cg-korea.org

KOREA
COMPUTER
GRAPHICS
SOCIETY

한국컴퓨터그래픽스학회

KCGS 2020
학술대회논문집

KCGS 2020
Beyond Visibility

한국컴퓨터그래픽스학회

2020 학술대회 논문집

2020년 7월 7일 ~ 2020년 7월 9일

온라인 학술대회

주관기관: 한국컴퓨터그래픽스학회

KCGS 2020 환영사

"Beyond Visibility." 올해 한국컴퓨터그래픽스학회 하계 학술대회에서 선택한 슬로건입니다. 컴퓨터그래픽스 분야에서 자주 등장하는 'visibility'라는 단어가 올해에는 조금 색다른 의미로 다가올 수 있음에 주목했습니다. 서로 손에 닿는 거리에서 눈을 마주보고 거리낌 없이 대화를 나누던 일상이 더 이상 당연한 것이 아니게 되었습니다. 일시적인 해프닝으로 지나가지 않을까 했던 예상은 섣부른 기대로 판명되는 듯합니다. 한편, 미래의 모습으로 막연히 떠올려왔던 온라인 중심의 사회가 거짓말처럼 눈앞의 현실로 자리 잡고 있습니다. 시간과 공간적으로 분리된 개인들을 서로 연결하는 기술의 중요성이 급속히 커지고 있습니다. 컴퓨터그래픽스 분야의 핵심 연구 주제인 시각적 표현과 상호작용 기술은 이러한 필요에 부응하여 포스트 코로나 시대의 중요한 기술로 부상하고 있습니다. 나아가서, 단순한 시각적 상호작용을 넘어서서 여러 감각을 포괄하는 상호작용의 기반 기술로 그 외연을 확장하는 계기가 마련될 수도 있습니다. 먼 미래에 돌이켜봤을 때 올해가 바로 그 단초가 되기를 바라는 마음으로 이번 학술대회를 준비했습니다.

올해 학술대회는 한국컴퓨터그래픽스학회 역사상 처음으로 온라인 형식으로 개최됩니다. 조직 위원과 프로그램 위원 누구도 이번 학술대회가 어떤 형태여야 하는지 미리 예측할 수 없었습니다. 일정을 미루더라도 오프라인 학회의 형식을 유지하는 안부터 올해는 한 회 건너뛰는 안까지, 생각할 수 있는 모든 대안을 테이블 위에 올려놓고 토론한 끝에 온라인 학회의 형태로 결론 내렸습니다. 하지만 이는 미지의 세계로 향하는 첫 걸음일 뿐이었습니다. 실시간 상호작용을 얼마나 포함시킬 것이냐, 어떤 플랫폼을 사용할 것이냐, 기존에 진행해왔던 여러 행사와 이벤트는 얼마나 유지할 것이냐 등 새로운 결정의 연속일 수밖에 없었습니다. 이 환영사를 쓰는 지금도 불확실성은 완전히 제거되지 않았습니다. 다만 여러 가지가 결정되었고 실행되는 중입니다. 온라인 학회의 중심 역할을 맡을 독립적인 홈페이지가 제작되었습니다. 이곳에서 학회에 등록하고, 프로그램을 확인하고, 프로시딩을 다운로드 받고, 논문 발표 영상을 시청하고, 실시간 질의/응답 세션에 참여할 수 있습니다. 학회장에서 서로 직접 만날 수 있던 기회는 소셜 네트워크 서비스가 대신할 것입니다. 학회장에서 커피 쿠폰을 사용했던 것처럼, 소셜 네트워크 서비스에서 활동하다 보면 우연히 커피 기프티콘을 받으실 수도 있습니다. 논문 질의/응답에 열심히 참여하시는 분이라면 경품을 기대해도 좋을 것 같습니다. 학회장에서 직접 건네 드렸던 경품이 아마도 이번에는 물류 서비스를 통해 전달된다는 차이만 있을 것입니다.

온라인 학회라 해도 가능하면 많은 프로그램을 유지하고자 했습니다. 이번 학회의 후원사인 엔씨소프트의 류제니 상무님과 KAIST의 윤성의 교수님이 기꺼이 초청강연을 맡아 주셨습니다. 세계 컴퓨터그래픽스 분야 연구를 선도하는 뛰어난 국내 연구자 세 분의 발표도 기다리고 있습니다. 아울러 신진연구자상 선정을 위해 두 후보의 발표가 진행될 예정이며, 그 중 수상자는 실시간 온라인 투표를 통해 결정될 것입니다. 이들 프로그램이 모두 실시간 플랫폼을 통해 진행되는데 반하여, 논문 발표 세션은 미리 녹화된 발표 동영상을 제공하는 스트리밍 서비스를 통해 진행될 것입니다. 프로그램 일정에 따라 발표 동영상이 순차적으로 공개될 예정이며, 각 세션이 끝나면 실시간 플랫폼을 통해 세션 체어를 중심으로 논문 저자들이 참여하는 질의/응답 시간이 진행됩니다. 많은 연구자분들이 최근 이루어진 훌륭한 연구 성과를 논문과 포스터의 형태로 제출해 주셨고, 이번 학술대회에서는 그 중 25편의 논문

과 5편의 포스터가 발표될 예정입니다. 여러모로 어려운 시기임에도 불구하고 온라인 학회의 불편함을 감수하면서 이들 프로그램에 기여해주신 모든 분들께 심심한 감사의 말씀을 드립니다.

우리 학술대회의 일부 프로그램은 아쉽게도 올해에는 여건상 생략하기로 하였습니다. 새로운 연구 분야에 진입하고자 하는 젊은 연구자들에게 훌륭한 배움의 기회가 되었던 여름학교, 그리고 잠재되어 있던 넘치는 열정을 확인하며 학회 구성원들의 마음을 한데 모을 수 있었던 해변가요제가 대표적입니다. 특히 해변가요제의 경우에는 동영상 스트리밍, 가상현실 플랫폼 등을 이용한 여러 대안들을 모색해 보았으나 시간적 한계로 인해 결국 내년을 기약하기로 하였습니다. 이와 같은 프로그램 상의 차이 외에도, 처음 진행되는 온라인 학회인 만큼 미처 예상치 못했던 문제점들이 여러 곳에서 드러날 수도 있을 것으로 우려됩니다. 부족한 부분들을 발견하실 때마다 바로 알려주시면 조직 위원과 프로그램 위원들이 힘을 모아 문제를 해결해 나가겠습니다. 모쪼록 이번 온라인 학술대회가 국내 컴퓨터그래픽스 분야의 최신 연구 결과를 공유하고 연구자들 간의 교류와 화합을 도모하는 본래의 역할을 온전히 감당할 수 있기를 기대합니다. 모든 참가자 여러분들의 건강과 행복을 기원합니다.

감사합니다.

2020년 6월 30일
KCGS 2020 학술대회 조직위원장 이강훈, 김민혁

학술대회 조직위원회

대회장 이제희 (서울대학교)

조직위원장 이강훈 (광운대학교), 김민혁 (KAIST)

프로그램위원장 이성길 (성균관대학교), 이윤진 (아주대학교)

수상위원회 위원장 김영준 (이화여자대학교)

조직위원 김덕수 (한국과학기술교육대학교), 김재일 (경북대학교),
이윤상 (한양대학교), 이지은 (한성대학교), 조동식 (원광대학교),
조성현 (POSTECH), 최명걸 (가톨릭대학교)

프로그램위원 계희원 (한성대학교), 고성안 (UNIST), 권구주 (배화여자대학교),
권태수 (한양대학교), 김익재 (KIST), 김정현 (고려대학교),
김종현 (강남대학교), 김진모 (한성대학교), 김창헌 (고려대학교),
김치용 (동의대학교), 김태영 (서경대학교), 김형석 (건국대학교),
김형석 (동의대학교), 남양희 (이화여자대학교), 노준용 (KAIST),
박경주 (중앙대학교), 박상일 (세종대학교), 박진아 (KAIST),
박진호 (숭실대학교), 백낙훈 (경북대학교), 변혜원 (성신여자대학교),
서진석 (동의대학교), 서진욱 (서울대학교), 손봉수 (중앙대학교),
송오영 (세종대학교), 신영길 (서울대학교), 안상철 (KIST),
오경수 (숭실대학교), 윤성의 (KAIST), 윤승현 (동국대학교),
이성희 (KAIST), 이아현 (ETRI), 이영호 (목포대학교),
이인권 (연세대학교), 이정진 (숭실대학교), 이종원 (세종대학교),
이주행 (ETRI), 이택희 (한국산업기술대학교), 이혜영 (홍익대학교),
이환용 (아주대학교), 임순범 (숙명여자대학교), 임인성 (서강대학교),
장윤 (세종대학교), 정순기 (경북대학교), 조동식 (원광대학교),
최민규 (광운대학교), 최수미 (세종대학교), 최아영 (가천대학교),
최정주 (아주대학교), 하종성 (우석대학교), 한다성 (한동대학교),
한정현 (고려대학교)

목 차

논문발표 4: 애니메이션/시뮬레이션 2020.7.7(화) 17:00-18:00

논문발표 5: HCI/그래픽스시스템/모델링 2020.7.8(수) 10:30-11:50

논문발표 6: 가상/증강현실 II 2020.7.8(수) 15:30-17:10

논문발표 7: 렌더링 2020.7.9(목) 10:30-12:10

포스터발표

초청강연

Deep Learning and Physically-based Rendering

윤성의 교수

KAIST 전산학부

http://sgvr.kaist.ac.kr/~sungeui/

강연 내용

실사 렌더링 기술은 상당히 오랫동안 연구되어 왔고, 영화/게임분야에서 필수적인 가시화 도구로 사용되고 있다. 하지만 실사 렌더링 기술이 아직 실시간 성능까지 확보하고 있지 못하기에 이 문제를 해결하기 위해 다양한 연구가 이루어지고 있다. 특히 최근에 많은 각광을 받고 있는 Deep learning 기술을 활용하여 성능 개선이 이루어 지고 있는데, 본 톡에서는 이와 관련된 최신 기술을 설명하고자 한다.

흥미롭게도 실사 렌더링 기술이 Deep leaning 기술의 도움을 받아 발전하는 측면도 있지만, 도리어 실사 렌더링 기술이 Deep learning의 핵심 문제를 해결해주는 경향도 있다. 특히 성능 향상을 위해 많은 데이터를 요구하는 Deep learning 기술의 특징이 있는데, 이런 문제를 실사 렌더링 기술로 해결 가능하다. 본 강의에서는 이런 측면의 새로운 기술도 다뤄보고자 한다.

강연자 이력

2007년 7월 ~ 현재: KAIST 전산학부 교수 (SGVR: Scalable Graphics, Vision, and Robotics Lab)
2006년 1월 ~ 2007년 6월: Post. Doctoral researcher, Lawrence Livermore National Lab.
2001년 8월 ~ 2005년 12월: UNC-Chapel Hill, 박사 졸업
1999년 3월 ~ 2001년 2월: 서울대 전산학과 석사졸업
1995년 3월 ~ 1999년 2월: 서울대 전산학과 학사졸업

초청강연

비디오게임 개발과정

류제니 AD

엔씨소프트

강연 내용

글로벌 컨텐츠는 글로벌 개발환경에서 만들어 질 수 있다는 것을 지난 20여년간 경험을 통하여 알 수 있었다. Triple A 퀄리티의 게임을 만드는 회사들은 누구나 다음의 3가지 목표를 공통적으로 갖게 된다. 1) 재미있는 게임, 2) 최고의 그래픽 퀄리티, 3) 높은 경쟁력을 갖춘 글로벌 컨텐츠. 이 세가지의 목표를 실행가능한 결과로 만들어 낸 요인들 중 "제작과정과 운영"은 빼놓을 수 없는 요소일 것이다. 대규모의 팀을 조직하고 같은 비젼을 공유하면서, 같은 방향, 같은 속도로 달릴 수 있었던 개발환경은 무엇이었는지, "People" 과 "Diversity"를 강조하는 글로벌회사들이 중요하게 생각하는 글로벌 컨텐츠적 요소들은 무엇인지, EA와 Activision 등에서의 경험과 사례를 바탕으로 소개하고자 한다.

강연자 이력

29년전 미국으로 이민. AAU에서 컴퓨터 아트 석사과정을 마치고 바로 게임개발자로 시작. 20여년간 게임아트 분야를 리드하며 다수의 비디오게임 개발에 코어멤버로 참여. 현재 NC소프트 임원.

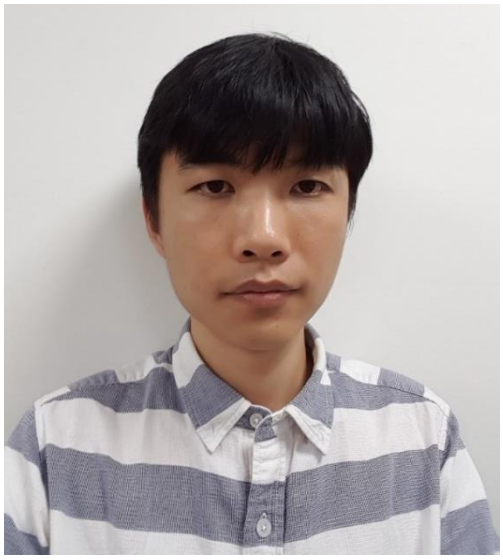
Jenny Ryu 게임개발 이력:

Castlevania: Resurrection	Dead to Right I	The Lord of the Rings: The Return of the King	The Godfather I	Call of Duty: Modern Warfare 3	Call of Duty: Advance Warfare	Call of Duty: WWII	NCsoft
Konami USA	Namco USA	EA	EA	Activision/ Sledgehammer Games (Infinity Ward, Treyarch, Raven)	Activision/ Sledgehammer Games (High Moon, Raven)	Activision/ Sledgehammer Games (Raven)	NCsoft Korea
Canceled 2000	Release Date: 2002	Release Date: Nov. 2003	Release Date: Mar 2006	Release Date: Nov. 2011	Release Date: Nov. 2014	Release Date: Nov. 2017	현재 개발중
Action-adventure	Action-adventure, Third-person shooter	Action RPG	Action-adventure	First-person shooter	First-person shooter	First-person shooter	-
Single player	Single player	Single player	Single player	Single Player/ Multiplayer	Single Player/ Multiplayer	Single Player/ Multiplayer	-
Sega Dreamcast	PC/PlayStation2	PC/PlayStation2	PC/ PlayStation2/ PlayStation3/ Xbox 360	PC/ PlayStation3/ Xbox 360	PC/ PlayStation 4/ Xbox One	PC/ PlayStation 4/ Xbox One	-
Character Artist/ Cinematic 디렉터	Character Artist/ 에니메이터	Senior Character Artist/ 3D 에니메틱 아티스트	Lead Character Artist	Lead Character Artist	Lead Character Artist/ Character Director	Art Director	Art Director/ 임원

부분적으로 참여했던 게임들:
- Call of Duty: Black Ops 4 (2018/ Treyarch) – 여자 주인공 모델링
- Call of Duty: Bleak Ops 3 (2015/ Treyarch) – 메인 케릭터 (미식축구 선수 Marshawn Lynch) Head 모델링작업
- Midnight Club II (2003/ Rockstar Game) – Cinematic 프로젝트 온영 (co-producer) 및 에니메이션 디렉터
- Seahorse GDG 게임회사 창업 (2006~2009) – Founder /Executive Producer/아트디렉터

도서출간 및 출강:
- Academy of Art University, SF USA 졸강 (1999 ~ 2005)
- Maya Modeling and Animation 튜토리얼 북 출판, Korea (2000)

2020 신진연구자상 후보



이경호 (Kyungho Lee)

whcjs13@ncsoft.com

이경호 박사는 현재 엔씨소프트의 Motion AI 팀에 재직 중이다. 2012년 서울대학교 컴퓨터 공학부에서 학사 학위를 취득하였고, 2019년 서울대학교 이제희 교수의 운동 연구실에서 박사 학위를 취득하였다. 박사 과정 동안 모션 캡쳐 데이터를 활용한 Data-driven Character Animation 분야에서 연구를 진행하였고, 모션 캡쳐 데이터로부터 Neural Network를 학습하여 사람 캐릭터의 실시간 동작 생성을 주제로 박사 학위 논문을 작성하였다.
엔씨소프트에 입사 후 학습 된 Neural Network의 반응 속도를 조절할 수 있는 기법에 대한 연구를 진행하였으며, 현재 연구 결과를 실제 게임에 적용 해 보기 위한 작업을 진행 중이다.

Interactive Character Animation by Learning Multi-Objective Control

Learning Time-Critical Responses for Interactive Character Control

2020 신진연구자상 후보

이 주 호 (Lee, Joo Ho) jhlee900958@gmail.com

이주호 후보자는 카이스트 전산학부 박사과정(지도교수: 김민혁)으로 재학중이다. 후보자의 연구 주제는 "고충실도 컴퓨터 그래픽스를 위한 물리적 외형 재현 기술에 대한 연구(Digital reproduction of Scene Appearance for High-Fidelity Computer Graphics)"로 디지털 컨텐츠 생성에 관련 전반에 걸쳐 연구하였다. 구체적으로 가려짐 현상을 극복하기 위한 이미지 채움 기술, 사실성을 높이기 위한 다중반사 표면 모델링, 그리고 RGBD 카메라를 이용한 실시간 기하 및 고충실도 텍스처 재구성 기술에 대해 연구하였다.

후보자의 박사학위 논문은 1편의 ACM SIGGRAPH Asia 논문과 2편의 IEEE CVPR 논문으로 이루어졌다. 그중 2020년 CVPR논문은 구두발표이며 학회 best paper finalists 26인에 호명되었다. 후보자는 오는 2020년 8월에 박사학위를 취득할 예정이다.

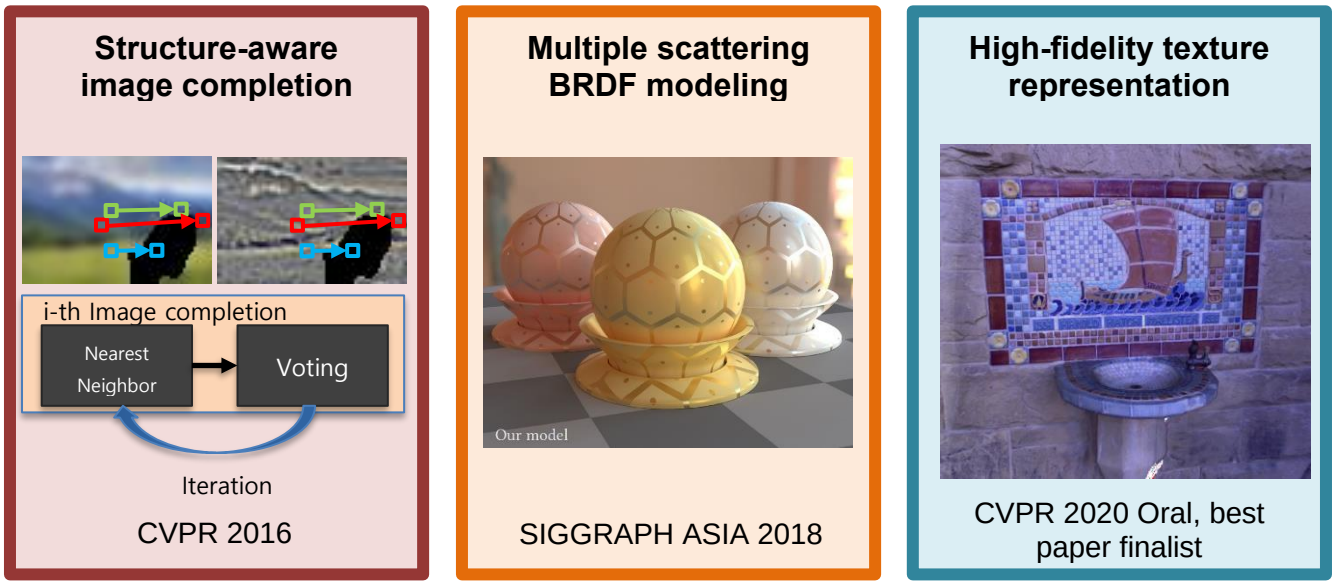


그림 1. 연구 요약. 좌) 라플라시엔 기반의 이미지 채움 기술 (CVPR 2016). 중) 다중반사를 고려한 표면반사 모델링 (SIGGRAPH ASIA 2018). 우) 실시간 고충실도 텍스처 재구성 기술 (CVPR 2020 Oral, best paper finalist).

캐리커처로부터 사실적인 얼굴 영상 자동 생성*

장원종[0], 정유철, 이승용 포항공과대학교
{wonjong, ycjung, leesy}@postech.ac.kr

Automatic Realistic Facial Image Generation from a Caricature

Wonjong Jang[0], Yucheol Jung, Seungyong Lee
POSTECH

요약

본 논문은 입력 캐리커처 영상과 얼굴의 핵심 특징을 공유하는 사실적인 얼굴 영상을 생성하는 시스템을 제안한다. 일반적으로 캐리커처와 얼굴 영상 사이에는 공통된 특징이 존재한다. 캐리커처로부터 얼굴 영상과 공통된 핵심 특징을 추출하기 위해 얼굴 식별을 목적으로 훈련된 깊은 신경망 네트워크를 이용한다. 본 논문에서 제안하는 특징 맵 변환 네트워크는 얼굴 식별 모델로부터 추출된 얼굴 특징 맵을 StyleGAN의 잠재 공간에 투영한다. 제안된 프레임워크로부터 생성된 얼굴 영상이 입력 캐리커처의 성별, 나이, 인종 등과 같은 특징을 잘 보존하는 것을 실험을 통해 확인하였다.

1. 서론

캐리커처는 주로 사람에 대해여 다른 사람과 비교되는 특징을 포착하여 익살스럽게 표현하는 방식을 의미한다. 하지만 캐리커처를 만들기 위해서는 전문적인 아티스트의 도움이 필요할 뿐만 아니라 이를 제작하기 위해 상당 시간이 소요된다.

최근 딥 러닝 기반으로 입력 얼굴 영상을 캐리커처로 자동으로 전이하는 시스템이 많이 제시되었다 [1] [2]. 이 프레임워크들은 스타일 전이와 이미지 와핑의 결합으로 문제를 해결하였다. 하지만 그들의 결과는 아티스트가 그린 실제 캐리커처와 비교하였을 때 대중이 일반적으로 기대하는 캐리커처를 생성하기에 무리가 있다.

데이터 기반 캐리커처 자동 생성이 어려운 주된 이유는 캐리커처가 가지는 다양성에 비해 수집할 수 있는 얼굴 영상과 캐리커처의 쌍의 양이 적기 때문이다. 따라서 캐리커처의 핵심적 특징을 지닌 얼굴 영상을 자동으로 생성할 수 있다면 캐리커처 생성에 도움이 될 것이다.

본 논문에서는 캐리커처로부터 얼굴 영상을 생성하기 위해 특징 맵 변환 네트워크를 제안한다. 특징 맵 변환 네트워크는 얼굴 식별에 특화된 FaceNet [3]으로부터 얻어진 얼굴 특징 맵을 사실적인 얼굴 영상 생성에 특화된 StyleGAN [4]의 잠재 공간으로 변환해주는 역할을 한다. 얼굴 인식에 필요한 핵심 정보를 담은 특징 맵을 사실적 얼굴 영상의 잠재 공간으로 투영함으로써 캐리커처의 얼굴 특징을 살리면서도 사실적인 얼굴 생성을 할 수 있다. 본 논문에서는 예시를 통해 시각적으로 해당 성질을 보인다.

