

양방향 LSTM을 이용한 손 접촉상태 검출 방법

강창구*, 이석원**

Detecting Hand Contact State using Bi-directional LSTM

Changgu Kang*, Sukwon Lee**

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2022년도 문화기술 연구개발 사업으로 수행되었음
(과제명: 복합리조트(Integrated Resort) 산업의 안전을 위한 AI 학습용 데이터 구축 및 지능형 서베일런스 플랫폼
기술개발, 과제번호: R2022020028, 기여율: 100%)

요 약

일상에서 사람들은 걷기, 문 열기, 물 마시기, 통화하기, 앉기 등 지속해서 환경과 다양한 물리적 상호작용과 함께 생활한다. 이와같은 상호작용에 관한 다양한 연구들이 진행되고 있다. 물리적 상호작용은 접촉에서 시작되며, 사람의 접촉 부분에 따라 상호작용 동작들이 구분될 수 있다. 이러한 중요성에도 불구하고 대부분의 연구들은 상호작용 접촉 자세 또는 행동유도성을 중심으로 진행되었으며 물리적 대상에 대한 접촉 유무를 판단하는 연구는 결핍되어 왔다. 이러한 상황을 극복하기 위하여 우리는 물리적 상호작용에서 가장 접촉이 많이 발생하는 손 접촉상태 검출 방법을 제안하였다. 제안된 방법은 양방향 LSTM 기반의 특징벡터 생성 부분과 접촉 검출 부분으로 구성된다. 그리고 접촉 검출 부분은 왼손과 오른손이 구분된 학습구조를 가진다. 우리는 수집된 모션 데이터를 이용하여 모델 학습을 수행하며, 학습된 모델의 테스트 결과 93.9% 정확도를 보여준다.

Abstract

In everyday life, people continue to live with various physical interactions with the environment, such as walking, opening the door, drinking water, talking on the phone, and sitting down. Various studies on such interactions are being conducted. The physical interaction starts with contact, and interaction operations may be classified according to the contact portion of the person. Despite this importance, most studies have focused on interactive contact posture or behavioral induction, and studies on determining whether or not to contact a physical object have been lacking. To overcome this situation, we propose a method for detecting hand contact states where the most frequent contact occurs in physical interactions. The proposed method consists of a bidirectional LSTM based feature vector generation part and a contact detection part. In addition, the contact detection part has a learning structure in which the left and right hands are divided. We perform model learning using the collected motion data, and the test results of the learned model shows 93.9% accuracy.

Keywords

hand contact detection, bi-directional lstm, attention, recursive neural network, interaction

* 경상국립대학교 컴퓨터과학부 교수
- ORCID: <https://orcid.org/0000-0003-4060-6835>
** 한국전자기술연구원 선임연구원(교신저자)
- ORCID: <https://orcid.org/0000-0002-3005-3707>

· Received: Oct. 11, 2022, Revised: Oct. 25, 2022, Accepted: Oct. 28, 2022
· Corresponding Author: Sukwon Lee
Senior Researcher, VR/AR Research Center, Broadcasting & ICT R&D Division, Korea
Tel.: +82-02-6388-6604, Email: sukwonlee@keti.re.kr

I. 서 론

일상에서 사람들은 걷기, 문 열기, 물 마시기, 통화하기, 앉기 등 지속해서 환경과 다양한 물리적 상호작용과 함께 생활하며, 이러한 상호작용 연구를 위해 다양한 3차원 모션 데이터와 환경 데이터들이 만들어졌다[1][2]. 상호작용에 관한 모션들은 물리적 대상의 행동유도성(Affordance)이 고려되며, 행동인식과 인간-환경 상호작용에 관한 연구들은 행동유도성을 발견하는 방법들을 제안하였다[3]-[5]. 일반적으로 상호작용 모션들은 행위자의 손 접촉으로부터 시작되며, 상호작용 대상(물체 또는 환경)의 행동유도성에 따라 적합한 접촉 위치를 찾게 된다. 하지만 대부분의 연구에서는 행동유도성을 중심으로 수행됐으며 행위자의 손 접촉 여부에 관한 연구들은 결핍되어 왔다.

