 기초적인

자세를 학습하지 못한 상태에서 빠르게 변화하는 운동에 적응하기 어렵기 때문이다. 이를 해결하기 위해서 본 연구에서는 커리큘럼 학습을 도입하였다. 커리큘럼 학습의 기본 원리는 쉬운 목표부터 학습한 뒤 어려운 목표를 학습하는 것이다. 예를 들어 사족 보행 로봇의 보행 학습이 있다고 가정하자. 보행 높이와 보폭이 0.05초마다 변하는 보행을 학습시키면 많은 경우 학습에 실패한다. 커리큘럼 학습을 이용하면 우선 하나의 고정된 보행 높이와 보폭의 보행을 학습한다. 이후 1초마다 이 변수들이 바뀌는 경우를 학습, 0.5초마다 바뀌는 경우를 학습하여 최종적으로 0.05초마다 보행 높이와 보폭이 바뀌는 보행을 학습할 수 있다.

#### 5. 결론 및 추후 연구

본 연구를 통하여 사람의 운동을 이용한 시뮬레이션 상의 사족 보행 로봇 제어를 하였다. 사람의 운동을 통하여 사족보행 로봇의 즉각적이고 격렬한 운동들: 보행, 뛰기, 몸 기울이기, 한쪽 다리 뻗어 물체 밀기 등을 그림 1과 같이 수행할 수 있었다.

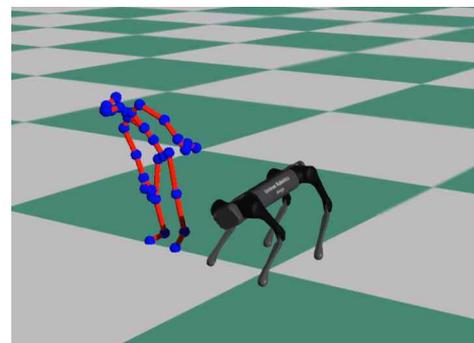


그림 1. 사람의 몸 기울임으로 제어한 로봇의 기울임 운동

본 연구의 추후 목표는 두 가지이다. 첫째는 역학을 고려한 매핑 함수 구축이다. 현재는 자세에서 자세로 매핑만을 고려했다. 역학적인 상태를 고려한 자세 매핑 방법을 찾는다면, 현재의 운동 상태를 고려한 제어를 만들 수 있을 거라 기대된다. 둘째는 실제 로봇에서 제어 확인이다. Unitree사의 A1, Alien-go 로봇을 이용하여 사람의 움직임으로 실제 로봇을 제어함이 추후 연구 목표다.

#### 참고문헌

[1] Y.Gong, R.Hartley, X.Da, A.Hereid, O.Harib, J.-K.Huang, and J.Grizzle, Feedback control of a cassie bipedal robot: Walking, standing, and riding a segway, American Control Conference (ACC), pp. 4559–4566, 2019.  
 [2] X.B.Peng, P.Abbeel, S.Levine, and M.vandePanne, Deepmimic: Example-guided deep reinforcement learning of physics-based character skills, ACMTrans. Graph., vol. 37, pp. 143:1–143:14, July 2018.

# 병적 보행 환자의 모션 캡처 데이터로부터 근골격 모델 추정

박범수<sup>0</sup>, 민세희, 박정남, 이제희  
 서울대학교  
 {zigui, sehee, jungnam04, jehee}@mrl.snu.ac.kr

## Estimating A Patient's Musculoskeletal Model Only Using Mocap Data

Bumsoo Park<sup>0</sup>, Sehee Min, Jungnam Park, Jehee Lee  
 Seoul National University

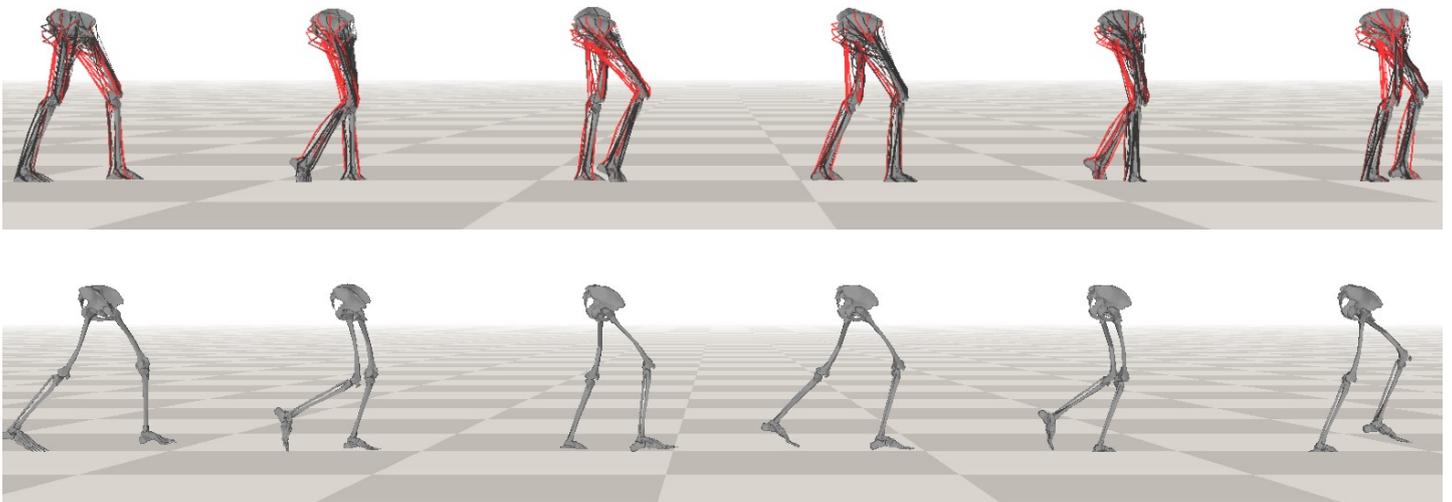


그림 1: 환자의 데이터(아래)와 이로부터 추정된 근골격 모델의 보행 시뮬레이션 결과(위)

### 요약

본 논문에서는 병적 보행 환자의 모션 캡처 데이터만으로 가상 환경에서 이와 유사하게 걷는 근골격 모델[1]을 추정할 수 있는 시스템을 구축하는 방법을 제안한다. 모션 캡처란 환자의 몸에 마커를 부착한 뒤 시간에 따른 움직임을 기록하는 것이다. 모션 캡처 데이터는 모션 캡처에서 기록한 마커의 3차원 좌표값들로 구성된다. 이것을 가능하게 만드는 구성 요소는 두 가지가 있다. 첫 번째는 근육 상태와 골격 구조를 매개 변수로 가지는 근골격 모델의 보행 시뮬레이터를 학습하여 사용한 것이다. 이를 이용하면 다양한 근골격 모델의 보행을 추가적인 학습 없이 바로 계산할 수 있다. 두 번째는 근골격 모델의 매개 변수를 입력으로, 시뮬레이션 결과와 보행 데이터의 차이값을 출력으로 하는 함수의 최솟값을 구하는 최적화 문제를 해결하는 방법으로 적합한 근골격 모델을 추정한다. 본 논문은 다양한 환자에 대해 이와 유사한 시뮬레이션 결과를 보이는 근골격 모델을 찾을 수 있음을 보였다.

### 1. 서론

병적 보행 환자는 어떤 질환에 의해 정상 보행을 하지 못하는 환자이다. 그 중 일부 근육이 수축하거나 약화된 질환을 가진 환자의 병적 보행은 몸을 구성하는 근육의 종류만큼 그 수가 많다. 하지만 환자의 근육은 밖으로 드러나있지 않아 직접 근육의 상태를 알아내는 것은 거의 불가능하며, 보행이나 움직임을 분석하여 그 상태를 간접적으로 확인하는 방법이 주로 사용된다. 병적 보행 환자를 가상 환경으로 옮기는 것은 드러나지 않은 근육을 직접 조작할 수 있다는 점에서 큰 장점을 가진다. 환자에 맞는 근골격 모델을 만드는 것은 환자를 가상 환경에서 시뮬레이션 할 수 있으며, 현실에서는 불가능한 작업을 시도할 수 있는 등 가상 환경의 이점을 모두 가져갈 수 있다는 장점이 있다.

본 논문에서는 병적 보행 환자의 모션 캡처 데이터만으로 가상 환경에서 이와 유사하게 걷는 근골격 모델을 추정할 수 있는 시스템을 구축하는 방법을 제안한다. 시스템을 구축하기 위해서는 먼저 매개 변수화된 근골격 모델의 보행을 학습한 시뮬레이터가 필요하다. 구축된 시스템의 실행 과정은 두 단계로 구성된다. 첫 번째는 모션 캡처 데이터에서 골격 모델과 환자의 보행을 복원하는 것이다. 두 번째는 앞 단계에서 복원된 환자의 보행에서 근골격 모델을 구성하는 근육의 상태를 추정하

\* 구두발표논문

\* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00878, SW컴퓨팅산업원천기술개발사업(SW스타랩))

는 것이다. 시스템의 모든 복원 과정이 끝나면 골격 모델과 추정된 근육의 상태가 결과물로 나오며, 두 결과물을 근골격 모델에 적용하는 것으로 환자의 보행과 유사하게 걷는 근골격 모델을 얻을 수 있다.

## 2. 문제 해결 방법

### 2.1. 보행 시뮬레이터 구축

근골격 모델을 매개 변수화 시키는 것으로 다양한 근골격 모델을 생성할 수 있다. 뼈의 크기와 길이를 매개 변수화 시킨 근골격 모델은 다양한 체형을 가진 사람을 모델링할 수 있고, 근육의 힘과 수축된 정도를 매개 변수화 시킨 근골격 모델은 다양한 질환을 가진 환자를 모델링할 수 있다. 본 논문에서는 뼈의 크기와 길이를 매개 변수로 가지는 골격 모델과, 근육의 수축 정도를 매개 변수로 가지는 근골격 모델을 사용한다.

시스템은 매개 변수화된 근골격 모델의 보행을 학습한 보행 시뮬레이터를 필요로 한다. 구체적으로, 보행 시뮬레이터의 목표는 임의의 근육 매개 변수로부터 생성된 근골격 모델을 물리 시뮬레이션 상에서 걸을 수 있도록 근육의 활성화 정도를 결정하는 것이다. 강화 학습은 매개 변수화된 근골격 모델이 특정 움직임을 따라하는 시뮬레이터를 만드는 것을 가능하게 해 준다.

### 2.2. 골격 및 움직임 복원

시스템의 첫 번째 단계는 골격에 부착된 마커의 위치와 모션캡처 데이터의 차이를 최소화 하는 골격 구조 및 움직임을 복원하는 것이다. 이를 최적화 문제로 변환하기 위해 3개의 변수 그룹인 골격 모델의 매개 변수, 자세, 마커의 부착 위치를 입력으로, 복원된 마커의 위치와 모션 캡처 데이터와의 L2-norm의 제곱합을 출력으로 하는 함수를 만든다.

이 문제는 Forward Kinematics가 식에 포함된 비선형 최적화 문제이며, 동시에 프레임의 개수에 골격 구조의 DOF(degree of freedom)이 곱해진 값 이상의 변수가 존재한다. 때문에 최적화 문제를 단 번에 해결하기에는 어려움이 있다. 우리는 위에서 언급한 변수 그룹 3개 중 하나만을 변수로 한 최적화 문제 3개를 해결한 뒤, 각각을 번갈아 적용하여 문제를 해결하였다. 이 방법은 자세 변수를 해결할 때 프레임에 독립적으로 해결할 수 있게 해 주어 변수 개수를 DOF만큼 줄일 수 있어 문제를 효율적으로 해결할 수 있다.

### 2.3. 근육 상태 추정

시스템의 두 번째 단계는 앞 단계에서 복원한 보행 데이터로부터 근골격 모델을 추정하는 것이다. 구체적인 목표는 환자의 보행과 가장 유사한 시뮬레이션 결과를 보이는 매개 변수를 찾는 것이다. 이를 해결하기 위해 먼저 매개 변수를 입력으로, 매개 변수로부터 만들어진 근골격 모델의 시뮬레이션 결과와 환자의 보행 데이터의 차이를 출력으로 하는 함수를 정의한다. 이 함수의 최솟값을 찾는 것으로 우리가 원하는 근골격 모델을 만들 수 있다.

시스템은 매개 변수화된 근골격 모델에 대한 보행 시뮬레이터를 보유하고 있기 때문에, 매개 변수로부터 만들어진 근골격 모델의 보행 시뮬레이션 결과를 추가적인 학습 없이 얻을 수 있다. 이는 매개 변수가 주어졌을 때 보행 시뮬레이션 결과와 복원된 보행 데이터의 차이값을 반환하는 함수를 빠르게 계산할 수 있다는 것을 의미한다.

이 함수는 강화 학습의 결과물을 시뮬레이션에 적용한 값을 사용하기 때문에 미분이 불가능하며, 그 결과가 일정하지 않고 잡음이 심하기 때문에 함수값을 사용하여 1차 미분값을 근사할 수 없다. 우리는 derivative-free optimization algorithm 중 하나인 Covariant Matrix Adaptation Evolutionary Steps(CMA-ES)를 사용하여 이 함수를 최적화하였다.

## 3. 결과 및 결론

우리는 먼저 모션 캡처 데이터에서 골격과 움직임을 복원할 수 있음을 보였다. 또한, 다양한 환자의 데이터에 대해서도 근골격 모델을 추정할 수 있음을 보였다.

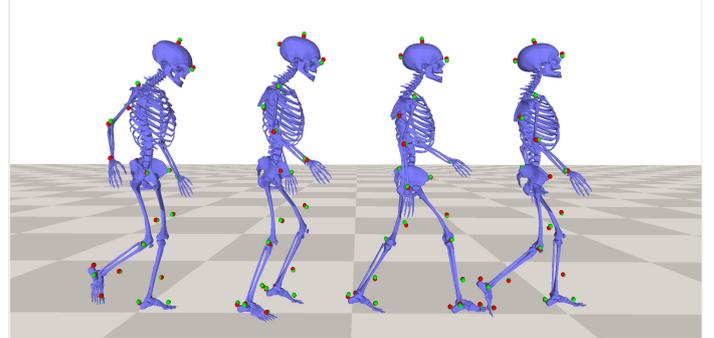


그림 2: 모션 캡처 데이터(초록색 점)과, 이로부터 복원된 골격 및 자세, 복원된 마커 위치(빨간색 점)의 모습.

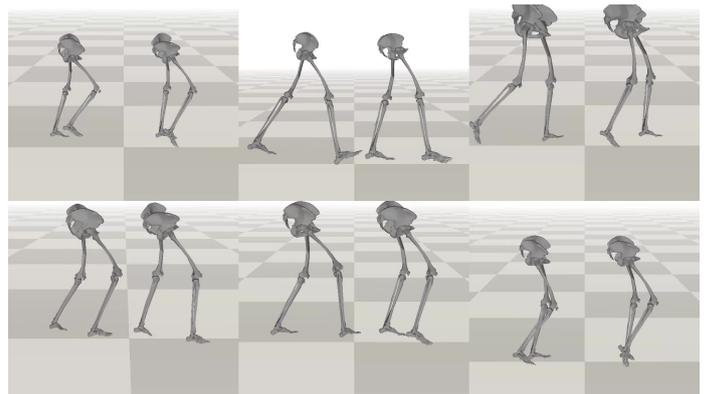


그림 3: 6개의 병적 보행 데이터(각 칸의 왼쪽)과 시스템이 추정한 근골격 모델의 시뮬레이션 결과(각 칸의 오른쪽).

## 참고문헌

- [1] Seunghwan Lee, Kyoungmin Lee, Moonseok Park, and Jehee Lee, Scalable Muscle-actuated Human Simulation and Control, *ACM Transactions on Graphics*, 37(4), 73, 2019.
- [2] Won, Jungdam and Lee, Jehee, Learning Body Shape Variation in Physics-based Characters, *Transactions on Graphics*, 38(6), 207, 2019.

# 마커를 이용한 관절 연체 모델 변형의 IK 계산 가속화\*

박현아<sup>0</sup>, 강다은, 권태수  
 한양대학교 일반대학원 컴퓨터소프트웨어학과  
 {qkrgusdk0303, teddysiah, taesoo}@hanyang.ac.kr

## An Accelerated IK solver for Articulated Soft Body Deformation using Markers

Hyunah Park<sup>0</sup>, Daeun Kang, Taesoo Kwon  
 Department of Computer Science, Hanyang University Graduate School, South Korea

### 요약

본 논문은 골격이 있으나 몸체의 변형이 비교적 자유로운 캐릭터 모델의 변형을 빠른 계산 속도로 구현하기 위해 가속화된 역기구학 풀이 방법을 제안한다. 기존의 풀이 방법은 ① 각 마커에 대하여, ② 해당 마커의 위치에 영향을 미치는 각 뼈대에 대하여, ③ 해당 뼈대에 영향을 미치는 상위(부모) 관절에 대하여 계산을 수행하는 삼중 중첩 반복 계산 구조를 갖는다. 그러나 해당 계산 과정에 중복되는 과정이 많은 것을 고려하여 사전에 계산된 값을 이용하는 방식을 통해 변형을 수행하는 데 필요한 계산 시간을 감소시켰다. 이러한 방식은 선형 블랜드 스키닝 기법으로 구성된 캐릭터 모델을 다루거나, 마커 없이 단순 촬영만으로 수행하는 모션 캡처 관련 연구에도 유용하게 활용할 수 있다.

### 1. 서론

컴퓨터 애니메이션에서는 관절 캐릭터 모델(articulated characters)뿐만 아니라 연체 성질을 가진 캐릭터 모델(flexible characters) 또한 빈번히 사용된다. 특히 연체 캐릭터의 경우 동작과 환경에 따라 그에 따른 몸체의 모양에 크고 작은 변화를 수반하는데, 몸체의 변형(deformation) 시 영향을 받는 모든 메쉬(mesh)를 재구성해야 하므로 계산 시간의 부담이 크다.

본 연구는 골격이 있으나 몸체의 변형이 비교적 자유로운 캐릭터 모델(articulated and soft-bodied models)의 변형을 빠른 계산 속도로 구현하는 것을 목표로 한다. 이를 위해 삼각형 메쉬(triangular meshes)로 이루어진 모델을 빠른 속도로 변형할 수 있는 역기구학(inverse kinematics, IK) 풀이 방법을 연구하고 해당 인터페이스를 개발하였다. 사용자는 인터페이스를 통해 모델 표면상의 임의의 정점(vertex)을 마커(marker)로서 지정하고, 해당 마커의 위치를 임의로 조정할 수 있다. 시스템은 마커의 목표 위치를 기준으로 가속화된 IK 풀이를 통해 삼각형 메쉬를 재구성하여 모델의 변형된 모습을 신속하게 구현한다.

### 2. 가속화된 모델 변형 IK 풀이 시스템

#### 2.1. 마커의 선택

사용자는 삼각형 메쉬로 이루어진 모델에서 임의의 위치를 “마커”로서 선택할 수 있다. 모델의 현재 관찰 시점(viewport)에서 드러나는 표면 부위에 존재하는 임의의 세 정점을 연결하여 삼각형을 생성하고, 해당 삼각형의 임의의 가중치를 갖는 중심점(Barycentric point with random weights)을 마커로 선택할 수 있도록 하였다. 관찰 시점에 따라 모델의 가시적이 표면 부위가 달라지므로, 마커로 지정 가능한 정점들의 집합 또한 달라진다. 사용자는 해당 집합 중 원하는 정점을 마커로 설정할 수 있고, 각 마커는 드래그를 통해 목표 위치를 지정할 수 있다. 이러한 마커의 설정은 모델 변형 시 각 메쉬의 정점이 이동할 위치를 결정하는 제약 조건의 역할을 한다.

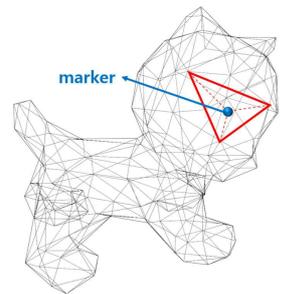


그림 1 모델 표면상의 임의의 삼각형에 대하여 임의의 가중치를 갖는 중심점을 마커로 설정

#### 2.2. 모델 변형을 위한 IK 풀이

사용자가 지정한  $k$ 개의 마커에 대하여, 각 마커의 목표 위치를 최대한 준수하기 위한 목적 함수(objective functions)는 마커의 현재 위치와 목표 위치의 거리 차를 합산하여 정의한다. 이때 각 마커의 에너지와 기울기(gradient)는 다음과 같이 구할 수 있다[1].

$$E(x_k) = \|x_k - x'_k\|_2^2$$

$$\frac{dE(x_k)}{dq} = 2(x_k - x'_k)^\top \frac{dx_k}{dq}$$

$x_k$ 와  $x'_k$ 는 각각  $k$ 번째 마커의 현재 위치와 목표 위치를 가리키며,  $\frac{dx_k}{dq}$ 는 뼈대(bone)의 모습(pose)  $q$ 에 대해,  $k$ 번째 마커에 대한 (관절의 위치)  $\times$  (관절의 목표 위치 마커의 목표 위치)<sup>T</sup> 형태의 야코비 행렬(Jacobian matrix)을 가리킨다.

\* 구두(포스터) 발표논문  
 \* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출 중.  
 \* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행되었음 (NRF-2020R1A2C1012847).

해당 IK 풀이는 ① 각 마커에 대하여, ② 해당 마커의 위치에 영향을 미치는 각 뼈대(bone)에 대하여, ③ 해당 뼈대에 영향을 미치는 상위 관절(parent joints), 즉 야코비 행렬의 열(column) 값에 대하여 계산을 수행하는데, 이른바 삼중으로 중첩된 반복 계산(3-nested loops)을 수행해야 한다.

### 2.3. 가속화된 IK 풀이

앞서 소개한 IK 풀이 방식은 삼중으로 중첩된 반복 계산으로 여전히 계산 시간의 부담이 크다. 그러나 수식 중 “해당 뼈대에 영향을 미치는 상위(부모) 관절”에 관한 계산과정은 주요 관절의 전역 좌표를 구하는 과정을 포함하고 있어 빈번히 중복된다. 따라서 해당 과정을 사전에 계산하고 기울기 값을 구하는 데 사용한다. 전처리된 수치만으로 기울기를 계산하므로 각 마커에 대한 반복 계산(loop)을 요구하지 않는다. 결과적으로 기존의 삼중 중첩 반복 계산을 이중 중첩 풀로 감축하여 변형 시 소요되는 계산 시간을 크게 감소시킬 수 있다.

### 3. 실험 결과 및 결론

본 연구의 작동 및 성능 확인을 위하여 i7-6700k CPU, 16G RAM, GeForce GTX TITAN X GPU를 장착한 컴퓨터를 사용하였고, 직육면체 스펀지, 이구아나, 고양이, 토끼, 인체 캐릭터 등 다섯 가지 삼각 메쉬 모델을 대상으로 실험을 수행하였다(표 1 참조). 표에서 “선택 가능한 마커의 수”는 실험 당시 변형을 시도하였던 임의의 모델 관찰 시점으로부터 관찰되는 선택 가능한 마커의 개수를 나타낸 것이다. 전처리과정은 실시간 성능으로 진행되었다.

그림 2는 다섯 가지 모델에 각각 1개, 5개, 10개, 50개, 120개의 마커를 설정한 후 변형을 시도한 결과이다. 마커를 1개만 설정한 경우, 해당 마커 이외의 다른 정점에 가해지는 제약조건이 없으므로 변형 시도 시 몸체의 변화가 거의 없고 단순 이동(translation)에 가까운 변형을 보여준다. 정점 선택에 따라 메쉬 변형 결과가 영향 받지 않았다. 마커가 모델에 고르게 분포할수록 결과가 좋아졌다. 마커의 개수가 많아질수록 더욱 유연하게 변형되는 양상을 보이며, 마커의 개수가 100개 이상 많아지면 모델 변형의 제약조건이 과도해져 변화가 거의 일어나지 않는다. 또한 모델을 구성하는 메쉬와 뼈대, 변형을 정의하는 마커의 수가 많아지더라도 변형을 수행하는 데 걸리는 시간이 기존의 삼중 중첩 반복 계산 방식보다 크게 단축됨을 확인할 수 있었다.

본 논문은 골격이 있으나 몸체의 변형이 비교적 자유로운 캐릭터 모델의 변형을 빠른 계산 속도로 구현하기 위해 가속화된 역기구학 풀이 방법을 제안하였다. 이러한 방식은 선형 블렌드 스킨닝(linear blending skinning) 기법으로 구성된 캐릭터 모델을 다루는 데 유리하며, 마커 없이 단순 촬영만으로 수행하는 모션 캡처(markless motion capture) 관련 연구에도 유용하게 활용될 수 있다.