2. 기존 연구

CariGANs [1]은 얼굴의 기하적 구조와 텍스처 전이를 학습하기 위해 각각의 역할을 하는 두 개의 CycleGAN 기반 모델을 훈련하였다. 그들의 CycleGAN 기반 모델은 캐리커처로부터 얼굴을 생성할 수 있지만 기하적 변환이 정면 영상에 과적합되어 있어 정면과 큰 포즈 차이를 지닌 얼굴 영상에는 한계를 보인다.

WarpGAN [2]은 얼굴 영상으로부터 캐리커처를 만들기 위해 얼굴의 기하적 구조의 변환과 텍스처 전이를 동시에 학습하는 새로운 프레임워크를 제안하였다. 직관적으로 생각하였을 때 캐리커처로부터 얼굴을 생성하는 가장 쉬운 방법은 WarpGAN [2]의 입력과 출력을 뒤집어 다시 학습하는 것이다. 하지만 실험을 통해 이러한 방식의 접근은 사실적인 얼굴 영상을 생성하는 데에 어려움이 있음을 확인하였다.

StyleGAN [4]은 progressive-GAN 구조와 AdaIN (Adaptive Instance Normalization) layer 를 이용하여 사실적인 고해상도 얼굴 영상을 생성해내는 방법을 제안하였다.

3. 시스템

3.1. 데이터

특징 맵 변환 네트워크를 학습하기 위해 같은 사람에

* 구두 발표논문
* 본 연구는 과학기술정보통신부의 재원으로 정보통신기술진흥센터(SW 스타랩, IITP-2015-0-00173)와 차세대정보·컴퓨팅기술개발사업(NRF-2017M3C4A7066317)의 지원을 받아 수행 되었음.

대한 캐리커처와 얼굴 영상의 쌍이 필요하다. 이를 위해 WebCaricature [5] 데이터 집합을 이용하였고, 훈련 데이터와 실험 데이터는 [2]와 같은 방법으로 분리하였다. 본 논문에 기재된 결과는 모두 실험 데이터로부터 만들어진 영상이다.

3.2. 특징 맵 변환 네트워크

특징 맵 변환 네트워크 F는 사전 훈련된 딥러닝 기반 얼굴 식별 네트워크인 FaceNet [3]로부터 얻어진 깊은 얼굴 특징 맵을 고화질 얼굴 영상 생성에 특화된 StyleGAN [4]의 잠재 공간 Z 으로 변환한다. 이 때, 입력 영상의 공간적인 정보를 활용하기 위해 특징 벡터가 아닌 특징 맵을 사용한다. 전체적인 네트워크의 흐름도는 그림 1에 표현되어 있다.

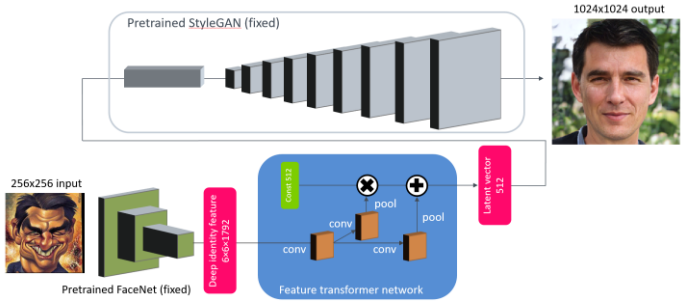


그림 1: 전체 네트워크 흐름도

네트워크 구조는 StyleGAN의 잠재 변수와 같은 차원을 가지는 특정 상수 벡터를 네트워크에서 얻어진 아핀 매개변수를 이용해 변환하도록 구성되어 있다. 즉, 네트워크는 총 세 개의 합성 곱 계층과 두 개의 풀링 계층을 거쳐 아핀 매개변수를 만들어 낸다.

본 논문의 시스템의 입력과 출력이 핵심적 얼굴 특징을 공유하도록 유도하기 위해 StyleGAN에서 무작위 생성된 사진 집합 M에 대하여 손실 함수를 다음과 같이 FaceNet [3]의 특징 공간에서 정의하였다.

$$L_p = \sum_{p \in M} ||\phi(G(F(\phi(p)))) - \phi(p)||_2^2$$

여기에서 ϕ는 FaceNet, G는 StyleGAN을 의미한다.

네트워크를 위의 손실 함수로만 학습하게 될 경우 캐리커처에 대해서 일반화되기 어려운 문제가 발생한다. 실제 캐리커처와 얼굴 영상 쌍을 이용한 손실 함수로 네트워크를 미세 조정해주어 캐리커처에 대해서 일반화될 수 있게 한다.

$$L_c = \sum_{(p,c) \in P} ||\phi(G(F(\phi(c)))) - \phi(p)||_2^2$$

여기에서 p와 c는 각각 WebCaricature [5] 데이터 집합에 존재하는 같은 사람에 대한 얼굴 영상과 캐리커처를 의미한다.

4. 결과

그림 2: 실제 캐리커처로부터 얼굴 생성 예시

그림 3: WarpGAN으로 생성된 캐리커처로부터 얼굴 생성 예시

그림 2에 제시된 것과 같이 캐리커처와 유사한 얼굴 영상이 생성되었음을 확인할 수 있다. 캐리커처에 해당하는 원본 얼굴과 비교하였을 때에도 성별, 나이, 인종 등과 같은 얼굴 특징들을 잘 복원해냈음을 볼 수 있다. 뿐만 아니라 WarpGAN이 생성한 캐리커처에 대해서도 이에 대해 훈련되지 않았음에도 불구하고 그림 3과 같이 얼굴 특징을 보존하는 결과를 보인다.

5. 결론 및 향후 계획

본 논문에서는 잘 사전 학습된 두 네트워크와 이 둘 사이의 변환을 제공하는 특징 맵 변환 네트워크를 이용하여 캐리커처의 얼굴 특징을 살리면서도 사실적인 얼굴 영상을 생성하는 방법을 제안하였다. 향후 캐리커처로부터 생성된 얼굴과 실제 얼굴 사이의 유사성을 더 높일 수 있는 방법에 대한 연구가 필요하다.

참고문헌

[1] Li, Wenbin, et al. "Carigan: caricature generation through weakly paired adversarial learning." *arXiv preprint arXiv:1811.00445* (2018).
[2] Shi, Yichun, Debayan Deb, and Anil K. Jain. "Warpgan: Automatic caricature generation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
[3] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
[4] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
[5] Huo, Jing, et al. "WebCaricature: a benchmark for caricature recognition." *arXiv preprint arXiv:1703.03230* (2017).

디블러링 알고리즘의 평가 및 학습을 위한 실제 블러 영상 데이터셋*

임재성[o1], 이해윤[2], 원주철[3], 조성현[4]
[1,2]대구경북과학기술원 정보통신융합전공 [3,4]포항공과대학교 컴퓨터공학과
jsrim@dgist.ac.kr haeyun@dgist.ac.kr jcwon@postech.ac.kr s.cho@postech.ac.kr

Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms

Jaesung Rim[o1], Haeyun Lee[2], Jucheol Won[3], Sunghyun Cho[4]
[1,2]Dept. of Information and Communication Engineering, DGIST [3,4]Dept. of Computer Science and Engineering, POSTECH

Abstract

In this work, we present a large-scale dataset of real-world blurred images and their corresponding sharp images captured in low-light environments for learning and benchmarking single image deblurring methods. To collect our dataset, we build an image acquisition system to simultaneously capture a geometrically aligned pair of blurred and sharp images, and develop a post-processing method to further geometric and photometric alignment. Our experiment shows that our dataset significantly improves deblurring quality for real-world low-light images.

1. Introduction

Images captured in low-light environments such as at night or in a dark room often suffer from motion blur caused by camera shakes or object motions as the camera requires a longer exposure time. With the advent of deep learning, several deep learning-based deblurring approaches [1, 2, 3, 4] have been proposed and shown a significant improvement. To learn image deblurring of real-world low-light photographs, they require a large-scale dataset that consists of real-world blurred images and their corresponding ground truth sharp images. However, there exist no such datasets so far due to difficulties involved with acquisition of real-world blurred images and their corresponding sharp images, which forces the existing approaches to resort to synthetic datasets such as the GoPro dataset [3].

In this paper, we present the first large-scale dataset of real-world blurred images for learning and benchmarking single image deblurring methods.

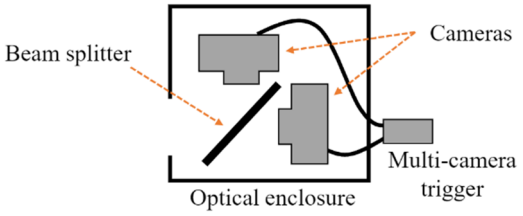


Figure 1: A diagram of our image acquisition system.

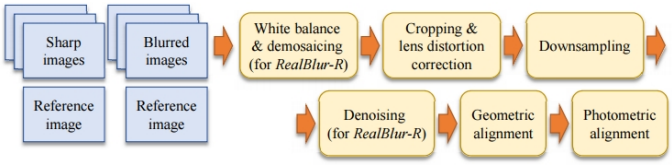


Figure 2: Overall procedure of our post-processing.

Using our dataset, we benchmark existing deblurring methods and their limitations. Our experiment shows that the dataset greatly improves the performance of deep learning-based deblurring methods on real-world blurred images.

2. Image Acquisition System and Process

To capture both blurred and sharp images simultaneously, we built a dual camera system shown in Figure 1. Our system consists of a beam splitter and two mirrorless cameras (Sony A7RM3 with Samyang 14mm F2.8 MF) so that the cameras can capture the same scene. One camera captures a blurry image with 1/2 shutter speed, while the other captures a sharp image with 1/80 shutter speed.

For each scene, we first captured a pair of two sharp images, which we refer to as a reference pair. We then captured 20 pairs of blurred and sharp images of the same scene to increase the amount of images and the diversity of camera shakes. We captured 4,738 pairs of images of 232 different scenes including reference

---

* 구두 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회에 제출 중
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020R1C1C1014863)

pairs. We captured all images both in the camera raw and JPEG formats, and generated two sets of datasets: *RealBlur-R* from the raw images, and *RealBlur-J* from the JPEG images processed by the camera ISP.

3. Post-Processing

Figure 2 shows an overview of our post-processing. For each pair of sharp and blurred images, we apply white balance and demosaicing to them in the first step. Images have invalid areas along the boundaries that capture outside the beam splitter or inside the system. Thus, we crop out such invalid regions. We correct lens distortions from the cropped images. Then, we downsample the corrected images, and perform denoising to the downsampled sharp image.

Although our image acquisition system has physically well-aligned cameras, there still exists some amount of geometric misalignment. To address this issue, for each scene, we first estimate the blur kernel k by minimizing the following energy function:

$$E(k) = ||k * \nabla s - \nabla b||^2 + \lambda||\nabla k||^2 \quad (1)$$

where s and b are sharp and blurred images, and ∇ is the image gradient operator. λ is the weight for the second term, which we set $\lambda = 10^3$. Then we align the sharp and blurred images using the center of mass of the blur kernel. We found that our approach using the center of mass of a blur kernel significantly improves the learning of image deblurring. Finally, we photometrically align blurred and sharp images using a linear model defined as $s = \alpha b + \beta$ where α and β are estimated from the reference pair.

4. Benchmarking

For the benchmark, we randomly select 182 scenes as our training sets and the remaining 50 scenes as our test sets. We found that using multiple training sets tends to achieve higher performance. Thus, we use dynamic scene blur dataset named GoPro [3] and synthetic uniform blur dataset as additional training sets. We trained SRN-DeblurNet [1] and DeblurGan-V2 [2] with pre-trained weight for stable training.

We then benchmark state-of-the-art deblurring methods including recent deep learning-based approaches using our test sets. Table 1 shows a summary of the benchmark. The deep learning-based approaches trained with our training sets show the best performance proving the benefits of training with real low-light blurred images. Figure 3 shows a qualitative comparison of the methods in Table 1. All

Table 1: Benchmark of state-of-the-art deblurring methods on real-world blurred images. Blue* : models trained with our dataset.

RealBlur-J		*RealBlur-R*	
Methods	PSNR/SSIM	Methods	PSNR/SSIM
SRN-DeblurNet* [1]	31.52/0.9217	SRN-DeblurNet* [1]	38.94/0.9674
DeblurGAN-v2* [2]	30.19/0.8957	DeblurGAN-v2* [2]	37.12/0.9520
DeblurGAN-v2 [2]	28.95/0.8831	Zhang et al. [4]	36.16/0.9539
Zhang et al. [4]	28.78/0.8805	SRN-DeblurNet [1]	36.04/0.9526
SRN-DeblurNet [1]	28.73/0.8829	DeblurGAN-v2 [2]	35.84/0.9505
Nah et al. [3]	28.43/0.8551	Nah et al. [3]	34.31/0.8886

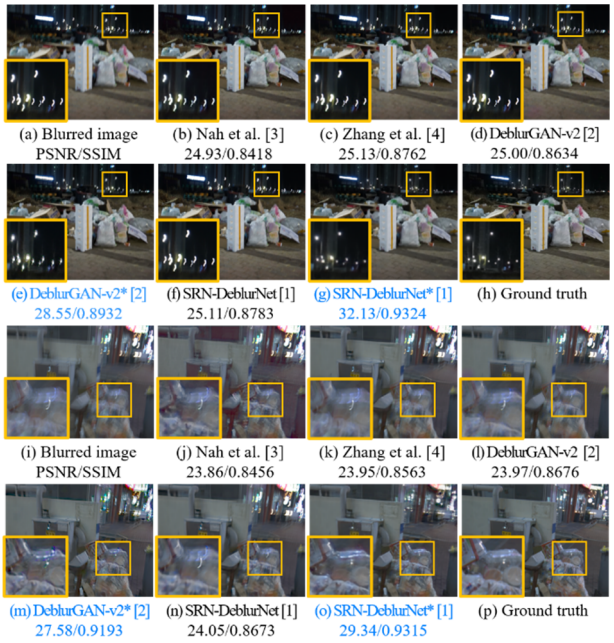


Figure 3: Qualitative comparison of different deblurring methods on the *RealBlur* test sets. Methods marked with `*' in blue are trained with our datasets.

the methods trained without real-world blurred images fail to restore light streaks as well as other image details. On the other hand, the methods trained with our datasets show better restored results.

References

[1] Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring, *CVPR*, 2017

[2] Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better, *ICCV*, 2019

[3] Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring, *CVPR*, 2017

[4] Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, R.W., Yang, M.H.: Dynamic scene deblurring using spatially variant recurrent neural networks, *CVPR*, 2018

교수/박사연구원급 발표

Image-Based Acquisition and Modeling of Polarimetric Reflectance

백승환[0,1], Tizian Zeltner[2], 구현진[1], 황인승[1], Xin Tong[3], Wenzel Jakob[2], 김민혁[1]

[1]KAIST, [2]EPFL, [3]Microsoft Research Asia

본 논문은 ACM SIGGRAPH 2020에 발표될 예정임.

교수/박사연구원급 발표

A Scalable Approach to Control Diverse Behaviors for Physically Simulated Characters

원정담, Deepak Gopinath, Jessica Hodgins

Facebook AI Research

본 논문은 ACM SIGGRAPH 2020에 발표될 예정임.

교수/박사연구원급 발표

Fast and Flexible Multilegged Locomotion Using Learned Centroidal Dynamics

권태수[0,1], 이윤상[1], Michiel van de Panne[2]

[1]한양대학교, [2]University of British Columbia

본 논문은 ACM SIGGRAPH 2020에 발표될 예정임.

무한공간탐험에서 직접적인 상호작용을 위한 회복 기법*

민대홍[0], 이동용, 조용훈, 이인권
연세대학교 컴퓨터과학과
bingo904@yonsei.ac.kr, pfldy2850@gmail.com, maxburst@gmail.com, iklee@yonsei.ac.kr

Shaking Hands in Virtual Space: Recovery in Redirected Walking for Direct Interaction between Two Users

Dae-Hong Min[0], Dong-Yong Lee, Yong-Hun Cho, In-Kwon Lee
Dept. of Computer Science, Yonsei University

요약

가상환경에서 직접적인 상호작용을 제공하기 위한 다양한 연구들이 있었다. 본 연구에서는 다중 사용자 환경에서 무한공간탐험법(Redirected Walking: RDW)을 적용 시 발생하는 직접적인 상호작용 문제에 대해 제시하고 이를 해결하였다. 예를 들어, 두 사용자가 각각 무한공간탐험법을 적용하며 이동하고 있다면, 가상 공간과 물리 공간에서의 위상 차이가 발생해 가상 공간과 물리 공간에서 동시에 직접적인 상호작용 하는 데에 문제가 발생한다. 우리는 가상 공간과 물리 공간에서의 상대적인 위치와 방향 차이를 조정해 가상과 물리 공간에서 동시적으로 직접적인 상호작용을 보장해주는 회복 알고리즘에 대해서 제안했고, 사용자 실험을 통해 회복 알고리즘을 사용했을 때 사용자들은 더 높은 현존 감과 공존감을 느낄 수 있었다.

1. 서론

VR 기술의 발전으로, 다중 사용자 VR 콘텐츠가 활발히 소비되고 있다. 다중 사용자가 가상 공간을 공유할 때, 다양한 상호작용이 발생한다. 직접적인 상호작용은 중간 매개체 없이 사용자가 물체를 직접 상호작용하는 것으로 사용자가 가상의 물체를 만지거나 잡는 행위 등을 의미한다. 우리는 두 사용자가 같은 가상 공간과 물리 공간을 공유하는 상황(SVPE)에서 RDW가 적용될 때 발생하는 직접적인 상호작용 문제에 대해 다루었다.

RDW[1]는 사용자가 인지하지 못 할 만큼 가상 공간을 왜곡시켜 사용자의 실제 이동한 경로에 변화를 준다. 따라서, SVPE에서 두 사용자에게 각각 RDW를 적용하면 (그림 1)과 같은 문제가 발생한다. 두 사용자가 가상과 물리 공간에서 같은 위치에서 시작하더라도 시간이 지나면 위치와 방향에 상대적인 차이가 발생한다. 결국, 가상 공간에서 두 사용자가 악수와 같은 직접적인 상호작용을 위해 만나더라도 실제 공간에서는 만나지 못하는

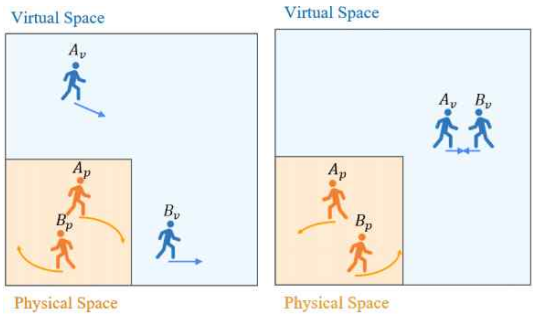


그림 1: RDW에 의해 두 사용자(A, B)의 물리 공간과 가상 공간 사이의 위치와 방향 차이 발생

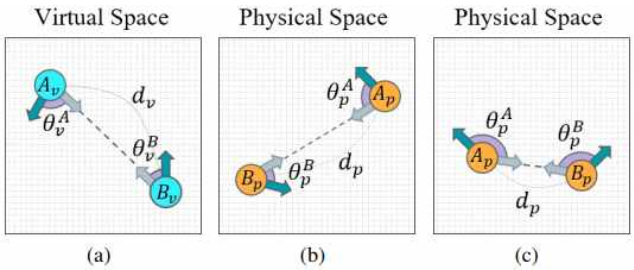


그림 2: 가상 공간에서 (a)와 같이 있을 때, (b)는 회복된 상태이고 (c)는 회복된 상태가 아니다.

상황이 발생해 실제로는 악수를 하지 못하게 된다. 우리는 회복 기술을 통해 위와 같은 문제를 해결하는 방안을 제안한다.

2. 방법

2.1. 회복된 상태

회복된 상태는 가상 공간과 물리 공간에서의 다중 사용자의 상대적인 위치와 방향이 같은 상태를 의미한다. 두 사용자의 경우 회복된 상태를 만족하기 위한 조건은 $d_v = d_p$, $\theta_v^A = \theta_p^A$, $\theta_v^B = \theta_p^B$이다(그림 2 참조).

2.2. 회복 알고리즘

회복 알고리즘은 두 사용자를 회복된 상태로 만들기 위한 일련의 과정으로 현재 사용자의 가상과 물리 공간에

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 IEEE VR 2020 에 발표되었음.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-2018-0-01419')

서의 상태와 관계없이 항상 회복된 상태를 보장한다. 오직 두 사용자만을 고려하며, 가상에서 서로 마주 보고 직선으로 걸어오는 상황을 고려한다. 회복 알고리즘은 가상 공간의 두 사용자 사이의 거리(d_v)를 기준으로 나타날 수 있는 모든 경우에서 회복된 상태를 만들기 위한 적절한 방법을 제공한다.

Case 1 $d_v > l_p$: d_v가 사각형 물리 공간의 한 변의 길이(l_p)보다 큰 경우 회복된 상태를 만들기 위해서는 가상의 두 사용자 사이의 거리가 먼저 좁아져 물리 공간의 크기 내로 들어와야 한다. 이 경우, Steer-to-Center (S2C)[1]와 같은 기존의 알고리즘을 사용하여 장애물과의 충돌을 최대한 줄이는 데에 초점을 두어 RDW를 적용한다.

Case 2 $\min(g_t)d_p \le d_v \le l_p$: 가상에서 두 사용자가 실제 공간 내의 크기만큼 가까워지고 어느 정도 충분한 간격이 있는 경우(Case 3 참고) 사용자들을 아래와 같은 2차 최적화 문제를 해결하여 RDW를 적용한다.

$$\text{minimize } F(g_t^A, g_c^A, g_t^B, g_c^B) = w_1(d_v - \hat{d}_p)^2 + w_2\left\{(\theta_v^A - \hat{\theta}_p^A)^2 + (\theta_v^B - \hat{\theta}_p^B)^2\right\}$$

$$\text{subject to } \min(g_t) \le g_t^i \le \max(g_t), \quad i = A,B$$

$$-\max(g_c) \le g_c^i \le \max(g_c), \quad i = A,B$$

우리는 매 순간 translational gain(g_t)값과 curvature gain(g_c)값[2]을 이용해 다음 순간의 물리 공간 내 두 사용자 사이의 거리($\hat{d}_p$)와 상대적인 각도($\hat{\theta}_p^A, \hat{\theta}_p^B$)를 구할 수 있다. 그 후, 위의 문제를 풀어 최대한 빠르게 회복된 상태를 만들기 위한 임곗값 내의 최적의 gain을 구해 RDW를 적용한다. 이처럼 진행하다보면, $d_v = d_p$인 순간이 오는데 이때에는 angle alignment reset을 시행한다. Angle alignment reset은 rotation gain을 이용하여 가상과 물리 공간에서의 상대적인 각도(θ_v, θ_p)를 맞춰주는 방법이다.