우리는 사람의 모션과 3차원 환경 데이터만 있으면 쉽게 사람의 손 접촉 상태를 확인할 수 있으며, RGB-D 카메라를 이용하면 두 데이터의 획득이 가능하다. 하지만 사람의 모션 캡처 연구와 3차원 장면 재생 연구들은 알고리즘의 복잡도와 컴퓨터 성능 문제로 인해 각각 구분되어 진행되었다. 구분되어 준비된 데이터가 존재한다고 하여도 좌표계를 맞춰야 하는 다른 문제를 일으키며, 모션 데이터는 사람의 표면 정보가 제외된 뼈대 형식으로 환경과의 기하학적 교차로 인한 판단에는 정확도가 낮다. 더욱이 IMU 센서 기반의 장비는 환경정보 획득이 불가능하다. 최근 이미지로부터 사람 손 접촉을 검출하는 연구는 진행되었다. 하지만 연속적인 정보가 아닌 정적인 정보를 제공하고 손에 대한 구분을 제공하지 않는다[6].

이러한 제한적인 요소들을 해결하기 위해 우리는 3D 장면 데이터가 제외된 모션 데이터만을 이용하여 사람의 양손이 구분된 접촉상태 검출 방법을 제안한다. 모션 데이터는 연속적인 특성이 있으므로, 제안된 방법은 순환신경망 구조를 가진다. 그리고 모션 데이터의 전체 모션 정보를 이용하여 특징벡터로 만들고 제안된 방법의 입력으로 이용하게 된다. 제안된 모델 학습을 위해 데이터로서 손의 접촉이 발생하는 수집된 모션은 ‘문’, ‘캐비닛’, ‘의자’,

‘전화기’, ‘물컵’ 들과 상호작용하는 동작들로 구성한다. 그리고 수집된 모션 데이터들은 프레임마다 접촉 상태를 표기한다.

본 논문의 구성은 다음과 같다. 2장은 상호작용 분석 및 접촉 검출에 관한 연구동향을 기술한다. 3장에서는 데이터수집과 제안된 손 접촉상태 검출 모델을 상세히 기술한다. 4장에서는 제안된 방법의 성능을 평가하고 실험결과를 설명한다. 마지막으로 5장에서는 결론에 대해 기술하고 추후연구에 대하여 기술한다.

II. 관련 연구

상호작용분석을 위해서는 사람의 자세 및 접촉정보가 필요하다. 하지만 환경의 복잡도에 따라 정확한 자세 및 접촉 데이터 획득이 어렵다. Hassan은 이러한 문제를 해결하기 위해 불완전하게 추정된 자세 위치를 환경 데이터를 이용하여 적합한 자세 위치를 조정해주는 연구를 수행하였다[7]. 본 연구에서 조정된 자세를 통해 환경과 상호작용 자세 간의 접촉정보를 정확히 검출할 수 있지만 정적인 자세 정보만을 이용하며 환경 데이터가 수반된다는 한계점을 가진다. 접촉정보에서 발 접촉 정보는 물리적인 제약을 제공함으로써 모션 캡처에 대한 정확도를 증가시킨다. Ma 는 발 접촉정보 검출 방법에 관한 연구를 수행하였다. 제안된 검출 방법에는 모션 데이터와 행위자의 무게정보가 함께 포함되며 모션데이터만 이용이 불가능하다[8]. Nar 는 복잡한 환경에서의 손 접촉 상태 검출 방법을 제안하였다 [6]. 제안된 방법은 Mask RCNN[9]과 어텐셔널 풀링 모듈(Attentional pooling module)[10]을 이용하여 입력된 이미지로부터 접촉이 발생한 손의 위치를 정확히 찾아내며, 매우 복잡한 상황의 이미지에서도 다수의 접촉 된 손들을 찾아낸다. 하지만 제안된 방법은 정적인 이미지에 동작하기 때문에 연속적인 접촉정보를 제공하지 못하며, 양손이 구분된 접촉정보를 제공하지 못한다. 본 연구에서는 모션 데이터를 이용하여 각 프레임별 접촉 상태를 검출하는 방법을 제안하며 오른손과 왼손이 구분된 접촉 상태 정보를 제공한다.

III. 데이터 수집과 손 접촉상태 검출 모델

3.1 데이터 수집

모션 데이터 수집을 위해 사용되는 장비는 크게 영상 기반의 장비와 IMU 센서를 이용한 장비로 구분된다. 우리는 환경에 있는 가구의 가림 현상에도 모션이 잘 캡처 될 수 있는 IMU 센서 기반의 장비인 퍼셉션뉴런(Perception neuron)을 이용한다. 표 1은 수집된 상호작용 대상과 세부 내용을 보여준다.