표 1 실험에 사용한 다섯 가지 모델의 정보

모델	뼈대의 수	정점의 수	표면 메쉬의 수	선택 가능한 마커의 수
스펀지	6	654	1040	520
이구아나	18	124	244	136
고양이	16	159	314	158
토끼	8	184	364	163
인체	26	13806	12764	7526

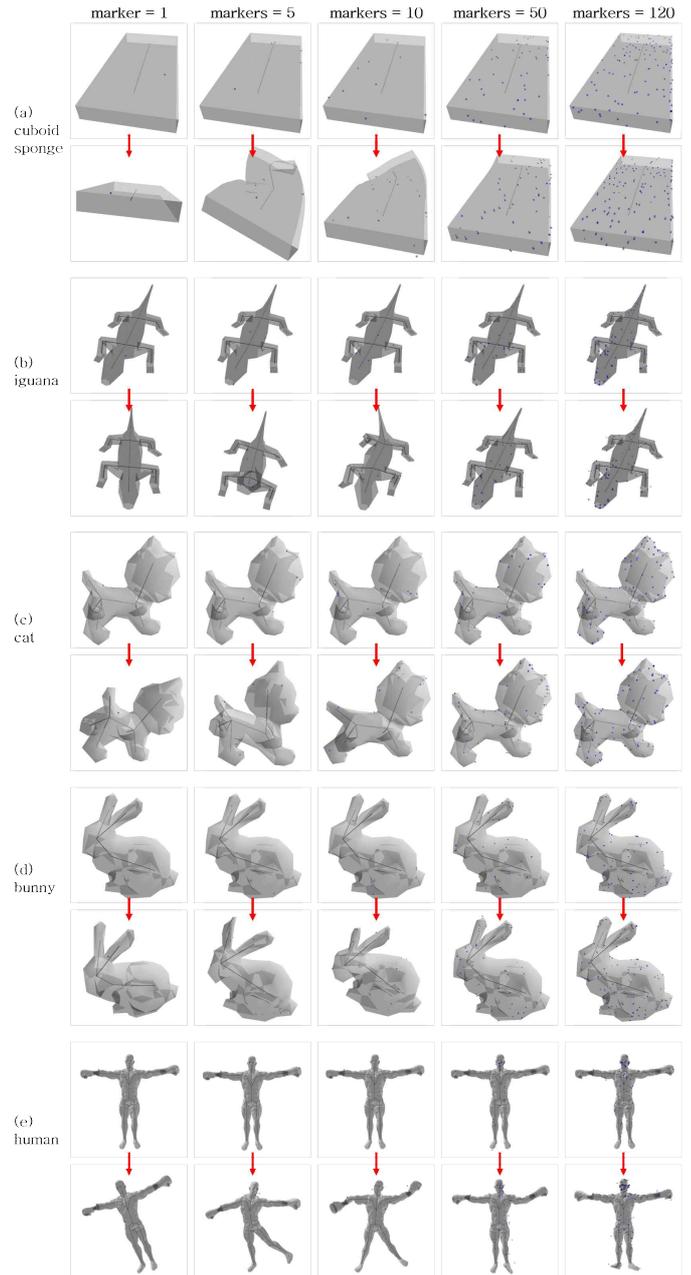


그림 2 가속화된 모델 변형 IK 풀이 시스템 실험 결과

### 참고문헌

[1] J. Kim, Y. Seol, H. Kim, & T. Kwon, T, Interactive character posing with efficient collision handling. *Computer Animation and Virtual Worlds*, 31(3), e1923, 2020.

# 가상 실내 환경을 위한 변화맹 기반 무한공간탐험\*

황준영<sup>0</sup>, 조용훈, 권순욱, 이인권  
연세대학교 컴퓨터과학과

tlavotl1@yonsei.ac.kr, maxburst@gmail.com, kwonars@yonsei.ac.kr, iklee@yonsei.ac.kr

## Infinite Space Exploration Based on Change Blindness for Virtual Indoor Environment

June-Young Hwang<sup>0</sup>, Yong-Hun Cho, Soon-Uk Kwon, In-Kwon Lee  
Dept. of Computer Science, Yonsei University

### 요약

가상 현실(Virtual Reality) 기술이 발전함에 따라 실제 걸음을 통한 가상 공간 탐험 기법은 꾸준히 연구되어 왔다. 무한공간탐험법(Redirected Walking: RDW)은 가상환경 내에서 사용자의 움직임을 왜곡하거나 혹은 가상환경의 구조 변형을 통해 강제로 실제환경과 가상환경의 동작이 달라지도록 하는 기법[1]이다. 본 연구에서는 변화 맹(Change blindness)를 이용한 공간 변형을 통해서 좁은 실제 공간에서도 무한히 많은 가상의 방들을 탐험하는 새로운 무한공간탐험법을 제안하고자 한다.

### 1. 서론

가상 현실(Virtual Reality)에서 사용자가 한 위치에서 다른 위치로 이동하는 행위는 가장 중요하면서 보편적인 행동 중 하나라고 할 수 있다. 이러한 행위를 실현하는 방법은 여러 종류가 있는데, 그 중에서도 실제 걸음을 통해서 무한히 넓은 가상 공간을 이동하는 기법의 효율성이 높다고 많은 연구에서 발표되었다. 따라서 가상 공간에서 실제 걸음을 통해서 탐험할 수 있도록 무한공간탐험법[2]이 최초로 제안되었다. 해당 연구에서는 사용자의 움직임을 왜곡을 시켜서 사용자 경로를 제어하여 항상 실제 공간의 중심으로 향하게 한다. 그러나 이러한 방식의 문제점은 지속적인 움직임 왜곡으로 인해 사용자가 시각과 전정감각 사이의 불일치(Visual-Vestibular Conflict)를 겪어서 멀미와 자세 불안정 같은 현상을 겪을 수 있다는 것이다.[3] 이러한 문제점을 극복하기 위해 변화 맹 현상을 이용하여 사용자가 인지하지 못한 채 가상 공간의 구조 자체에 변형을 주는 새로운 무한공간탐험법을 제안하고자 한다.

### 2. 방법

#### 2.1 변화 맹(Change Blindness)

변화 맹(Change Blindness)이란, 시각 자극의 변화가 있음에도 이를 인지하지 못하는 현상을 뜻한다. [3]에서는 이 현상을 이용하여 실험자가 문을 열고 가상의 방에 들어가서 책상 위에 있는 컴퓨터를 조작하는 사이에 시야 밖에 있는 문의 위치를 회전시키더라도 실험자가 이를 인지하지 못한다는 결과를 보여줬다. 따라서 본 연구에서는 가상 공간이 무한히 많은 가상의 방들로 이루어져 있을 때 변화 맹을 이용해 가상의 벽들을 이동시키는 새로운 방향 전환 기법을 제안하고자 한다.

#### 2.2 알고리즘

실제 공간  $\mathcal{R}$  은  $O_R$  를 중심으로 하고  $(W_R, H_R)$  크기를 가지는 직사각형이다. 가상 공간  $\mathcal{V}$  은  $n$  개의 방  $v_1, v_2, \dots, v_n$  들로 이루어져 있고 각 방은  $O_1, O_2, \dots, O_n$  을 중심으로 하고  $(W_1, H_1), (W_2, H_2), \dots, (W_n, H_n)$  크기를 가지는 직사각형이다. 이 때, 각 방의 크기는 실제 공간 크기보다 작으며 적어도 한 개 이상의 방과 문으로 연결되어 있다. (그림 1 참조)

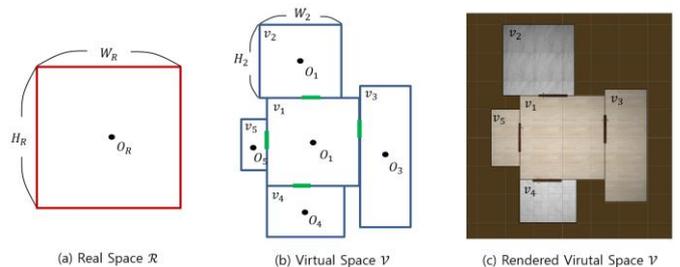


그림 1: (a)는 실제 공간  $\mathcal{R}$  을 나타냄. (b) 가상 공간  $\mathcal{V}$  을 나타냄. 이때 초록색 직사각형은 연결된 문을 뜻함. (c)는 (b)를 실제로 구현한 모습을 나타냄.

처음에는 가상의 방  $v_1, v_2, \dots, v_n$  중 하나가 실제 공간  $\mathcal{R}$  의 중심  $O_R$  과 일치하도록 가상 공간이 이동하게 되며, 사용자는 실제 공간과 중심이 일치한 가상의 방에서 시작하게 된다. 이때, 가상의 방  $v_c$  는 현재 사용자가

\* 구두(포스터) 발표논문

\* 본 논문은 요약논문 (Extended Abstract) 으로서, 초기 연구 진행 결과를 포함하고 있음.

\* 본 연구는 정부(과학기술정보통신부)의 재원으로, 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구 결과로 수행되었음 (IITP-2021-2018-0-01419)

있는 방을 의미한다. (그림 2 참조)

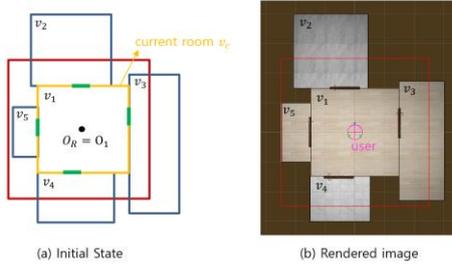


그림 2: (a)는 최초 시작 시, 실제 공간 중심에 가상 공간의 방  $v_1$ 이 위치한 모습. (b)는 이를 실제로 나타낸 것.

만약 사용자가 현재 있는 방과 연결된 다른 방을 방문하려고 한다면, 실제 공간과 충돌할 가능성이 생긴다. 이를 방지하기 위해 본 연구에서는 크게 두 단계를 적용한다. 첫번째 단계는 축소(Reduce) 단계로, 사용자가 현재 있는 방  $v_c$ 에 연결된 방들이 실제 공간 안에 있도록 축소시키는 단계이다. 이를 위해 우선  $v_c$ 와 연결된 각 방이 실제 공간  $\mathcal{R}$  안에 있는지를 확인한다. 만약 안에 있지 않다면, 해당하는 방의 벽면 중에서 실제 공간  $\mathcal{R}$  밖에 있는 벽면들을 실제 공간 안쪽으로 축소를 시켜서 연결된 방들을 실제 공간 안쪽으로 축소시킨다. (그림 3 참조)

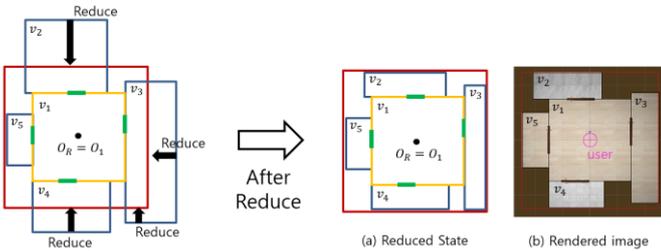


그림 3: 실제 공간  $\mathcal{R}$  안에 들어가지 않는 연결된 방들을 찾아낸 뒤에 이들을 축소시킴. (a)는 축소 단계가 완료된 모습. (b)는 이를 실제로 구현한 모습을 나타냄.

두번째 단계는 복원(Restore) 단계로, 사용자가 연결된 다른 방을 방문하여 현재 사용자가 있는 방  $v_c$ 가 변했을 때 적용된다. 이때는  $v_c$ 를 원래 크기로 확장하면서 동시에 실제 공간 중심에 위치하도록 방의 벽면들을 지속적으로 이동시키는 단계이다. 우선 각 벽면이 이동해야 되는 위치는 현재 방의 위치  $O'_c$ 에서 실제 공간 중심  $O_R$ 로 향하는 벡터와 방의 중심을 기준으로 축소된 크기  $(W'_c, H'_c)$ 를 원래 크기  $(W_c, H_c)$ 로 확대하는 스케일 벡터의 합으로 구해낸다. (그림 4 참조)

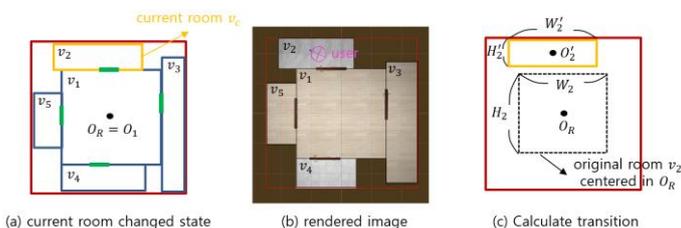


그림 4: (a)는 축소 연산 이후 현재 사용자가 있는 방이 변경된 상황, (b)는 이를 실제로 구현한 모습을 나타냄. (c)는 복원을 위해 각 벽면이 이동해야 되는 위치를 계산한 것.

각 벽면을 이동시킬 때 이동해야 되는 위치로 한 번에 움직이게 되면 사용자가 벽의 이동을 인지할 가능성이 크다. 한편으로, 사용자 시야각 안에 있는 벽면이 움직이게 되면 이 또한 벽의 이동을 인지할 가능성이 크다. 따라서 본 연구에서는 사용자 시야 밖에 있는 벽면들을 이동해야 되는 위치로 여러 번 나누어서 짧게 이동하는 방식을 사용한다. 이로 인해 벽이 이동한 뒤에 사용자가 해당 벽면을 바라봐도 변화 맹 현상으로 인해 직전에 봤던 모습과 큰 차이를 느끼지 못해 벽의 이동을 인지하지 못하게 된다. 이 방식을 적용하여 사용자가 현재 방을 탐험하는 동안 모든 벽면이 주어진 위치로 도달하게 되면 복원 단계가 완료된다. (그림 5 참조) 복원 단계가 완료되면 첫번째 단계인 축소 단계로 돌아가게 되고 앞선 과정을 반복하게 되면서 사용자는 좁은 실제 공간만을 사용하여 무한히 많은 가상의 방들이 연결된 가상 공간을 돌아다닐 수 있게 된다.

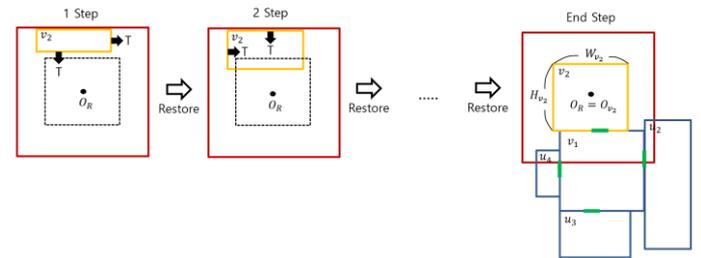


그림 5: 사용자가 현재 방을 탐험하는 동안 현재 사용자가 있는 방이 실제 공간 중심에 있고 원래 크기를 가지도록 사용자 시야 밖 벽면이 이동하여 점차 복원되는 모습.

### 2.3 벽면 이동에 대한 인지 한계(Threshold) 실험

무한공간탐험법에서 움직임 왜곡에 대한 인지 한계 실험이 진행된 것처럼[4], 사용자 시야 밖에 있는 벽면 이동에 대한 인지 한계를 확인해보기 위해 실험을 진행했다. (부록 그림 1 참조) 총 15명의 실험자를 모집하였고 실험자가 하나의 방에서 주어진 지시대로 행동을 하는 도중 실험자 시야 뒤에 있는 벽면을 이동시킨 뒤에 이를 인지했는지를 확인하였다. (부록 그림 2 참조) 그 결과, 실험자와 이동하는 벽면 사이의 거리에 따라 벽면이 이동할 수 있는 인지 한계를 도출해낼 수 있었다. (부록 그림 3 참조)

### 3. 결론

본 연구에서는 변화 맹을 통한 가상의 방 변형 기법을 제시하여 좁은 실제 공간에서 무한히 연결된 가상의 방들을 걸을 수 있는 새로운 무한공간탐험법을 제안하였다. 해당 기법을 통해서 모델 하우스, 박물관 등과 같이 여러 개의 방들로 이루어져 있는 구조를 가상 공간에서 실제로 걸어서 탐험할 수 있도록 만들 수 있을 것이다.

## 참고문헌

- [1] Nilsson, Niels Christian, et al., 15 years of research on redirected walking in immersive virtual environments, *IEEE computer graphics and applications*, 38(2):44-56, 2018.
- [2] Razzaque, Sharif, Zachariah Kohn, and Mary C. Whitton. Redirected walking. *Chapel Hill: University of North Carolina at Chapel Hill*, 2005.
- [3] Suma, Evan A., et al, Leveraging change blindness for redirection in virtual environments, *2011 IEEE Virtual Reality Conference*. IEEE, 2011.

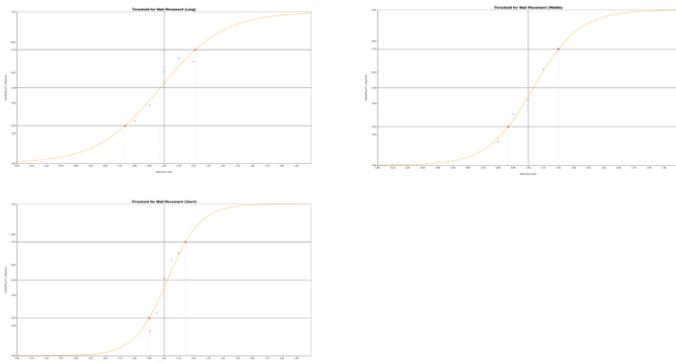
## 부록



**그림 1:** 인지 한계 측정 실험을 진행한 방의 모습. (a)는 위에서 바라본 모습이고 (b)는 방 안을 바라본 모습을 나타냄



**그림 2:** 실험 도중 파란색 선에 위치한 벽이 움직이는 예시. (a)는 벽이 움직이기 전의 모습이고 (b)는 벽이 사용자를 향해 움직인 후의 모습을 나타냄.



**그림 3:** 실험자들의 응답을 통해 추론한 심리 측정 곡선 (psychometric curve). 왼쪽 위부터 순서대로 벽과 거리가 먼 경우(Long), 중간인 경우(Middle), 가까운 경우(Short)를 나타냄.

# 고정된 위치에서의 Reset과 블록 이어 붙이기를 이용한 무한 가상 공간 탐험\*

권순욱<sup>0</sup>, 황준영, 조용훈, 이인권  
연세대학교 컴퓨터과학과

kwonars@yonsei.ac.kr, tlavot11@yonsei.ac.kr, maxburst@gmail.com, iklee@yonsei.ac.kr

## Infinite Virtual Space Exploration Using Block Concatenation and Reset at Fixed Locations

Soon-Uk Kwon<sup>0</sup>, June-Young Hwang, Yong-Hun Cho, In-Kwon Lee  
Dept. of Computer Science, Yonsei University

### 요약

가상공간에서 사용자가 이동할 때 실제공간에서도 동시에 걷는 방법은 사용자들에게 높은 현존감을 제공할 수 있으나, 실제 공간의 유한한 물리적 제약으로 인해 가상 공간을 무한히 걷는 것에는 제약이 있다. 이를 극복하기 위해 RDW(Redirected Walking)와 같은 방법들이 연구되어 왔다. RDW는 사용자들에게 몰입감을 제공할 수 있으나 실제공간의 유한성 때문에 의도하지 않게 명시적으로 사용자를 Reset 시켜야 하며, 이 Reset point는 사용자가 기대하지 않는 임의의 시점과 지점에서 자주 일어날 수 있다는 것이 단점이라 할 수 있다. 이에 본 논문에서는 사용자가 원하는 시점에 Reset을 할 수 있도록 하는 "고정위치 Reset" (RFL: Reset at Fixed Locations) 방법을 제안한다. RFL은 2:1 Turn을 수행할 Reset의 위치들을 고정함으로써 가상의 단위 블록을 만들고, 이를 연결하는 방식으로 가상공간을 생성한다. 사용자는 RFL을 통해 만들어진 가상공간에서 스스로 Reset을 수행하지 않는 이상 Reset 없이 무한한 가상 공간을 탐험하게 된다. 또한 시뮬레이션 실험을 통해 RFL 방법이 기존의 S2C나 APF방법과 같은 RDW 기법들보다 적은 수의 Reset을 발생시키는 것을 확인하였다.

### 1. 서론

가상 현실 기술이 대두됨에 따라 넓은 가상 현실을 좁은 실제 공간을 활용하여 보행할 수 있도록 하는 여러

방향전환보행(Redirected Walking) 기술이 연구되어 왔다. 하지만 기존의 방향전환보행 기술은 물리 공간의 경계에서 Reset을 갑작스럽게 발생시키기 때문에 보행자의 충돌 위험을 사전에 알리지 못하게 된다. 또한 충돌 직전 Reset을 지시하는 것과 같이 갑작스레 사용자에게 행동을 지시하는 명시적 기술(Overt Technique)은 가상 현실 시스템의 현존감을 저하시킬 수 있다.[1] 이에 우리는 Reset의 위치를 고정하여 가상공간을 생성하는 방식인 RFL을 제안한다. 이를 통해 사용자는 가상 현실 시스템으로부터 Reset 지시를 받지 않더라도 가상 공간 상에서 스스로 충돌을 예측할 수 있게 되며, 자신이 원하는 때에 Reset을 수행하여 충돌을 피해 자유롭게 무한한 가상공간을 돌아다닐 수 있게 된다.

### 2. 방법

RFL 방법은 아래의 단계로 나뉘게 된다.

- 단계 1: 실제 공간 내의 적절한 Reset 위치 선정
- 단계 2: Reset 위치에 따라 가상공간의 단위 블록 생성
- 단계 3: 단위 블록들을 이어 붙여 가상공간 생성

#### 2.1. 실제 공간 내의 적절한 Reset 위치 선정

가상공간의 형태를 결정하기 위해 실제 공간 내에서 적절한 Reset의 위치를 선정한다. 이 Reset 위치들은 실제공간에서의 충돌을 피하도록 가상공간의 벽과 장애물의 형태 및 위치를 결정하며, 사용자는 원하는 때에 이곳에서만 Reset이 가능하다. RFL의 Reset 방식은 2:1 Turn을 활용하며, 2:1 Turn 시 위상이 180도 바뀌는 성질을 통해 하나의 단위 블록을 결정한다. 블록을 2차원으로 이어 붙이기 위해 실제 공간 내 총 4개의 고정된 Reset 위치가 필요하다. 실제 공간을 최대한 활용하기 위해 Reset 위치를 경계선에 위치시킬 수 있다.

\* 구두발표논문

\* 본 논문은 요약논문 (Extended Abstract) 으로서, 초기 연구 진행 결과를 포함하고 있음.

\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구 결과로 수행되었음 (IITP-2021-2018-0-01419\*)

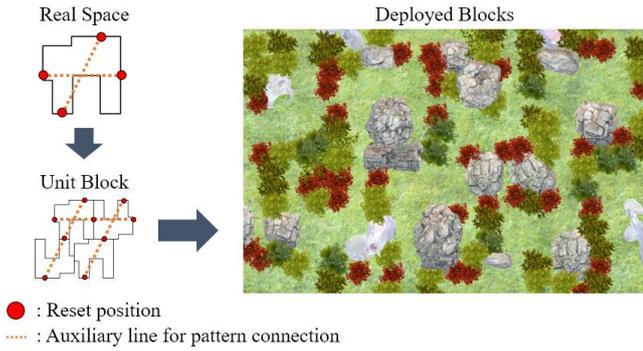


그림 1: RFL의 과정

## 2.2. Reset 위치에 따라 가상공간의 단위 블록 생성

단계 1에서 설정된 Reset의 위치를 축으로 하여 180도 회전시켜 단위 블록을 생성한다. 여기서 단위 블록 상단의 두 개는 4개의 Reset 위치를 가지며, 하단의 두 개는 2개만을 가진다. 그림 1의 좌측 하단은 반복을 위한 하나의 단위 블록의 예시이다.

## 2.3. 단위 블록들을 이어 붙여 가상공간 생성

단계 2에서 생성된 단위 블록을 반복하여 원하는 크기만큼의 가상공간을 생성한다. 그림 1의 오른쪽에 나타난 예시는 단위 블록을 이어 붙여 가상물체를 배치한 그림이다. 이를 통해 만들어진 가상공간에서 사용자는 충돌을 피해 가상공간을 무한히 탐험할 수 있게 된다.

## 3. 실험 및 결과

### 3.1 시뮬레이션 실험

시뮬레이션 실험에서 실제 공간은 그림 2의 좌측과 같이 Convex 형태의 사각형, Concave 형태의 T자형 및 임의의 방 형태의 세 공간으로 구성하여 RFL에 의해 가상 물체가 배치된 상태에서 Reset 발생 횟수를 비교하였다. 모든 실제 공간의 크기는 16m<sup>2</sup>보다 작다. 가상 공간의 경로 생성 방식은 1.5m 근방에 위치한 임의의 Waypoint를 Uniform 랜덤 방식으로 생성하여 총 500개의 Waypoint를 달성하도록 하나의 경로를 구성

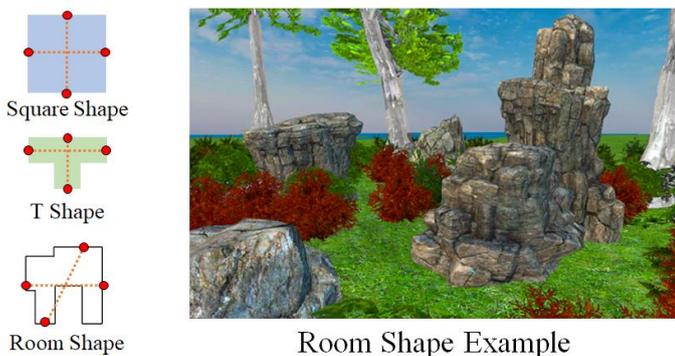


그림 2: 시뮬레이션 실험에 활용된 실제 공간과 예시 화면

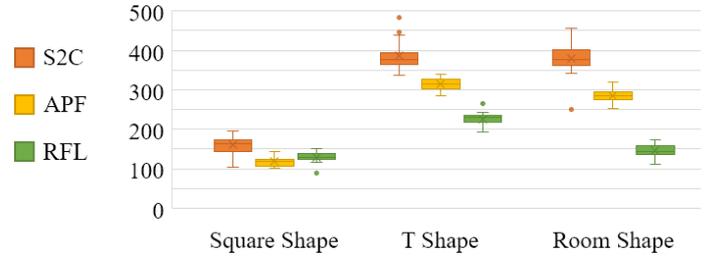


그림 3: 시뮬레이션 실험 결과

하였으며, 각 알고리즘마다 30개의 경로를 생성하여 기존의 S2C[2] 및 APF[3] 방식과 Reset 발생 횟수를 비교하였다. S2C의 Reset 방식은 2:1 Turn 방식으로 설정하였으며, APF의 Reset 방식은 APF-R 방식으로 진행하였다.