Case 3 $d_v \le \min(g_t)d_p$: 가상에서 두 사용자가 사이의 거리가 실제보다 훨씬 가까워진 경우에는 gain을 최대로 적용해도 가상에서 만났을 때 실제로 만남을 보장하지 못하게 된다. 이 경우 사용자가 회복 과정을 알아차리더라도 강제로 사용자의 위치를 조정하는 overt recovery 기술을 사용한다. Overt recovery 기술은 freeze-backup, distractor, change blindness, seven-league boots와 같은 RDW의 명시적 기술[2]들의 원리를 이용해 회복된 상태를 강제로 보장한다.

3. 실험 및 결과

3.1 사용자 실험

우리는 회복 알고리즘에 대한 사용자들의 반응을 확인하기 위해 사용자 실험을 진행했다. 회복 알고리즘을 사용한 경우와 S2C 알고리즘을 사용해 회복된 상태를 보장하지 않은 경우에 대해 실험을 진행했다. 총 20명(남:18명, 여:2명)을 대상으로 실험을 진행했으며, 평균 나이는 25세였다(21세~26세). 각 실험이 끝난 후 멀미 정도 (SSQ), 시스템 사용성 (SUS[3]), 현존감 (SUSPQ[4]), 공존감 (CPQ[5])에 대한 설문을 진행했다.

3.2 사용자 실험 결과

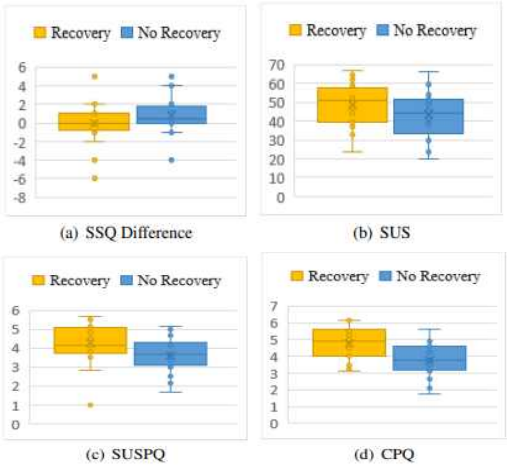


그림 3: 사용자 실험 결과

멀미 정도는 통계적으로 유의미한 차이는 없었다. 나머지 항목에는 모두 회복을 사용한 경우 더 높은 점수를 받았으며, 통계적으로 유의미한 차이가 있었다. 즉, 회복 기술은 사용자들에게 더 높은 시스템 만족도와 현존감을 제공한다. 특히, 공존감 설문의 결과를 통해 회복 기술은 사용자들에게 상호작용 시 실제로 상대방과 같이 존재하는 느낌을 더 많이 주는 것을 확인할 수 있다.

4. 결론

본 연구는 다중 사용자가 같은 가상과 물리 공간을 공유하는 상황에서 RDW를 적용할 때 직접적인 상호작용이 가능하도록 하는 회복 기술을 제안했다. 이를 통해 사용자들은 가상에서 더 높은 현존 감을 느낄 수 있었고, 다른 사용자와 직접적인 상호작용 시 공존감을 더 많이 느낄 수 있었다. 세 명이상의 회복 기술 등이 앞으로 더 연구되어야 할 것이다.

참고문헌

[1] S. Razzaque. Redirected Walking. University of North Carolina at Chapel Hill, 2005.

[2] E. A. Suma, G. Bruder, F. Steinicke, D. M. Krum, and M. Bolas. A taxonomy for deploying redirection techniques in immersive virtual environments. In 2012 IEEE Virtual Reality Workshops (VRW), pp. 43–46. IEEE, 2012.

[3]] J. Brooke et al. Sus-a quick and dirty usability scale. Usability evaluation in industry, 189(194):4–7, 1996.

[4] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. Presence: Teleoperators & Virtual Environments, 9(5):497–503, 2000.

[5] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. ACM Transactions on Computer-Human Interaction (TOCHI), 7(4):443–460, 2000.

다중 사용자 환경과 비정형 공간에서의 강화학습 기반 무한공간탐험 최적 계획법*

이동용[0], 조용훈, 민대홍, 이인권
연세대학교 컴퓨터과학과
pfldy2850@gmail.com, maxburst@gmail.com, bingo904@yonsei.ac.kr, iklee@yonsei.ac.kr

Optimal Planning of Redirected Walking Based on Reinforcement Learning in Multi-user Environment with Irregularly Shaped Physical Space

Dong-Yong Lee[0], Yong-Hun Cho, Dae-Hong Min, In-Kwon Lee
Dept. of Computer Science, Yonsei University

요약

무한공간탐험법(Redirected Walking: RDW)은 사용자의 실제 보행을 기반으로 가상환경을 탐험하는 기법으로, 가상 공간 내의 다른 이동 방법과 비교하여 가장 높은 현존감을 사용자에게 제공한다. 본 논문에서는 강화학습에 기반한 무한공간탐험 최적 계획법인 Multi-user-Steer-to-Optimal-Target(MS2OT)을 제안하고, 기존 알고리즘들과의 성능 비교 실험을 진행했다. MS2OT는 3D Convolutional Neural Network (3D CNN)로 구성된 Deuling Double Deep Network (D3QN) 구조로 이루어져 있다. 성능 비교 실험 결과, MS2OT를 사용한 경우, 기존 방법들과 비교하여 물리 공간 경계와의 충돌 (reset) 횟수가 크게 감소되었으며, 사용자의 탐험 거리도 유의미하게 향상되었다.

1. 서론

몰입형 가상현실 시스템에서, 가상공간과 물리공간에서 동시에 걷는 환경을 제공하려 할 때, 공간 크기의 차이로 인해 사용자에게는 물리적 충돌의 가능성이 있다. 무한공간탐험법(Redirected Walking: RDW)은 가상공간의 사용자 시점에 인지할 수 없는 회전 왜곡 등을 통해 물리적 탐험 경로를 제어하며, 사용자를 항상 물리공간의 중심으로 유도한다[1]. 본 논문은 직사각형이 아닌 비정형 물리공간과 다중 사용자 환경에 대해서 실시간으로 적용할 수 있는 (그림 1) 방법인 MS2OT를 제시한다. MS2OT는 기존의 강화학습 기반 방법론인 S2OT[2]를 비정형 물리공간과 다중 사용자 환경에 적용될 수 있도록 확장하여, 더 많은 수의 조향 목표 지점과 행동 유형을 가지며, 개선된 보상 함수가 사용되었다. 3D CNN을 사용하고, 물리 공간과 다중 사용자를 이미지 데이터에 표현하여, 일반성을 높였다.

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 IEEE VR 2020 에 발표되었음.
* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. NRF-2020R1A2C2014622) 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-2018-0-01419')

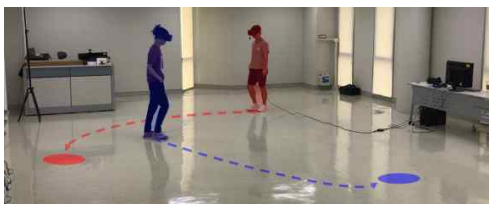


그림 1: MS2OT를 이용한 무한공간탐험

다른 방법과의 성능 비교를 위해, 우리는 사용자 실험과 시뮬레이션 실험을 모두 수행했다. MS2OT는 S2C, S2OT, APF와 비교하여 전체 reset횟수를 유의미하게 감소시켰다. 특히 사용자간 충돌을 크게 감소시켰으며, 최대 32명의 사용자를 실시간으로 처리할 수 있었다.

2. 방법

2.1. 상태 (State)

강화학습의 state는 0.02초 간격으로 집계된 64개의 연속된 이미지 데이터로, 사용자 위치 정보와 다른 사용자들의 위치 정보로 이루어진 두 채널 및 물리 공간을 나타내는 세 번째 채널로 이루어진다.

2.2. 행동 (Action)

조향행동과 재설정행동의 두 가지 행동 유형을 사용한다. 조향행동은 사용자를 지정된 조향목표로 유도하는 행동으로써, 32×32 그리드로 이루어진 물리공간의 지점들로 사용자를 유도한다. 이 과정에서 사용자들이 인지할 수 없을 정도의 회전 왜곡을 사용기 때문에, 사용자들은 물리 탐험경로의 왜곡을 인지하지 못한다. 재설정행동은 충돌이 발생하지 않았음에도 사용자의 물리방향을 재설정하여 미래의 충돌 가능성을 줄이게 한다.

2.3. 보상함수 (Reward Function)

보상함수는 사용한 행동의 유형과 그로 인해 발생된 사용자의 안전성을 기반으로 설정되며, 조향행동 보상

$R_s = R_w + R_u + R_r + 5$과 재설정행동 보상 $R_p = 0$으로 이루어져 있다. 특히, 조향행동 결과로 인해 충돌이 발생한다면, 전체 보상함수에 -10의 감점을 적용한다. $R_w = (\sum \| p - l_j \|)/M_w$는 사용자와 물리공간의 벽간의 안전성에 대한 보상함수이며, p는 사용자의 위치, l_j는 벽 세그먼트, M_w는 사용자와 벽간의 최대거리이다. 따라서 R_w는 사용자와 벽과의 거리가 멀어질수록 높은 보상을 할당한다. $R_u = (\sum \| p - o_j \|)/M_u$는 사용자간의 안전성에 대한 보상함수이며, o_j는 다른 사용자의 위치, M_u는 사용자 간 최대거리이다. 마지막으로 $R_r = 1 - \| S_{a_t} - S_{a_{t-1}} \| /32\sqrt{2}$는 조향행동의 왜곡지수에 대한 보상함수이며, S_{a_t}는 t시점에서 사용된 조향행동 S_a의 조향목표 위치이다. 조향행동 왜곡지수 보상함수는 조향왜곡 변동량이 적을수록 높은 보상을 제공한다.

2.4. Q-네트워크 (Q-Network)

Q-Network[3]는 상태 이미지정보에 사용자 보행정보를 GRU 레이어와 타일링 방법을 통해 임베딩하여 64×64×64×11 크기의 입력정보를 생성한다. 생성된 입력정보는 4개의 3×3×3 컨볼루션 레이어와 LeakyReLU, 그리고 2×1×1 MaxPool 레이어를 거쳐 공통 속성정보로 변환된다. 이렇게 변환된 속성정보는 서로 다른 두 개의 네트워크를 통해 adv와 val 점수로 변환된다. adv 점수는 행동들이 상태에 적합한지에 대한 평가 점수이며, val 점수는 상태 자체에 대한 적절성을 평가하는 점수이다. 행동을 선택하는 Q 함수는 adv와 val의 합으로 이루어진다; $Q = adv + val$ (그림 2).

3. 실험 및 결과

3.1 사용자 실험

사용자 실험에서 사용자들은 미리 구성된 $1400m^2$ 크기의 가상공간에서 자유롭게 탐험한다. S2C, S2OT, APF, MS2OT의 성능을 비교 사용하였고, 1명과 2명의 사용자 수의 경우를 실험하였다. 실험 결과에 대한 분석은 ANOVA test를 진행하였고, 사후검정으로 Tukey's HSD test를 실시하였다. 단일 사용자 환경에서는 MS2OT ≈ S2OT ≈ APF < S2C, 다중 사용자 환경에서는 MS2OT ≈ APF < S2OT ≈ S2C 순으로 MS2OT가 가장 작은 reset횟수를 기록하였으며, SSQ 설문으로 멀미 정도의 변화는 거의 없었다.

3.2 시뮬레이션 실험

시뮬레이션 실험에서는 미리 수집된 44개의 실제 사용자 보행경로를 이용하여 시뮬레이션이 구현되었으며, 물리공간의 크기와 모양에 따른 분리하여 테스트가 수행되었다. 네가지 방 크기와 2명에서 10명의 시뮬레이션 사용자에 대해 실험하였고, (4개 알고리즘)×(4개 크기)×(9사용자수)의 전체 충돌횟수를 ANOVA 테스트한 결과, 알고리즘과 방의 크기와 사용자 수 모두 유의미한 효과가 있었고, 각각의 상호작용 효과 또한 모두 유의미

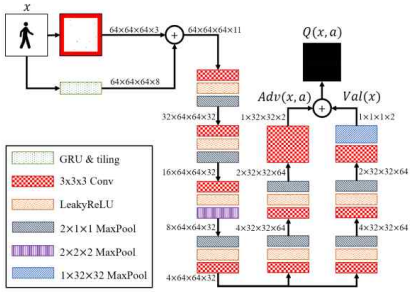


그림 2: MS2OT 강화학습 모델

했다. 사후검정을 통하여 전체 충돌량은 MS2OT < APF < S2OT < S2C 순이었음을 확인하였다. 사용자 수와 방 크기에 따른 각 알고리즘의 회귀분석 결과이며:

- S2C: $T = -0.003S + 0.748U + 3.723, (R^2 = .871)$
- S2OT: $T = -0.003S + 0.670U + 3.383, (R^2 = .841)$
- APF: $T = -0.002S + 0.441U + 2.367, (R^2 = .655)$
- MS2OT: $T = -0.002S + 0.374U + 2.267, (R^2 = .652)$

여기서, T는 충돌횟수, S는 방의 크기, U는 사용자의 수이며, MS2OT는 사용자 수가 증가함에 따른 충돌횟수 증가량이 0.374로 가장 작은 수준이었다.

방의 모양에 대한 실험은 십자형, L자형, T자형, 직사각형, 사다리꼴, 삼각형의 여섯 가지에 대해서 비교하였다. 이 중 T자형과 삼각형 두 가지 모양은 학습에 사용되지 않았던 것이다. 4(알고리즘)×6(방의 모양)의 충돌량에 대한 ANOVA 검정 결과, 알고리즘과 방의 모양에는 모두 유의미한 효과가 있었고, 상호작용 효과 또한 유의미했다. 사후검정 결과로 T자형을 제외한 모든 모양에서 총 충돌량은 MS2OT < APF < S2OT < S2C 순으로 증가했다.

MS2OT의 연산시간을 회귀분석을 통해 분석한 결과는 $T = 0.3992U + 6.8668$ 이며, 이때 U는 사용자의 수이다. 이에 따르면, 50fps 기준 컨텐츠의 경우에 MS2OT는 최대 32명을 실시간으로 처리할 수 있었다.

4. 토의 및 결론

본 연구에서는 무한공간탐험법을 다양한 모양의 공간에서 다중 사용자 환경으로 확장한 강화학습 기반 방식으로 MS2OT를 제시했다. 발전된 보상함수를 사용하여 이미지 처리된 상태정보에 조향, 재설정행동을 최적 선택함으로써 충돌을 피하는 최적 무한공간탐험 기법을 제공했다. 성능을 평가하기 위해 사용자 및 시뮬레이션 실험을 했고, MS2OT는 타 방식들에 비해 유의미하게 낮은 충돌 횟수를 발생시키며 가장 좋은 성능을 보였다.

참고문헌

[1] S. Razzaque. et. al, *Redirected Walking*. University of North Carolina at hapel Hill, 2005.

[2] D. Lee, et. al, "Real-time Optimal Planning for Redirected Walking Using Deep Q-Learning", *IEEE VR* 2019.

[3] C. J.Watkins and P. Dayan. “Q-learning”, *Machine learning*, 8(3-4):279-292, 1992.

인터랙티브 캐릭터 제어를 위한 시간 임계적 반응 학습*

이경호[1], 민세희[2], 이선민[20], 이제희[2]

[1]NCSoft
[1]whcjs13@ncsoft.com

[2]서울대학교
[2]{sehee, sunmin.lee, jehee}@mrl.snu.ac.kr

Learning Time-Critical Responses for Interactive Character Control

Kyungho Lee[1], Sehee Min[2], Sunmin Lee[20], Jehee Lee[2]
[1]NCSoft
[2]Seoul National University

요약

상호작용이 가능한 캐릭터 애니메이션에서, 사용자의 입력에 따라 빠르게 반응하는 캐릭터 제어기를 만드는 것은 중요한 문제이다. 최근 많은 캐릭터 애니메이션 연구에서 딥러닝을 이용한 상호작용가능한 캐릭터 제어기를 제안하였다. 캐릭터 제어기의 반응성은 사용한 데이터셋과 학습 방법, 이 때 사용한 여러 파라미터들에 의해 결정되므로, 반응성을 사용자가 원하는 대로 조절하는 것은 복잡한 문제이다. 본 연구는 사용자가 캐릭터 제어기의 반응 시간을 조절하고 이를 보장할 수 있는 적응적 학습 프레임워크를 제안한다.

1. 서론

사용자의 입력에 반응하여 실제 사람과 같이 자연스럽게 움직이는 캐릭터 제어 시스템을 만드는 것은 컴퓨터 애니메이션 분야에서 오랫동안 연구되어 온 주제이다. 특히 최근에는 딥러닝을 활용한 연구들이 많은 성과를 보이고 있다. 하지만 딥러닝은 종종 결과를 분석하거나 원하는 방향으로 수정하기 어려운 블랙박스처럼 여겨진다. 실시간 캐릭터 제어 시스템은 움직임의 퀄리티, 반응성, 다양성과 같은 다양한 지표로 평가될 수 있다. 각 지표는 사용한 데이터, 학습 모델 및 알고리즘과 이 과정에서 사용된 여러 파라미터에 의해 복합적으로 결정된다. 따라서 우리가 원하는 수준을 만족하는 캐릭터 제어기를 만드는 것은 복잡한 문제이다.

* 구두발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 현재 타 학술대회(논문지)에 제출중.
* 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW컴퓨팅산업원천기술개발사업(SW스타랩)의 연구결과로 수행되었음. (IITP-2017-0-00878)

우리는 여러 지표 중 특히 캐릭터의 반응성에 초점을두고 연구를 진행하였다. 본 연구는 '시간 임계적 반응성'의 새로운 개념을 정의하고, 이를 원하는 수준으로 향상하는 적응적 학습 프레임워크를 제안한다. 이 학습 프레임워크를 통해 생성된 캐릭터 제어기는 사용자의 입력 후 '임계 반응 시간' 내에 캐릭터가 반응함을 보장하며, 임계 반응 시간을 사용자가 조절할 수 있다.

2. 시간 임계적 반응성

사용자의 제어 입력 후 캐릭터가 사용자의 입력을 만족하도록 움직이는데 걸린 시간을 '완료 시간', 완료시간의 기대값을 캐릭터 혹은 캐릭터 제어기의 '반응성'이라고 정의한다. 캐릭터가 무조건 높은 '반응성'을 가지기를 원한다면, 하나의 쉬운 해법은 모션 데이터를 실제보다 빠르게 배속 재생하는 것이다. 하지만, 이는 부자연스러운 움직임을 생성하고 이미 충분히 빠른 부분조차 불필요하게 배속한다는 단점이 있다.

이를 보완하고자 우리는 새롭게 '시간 임계적 반응성'을 정의하고 이를 향상하고자 한다. 사용자가 원하는 캐릭터의 완료 시간을 '임계 반응 시간'이라고 하고, '초과 시간'을 max(완료 시간- 임계 반응 시간,0)으로 정의한다. 이 때, 캐릭터 제어기의 '시간 임계적 반응성'을 초과 시간의 기대값으로 정의한다. 우리의 목적은 이 값을 최소화하면서 최대한 자연스럽게 움직이는 캐릭터 제어기를 만드는 것이다.

3. 시스템 개요

임계 반응시간 내에 사용자의 요구를 완료하는 캐릭터 제어기를 만들려면 모션캡처 데이터를 그대로 사용하거나 데이터베이스를 랜덤워크하여 생성된 모션 시퀀스로 지도학습을 하는 기존의 방법과 다른 접근이 필요하다. 우리의 시스템은 크게 모션 시퀀스 생성, 생성된 모션 시퀀스를 이용한 제어기의 지도 학습, 제어기 평가 및

적응적 개선의 세 요소로 구성된다(그림1 참조). 먼저, 사용자의 입력 분포를 가정하고, 사용자의 입력에 따라 임계 반응 시간 내에 주어진 임무를 완료하는 가장 자연스러운 모션 시퀀스를 모션 캡쳐 데이터베이스로부터 생성한다. 그 다음, 생성된 모션 시퀀스를 [1]에서 제안한 다목적 제어모델을 활용하여 RNN으로 제어기를 학습한다. 학습 후에는 이 제어기가 사용자의 입력에 따라 실제로 임계 반응 시간 내에 주어진 임무를 완료하는지를 평가하고, 임계 반응 시간 내에 하지 못하는 상황에 대해 학습 데이터의 분포를 조정하여 추가적인 학습을 진행한다. 모든 상황의 사용자 입력에 대해 시간 임계적 반응성이 보장될 때까지 이 일련의 과정을 반복하여 제어기를 향상한다.

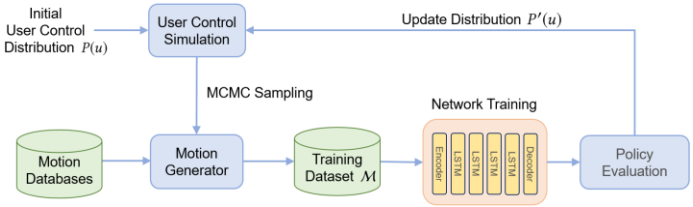


그림 1: 시스템 구성도

4. 모션 생성

캐릭터의 초기 상태와 사용자의 입력이 주어졌을때, 모션 생성기는 초기 상태에서부터 사용자의 입력을 만족할 때까지의 하나의 에피소드 모션 시퀀스를 생성한다. 시간 임계적 반응성을 만족하면서도 최대한 자연스러운 움직임을 생성하기 위해, 주어진 임계 반응 시간보다 조금 긴 시간 제한을 두고 모션 시퀀스를 생성한 후 이를 변형하는 방식을 택한다.

모션 시퀀스를 생성하는 방식은 모션 데이터베이스의 모든 포즈 중 가장 적절한 포즈를 매 N 프레임마다 고른다는 점에서 모션 매칭[2]과 비슷하다. 동작 종류와 방향, 위치, 속도 등의 입력을 동시에 만족하는 에피소드를 만들기 위해 모든 프레임에서 특정 동작을 수행하는 비용을 미리 계산해둔다. 이를 바탕으로 주어진 동작의 수행여부, 모션의 자연스러움, 위치, 방향, 속도 등의 오차, 시간 초과 등을 평가하여 데이터 베이스에서 가장 적절한 포즈를 골라 모션 시퀀스를 생성한다.

생성된 모션 시퀀스가 임계 반응 시간보다 긴 경우 차선의 부분 모션 시퀀스를 찾아 임계 반응 시간 내에 끝나도록 라플라시안 모션 에디팅[3]으로 변형한다.

5. 평가 및 적응 학습

모션 시퀀스 생성 단계에서 사용자의 입력에 시간 임계적 반응성을 보장하도록 모션 시퀀스를 생성하였음에도, 학습 데이터 분포의 불균형으로 학습 결과 이를 만족하지 못할 수 있다. 시간 임계적 반응성을 확실히 보장하기 위해서는 실제로 학습된 제어기를 평가하고 이를 토대로 사용자의 입력 분포를 조정하여 학습을 반복하는 적응적 알고리즘이 필수적이다.