표 1. 상호작용 물체와 세부 내용
Table 1. Interaction object and motion details

Object	Motion details
Door	Open the door and enter
Cabinet	Open and close the cabinet
Chair	Sit with hands on desk
Telephone	Answer a phone
Cup	Drink water

수집된 데이터의 움직임은 손을 이용하여 ‘캐비닛’, ‘책’, ‘전화기’, ‘의자’, ‘물컵’을 이용한 동작들이다. ‘문’의 경우 미닫이문을 이용하며 사람이 문을 열고 들어가는 행동을 캡처하였다. 그리고 ‘캐비닛’은 양문이 있는 캐비닛 문을 열고 캐비닛 안에 있는 서류를 접촉 후 다시 문을 닫는 행동을 캡처하였다. 다음으로 ‘의자’의 상호작용 행동은 책상에 손을 딛고 앉는 행동을 캡처한다. 마지막으로 ‘전화기’와 ‘물컵’은 각 물체를 들어서 목적에 맞게 수행한 후 다시 제자리에 놓는 행동을 캡처한다. 행동마다 20개의 모션 데이터를 수집하였다. 수집된 모션 데이터를 이용하여 수동으로 왼손과 오른손의 접촉상태를 표기한다. 접촉상태는 0과1로 표기되며 0은 미접촉, 1은 접촉으로 정의한다.

3.2 손 접촉상태 검출 모델

손 접촉상태 검출 모델의 입력으로 각 관절의 각을 입력으로 사용한다. 각 관절은 매우 세분되어 있지만 그중에서 주요한 각 팔(어깨, 팔꿈치, 손목)과 다리(고관절, 무릎, 발목)에 3개의 관절 각을 입력으로 모두 사용한다.

손 접촉상태 검출 모델은 크게 3개의 모듈로 구성된다. 첫 번째 모듈은 연속된 관절의 각도 값으로 접촉 검출을 위한 특징을 만든다. 모션 데이터는 연속적인 특징을 가지므로 순환신경망 구조 중 양방향 LSTM(Bidirectional LSTM) 구조[11]를 적용하였다. 양방향 LSTM 모두 4개의 층으로 구성하였으며, 모두 32x1의 출력크기를 가진다.

두 번째 모듈에는 특징벡터의 크기가 줄어들고 줄어든 크기에 대한 정보손실 및 기울기 소실 문제를 해결하기 위해서 어텐션 레이어를 적용하였다 [12]. 어텐션 레이어에는 내적 어텐션을 적용하여 어텐션 스코어를 계산하게 된다. 양방향 LSTM의 정방향과 역방향의 기억값의 병합 후 어텐션 레이어 입력값 내적으로 스코어가 계산된다. 다음으로 소프트맥스(Softmax) 함수를 적용하여 전체합이 1인 어텐션 가중치를 결정한다. 마지막으로 양방향 LSTM의 출력 값과 어텐션 가중치를 곱하고 차원 축소 합을 적용하여 64x1 크기의 문맥 벡터를 만든다.

마지막 모듈은 입력된 특징들을 해석하여 왼손과 오른손에 대한 접촉정보를 검출하기 위하여 모든 노드들이 연결된 완전 연결 레이어 구조를 가진다. 완전 연결 레이어는 4개의 층으로 구성되며, 각 중간층은 8x1, 마지막 층은 1x1 출력크기를 가진다. 양손은 서로 독립된 접촉상태를 가지므로 서로 독립된 완전 연결 레이어를 가진다. 손실함수와 최적화 알고리즘은 최소제곱오차와 아담(Adam)[13]을 각각 적용한다.

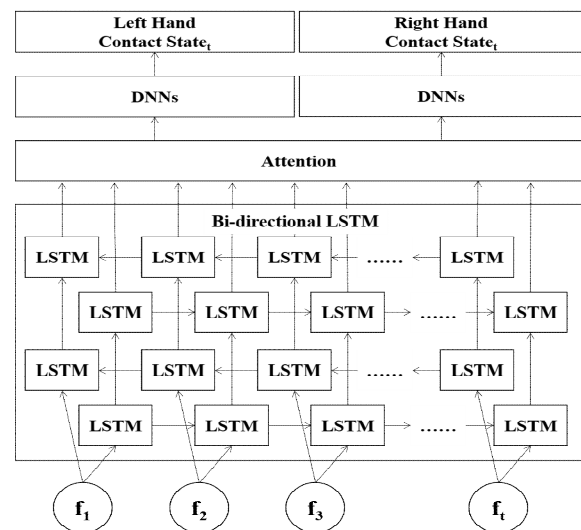


그림 208. 손 접촉상태 검출 모델
Fig. 1. Model for detecting hand contact state

IV. 실험

우리는 손 접촉상태 검출 모델에 대한 학습과 학습된 모델을 이용하여 접촉상태 검출 실험을 수행하였다. 수집된 120개 모션데이터를 이용하여 모두 5그룹으로 나누고 4개의 그룹은 학습 데이터 그리고 1개 그룹을 테스트 데이터로 이용하였다. 각 그룹당 24개의 모션 데이터가 포함된다.