### 3.2 시뮬레이션 실험 결과

시뮬레이션 실험의 결과는 그림 3과 같다. 3가지 방법에 대해 One-way ANOVA test를 진행하였으며, Tukey's HSD test를 통해 사후 검정을 수행하였다. 유의수준 0.01에서 각 실제 공간마다 Reset 발생 횟수의 관계는 다음과 같다.

- 사각형: RFL ≈ APF < S2C
- T자형: RFL < APF < S2C
- 임의의 방 형태: RFL < APF < S2C

### 4. 토의 및 결론

본 연구에서는 가상현실의 사용자가 원하는 시점에 Reset을 할 수 있도록 하는 Reset 위치를 고정시키는 방법인 RFL을 제시하였다. RFL의 시뮬레이션 실험을 통해 임의의 형태의 실제 공간에서 타 방식들에 비해 Reset 발생 횟수가 유의미하게 낮은 결과를 얻었다. 본 방법은 갑작스럽게 Reset을 발생시키지 않기 때문에 사용자 실험에서 기존의 방법들보다 몰입감과 현존감이 뛰어날 것으로 기대된다. 추가적으로 본 연구는 가상공간이 실제공간의 반복형태이기 때문에 가상공간이 정형화되었다는 한계점이 있다. 추후 반복적인 형태를 활용하지 않은 방식에 대한 추가적인 연구가 필요하다.

### 참고문헌

[1] E.A. Suma et al., A taxonomy for deploying redirection techniques in immersive virtual environments, *Proceedings of IEEE Virtual Reality Short Papers and Posters*, pp. 43–46, 2012.  
 [2] S. Razzaque. et. al, Redirected Walking. University of North Carolina at Chapel Hill, 2005.  
 [3] J. Messinger, E. Hodgson, and E. R. Bachmann. Effects of tracking area shape and size on artificial potential field redirected walking. In *IEEE Conference on Virtual Reality and 3D User Interfaces*, 2019.

# 무한공간탐험을 위한 옆걸음 및 뒷걸음의 인지 한계점 측정\*

조용훈<sup>0</sup>, 이인권  
연세대학교 컴퓨터과학과  
cyhterry@naver.com, iklee@yonsei.ac.kr

## Estimation of Detection Thresholds for Non-Forward Steps in Redirected Walking

Yong-Hun Cho<sup>0</sup>, In-Kwon Lee  
Dept. of Computer Science, Yonsei University

### 요약

대부분의 가상현실(VR) 경험은 굉장히 제한된 좁은 현실 공간 안에서 이루어진다. 무한공간탐험은 좁은 현실 공간의 한계를 극복하기 위해 사용자의 시점을 조금씩 변경시킨다. 대부분의 무한공간탐험 알고리즘은 앞걸음에 초점이 맞추어져 있는데, 진정한 가상현실에서의 보행의 자유를 위해서는 옆걸음 및 뒷걸음의 활용이 필수적이다. 본 논문에서는 사용자가 왜곡을 인지할 수 없는 한계점인 인지 한계점(Detection Thresholds)을 옆걸음과 뒷걸음에 대해 측정한다.

### 1. 서론

가상현실에서는 현실에서 불가능한 다양한 경험이 가능하다. 하지만 가상현실을 체험할 때, 대부분 컨트롤러를 사용하거나, 텔레포트 방식을 이용한 이동방식을 사용하기 때문에, 실제 현실과 괴리감이 있으며, 몰입도가 떨어질 수 있다. 실제 보행을 통해 이런 한계를 극복할 수 있지만, 사용자가 사용가능한 공간은 제한되어 있어 보행하기에 적합하지 않다.

무한공간탐험(Redirected Walking)은 보행 시 사용자의 시점을 조금씩 왜곡하여 좁은 현실공간에서 넓은 가상현실을 탐험할 수 있게 해준다[1]. 대표적으로 이동 계인(Translation Gain)과 곡률 계인(Curvature Gain)이 존재하는데, 이동 계인은 사용자가 이동 시 가상에서의 사용자 속도를 빠르게나 느리게 조절하며, 곡률 계인은 사용자가 직선으로 이동시 아주 조금씩 사용자의 가상 시점을 회전시켜, 곡선으로 이동할 수 있게 해준다[2]. 이러한 계인을 크게 적용하면 적용할수록 공간의 효율성은 높아지지만, 사용자가 왜곡을 느끼기 시작하면 반대로 몰입도가 떨어지고, 사이버멀미가 발생할 수 있기 때문에, 여러 연구에서 사용자가 인지할 수 없는 최대 계인 범위인 인지 한계점을 측정하였다. 이동 계인의 경우 0.86에서 1.26까지 적용이 가능한데, 즉, 사용자가 눈치채지 못하게, 실제 1m 이동 시, 가상현실에서의 이동거리를 0.86m로 줄이거나 1.26m로 늘릴 수 있다. 곡

률 계인의 경우, 1m 이동 시 약 2.6° 시점을 회전 시킬 수 있다.

하지만 대부분의 무한공간탐험 기법은 앞걸음에만 초점이 맞추어져 있는데, 가상현실에서의 다양한 이벤트를 소화하기 위해서는 옆걸음 및 뒷걸음에 대한 고려가 필요하다. 가파른 난간을 걸을 때나 조용히 걸을 때 옆걸음이 자주 사용되고, 적에게서 물러서거나 할 때 뒷걸음을 사용할 수 있는데, 이러한 이벤트 시 계인이 적용될 수 있다면, 많은 가상현실 어플리케이션에 대해 효율적인 공간 사용이 가능해진다.

본 논문에서는 옆걸음과 뒷걸음에 대해 인지 한계점을 측정하기 위한 실험을 설계하고 시행하였으며, 그 결과 앞걸음과 유사한 이동 계인, 그리고 앞걸음보다 큰 곡률 계인을 적용할 수 있는 것을 확인하였다.

### 2. 인지 한계점 측정 실험 설계

인지 한계점 측정을 위해 실험에서 사용될 계인의 범위를 정해야 하는데, 프리-테스트 과정을 거쳐 이동 계인의 범위  $g_t$ 와 뒷걸음에서의 곡률 계인 범위  $r_b$ , 옆걸음에서의 곡률 계인 범위  $r_s$ 를 다음과 같이 정의하였다 :

$$g_t = \{0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4\}, \quad (1)$$

$$r_b = \{\pm 38.2m, 19.1m, 12.7m, 9.5m, 7.6m\}, \quad (2)$$

$$r_s = \{\pm 22.9m, 11.5m, 7.6m, 5.7m, 4.6m\}. \quad (3)$$

이동 계인의 범위는 뒷걸음 및 옆걸음에서 같은 범위를 사용하였으며, 이동 계인  $g_t$ 는 실제 이동거리 대비 가상 이동거리를 나타내고, 곡률 반지름  $r$ 은 휘어져 걷는 곡선의 반지름을 뜻하는데, 반지름이 낮을수록 더 크게 사용자를 휘게 할 수 있다.

실험은 총 6가지로, 뒷걸음과 왼쪽 옆걸음, 그리고 오른쪽 옆걸음에 대해 이동 계인과 곡률 계인을 각각 측정한다. 사용자는 뒷걸음 혹은 옆걸음으로 4m를 이동하게 되는데, 이 때 랜덤하게 계인 범위 내의 값 중 하나가 적용된다.

인지 한계점 측정 실험 방법은 통상적으로 사용되는 2중강제택일(2-Alternate-Forced-Choice) 방법을 사용하여 진행된다. 사용자는 두 가지 선택지 중 한 가지를 강제로 선택하여야 하는데, 이동 계인 실험에서는 실제 이동거리가 가상에서의 이동거리보다 큰지 작은지를 선택하게 되고, 곡률 계인 실험에서는 내 걸음이 왼쪽으로 휘어졌는지, 오른쪽으로 휘어졌는지를 선택한다.

총 18명의 사용자에게 실험을 진행하였으며 각 사용자

\* 구두발표논문

\* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문 IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR) 2021에 게재 되었음.

\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구 결과로 수행되었음 (IITP-2021-2018-0-01419\*).

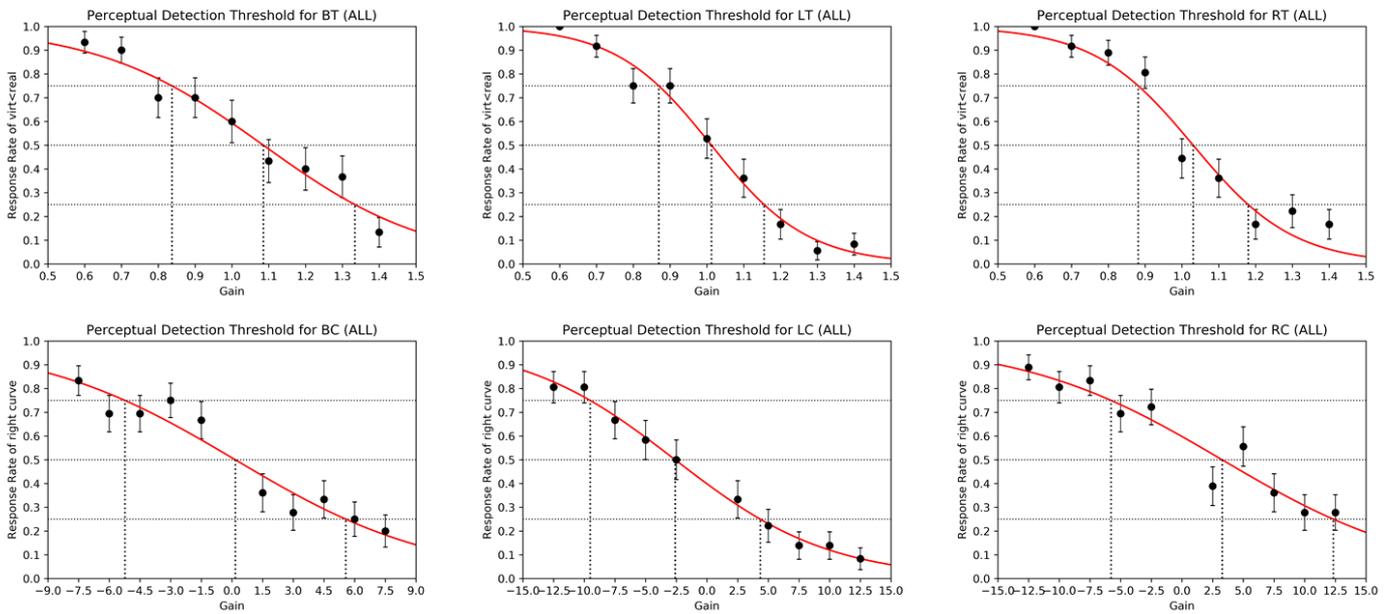


그림 1 인지 한계점 실험 결과. 윗줄부터 뒷걸음, 왼쪽 옆걸음, 오른쪽 옆걸음에 대한 이동 계인, 아랫줄부터 뒷걸음, 왼쪽 옆걸음, 오른쪽 옆걸음에 대한 곡률 계인 결과

는 뒷걸음, 옆걸음으로 4m 이동 후, 해당 이중강제택일 질문에 답변을 하게 되며, 이를 앞서 언급한 6가지의 실험에 대해 반복한다.

### 3. 결과

[그림 1]은 수집된 데이터를 시그모이드(Sigmoid) 함수에 피팅(fitting)한 결과이다. X축은 적용된 계인값, Y축은 이동거리가 실제보다 가상에서 더 크다고 높다고 대답한 확률이다. 곡률계인 실험에서는 오른쪽으로 경로가 휘었다고 대답한 확률을 의미한다. 해당 시그모이드 함수값이 0.25, 0.75인 지점의 X값을 인지 한계점으로 정한다. 즉, 사용자가 정답을 맞출 확률이 25%~75% 사이인 계인은 사용자가 인지하지 못한다고 상정한다.

#### 3.1. 이동 계인 인지 한계점

뒷걸음, 왼쪽 옆걸음, 오른쪽 옆걸음의 이동 계인 인지 한계점은 각각 [0.84, 1.33], [0.87, 1.16], [0.88, 1.18] 이다. 앞걸음의 인지 한계점이 [0.86, 1.26]으로 측정된 것과 비교하면, 전반적으로 비슷한 것을 볼 수 있다. 다만, 속도를 높이는 계인을 적용하는 경우, 뒷걸음의 인지 한계점이 옆걸음 보다 좀 더 크게 나온 것을 알 수 있는데, 이는 뒷걸음시 보는 방향과 반대방향으로 걷기 때문에, 거리가 멀어져 조금 더 둔감해지는 것으로 해석할 수 있다.

#### 3.2. 곡률 계인 인지 한계점

뒷걸음, 왼쪽 옆걸음, 오른쪽 옆걸음의 곡률 계인 인지 한계점은 각각 [-10.65m, 10.30m], [-6.02m, 13.19m], [-9.92m, 4.65m] 이다. 앞걸음의 인지 한계점인 22m와 비교했을 시, 뒷걸음과 앞걸음 모두 인지 한계점이 크게 증가하였음을 알 수 있다. 옆걸음의 경우, 왼쪽과 오른쪽 둘다 비대칭을 띄고 있는 것을 볼 수 있는데, 이를 해석해보면, 둘다 몸의 앞쪽으

로 휘는 것에 둔감하고, 몸의 뒤쪽으로 휘는 것에 더 민감하다 볼 수 있다. 이는 옆걸음 시, 고개를 90도로 돌려 옆을 바라보면서 가기 때문에 나타나는 현상으로 생각되는데, 목을 옆으로 돌려서 걷기 때문에, 몸위 뒤쪽보다 앞쪽으로 이동하는 것이 편하게 되고, 이에 따라 앞쪽이 보다 많은 왜곡을 줄 수 있게 된다.

또한, 뒷걸음 및 옆걸음의 경우 앞걸음보다 속도가 느린데, 속도가 느리면 인지 한계점이 증가한다는 기존 연구와 결을 같이 한다[3].

### 4. 결론

본 논문에서는 옆걸음과 뒷걸음에 대한 이동 계인, 곡률 계인에 대한 인지 한계점을 측정하였다. 이동 계인의 인지 한계점은 앞걸음과 유사하고, 곡률 계인의 인지 한계점은 앞걸음보다 훨씬 더 크게 측정되었다. 따라서, 옆걸음과 뒷걸음을 이용하여 보다 효율적인 무한공간탐험을 할 수 있으며, 4방향의 보행의 자유가 보장되기 때문에, 조금 더 다채로운 가상현실 어플리케이션에서 무한공간탐험을 사용하는 것이 가능하다.

### 참고문헌

[1] S. Razzaque, D. Swapp, M. Slater, M. C. Whitton, and A. Steed. Redirected walking in place. In EGVE, vol. 2, pp. 123-130, 2002.  
 [2] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. IEEE transactions on visualization and computer graphics, 16(1):17-27, 2009.  
 [3] Mohler. Velocity-dependent dynamic curvature gain for redirected walking. IEEE transactions on visualization and computer graphics, 18(7):1041-1052, 2012.

# Unity MARS를 이용한 증강현실 기반 불 특수효과 콘텐츠 개발\*

김보성, 송오영

세종대학교 소프트웨어학과

boseongkim96@gmail.com, oysong@sejong.ac.kr

## Fire Special Effect Content Development by Augmented Reality Using Unity MARS

Boseong Kim, Oh-young Song

Dept. of Digital Contents, Sejong University

### 요약

증강현실(Augmented Reality, AR) 기술은 사용자가 자각하는 현실에 컴퓨터가 만든 정보를 추가하는 것으로, 현실 세계에 가상 물체를 겹쳐 보여주는 기술이다. 실제로 위험한 현상이나 실제 세계에서 구현되기 어려운 환경, 혹은 위험한 환경을 증강현실 기술을 통해 재현한다면 훈련, 교육 분야에 널리 활용될 수 있다. 예를 들어, 불에 대한 특수효과는 화재 상황과 같은 실제 현실에서 구현할 수 없는 상황은 증강현실을 활용하면 현실에 불이 실제로 난 듯한 효과 구현이 가능하다. 이 논문은 증강현실 기반 불 특수효과 생성 기법을 제시하고 불의 물리적 특성을 상황에 따라 조절할 수 있는 터치기반 사용자 인터페이스도 소개한다.

### 1. 서론

전세계에 있는 스마트폰의 카메라와 센서를 활용하여 간단한 증강현실 앱을 활용하거나 개발할 수 있도록 증강현실 기술은 보편화되었다. 모바일 환경의 증강현실 제작에 있어 iOS와 Android는 각각 ARKit과 ARCore이라는 개발 프레임워크를 제공하고 있으며, 다양한 개발 소프트웨어 엔진들이 해당 프레임워크를 이용한 증강현실 콘텐츠 개발에 대한 지원을 하고 있다. 이 논문에서는 Unity 게임 엔진에서 개발한 워크플로 저작(authoring) 도구인 Unity MARS[1]를 활용하여, 불 특수효과를 활용하여 상호작용할 수 있는 AR 어플리케이션을 개발 과정과 결과를 제시한다. 불은 대표적인 시각적 효과가 뛰어난 특수효과이기에 해당 특수효과를 구현하였다. Unity MARS에서 기본 제공되는 로봇 에셋을 이용해 이 로봇이 증강현실을 돌아다니면서 구현되어 있는 불을 끌 수 있도록 만들었으며, 실감적인 경험, 자연스러움, 사용자와의 상호작용을 중점에 두었다.

### 2. Unity MARS 개발

#### 2.1. 개발 환경 설정

연구 목적의 개발에 있어 빌드가 이루어지는 플랫폼의 선정은 매우 중요하다. 이 논문에서는 별도의 최적화 과정이 없어도 영향을 크게 받지 않도록 강력한 성능을 가지고 있으며, 가장 앞선 AR 기술을 가지고 있는 iOS 기기를 활용한 빌드를 진행하였다. iPhone 12 pro와 iPad pro를 비롯한 iOS 기기들은 타 운영체제를 탑재한 모바일 기기에 비해 월등한 성능을 가지고 있으며, 깊이 인식에 큰 도움을 주는 lidar scanner를 탑재하여 증강현실 개발에 있어 큰 우위를 가지고 있다. 개발 엔진은 Unity MARS를 이용한다. Unity MARS의 AR 프레임워크인 AR Foundation은 빌드의 대상이 된 플랫폼이 iOS 기기라면 자동으로 ARKit 프레임워크 빌드로 전환시켜 준다.

#### 2.2 불의 특수효과 표현

Unity 엔진은 사실적으로 구현되어 있는 다양한 특수효과를 가진 particle system 기반 asset을 구할 수 있다. Particle system은 전문 지식 없이도 간단하게 특수효과 물리적 특성(크기, 속도와 같은 물성치)에 대한 기본적인 조절(그림 1)을 할 수 있어 시각적 효과 구현에 용이하다. 이 논문에서는 불의 세기를 임의로 조절할 수 있는 asset을 이용하였다. 로봇이 특정 지점에서 불길이 일어난 곳을 끄도록 하기 위해 세기를 줄이는 동시에 시작 속도를 빠르게 하여 특정 부분에서 일어난 불길이 타고 있는 것과 같이 연출되도록 하였다.



그림 1: Fire Asset의 물리적 특성 조절

\* "본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT연구센터지원사업의 연구결과로 수행되었음" (IITP-2021-2016-0-00312)

\* 본 연구는 산업통상자원부와 한국산업기술진흥원의 "국제공동기술 개발사업"의 지원을 받아 수행된 연구결과임. (과제번호 P0016038)

### 2.3. 소방 로봇

Unity MARS에서는 AR 개발을 간편하게 할 수 있도록 기본적인 asset 및 시나리오 scene을 제공하고 있다. 이 논문에서는 기본적으로 제공되는 asset 중 하나인 Unit-E로 명명된 로봇을 활용한다. Unit-E는 터치 조작을 활용하여 사용자와 증강현실 간의 상호작용을 가능하도록 해주며, tab을 하여 지정해서 터치한 특정한 위치로 순간이동을 하거나, swipe를 하여 터치를 하는 쪽으로 따라가도록 해 줄 수 있다. 해당 로봇을 활용해, 증강현실 세계를 여기저기 이동하며 불을 끄고 다닐 수 있도록 개발하였다. 불을 끄러 다니는 소방관과 같은 연출을 위해, 해당 로봇의 물 분출구를 설정한 앞부분에서 물줄기가 나오는 특수효과를 추가하였다. 해당 특수효과는 particle system 컴포넌트를 사용하였으며, 직관적으로 물줄기 폭과 세기를 임의로 조절할 수 있다.



그림 2: Unit-E를 활용한 소방 로봇

### 2.4. 시나리오 구성

해당 프로그램은 증강현실 세계에 불을 적용시킨 이후 해당 불길을 로봇이 끄면서 이동해야 한다. Unity MARS는 수평 표면(horizontal plane)과 수직 표면(vertical plane)을 인식할 수 있도록 하며, 이는 iOS 기기에 탑재된 lidar scanner를 통해 빠른 인식이 가능하다. 증강현실 상의 광범위한 화재 표현을 위해, 간단히 의미 있는 수평 표면이 인식되었을 경우 해당 표면에 불 이펙트가 발생되도록 진행하였다.

이후 로봇에 해당 불길과 닿았을 때 불길을 끄도록 해야 한다. Unity에서 해당 상황이 필요할 때는 보통 좌표 계산, 또는 불길이 있는 곳에 따로 가상의 collider를 세워 충돌 여부를 인식하는 두 가지 방법 중 하나를 이용한다. 이 논문에서는 추후 보정 작업을 할 시 수정이 용이하도록 로봇과 불길의 좌표가 일치하게 닿을 경우 불 특수효과가 사라지도록 하였다. Unity는 Vector3를 통해 x, y, z 좌표를 제공하고 있으며, 해당 특수효과와 소방 로봇 모두 동일하게 수평 표면 위에서 동작하고 있다. 따라서 제작한 불 특수효과에 충돌을 인지할 수 있는 FireController script를 작성하여, 소방 로봇과 불길이 겹쳤을 경우 Trigger가 작동하도록 설계해 해당 특수효과가 Destroy 되도록 하였다.

### 3. 결과

이 Unity 프로젝트는 macOS 운영체제를 탑재한 컴퓨터를 통해 iOS 기기로 타겟하여 빌드되었다. 사용자는 터치 인터페이스를 이용하여 로봇을 증강현실 세계 안에서 자유롭게 조종하며 상호작용할 수 있다. 현실에서 실제로 화재가 일어난 것과 같은 실감적인 연출을 보여주며, 로봇을 이용하여 직접 불을 끄면서 특수효과에 대한 체험이 가능하다.



그림 3: 실제 모바일 기기에서의 빌드

기본 프로그램은 설계한 목표에 따라 특수효과가 자연스럽게 실감적으로 증강현실 세계에 표현되었으며, 사용자가 원하는 위치로 이동해 불을 끌 수 있는 것이 확인되었다. 이를 응용해 Unity 프로젝트 내에서 불의 세기 등을 원하는 대로 바꿔 표현할 수 있으며, 특수효과 asset을 교체하여 불이 아닌 다른 특수효과를 적용할 수 있다.

### 4. 결론

증강현실은 실제 세계에서 표현하지 못하는 다양한 시나리오 및 특수효과를 마치 현실에서 일어난 것과 같이 표현할 수 있다. 본 개발을 통해 AR이 가질 수 있는 장점과 사실적인 특수효과에 대한 표현이 가능했다. 또한 특수효과에 대한 변화를 통해 원하는 효과에 대한 구현을 할 수 있었다. 이와 같이 증강현실과 특수효과에 대한 접목은 몰입감과 실존감에 있어 큰 강점을 가지고 이를 활용한 콘텐츠를 진행할 수 있다.

본 개발은 이후 다양한 연구에 대한 진행이 가능하도록 설계되어 있다. 예시로, 불과 같은 특수효과가 만들어지는 위치를 따로 설계해, 불이 발생하는 물체의 물성치에 따라 불의 세기가 달라지는 등의 효과를 입힐 수 있다. 향후 연구로 딥러닝 기반의 물체 인식 기술을 활용하여 AR 상의 인식된 물체에 따라 구분된 특수효과를 넣는 기술 개발을 진행 중이다.

### 5. 참고 문헌

[1] <https://unity.com/kr/products/unity-mars>  
 [2] Du, Ruofei, et al. "DepthLab: Real-Time 3D Interaction With Depth Maps for Mobile Augmented Reality." *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 2020.

# 환자 특이적 수술 내비게이션 개발\*

한예진, 홍기윤, 김성재, 성낙준\*  
(주)휴툼

{yennnyhan, gyoonh, jay.kim, njsung}@hutom.io

## Development of Patient Specific Surgical Navigation

Yejin Han, Geeyun Hong, Sungjae Kim, Nak-Jun Sung\*  
Hutom Corp.

### 요약

디지털 트윈은 현실에서 발생할 수 있는 상황들을 시뮬레이션해 미리 결과를 예측할 수 있는 기술을 말한다. 본 연구진은 최근 헬스케어 분야에서도 다양하게 적용되고 있는 디지털 트윈 기술을 활용해 장기·혈관의 변형이 빈번한 복부 암종을 대상으로 환자 특이적 해부학적 정보와 수술 단계에 대한 정보를 함께 제공하는 수술 내비게이션을 연구 및 개발하였다.