제어기를 평가하기 위해 랜덤한 입력에 대한 완료 시간을 측정하고 커널 밀도 추정을 통해 사용자의 입력에 대한 확률 밀도 함수와 특정 입력에 대한 완료시간의 기대값을 추정한다. 완료시간의 기대값이 임계 반응시간을 초과하는 경우, 사용자의 입력 밀도함수를 이에 비례하게 설정한다. 이 과정을 통해 사용자 입력의 밀도함수가 변경되면 MCMC(Markov Chain Monte Carlo) 샘플링을 통해 주어진 함수에 근사하는 사용자의 입력을 생성할 수 있다. 조정된 사용자 입력 분포로부터 생성된 모션 시퀀스로 다시 학습을 진행하고 평가하는 일련의 과정을 반복적으로 진행하여 제어기를 향상한다.

6. 결과 및 결론

본 논문은 사용자의 입력에 따라 임계 반응 시간 내에 반응함을 보장하며, 임계 반응시간을 조절할 수 있는 시스템을 제안하였다. 이를 통해 방향, 위치, 속도, 동작 종류 등을 제어할 수 있다. 우리는 보행뿐만 아니라 역동적인 무술 동작, 장애물을 피하기 위한 파쿠르 동작, 아티스트가 만든 과장된 점프 동작 등의 예제를 통해 시스템이 잘 동작함을 확인하였다(그림2 참조).

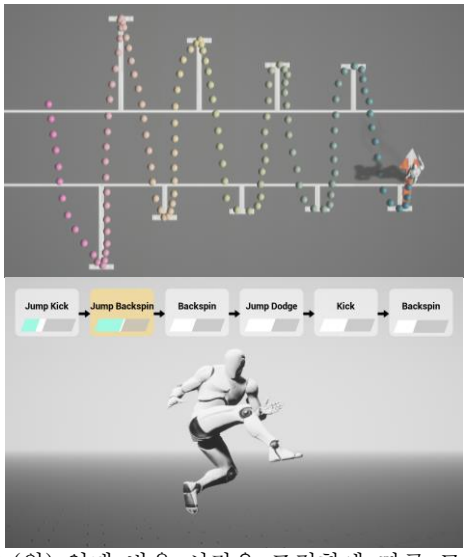


그림 2: (위) 임계 반응 시간을 조절함에 따른 모션 변화
(아래) 무술 예제

본 논문은 시간 임계적 반응성에 집중하여 캐릭터 제어기를 평가하였으나, 본 논문에서 제시한 평가를 통한 적응적 학습 프레임워크는 다른 지표를 만족하는 제어기를 만들기 위해서도 범용적으로 활용될 수 있다. 또한 캐릭터 제어기로 다목적 제어모델과 RNN 네트워크를 사용하였는데 이와 무관하게 활용될 수 있다.

참고문헌

[1] Kyungho Lee, Seyoung Lee, and Jehee Lee, Interactive character animation by learning multi-objective control, ACM Trans. Graph, 37(6): Article 180, 2018.

[2] Simon Clavet. Motion Matching and The Road to Next-Gen Animation. GDC, 2016.

[3] Manmyung Kim, Kyunglyul Hyun, Jongmin Kim, and Jehee Lee. 2009. Synchronized multi-character motion editing. ACM Transactions on Graphics 28, 3, Article 79 (2009).

변형 가능한 모델의 충돌 탐지를 위한 효율적인 법선 원뿔 컬링 방법*

이창진[0], 한동훈, 이상빈, 고형석
서울대학교 전기정보공학부
{OSgood, dhhan, sblee, ko}@graphics.snu.ac.kr

Efficient normal cone culling method for collision detection of deformable model

Chang-Jin Lee[0], Dong-Hun han, Sang-Bin Lee, Hyeong-Seok Ko
Dept. of Electrical and Computer Engineering, Seoul National University

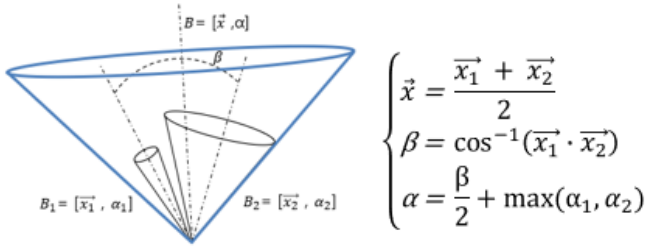
요약

본 논문에서는 변형 가능한 모델의 충돌 탐지에서 이용되는 법선 원뿔 컬링 방법[1][2]을 발전시켜 새로운 법선 원뿔 구성 방법론을 제시한다. 기존 법선 원뿔 구성 방법은 법선 원뿔을 구성함에 있어 불필요한 부분이 법선 원뿔로 구성되어 성능이 저하되는 측면이 있다. 따라서 좀 더 효율적인 법선 원뿔 구성 방법을 제시함으로써 더 효율적인 컬링을 가능하게 한다.

1. 서론

변형 가능한 모델 시뮬레이션에 있어서 충돌 탐지는 필수적인 요소이다. 또한 많은 연산과 시간을 필요로 하는 부분이다. 따라서 충돌 탐지 연산을 최소화하기 위해 법선 원뿔 컬링 방법[1][2]이 제시되었다.

기존 제시된 방법[1]에서는 그림 1과 같은 병합 방법을 이용하여 법선 원뿔 트리를 구성한다. 그림 1의 $B = [\vec{x}, \alpha]$의 표현은 법선 원뿔을 의미하며, $\vec{x}$는 법선 원뿔의 축을 의미하고 α는 법선 원뿔의 각도를 의미한다. 또한 그림 1의 B는 부모 원뿔을 의미하고, B_1, B_2는 각각 자식 원뿔을 의미한다. 기존 그림 1의 병합 방법을 이용했을 경우 부모 원뿔에는 불필요한 부분이 포함된다. 따라서 본 논문에서는 새로운 병합 방법을 제시한다.



$$\begin{cases} \vec{x} = \dfrac{\overrightarrow{x_1} + \overrightarrow{x_2}}{2} \\ \beta = \cos^{-1}(\overrightarrow{x_1} \cdot \overrightarrow{x_2}) \\ \alpha = \dfrac{\beta}{2} + \max(\alpha_1, \alpha_2) \end{cases}$$

그림 1: 기존 방법[1]의 법선 원뿔 병합 방법

* 구두 발표논문
* 본 연구는 교육과학기술부(MEST)의 지원을 받는 국립연구재단(NRF-2018M3E3A1057302)과 서울대 자동화시스템연구소(ASRI)의 지원으로 수행되었음.
* 현재 타 학술대회(논문지)에 제출 중

기존 제시된 방법[2]에서는 연속 충돌 탐지에 법선 원뿔 컬링 방법을 적용하였다. 그림 2는 한 타임스텝 $t \in [0,1]$ 안에서 삼각형의 움직임과 그에 따른 법선의 움직임을 나타낸다. 또한 법선의 경로를 감싸 구성된 법선 원뿔을 의미한다. 식(2)는 한 타임스텝 $t \in [0,1]$ 안에서 법선의 움직임을 베지어 식 형태로 전개한 식이다. 베지어 식의 특성상 3개의 제어점 $\boldsymbol{n_0}$, $\frac{\boldsymbol{n_0}+\boldsymbol{n_1}-\delta}{2}$, $\boldsymbol{n_1}$ 안으로 $\boldsymbol{n_t}$가 포함되게 된다. 이러한 특성을 이용하여 연속 충돌 탐지에서 법선 원뿔을 구성한다. 제어점의 볼록 껍질(convex hull)의 특성과 법선 원뿔 구성 시 앞서 방법[1]의 병합 방법을 이용하므로 불필요한 부분이 법선 원뿔에 포함된다. 따라서 본 논문에서는 볼록 껍질의 크기를 줄이고 방법[3]의 알고리즘을 적용함으로써 좀 더 효율적인 법선 원뿔 컬링 방법을 제시한다.

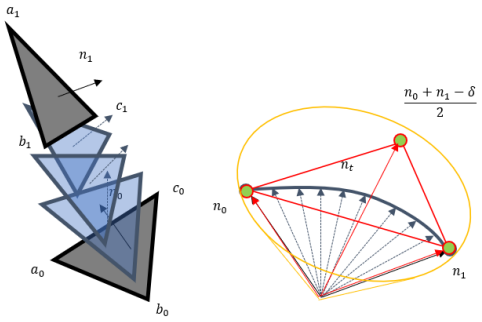


그림 2: 기존 방법[2]의 연속 충돌 탐지 법선 원뿔 구성

$$\delta = (b_1 - b_0 - a_1 + a_0) \times (c_1 - c_0 - a_1 + a_0) \quad (1)$$

$$\boldsymbol{n_t} = n_0 \cdot (1-t)^2 + \frac{(n_0+n_1-\delta)}{2} \cdot 2t \cdot (1-t) + n_1 \cdot t^2 \quad (2)$$

2. 효율적인 법선 원뿔 구성

2.1. 병합 방법

본 논문에서는 새로운 법선 원뿔 병합 방법을 제시한다. 자세한 병합 방법은 그림 3의 알고리즘1을 따르고 있다. 알고리즘은 두 자식 원뿔의 관계가 서로 어떤 포함관계에 있는지에 따라 다른 방식으로 처리하여 부모 원뿔을 구성한다.

알고리즘 1. Merging method $(B_{left\_child}[\overrightarrow{x_1}, \alpha_1], B_{right\_child}[\overrightarrow{x_2}, \alpha_2])$

```
Merging method(B_left_child[x1, α1], B_right_child[x2, α2]) :
    β = cos^-1(x1 · x2)
    α = (β + α1 + α2)/2
    weight_left = α − α1
    weight_right = α − α2
        if weight_left ≤ 0
            x = x1
            α = α1
        else if weight_right ≤ 0
            x = x2
            α = α2
        else
            x = sin(weight_right) * x1 + sin(weight_left) * x2
            x = normalize(x)
        end if
    return B_parent[x, α]
```

그림 3: 본 논문에서 제시하는 병합 방법

본 논문의 병합 방법을 적용해서 병합을 실시한 결과는 그림 4와 같다.

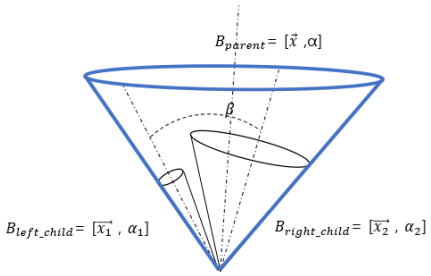


그림 4: 본 논문의 병합 방법을 이용한 병합 모습

2.2. 연속 충돌 탐지 법선 원뿔 구성 방법

본 논문에서는 식(2)의 3개의 제어점에 De cateljau's algorithm을 적용함으로써 그림 5와 같이 5개의 제어점 $\boldsymbol{n}_0$, $\frac{3\boldsymbol{n}_0+\boldsymbol{n}_1-\delta}{4}$, $\boldsymbol{n}_{0.5}$, $\frac{\boldsymbol{n}_0+3\boldsymbol{n}_1-\delta}{4}$, $\boldsymbol{n}_1$을 구한다. 5개의 제어점에 벡터집합의 최적화 원뿔을 구하는 알고리즘[3]을 적용함으로써 좀 더 효율적인 법선 원뿔을 구성한다.

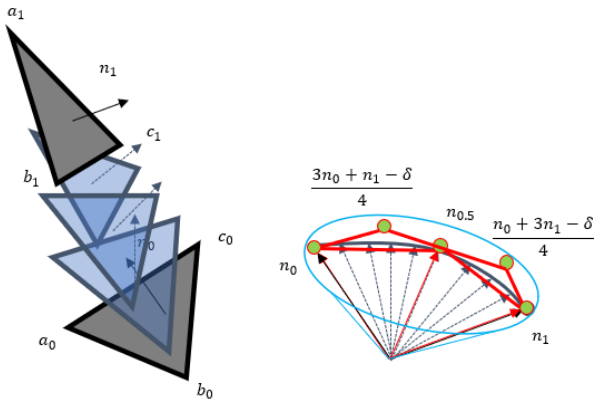


그림 5: 본 논문의 연속 충돌 탐지 법선 원뿔 구성

3. 결과

본 논문에서는 점 10,783개, 삼각형 20,849개, 선 31,633개로 이루어진 원피스를 가지고 실험을 진행하였다. 그림 6은 본 논문의 방법을 적용했을 때 시뮬레이션 전과 후를 나타낸다. 법선 원뿔 컬링 적용 후 충돌 가능성이 있는 삼각형 개수를 프레임별로 비교한 그래프는 그림 7과 같다. 또한 기존의 방법과 평균 충돌 가능성이 있는 삼각형 개수를 비교했을 때, 41,213개에서 28,832개로 감소하였다. 기존의 방법과 법선 원뿔 컬링 방법을 통한 실제 시간을 비교했을 때, 33.68ms에서 30.29ms로 약 10%정도 가속화 된 모습을 보였다.

충돌 가능성이 있는 삼각형의 개수가 41,213개에서 28,832개로 줄었음에도 불구하고 실제 걸린 시간의 가속화가 10% 정도에 머문 이유는 알고리즘[3]을 적용함에 있어 오버헤드가 생겼기 때문이다.

반면 그림 7을 분석한 결과 알고리즘[3]은 동적인 프레임에서 많은 효과를 얻었으므로 추후에는 동적인 프레임에서만 선택적으로 알고리즘[3]을 적용할 수 있게 함으로써 추가 연구를 진행할 계획이다.

그림 6: 원피스의 시뮬레이션 전과 시뮬레이션 후

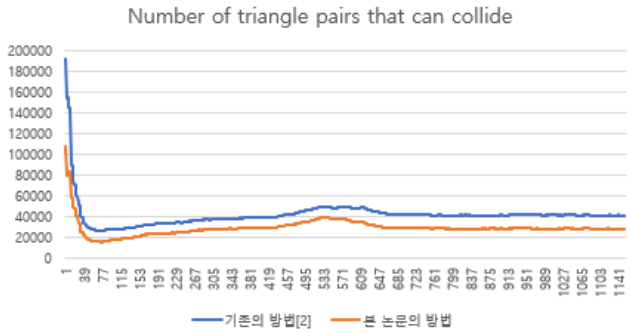


그림 7: 프레임별 법선 원뿔 컬링 적용 후 결과

참고문헌

[1] Provot X. Collision and self-collision handling in cloth model dedicated to design garments. In Computer Animation and Simulation' 97, Springer, Vienna, 177-189, 1997.

[2] Tang, M., Curtis, S., Yoon,S.-E, & Manocha, D. ICCD: Interactive continuous collision detection between deformable models using connectivity based culling. IEEE Transactions on Visualization and Computer Graphics 15 (4), 544-557, 2009.

[3] Barequet, Gill and Gershon Elber. Optimal bounding cones of vectors in three dimensions. Inf. Process. Lett. 93, 2, 83-89, 2005.

SketchScope: 터치스크린에서의 볼륨 데이터 탐구 인터페이스*

김현수[0], 박진아
한국과학기술원 전산학부
{khskhs, jinahpark}@kaist.ac.kr

SketchScope: A Volume Data Exploration Interface on a Touchscreen

Hyunsoo Kim[0], Jinah Park
School of Computing, KAIST

요약

본 논문은 3차원 볼륨 데이터의 내부를 부분적으로 들여다보며 손쉽게 탐구할 수 있도록 개발한 터치스크린 상의 인터페이스인 SketchScope를 소개한다. 본 연구는 direct manipulation widget을 스케치 기반의 인터페이스에 접목하여 볼륨 데이터를 처음 접하는 사용자도 쉽고 직관적으로 이용할 수 있도록 하였고, 특히 사용자가 스케치한 관심 부위 부근에 위젯이 위치할 수 있도록 ROI-based RANSAC을 고안하였다. 본 인터페이스를 사용하여 '한국의 미라'에 대한 컨텐츠를 제작하여 박물관에 프로토타입을 전시하였고 관람객들에게 새로운 체험이 가능하도록 하였다.

1. 서론

볼륨 데이터는 3차원 데이터라는 특성 때문에 조작이 어려워 일반인을 대상으로 하는 활용이 제한적이었다. 여러 연구에서 볼륨 데이터를 쉽게 탐구하는 방법을 제시하고 있으나, 조작의 자유도나 사용성에 한계가 있다 [1,2]. 본 논문에서는 관련 지식이 없는 일반인들이 터치스크린을 통해 볼륨 데이터의 내외부를 직접 탐구해 볼 수 있도록 고안된 스케치 기반의 볼륨 탐구 인터페이스 SketchScope를 제시한다.

Shneiderman이 제안한 direct manipulation[3] 은 이후 다양한 인터랙션 디자인에 영향을 끼쳤다. 특히, Stoppel 외 1인은 direct manipulation 방법론이 적용된 smart surrogate widget[4]를 사용하여 간단한 위젯으로 볼륨 데이터의 내부를 관찰할 수 있는 인터페이스를 제공할 수 있음을 보였다. 우리는 이러한 위젯 기반의 볼륨 탐구 인터페이스를 터치스크린에 적용하여 사용성을 높이고자 하였다. 특히, 터치스크린 상에서 가장 직관적인 조작 방식으로 많은 연구가 진행된 스케치 기반의 인터랙션 기법을 사용하였다 [5,6].

2. SketchScope

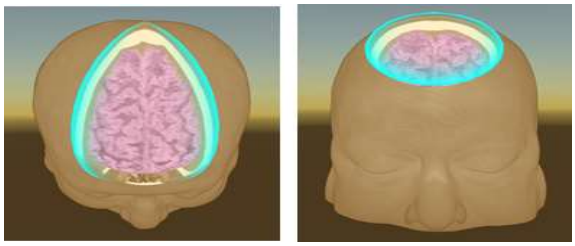


그림 1: 웨지 위젯(좌)과 아이리스 위젯(우)의 예시

SketchScope는 smart surrogate widget을 기반으로 하여, 쐐기 모양으로 볼륨을 잘라내는 웨지 (Wedge) 위젯과 원뿔모양으로 잘라내는 아이리스 (Iris) 위젯을 제공한다. 사용자는 위젯 생성, 위젯 편집 모드를 전환해 가며 위젯을 직접 구성하여 볼륨 데이터를 탐구할 수 있다.

2.1 위젯의 생성

SketchScope는 직관적인 위젯 생성을 위해 사용자가 볼륨을 잘라서 내부를 들여다보는 메타포를 차용했다. 사용자가 볼륨 위에 직선을 그리면 직선을 따라 잘려진 모양의 웨지 위젯이 생성되고, 원을 그리면 아이리스 위젯이 생성된다. 이러한 위젯의 예시는 [그림 1]에서 볼 수 있다. 위젯 생성 단계는 세부적으로 집중점 생성, 집중점 검색, 최적의 구형 위젯 생성, 위젯 타입 결정 단계로 나뉜다.

집중점 생성 단계에서는 Weiber 외 3인의 WYSIWYP[7]의 방식을 단순화여 집중점을 생성한다. 뷰포인트에서 볼륨으로 여러 레이를 생성하여 레이 위에서 볼륨의 불투명도가 극대인 집중점을 생성하는 것은 같지만, WYSIWYP이 한 레이에서 여러 집중점을 생성할 수 있는 반면에 SketchScope에서는 하나의 집중점만 생성할 수 있는 점에 차이가 있다. 레이 위에서 하나의 집중점을 찾는 알고리즘은 [그림 2]에서 볼 수 있다. 레이는 2차원 그리드로 균일하게 생성이 되며, [그림 2]의 알고리즘을 GPU로 실행하여 생성된 집중점의 3차원 좌표를 Congote 외 5인이 보여준 방식[8]과 같이 부호화하여 텍스처에 저장한다.

집중점 검색 단계에서는 사용자가 터치스크린 상에서 스케치를 진행할 때 터치 포인터가 가리키는 곳 인근의 집중점을 검색하여 관심 부위 집중점으로 지정한다. 집중점의 빠른 검색을 위해 octree가 사용되었다.

사용자가 터치 입력을 끝내면, 볼륨 전반적으로 분포되어있는 집중점과, 사용자의 스케치를 지나는 관심 부위

---

* 구두 발표논문
* 본 논문은 요약논문 (Extended Abstract)으로서, 본 논문의 원본 논문은 제9회 국제과학관심포지엄 (International Symposium of Science Museums 2019) 에 발표되었음.
* 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-과학문화전시서비스 역량강화지원사업의 지원을 받아 수행된 연구임 (NRF-2018X1A3A1069693).

```
Procedure find_focal_point(ray, volume) =
    Current_focal_point = null
    Start = ray.intersect(volume.bounding_box)
    Sample_point = start
    While sample_point is in volume do
        intensity = volume.sample(sample_point)
        opacity = transfer_function(intensity)
        if opacity is local maximum and rand() < γ then
            current_focal_point = sample_point
        end if
        sample_point += sample_interval
    end while
    return current_focal_point
```

그림 2: 집중점을 선택하는 알고리즘

집중점 정보를 종합하여 최적의 구형 위젯을 생성한다. 이를 위해 기존의 Random Sample Consensus(RANSAC) 알고리즘에 관심 부위에 대한 고려를 추가한 Region-of-interest(ROI)-based RANSAC을 고안하였다. ROI-based RANSAC의 목적함수는 다음과 같다:

$$\arg\max_{P_\theta \subset P} \sum_{p_i \in P} D(p_i|P_\theta,\ \delta) + \lambda \sum_{p_j \in P^*} D(p_j|P_\theta,\ \gamma). \qquad (1)$$

식 (1)에서 P는 전체 샘플(집중점)의 집합이고, $P^* \subset P$는 관심 부위 집중점의 집합, P_θ는 모델(구형 위젯)을 추정하기 위한 최소한의 부분집합이며, $D(p_i|P_\theta,\delta)$는 P_θ가 정의하는 모델에 대한 집중점 p_i의 Loss가 δ 이하이면 1, 초과이면 0을 출력하는 함수이다. ROI-based RANSAC은 전체 집중점에 대해 최적인 구를 검색하는 기존 RANSAC의 첫째 항과 더불어, 관심 부위 집중점에 대해 최적인 구를 구하는 둘째 항을 계수 λ를 곱해 추가하여 볼륨의 모양에 맞으면서 사용자의 관심부위에 위치한 구형 위젯을 생성할 수 있다.
최적의 구형 위젯을 생성하면 사용자가 입력한 스케치의 모양을 분석하여 웨지/아이리스 타입의 위젯을 생성한다. 이를 위해 스케치 위에 위치한 관심 부위 집중점을 구를 지나는 평면으로 정사형한 뒤 2D Principal Component Analysis를 통해 장축과 단축의 고윳값 $\lambda_1,\lambda_2\,(\lambda_1 \geq \lambda_2)$을 구한다. 고윳값을 계산하면 다음과 같이 스케치를 분류할 수 있다:

$$\frac{\lambda_1}{\lambda_2} \geq K \rightarrow Wedge,$$
$$\frac{\lambda_1}{\lambda_2} < K \rightarrow Iris.$$

여기서 K는 웨지와 아이리스 위젯을 구분하는 1보다 큰 계수이다. 위젯 타입이 아이리스 위젯일 경우, 스케치의 크기를 토대로 초기에 열린 각도를 결정하는 과정을 거친다.