모델 학습을 위하여 학습 데이터 시퀀스 (Sequence) 길이는 32, 배치(Batch) 크기는 32, 학습률은 0.001로 각각 설정하였다. 에포크(Epoch)를 20로 설정하고 학습을 수행하였으며 그림 2는 왼손과 오른손의 학습 과정에서 나타난 손실변화를 보여준다. 왼손과 오른손의 손실 값은 지속해서 작아지면서 점점 수렴되고 있음을 확인할 수 있다. 주로 상호작용 행동들이 왼손보다 오른손에 접촉이 많이 발생하기 때문에 왼손의 손실이 더 작게 나타난다.

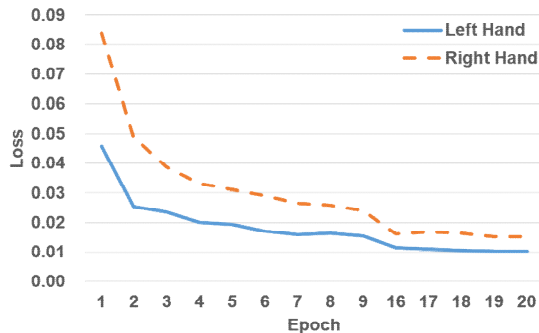


그림 2. 모델 학습과정에서 손실변화
Fig. 2. Training loss

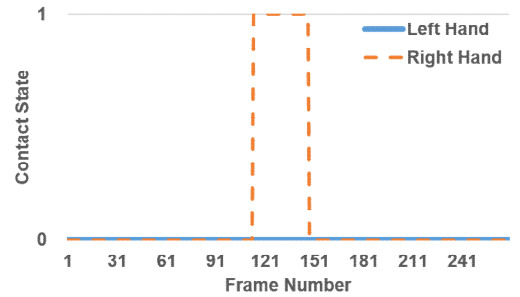
다음으로 테스트 모션 데이터를 이용하여 양손에 대한 접촉상태 검출 실험을 수행하였다. 표 2는 접촉상태 검출 결과에 대한 오차행렬을 보여준다.

표 2. 접촉상태 검출 결과의 오차 행렬
Table 2. Confusion matrix for experiments

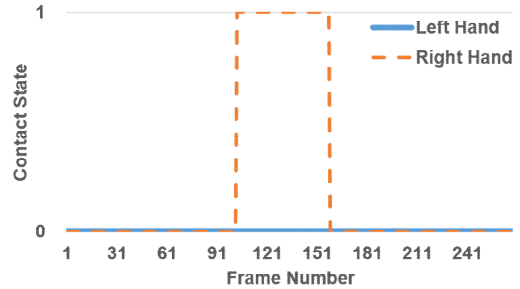
Left hand		Detected values	
		1	0
Actual values	1	2205	114
	0	466	9195
Right hand		Detected values	
		1	0
Actual values	1	3508	408
	0	461	7603

실험 데이터의 전체 프레임 수는 11,980이며 검출된 접촉상태의 정확도는 오른손 약 92.7%, 왼손 약 95.1% 로 각각 나타났다. 양손 전체 접촉상태에 정확도는 약 93.9% 이다.

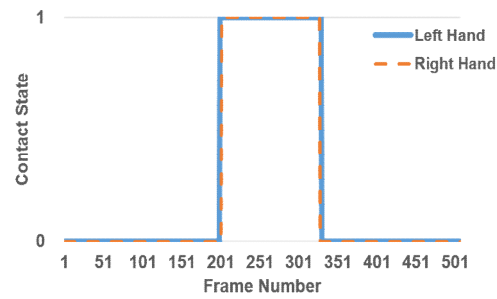
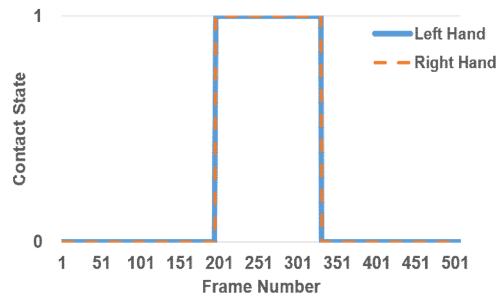
그림 3은 ‘문 열고 들어가기’, ‘양문 캐비닛을 열고 닫기’, ‘손을 책상에 딛고 앉기’, ‘수화기를 받고 제자리에 놓기’ 상호작용 동작에 대한 접촉상태 검출 결과 그래프를 보여준다.