### 1. 서론

디지털 트윈(Digital Twin)은 컴퓨터에 현실 속 사물의 쌍둥이를 만들어 현실에서 발생할 수 있는 상황들을 시뮬레이션해 미리 결과를 예측할 수 있는 기술을 말한다 [1-2]. 디지털 트윈 기술은 다양한 산업 분야에 적용되고 있으며 특히 최근 헬스케어 분야에서도 다양하게 적용되고 있다.

CT(Computed Tomography)와 MRI(Magnetic Resonance Imaging)는 환자의 해부학적인 특징을 잘 표현할 수 있는 영상 촬영 기법으로, 헬스케어 분야의 디지털 트윈 기술은 주로 두 영상을 활용한 3차원 재구성(3D Reconstruction)을 이용한다.

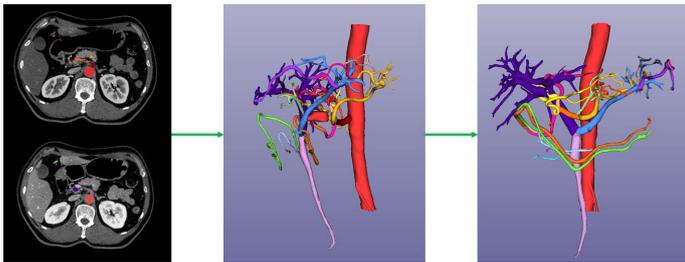


그림 1: CT 이미지를 활용한 3차원 재구성 결과

본 연구에서는 관심 장기와 혈관의 변형이 빈번하게 발생하는 복부 암종에 관한 연구를 진행하였다. 복부 암종은 집도의가 수술 중 환자의 구조적 변이에 대처하는 것이 굉장히 중요하지만, 이는 경험이 많은 외과의도 예측하기 어려운 문제 중 하나로 손꼽힌다.

본 연구진은 환자의 해부학적 구조를 수술 중 외과의에게 효과적으로 제공할 수 있는 디지털 트윈 기반 환자 특이적 3D 모델링 기술과 이를 활용한 수술 내비게이션 서비스를 연구·개발하였다.

### 2. 관련 연구

#### 2.1. 이미지 유도 수술

이미지 유도 수술(ISG: Image-Guided Surgery)은 환자의 영상자료와 실물을 일치시켜 해부학적 정보를 수술 중 제공하는 기법으로 주로 의사의 판단을 돕는데 사용된다[3-5]. 수술 내비게이션 서비스는 이러한 기법을 더욱 발전시킨 것으로, 환자의 해부학적 정보를 제공해 줄 뿐 아니라 수술 단계에 대한 정보를 제공해 수술 중 의사의 판단과 불필요한 의사결정을 줄여줄 수 있다 [6-7].

#### 2.2. 3D 재구성

3D 재구성(3D Reconstruction)은 여러 장의 2D 이미지로부터 3D 이미지를 만들어내는 것으로, 의료 분야에서는 환자의 해부학적 정보를 확인하는 데 주로 사용되고 있다[8-9].

### 3. 수술 내비게이션

#### 3.1. 환자 특이적 3D 모델링

수술 상황에 맞는 환자 특이적 3차원 모델을 생성하기 위해서는, 환자당 약 200~500장의 CT 이미지에 대한 정교한 장기, 혈관 분할이 필요하다. 그러나 이는 고비용의 인력과 시간을 요구하기 때문에, 장기의 변형이 빈번하게 발생하는 복부 수술의 경우 CT 이미지에서 관심 장기와 혈관을 자동으로 분할하는 기법이 핵심 요소로 작용한다. 본 연구진은 위암의 관심 장기를 자동으로 분할할 수 있는 딥러닝 기술을 연구하였다.

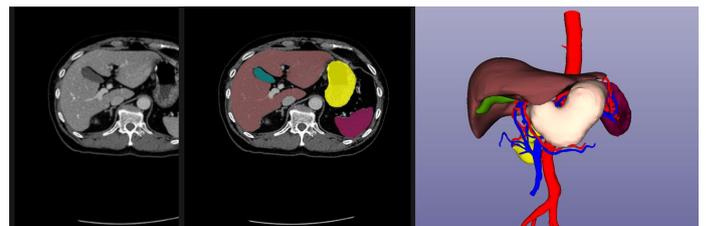


그림 2: CT 이미지를 활용한 3차원 재구성 결과

\* 구두(포스터) 발표논문

\* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 타 저널에 제출 중

### 3.2. 수술 내비게이션

환자 특이적으로 모델링 된 3차원 해부학적 정보는 Unity Engine을 이용해 개발한 수술 내비게이션 시스템(RUST™)을 활용해 외과의에게 제공된다. 해부학적 구조를 관찰하는 시점은 Intuitive Surgical사의 수술 로봇인 daVinci Xi®의 Endoscope(내시경 카메라)와 동일하게 제공하고 있으며, 실제 수술과 같은 기구학적 메커니즘을 통해 해부학적 구조를 관찰할 수 있도록 구현하였다.

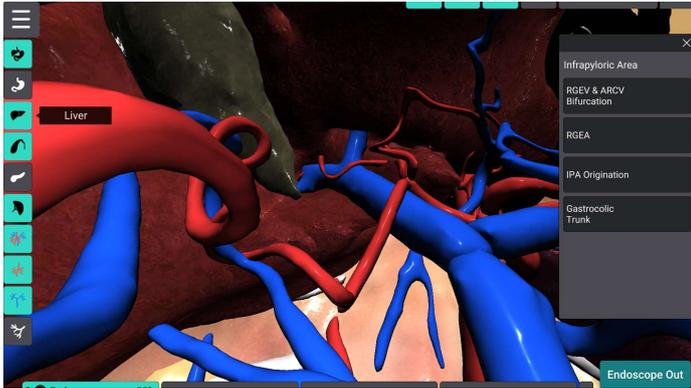


그림 3: 수술 내비게이션(RUST™)

본 연구진은 위아전절제술(Subtotal Gastrectomy)의 주요한 혈관과 장기의 위치를 확인할 수 있도록 POI(Point of Interest) & TP(Target Point)의 개념을 도입하였으며, 이를 활용해 수술 중 시점 및 현재 수술의 단계와 동일한 해부학적 시야를 제공하는 내비게이션 서비스를 구현하였다.

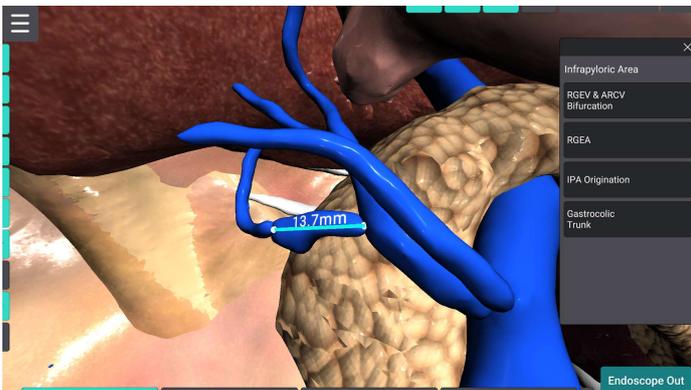
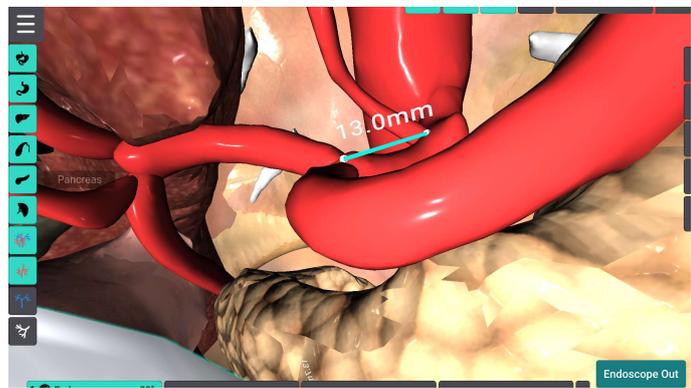


그림 4: 수술 단계에 따른 내비게이션(POI&TP)

### 4. 결론

본 연구진은 최근 활발하게 연구되는 디지털 트윈 기술을 수술 분야에 접목해, 수술 중 외과의에게 도움이 되는 내비게이션 시스템을 연구하였다. 환자의 CT 영상을 활용해 해부학적 정보를 재구성하고, 이를 외과의에게 제공해 장기와 혈관 변이에 대한 대응을 가능하게 하며, 이를 활용해 수술 단계에 맞는 정보를 제공한다. 내비게이션 시스템은 최소침습수술과 유사한 기구학적 메커니즘을 기반으로 동작하도록 구현하였으며, 시스템에 관한 임상 연구를 계획 중에 있다.

또한 실시간으로 수술 중 입력되는 내시경 영상을 분석해 현재의 수술 단계를 파악하는 기술을 연구하고 있으며, 이를 활용해 수술 내비게이션을 실제 영상과 정합해 자동화된 수술 정보 제공을 목표로 연구를 진행하고 있다.

### 참고문헌

- [1] Zheng, Yu, Sen Yang, and Huanhong Cheng, "An application framework of digital twin and its case study," *Journal of Ambient Intelligence and Humanized Computing* 10(3), pp.1141-1153, 2019.
- [2] Zhuang, C. B., Liu, J. H., Xiong, H., Ding, X., Liu, S., and Weng, G, "Connotation, architecture and trends of product digital twin," *Computer Integrated Manufacturing Systems*, 23(4), pp.753-768, 2017.
- [3] Grimson, W. E. L., Kikinis, R., Jolesz, F. A., and Black, P. M, "Image-guided surgery," *Scientific American*, 280(6), pp.62-69, 1999.
- [4] Labadie, Robert F., Bryan M. Davis, and J. Michael Fitzpatrick, "Image-guided surgery: what is the accuracy?," *Current opinion in otolaryngology & head and neck surgery*, 13.1, pp.27-31, 2005.
- [5] Barone, Damiano Giuseppe, Theresa A. Lawrie, and Michael G. Hart, "Image guided surgery for the resection of brain tumours," *Cochrane Database of Systematic Reviews*, 2014.
- [6] Soler, L., Hostettler, A., Collins, T., Pessaux, P., Mutter, D., and Marescaux, J, "The Visible Patient: Augmented Reality in the Operating Theater," *In Digital Surgery*, pp. 247-260, 2021.
- [7] Widmann, G., Widmann, R., Widmann, E., Jaschke, W., and Bale, R, "Use of a surgical navigation system for CT-guided template production," *International Journal of Oral & Maxillofacial Implants*, 22(1), 2007.
- [8] Zachow, Stefan, Michael Zilske, and Hans-Christian Hege, "3D reconstruction of individual anatomy from medical image data: Segmentation and geometry processing," *ANSYS Conference & CADFEM Users' Meeting*, 2007.
- [9] Zheng, Liping, Guangyao Li, and Jing Sha, "The survey of medical image 3D reconstruction," *Fifth International Conference on Photonics and Imaging in Biology and Medicine*, Vol.6534, 2007.

# 모바일 매니플레이터 로봇을 이용한 TSP 펜아트\*

임은정<sup>0</sup>, 박지윤<sup>0</sup>, 송다은, 김영준  
이화여자대학교 컴퓨터공학과

{ejunglim12, jiyoopark13, daeunsong}@ewhain.net, kimy@ewha.ac.kr

## TSP Pen Art using a Mobile Manipulator Robot

Eunjung Lim<sup>0</sup>, Jiyoon Park<sup>0</sup>, Daeun Song, Young J. Kim  
Dept. of Computer Science and Engineering, Ewha Womans University

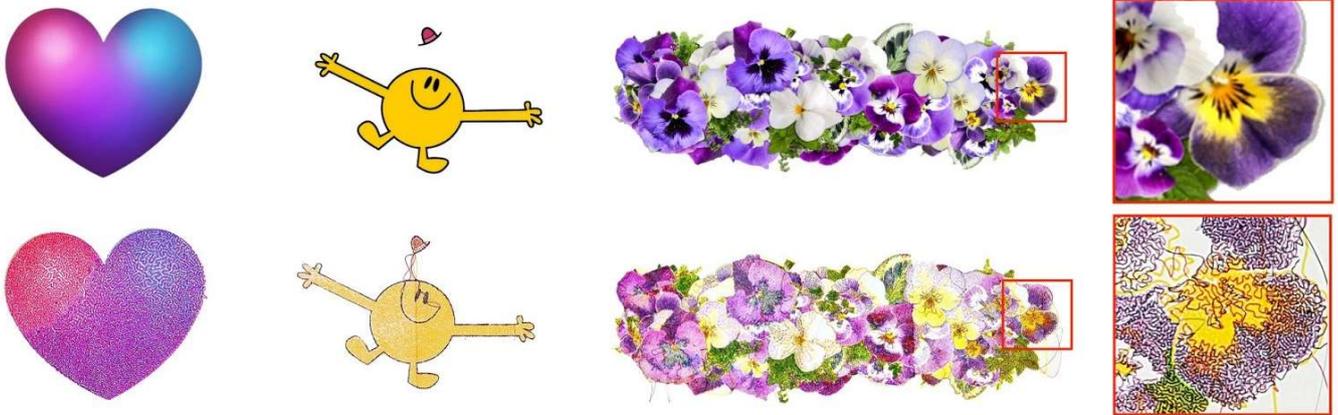


그림 1: (왼쪽부터) 하트, 캐릭터, 꽃 입력 이미지(윗 행)와 로봇 TSP 펜아트 드로잉 결과(아래 행)

### 요약

본 연구에서는 모바일 매니플레이터 로봇을 이용한 TSP 펜아트 드로잉 시스템을 제안한다. 입력된 임의의 컬러 래스터 이미지를 로봇으로 드로잉이 가능한 벡터 이미지로 변환하기 위하여 주어진 문제를 Traveling Salesman Problem (TSP)로 변환한다. 이 벡터 이미지를 전 방향 이동이 가능한 모바일 매니플레이터 로봇을 이용하여 넓은 캔버스 영역에 그려낸다. 제안하는 시스템을 통해 본 연구에서는 다양한 입력 이미지에 대해서 대영역 실사 펜아트 드로잉이 가능함을 보였다.

### 1. 서론

현재 로봇은 산업 영역뿐만 아니라 미술 분야에서도 활발히 연구되고 있다. 캐리커처를 그려주는 화가를 모방한 로봇, 디지털 이미지를 수채화로 그려주는 로봇, 벽화와 같은 대형화를 그리는 로봇 등이 존재한다. 하지만 대부분의 경우, 로봇의 베이스가 고정되어 제한된 영역에 그림을 그린다. 본 연구에서는 로봇의 베이스가 고정되어 그림을 그린 기존 연구[1]를 확장하여 베이스가 자유롭게 이동하는 모바일 매니플레이터 로봇을 이용한 로보틱 펜아트 드로잉 시스템을 제안한다.

제안하는 로봇 펜아트 시스템은 크게 드로잉 입력 전처리, 펜 드로잉, 그리고 대영역 캔버스를 위한 모바일 주행으로 이루어져 있다.

본 연구에서는 래스터 이미지를 입력 데이터로 활용한다. 래스터 이미지로부터 매니플레이터 로봇 말단에서 그림을 그릴 수 있는 경로를 생성하기 위해 TSP 아트 드로잉 기법을 사용한다. TSP 아트는 점묘법으로 표현된 그림의 점들을 Traveling Salesman Problem을 이용하여 연속된 선으로 그림을 그리는 기법이다. 이는 3차원 경로를 입력 받는 매니플레이터 로봇이 그려내기 어렵고, 정교한 표현이 가능한 로봇 드로잉에 적절하다.

### 2. 드로잉 입력 전처리

물리적인 펜으로 디지털 컬러 이미지의 다채로운 색상을 표현하기 위해 컬러 스페이스를 CMYK 네 가지 색의 조합으로 표현하는 4도 인쇄법을 사용하였다. 입력 이미지로부터 각각 C, M, Y, K 색상 값을 추출하여 저장한다. 본 연구의 그림 도구인 마커펜을 이용하여 색의 복제율을 실험한 결과, 검정색은 다른 색채를 압도하는 것을 확인했다. 따라서, 검정색의 기여분을 최소화하기 위해 입력 이미지를 각 색상 값으로 나누어 저장할 때, 이미지의 임계값과 채도를 조정하였다.

<sup>0</sup> 해당 저자들은 본 연구에 동등하게 기여하였다.

\* 학부생 주저자 논문, 구두 발표 논문, 요약논문(Extended Abstract).

\* 본 연구는 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2021-2020-0-01460).

### 3. 펜 드로잉

#### 3.1. TSP 아트

CMYK 로 분리된 색상 별 이미지 데이터를 TSP 아트로 변환하기 위해 가중 보로노이 점묘법을 사용한다. 가중 보로노이 점묘법은 보로노이 다각형을 계산하고 각 다각형의 중심점을 찍는 기법이다 [2]. 가중 함수에 따라 점들의 밀도가 결정되어 이미지의 명도 표현이 가능하다.

이후 Concorde TSP solver [3]를 이용하여 보로노이 사이트들을 잇는 2 차원 로봇 경로를 생성한다. Concorde 는 cutting-plane method 를 재귀적으로 이용하여 TSP 문제를 효율적으로 해결한다. 이는 TSP 를 LP-relaxation 하여 푸는 방법으로 현존하는 빠른 TSP 해결법 중 하나이다. 한편, 보다 매끄러운 로봇 경로를 생성하기 위해 TSP 노드들 사이에 추가 경로 좌표들을 삽입한다. 생성된 경로상의 모든 경로 좌표들을 매니플레이터 로봇이 드로잉 가능한 워크스페이스 영역별로 분할하여, 로봇이 다수의 영역 간을 주행하며 대영역 드로잉을 완성하도록 한다.

#### 3.2. 로보틱 드로잉

본 연구에서는 매니플레이터의 정확도와 안정성이 보장되는 7 자유도 로봇인 KUKA IIWA 7 R800을 이용해 로봇 드로잉을 구현하였다. 특히, 고정된 캔버스에 펜화를 그릴 때, 캔버스로부터 가해지는 외력에 적응하기 위해 임피던스 제어를 사용한다 [1].

TSP 아트를 통해 계산한 점들의 좌표는 2차원 정보이지만 캔버스에 물리적인 드로잉을 수행하기 위해서는 3차원 정보가 필요하다. 이를 위해 로보틱 드로잉을 수행하기 전, 마커펜을 캔버스에 수직으로 운용하여 닿는 곳의 좌표 값을 저장한다. 해당 좌표로부터 캔버스와 매니플레이터 로봇 사이의 깊이 정보 값을 획득하여 3차원 드로잉 경로를 생성한다.

본 연구에서는 로봇 제어를 위해 Robot Operating System (ROS)를 미들웨어로 사용한다. 캔버스는 1절지 종이를 벽에 부착하여 사용한다. 그림 도구는 C, M, Y, K 색상 값을 대표하는 마커펜 4개이며, 이를 하나씩 로봇 말단에 부착하여 색상 별 경로에 따른 드로잉을 수행한다.

### 4. 대영역 캔버스를 위한 모바일 주행

본 연구에서는 전 방향 이동이 가능한 Clearpath Ridgeback 을 모바일 플랫폼으로 이용하였다. 모바일 플랫폼은 벽면에 부착된 캔버스를 마주하며 좌우로 이동해 대영역 드로잉을 가능케 한다.

펜과 캔버스의 지속적인 접촉과 안정적인 좌우 이동을 보장하기 위해 로보틱 드로잉을 하기 전 로봇을 벽에 평행하게 정렬시킨다. 이를 위해 로봇의 정면에 부착된 라이더 센서 정보를 이용하여 정면을 기준으로 좌우 30도의 거리 센서값을 입력받아 이들 지점까지의 거리

의 차이가 5mm 이내가 되도록 로봇을 회전시켜 캔버스와의 각도를 정렬시킨다.



그림 2: 매니플레이터와 모바일 플랫폼이 결합된 로봇

대영역 드로잉을 완성하기 위하여 분할된 영역별로 드로잉을 수행한다. 다음 영역으로 이동하기 위해 모바일 로봇은 각 영역의 크기만큼 캔버스와 평행하게 느린 등속 주행을 한다.

### 5. 결과

그림 1 은 제안하는 시스템을 통해 그려낸 TSP 펜아트의 결과이다. 하트에는 모바일 로봇이 활용되지 않았으며, 캐릭터와 꽃은 모바일 로봇을 각각 2, 3 번 오른쪽으로 0.35m 씩 움직여 완성한 그림이다.

표 1: 각 이미지의 메타데이터

이미지	크기(m)	점 개수	TSP 소요시간	그림 소요시간
하트	0.40 × 0.35	34,849	6 분	120 분
캐릭터	0.60 × 0.40	31,151	34 분	110 분
꽃	0.85 × 0.30	95,155	97 분	585 분

### 6. 결론

본 연구에서는 모바일 매니플레이터 로봇을 이용한 TSP 펜아트 드로잉 시스템을 제안하였다. 다양한 이미지로부터 TSP 아트를 만들어 낼 수 있고, 시스템의 모델리티를 이용하여 복잡한 그림을 대영역 캔버스에 그릴 수 있다.

향후 연구로 카메라를 이용한 모바일 로봇의 이동 정확도 향상과 자동 펜 교체 등의 사람의 개입 최소화를 계획하고 있다.

### 참고문헌

- [1] D. Song, T. Lee, and Y.J. Kim, Artistic pen drawing on an arbitrary surface using an impedance-controlled robot, *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.
- [2] A. Secord, Weighted Voronoi Stippling, *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering (NPAR 2002)*, p37-43, 2002.
- [3] D. Applegate, R. Bixby, V. Chvatal, and W. Cook, Concorde: A code for solving traveling salesman problems. <http://www.tsp.gatech.edu/concorde.html>.

# 사람의 EOS X-ray 스캔으로부터 표면 메시 추정\*

장필식<sup>0</sup>, 이제희  
서울대학교  
{phil, jehee}@mrl.snu.ac.kr

## Estimating surface mesh from EOS X-ray scan of human body

Philsik Chang<sup>0</sup>, Jehee Lee  
Seoul National University

### 요약

사람의 표면 메시 데이터는 인체 시뮬레이션을 수행하기 위해 필요한 중요한 정보이다. 이를 얻기 위해 우리는 별도의 3차원 스캐너 없이, 병원에서 촬영한 EOS X-ray 스캔으로부터 3차원 표면 메시를 복원하는 방법을 제시한다. 일단 EOS 이미지에서 주요 관절 위치를 마킹하고 이미지 thresholding을 통해 실루엣을 얻어낸 다음, 이 정보들을 이용해 메시를 추정하고자 한다. 우리는 적은 수의 파라미터로 사람의 표면 메시를 매개화한 STAR 모델을 사용하여, 메시의 실루엣을 EOS 이미지의 실루엣과 가장 유사하게 맞추었다. 실루엣의 차이를 최소화시키는 STAR 모델의 파라미터들은 CMA-ES 최적화 방법을 통해 찾을 수 있었다.

### 1. 서론

사람에서부터 얻을 수 있는 신체 정보 중 표면의 기하학적인 모양은 현실적인 인체 시뮬레이션을 수행하는데 중요한 정보이다. 사람의 표면 모양에는 키, 몸무게, 팔 다리의 길이와 같은 정보뿐만 아니라, 근육과 지방의 분포가 모두 녹아 들어있다. 이러한 특징들은 사람의 운동 능력에 많은 영향을 주기 때문에, 특정한 환자에 대한 개인화된 인체공학적인 분석을 하는데 큰 도움이 된다. 표면 정보를 직접적으로 얻으려면 3D full-body scanner 장비가 필요한데, 이러한 장비 없이 병원에 일반적으로 구비되어 있는 X-Ray 장비만으로도 표면 데이터를 얻는 방법을 모색했다.

EOS 이미징은 방사능 투여량이 적은 X-Ray 방법으로, 환자의 일반 X-Ray 혹은 CT 스캔에 비해 사람의 해부학적인 정보를 적은 시간과 비용으로 얻을 수 있는 방법이다. [1] EOS는 주로 뼈에 문제를 가지고 있는 환자들의 척추, 허리와 무릎에 대한 진단을 하기 위해 사용되어 왔다. 우리는 여기서 나온 데이터를 가지고 뼈의 크기와 모양에 대한 정보뿐만 아니라, 사람의 표면 정보도 같이 추출할 수 있는 방법을 고안해내고자 한다.

### 2. 접근 방법

일반적으로 임의의 2차원의 이미지로부터 3차원의 형상을 복원하는 것은 매우 어려운 작업이다. 하지만 형상이 사람으로 국한되어 있는 경우에는, 사람의 모양을 적은 양의 변수들로 매개화시킨 모델이 미리 있으면 문제가 훨씬 쉬워진다. 이전에 Loper et al.[2] 와 Osman et al.[3] 은 적은 수의 파라미터로 다양한 신체 형상을 생성할 수 있는 모델을 대량의 3D 스캔 데이터셋을 가지고 학습한 바가 있다. 우리는 이 중에서 Osman et al.의 STAR 모델[3]을 EOS 이미지의 실루엣에 가장 가깝게 맞추는 방법을 고안했다.

EOS 기기의 아웃풋은 정면과 측면에서 찍은 흑백 X-ray 사진으로, 해상도가 일반 X-Ray에 비교해서 떨어지지만 몸의 형체를 추정하는 데에는 충분한 정보를 제공한다. 우리는 EOS 이미지에서 정면과 측면의 영역을 지정하고 이미지 상에서의 관절 위치를 마킹하는 GUI를 만들었다. GUI를 통해 전처리 과정을 완료한 후 정면과 측면 이미지에 thresholding을 걸어 사람의 실루엣에 해당되는 흑백 바이너리 이미지를 얻어냈다. 이 두 장의 바이너리 이미지는 마킹한 관절 위치와 함께 메시를 추정하는데 사용된다. (그림 1 참고)

\* 구두발표논문

\* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00878, SW컴퓨팅산업원천기술개발사업(SW스타랩)).