2.2 위젯의 편집

SketchScope는 사용자가 위젯의 열린 정도를 조절하고, 삭제할 수 있는 위젯 편집 모드를 제공한다. Smart surrogate widget이 제안한 인터페이스[4]는 터치스크린보다 비교적 정밀한 마우스를 필요로 하여 터치스크린에 그대로 사용하기에 한계가 있다. 터치스크린에 적합한 인터페이스를 제공하기 위해, 위젯의 편집을 단순화하였다. 위젯 편집 모드에서는 위젯의 경계가 [그림 1]과 같이 하늘색으로 강조되는데, 사용자가 강조된 경계를 드래그하여 열린 정도를 조절할 수 있다. 사용자가 열린 정도를 완전히 없애면 위젯을 삭제한다.

3. 결론

본 논문은 일반인들에게 비교적 생소했던 볼륨 데이터와 볼륨 렌더링 기법에 대한 관련 지식이 없는 사람들도 직관적으로 데이터를 관찰할 수 있는 터치스크린 상의 인터페이스인 SketchScope를 소개하였다. 위젯을 통해 볼륨을 잘라 내부를 본다는 메타포를 차용하고, 자르는 과정을 스케치 기반 인터페이스를 통해 직관적으로 구현하여 전문가가 아닌 일반 사용자가 간단한 튜토리얼만으로 볼륨 데이터를 탐구할 수 있다.
우리는 충청남도 공주시의 한국자연사박물관과 함께 SketchScope를 적용한 미라 인체의 CT 볼륨 데이터 탐험 프로토타입 전시물을 개발하고 설치하여 리빙랩을 운영하였다. 실제 관람객 사용을 관찰한 결과, 직관적인 터치 인터페이스를 활용하는 SketchScope는 전시물에 대한 관심과 흥미를 높일 수 있음을 확인하였다. 하지만 위젯 생성/편집 간 모드 전환에 대한 사용성의 어려움에 대하여 양 모드를 통합하여 모드 전환을 제거하는 방향으로 인터페이스를 개선하면 사용성이 더 좋아질 것으로 기대된다.

참고문헌

[1] Luigi Gallo, Giuseppe De Pietro, and Ivana Marra, 3D Interaction with Volumetric Medical Data: experiencing the Wiimote, *Proceedings of the First International Conference on Ambient Media and Systems* (2008), PP.1-6.

[2] Daniel Jonsson, Martin Falk, and Anders Ynnerman, Intuitive Exploration of Volumetric Data Using Dynamic Galleries, *IEEE Trans. Vis. Comput. Graph. 22, 1* (January 2016), PP 896-905.

[3] Ben Shneiderman, Direct Manipulation: A Step Beyond Programming Languages, *Computer (Long. Beach. Calif). 16, 8* (August 1983), PP.57-69.

[4] Sergej Stoppel and Stefan Bruckner, Smart Surrogate Widgets for Direct Volume Manipulation, *IEEE Pacific Vis. Symp.* 2018-April, (2018), PP.36-45.

[5] Takeo Igarashi, Satoshi Matsuoka, and Hidehiko Tanaka, Teddy: A Sketching Interface for 3D Freeform Design, *In Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*, PP.409-416.

[6] Ivan E. Sutherland, Sketch pad: a man-machine graphical communication system, *In Proceedings of the SHARE design automation workshop on - DAC '64*, P.38.

[7] Alexander Wiebel, Frans M. Vos, David Foerster, and Hans Christian Hege, WYSIWYP: What you see is what you pick, *IEEE Trans. Vis. Comput. Graph. 18, 12* (December 2012), PP.2236-2244.

[8] John Congote, Alvaro Segura, Luis Kabongo, Aitor Moreno, Jorge Posada, and Oscar Ruiz, Interactive visualization of volumetric data with WebGL in real-time, *Proc. - 16th Int. Conf. 3D Web Technol. Web3D 2011*, PP.137–145.

초음파 비접촉식 SPH 유체 촉감 렌더링*

장재현[0], 박진아
한국과학기술원 전산학부
{jaehyunjang, jinahpark}@kaist.ac.kr

SPH Fluid Tactile Rendering for Ultrasonic Mid-air Haptics

Jaehyun Jang[0], Jinah Park
School of Computing, KAIST

요약

일상생활에서 손과 유체의 직접적인 상호작용을 통해 사용자는 유체의 특성, 손의 자세와 움직임에 따라 유체가 신체 표면에서 생성하는 다양한 압력 필드를 체험한다. 가상 환경 유체 시뮬레이션 내 사용자 상호작용을 통한 유체 시뮬레이션의 현실감을 개선하기 위해 SPH 유체 시뮬레이션에서 계산된 압력 필드를 초음파 비접촉식 햅틱 장치를 통해 유사한 음압 필드를 제공하는 유체 촉감 렌더링 기술을 제안한다. 유체와 상호작용 할 시 손의 자세와 움직임에 따라 변화하는 압력 필드를 제공하여 유체 시뮬레이션의 몰입감을 증대할 수 있었다.

1. 서론

대화형 가상현실 시뮬레이션에서 사용자 상호작용에 따른 사실감을 제공하기 위해 시각 및 햅틱 렌더링을 통해 멀티모달 피드백을 생성하게 된다. 유체 시뮬레이션 분야의 경우 스타일러스 기반의 유체 햅틱 렌더링 기법들이 제시되었다. [1][2] 이 중 6-자유도 유체 햅틱 렌더링 기법[2]은 햅틱 장치와 커플링 된 가상 프록시에 작용하는 힘과 토크를 질량 중심에서 해석하여 햅틱 인터페이스를 통해 역감 피드백으로 제공한다. 이를 통해 도구와 유체의 상호작용과 같은 시나리오를 사용자에게 전달할 수 있게 한다.

도구 상호작용과 더불어 사용자는 일상생활에서 유체를 신체와 직접 접촉하여 유체가 발생하는 다양한 압력 필드를 촉각적으로 체험하게 된다. 가상 환경 내 발생하는 유체의 압력을 사용자에게 제공하기 위해 최근 반-라그랑지안 유체 시뮬레이션과 초음파 비접촉식 햅틱 장치를 결합하여 가상 손 표면 위에 정의된 보로노이 영역에 대해 압력 필드를 계산하고 해당 압력 필드를 음압 필드로 제공하기 위해 다수의 집중점으로 추정하는 기법을 제시하였다. [3]

본 연구는 입자법에 기반하여 강체-유체 상호작용에 따른 유체의 압력 필드를 생성하고 이를 초음파 비접촉식 햅틱 장치를 통해 음압 필드로 제공하기 위한 다수의 초음파 초점을 생성하는 촉감 렌더링 기법을 제시한다. 이를 위해 유체 압력을 계산하는 영역을 포인트 클라우드로 획득하고, 유체 압력에 대해 지역 최댓값을 병렬 계산을 통해 획득한다. 획득한 지역 최댓값을 기준으로 초음파 초점을 변환 후 사용자는 음압 필드를 통해 유체의 촉감을 체험할 수 있게 된다.

2. 포인트 클라우드 획득 및 SPH 시뮬레이션

2.1. 포인트 클라우드 획득

유체 시뮬레이션 내 가상 손의 상호작용과 압력 필드 추정을 위한 표면 획득을 위해 가상 카메라는 그림 1과 같이 설정한다. 가상 카메라의 방향은 초음파 장치가 피드백을 제공하는 방향으로 지정하여 생성된 포인트 클라우드가 피드백 영역을 표현함과 동시에 SPH 시뮬레이션 내에서 유체 입자와 상호작용 할 수 있다.

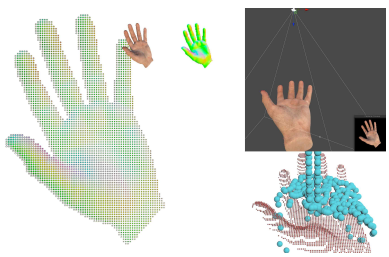


그림 1: 포인트 클라우드 획득

2.2. SPH 시뮬레이션

유체 시뮬레이션 및 강체-유체 상호작용을 계산하기 위해 SPH 연산 모델을 이용하였다. SPH 연산 모델은 식 1과 같이 매 프레임마다 주변 입자들의 물리량을 커널 함수에 대해 가중치를 적용하여 각 입자들의 물리량을 계산하는 방식이다. 본 연구에서는 나비에-스톡스 연산을 SPH 모델을 이용하여 계산하였다.

$$A(x_i) = \sum_j \frac{m_j}{\rho_j} A_j W(x_i - x_j, h) \quad (1)$$

---

* 구두(포스터) 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 IEEE Transactions on Haptics 13(1). 2020 에 게재되었음.
* 본 연구는 정보통신기획평가원의 HD 촉감 기술 기반 초실감 콘텐츠 재현 기술 개발 지원 (2017-0-00179)으로 수행되었음.

강체-유체 상호작용을 고려하기 위해 Akinci 외 4인[4]의 방식을 이용하여 강체에 적용되는 합력을 계산하였다. 각 입자에 적용되는 유체의 압력은 SPH 시뮬레이션으로 구한 합력과 포인트 클라우드 획득 과정에서 구한 법선의 내적이다.

3. 유체 촉감 렌더링

3.1. 지역 최댓값 연산 및 필터링

SPH 시뮬레이션으로 연산한 압력장 내 지역 최댓값들을 병렬 연산으로 획득한다. 그림 2와 같이 포인트 클라우드 내 각 입자에 대해 병렬 연산을 진행하며 순회 연산마다 이웃 입자를 기준으로 최댓값을 찾게 된다. 전체 순회 연산이 끝난 후 처음 순회를 시작한 위치와 마지막 위치가 동일한 경우 지역 최댓값으로 판단한다.

초음파 디스플레이는 최대 N개의 초음파 초점을 생성하여 진폭 변조를 할 수 있으므로 병렬 계산한 지역 최댓값들을 N개의 초점으로 필터링하게 된다. 최댓값부터 초점을 할당하기 위해 지역 최댓값을 정렬하고 기존의 초점들과 거리를 비교하여 지정된 거리 이상일 때 새로운 초점을 할당한다.

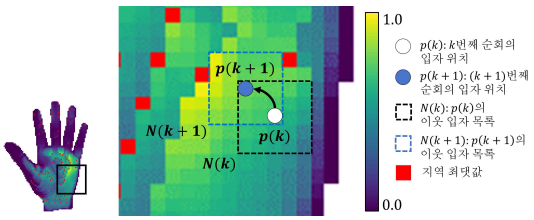


그림 2: 지역 최댓값 병렬 연산 커널

3.2. 음압 크기 변환

포인트 클라우드 내 정의된 압력 값은 SPH 시뮬레이션에 의해 결정되므로 실제 초음파 초점을 통해 음압으로 표현되기 위해서는 장치에서 정의된 압력 범위 내에서 결정되어야 한다. 음압 크기 변환을 위해 SPH 시뮬레이션 내에서 정의된 매개변수를 이용하여 표준화를 진행하게 된다. 표준화를 위해 본 연구에서는 유체 유동의 속도가 0인 정체 압력을 계산하여 이를 최댓값으로 간주한다.

$$P_{d_i} = \max\left(\frac{P_{r_i}}{\rho_0 g H + \frac{1}{2}\rho_0 v_{\max}^2}, 1.0\right) = \max\left(\frac{P_{r_i}}{P_s}, 1.0\right) \quad (2)$$

3.3. 적응적 변조

진폭 변조는 각 초점에 대해 각각 다른 변조 주파수를 할당할 수 있으며 모두 고정된 변조 주파수로 변조하는 것 대신 다른 변조 주파수로 할당하는 것이 구분감 있는 피드백을 줄 수 있다.[5] 적응적 변조는 새롭게 할당되는 초점 위치가 이전에 할당된 위치의 거리적인 차이가 적어 지속적으로 초점이 할당되는 경우(200 Hz)와 이외의 새로운 위치에 할당하는 경우(100 Hz)로 구분하여 변조 주파수를 제공한다.

4. 결과

제안한 유체 촉감 렌더링 기법은 Intel i9-9900 프로세서와 32 GB 램, Nvidia RTX 2080 (VRAM 8GB)의 환경에서 300 프레임 샘플을 획득하였을 때 SPH 시뮬레이션과 촉감 렌더링을 모두 합해 90-100Hz의 렌더링 레이트를 보여준다. 이는 부드럽고 안정적인 유체 햅틱 렌더링 업데이트율을 보여준다. [2]

그림 3은 총 4개의 유체 분사 상황에 대한 피드백 할당 결과이다. 생성된 유체 압력에 대해 유체 촉감 렌더링 기법이 생성한 초음파 초점 또한 이를 유사하게 추정함을 알 수 있다.

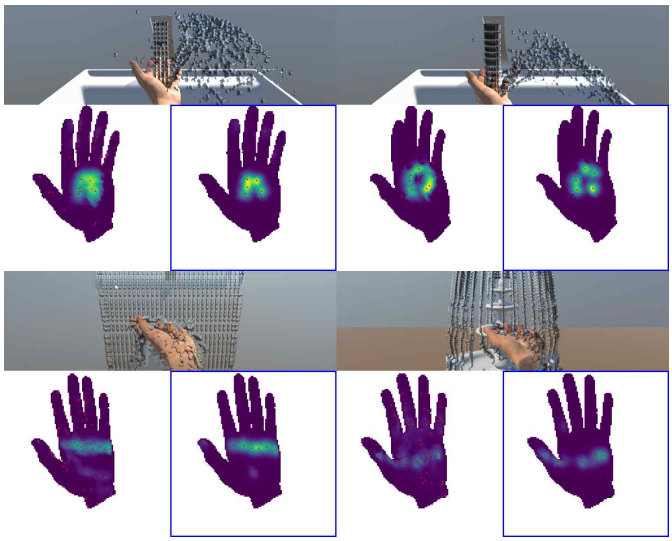


그림 3: 유체 분사 상황에 대한 유체 압력 필드 (좌) 초음파 초점 필드 (우, 파란 박스)

참고문헌

[1] M.Yang, J.Lu, A.Safonova, and K.J.Kuchenbecker, GPU methods for real-time haptic interaction with 3D fluids, 2009 IEEE International Workshop on Haptic Audio Visual Environments and Games, pp.24-29, 2009.

[2] G.Cirio, M.Marchal, S.Hillaire, and A.Lecuyer, Six degrees-of-freedom haptic interaction with fluids, IEEE Transactions on Visualization and Computer Graphics, 17(11):1714-1727, 2011

[3] H.Barreiro, S.Sinclair, and M.A.Otaduy, Ultrasound Rendering of Tactile Interaction with Fluids, 2019 IEEE World Haptics Conference, pp.521-526, 2019

[4] N.Akinci, M.Ihmsen, G.Akinci, B.Solenthaler, and M.Teschner, Versatile rigid-fluid coupling for incompressible SPH, ACM Transactions on Graphics, 31(4):1-8, 2012

[5] T.Carter, S.A.Seah, B.Long, B.Drinkwater, and S.Subramanian, Ultrahaptics, ACM Symposium on User Interface Software and Technology, pp.505-514, 2013

몬테카를로 렌더링의 노이즈 제거를 위한 고차원 피처 추출*

조인영[0], Yuchi Huo, 윤성의
한국과학기술원 전산학부
{ciy405x, eehyc0, sungeui}@gmail.com

High-dimensional Feature Extraction for Denoising Monte Carlo Renderings

In-Young Cho[0], Yuchi Huo, Sung-Eui Yoon
School of Computing, Korea Advanced Institute of Science and Technology

요약

몬테카를로 광선 추적 알고리즘은 현실감 있는 이미지를 렌더링하는 대표적인 알고리즘이다. 그러나 고품질 이미지를 얻기 위해서 한 픽셀 당 충분한 광선 샘플이 필요하여 수렴 시간이 오래 걸린다는 단점이 있다. 이에 광선을 적게 추출하여 노이즈 있는 이미지를 고속으로 구한 뒤, 노이즈 제거 모듈을 이용해 후처리하는 알고리즘에 관한 연구가 많이 진행되었다. 한편, 렌더링에서는 정의된 장면에서 텍스쳐 맵, 노말 맵, 깊이 맵 등의 이미지 공간 정보를 추출할 수 있는데, 최신 연구에서는 이를 노이즈 제거 모델의 구성 혹은 학습에 피처로 활용해왔다. 본 논문에서는 노이즈 제거 문제에서 기존 저차원 피처의 한계를 분석하고, 이를 보완하기 위해 몬테카를로 광선 추적 과정에서 얻을 수 있는 고차원 피처를 제안한다. 본 연구진은 이 고차원 피처를 최신 노이즈 제거 심층 모델에 적용하여 성능을 확인하였다.

1. 서론

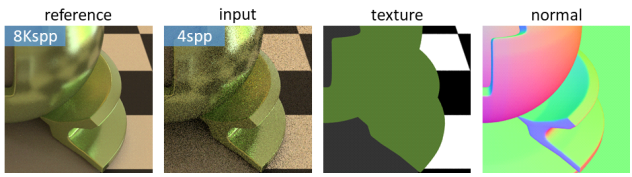


그림 1: 정답 이미지, 노이즈한 이미지와 기하 피처

노이즈 제거 알고리즘 연구의 최근 흐름은 텍스쳐 맵, 노말 맵, 깊이 맵 등의 기하 피처(G-buffer)는 단순하게 구성하는 한편 모델을 고도화하는 방향으로 진행됐다 [1, 2, 3]. 그러나 그림 1과 같이 일부 그래픽 효과에서 기하 피처의 한계를 찾아볼 수 있다. 그림 1에서 금속 재질의 물체에 고진동수 패턴이 반사된다. 그러나 텍스쳐 및 노멀 맵을 포함한 기하 피처에서는 이런 반사 패턴을 찾을 수 없다. 즉, 이런 그래픽 효과에 한해서는 기하 피처가 노이즈 제거에 도움이 되는 부가 정보를 제공하기 힘들다.

* 구두 발표논문
* 이 성과는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2019R1A2C3002833).

2. 고차원 다중 반사 피처 추출

2.1. 다중 반사 피처의 필요성

우리는 복잡한 그래픽 효과의 노이즈를 제거하는 문제에서, 기존의 기하 피처의 한계를 보완하기 위해 새로운 고차원 피처를 제안한다. 몬테카를로 광선 추적법은 물체나 벽 등에 반사되는 간접광, 색상 혼합, 렌즈로 인한 굴절 등의 전역 조명 효과를 표현하기 위한 알고리즘이다. 알고리즘은 한 면에 투사된 입사 광선에 대해 반사 광선을 반복적으로 추출하면서 빛 경로(light path)를 구성한다. 따라서 다양한 그래픽 효과에서 발생한 노이즈를 제거하기 위해서는 이런 다중 반사에 관한 정보를 포함하는 표현력 높은 피처를 사용해야 한다.

2.2. 다중 반사 피처의 추출법

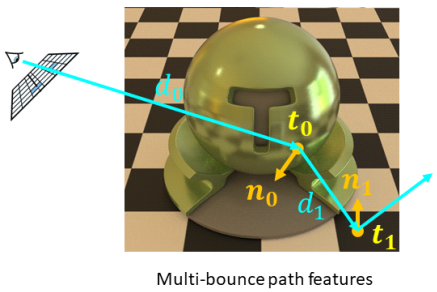


그림 2: 몬테카를로 광선 추적 과정에서 다중 반사 피처의 추출법.

일반적으로 기하 피처는 눈에서 픽셀을 향하게 발사된 광선과 장면 내 물체와의 첫 번째 교차점에서 구한다. 이를 확장하면, 몬테카를로 방법으로 추출한 빛 경로의 각 꼭짓점에서도 텍스쳐, 노말, 깊이에 관한 정보를 추출 및 저장할 수 있다(그림 2 참고). 또한, 이는 전체 렌더링 및 노이즈 제거 과정에서 볼 때 무시할 만한 오버헤드만 발생시킨다.
본 연구에서는 한 경로의 꼭짓점마다 텍스쳐, 노말, 직접 광 라디언스(radiance)의 세 가지 피처를 추출한다. 직접 광 라디언스는 각 꼭짓점에서 다중 중요도 샘플링을 위해 광원에서 추출하여 계산하는 라디언스이다[4]. 한 경로에서 다중 반사 피처를 구하는 꼭짓점의 개수를 6으로 제한하여 각 경로당 54차원의 벡터를 구하였다.

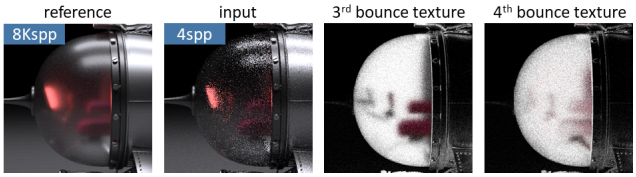


그림 3: 추출한 고차원 피처.

위 그림 3은 거친 유리 돔 내부에 있는 조종석을 표현한다. 일반적으로 기하 피처는 위 예시처럼 유리 돔 내부의 물체를 표현하기 힘들지만, 본 논문에서 제시한 다중 반사 피처를 이용했을 때는 그림 3의 3번째 및 4번째 열과 같이 돔 내부의 조종석까지 표현할 수 있다.

2.3 노이즈 제거 심층 모델 학습

제안한 다중 반사 피처를 이용하여 최신 노이즈 제거 심층 모델인 SKSN[3]을 훈련했다. 몬테카를로 광선 추적법의 확률론적 특성상, 이미지상에서 가까운 픽셀 또는 같은 픽셀에서 추출한 다중 반사 피처라도 서로 간의 상관관계가 복잡하다. 그러므로 일반적으로 노이즈 제거 연구에 활용되는 픽셀 기반 심층 모델을 활용하기 힘들다. 따라서 본 연구에서는 각 샘플 기반 심층 모델인 SKSN을 활용하였다. 학습에는 SKSN에서 제안한 하이퍼 파라미터를 이용하였고, GTX1080ti 4개로 약 3일간 학습하였다.

3. 결과

	RelMSE	Tone-mapped RelMSE	MSE	RelL1	L1
G-buffer	0.155176	0.153839	**0.035666**	0.263642	**0.039345**
ours	**0.081201**	**0.151985**	0.039978	**0.249911**	0.044445

표 1: 테스트 데이터에서 다양한 오차 측도 비교.

표 1과 같이, 가장 일반적으로 활용되는 상대 평균 제곱 오차(relMSE) 측도에서는 본 논문에서 제안한 피처로 학습시킨 모델이 큰 향상을 보인다. 다만, 다른 측도에서는 비교적 큰 차이가 없음을 알 수 있다. 매우 고차원인 다중 반사 피처의 특성상, 모델이 고차원 공간을 완전하게 탐험하는 것은 어려웠다고 해석한다.

반면, 시각적으로는 기하 피처로 훈련한 모델과 제안한 다중 반사 피처로 훈련한 모델이 생성한 노이즈 제거 이미지는 흥미로운 차이를 보였다(그림 4 참고). 기하 피처의 한계를 분석하며 예상했던 바와 같이, 반사나 굴절처럼 기하 피처가 표현하기 어려운 그래픽 효과에 관해서는 본 연구에서 제시한 피처가 이미지 재구성에 효과가 있음을 알 수 있다. 제시한 피처로 훈련한 모델은 LIVINGROOM 장면에서는 거울에 비친 조각을 더 뚜렷하게 재구성했으며, STAIRCASE 장면에서는 유리판 뒤편에 있는 계단을 더 뚜렷하게 재구성했다.

한편, 제안한 피처는 기하 피처보다 차원이 높아 심층 모델의 추론 시간을 저하했다. 기하 피처를 사용한 모델은 데이터 로드에 1.01s, 추론에 8.02s를 소요했으며, 제안한 피처의 경우 각각 4.03s와 9.10s를 소요했다.