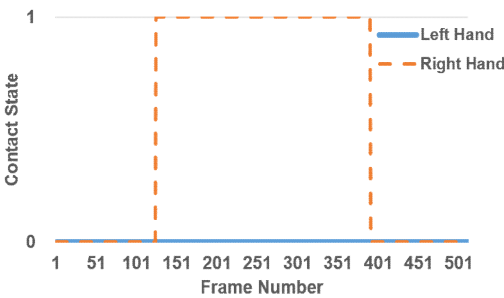
(a) 문 열고 들어가기
(a) Open and enter the door



(b) 양문 캐비닛을 열고 닫기
(b) Open and close the two-door cabinet



(c) 손을 책상에 두고 앉기
(c) Sit with your hand on the desk



(d) 수화기를 받고 제자리에 놓기
(d) Receive and place the phone in place

그림 3. 접촉상태 검출 결과 그래프
Fig. 3. Graph for the result of contact detection

행동에 대한 실험 데이터에 대한 양손 접촉 상태 검출 결과 그래프(위)와 실제 값 그래프(아래)를 보여준다. Y 축은 0과 1로 표현된 접촉 상태를 나타내며, X 축은 모션의 프레임 숫자를 나타낸다.

(a), (b), (c), (d)의 그래프에서 실험 결과 그래프와 실제 값 그래프가 모두 유사한 형태를 가지고 있음을 확인할 수 있다. 두 그래프를 비교해보면 대부분의 에러가 접촉상태 변화에서 발생하고 있음을

확인할 수 있다. 이유는 모션 데이터 접촉 여부를 판단할 때 사람의 표면적과 손가락 관절이 고려되지 않고 전체 모션의 행태만으로 접촉상태를 구분하기 때문에 접촉이 발생한 시점을 정확히 확인하기 어렵다. 그러므로 접촉상태가 변화되는 시점을 기준으로 오차가 발생하게 된다.

그림 4는 테스트 데이터중 문열기, 양손으로 캐비닛 열기, 손 딛고 책상에 앉기, 물 마시기 동작에 대한 접촉상태 검출 결과 모션들을 보여주며 접촉이 발생한 손은 파란색으로 표시하였다.

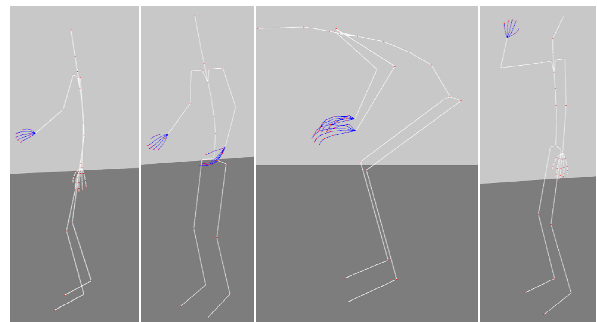


그림 4. 접촉상태 검출 결과 모션: 문 열기, 양손으로 캐비닛 열기, 책상에 손 딛고 앉기, 물 마시기 동작
Fig. 4. Motion for the result of contact state detection: Open the door, open the cabinet with both hands on a desk, and drink water

V. 결론 및 향후 연구

물리적 상호작용은 접촉에서 시작되며, 접촉되는 부분에 따라 상호작용 동작들이 구분될 수 있다. 이러한 중요성에도 불구하고 대부분의 연구들은 상호작용 접촉 자세 또는 행동유도성을 중심으로 진행되었으며 물리적 대상에 대한 접촉 유무를 판단하는 연구는 결핍되어 왔다. 이러한 상황을 극복하기 위하여 우리는 모션 데이터로부터 양방향 LSTM을 이용한 손 접촉상태 검출 방법을 제안하였다.