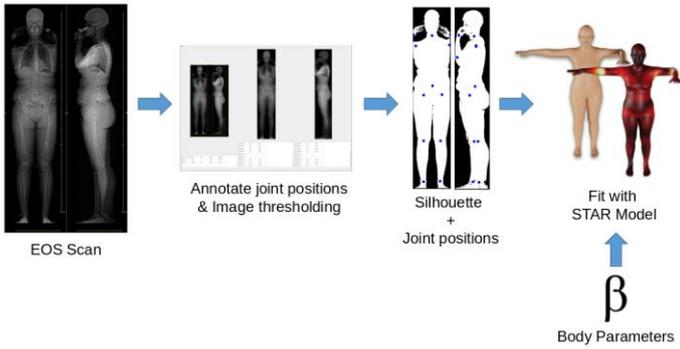


그림 1: EOS 이미지에서 메시를 얻어 내는 전체 과정.

STAR 모델[3]은 3가지 종류의 파라미터(모두 합쳐서  $X = [\beta, \theta, T]$ )를 입력으로 받아 메시를 생성해 주는데,  $\beta \in \mathbb{R}^{20}$ 는 사람의 신체 모양에 영향을 주는 신체 파라미터들,  $\theta \in \mathbb{R}^{72}$ 는 24개의 관절의 회전값을 모은 벡터, 그리고  $T \in \mathbb{R}^3$ 은 루트 관절의 위치를 나타낸다. 모델의 아웃풋으로는 사람 메시의 vertex 위치들  $V(X)$ 와 관절 위치들  $J(X)$ 가 나온다. 이를 정면과 측면 방향으로 렌더링 한 후 바이너리 이미지로 변환하면, 주어진 파라미터들에 대한 실루엣  $S^{front}(X), S^{side}(X)$ 를 얻어낼 수 있다.

이렇게 얻어낸 영역을 EOS 이미지의 실루엣과 가장 가깝게 맞추는 과정은 Won et al.[4]의 아이디어를 바탕으로 수행하였다. 정면 혹은 측면의 영역  $S$ 를 실루엣의 영역  $S_0$ 를 최대한 비슷하게 만들려면  $S$ 와  $S_0$ 가 서로 겹치지 않는 영역 (symmetric difference)의 넓이  $L_{diff}(S, S_0) = Area(S \setminus S_0) + Area(S_0 \setminus S)$ 를 최소화하면 된다. 최종적으로 최소화하고 싶은 목적함수는 다음과 같다:

$$L_{tot}(X) = L_{diff}(S^{front}(X), S_0^{front}) + L_{diff}(S^{side}(X), S_0^{side}) + \lambda_p \|J(X) - J_0\|^2 + \lambda_\beta L_\beta + \lambda_\theta L_\theta$$

식의 첫번째와 두번째 항은 모델을 정면과 측면 실루엣에 가깝게 매칭하는 역할을 하고, 세 번째 항은 모델에서 나온 관절 위치를 EOS 이미지에서 마킹한 위치( $J_0$ )와 가까워지게 한다.  $L_\beta$ 와  $L_\theta$ 는  $\beta, \theta$ 에 대한  $L_2$  정규화 항이다. 최적화의 효율성을 위해 관절들 중 일부는 회전값을 고정하였다.

이렇게 설계한 목적함수는 계산하는 과정에 렌더링이 들어가기 때문에 그라디언트를 계산하기 힘들다. 따라서 우리는 목적함수를 최소화하기 위해 gradient-free method 중 하나인 공분산 행렬 적용 진화 전략 (CMA-ES)를 사용하였다. CMA-ES는 진화 알고리즘의 일종으로, non-convex한 고차원 함수의 최적화에 대해 좋은 성능을 보인다. (구현은 PyCMA[5],  $\sigma = 0.2$  사용)

### 3. 실험 결과

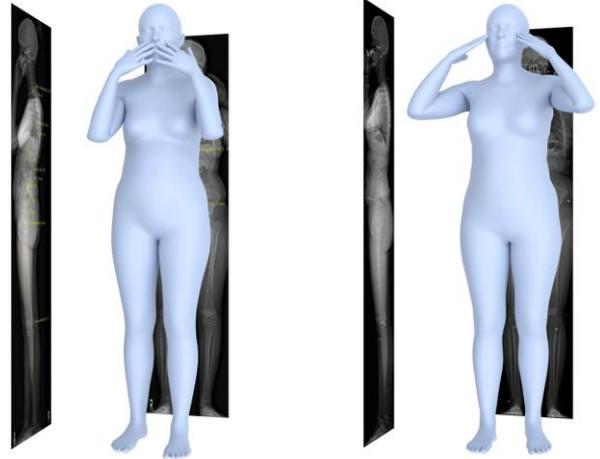


그림 2: 두 명의 환자의 EOS 이미지를 대상으로 메시를 얻어낸 결과.

그림 2는 여성 환자 두 명에 대한 EOS 이미지를 바탕으로 최적화를 돌린 결과를 보여준다. 최적화를 수행하는데 한 사람 당 10분 정도의 시간이 소요되었다. 다양한 카메라 시점에서 메시를 본 더 자세한 결과는 첨부된 동영상에서 볼 수 있다.

본 연구는 EOS 이미지에서 사람의 형체 일부분이 잘려나간 경우에는 자연스러운 사람의 형체를 얻지 못한다는 한계점이 있다. EOS 기기가 캡처할 수 있는 영역의 크기가 한정되어 있기 때문에, 환자의 체형이 기기에 비해 너무 크거나 환자가 똑바로 서 있는 자세를 갖추기 힘든 조건에서는 정확한 메시를 얻기 힘들다.

### 참고문헌

[1] “EOS Imaging,” <https://www.eos-imaging.com/>.  
 [2] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J.Black, “SMPL: a skinned multi-person linear model,” ACM Transactions on Graphics, vol. 34, pp. 1–16, 2015.  
 [3] A. A. A. Osman, T. Bolkart, and M. J. Black, “STAR: SparseTrained Articulated Human Body Regressor,” Computer Vision – ECCV 2020, vol. 12351, pp. 598–613, 2020.  
 [4] J. Won and J. Lee, “Shadow theatre: discovering human motion from a sequence of silhouettes,” ACM Transactions on Graphics, vol. 35, pp. 1–12, 2016.  
 [5] N. Hansen, Y. Akimoto, and P. Baudis, CMA-ES/pycma on Github, 2019.

# 맞춤형 의료기기 제작을 위한 클라우드기반 협업서비스 개발

정영진<sup>1</sup>, 김태한<sup>1</sup>, 서안나<sup>1\*</sup>  
<sup>1</sup>기업부설연구소, 주식회사 시안솔루션  
{youngjin, taehan, anseo}@seeann.net

## Development of cloud-based collaboration service for custom medical device production

Young-Jin Jeong<sup>1</sup>, Tae-Han Kim<sup>1</sup>, Anna Seo<sup>1\*</sup>  
<sup>1</sup>research institute, SEEANN Solution Corporation.

### 요약

임상현장에서 기존 상용의료기기 형상의 한계로 수술적 치료를 적용할수 없는경우 3D프린팅 기술을 활용하여 개인맞춤형 의료기기를 제작하고 있다. 정확한 맞춤형 3D프린팅 의료기기(보조기, 임플란트) 제작을 위해서는 의료영상 기반 인체 해부학적 구조 및 맞춤형 의료기기의 3차원 모델 생성 및 형상 검증 작업이 반드시 필요하다. 이러한 협업은 의료 3D모델을 생성하는 모델러와 생성된 3D 모델을 검증하는 의료전문가 사이에 이루어지며, 정확한 제품을 제작한 과정이다. 기존 정확한 의료기기 제품 제작을 위해서는 모델러가 작업을 완료한 후 의료진이 있는 병원에 방문하여 설계한 내용을 설명하고, 수정할 사항을 의료전문가로부터 조언을 듣는 방식으로 이루어졌다. 이러한 확인절차는 많은 시간을 필요하게되고 제품 제작까지 많은 비용을 소비된다. 본 논문에서 제안하는 클라우드기반 협업서비스(Reconeasy Collabo)는 기존의 협업방식의 한계를 극복하고 빠른 확인절차와 제품 제작시간을 줄이는데 기여할수 있을것이다.

### 1. 서론

최근 3D프린팅 산업의 발달과 양질의 의료서비스 수요에 따라 환자 맞춤형 보조기나 수술용 임플란트, 수술계획용 인체모형 제작에 대한 요구가 늘어나고 있다 [1]. 의료인이 이러한 필요성을 느끼더라도 3차원 모델링 SW를 사용해야하기에 의료영상의 영상분할과 3차원 모델링, 기구 설계 등을 전문 모델러에게 의뢰하는 경우가 많다. 이러한 업무를 수행하게된 3D모델러는 영상분할과 3차원 모델링의 각 단계마다 인체해부학적 구조와 관련된 내용을 의료전문가의 확인을 받아야한다. 그러나 이를 도와줄 협업 플랫폼이 미비하여 이러한 협업 업무를 정확히 보조할수 없는 실정이다. 이에 의료인과 3차원 모델링 작업자/의료전문가/맞춤형 3D프린팅 의료기기 제작자 간의 협업을 위해 클라우드(cloud)기반 프로그램을 개발하였다[그림1].

## 2. Cloud 기반의 3D Model viewer(Reconeasy Collabo)

### 2.1. Cloud 기반의 편리한 사용성

기존의 협업 플랫폼은 특정 모델링 소프트웨어에 종속된 형태로 서비스되어, 의료인과 모델러 모두 해당 모델링 소프트웨어를 사용하고 있지 않은 경우 협업에 어려움이 있었다. 이에 Collabo는 web 서비스 형태로 개발하여 플랫폼의 제약으로부터 자유롭도록 하였고, PC와 모바일 환

경 모두에서 언제든 협업을 위한 업무를 수행할수 있다. 협업을 위해 공유되어야할 3차원 모델 데이터를 cloud 저장공간에 업로드하고, 해당 데이터가 업로드된 Viewer의 주소(URL)을 복사해서 협업 대상에게 전달하면 Collabo를 통한 협업 준비는 끝이다[그림2].

언제 어디서든 web browser에서 해당 URL로 이동하면 Viewer가 열리고 협업을 위해 의견을 주고 받아야할 3차원 모델이 나타난다. Cloud 저장공간을 통해 3차원 모델 데이터를 공유하기 때문에 모바일 환경에서 무거운 모델 데이터를 주고 받아야하는 불편함도 해소되었다.



그림 1. SeeAnnSolution Collabo

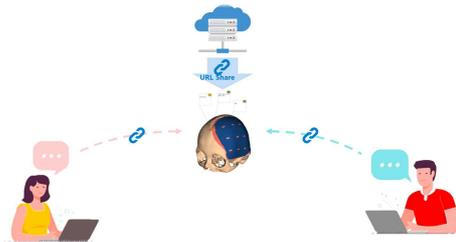


그림 2. 협업솔루션 구성

### 2.2. 협업을 위한 메시지 전달

보통 맞춤형 임플란트나 보조기 모델링 제작에서 모델러가 작업한 3차원 모델을 보고 의료인이 영상분할의 실수와 모델링 및 설계에서 아쉬운 부분을 이야기하고 서로의 의견을 교환하며 해부학적 구조 확인과 수정을 반복하게 된다. 이때 의료인의 수정을 요구하는 영역이 정확히 어디인지 인체해부학 지식이 부족한 모델러가 이해하기 쉽지 않다. 그리하여 Collabo에서는 3차원 모델의 특정 부위를 포인팅하고, 포인팅된 지점마다 메모를 남길수 있도록 기능을 제공하여 의료용 모델링 작업에서 의견 교환이 정확히 이루어질수 있도록 하였다[그림3a].

모바일 환경에서 협업을 진행하는 경우 작은 스마트폰 화면에 손가락으로 포인팅하고 타이핑으로 메모를 작성하기 어려울 수 있기 때문에 Viewer 화면에 손글씨로 의견을 작성하거나 그림을 그려 내용을 설명할 수 있는 기능을 개발하였다[그림3b].

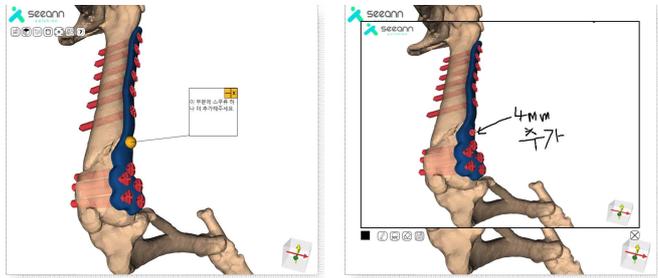


그림 3. 메모 작성. a) pointing-text 메모, b) 이미지 메모

### 2.3. 3D Model Viewer 시각화 기능

복잡한 인체구조와 임플란트, 보조기 등의 구조를 명확히 파악할 수 있도록 Collabo에서는 다양한 시각화 기능을 제공한다. 3차원 모델의 표면을 표현하는 방법을 점(point cloud), 선(wire frame), 면(surface)으로 선택할 수 있고[그림4], 3차원 모델이 여러 부위로 이루어져 있을 경우, 각각의 색상이나 투명도를 조절하여 내부에 위치한 구조를 파악할 수 있도록 하였다[2]. 또한, 각 모델의 내부 면을 확인하거나, 모델의 위치를 이동시켜 볼 수 있도록, 세부 모델별로 이동/회전 기능이 제공된다[그림5].



그림 4. 표면 시각화: a) point cloud, b) wire frame c) surface

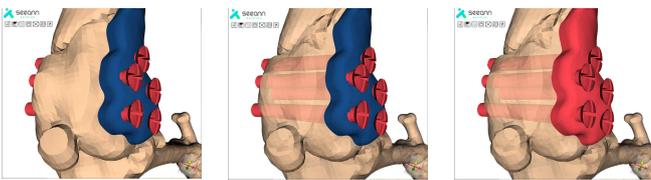


그림 5. 3D모델표현: a)기본, b)투명화, c)색상변경

## 3. 협업 활용 사례

### 3.1. 수술 계획용 소아심장 모형 모델링

소아심장 수술의 수술계획용 심장 모형 모델링에서 협업 솔루션이 사용된 사례이다. 심장영역의 복잡한 해부학적 구조로 수술에 어려움이 있는 경우가 많은 선천성 소아심장기형 수술에서 수술계획용으로 소아심장 3D모델과 3D프린팅 모형 제작이 필요한 경우 3일 이내 작업이 완료되어야 하는 경우가 있다.

그러나 긴급으로 수행되는 작업이라도 정확한 심장 해부구조 모델링 결과물에 대한 의료인의 확인은 반드시 거쳐야 하나 의료인의 바쁜 스케줄 문제로 제때에

검수가 이루어지지 않은 경우가 발생할 수 있다. 이러한 사례의 경우는 모델러가 모델링 완료 후 Collabo SW를 통해 검수가 필요한 부분에 대한 메모를 작성해 두어, 의료인의 응답을 무한정 기다리지 않고 다른 작업을 진행할 수 있었고, 의료인이 잠시 시간을 내어 메모를 확인 후 피드백을 주면 모델러는 이것을 받아 정정 작업을 신속히 완료할 수 있어, 임상속도에 맞춘 3D모델링/3D프린팅이 가능하게 해준다[그림6].

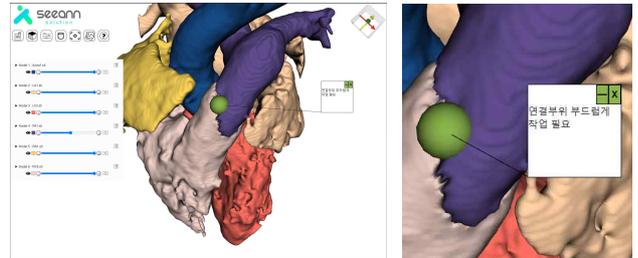


그림 6. 소아심장 모델링 a)사용 예, b)메모 사용 예

### 3.2. 동물 임플란트 모델링

골절된 동물의 맞춤형 수술용 임플란트 설계 사례이다 [그림3]. 제작된 3차원 모델에서 의료인이 Collabo를 통해 직접 스크류의 위치를 조절하여 수정이 필요할지 여부를 판단할 수 있었고, 이미지 메모를 통해 말로 전달하기 어려운 정확한 위치 및 방향에 관한 요구사항을 전달 가능하다.

## 4. 결론

모델러와 의료인간의 Cloud 기반 협업솔루션 Collabo SW의 활용으로 맞춤형 임플란트, 수술 가이드, 보조기 제작 및 설계에서 의료인과 모델러간의 의견 교환이 원활히 이루어져 기존 협업절차방식보다 제작 시간이 현저히 단축되었으며, 설계의 정확성 또한 증대되었다.

기존의 플랫폼 종속적인 협업 환경으로 접근이 불편했던 점 또한 PC/모바일 웹 브라우저에서 바로 3차원 모델을 공유하고 의견을 남길 수 있고, 매우 직관적인 인터페이스 사용으로 인한 편리성에 관한 의견이 많았다. 또, Collabo가 구동되는 환경의 하드웨어 자원을 거의 사용하지 않는 cloud web 서비스 형태로 인해 3차원 모델의 시각화와 움직임이 매우 가벼워서 사용자의 만족도가 높았다[3].

향후 3D모델과 2차원 의료영상 퓨전 기능과 메쉬 수정 기능의 추가로 dicom viewer와 협업 SW를 따로 사용하지 않고 하나의 화면에서 서비스하여 임상적 유용성을 높일 것이다.

## 참고문헌

- [1] 박준홍, 이준상. (2020). 3D 프린팅 환경에서 개인 맞춤형 컨트롤러 제품디자인 구매 의도의 영향요인에 관한 연구. *한국정보통신학회논문지*, 24(7), 873-878.
- [2] Bartosz Sawicki, Bartosz Chaber, (2013), Efficient visualization of 3D models by web browser, *Computing volume 95*, 661-673
- [3] 나중희. (2011). 클라우드 컴퓨팅의 서비스 특성에 관한 질적연구. *한국디지털콘텐츠학회 논문지*, 12(3), 319-327.

# GPU를 이용한 삼각형 집합의 근사적 외부 경계 계산

최규진<sup>0</sup>, 경민호  
 아주대학교 라이프미디어협동과정  
 {paganinist, kyung}@ajou.ac.kr

## Fast Outer Boundary Approximation of Triangle Set on GPU

Gyu-Jin Choi<sup>0</sup>, Min-Ho Kyung  
 Lifemedia Interdisciplinary Program, Ajou University

### 요약

본 연구에서는 삼각형 집합의 외부 경계를 GPU를 이용하여 다각형 메시로 빠르게 계산하는 볼륨 기반의 새로운 방법을 제안한다. 외부 경계 추출을 위해서는 삼각형 집합이 포함된 그리드 공간을 내·외부 볼륨으로 구분해야 한다. 그런데 볼륨 기반의 처리방식에서는 이 과정에 가장 많은 시간이 소요된다. 우리는 처리되는 그리드 공간을 삼각형 집합의 복셀 주변으로 한정하고, union-find 연산으로 내·외부 연결 요소(connected component)들을 찾았다. 알고리즘의 모든 과정들은 GPU에서 효율적으로 구현했고, 성능과 정확성을 확인하기 위하여 swept volume 계산에 적용하였다. 그리고 실험을 통해 기존 연구에 비해 높은 성능을 보이는 결과를 확인했다.

### 1. 서론

기하 알고리즘에서는 삼각형 집합의 외부 경계 표면을 최종 결과로 계산하는 경우가 자주 있다. 특히 삼각형 메시로 구성된 솔리드 모델의 swept volume, Minkowski sum, 그리고 surface offsetting 등은 삼각형 집합의 외부 경계를 구하는 계산 과정이 마지막 단계로 들어간다. 이처럼 다양한 응용 분야의 핵심 과정임에도 삼각형 집합의 외부 경계 표면을 구하는 것은 쉬운 문제가 아니다. 삼각형 집합의 외부 경계를 구하는 방법은 크게 삼각형 간의 교차 계산에 기반한 방식과 삼각형 집합을 볼륨 그리드로 이산화하여 근사적인 경계를 구하는 방식이 있다. 우리는 기존의 볼륨 기반 방법에서 많은 시간이 소요되던 내·외부 볼륨 분할을 효율적으로 처리하는 새로운 방법을 제안한다. 우리는 처리되는 그리드 공간을 삼각형 집합의 복셀 주변에 한정하고, label equivalence 방식으로 주변 복셀들의 내·외부 연결 요소들을 탐색했다. 이 방식은 복셀 데이터의 처리량을 획기적으로 줄이고, GPU의 성능을 효율적으로 사용할 수 있다. 우리는 제안한 알고리즘을 swept volume 계산에 적용하여 성능과 정확성을 측정했으며, 기존

연구에 비해 높은 성능을 보이는 결과를 확인했다.

### 3. 경계 계산 알고리즘

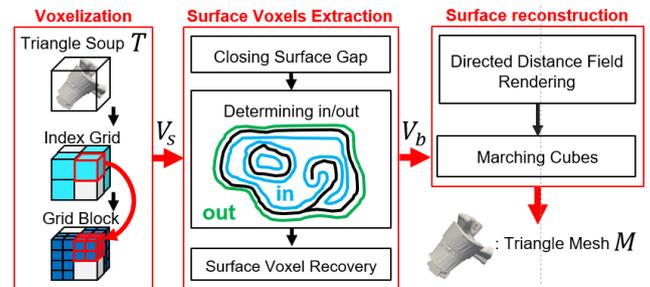


그림 1 : 경계 계산 알고리즘의 전체 진행 과정

입력 삼각형 집합  $T$ 로부터 경계 메시  $M$ 을 얻기 위해 우리가 제안하는 방법은 크게 세 단계로 구성되어 있다 (그림 1). 첫 번째로,  $T$ 를 복셀화하여 복셀 집합  $V_s$ 를 얻는다. 두 번째로,  $V_s$ 를 확장하여 주변 복셀 집합을 생성하고, 여기서 내·외부의 연결 요소들을 label equivalence 방법으로 찾는다. 이 내·외부 연결 요소들을 다시 확장하여 그사이에 존재하는 외부 경계 복셀 레이어  $V_b$ 를 추출한다. 마지막으로,  $V_b$ 의 복셀들에 대하여 directed distance를 구하고, 이 값들을 이용하여 외부 경계 표면을 메시 곡면  $M$ 으로 복원한다.

#### 3.1. 삼각형 집합의 복셀화

이 단락에서는  $T$ 를 복셀화하여  $V_s$ 를 얻는 과정을 서술한다. 볼륨 그리드의 과도한 메모리 사용량을 피하고자 복셀 데이터는 저해상도 인덱스 그리드와 고해상도 그리드 블록 배열로 저장한다(그림 1). 먼저  $T$ 를 저해상도의 볼륨으로 복셀화하고, 이 복셀들을 수집하여 고해상도 그리드 블록에서  $T$ 가 지나가는 블록들을 결정한다. 복셀화를 위해서 OpenGL의 fragment shader를 사용하였으며, 카메라의 시점을  $T$ 의 바운딩 볼륨 중앙으로 설정하고  $z$ 축 방향으로 직교 투영했다. 이때, 앨리어싱에 의한 복셀의 손실을 피하고자 삼각형의 면이 투영 방향을 바라보도록  $x$ 축 또는  $y$ 축으로 90도 회전한다.

\* 구두발표논문

### 3.2. 경계 복셀 레이어 계산

복셀 집합  $V_s$ 로부터 외부 경계 복셀 레이어  $V_b \subset V_s$ 를 추출하는 과정은 네 단계로 진행된다. 첫 번째로,  $V_s$ 을  $3 \times 3 \times 3$  크기의 박스 필터로 팽창(dilation)하여 복셀 레이어  $L_{fat}$ 을 만든다. 만약 외부면을 이루는 삼각형 간의 간격이 너무 먼 경우, 래스터화 과정에서 외부 경계를 이루는 복셀들이 손실될 수 있다.  $L_{fat}$ 은 이런 손실된 영역을 복구하고,  $L_{fat} \cup V_s$ 이 전체 그리드 공간을 내·외부 볼륨으로 분할하기 위해 필요하다. 두 번째로,  $L_{fat} \cup V_s$ 을 다시  $3 \times 3 \times 3$  크기의 박스 필터로 팽창하여 추가적인 복셀 레이어  $L_{fringe}$ 를 생성한다.  $L_{fringe}$ 는 분할된 공간들과  $L_{fat} \cup V_s$  사이의 경계가 된다. 세 번째로, 연결 요소 라벨링 방법으로  $L_{fringe}$ 의 복셀들을 독립적인 연결 요소로 구분한다. 이 때, 각 연결 요소는 외부 혹은 내부에 속하는 레이어로 구분된다. 외부 레이어  $C_{out}$ 은 간단하게  $x$ 축의 좌표가 가장 작은 복셀을 가진 연결 요소로 결정되고, 나머지 연결 요소들은 내부 레이어  $C_{in}$ 으로 분리한다. 마지막 네 번째 단계에서는  $V_b$ 를 추출한다.  $T$ 의 외부 경계는  $L_{fat}$ 으로 채워진  $C_{out}$ 과  $C_{in}$ 의 사이에 놓여있다. 따라서  $V_b$ 에 들어가는 경계 복셀들은 외부  $C_{out}$ 을 팽창시켜 처음 만나는  $V_s$ 의 복셀들로 구해진다. 이 때 팽창된  $C_{out}$ 이 삼각형 사이의 틈새를 통해 내부로 흘러 들지 못하도록  $C_{in}$ 을 함께 팽창시킨다.