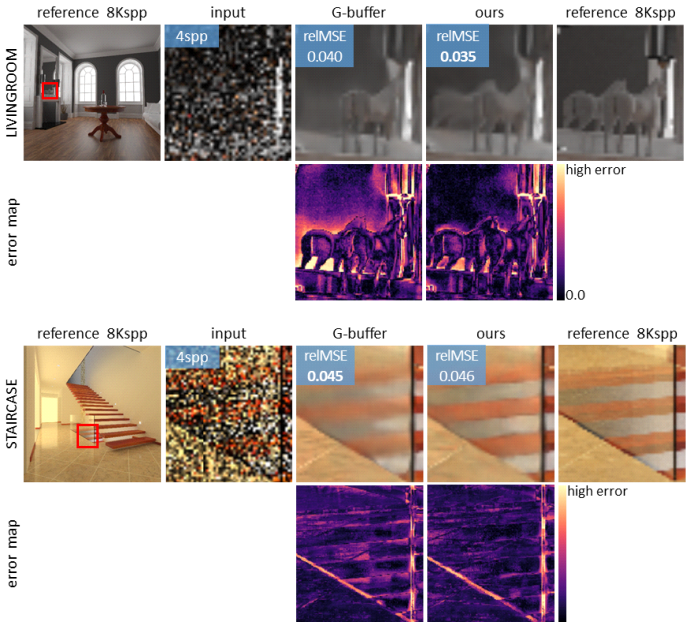


그림 4: SKSN 네트워크를 기하 피처와 본 연구에서 제시한 고차원 피처를 이용해 각각 학습시킨 뒤 테스트 데이터에서 비교한 결과. relMSE는 고해상도 이미지 전체에서 측정함.

4. 결론

본 연구에서는 기존 기하 피처가 갖는 본질적인 한계를 탐구하여, 전역 조명 효과를 더 잘 표현할 수 있는 다중 반사 피처를 제안했다. 제시된 피처를 이용하여 최신 심층 모델을 훈련했을 때, 수치적으로는 기하 피처로 훈련한 모델과 비슷하거나 더 나은 결과를 얻었고, 시각적으로는 반사 및 굴절 효과가 지배적인 장면에서 정답 이미지에 더 가까운 이미지를 생성했다. 다중 반사 피처는 고차원인 특성을 가져 학습에 어려움이 있는 것으로 예상하고, 고차원 피처를 효과적으로 학습하는 방법에 관한 후속 연구가 필요하다.

참고문헌

[1] S.Bako, T.Vogels, et al., Kernel-predicting convolutional networks for denoising Monte Carlo renderings, *ACM Transactions on Graphics*, 36(4):97-1, 2017.

[2] C.R.A.Chaitanya, A.S.Kaplanyan et al., Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder, *ACM Transactions on Graphics*, 36(4):1-12, 2017.

[3] M.Gharbi, T.M.Li, et al., Sample-based Monte Carlo denoising using a kernel-splatting network, *ACM Transactions on Graphics*, 38(4):1-12, 2019.

[4] E.Veach, and L.J.Guibas, Optimally combining sampling techniques for Monte Carlo rendering, *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp.419-428, 1995.

GPU 기반 실시간 멀티 바운스 주변 폐색 렌더링*

안현장[01], 최재원[2], 이성길[3]
[1]성균관대학교 독어독문학과, [2]성균관대학교 전자전기컴퓨터공학과, [3]성균관대학교 소프트웨어학과
{hyoenjang.an, jaewonchoi, sungkil}@skku.edu

GPU-Based Real-Time Multi-Bounce Ambient-Occlusion Rendering

Hyeonjang An[01], Jaewon Choi[2], Sungkil Lee[3]
[1]Department of German Language and Literature, Sungkyunkwan University.
[2]Department of Electrical and Computer Engineering, Sungkyunkwan University
[3]Department of Software, Sungkyunkwan University

요약

주변 폐색 렌더링은 간접 조명의 실시간 근사를 위해 많이 사용되나, 싱글 바운스 폐색 추정이 보편적이다. 본 논문은 보다 정확한 가시도 표현을 위해 물체 표면과 주변 기하의 폐색 관계를 멀티 바운스로 추정하는 방법을 제안한다. 멀티 바운스 기반 접근은 주변의 기하 관계를 보다 정확히 반영하여 더 사실적인 음영을 표현한다. 실시간 성능을 위한 GPU 기반 광선 추적은 기존보다 정교한 실시간 주변 폐색 렌더링을 가능하게 한다.

1. 서론

컴퓨터 그래픽스에서 주변 폐색은 난반사 물체의 간접 조명을 가시도 만으로 효과적으로 근사하는 기법이다. 이는 전역 조명을 완전히 계산하지 않고, 물체의 표면에서 반구 형태의 brute-force 방식으로 광선과 기하의 교차 검사만을 수행하기 때문에 저비용으로 간접 조명 효과를 낼 수 있다. 실시간 주변 폐색 효과 렌더링은 주로 screen space ambient occlusion (SSAO)[2] 기법으로 화면상에서 근사되었지만, 이는 G-buffer를 활용한 2차원 깊이 비교 근사로 정확성이 낮다.

본 논문은 이전보다 사실적인 음영의 표현을 위한 GPU 기반의 멀티 바운스 주변 폐색 렌더링 기법을 제안한다. 표면에서 보이는 주변 폐색을 정확하게 추적하기 위해 교차 후 광선을 추가로 바운스하여 주변 폐색 정도를 누적한다. 또한, 광선 추적법을 사용하여 발생하는 연산의 과부하는 병렬 연산이 가능한 GPU로 완화한다.

2. 관련연구

2.1 주변 폐색 렌더링

Shanmugam[1]은 동적인 장면과 복잡한 기하에도 적용 가능한 GPU 기반의 주변 폐색 근사 기법을 제안하였다. 주변 폐색 기법의 근사 알고리즘를 제시하고, GPU 기반의 깊이 버퍼와 법선 버퍼를 활용하여 이를 적용하는 방법을 고안하였다. Bavoil[2]은 실시간으로 주변 폐색을 렌더링하기 위한 SSAO 기법을 제시하였다. SSAO는 주변 폐색의 근사방법으로, G-buffer로 픽셀 주변을 샘플링하여 깊이를 구해내고 그 비교를 통해 폐색정도를 계산한다. 음영표현이 부정확하다는 단점이 있지만, 동적인 장면에서도 실시간 렌더링이 가능하다.

2.2 광선 추적 가속화

Horn[3]은 GPU에서의 KD트리 탐색 가속 방법으로, KD-restart 알고리즘을 개선한 세 가지 방법을 제시하였다. 기존의 방법은 큰 용량의 스택을 필요로 하거나, 트리의 루트 노드를 반복적으로 방문하기 때문에 추가적인 부하를 유발한다. Horn은 이 부하를 완화하기 위해 광선을 패킷화하는 방법, 서브 트리만을 방문하는 방법, 저용량 고정 스택을 이용하는 방법을 제시하였다.

3. 알고리즘

본 연구의 멀티 바운스 주변 폐색 알고리즘은 반복적으로 바운스되는 광선을 통해 주변 차폐여부를 정교하게 확인한다. 연산의 과부하는 GPU 병렬연산과 이에 적합한 KD-pushdown 탐색 알고리즘을 적용하여 해결한다.

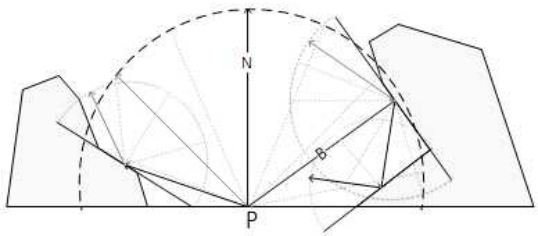


그림 1: 멀티 바운스 주변 폐색

* 구두 발표논문
* 학부생 주저자 논문임.
* 본 연구는 미래창조과학부의 재원으로 한국연구재단의 중견연구자지원사업(2019R1A2C2002449) 및 과학기술인문융합연구사업(2017M3C1B6070980)의 지원을 받아 수행되었음.

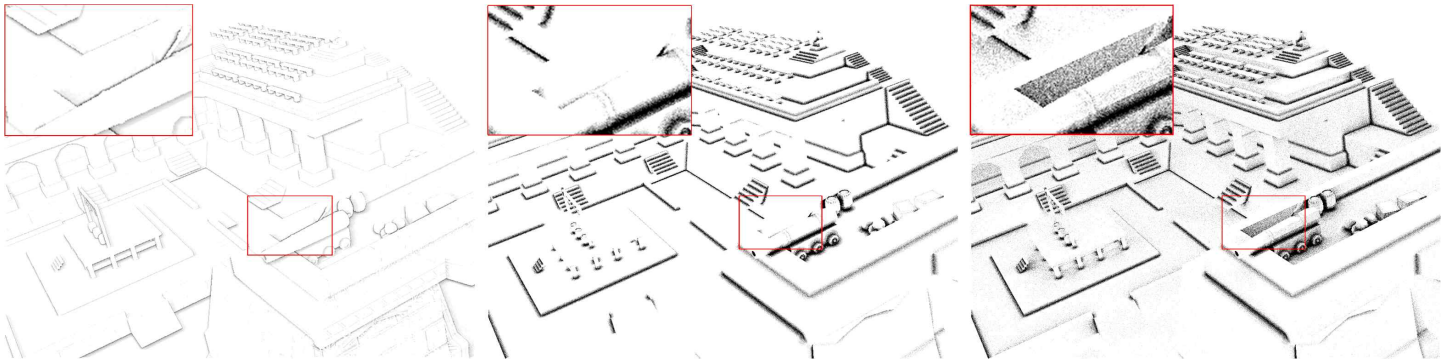


그림 2: 주변 폐색 기법에 따른 주변 폐색 맵의 예시 결과 비교. (좌) SSAO, (중간) 싱글 바운스, (우) 더블 바운스

3.1 멀티 바운스 주변 폐색

그림 1과 같이, 본 기법은 스크린 공간에서 쏜 광선이 교차하는 표면에서만 주변 폐색을 반복 계산한다. 단순히 교차를 검사하여 차폐 여부를 확인하는 것이 아니라, 교차하는 표면에서 광선을 바운스하여 다른 표면과의 차폐 여부도 함께 확인한다. 해당 알고리즘은 하단의 공식으로 정의할 수 있다.

$$B = \frac{1}{\pi}\int_{\Omega} B_i V(p,\omega_i)(n \bullet \omega_i)dw_i \qquad (1)$$

$$A(p) = 1 - \frac{1}{\pi}\int_{\Omega} B_i V(p,w_i)(n \bullet w_i)dw_i \qquad (2)$$

B는 점p에 도달한 광선$\vec{\omega}$에 대하여 반구 형태로 바운스된 광선들을 적분하는 함수이며 바운스된 광선들은 다시금 B_i로 교차한 지점에서 B와 동일한 역할을 한다. V는 가시도 함수로 점 p에서 출발한 광선 $\vec{\omega}$의 도달 지점에서 물체 표면의 교차 여부를 판단하여 0과 1을 반환한다. $(n \bullet \omega_i)$는 광선 입사각에 따른 감쇠 계수이며, 최종으로 p에서의 주변 폐색값 $A(p)$는 1에서 p에 모인 B의 반구 형태의 적분 결과를 뺀 것이 된다.

3.2 GPU 기반 광선추적법

본 기법은 픽셀에서 쏜 광선이 표면에 바운스(b)할 때마다 brute-force로 재생성된 광선의 교차를 검사한다. 이에 연산 부하는 바운스마다 재생성된 광선 수(r)의 누적 곱만큼 증가하여 총 수행 시간은 화면 해상도$\times r^b$ 가 된다. 본 연구는 GPU의 병렬 연산을 이용하여 이를 완화한다. 셰이더 상의 한정된 반복문 내에서 광선을 재생성하고, 이를 통해 주변 폐색을 검사하여 결과를 텍스처에 누적한다. 물체 표면과의 교차 연산은 물체를 GPU 상의 KD 트리 가속 구조로 구성하고 KD-pushdown 알고리즘을 적용하여 최적화한다.

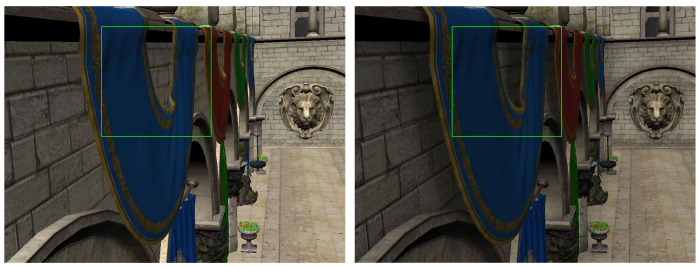


그림 3: 멀티 바운스 주변 폐색 렌더링의 적용 전(좌)과 후(우)

표 1: 렌더링 성능 측정 결과(샘플 수:4, 해상도:512×512)

모델	SSAO[2]	1 바운스	2 바운스
코넬 박스	0.048ms	1.524ms	10.431ms
유적	0.049ms	6.195ms	36.673ms
스폰자	0.169ms	21.297ms	147.508ms

4. 결과 및 토론

본 기법은 Intel Core i7-4790, NVIDIA GeForce RTX 2080 Ti에서 구현되었다. 그림 2의 1 바운스와 2 바운스가 그려낸 음영맵을 비교하면, 2 바운스의 경우 1 바운스보다 옅은 차폐와 깊은 차폐를 명확히 구분해 낸다는 사실을 확인할 수 있다. 이는 1 바운스의 경우 구할 수 있는 차폐 정보가 한정되어 있지만, 멀티 바운스의 경우 차폐 정보가 바운스마다 점진적으로 구체화되기 때문이다. 성능은 표 1에서 보듯이 1 바운스의 경우 비교적 빠르며 2 바운스부터는 연산량이 급증한다.

본 논문에서는 광선을 개별로 쏘아 brute-force로 처리하기 때문에 바운스 카운트에 따라 연산 부하가 급증한다. 추후 연구에서는 GPU 기반 광선추적법을 최적화하여 연산 부하를 효율적으로 처리하는 연구를 진행할 예정이다.

참고문헌

[1] Shanmugam, Perumaal, and Okan Arikan. "Hardware accelerated ambient occlusion techniques on GPUs." Proceedings of the 2007 symposium on Interactive 3D graphics and games. 2007.

[2] Bavoil, Louis, Miguel Sainz, and Rouslan Dimitrov. "Image-space horizon-based ambient occlusion." ACM SIGGRAPH 2008 talks. 2008. 1-1.

[3] Horn, Daniel Reiter, et al. "Interactive kd tree GPU raytracing." Proceedings of the 2007 symposium on Interactive 3D graphics and games. 2007.

[4] Díaz, José, et al. "Real-time ambient occlusion and halos with summed area tables." Computers & Graphics 34.4 (2010): 337-350. Symposium on Interactive 3D Graphics and Games. 2013.

[5] Garanzha, Kirill, and Charles Loop. "Fast ray sorting and breadth-first packet traversal for GPU ray tracing." Computer Graphics Forum. Vol. 29. No. 2. Oxford, UK: Blackwell Publishing Ltd, 2010.

몰입형 가상 환경에서 볼륨 캡처 아바타가 사회적 존재감에 미치는 영향*

조성익[0], 김승욱, 이종민, 안정현, 한정현
고려대학교 컴퓨터학과
{irunarae, wook0249, leejm8304, miru3137, jhan}@korea.ac.kr

Effects of volumetric capture avatars on social presence in immersive virtual environments

SungIk Cho[0], Seung-wook Kim, JongMin Lee, JeongHyeon Ahn, JungHyun Han
Dept. of Computer Science & Engineering, Korea University

요약

본 논문에서는 실시간 볼륨 캡처 아바타가 몰입형 가상 환경에서 사용자의 사회적 존재감에 미치는 영향을 다룬다. 이를 위해 상대방을 '볼륨 캡처 아바타', '사전 스캔 아바타', '2D 비디오'로 표현하고 각 경우에서 사용자가 느끼는 사회적 존재감의 차이를 비교하였다. 실험 결과 사용자는 동적 과제 수행 시에는 볼륨 캡처 아바타에게 가장 높은 사회적 존재감을 느끼고, 정적 과제 수행 시에는 볼륨 캡처 아바타와 2D 비디오로부터 사전 스캔 아바타 대비 높은 사회적 존재감을 느끼는 것으로 나타났다. 이를 통해 본 논문은 혼합현실이나 원격회의 등의 응용 분야에서 볼륨 캡처 아바타의 활용 가능성을 보였다.

1. 서론

3차원 재구성 관련 연구의 진전에 따라 사람의 신체와 행동을 동시에 캡처하는 실시간 volumetric capture(볼륨 캡처) 기법들이 다수 제시되었다. 이 기법들은 시각적·행동적 현실감이 높은 결과물을 생성할 수 있는데, 기술 접근성 등의 문제로 human-computer interaction 분야에서 이를 활용한 사용자 평가는 거의 이루어지지 않았다. 이에 본 논문은 특정 연기자(actor)를 실시간으로 볼륨 캡처한 아바타, RGB 영상을 이용한 2D 비디오, 전문 스튜디오의 스캔을 통해 생성한 3D 아바타의 3가지 방식으로 표현하는 시스템을 구성하고, 이를 상대로 동적/정적 과제를 수행하는 과정에서 사용자가 느끼는 사회적 실재감을 측정하였다.

2. 실험 환경

2.1. 실험 환경 구성

본 논문의 실험은 로컬 사용자와 원격 사용자가 상호작용하는 시나리오를 상정하였다. 로컬 사용자는 VR HMD를 착용하고, 원격 사용자는 그림 1과 같이 캡처된 후 로컬 사용자의 가상 공간으로 텔레포트 되도록 구성하였다. 두 사용자는 칸막이로 나누어진 방에 위치하였으며, 가상 공간은 로컬 사용자의 3.3m×2m 공간과 맞대응되도록 구성되었다.

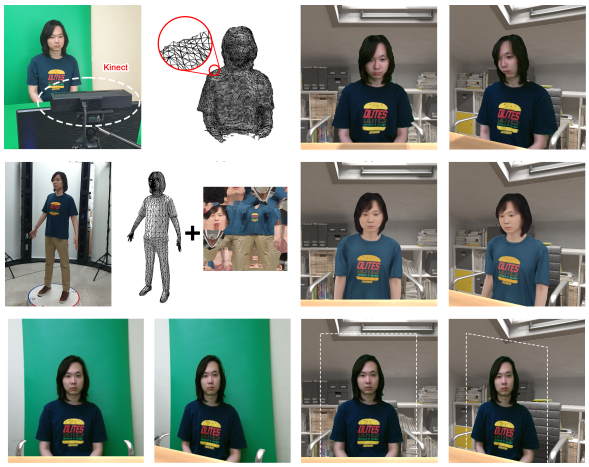


그림 1: 원격 사용자 캡처. (상단) 볼륨 캡처 아바의 캡처 환경, 메시 및 렌더링 화면. (중단) 사전 스캔 아바타의 생성 과정, 결과물 및 렌더링 화면. (하단) 2D 비디오의 크로마키 적용 전과 후의 렌더링 화면.

2.2. 원격 사용자 캡처

볼륨 캡처 아바타 : 볼륨 캡처 아바타는 [1]의 기법을 확장하여 생성하였다. Kinect V2 RGB-D 카메라를 이용하여 원격 사용자의 상체를 캡처하고, 투영 텍스처 매핑과 voxel 색상을 이용하여 캡처 대상의 색상 정보를 생성하였다. 메시 최적화 과정을 통해 30fps의 속도로 매 프레임 약 3만 개의 정점 정보와 512×424의 텍스처 정보를 1Gbit LAN을 통해 전송하도록 구성하였다.
사전 스캔 아바타 : 사전 스캔 아바타는 전문 스튜디오에서 연기자를 촬영 뒤 사진 기반 재구성 기법과 후처리를 통해 생성하였다. 전문가 작업을 통해 얼굴 애니메이션용 blendshape를 생성하고 Oculus LipSync API를 이용하여 아바타의 입 모양이 원격 사용자의 마이크 입력에 대응하도록 하였다. Adobe Mixamo로 생성한 뼈

* 포스터 발표논문
* 본 논문은 요약논문(Extended Abstract)으로서, 본 논문의 원본 논문은 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)에 발표되었음.
* 이 성과는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.
(No. NRF2016R1A2B3014319, NRF2017M3C4A7066316)

대 정보와 HMD 및 컨트롤러의 위치를 이용한 inverse kinematics로 신체 움직임을 구현하였다.
2D 비디오 : 2D 비디오는 Kinect V2 카메라의 RGB 영상을 가상 공간으로 전송하여 구현하였다. 원격 사용자 촬영 과정에서 녹색 배경을 사용하여 크로마키 기법을 적용하였으며, world-aligned billboard에 부착하여 가상 공간에 렌더링하였다.

3. 실험 구성

원격 사용자의 볼륨 캡처 아바타, 사전 스캔 아바타, 2D 비디오를 상대로 로컬 사용자가 동적/정적으로 상호작용을 수행하는 2개의 실험을 구성하였으며, 총 24명의 참가자가 실험을 진행하였다. 사회적 존재감 측정에는 Harms and Biocca의 사회적 존재감 설문(HSP)[2]에서 '공존감(CP)', '관심 할당(AA)', '인식된 메시지 이해(PMU)' 항목을, Temple Presence Inventory(TPI)[3]에서 '사회적 존재감(SP)', '정신적 몰입(EG)', '사회적 풍요성(SR)' 항목을 추출하여 7점 리커트 척도의 설문지를 구성하였다. 각 설문 항목은 크론바흐 알파 값을 통해 유효성을 확인하였고, 설문 분석에는 각 설문지의 누적합계인 AggHSP와 AggTPI를 포함하였다. 통계 분석을 위해 샤피로-윌크 검정을 수행한 후 정규성 여부에 따라 반복측정 분산분석 혹은 프리드만 검정을 사용하였다. 사후검정으로는 대응표본 t 검정과 윌콕슨 부호순위 검정을 각각 사용하였다. (신뢰수준 95%)

3.1. 실험1 – 과정

실험1에서는 3~4개의 정육면체로 구성된 조각 7개를 사용하여 3×3×3의 cube를 조립하는 'soma cube'를 과제로 제시하였다. 본 실험에서는 네 가지 색으로 구성된 14개의 조각을 가상 공간에 배치하였다. 로컬 사용자는 원격 사용자의 지시에 따라 필요한 조각을 탐색한 뒤, 각 조각의 색상과 모양을 원격 사용자에게 확인받은 후 이를 조립하도록 하였다. 이 과정에서 로컬 사용자가 가상 공간의 동적으로 움직이며 다양한 위치에서 원격 사용자를 응시하도록 설정하였다.