CCL 방법을 GPU에서 효율적으로 구현하기 위해서 Oliveira 등[1]이 제안한 union-find 기반의 GPU 알고리즘을 사용했다. 이 알고리즘은 각 복셀들이 고유의 라벨로부터 시작하여 비동기적으로 다른 라벨을 가진 이웃 복셀과 라벨을 병합하며 연결 요소를 구성해 간다.

### 3.3. 경계 메쉬 복원

우리는  $V_b$ 로부터 경계 메쉬를 복원하기 위해 Kobbelt 등[2]이 제안한 방향 거리장(directed distance field)을 이용한다. 방향 거리장은 일반적으로 ray casting을 사용하여 구할 수 있는데, 효율적인 GPU 알고리즘으로 구현하기가 어려운 단점이 있다. 우리는 대신에  $T$ 의 복셀화 과정을 변형하여  $V_b$ 에 포함된 복셀에 대해서만 방향 거리 값을 GPU 배열에 기록했다. Fragment shader에서 경계 복셀에 포함되는 fragment에 해당 복셀의 중심으로부터  $x, y, z$ 축 방향의 거리를 구하고, 그 순간 복셀에 기록된 값과 비교하여 최대, 최소 거리를 갱신한다. 마지막으로 이 값을 마칭 큐브 알고리즘[3]의 교차점으로 사용하여 근사적인 외부경계를 생성했다.

## 4. 실험 및 결과

실험을 위해 Intel i7-8700, Nvidia RTX 2060 Super 사양의 시스템을 사용했다. 우리는 제안한 알고리즘을 swept volume 경계를 구하는데 사용했다. Swept volume은 메쉬 모델의 각 모서리를 스융 경로를 따라

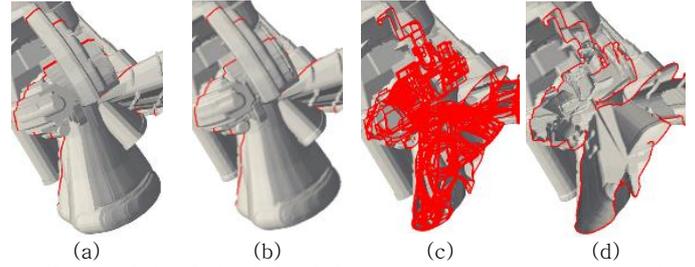


그림 2 : 외부 경계 계산 결과. (a)는 robot sweeping 모델, (b)는 외부 경계, 그리고 (c), (d)는 각각 (a), (b)의 단면임.

이동시켜 ruled surface을 만들고, 곡면을 삼각화하여 생성했다[4]. 표 1은 Index grid와 Grid block의 해상도가 각각  $128 \times 128 \times 32$ ,  $8 \times 8 \times 32$ 일 때의 알고리즘의 소요 시간이고, 그림 2는 계산된 표면의 시각적 차이와 절단면에 따른 내부 모습이다. 기존 연구들과 구현 방식, 그리고 실험 데이터가 달라 직접적인 성능 비교는 어렵지만, 동일한 해상도에서 von Ditzigielewski 등[5]이 372s가 소요되었던 것에 비해, 우리는 훨씬 많은 삼각형이 포함된 집합의 표면을 1s 이내에 계산하였다. [5]와 다르게 우리의 알고리즘에는 sweeping 과정이 포함되지 않았음을 감안하더라도 수십 배의 속도 향상이 있음을 알 수 있다.

표 1 : 삼각형 외부 경계 계산에 소요된 시간과 계산 오차

Input model	# of triangles	Elapsed time	Hausdorff distance
robot sweeping	8M	714ms	1.7624mm

## 5. 결론

우리는 GPU에서 삼각형 집합의 외부 경계를 빠르게 계산하는 볼륨 기반의 새로운 알고리즘을 제안하고 성능을 검증했다. 다음에는 제안한 알고리즘이 실시간에 가까운 속도로 동작하도록 최적화하고, 스융 볼륨 계산의 모든 과정을 GPU로 가속하는 연구에 적용할 것이다.

## 참고문헌

- [1] Oliveira, Victor MA, and Roberto A. Lotufo. "A study on connected components labeling algorithms using GPUs." *SIBGRAPI*. Vol. 3. 2010.
- [2] Kobbelt, Leif P., et al. "Feature sensitive surface extraction from volume data." *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 2001.
- [3] Lorensen, William E., and Harvey E. Cline. "Marching cubes: A high resolution 3D surface construction algorithm." *ACM siggraph computer graphics* 21.4 (1987): 163-169.
- [4] Abrams, Steven, and Peter K. Allen. "Computing swept volumes." *The Journal of Visualization and Computer Animation* 11.2 (2000): 69-82.
- [5] Von Ditzigielewski, Andreas, Michael Hemmer, and Elmar Schömer. "High precision conservative surface mesh generation for swept volumes." *IEEE Transactions on Automation Science and Engineering* 12.1 (2013): 183-191.

# 물리적인 상호작용을 위한 유연한 휴머노이드 제어와 시뮬레이션

이승환<sup>0</sup>, 이제희<sup>0</sup>  
서울대학교 컴퓨터공학부  
{lsw9021, jehee}@mrl.snu.ac.kr

## Simulation and Control of Compliant Humanoid for Physcal Interactions.

Seunghwan Lee<sup>0</sup>, Jehee Lee<sup>0</sup>  
Dept. of Computer Science and Engineering, Seoul National University

### 요약

사람은 주변환경과 물리적인 상호작용을 한다. 주변환경이 바뀌면 사람은 유연하게 반응하며 움직인다. 이 초록에서는 주변환경과 상호작용할 때에 유연하게 움직이는 동작 제어 시뮬레이션 프레임워크를 제안한다. 최근 강화학습에 기반한 모방 학습은 물리기반 캐릭터 애니메이션에서 효과적인 방법론으로 떠오르고 있다. 내부에 존재하는 미분-비례제어와 고정된 참조 모션의 사용은 제어 정책이 딱딱하게 학습되게 한다. 이 딱딱한 정책은 물리적인 상호작용에서 특히 치명적으로 작용한다. 우리는 적대적 모방학습과 변형가능한 참조모션 모델을 사용하여 유연하게 움직일 수 있는 프레임워크를 디자인하였다. 우리는 사람이 행동하는 다양한 상호작용에서 유연한 정책 모델이 효과적임을 증명하고자 한다.

### 1. 서론

사람은 매일 물리적인 상호작용을 한다. 문을 열고 닫기도 하고, 지면 반발력을 이용해 달린다. 사람과 주변환경은 힘을 통해 상호작용하는데 사람은 이 힘을 인지하고 사고 과정을 통해 행동한다. 주변 환경이 달라지면 그에 맞춰 움직임을 바꿔가며 유연하게 대처한다.

컴퓨터 그래픽스 분야에서는 물리 시뮬레이션을 통해 사람의 운동 능력을 학습하고 움직임을 재현하는 연구가 지난 몇 십년간 이어져왔다. 최근 강화학습(Deep Reinforcement Learning)은 효과적인 방법론으로 떠오르고 있다. 심층 신경망(Deep Neural Network)으로 구성된 정책(Policy)은 사람의 동작을 학습하고 가상 물리 환경상에서 재현할 수 있다. 핵심 아이디어는 모방 학습(Imitation Learning)이다. 모션 캡처로 얻은 참조 모션을 모방하도록 학습하는데, 참조 모션이 지시서 역할을 하며 학습 과정을 쉽게 한다. 하지만 참조 모션의 존재는 정책이 다른 동작을 학습하는 것을 막는 방어벽의 역할을 하기 때문에 주변 환경에 따라 유연하고 다양한 동작을 학습하는 것에 어려움이 있다.

이 초록에서는 물리적인 상호작용에서 유연한 동작 제어와 시뮬레이션 프레임워크를 제안한다. 우리는 참조 모션이 물리 환경에 따라 변형 가능하도록 디자인하였다. 적대적 모방 학습(Adversarial Imitation Learning)

을 하여 휴머노이드가 다양한 자세를 학습하게 하고, 스프링-댐퍼(Spring-damper) 모델을 차용하여 매 순간 상호작용하는 힘에 따라 참조 모션을 유연하게 변형한다. 우리의 프레임워크가 다음과 같은 물리적인 상호작용에서 유연함을 보여주고자 한다.

- 스프링-댐퍼 모델이 물체와 상호작용하는 힘을 최소화하며 상호작용할 수 있다.
- 미는 힘에 버티고 밸런스를 유지하기 위한 발목 전략(Ankle Strategy), 발디딤 전략(Stepping Strategy)을 재현할 수 있다.

### 2. 모방 학습(Imitation Learning)

우리의 프레임워크는 모방 학습을 기반으로 한다[1]. 모방하고자 하는 참조 모션을 입력으로 받아 정책이 시뮬레이션 상에서 참조 모션을 잘 따라가도록 학습한다. 휴머노이드는 모델의 자유도보다 작동 자유도가 적은 저구동(Under-actuated) 모델이다. 정책은 저구동 모델이 균형을 잃지 않고 참조 모션을 모방하기 위해 휴머노이드의 상태에 따라 적절히 참조 모션을 수정한다. 수정된 모션은 미분-비례 제어(PD Control)를 사용하여 따라간다.

우리는 정책을 학습하기 위해 강화학습을 이용한다. 강화학습 알고리즘은 현재 정책을 최적화하기 위해 수많은 시뮬레이션 데이터를 샘플링하는데, 미리 정의된 보상 함수(Reward Function)를 최대화 하는 방향으로 학습된다. 참조 모션을 모방하도록 하기 위해서 현재 휴머노이드의 자세가 참조 모션과 비슷할수록 높이 보상한다. 우리는 강화학습의 알고리즘 중 하나인 PPO(Proximal Policy Optimization)를 사용하여 정책을 학습하였다[2].

#### 2.1. 적대적 모방 학습(Adversarial Imitation Learning)

모방 학습은 참조 모션 하나를 모방하는 것에 효과적이다. 우리는 휴머노이드가 상황에 따라 유연하게 움직이기 위해서 다양한 자세를 학습하고자 한다. 그러기 위해서 구조화되지 않은 모션 데이터들을 이용한다. 모방 학습에 사용된 보상 함수는 지점간(Point-to-Point) 함수로 작동한다. 휴머노이드의 자세는 현재 정해진 참조 모션의 자세와 매칭되어 보상하기 때문에 휴머노이드가 다양한 자세를 학습하는 것을 제한한다. 우리는 적대적

신경망(Adversarial Network)을 이용하여 보상 함수를 디자인하였다[3]. 적대적 신경망은 모션 데이터에 있는 자세들과 휴머노이드의 자세를 구분하고자 학습하고, 휴머노이드의 현재 상태가 모션 데이터들과 구분이 되지 않을 때 높은 보상을 한다. 이 결과로 정책은 모션 데이터의 자세들을 모방하며 학습된다. 이 방법으로 모방 정책이 모션 데이터에 있는 다양한 자세를 학습할 수 있다.

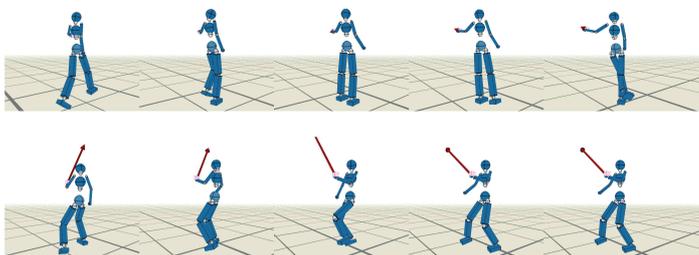
### 3. 변형가능한 참조 모션(Deformable Reference Motion)

이 섹션에서는 물리적인 상호작용에서 유연하게 움직이는 휴머노이드 모델링을 다룬다. 유연함이란 다양한 방식으로 정의된다. 물리적인 관점에서 우리는 외부 물체와 휴머노이드가 주고 받는 힘이 작을 때 유연하다고 정의하였다. 주고 받는 힘이 작으면 주변 환경이 달라져도 휴머노이드는 큰 어려움 없이 유연하게 대처할 수 있다.

적대적 모방 학습은 참조 모션을 고정시킨 상태로 학습이 진행된다. 주변 환경이 바뀌면 휴머노이드는 큰 힘을 내며 딱딱(Stiff)하게 저항한다. 이 문제를 해결하기 위한 방법으로 방대한 양의 모션 데이터를 이용할 수도 있지만 현실적으로 불가능하다. 대신에 우리는 참조 모션이 변형 가능하게 디자인하였다. 주변 환경의 변화에 따라 모방 학습 해야할 모션 데이터를 변형한다. 우리는 스프링-댐퍼 모델을 차용하여 모션 데이터를 변형하였다. 상호작용하는 힘이 관절마다 있는 스프링-댐퍼 모델을 통해 모션 데이터를 수정하고, 결과적으로 모방 학습에서 기존의 데이터보다 현재 주변 환경에 더 적절한 모션 데이터로 학습할 수 있다.

### 4. 결론

우리는 두 가지 상황을 재현하였다. 첫 번째는 손을 잡고 무작위로 끌었을 때 반응을 학습하였다[그림 1]. 변형가능한 참조 모션 모델을 사용하면 손과 물체에 걸리는 힘이 적은 반면 고정된 참조모션으로 학습하면 상호작용 힘이 커진다. 두 번째는 캐릭터가 밸런스를 유지하고 있는 상태에서 힘을 주어 밀었을 때 캐릭터의 반응을 학습한 것이다[그림 2]. 캐릭터가 맞았을 때는 최대한 발목 전략을 이용해 몸을 뒤로 제친다. 외력이 커서 발목 전략으로 버틸 수 없을 때에는 발디딤 전략으로 전환하여 밸런스를 유지한다.



[그림 1] 위 : 변형가능한 참조 모션 모델 아래 : 고정된 참조 모션. 손을 끌었을 때 반응을 학습한 것이다. 화살표는 손에 걸리는 힘을 보여준다.



[그림 2] : 외부에서 미는 힘이 들어왔을 때 반응.

### 4. 결론

최근 모방 학습은 물리에 기반한 애니메이션 분야에서 탁월한 성능을 보여주고 있다. 내부에 존재하는 미분-비례 제어, 고정된 참조 모션은 정책이 딱딱하게 학습하게 된다. 이 딱딱한 정책은 물리적인 상호작용에서 치명적으로 작용한다. 이 초록에서는 물리적인 상호작용에서 유연하게 움직이기 위해 변형가능한 참조 모션 모델을 제안하였다.

이 초록에서 다루지 못한 향후 작업이 있다. 우리는 관절마다 스프링-댐퍼 모델을 임베드하여 모션을 변형하였다. 사람은 관절마다 따로 움직이지 않고 주위 관절을 고려하여 조화롭게 움직인다. 모든 관절을 고려하여 한번에 모션을 변형하는 방법론을 고려해야 한다[4]. 사람의 움직임은 이방성(Anisotropic)을 가진다. 이방성을 가지는 변형 모델을 도입하는 것도 흥미있는 연구 주제이다.

상호작용 힘에 따라 얼마나 모션을 변형할 지는 변형가능한 참조 모션의 파라미터에 의해 결정된다. 현재는 이 파라미터를 적절히 수동으로 조절하였다. 사람과 비슷한 파라미터를 찾는 일 역시 흥미있는 연구 주제이다.

### 참고문헌

[1] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Transactions on Graphics 37, 4, Article 143

[2] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. CoRR abs/1707.06347

[3] 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 40, 4, Article 1

[4] Yifeng Jiang, and C. Karen Liu, 2018. Data-Driven Approach to Simulating Realistic Human Joint Constraints IEEE International Conference on Robotics and Automation (ICRA)

# 상태 표현 방식에 따른 심층 강화 학습 기반 캐릭터 제어기의 학습 성능 비교\*

손채준<sup>0</sup>, 이윤상

한양대학교

{thscowns, yoonsanglee}@hanyang.ac.kr

## Comparison of learning performance of character controller based on deep reinforcement learning according to state representation

Chaejun Sohn, Yoonsang Lee

Hanyang University

### 요약

물리 시뮬레이션 기반의 캐릭터 동작 제어 문제를 강화 학습을 이용하여 해결해 나가는 연구들이 계속해서 진행되고 있다. 강화학습을 사용하여 문제를 풀기 위해서는 네트워크 구조, 하이퍼파라미터 튜닝, 상태(state), 행동(action), 보상(reward)이 문제에 맞게 적절히 설정이 되어야 한다. 많은 연구에서 다양한 조합으로 상태, 행동, 보상을 정의하였고, 성공적으로 문제에 적용하였다. 상태, 행동, 보상을 정의함에 다양한 조합이 있다 보니 학습 성능을 향상시키는 최적의 조합을 찾기 위해서 각각의 요소들이 미치는 영향을 분석하는 연구도 진행되고 있다. 우리는 지금까지 이뤄지지 않았던 상태 표현 방식에 따른 강화학습에 미치는 영향을 분석하였다. 첫째로, root attached frame, root aligned frame, projected aligned frame 3가지로 좌표계를 정의하였고, 이에 대해 표현된 상태를 이용하여 강화학습에 미치는 영향을 분석하였다. 둘째로, 상태를 정의할 때, 관절의 위치, 각도로 다양하게 조합하는 경우에 학습성능에 어떠한 영향을 미치는지 분석하였다.

### 1. 서론

물리 시뮬레이션 기반의 사람 캐릭터 동작 제어는 상당히 어려운 문제로 여겨져 왔다. 그 이유 중 하나로 사람 캐릭터가 많은 관절을 가지고 있어 높은 자유도(degrees of freedom, DOFs)를 가지기 때문에 상태(state)의 차원이 크다는 것을 들 수 있다. 이러한 어려움을 효과적으로 극복할 수 있는 방법으로서, 최근에는 심층신경망 정책을 기반으로 하는 심층강화학습을 캐릭

터 동작 제어 문제에 적용하는 연구들이 많이 발표되고 있다. 강화학습을 사용하여 문제를 풀기 위해서는 네트워크 구조, 하이퍼파라미터 튜닝(hyperparameter tuning), 상태, 행동(action), 보상(reward)이 문제에 맞게 적절히 설정이 되어야 사람 캐릭터가 걷거나 뛰는 동작을 만들어 낼 수 있다.

강화학습이 다루는 문제의 기본적인 형태를 살펴보자면 매 스텝마다 환경은 상태와 보상을 제공하고 에이전트(agent)는 이 상태에 근거하여 최적의 행동을 취한다. 강화학습의 목표는 가장 높은 누적보상을 갖게 하는 행동을 출력하는 에이전트 정책(policy)을 찾는 것이다. 캐릭터 제어 문제에서 상태는 일반적으로 관절 위치(joint position) 또는 관절 각도(joint angle), 속도, 위상(phase) 등을 이용하여 표현된다. 행동은 PD제어(PD control)를 위한 관절 각도 또는 토크(torque)가 이용되며, 보상은 보통 참조 동작(reference motion)과 시뮬레이션된 캐릭터의 차이와 에너지 최소화 항 등으로 구성된다. 이러한 상태, 행동, 보상은 각 연구마다 조금씩 다른 조합으로 정의되고 있다. 예를 들어 [1]에서는 관절 위치, 관절 속도, 각속도, 위상 정보들을 이용해서 상태를 정의하여 캐릭터 동작제어 문제를 풀었고, [2]에서는 관절 각도, 관절 속도, 말단부(end-effector) 위치 정보 및 미래의 참조 동작을 상태로 정의하여 문제를 성공적으로 풀었다. 이렇듯 다양한 조합으로 문제가 정의될 수 있다 보니 캐릭터 동작 제어 문제를 강화학습을 이용하여 풀 때 학습성능에 영향을 미치는 요소들을 알아볼 필요성이 있었다. 이에 [3]에서는 강화학습을 통해 학습시킨 정책의 성능에 행동의 설계(design)가 미치는 영향에 대해 분석하였고, [4]에서는 상태, 행동, 보상의 설계가 학습에 미치는 영향을 분석하였다. 상태의 경우, 같은 속성을 상태로 정의하더라도 어떤 좌표계에 대해 표현된 상태 값인지에 따라 그 특성이 달라질 수 있다. 예를 들면 평면에 투영된 좌표계에 의해 표현된 상태 값은 높이에 대한 정보가 포함되므로, 투영되지 않은 좌표계에 의해 표현된 상태와 특성이 다를 수 있다. 그러나 기존 연구들은 상태를 표현하기 위한 좌표계의 선택에 집중하여 분석하지는 않았다. 이 연구

\* 구두 발표논문, 요약 논문(Extended Abstract)

\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2019R1C1C1006778, NRF 2019R1A4A1029800).

에서는 다양한 좌표계에서 표현된 상태가 강화학습을 통한 문제 해결에 어떠한 영향을 미치는지 알아볼 것이다. 또한, 캐릭터의 상태를 나타내기 위해 관절 위치 혹은 관절 각도가 사용될 수 있다. 이에 [4]에서 상태를 관절 위치, 각도를 이용하여 정의할 때 학습에 미치는 영향을 보여주었다. 그러나 이는 두가지 모두를 사용한 것과 각도만 사용하였을 경우 학습성능을 비교하였고, 참조 동작을 따라하도록 하는 문제에 대해서는 학습 성능을 비교하지 않았다. 따라서 우리는 참조동작을 따라하도록 하는 캐릭터 제어문제를 강화학습을 이용하여 풀 때, 관절의 위치와 각도를 다양하게 조합하는 경우의 학습 성능을 분석할 것이다. 이를 통해 우리는 물리 시뮬레이션 기반의 사람 캐릭터 동작 제어 문제를 강화학습을 이용하여 풀 때, 상태를 정의하는 가장 좋은 방법 및 상태를 표현하는 좌표계의 선택에 중요한 요소가 무엇인지 보이고자 한다.

## 2. 관련연구

물리 시뮬레이션 기반의 캐릭터 제어 문제에 강화학습을 적용하는 연구는 계속해서 늘어나고 있다. 효과적으로 강화학습을 이 문제에 적용하기 위해서는 상태, 행동, 보상을 적절히 설계해야 한다.

상태는 크게 캐릭터의 동작을 나타내는 정보인 동역학적 상태(dynamic state), 캐릭터가 따라 가야할 동작에 대한 정보인 추적 정보(tracking information), 캐릭터가 움직이는 방향, 속도 등을 설정하기 위한 사용자 지정 정보(user-specified goal), 추가 정보(additional information)로 나눌 수 있다. 첫째로, 동역학적 상태를 정의하기 위해 [1]에서는 관절 위치, 관절 각도, 속도, 각속도를 이용하여 정의하였다. [2]에서는 관절 각도, 속도, 말단부(end effector) 정보를 이용하여 정의하였다. [5]에서는 시뮬레이션 캐릭터의 속도, 무게중심의 속도를 이용하였다. [6]에서는 속도, 상체의 기울기(up-vector of the trunk), 말단부 위치, root의 높이가 사용되었다. 둘째로, 추적 정보를 [1]에서는 위상 정보를 이용하였고, [2]에서는 현재, 미래의 동작을 이용하여 추적정보를 구성하였다. [5]에서는 참조 동작의 위치, 속도와 시뮬레이션 되는 캐릭터와의 위치, 속도의 차이를 이용하였다. 사용자 지정 정보는 어떠한 일을 수행하는가에 따라 다르게 구성되기도 한다. 예를 들어, [1]에서 사용된 목표 방향(target direction), 무게중심의 속도가 사용되기도 하고, [2], [5]처럼 root의 목표(desired) 속도 정보를 이용하여 문제를 풀기도 하였다. 그 외의 추가 정보로는 [2]에서 사용한 캐릭터 크기 변수(body shape parameter), [5]에서 이전 스텝의 에이전트가 선택한 행동을 이용한 경우가 있었다.

행동으로는 일반적으로 PD 제어 목표 자세, 토크, 근육 활성화 정도 등이 사용된다. [1]에서는 목표 자세를 사용하여 행동을 정의하였다. [2], [5], [6]에서는 목표 자세를 구성하기 위한 참조동작으로부터의 차이(offset)를 이용하여 정의하고 있다. [7]에서는 근육을 통해 움직이는 캐릭터를 모델링하였고, 근육의 활성화정도를 행동으

로 이용하여 캐릭터의 움직임을 생성하였다.

보상의 구성요소는 크게 3가지로 나눌 수 있었다. 참조 동작과 캐릭터가 비슷하도록 하는 모방 항(imitation term), 특정 일을 수행하도록 하는 수행 항(task term), 최소한의 에너지를 사용하도록 하는 에너지 항(energy term)으로 구성할 수 있다. 뿐만 아니라, 이러한 항들을 어떻게 조합할지에 대한 방법들이 제안되어왔다. [1]와 [2]에서는 모방항을 정의할 때, 참조동작과 캐릭터의 관절 각도, 관절의 속도, 말단부 위치를 이용하였다. 에너지 항은 [2]에서 사용되었고, 가능하면 최소한의 에너지를 사용하도록 캐릭터를 제어할 수 있게 한다. 마지막으로 이렇게 정의된 항들을 조합하여 최종보상으로 구성하는 방법에는 [1],[5],[2]에서 사용된 항들을 더하는 방식과 [6]에서 사용된 모든 항들을 곱하는 방식이 존재한다. [6]에서는 최종 보상을 곱셈으로 정의하는 것이 모든 항에서 높은 보상을 가져야 높은 누적보상을 가질 수 있기 때문에 더하는 방식보다 더 좋은 성능을 가질 수 있다고 한다. 이렇게 물리 시뮬레이션 기반의 사람 캐릭터 동작 제어 문제를 강화학습을 이용하여 풀기 위해 고려해야할 사항들이 많이 존재한다.