3.2. 실험1 – 결과 및 분석

사회적 존재감에 대한 로컬 사용자의 설문지 분석 결과, 대부분의 설문 항목에서 사전 스캔 아바타의 점수가 상대적으로 낮게 나타났다. 전문 스튜디오에서의 촬영과 전문가의 후처리를 거쳤음에도 불구하고 실험 후 인터뷰에서 많은 실험 참가자들이 사전 스캔 아바타가 인공적인 느낌이 든다고 응답하였으며, 이는 uncanny valley에 의한 것으로 추정된다. HSP 설문에서는 '관심 할당' 이외의 모든 항목에서 볼륨 캡처 아바타가 사전 스캔 아바타(AggHSP:$p < .001$, CP:$p = .002$, AA:$p = .012$, PMU:$p = .003$)와 2D 비디오(AggHSP:$p = .016$, CP:$p = .002$, PMU:$p = .011$)보다 통계학적으로 유의미한 수준의 높은 사회적 존재감을 보였으며, TPI 설문 항목에서는 볼륨 캡처 아바타의 사회적 존재감이 사전 스캔 아바타(AggTPI:$p = .020$, SP:$p = .010$, SR:$p = .012$), 2D 비디오(SP:$p = .040$)보다 높거나, 상호 간의 차이가 없는 것으로 나타났다.

3.3. 실험2 – 과정

실험2는 로컬과 원격 사용자 모두 가상 환경 속 책상에 마주 앉은 정적인 환경에서 'desert survival task'를 수행하는 것으로 구성하였다. 이는 비행기 사고로 사막에 불시착한 상황에서 제시된 도구들의 중요도 순위를 협의를 통해 결정하는 것으로, 로컬 사용자와 원격 사용자는 각자 본인이 생각한 초기 순위를 바탕으로 상대방과 합의를 진행하였다. 원격 사용자인 연기자는 각 도구의 장단점을 서술한 스크립트를 이용하여 로컬 사용자를 설득하는 방향으로 대화를 진행하였다.

3.2. 실험2 – 결과

사회적 존재감에 대한 로컬 사용자의 설문지 분석 결과, HSP의 모든 항목에서 통계학적으로 유의미한 차이가 발견되지 않았다. 몇몇 실험 참가자들은 이에 대해 사전 스캔 아바타가 실험1과 비교하여 대화 과정에서 더 눈을 잘 마주치는 것처럼 느꼈으며, 볼륨 캡처 아바타와 2D 아바타의 차이를 잘 느끼지 못했다고 응답하였다. 이는 비대칭적인 하드웨어 환경으로 인해 볼륨 캡처와 2D 비디오의 경우 원격 사용자가 RGB-D 카메라를 바라보는 것으로 시선을 처리하였는데, 이에 의한 부자연스러움이 영향을 미친 것으로 추정된다. TPI 설문에서는 모든 항목에서 통계학적으로 유의미한 수준으로 사전 스캔 아바타의 사회적 존재감이 볼륨 캡처 아바타(AggTPI:$p = .010$, SP:$p = .015$, EG:$p = .031$, SR:$p = .013$)와 2D 비디오(AggTIP:$p < .001$, SP:$p < .001$, EG:$p < .001$, SR:$p = .002$)보다 낮게 나타났으나, 볼륨 캡처 아바타와 2D 비디오 사이에서는 유의미한 우열 관계가 나타나지 않았다.

4. 결론

본 논문에서는 볼륨 캡처 아바타가 사회적 존재감에 미치는 영향을 실험을 통해 알아보았다. 동적 실험 환경에서는 사전 스캔 아바타와 2D 비디오보다 볼륨 캡처 아바타의 사회적 존재감이 높은 것으로 나타났으며, 정적 실험 환경에서는 볼륨 캡처 아바타와 2D 비디오가 사전 스캔 아바타보다 높은 사회적 존재감을 제공하는 것으로 나타났다. 본 실험 결과는 볼륨 캡처 아바타를 향후 혼합현실이나 원격회의 등의 응용 분야에 접목하기 위한 기초 연구로 활용할 수 있을 것이다.

참고문헌

[1] R. A. Newcombe et al. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.
[2] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. In Seventh Annual International Workshop: Presence 2004., 2004.
[3] M. Lombard et al. Measuring presence: the temple presence inventory. In Proceedings of the 12th Annual International Workshop on Presence, pp. 1–15, 2009.

다중 이미지를 이용한 광원 추정*

조창호[0], 하인우, 윤성의
한국과학기술원 전산학부
timegate@kaist.ac.kr, iw.ha@samsung.com sungeui@kaist.edu

Light estimation using multiple images

Changho Jo[0], Inwoo Ha, Sung-eui Yoon
School of computing, KAIST

요약

광원 추정(Light estimation) 문제는 가상 현실 어플리케이션에서의 현실적인 표현을 위한 주요한 문제이며, 제한된 시야로부터 가상의 물체를 렌더링하기 위해 반드시 필요한 과정이다. 본 논문에서는 기존까지의 단일 이미지 기반으로 접근해왔던 광원 추정 문제를, 다중 이미지를 활용하여 더 많은 정보로부터 풀어내는 것을 제안한다.

1. 서론

최근 들어, 많은 어플리케이션들은 가상의 물체들을 현실로 가져와 새로운 경험들을 제공하고 있다. 하지만 여전히 가상 물체와 실제 물체 사이의 격차를 제거하기 위한 많은 문제들이 남아있으며, 역 렌더링(Inverse rendering) 문제[1]에 그 기반을 둔다. 즉, 가상의 물체를 현실 환경에 놓기 위해서는 주어진 장면으로부터 그 환경의 기하, 반사 성질, 그리고 빛의 3가지 요소를 분리해서 찾아내야 하는 어려움이 있게 된다. 더욱이, 모바일 어플리케이션이라면 오브젝트를 둘러싼 모든 방향에서 오는 빛을 전체 파노라마의 6%밖에 되지 않는 모바일 기기의 화면에서 예측해야만 한다.

그러나 실제로는 여러 장의 관련된 이미지들을 활용할 수 있음에도 불구하고, [2]와 [3]과 같은 기존 연구들은 모두 단일 이미지 기반이었다. 따라서 우리는 이 다중 이미지들을 활용해 더 좋은 결과를 내고자 하였고, 이를 위해 매터포트3D 데이터셋을 가지고 다중 이미지들을 잘 활용할 수 있는 ConvLSTM 인코더-디코더 구조의 네트워크를 활용하여 문제를 풀었다.

2. 광원 추정이란 무엇인가

2.1. 광원 추정 문제와 기존 해결 방법

광원 추정 문제란 주어진 장면으로부터 HDR 파노라마와 같은, 가상의 물체를 렌더링하기에 충분한 결과를 만드는 과정이다. [2]는 처음으로 기하, 물질 속성, 빛 등 어떠한 것에도 큰 가정을 두지 않고 광원을 추정해냈으며, 가상 물체가 놓일 특정 위치를 중심으로 일정 크기로 잘라낸 이미지를 인풋으로 하고, 그들이 직접 만든 광원 분류기로 찾아낸 광원 마스크와 위에서 정한 위치로 구형 워핑을 한 파노라마를 아웃풋으로 하여 하나의 심층망으로 구현하였다. 그러나 이 연구는 세부적인 광원들까지 맞추지는 못한다는 한계를 가진다.

다음으로 [3]는, 어려운 역 렌더링 문제에 기반한 광원 추정 문제를 하나의 블랙박스 네트워크로 풀기보다, 서브 태스크들로 나누어 풀 때 더 좋은 결과를 낸다는 것을 보여주었다. 광원 추정 문제를 기하 추정 문제, 기하를 이용한 특정 위치로의 워핑, 파노라마 장면 완성, 완성된 장면의 HDR로의 전환, 이 4가지로 분리해서 세부적인 광원들까지 더 잘 찾아내는 결과를 보여주었다. 하지만 다중 이미지를 활용할 수 있는 경우에도 단일 이미지 기반으로만 광원을 추정한다는 한계를 가진다.

3. 다중 이미지를 이용한 광원 추정

3.1. 매터포트3D 데이터셋

매터포트3D 데이터셋은 다양한 시각에서 촬영한 194,400개의 HDR RGB-D 이미지들을 제공한다. 이 이미지들은 10,800개의 파노라마로 구성되어있고 90개 빌딩의 실내 모습을 보여주며, 데이터셋은 또한 실내 환경에 관련된 특징들을 학습하는 데 사용될 수 있도록 세그멘테이션 정보와 메쉬 정보도 가지고 있다.
더 자세하게는 하나의 파노라마를 모두 덮을 수 있는 18장의 이미지를 제공하는데, 우리는 그림1과 같이 18장의 이미지 중 광원의 직접적인 정보를 찾기 힘든 수직적으로 가운데에 있는 6장의 이미지들만을 활용하여 문제를 풀었다. 본 논문의 실험에서는 LDR 파노라마만을 사용하였다.

3.2. ConvLSTM

ConvLSTM은 비디오 예측 모델에서의 유명한 네트워크 블록들 중 하나이다. 기존의 Fully-connected LSTM 네트워크보다 모델의 복잡성을 줄이면서 연속적인 이미지들의 공간적인 정보를 더 잘 저장하며, 계속해서 실험적으로도 좋은 결과를 보여주었기에 계속해서 사용돼 왔다. 우리도 연속적인 여러 이미지들의 시공간적 정보

* 포스터 발표논문
* 본 연구는 과학기술정보통신부/정보통신기획평가원의 지원(IITP-2015-0-00199) 및 과학기술정보통신부/한국연구재단-차세대정보·컴퓨팅기술개발사업의지원(No. NRF-2017M3C4A7066317)을 받아 수행된 연구임.

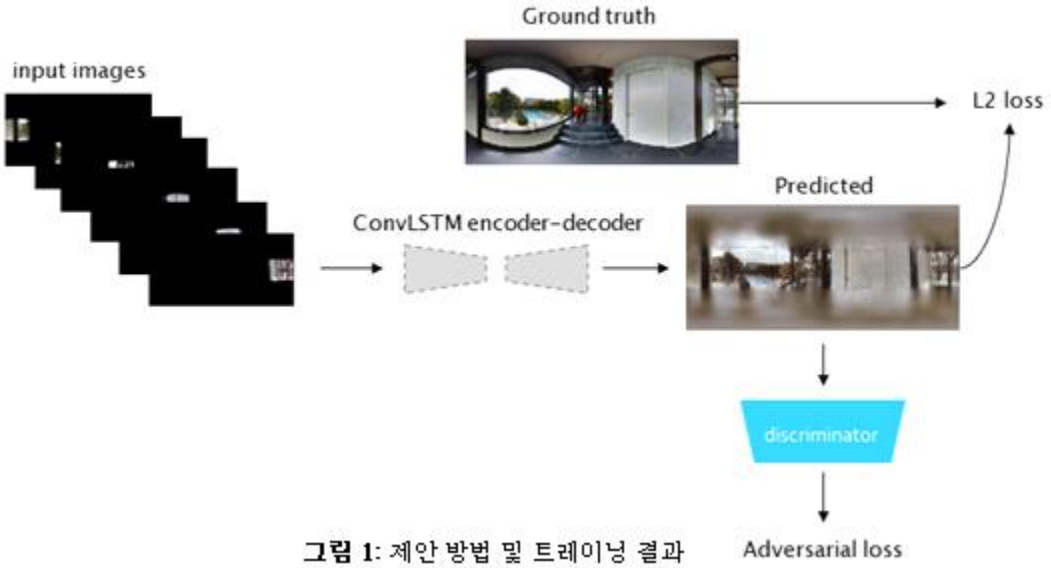


그림 1: 제안 방법 및 트레이닝 결과

를 완전히 활용하기 위해 ConvLSTM을 사용하였다.

3.3 제안 방법

우리는 [4]에서 사용한 ConvLSTM 인코더-디코더 구조를 RGB 이미지의 3채널을 받아들일 수 있도록 변형하여 사용하였다. 트레이닝 과정에서는 매터포트3D의 각각의 파노라마를 네트워크의 아웃풋으로, 앞서 언급한 6장의 이미지를 네트워크의 인풋으로 하였고, 테스트 과정에서는 보여주지 않았던 위치에서의 6장의 이미지만을 주고 전체 파노라마를 출력하도록 하였다. 여기에서의 인풋으로 들어가는 6장의 이미지들은 각각 파노라마의 일부분과 공간적으로 같을 수 있도록, 카메라가 바라보는 방향에 맞게 파노라마처럼 구형 왜곡을 적용해주고, 공간적으로 대응되는 곳에 위치시켰다.

3.3. 실험 결과

트레이닝은 9,810개의 파노라마로 20 epoch 동안 진행하였으며, 손실 함수는 [3]과 같이 L2 loss와 adversarial loss를 사용하였다. c개의 채널과 k개의 필터를 가진 Convolution-Relu를 C(c, k)로 표기할 때, adversarial loss를 위한 discriminator는 C(3, 16)-C(16, 32)-C(32, 64)-C(64, 128)-C(128, 256)-C(256, 512)와 같이 구성하였다.

트레이닝 결과에서는 인풋에서 제공된 수직적으로의 가운데 부분뿐만 아니라 파노라마의 위와 아래 부분도 괜찮은 결과를 내었지만, 테스트 결과에서는 위와 아래 부분으로 중간 부분의 이미지에서의 컨벌루젼이 충분히 되지 않는 결과를 얻었다.

그림 2: 테스트 결과

4. 결론

이 논문은 다중 이미지를 활용해 광원 추정 문제를 더 정확히 풀어낼 수 있음을 제안하였다. 실제 어플리케이션에는 하나의 이미지로부터 광원을 추정해야 하는 기존 연구들과는 달리 다중 이미지를 활용해 광원을 추정하기 위한 더 많은 정보들을 얻을 수 있기 때문이다. 비록 기존 비디오 예측 모델에서 주로 사용하던 ConvLSTM 블록을 변형하여 사용한 이 실험은 좋은 테스트 결과를 내지 못했지만, 이러한 광원 추정 문제에 맞는 연속적인 이미지들을 더 잘 활용할 수 있는 네트워크를 도입하고, 매터포트3D 데이터셋에서 제공하는 공간적인 정보, 시맨틱한 정보까지 모두 잘 이용한다면 더 좋은 결과를 낼 수 있을 것이다.

참고문헌

[1] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 117–128. ACM, 2001.

[2] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to predict indoor illumination from a single image. ACM Trans. Graph. 36, 6, Article 176 (November 2017), 14 pages.

[3] S. Song and T. Funkhouser. Neural illumination: Lighting prediction for indoor environments. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6918–6926, 2019

[4] WANG, Rui; PIZER, Stephen M.; FRAHM, Jan-Michael. Recurrent neural network for (un-) supervised learning of monocular video visual odometry and depth. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. p. 5555-5564.

확장 컨볼루션과 뼈대 기반 손실 함수를 이용한 모션 리타겟팅*

김상빈[0], 박인범, 권성수, 한정현
고려대학교 컴퓨터학과
{sang-bin, nulledge, tjdtn5683, jhan}@korea.ac.kr

Motion Retargetting based on Dilated Convolutions and Skeleton-specific Loss Functions

SangBin Kim[0], Inbum Park Seongsu Kwon, JungHyun Han
Dept. of Computer Science & Engineering, Korea University

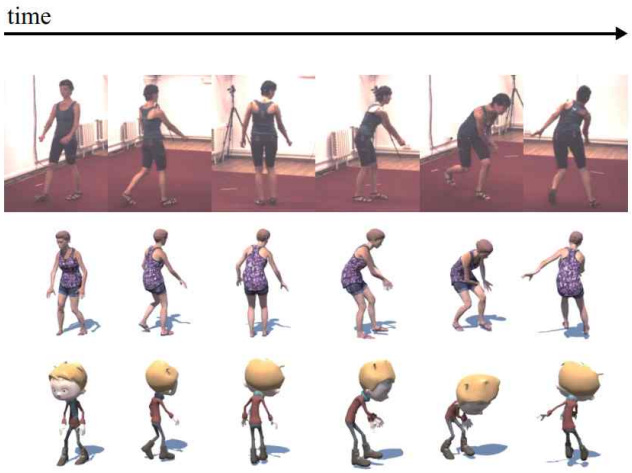


그림 1: 원본 모션 캡처 데이터(상단)를 골격 크기가 다른 다양한 캐릭터(하단)에 리타겟팅한 결과.

요약

모션 리타겟팅은 원본 캐릭터의 모션을 대상 캐릭터에 맞추는 과정이다. 본 논문은 확장 컨볼루션에 기반한 모션 리타게팅 모델을 제시한다. 제안된 모델은 비감독 방식으로 다양한 휴머노이드 캐릭터에 대해 사실적인 동작을 생성한다. 제안한 모델로 생성한 리타겟팅 모션은 원본과 대상 캐릭터 간의 골격 크기 차이에도 입력 모션의 디테일을 보존할 뿐만 아니라, 자연스럽고 안정적인 움직임을 생성한다. 가상 캐릭터 데이터셋과 모션 캡처 데이터셋을 이용하여 다양한 실험을 수행하고, 기존 방법에 대한 정성적 및 정량적 비교를 통해 제안한 방법의 성능과 효과를 입증하였다.

1. 서론

딥 러닝은 여러 분야에 적용되어 높은 성능을 입증하고 있지만, 모션 리타겟팅에 적용된 사례는 근래 들어 몇 가지 연구가 보고되었다. 최신 연구 중 하나인 [1]은 순환 신경망(RNN)과 정기구학(forward kinematics)를 결합한 구조를 제안하였다. 기 제안된 방법은 좋은 성능을 보여주었지만, 모션 리타겟팅시에 흔히 마주칠 수 있는 상황인 크기가 다른 캐릭터에게 적용할 때, 불안정한 동작을 생성하였다. 우리는 이러한 한계를 극복하기 위해 다음과 같은 모델을 설계하였다. 우선, 학습을 최적화하기 쉬우면서도 파라미터를 효율적으로 활용할 수 있는 확장 컨볼루션(dilated convolution) 신경망을 사용하였다. 또한, 다양한 크기의 캐릭터를 다룰 수 있는 뼈대 기반 손실 함수를 도입하였다. 이를 통해 제안된 모델은 사실적이고 강건한 모션 리타겟팅을 수행할 수 있었으며, 다양한 데이터셋을 이용한 실험을 통해 성능을 측정하였다.

* 포스터 발표논문
* 본 논문은 요약논문(Extended Abstract)으로서, 본 논문의 원본 논문은 Computer Graphics Forum (Proc. Eurographics) 39(2). 2020에 게재 되었음.
* 이 성과는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No. NRF2016R1A2B3014319, NRF2017M3C4A7066316)

2. 방법

2.1. 모션 리타겟팅 모델

그림 2는 전체 모션 리타겟팅 구조를 나타낸다. 모델 입력 데이터는 소스 모션 시퀀스 $x^A_{1:T}$와 타겟 스켈레톤 $\bar{s}^B$이다. $x^A_{1:T}$는 관절 회전 q^A_t와 루트 관절 위치 r^A_t로 이루어진다. 정기구학을 계산하는 FK 모듈을 이용하여 로컬 관절 위치 p^A_t를 계산하고, 입력 시퀀스의 양 끝을 시작과 끝 프레임을 이용하여 패딩 한다. 시간 확장 컨볼루션 네트워크(TDCN)에 이 데이터를 입력받아 타겟 스켈레톤 $\bar{s}^B$에 맞게 리타겟팅 된 관절 회전 $q^B_{1:T}$와 루트 관절 위치 $r^B_{1:T}$를 출력한다. 최종적으로 FK 모듈을 통해 리타겟팅 모션 시퀀스 $\hat{x}^B_{1:T}$를 얻는다.

2.2. 시간 확장 컨볼루션 네트워크(TDCN)

RNN 대신, 고정된 크기의 수용장(receptive field)를 가진 CNN을 사용하면 시계열 입력 데이터를 처리할 때도 모델이 고정된 길이의 종속성을 가질 수 있다. 또한, 네트워크를 조정하여 원하는 대로 길이를 제어할 수 있다. CNN을 확장한 확장 컨볼루션은 입력 프레임을 건너뛸 수 있도록 설계되었고, 같은 크기의 수용장을 유지하면서도 적은 파라미터를 가진다. 이를 통해 모델을 효과적으로 학습시킬 수 있으며 CNN과 비슷한 성능을 유지할 수 있다. 이는 소리 합성, 자연어 처리, pose estimation[2]과 같은 여러 분야에서 활용되어 성능이 입증되었다. 서술한 특징은 모션 시퀀스와 같은 시계열 데이터를 처리하는 데 적합하므로, 안정적인 모션 리타겟

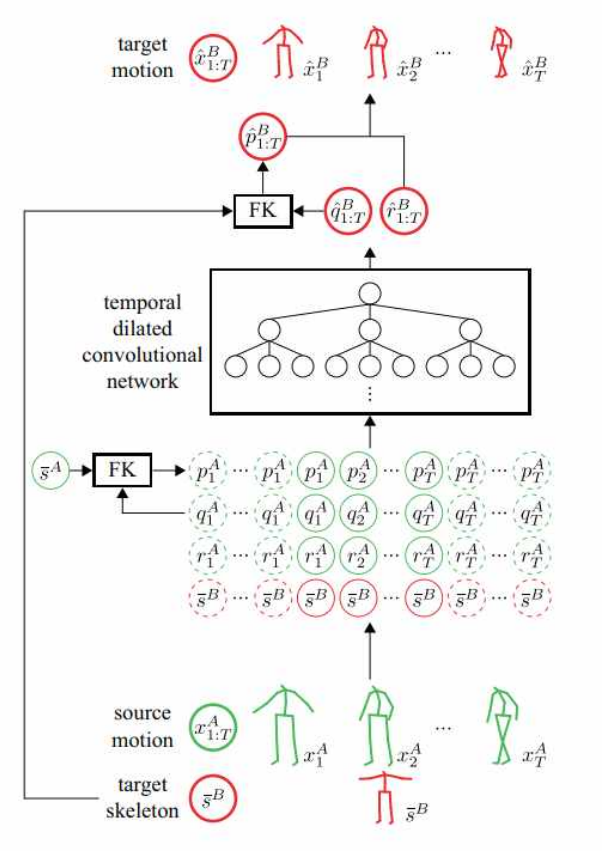


그림 2: 전체 모션 리타겟팅 모델 구조.

팅을 위해 TDCN 기반의 모델을 설계하였다. 또한, 모델을 골격 크기 또는 모션 길이와 관계없이 자연스럽고 안정적인 모션 리타겟팅을 수행하도록 학습하였다.

3. 적대적 학습

일반적인 지도학습을 통해 학습하기 위해서는 모든 캐릭터에 대해 대응되는 대상 캐릭터의 모션 쌍이 필요하다. 이러한 한계를 극복하기 위해 적대적 학습과 사이클 일관성(cycle consistency)을 이용한 비지도 학습을 적용하였다. 이를 위한 손실 함수는 적대적 손실, 높이 손실 등 6가지를 사용하였다. 먼저, '적대적 손실'은 안정성을 위해 최소 제곱 GAN 손실을 활용하고, 구별자(discriminator)에 PatchGAN을 적용하였다. 이를 통해 학습 과정 중 구별자의 피드백이 전체 생성된 모션의 평균으로 수렴하지 않으면서 원본 동작의 디테일을 잘 보존할 수 있도록 하였다. '높이 손실'은 입력 및 대상 캐릭터의 키를 이용하여 모션을 정규화(normalization)하고 입력 모션과 생성된 모션의 차이가 줄어들도록 smooth L1 손실을 적용하였다.