많은 연구에서 강화학습을 이용하게 되면서 강화학습의 성능에 영향을 미치는 요소들의 적용 여부에 따른 결과를 비교 분석하는 연구들도 많이 수행되었다. [8]에서는 강화학습에 영향을 미치는 요소들을 코드수준 최적화 기법(code-level optimization)이라고 정의하고 알고리즘 별로 이러한 최적화 요소의 존재 여부에 따른 성능을 비교 분석한다. 또한, 문제의 환경에 따른 영향을 분석한 연구들도 존재하였다. [3] 연구에서는 행동을 어떻게 정의하는 지에 따른 강화학습의 성능을 분석하고 있다. 또, [4]에서는 환경을 정의하는 방식과 그 외의 다른 요소(episode termination, control frequency)에 따른 학습된 정책의 성능을 분석하였다. 환경의 상태를 다양하게 정의해서 분석하는 연구들이 존재했지만, 상태의 표현방식에 따른 성능을 분석하는 연구는 부족했고, 이 논문에서 상태의 표현방식에 따른 학습된 정책의 성능을 분석할 것이다.

## 3. 방법

우리는 물리 시뮬레이션 기반의 캐릭터를 참조동작을 따라하도록 제어하는 문제를 강화학습을 이용하여 풀 때, 상태를 여러가지 다른 방식으로 정의하고 이것이 학습에 어떠한 영향을 미치는지 확인하였다. 이를 위해 우리는 [1]에서 제안한 강화학습 기반 캐릭터 제어 프레임워크의 구조를 사용하였다. 우리는 [1]에서 정의된 것처럼 관절 위치, 관절 각도, 속도, 각속도, 위상, root의 높이를 상태로 사용하였다. 첫째로, 우리는 상태가 서로 다른 좌표계에 대해 표현되었을 경우에 학습 성능을 비교하였다. 둘째로, 상태로 관절 위치와 각도의 다양한 조합을 사용하는 경우의 학습 성능을 비교하였다. 3.1에서는 상태를 정의하기 위한 3가지 다른 좌표계와 좌표계에 대해 표현되는 상태를 설명할 것이다. 3.2에서는 관절의 위치 및 관절 각도가 상태에 포함되는지 여

부에 따른 실험을 설명할 것이다.

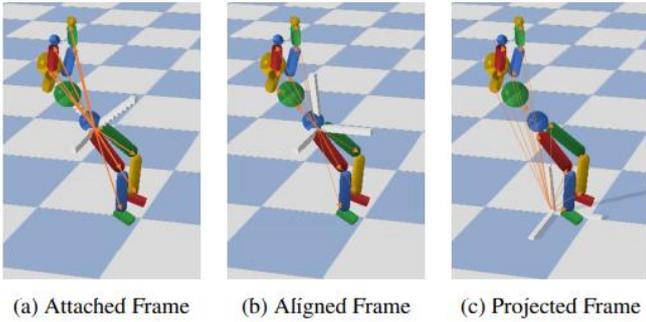


Figure 1: The three reference frames used in the experiments. The orange arrows depict joint positions represented in each frame.

### 3.1. 상태를 나타내는 좌표계

어떤 좌표계에 대해 표현된 상태 값인지에 따라 특성이 다르기 때문에 우리는 총 3가지의 좌표계를 이용하여 상태를 다르게 표현하여 사용하였다. Figure 1에 표현한 것처럼, (1) Root attached frame, (2) Root aligned frame, (3) Projected root aligned frame 세 가지 좌표계를 이용하였다.

Root attached frame은 root의 로컬 좌표계를 의미한다. Root aligned frame은 상체가 바라보는 방향과 y축 및 이 두가지를 외적 한 것을 각 축으로 가지는 좌표계를 의미하고 위치는 root의 위치로 정의한다. Projected root aligned frame은 root aligned frame의 원점을 xz 평면에 projection 시킨 좌표계이다. Figure 1에서처럼 우리는 관절 위치 뿐 아니라 관절 각도, 속도, 각속도를 각각의 좌표계에 대해서 표현하여 상태를 정의하였다. 우리는 앞서 설명한 총 3가지의 좌표계를 이용하여 관절 위치, 관절 각도, 속도, 각속도를 각각의 좌표계에 대해서 표현하여 상태를 정의하였다.

### 3.2. 관절의 위치 혹은 각도의 조합

캐릭터의 위치 상태를 나타내기 위해 관절 위치, 관절 각도 중 하나만 알아도 된다. 이전 연구들에서 상태를 정의할 때 관절 위치, 관절 각도를 모두 사용한 연구 [1]도 존재하였고, 둘 중 하나만 사용한 연구 [2]도 존재했다. 이러한 경우들 간의 성능을 비교하는 연구 [4]도 존재하였으나 관절 위치, 각도 모두 사용한 것과 관절 각도를 사용한 것의 비교였고, 참조 동작을 추적하는 문제가 아닌 사람이 걷는 문제를 강화학습을 이용해 학습하는 경우 성능을 비교하였다. 우리는 관절 위치만 사용한 것, 관절 각도만 사용한 것, 관절 위치, 각도 모두 사용한 것 세가지로 나누어서 참조 동작을 추적하는 문제의 강화학습을 진행하여 결과를 비교했다. 이 실험에 대해서 모든 상태는 root aligned frame에 의해 표현된다.

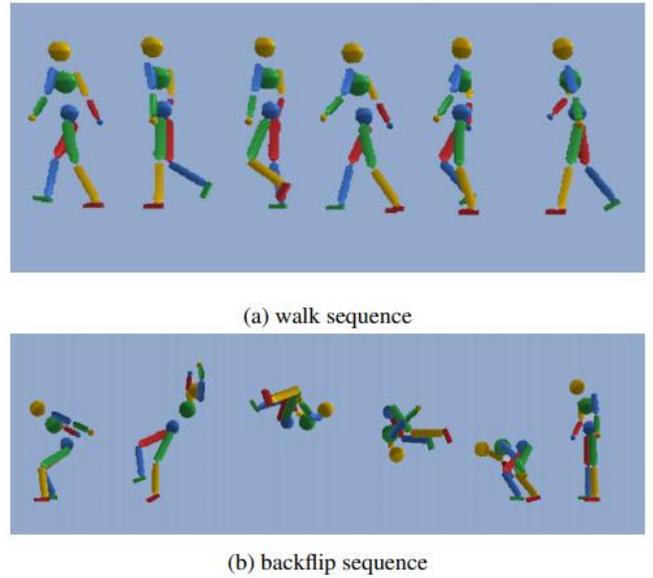


Figure 2: Reference motion sequence for each environment

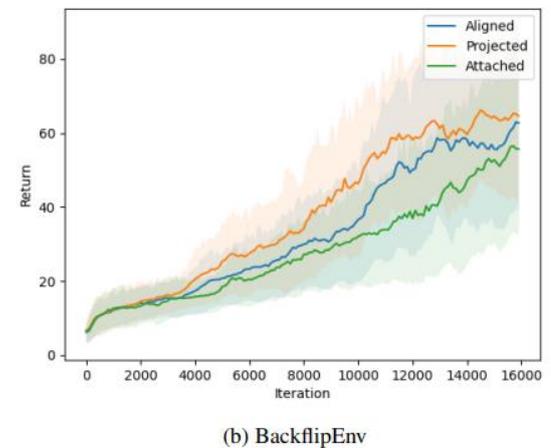
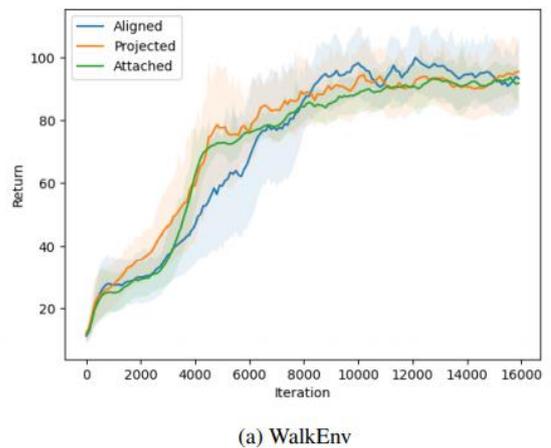


Figure 3: This is a learning graph when the state is represented by 3 frames(root aligned, projected root aligned, root attached).

#### 4. 결과

시물레이션은 bullet physics[9]를 이용하였다. 우리는 pybullet에 정의되어 있는 deepmimic 환경을 이용하였다. 기본적으로 [1]에서 사용된 것처럼 aligned frame에 대해 표현된 상태를 사용하였고, PPO를 이용하여 학습을 진행하였다. hyperparameter는 [1]에서 사용한 값 그대로 사용하였고, HumanoidDeepMimicWalkBulletEnv(이하 WalkEnv)와 HumanoidDeepMimicBackflipBulletEnv(이하 BackflipEnv) 환경을 학습시켰다. 각 환경마다 16000 iteration(65,000,000 sample)씩 학습을 진행하였고, 100 iteration 마다 네트워크 파라미터를 저장하였다. 저장된 네트워크마다 50번씩 에피소드를 시물레이션 하여 누적보상의 평균과 표준편차를 계산하였고, 이를 이용하여 그래프를 만들었다(Figure 3,4). 그래프에서 실선은 누적보상의 평균을 의미하고, 음영은 각 시점의 평균에 해당 시점의 표준편차를 더한 값과 뺀 값 사이의 영역으로 각 시점에서 나타날 수 있는 누적보상의 범위를 의미한다.

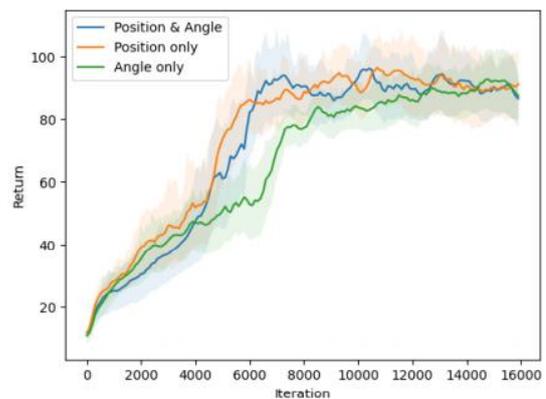
##### 4.1. 상태를 나타내는 좌표계에 따른 성능 비교

WalkEnv, BackflipEnv에 대해 상태를 각각 attached frame, aligned frame, projected aligned frame에 대해 표현하여 학습을 진행하였다. Figure 3에서 알 수 있듯이 WalkEnv에 대해서는 상태를 각 프레임에 따라 표현하는 것이 크게 의미가 있지는 않았다. 그러나, BackflipEnv에서는 상태를 각각 다른 좌표계에 대해서 표현하는 것이 큰 영향을 미치는 것을 확인할 수 있었다. 이유에 대해 분석하기에 앞서, 학습한 환경들의 참조동작(Figure 2)을 살펴보면 WalkEnv의 경우 root의 높이는 거의 변하지 않고, 회전이 큰 동작이 없으므로 나아가기만 한다. 반면 BackflipEnv의 경우 root의 높이가 계속해서 변하고, 회전이 큰 동작이다. 이러한 환경의 특성 차이로 인해 Figure 3a와 3b에서 나타난 경향성에 차이가 난다.

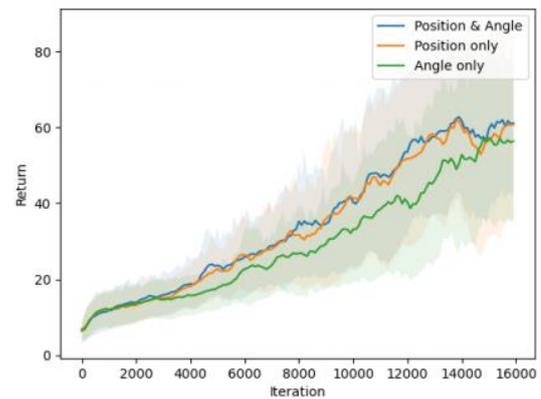
Figure 3b를 보면 BackflipEnv를 학습시킬 때 projected frame에 의해 표현된 상태를 이용하여 학습시킨 결과가 다른 좌표계를 이용하였을 경우보다 더 좋은 성능을 보여준다. projected frame에 의해 표현되는 관절의 위치 정보에는 나머지 좌표계에 의해 표현되는 상태에 비해 root의 높이가 더해져서 상태로 정의된다. 그렇기에 root의 높이 정보가 계속해서 달라지는 BackflipEnv의 경우 projected frame에 대해 상태를 표현하였을 때, 캐릭터의 동역학적 상태를 가장 잘 표현하고, 결과적으로 학습도 가장 잘 되는 것으로 볼 수 있다. 또한 Figure 3b에서 보면 상태를 aligned frame에 대해 표현하는 경우가 attached frame에 대해 표현하는 경우보다 학습성능에 미치는 영향이 큰 것을 알 수 있다. 이는 aligned frame 기준으로 표현된 상태가 attached frame에 의해 표현된 상태보다 회전에 대한 정보를 더 잘 표현하기 때문으로 볼 수 있다. 예를 들면, 점프의

중간단계에서 상태를 aligned frame에 대해 표현한다면 aligned frame에서 y축은 고정이 되어있기 때문에 몸이 기울어진 정도를 바로 알 수 있지만, attached frame은 캐릭터가 회전을 하면 좌표계도 같이 회전을 하게 되고 따라서 attached frame에 대해 표현된 상태는 그만큼 회전에 대한 정보가 부족하다고 할 수 있다. 이러한 이유로 aligned frame에 의해 정의된 상태를 이용하여 학습시킨 결과가 더 좋은 성능을 보여주는 것으로 추측된다.

이를 통해 최적의 학습을 위한 상태 정의를 위한 좌표계의 선택은 참조동작이 가지는 특성에 따라 달라져야 한다는 것을 알 수 있다. 즉, 높이의 변화가 큰 부분이 많은 참조 동작에 대해서는 projected frame을 사용하는 것이 유리하고, 몸 전체의 회전 동작이 많은 참조 동작에 대해서는 projected frame 혹은 aligned frame을 사용하는 것이 유리하다고 볼 수 있다.



(a) WalkEnv



(b) BackflipEnv

Figure 4: This is a learning graph when the dynamic state is defined as Position only or Angle only or Position and Angle

##### 4.2. 관절의 위치 혹은 각도의 조합에 따른 성능 비교

WalkEnv, BackflipEnv에 대해 각각 관절 위치만 사용한 경우, 관절 각도만 사용한 경우, 관절 위치, 각도 모두 사용한 경우로 상태를 정의하여 학습을 진행하였다. Figure 4에서와 같이, WalkEnv와 BackflipEnv에서 모

두 최종적으로는 비슷한 수준의 누적보상을 보여주었다. 두 환경 모두 위치 정보를 사용하여 학습한 경우가 각도만 사용한 경우보다 빠르게 최종수준에 도달하였다. 이를 통해 환경의 참조 동작이 갖는 특성, 즉 동작의 회전과 root 높이 변화에 상관없이 캐릭터 관절의 위치정보가 학습에 더 큰 영향을 주는 것을 알 수 있었다. 강화학습에서 에이전트가 캐릭터의 상태를 학습하여 가장 높은 누적보상을 갖도록 행동을 만들어내는데, 이 때 캐릭터의 위치정보가 에이전트가 학습을 더 쉽게 할 수 있도록 만드는 것으로 여겨진다

## 5. 결론

물리 시뮬레이션 기반의 캐릭터 제어 문제에 강화학습을 적용하는 시도들이 계속해서 이루어지고 있다. 최근에는 참조동작을 따라하도록 환경을 구성하고 이를 강화학습을 이용하여 푸는 연구들이 주를 이룬다. 이에 우리는 이러한 문제 정의속에서 상태를 표현하는 방식이 강화학습 에이전트의 학습에 미치는 영향에 대해서 알아보았다. 실험 결과로부터 참조동작이 가지는 특성에 따라서 상태를 정의하는 좌표계를 선택해야 학습 속도 및 성능을 높일 수 있다는 것을 알 수 있다. 또한, 상태를 정의함에 관절의 위치 정보와 각도가 학습하는 데에 미치는 영향을 알아보았다. 이 연구를 통해 상태 표현 방식이 강화학습에 미치는 영향을 알아볼 수 있었다. 이 연구에서는 두 가지 참조동작만을 이용하여 실험을 진행하였으나, 향후 더 많은 참조동작들에 대해 분석을 수행할 계획이다. 또한, 이 연구에서는 학습 속도 측면에서 가장 유리한 참조동작 별 상태 표현 좌표계의 선택에 대해 정성적인 분석만을 수행하였으나, 향후 이를 정량적으로 분석할 수 있는 방법에 대한 연구를 진행할 계획이다.

## 참고문헌

- [1] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "Deepmimic," ACM Transactions on Graphics, vol. 37, no. 4, p. 1–14, Aug 2018. [Online]. Available:
- [2] J. Won and J. Lee, "Learning body shape variation in physics-based characters," ACM Trans. Graph., vol. 38, no. 6, Nov. 2019. [Online]. Available: <https://doi.org/10.1145/3355089.3356499>
- [3] X. B. Peng and M. van de Panne, "Learning locomotion skills using deepri," Proceedings of the ACM SIGGRAPH / Euro graphics Symposium on Computer Animation, Jul 2017. [Online]. Available: <http://dx.doi.org/10.1145/3099564.3099567>
- [4] D. Reda, T. Tao, and M. van de Panne, "Learning to locomote: Understanding how environment design matters for deep reinforcement learning," Motion, Interaction and Games, Oct 2020. [Online]. Available: <http://dx.doi.org/10.1145/3424636.3426907>

[5] K. Bergamin, S. Clavet, D. Holden, and J. R. Forbes, "Drecon: Data-driven responsive control of physics-based characters," ACM Trans. Graph., vol. 38, no. 6, Nov. 2019. [Online]. Available: <https://doi.org/10.1145/3355089.3356536>

[6] S. Park, H. Ryu, S. Lee, S. Lee, and J. Lee, "Learning predict-and-simulate policies from unorganized human motion data," ACM Trans. Graph., vol. 38, no. 6, 2019.

[7] S. Lee, M. Park, K. Lee, and J. Lee, "Scalable muscle-actuated human simulation and control," ACM Trans. Graph., vol. 38, no. 4, July 2019. [Online]. Available: <https://doi.org/10.1145/3306346.3322972>

[8] L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph, and A. Madry, "Implementation matters in deep policy gradients: A case study on ppo and trpo," 2020.

[9] "Bullet physics library," <https://pybullet.org/>, 2015

# 단일 보행 패턴으로부터 신체 조건 유추\*

민세희<sup>0</sup>, 이제희  
서울대학교

{sehee, jehee}@mrl.snu.ac.kr

## Inferring Body Conditions from a Single Gait Cycle

Sehee Min<sup>0</sup>, Jehee Lee  
Seoul National University



[그림 1] 보행 매니폴드. 걸음걸이 다양체. 신체조건과 짝지어지지 않은 임상 보행 데이터(빨강)와 임의의 신체조건을 통해 시뮬레이션된 결과 보행 데이터(파랑)

### 요약

걸음걸이는 인간의 신체 상태를 분석할 수 있는 효과적인 수단이다. 우리는 하나의 보행 사이클로부터 신체 조건을 얻어낼 수 있는 개인화된 보행 분석 프레임워크를 제안한다. 우리의 프레임워크는 고차원 시계열 자세 시퀀스에서 특징적인 임베딩을 얻어낼 수 있는 보행 매니폴드를 포함한다. 임의의 보행 주기에 영향을 주는 뼈, 근육, 관절 한계와 같은 신체 매개변수는 저차원 잠재 공간에서의 유사성을 측정함으로써 근사할 수 있다. 우리의 물리 기반 보행 시뮬레이터는 실제 보행을 캡처한 모션 캡처 데이터와 함께 보행 공간에 포함될 데이터를 생성한다.

### 1. 서론

키, 몸무게, 신체구조, 근력과 같은 신체적 특성뿐만 아니라 기분, 노화, 피로 등의 수많은 조건이 보행에 영향을 끼친다. 걸음걸이 패턴이 고유의 신체 특성을 반영하기 때문에 병증 진단, 개인 식별 등의 다양한 분야에서 보행 분석이 연구되어 왔다.

모션 캡처 혹은 카메라를 사용한 보행 분석 시스템은 보행을 수치화할 수 있으며 환자에게 부담이 없다는 장점이 부각되며 병원에도 진단 수단으로 자리 잡았다. 현재의 방식은 캡처 데이터의 모든 키네마틱 정보를 활용하는 대신 간단한 수작업 표지 (보행 주기 중 최대 무릎 굽힘 각도, 웅크림 정도)를 계산하는 것이다. 때문에

걸음걸이만으로 정확한 병증의 정도나 특정 질환(근육 구축, 뼈의 휘어짐 정도 등)을 보행 분석만으로 얻어내기에는 무리가 있다. 이는 보행 패턴과 신체 조건이 너무 고차원이고 이 둘 간의 관계를 짝지어 둔 데이터가 턱없이 부족하기 때문이다. 또한 보행 분석 임상 데이터의 경우, 질환이 심한 환자의 수술 전후 상태를 비교하기 위해 촬영되면서 편향되기도 한다.

본 논문에서는 개인화된 보행 분석을 위한 새로운 딥러닝 기반 프로임워크를 제안한다. 우리는 보행 패턴과 신체 조건 사이의 강한 상관관계를 알아내기 위해서 차원을 낮춰줄 수 있는 공간, 보행 매니폴드를 제안한다. 이 압축된 공간에서는 수천 차원(초당 30프레임 이상, 각 관절의 회전값 및 진행속도로 구성)의 보행 시퀀스가 16차원으로 표현되어 보행 패턴 사이의 유사도를 측정하는 것이 용이하다. 우리의 보행 매니폴드는 병원에서 캡처한 임상 데이터뿐만 아니라 물리 시뮬레이션을 통한 부트스트래핑을 사용하여 다양하면서도 균형 잡힌 보행 데이터베이스를 구축할 수 있다.

### 2. 접근 방식

이 논문의 목표는 단일 보행 사이클로부터 해당 신체 변수들을 얻어내는 것이다. 우리의 핵심 아이디어는 보행 분석과 보행 시뮬레이션이 역연산의 관계를 가진다는 것에 착안한다. 보행 분석의 경우 보행 패턴으로부터 신체 조건을 유추하는 것이라면, 보행 시뮬레이션은 다양한 신체 조건의 변화에 따른 걸음걸이를 재현하는 것

에 중점을 둔다.

우리는 고차원 변수의 표현력을 유지하면서도 차원을 낮출 수 있도록 Wasserstein Autoencoder[1]를 기반으로 하는 인공 신경망 구조를 사용한다. 보행 사이클은 시계열 자세 집합으로, 매 순간의 자세는 관절의 회전으로 표현된다. 이 논문에서 우리는 하지 근골격 장애를 가지는 임상 데이터와의 비교를 위해 11개의 관절이 있는 하체 캐릭터를 사용하여 시뮬레이션 했다. 보행 데이터는 왼발로 지면을 딛기 시작해 오른발로 지면과 접촉한 뒤 왼발이 다시 착지할 때까지를 30프레임으로 정규화한다. 정상 보행뿐만 아니라 임상 목적으로 수집된 뇌성마비 환자의 모션 캡처 데이터는 보행의 다양성을 실험하기 위해 포함되어 까치발, 웅크림, 팔자걸음 등 여러가지 패턴으로 이루어져 있다. 이 환자들은 골반 경사, 대퇴골 전방경사, 무릎 굽힘, 발목 구부러짐 등의 근골격계 질환을 가지고 있다. 해당 데이터는 모션 캡처된 보행으로만 이루어져 있으며 정확한 골격의 크기나 길이, 휘어짐이나 근육의 힘, 길이 등의 자세한 정보와 짝지어져 있지는 않다.

### 2.1. 보행 매니폴드

우리 네트워크는 압축-재구성 과정을 통해 보행 사이클에서 핵심 키네마틱 특징을 학습한다. 네트워크의 학습 과정에서 입력으로 주어진 걸음걸이와 압축-재구성을 거쳐 복원된 걸음걸이 사이의 차이를 최소화한다. 또한 고차원 원본 공간에서의 데이터 분포가 저차원 매니폴드에서도 유지될 수 있도록 하는 최대 평균 불일치(Maximum mean discrepancy) 값 역시 최소화한다. 우리는 정확성을 높이고 재구성 데이터가 평균값으로 무뎠다는 것을 방지하기 위해서 입력 데이터로부터 3차원 관절 위치를 계산하는 정역학 레이어를 추가한다. 학습된 네트워크를 사용해서 보행 동작간의 유사성을 측정하는 것이 가능하다.

### 2.2. 물리 시뮬레이션을 통한 보행 데이터베이스 구성

우리는 정형외과에서 뇌성마비 환자를 대상으로 하는 수술의 효과를 판단하기 위해 모션 캡처된 보행 데이터베이스를 사용한다. 이렇게 캡처된 임상 데이터는 수술의 환자 통제 상태를 평가하기 위해서 중증 환자를 대상으로 캡처하기 때문에 편향되어 있다. 임상 데이터만으로 보행 매니폴드를 구축할 경우 데이터의 분포에 따라 심하게 편향되어 압축-재구성의 성능이 떨어지게 되고 보행 매니폴드 상에서 표현 가능한 걸음걸이가 제한된다는 문제가 있다. 우리의 알고리즘은 이 문제를 해결하기 위해서 역으로 신체 조건의 변화를 가상의 물리 환경에서 연산하여 해당 인체의 보행을 시뮬레이션하는 방법에 착안한다. 프레임워크에는 하반신 172개의 근육을 가지는 근육 구동 인체 모델 중 하지 근육 구축, 관절 운동 범위 제한, 골반 기울기 정도 및 골변형 등의 신체 매개 변수의 변화에 따라 보행 주기를 시뮬레이션할 수 있는 시뮬레이터가 포함되어 있다. 그 중 10개의

주요 근육 (*Psoas Major*, *Biceps Femoris*, *Rectus Femoris*, *Soleus*, *Tibialis*)은 근육의 구축 범위를 재현할 수 있는 매개변수로 포함되어 있다. 총 신체 파라미터의 개수는 16개다.