4. 실험 및 평가

학습을 위해 Adobe사에서 제공하는 Mixamo 가상 캐릭터 데이터셋을 사용하였다. 해당 사이트에서 다양한 크기의 9개 캐릭터와 1646개의 모션을 선정하여 학습 데이터셋을 구성하였다. 학습 과정에서 모션은 랜덤하게 선택되고, 선택된 모션으로부터 다시 랜덤하게 81개의 프레임을 추출하였다. 또한, 데이터 증강을 위한 무작위 스케일링을 적용하여, 모델이 다양한 크기의 캐릭터를 처리하도록 하였다. 테스트를 위해서는 짧고 긴 두 가지

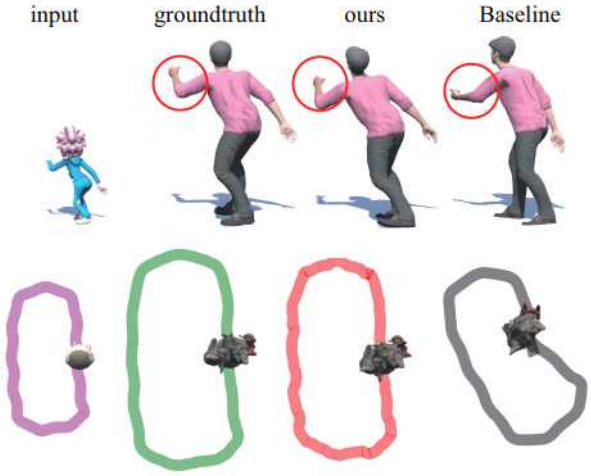


그림 3: 정성적 비교를 위한 모션 리타겟팅 결과.

종류의 데이터셋을 활용하였다. 우선 짧은 모션 시퀀스는 [1]의 테스트 데이터셋과 같으며 모두 120프레임으로 구성되었다. 둘째로, 120 frame으로 나뉘지 않은 긴 모션 시퀀스에 대해서도 테스트를 진행하여 모델이 이에 대해서도 안정적으로 작동할 수 있는지 정량적, 정성적으로 평가를 진행하였다.

정량적 평가는 관절들의 글로벌 관절 위치 차이를 이용하여 평균 제곱 오차를 계산하였다. 제안한 모델은 모든 정량적 평가 항목에서 [1]의 모델을 포함한 모든 베이스라인과 ablation 모델들에 대해 모두 높은 성능을 보였다. 특히, '높이 손실'을 이용한 베이스라인과 ablation 모델을 비교할 때, 성능의 차이가 가장 크게 나타나, 명시적으로 캐릭터의 크기를 다루는 것이 성능에 높은 영향을 미치는 것을 확인하였다. 그림 3은 정성적 평가결과를 나타낸다. 상단의 결과를 살펴보면 우리 모델의 리타겟팅 결과가 groundtruth 결과를 잘 따라가는 것을 알 수 있다. 하단의 결과는 긴 모션 시퀀스에 대해 리타겟팅을 한 결과인데 안정적인 캐릭터 움직임을 생성하는 것을 알 수 있다.

4. 결론

본 논문은 시간 확장 컨볼루션(TDCN)에 기반한 효과적이고 효율적인 리타겟팅 모델을 제안하였다. TDCN은 기존 방법보다 학습이 빠르며, 다양한 캐릭터로 훈련할 때도 안정적으로 학습할 수 있다. 또한, 새로 도입한 뼈대 기반 손실 함수들은 모델이 캐릭터 간 크기 차이를 명시적으로 학습할 수 있도록 도와 다양한 크기의 캐릭터들 사이에 안정적이고 자연스러운 리타겟팅이 가능하다.

참고문헌

[1] Villegas, Ruben, Jimei Yang, Duygu Ceylan, and Honglak Lee. "Neural kinematic networks for unsupervised motion retargetting." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8639-8648. 2018.

[2] Pavllo, Dario, Christoph Feichtenhofer, David Grangier, and Michael Auli. "3D human pose estimation in video with temporal convolutions and semi-supervised training." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7753-7762. 2019.

초현실주의 회화 감상을 위한 가상현실 콘텐츠 개발*

손서영[0], 이승연
세종대학교 소프트웨어학과, 디지털콘텐츠학과
{17011663, lsyn97}@sju.ac.kr

Development of Virtual Reality Content for Surrealistic Painting Appreciation

Seo Young Son[0], Seung Yeon Lee
Dept of Software, Dept of Digital Contents, Sejong University

요약

본 연구에서 제안하는 가상현실 콘텐츠는 회화 감상에 진입장벽이 높다고 여겨지는 초현실주의 미술 전시를 대중들에게 쉽게 접근하는 기회를 제공하고 개별적인 작품들을 하나의 세계관으로 감상할 있도록 한다. 사용자는 제작된 콘텐츠를 네비게이션 하면서 작가의 세계관을 자연스럽게 이해할 수 있다.

1. 서론

2017년 문화체육관광부가 발표한 통계[1]에 따르면 매년 문화 예술행사 관람률이 지난 10년동안 꾸준히 상승세를 보이고 있다. 하지만 다른 분야와 다르게 미술관람 분야의 상승세는 미미하였다. 설문조사[2]에 따르면 대중들은 미술관람을 하지 않은 이유를 크게 “미술에 관해 지식이 없다”, “주변에 갤러리가 없다”, “시간상의 이유로 관람을 할 수 없다” 등으로 응답하였다. 이처럼 어렵게만 느껴지는 미술에 대한 장벽을 낮추고자 미술 전시회를 가상 콘텐츠로 제작하여 언제 어디서나 전시회를 현장감 있게 감상할 수 있도록 하는 것이 본 연구의 목표이다. 고전적인 미술 감상에서 벗어나 시간이 부족한 현대인들과 미술교육의 형태로써 가능할 수 있을 것으로 기대한다.

2. 관련 회화 감상 콘텐츠 분석

근현대 회화 작품을 가상현실 콘텐츠로 제작하는 시도는 활발히 진행되고 있다. 그림 1은 2016년 The Dali Museum에서 초현실주의 작가 달리의 작품을 교육 및 홍보 목적으로 제작한 360도 창작물이다.

그림 1: <달리의 꿈> 360° 영상

그림 2: <밤의 카페>

위 작품은 <밀레의 만종에 대한 고고학적 회상> 작품을 다방면에서 관찰할 수 있으며 사용자들의 인터뷰에 따르면, “충분한 몰입감을 선사하며 작품을 보는 관점을 바꾸었다”고 밝혔다. 그림 2는 2016년 Borrowed Light Studio에서 개발한 낭만주의 화가 고흐의 작품을 하나의 세계관으로 엮은 가상현실 콘텐츠이다. 그림 1과 다르게 화풍을 텍스쳐로 표현하였으며 컨트롤러 동선을 이동하여 보다 자유롭게 작품을 감상할 수 있다. 체험자들의 후기에 의하면, 압도적으로 긍정적인 반응을 보였고, 그림 1의 콘텐츠와 마찬가지로 회화 감상에 흥미를 보였다. 하지만 앞선 두 사례 중 그림 1은 고정적인 동선 한해서 작품을 관찰할 수 있어, 자유로운 감상을 하지 못하는 한계를 지니고 있으며, 그림 2는 고가의 장비를 필요로 한다는 점에서 콘텐츠 소비가 제한적이다.

3. 대중을 위한 초현실주의 가상현실 전시

3.1. 르네 마그리트 가상전시관 기획

초현실주의 작가로 르네 마그리트로 선정하였으며, 각기 여러 작품들의 감상을 원활히 하도록 자연스러운 동선을 기획하였다.

디바이스 시장 중 가장 많이 차지하고 있는 플랫폼이 안드로이드임을 비추어 시중에서 저렴한 편이고 착용과 접근성이 쉽다고 평가받는 구글 카드보드로 개발하는 것으로 초점을 맞췄다. 구글 카드보드는 다른 VR 디바이스와 달리 일반적인 어플리케이션 설치와 구글 카드보드에 장착하여 시점에 따라 플레이어가 움직이는 매우 간단한 조작법으로 기획하게 되었다.

필요한 모델 에셋들은 직접 3D 모델링 툴로 직접 제작하는 방향으로 기획하였다. 모델의 효율적 시각화로

* 포스터 발표논문
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터 지원사업의 연구결과로 수행되었음 (IITP-2020-2016-0-00312)
* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2015-0-00938).

초현실주의에 걸맞게 사실적으로 표현하되 안드로이드 플랫폼에서 제작하는 만큼 low polygon으로 최대한 모델링하였다.

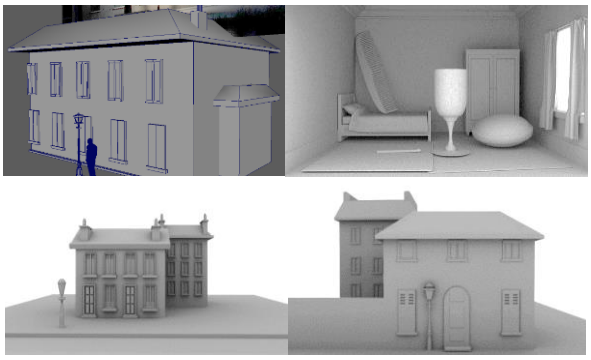


그림 3: 르네 마그리트 회화의 3D 모델링화

3.2. 가상전시관 구현

여러 작품을 통합하여 만들어낸 임의의 작품 단지로부터 사용자가 바라보는 대로 작품에 다가갈 수 있도록 하는 Lookwalk 기능을 구현하였다. 사용자가 작품에 접근할 수 있는 영역은 Collider object를 설치하여 제한하였다

그림 4: vertex shader 쉐이더

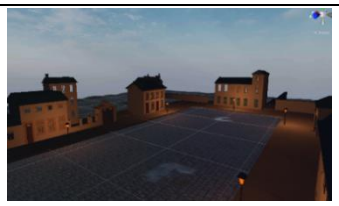


그림 5: vertex shader를 정보 받아 표현된 지형 텍스처의 모습

제한적인 하드웨어임을 감안하여 비추어지는 텍스쳐의 질감을 더 잘 돋보이기 위해 쉐이더를 설정하였다. 쉐이더는 최종적으로 디스플레이에 출력하는 픽셀들의 색을 최종적으로 결정하는 함수인데, 텍스처로 작업하기 불가능한 부분 및 연산 시간을 단축을 이루었다. 위 이미지에서 배경이 되는 풍경에서 지형부분들을 vertex color 정보를 받아들여 색상 정보에 따라 각기 다른 지형 텍스쳐를 적용하였다.

그림 6: Emission before bloom effect

그림 7: Emission after bloom effect

위 이미지는 bloom효과를 사용한 전 후 이미지이다. Unity 내에서는 기존 렌더링된 씬에서 추가적으로 렌더링 효과를 더하는 작업인 Post processing을 지원한다. 아래는 Post processing에 bloom 효과를 오버라이드하여 추가한 전후 모습이다.

물리 기반 렌더링으로 처리되는 반사와 굴절 효과를 위해 노이즈 기반의, bump map을 설정하였다. 아래 이미지는 물이 시간에 따라 흐르는 효과를 구현한 쉐이더의 모습이다.

그림 8: 시간에 따라 파도가 일렁이는 물 쉐이더

3.3. 가상회화 콘텐츠 전시 및 결과

본 콘텐츠는 세종대학교 광개토관 갤러리관에서 2019년 12월 4일에서 5일까지 이틀 간 전시되었다. 콘텐츠를 체험하는 관객들이 반응을 조사한 결과, 모든 관객들은 콘텐츠에 만족하였으며, 회화를 하나의 세계관으로 인식하여 크게 흥미를 보였다. 공통적으로 “회화의 세계를 걷는 것이 신기하다”, “단편적으로 보았던 회화를 산책하는 것과 같아 신선하다”라는 긍정적인 반응을 보였다

4. 결론 및 향후 계획

실사용자들의 반응을 토대로 가상전시관은 크게 이목을 끄기에 충분하며, 손쉽게 어디에서나 회화를 즐길 수 있는 기회를 제공한다. 본 논문에서 제시한 콘텐츠는 더 나아가 전시작품들에 대한 설명이나 도슨트 안내가 텍스트 또는 오디오로 제공할 수 있게 기능을 추가하여 작품 속을 이동하면서 동시에 전문적인 전시 해설을 듣거나 볼 수 있게 발전시키고 매끄러운 진행을 위한 동선을 재배치할 것이다. 또한 좋은 퍼포먼스와 저사양 환경에서도 즐길 수 있는 콘텐츠로 나아가기 위해 Vulkan API 적용을 할 예정이며 경량 렌더링 파이프라인(Lightweight rendering pipeline)을 도입하여 콘텐츠를 개선할 예정이다. 가상전시 콘텐츠는 남녀노소 누구나 쉽게 즐길 수 있는 하나의 장르로 나아가길 기대해본다

참고문헌

[1] 2017년 문화예술행사 관람률 변화 추이 중 국민문화예술활동조사, 문화예술행사 관람률 변화 추이
[2] 김가영, 미술 전시 관람객의 작품 감상 활성화를 위한 사용자 경험 디자인, p.21, 2015

기하학 컨볼루션 신경망을 이용한 3D 구강 메시의 개별 치아 영역 분할*

최규진[0], 이윤진, 경민호
아주대학교 라이프미디어협동과정
{paganinist, yunjin, kyung}@ajou.ac.kr

3D Tooth Instance Segmentation using Geometric Convolutional Network

Gyujin Choi, Yunjin Lee, Minho Kyung
Lifemedia Interdisciplinary Program, Ajou University

요약

디지털 치아 교정 시스템에서는 3D 스캐너로 얻은 환자의 3D 구강 스캔 데이터로부터 3D 프린팅을 위한 교정기 모델을 생성한다. 이를 위해서는 구강 스캔 데이터로부터 개별 치아를 분할하는 과정이 필수적으로 선행되어야 한다. 본 연구에서는 메시에서 오차나 위상에 강건한 특징을 추출하고, 메시를 점군으로 변환한 후 이 특징들을 사용하여 개별 치아 영역 분할을 찾는 방법을 제안한다. 본 연구에서 제안한 치아 영역 분할 방법은 기존 연구들에 비해 정확하게 개별 치아 영역들을 분리함을 실험으로 확인하였다.

1. 서론

디지털 치아 교정을 위해서는 환자의 구강 내부에 대한 3D 모델이 필요하다. 3D 구강 모델은 대부분 치과에서 제작한 인상을 3D 스캐너로 스캔하여 얻는다. 이 때, 구강 인상에는 컬러 정보가 없기 때문에 개별 치아를 구분하기 위해서는 순수하게 형상 정보만을 이용해야 하는 어려움이 있다. 치아 영역 분할이 어려운 이유는 다음과 같다. 첫번째로, 환자에 따라 치아 형태와 배열이 치과학에서 정립된 구강 이론에서 많이 벗어나 있다는 점이다. 예를 들어 덧니가 있거나 밀집된 치아는 위치와 방향이 불규칙해 포물선 형태의 아치에 맞지 않는 치열을 이룬다. 또한 교합 면의 마모로 인해서 해부학적 랜드마크가 사라지기도 한다. 두 번째로, 스캐닝 장치와 메시 전처리 과정에서 발생하는 오차로 인한 정보의 상실이다. 표면 정점에 발생한 스캐닝 오차는 치간 혹은 치은 경계를 희미하게 만들어 영역 분할을 어렵게 한다. 최근에 3D 구강 모델의 객체 영역 분할(instance segmentation) 성능 향상을 위해 다양한 연구들이 이루어지고 있다. 이 연구들에서는 메시 대신 점군을 이용하거나[2], 혹은 저자들이 직접 고안한 메시 표면의 형상 서술자(shape descriptor)를 사용하였다[3]. 특히 Xu 등[3]이 제안한 방법은 전처리 단계에서 형상 서술자의 계산에 많은 시간을 요구하고, 충분한 정확성을 얻기위해 대량의 훈련 데이터를 사용하였다. 본 논문에서는 이를 개선하기 위해 Hanocka 등[4]이 제안한 MeshCNN을 사용하여 메시 표면에서 간선을 기반으로 오차나 위상에 강건한 특징을 추출한 후, Pham 등[5]이 제안한 JSIS3D의 구조를 사용하여 객체 영역분할을 계산하였다.

* 포스터 발표논문

2. 개별 치아의 영역 분할 방법

2.1. 데이터 증강 및 전처리

변이가 많은 구강 모델에서도 네트워크가 강건하게 치아를 분할할 수 있도록 하기 위해 데이터를 증강하였다. 본 논문의 방법은 데이터로부터 유사 불변한 특징을 학습한다. 따라서 이동, 회전, 등방성 확대한 데이터는 새로운 특징을 가진 데이터를 만들어내지 못한다. 그래서 x, y, z축에 대한 비등방성 확대, 메시 표면 노이즈, 메시의 테셀레이션 변경으로 데이터를 만들어냈다. 마지막으로, 학습, 추론 시간을 단축하고 네트워크에 균일한 크기의 데이터를 입력하기 위해 신경망에 입력되는 모든 데이터의 정점의 개수를 간선 병합(edge collapse)을 이용하여 9192개로 통일하였다.

2.2. 네트워크 설계

본 연구에서 실험에 사용한 네트워크는 크게 두 부분으로 나뉜다(그림 1 참고). 첫번째는 Hanocka 등[4]이 제안한 방법을 바탕으로 입력메시에 대하여 간선 기반으로 특징을 추출하는 부분이며, 3개의 EdgeConv레이어로 이루어져 있다. EdgeConv 레이어에서는 간선 주변의 기하학적 특성을 나타내는 5차원의 특징 벡터가 사용된다. 특징 벡터는 이면각, 간선에 인접한 두개의 삼각형의 대각, 그리고 간선에 인접한 두개의 삼각형의 반대쪽

꼭지점에서 간선으로 내린 수선의 발까지의 길이와 간선의 길이의 비율로 구성된다. 간선 단위의 특징은 각각 레이어에서 Wang 등[6]이 제안한 방법을 이용하여 정점 단위의 특징으로 변환하였다. 두번째는, Pham 등[5]이 제안한 방식을 이용하여 객체 영역 분할을 계산하는 부분이다. 메시 표면에 대한 최종적인 객체 영역 분할은 임의의 면을 이루는 정점들의 확률을 평균내어 가장 높은 확률값을 가지는 객체, 혹은 범주로 면을 할당하였다. 네트워크의 모든 레이어에는 배치 정규화와 ReLU 활성화 함수를 적용하였다.

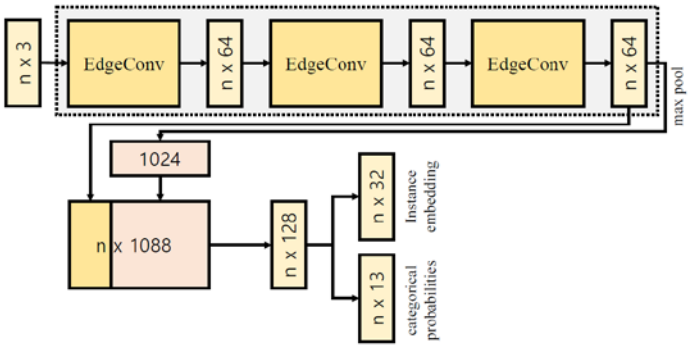


그림 1: 본 논문에서 제안하는 신경망의 구조

2.3. 학습

신경망을 최적화하기 위해 SGD(Stochastic Gradient Descent)를 사용하였다. 학습 속도(learning rate)는 0.01로 설정하였으며, 0.001이 될 때까지 코사인 어닐링(cosine annealing)[7]을 사용하여 감소시켰다. 배치 정규화(batch normalization)의 모멘텀(momentum)은 0.9, 그리고 배치 정규화를 적용하였기 때문에 가중치 감쇠(weight decay)는 사용하지 않았다. 배치의 크기는 32, 모멘텀은 0.9로 설정하였다.

3. 실험 및 결과

총 8개의 구강 메시를 사용하여 네트워크를 훈련하였다. 학습된 네트워크의 분할 성능은 학습에 사용하지 않고 제외해 둔 8개의 구강 메시를 테스트 데이터 셋으로 사용하여 분할 성능을 검증하였다. JSIS3D 만을 이용한 분할 결과에서는 치은선의 경계가 제대로 분할되지 않거나 특정한 치아를 인식하지 못하는 등의 문제가 있었으나, 본 논문의 방법을 사용한 이후에는 이러한 문제를 상당 부분 해결할 수 있었다(그림 2 참고).

표 1: 분할 성능

Network Arch	IoU	Mean Elapsed time
JSIS3D[5]	0.863	2.89s
Proposed	0.941	3.01s

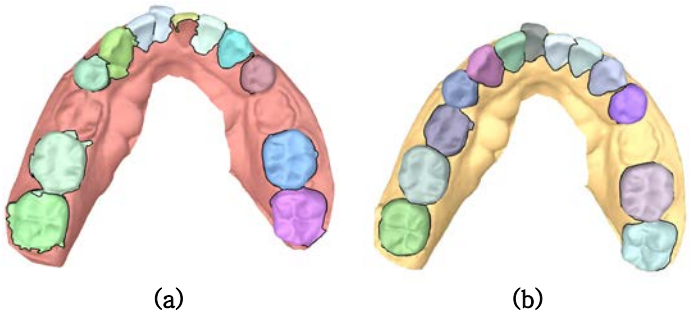


그림 2: (a)는 JSIS3D, (b)는 논문에서 제안하는 방법으로 테스트 셋의 샘플을 분할한 결과이다.

4. 결론 및 향후 연구 방향

본 논문에서는 3D 구강 메시에서 개별 치아 영역을 분할하는 새로운 방법을 제안하였다. 제안한 방법은 기존 연구에 비해 적은 훈련 데이터만으로도 정확한 분할 결과를 보여 주었다. 추후에는 영역 분할 뿐만 아니라, 치아의 종류까지도 분류할 수 있도록 신경망을 개선할 것이다. 또한, 메시 표면에서 객체 영역 분할을 직접적으로 계산할 수 있는 방법도 연구해 볼 것이다.

참고문헌

[1] Kim, Taeksoo, et al. "Tooth Segmentation of 3D Scan Data Using Generative Adversarial Networks." *Applied Sciences* 10.2 (2020): 490.
[2] Zanjani, Farhad Ghazvinian, et al. "Mask-MCNet: Instance Segmentation in 3D Point Cloud of Intra-oral Scans." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2019.
[3] Xu, Xiaojie, Chang Liu, and Youyi Zheng. "3D tooth segmentation and labeling using deep convolutional neural networks." *IEEE transactions on visualization and computer graphics* 25.7 (2018): 2336-2348.
[4] Hanocka, Rana, et al. "MeshCNN: a network with an edge." *ACM Transactions on Graphics (TOG)* 38.4 (2019): 1-12.
[5] Pham, Quang-Hieu, et al. "JSIS3D: joint semantic-instance segmentation of 3d point clouds with multi-task pointwise networks and multi-value conditional random fields." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
[6] Wang, Yue, et al. "Dynamic graph cnn for learning on point clouds." *ACM Transactions On Graphics (TOG)* 38.5 (2019): 1-12.
[7] Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts." *arXiv preprint arXiv:1608.03983* (2016).

한국컴퓨터그래픽스학회 2020 학술대회 논문집

2020년 7월 1일 인쇄
2020년 7월 1일 발행

발행인: 이제희
편집인: 이성길, 이윤진
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

주최 | 한국컴퓨터그래픽스학회

후원 |