### 3. 결과 및 추후 연구

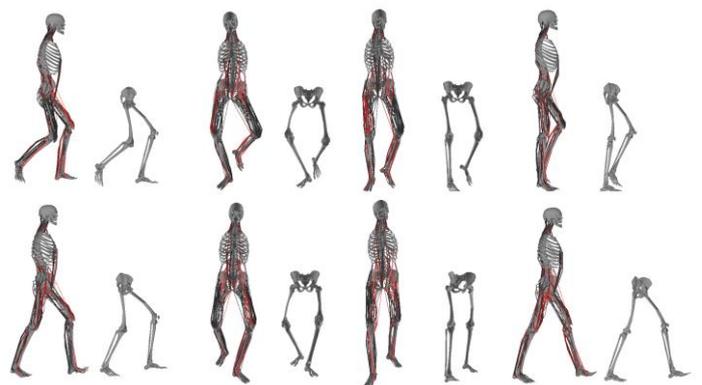
보행 매니폴드를 학습하기 위해 마련한 보행 데이터 셋은 물리 시뮬레이션을 통해 생성된 20,000개의 데이터와 임상 목적 데이터 200개가 혼합되어 있다. 데이터 셋이 형성하는 보행 공간은 UMAP[2]을 사용하여 비선형성을 유지하면서 2차원에 시각화될 수 있다.

그림 1은 보행 매니폴드 상에서 가까운 거리에 있는 데이터들이 복원했을 경우에도 유사한 보행 패턴을 지니며, 이를 통해 가까운 데이터의 신체조건으로부터 해당 신체의 조건을 유추해낼 수 있음을 보인다.

우리는 보행 매니폴드가 고차원 원본 보행 공간을 대체할 수 있음을 보이기 위해서 보행 보간을 실험했다. 보행 매니폴드에서 두개의 다른 보행 사이클을 선형 보간한 결과 합리적인 걸음걸이가 생성됐다. 네트워크를 일반적인 VAE와 비교했을 경우 우리의 네트워크는 각 관절 당 0.7cm의 유사도를 보였으며, VAE의 경우 7.1cm로 거리가 있었다.

본 연구의 미래에는 하지의 근골격계 질환뿐만 아니라 더 많은 병리학적 증상들을 분석할 수 있도록 시뮬레이션에 대한 매개변수의 종류를 늘리는 것이 있다. 근골격계 뿐만 아니라 신경계, 성장, 노화의 매커니즘 등을 포함하여 모델링 및 시뮬레이션할 수 있다면 다양한 질병을 발견할 가능성이 높아진다.

또한



[그림 2] 보행 주기 및 유추된 신체 조건을 이용한 시뮬레이션 결과와의 비교.

### 참고문헌

- [1] Tolstikhin, I., Bousquet, O., Gelly, S., & Schoelkopf, B. (2017). Wasserstein auto-encoders. arXiv preprint arXiv:1711.01558.
- [2] McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

# 2D 이동 경로 지정을 통한 모션 매칭 기반 캐릭터 모션 생성\*

이정민<sup>0</sup>, 이윤상  
 한양대학교 컴퓨터소프트웨어학과  
 {j0064423, yoonsanglee}@hanyang.ac.kr

## Motion generation based on motion matching with 2-dimensional trajectories

Jeongmin Lee<sup>0</sup>, Yoonsang Lee  
 Department of Computer Science, Hanyang University

### 요약

사람의 동작을 합성하는 기술은 게임, 애니메이션 등 다양하게 응용될 수 있으며 그동안 이 문제를 해결하기 위해 다양한 방법들이 제안되어왔다. 그 중 모션 매칭[1]은 모션 캡처 데이터로부터 특성 데이터베이스를 만들어 키보드 또는 조이스틱 입력으로 현재 캐릭터의 상태와 목표 방향에 적합한 동작을 쉽고 빠르게 찾아 재생할 수 있다. 이 연구에서는 방향 뿐만 아니라 전체 경로를 마우스로 입력해서 캐릭터 미래 경로의 속도와 방향을 모두 고려하도록 모션 매칭의 기능을 확장하는 방법을 제안한다.

### 1. 서론

모션 매칭 [1]은 모션 캡처 데이터를 활용해 캐릭터의 동작을 합성하는 기법의 하나로, 간단한 특성 데이터 베이스를 생성해 복잡한 계산이나 인공지능 없이도 쉽고 빠르게 자연스러운 동작을 합성할 수 있다. 키보드 또는 조이스틱으로 다음 순간 캐릭터가 이동할 방향을 입력하면 모션 매칭은 특성 데이터베이스에서 가장 자연스러운 다음 동작을 출력한다. 만일 마우스로 평면 상의 경로를 지정하면 순간적인 방향 뿐만 아니라 전체 경로를 고려하게 되어 키보드 또는 조이스틱에 비해 즉석으로 더 자유롭고 중장기적인 동작을 합성하도록 할 수 있다. 그러나 이동 경로를 입력으로 주게 되면 캐릭터가 경로를 이탈하거나 급격히 방향을 전환하는 문제가 생기기 쉽다. 이 연구에서는 경로를 벗어난 캐릭터가 보다 부드럽게 원하는 궤도를 따라가면서 RNN기반의 네트워크로 더 자연스러운 쿼리를 생성해 모션 매칭의 성능을 확장하는 방법을 제안한다.

### 2. 마우스 경로를 활용한 쿼리 합성

모션 매칭 [1]은 구조화되지 않은 모션 캡처 데이터의 모든 동작 프레임에 대해 27차원의 간단한 특성 데이터 베이스를 생성한 다음, 캐릭터의 현재 동작과 가장 비슷하면서 가고자 하는 방향에 맞는 동작을 빠르게 찾아 재생하는 기법이다. 모션 매칭의 쿼리로서 쓰이는 27차원 특성은 캐릭터 루트(root)의 지역 좌표계로 표현된 양쪽 발의 위치와 속도  $p_i^{lfoot} \in \mathbb{R}^3, v_i^{lfoot} \in \mathbb{R}^3, p_i^{rfoot} \in \mathbb{R}^3, v_i^{rfoot} \in \mathbb{R}^3$  그리고 루트의 속도  $v_i^{hip} \in \mathbb{R}^3$ 를 포함하며, 남은 12차원의 정보는 총 3개의 미래 프레임에서 평면 상에 투사한 2차원 캐릭터 루트의 위치  $t_i \in \mathbb{R}^2$ 와 그 때 캐릭터가 바라보는 방향  $d_i \in \mathbb{R}^2$ 을 캐릭터 루트의 지역 좌표계에 대해 구한 것이다.

$$feature_i = \{p_i^{lfoot}, v_i^{lfoot}, p_i^{rfoot}, v_i^{rfoot}, v_i^{hip}, t_i, d_i\} \in \mathbb{R}^{27}$$

만일 마우스 경로 상의 점들로 미래 경로의 위치를 곧바로 계산한다면 현재 캐릭터의 위치를 고려하지 못하고, 미래경로상의 캐릭터의 방향을 경로 각 지점에서의 접선의 방향으로써 간단히 구한다면 캐릭터가 과거에 바라본 방향을 모두 무시하게 된다. 따라서 마우스 경로를 활용한 모션 매칭 데이터베이스가 자연스러운 동작을 합성하려면 다음 세가지 조건을 만족해야 한다.

- 경로를 벗어난 캐릭터는 다시 경로 위로 돌아오도록 해야 한다.
- 캐릭터가 미래에 바라보고 있을 방향은 현재 캐릭터가 진행해온 동작을 반영해야 한다.
- 캐릭터의 현재 속도 및 미래 경로는 주어진 경로의 형태와 간격을 반영해야 한다.

앞에서 언급된 단순한 쿼리에서 미래경로를 수정하면 더 적합한 쿼리를 만들 수 있다. 입력으로 주어진 마우스 경로에서 현재 캐릭터의 위치와 가장 가까운 점으로부터 경로의 끝까지 잘라낸 단편을  $traj_{mouse}$ 라고 하고  $traj_{mouse}$ 을 현재 위치에서 시작하도록 평행 이동 시킨

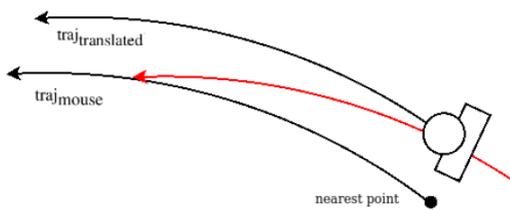
\* 구두발표논문, 요약논문 (Extended Abstract)

\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행되었음(No. 2019R1C1C1006778, NRF-2019R1A4A1029800).

경로를  $traj_{translated}$  이라고 할 때, 두 경로  $traj_{mouse}$ 와  $traj_{translated}$ 를 적당한 스케일 함수로 합성하면 현재 위치에서 자연스럽게 마우스 경로까지 지나도록 하는 경로를 합성할 수 있다(Figure 1).

$$t_i = s(i) * traj_{mouse} + (1 - s(i)) * traj_{translated}$$

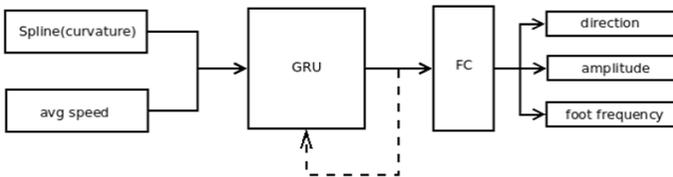
$$s(i) = \frac{1}{2} \cos\left(\frac{\pi i}{RANGE}\right) + \frac{1}{2}, (0 \leq i \leq RANGE)$$



**Figure 1:** A synthesized future trajectory. The RANGE can be of any value, but here it is given as the length of a future trajectory.

이렇듯 마우스 경로의 위치  $t_i$ 는 현재 위치를 기준으로 쉽게 구할 수 있지만 그 때 캐릭터가 바라보는 방향  $d_i$ 은 현재 방향만 고려해서 자연스러운 값을 구하기 어렵다. 인공 신경망을 활용하면 비교적 효율적으로 이 문제를 해결할 수 있으며 이 문제의 경우 시계열 데이터에 대한 문제가 되므로 순환형 네트워크를 쓰는 것이 자연스럽다.

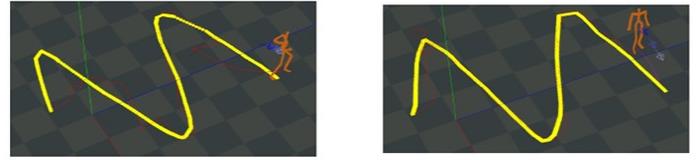
캐릭터의 경로 정보로 동작 네트워크를 구성하는 QuaterNet [2]은 한 개의 GRU유닛으로 구성된 페이스 네트워크(Pace network)로 경로 정보를 구한다(Figure 2). 페이스 네트워크는 진행해온 경로의 곡률과 재생 중인 동작의 평균 속도를 입력으로 캐릭터가 다음 순간 바라보는 방향, 속도, 그리고 발걸음 빈도를 출력한다. 이 출력 중 방향을 쿼리에 덧씌우면 최종적으로 전체 경로에 대해 자연스럽게 이어지는 캐릭터 미래경로를 계산할 수 있다.



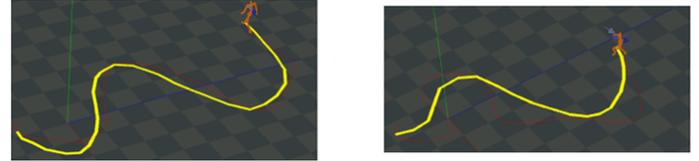
**Figure 2:** Overall structure of a pace network.

### 3. 실험 결과

이 연구에서 제안하는 두가지 방법의 효과를 보이기 위한 두가지 실험을 보인다. 실험에 쓰인 모션 캡처 데이터는 [3]의 걷거나 뛰는 모션(30Hz)이다.



**Figure 3:** A comparison of future trajectory positions using methods between a direct synthesis and a recalculation.



**Figure 4:** A comparison of future trajectory directions using methods between a heuristic synthesis and a pace network output.

첫번째 실험(Figure 3)은 이동 경로를 목표 경로의 가장 가까운 점으로 우선 직행하도록 할 때와 목표 경로까지 부드럽게 이어붙였을 때의 비교 결과이다. 두번째 실험(Figure 4)은 미래 경로상의 캐릭터의 방향을 단순히 미래 경로 각 지점에서의 접선의 방향으로 계산했을 때와 페이스 네트워크로 계산했을 때의 비교 결과이다.

### 4. 결론

이 연구에서는 모션 캡처 데이터를 활용하는 모션 매칭 기법에 사용자가 즉석으로 경로를 입력할 수 있도록 기능을 확장하는 방법을 제안한다. 입력된 경로는 캐릭터가 경로에서 벗어나는 경우를 고려하고, 다음 순간 향하는 방향과 속도를 기존에 캐릭터가 이동해온 동작과 자연스럽게 이어지도록 순환형 네트워크인 페이스 네트워크를 적용했다.

### 참고문헌

[1] S. Clavet, "Motion matching and the road to next-gen animation," *GDC 2016*, 2016.  
 [2] D. Pavllo, D. Grangier, and M. Auli, "Quaternet: A quaternion-based recurrent model for human motion," *CoRR*, vol. abs/1805.06485, 2018. [Online]. Available: <http://arxiv.org/abs/1805.06485>  
 [3] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," in *SIGGRAPH 2016*, 2016.

# Constrained Projective Dynamics: 에너지와 운동량을 보존하는 실시간 물리 기반 변형체 시뮬레이션\*

기민형<sup>1</sup>, 엄기원<sup>2</sup>, 정우석<sup>1</sup>, 한정현<sup>1</sup>

<sup>1</sup>고려대학교 컴퓨터학과

<sup>2</sup>텔레콤파리대학 이미지, 데이터, 신호학과

minh702@korea.ac.kr<sup>1</sup>, kiwon.um@telecom-paris.fr<sup>2</sup>, www0131@korea.ac.kr<sup>1</sup>, jhan@korea.ac.kr<sup>1</sup>

## Constrained Projective Dynamics: Real-Time Simulation of Deformable Objects with Energy-Momentum Conservation

Min Hyung Kee<sup>1</sup>, Kiwon Um<sup>2</sup>, WooSeok Jung<sup>1</sup>, JungHyun Han<sup>1</sup>

<sup>1</sup>Dept. of Computer Science and Engineering, Korea University

<sup>2</sup>IDS Department, LTCI, Telecom Paris, IP Paris

### 요약

본 논문은 에너지와 운동량을 보존하는 새로운 물리 기반 시뮬레이션 기법을 제안한다. 이 기법은 Projective Dynamics의 수치적 시간 적분에 대한 비구속 최적화 문제를 위치기반 에너지-운동량 구속 조건에 기반한 구속 최적화 문제로 확장한다. 이를 통해 시뮬레이션의 안정성과 효율성을 저해하지 않고 Projective Dynamics의 수치 감쇠 문제를 해결한다.

### 1. 서론

컴퓨터 그래픽스 분야에서 변형체의 생생한 움직임을 시뮬레이션하는 것은 지속적으로 연구되는 주제이다. 일반적으로 실시간 물리 기반 시뮬레이션에서는 안정성과 효율성에 대해 엄격한 기준이 요구된다. 따라서, 큰 time step에 대해서도 안정적인 시뮬레이션 결과를 보장하는 backward Euler 등의 implicit 수치적 시간 적분 기법이 자주 사용된다. Backward Euler 기반 기법 중 가장 널리 사용되는 것 중 하나는 Projective Dynamics [1](이후, PD로 약칭)이다. 이 기법은 시간에 대한 수치 적분 문제를 최적화 형태로 재구성하여 안정성과 효율성 관련 요구

사항을 모두 충족한다. 하지만, 이 기법은 에너지와 운동량에 대해 수치 감쇠가 발생하는 문제점이 있어 지속적인 진동 및 회전 등 에너지-운동량 보존이 요구되는 운동을 시뮬레이션 할 수 없다. 본 논문에서는 이를 해결하기 위해 위치기반 에너지-운동량 구속 조건을 사용하여 PD를 구속 최적화 문제로 확장한 Constrained Projective Dynamics를 제안한다.

### 2. Projective Dynamics

현재  $n$ 번 프레임에서의 위치  $\mathbf{x}_n$ 과 속도  $\mathbf{v}_n$ 이 주어질 때, PD는 backward Euler에서 유도된 목적 함수  $g(\mathbf{x}_{n+1})$ 에 대한 비구속 최적화 문제를 풀어 time step  $h$  이후 결과인  $\mathbf{x}_{n+1}$ 과  $\mathbf{v}_{n+1}$ 을 구한다.

$$g(\mathbf{x}_{n+1}) = \frac{1}{2h^2} \|\mathbf{x}_{n+1} - \mathbf{y}_n\|_{\mathbf{M}}^2 + E(\mathbf{x}_{n+1})$$

여기서  $\mathbf{y}_n = \mathbf{x}_n + h\mathbf{v}_n$ ,  $\mathbf{M}$ 은 질량 행렬, 그리고  $E$ 는 탄성 에너지이다.

### 3. Constrained Projective Dynamics

#### 3.1. 위치기반 에너지-운동량 보존 구속 조건

선운동량  $P$ , 각운동량  $L$ , 에너지  $H$ 에 대한 위치기반 보존 구속 조건은 다음과 같이 정의된다.

$$C_P(\mathbf{x}_{n+1}) = \frac{1}{h} \sum_i \mathbf{M}^i (\mathbf{x}_{n+1}^i - \mathbf{x}_n^i) - P$$

$$C_L(\mathbf{x}_{n+1}) = \frac{1}{h} \sum_i \mathbf{x}^i \times \mathbf{M}^i (\mathbf{x}_{n+1}^i - \mathbf{x}_n^i) - L$$

$$C_H(\mathbf{x}_{n+1}, \alpha) = \frac{1}{2h^2} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|_{\mathbf{M}}^2 + E(\mathbf{x}_{n+1}) - (1 - \alpha)H - \alpha K$$

여기서  $K$ 는 운동량에서 기인한 운동 에너지이다. 편의상, 위의 세 구속 조건을 통틀어  $\mathbf{c}(\mathbf{x}_{n+1}, \alpha)$ 로 표기한다.

\* 구두발표논문

\* 본 논문은 요약논문(Extended Abstract)으로, 본 논문의 원본 논문은 ACM Transactions on Graphics 40(3). 2021 에 게재 승인되었음.

\* 이 논문은 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터 육성지원사업 (IITP-2021-2020-0-01460) 지원 하에 수행된 연구임.

### 3.2. 구속 최적화

PD를 3.1 에서 정의한 구속조건을 사용하여 구속 최적화 문제로 확장하면 아래와 같이 정의된다.

$$\begin{aligned} & \underset{\mathbf{x}_{n+1}, \alpha}{\operatorname{argmin}} g(\mathbf{x}_{n+1}) + \frac{1}{2}\epsilon\alpha^2 \\ & \text{s.t. } \mathbf{c}(\mathbf{x}_{n+1}, \alpha) = \mathbf{0} \end{aligned}$$

이 문제는 Sequential Quadratic Programming(SQP)을 이용해 풀 수 있다.

### 4. 결과

아래 그림은 변형체 Cube를 우리 기법과 PD로 각각 시뮬레이션한 것을 보여준다. Cube는 수직 방향으로 눌린 뒤 각운동량이 가해진 상태로 초기화되었다. 그림 1 에서 볼 수 있는 것처럼, 우리 기법은 에너지와 각운동량을 보존해 Cube가 탄성 운동과 회전 운동을 유지하도록 하는 반면 PD는 수치 감쇠 때문에 이를 유지하지 못한다. 그림 2 의 그래프는 이를 정량적으로 보여준다.

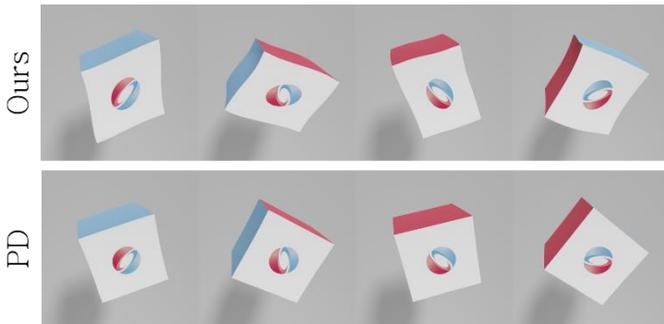


그림 1: 탄성+ 회전 운동 시뮬레이션

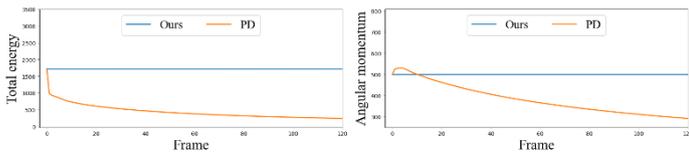


그림 2: 시간에 따른 에너지 및 각운동량 그래프

그림 3 은 변형체 공이 자유 낙하하여 바닥에 충돌하는 실험을 시뮬레이션한 것을 보여준다. 에너지가 보존되어 공이 충돌 이후 초기 높이로 튀어 오르는 것을 반복한다.

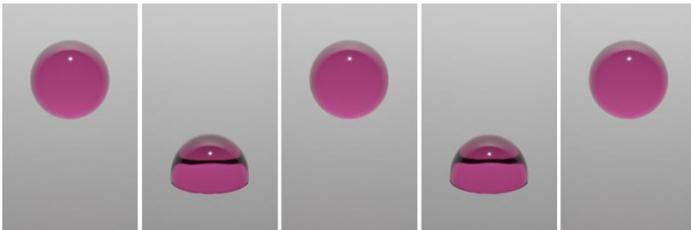


그림 3: 자유 낙하 및 충돌 시뮬레이션

그림 4 는 꼭지점을 고정한 트램폴린에 수직 방향으로 선운동량을 가한 상태로 초기화 한 뒤 우리 기법과 기존 에너지-운동량 보존 기법인 FEPR [2]로 각각 시뮬레이션한 것을 보여준다. FEPR은 임의의 시뮬레이션 기법(여기선 PD를 사용함)의 결과에 후처리를 가해 에너지와 운동량을 보존하는 기법이다. 그림 4 는 두 기법이 시각적으로 비슷한 결과를 생성하는 것을 보여준다. 하지만, 그림 5 에서 볼 수 있는 것처럼, 우리 기법은 PD+FEPR보다 훨씬 적은 solver iteration에 수렴한다.

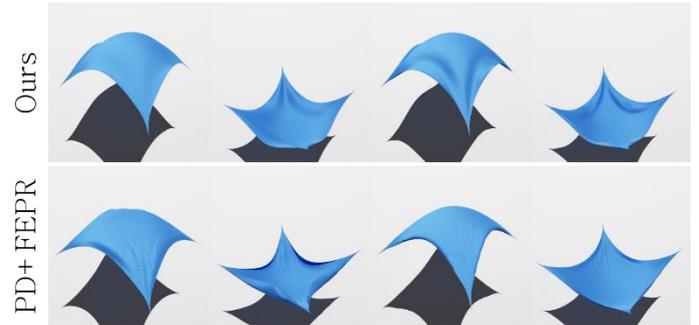


그림 4: 진동 시뮬레이션

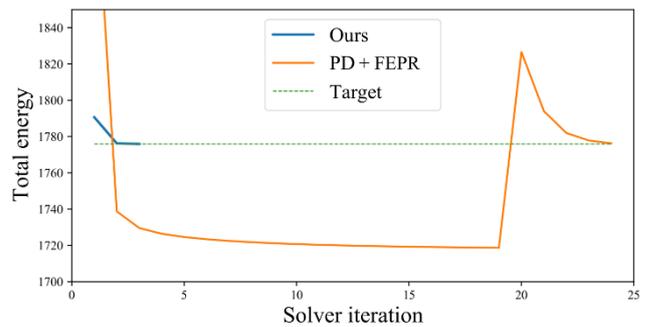


그림 5: Solver iteration에 따른 에너지 그래프

### 5. 결론

본 논문에서는 물리 기반 변형체 시뮬레이션의 수치 적분 문제를 새로운 구속 최적화 문제로 치환한 Constrained Projective Dynamics, 그리고 이를 위한 위치기반 에너지-운동량 구속 조건을 제안한다. 이는 PD의 수치 감쇠 문제를 효율적으로 해결하여 진동이나 빠른 회전 등 생생한 변형체 운동의 실시간 시뮬레이션을 가능하게 한다.

### 참고문헌

[1] Bouaziz, Sofien, et al. "Projective dynamics: Fusing constraint projections for fast simulation." ACM transactions on graphics (TOG) 33.4 (2014): 1-11.  
 [2] Dinev, Dimitar, et al. "FEPR: Fast energy projection for real-time simulation of deformable objects." ACM Transactions on Graphics (TOG) 37.4 (2018): 1-12.

---

## 한국컴퓨터그래픽스학회 2021 학술대회 논문집

---

2021년 7월 1일 인쇄

2021년 7월 1일 발행

발행인: 이제희

편집인: 윤성의, 이성길

발행처: 사단법인 한국컴퓨터그래픽스학회

주소: 서울특별시 관악구 관악로 1, 302동 312-1호

서울대학교 운동연구실

전화: 02-880-1864

FAX: 02-886-7589

URL: <http://cg-korea.org>

---

KOREA COMPUTER GRAPHICS SOCIETY

한국컴퓨터그래픽스학회



# KCGS 2021