수집된 모션 데이터를 이용하여 모델 학습을 수행하였으며, 제안된 방법을 통하여 평가하였다. 실험 결과 약 93.9% 양손의 검출 정확도를 보여주며, 대부분의 에러가 접촉변화시점에서 발생함을 확인할 수 있었다. 향후 연구로서 우리는 상호작용에 이용되는 몸의 다양한 부분의 접촉상태 검출 연구를 수행할 계획이며, 최종적으로 접촉상태 기반의 상호작용 행동 검출 방법을 연구할 계획이다.

References

- [1] Changgu Kang and Sung-Hee Lee, "Scene Reconstruction and Analysis from Motion", *Graphical Models*, Vol. 94, No. 2, pp. 25-37, Nov, 2017. <https://doi.org/10.1016/j.gmod.2017.10.002>.
- [2] Savva Manolis, et al., "SceneGrok: Inferring action maps in 3D environments", *ACM Transactions on Graphics(TOG)*, Vol. 66, No. 6, pp. 1-10, Nov. 2014. <https://doi.org/10.1145/2661229.2661230>.
- [3] Hu Ruizhen, et al., "Interaction context(ICON) towards a geometric functionality descriptor", *ACM Transactions on Graphics(TOG)*, Vol. 34, No. 4, pp. 1-12, Aug. 2015. <https://doi.org/10.1145/2766914>.
- [4] Nguyen Anh, et al., "Detecting object affordances with convolutional neural networks", *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*, pp. 2765-2770, Oct. 2016. <https://doi.org/10.1109/IROS.2016.7759429>.
- [5] Kovic Mia, et al., "Affordance detection for task-specific grasping using deep learning", *IEEE-RAS 17th International Conference on Humanoid Robotics(Humanoids)*, Birmingham, UK, pp. 91-98, Nov. 2017. <https://doi.org/10.1109/HUMANOIDS.2017.8239542>.
- [6] Narasimhaswamy Supreeth, Trung Nguyen, and Minh Hoai Nguyen, "Detecting hands and recognizing physical contact in the wild", *34th Conference on Neural Information Proc. Systems(NeurIPS 2020)*, pp. 7841-7851, 2020. <https://doi.org/10.48550/arXiv.2010.09676>.
- [7] Hassan Mohamed, et al., "Resolving 3D human pose ambiguities with 3D scene constraints", *Proc. of the IEEE/CVF international conference on computer vision*, pp. 2282-2292, 2019. <https://doi.org/10.1109/ICCV.2019.00237>.
- [8] H. Ma, W. Yan, Z. Yang, and H. Liu, "Real-Time Foot-Ground Contact Detection for Inertial Motion Capture Based on an Adaptive Weighted Naive Bayes Model", *IEEE Access*, Vol. 7, pp. 130312-130326, 2019. <https://doi.org/10.1109/ACCESS.2019.2939839>.
- [9] He Kaiming, et al., "Mask R-CNN", *Proc. of the IEEE international conference on computer vision*, 2017. <https://doi.org/10.48550/arXiv.1703.06870>.
- [10] Girdhar Rohit and Deva Ramanan, "Attentional pooling for action recognition", In *Proc. of the 31st International Conference on Neural Information Processing Systems*, pp. 33-44, Dec. 2017. <https://doi.org/10.48550/arXiv.1711.01467>.
- [11] Graves Alex, Navdeep Jaitly, and Abdel-rahman Mohamed, "Hybrid speech recognition with deep bidirectional LSTM", In *2013 IEEE workshop on automatic speech recognition and understanding*, pp. 273-278, Dec. 2013. <https://doi.org/10.1109/ASRU.2013.6707742>.
- [12] Devlin Jacob, et al., "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv preprint arXiv:1810.04805* Oct. 2018. <https://doi.org/10.48550/arXiv.1810.04805>.
- [13] Kingma Diederik P. and Jimmy Ba., "Adam: A method for stochastic optimization", *3rd International Conference for Learning Representations*, San Diego, Vol. 1, 2014. <https://doi.org/10.48550/arXiv.1412.6980>.

저자소개

강 창 구 (Changgu Kang)



2010년 2월 : 광주과학기술원
정보기전공학부(공학석사)
2017년 8월 : 광주과학기술원
전기전자컴퓨터공학부(공학박사)
2018년 3월 ~ 현재 :
경상국립대학교 컴퓨터과학부
부교수

관심분야 : 컴퓨터 그래픽스, 증강현실, 인공지능

이 석 원 (Sukwon Lee)



2013년 2월 : 광주과학기술원
정보기전공학부(공학석사)
2019년 2월 : 한국과학기술원
문화기술대학교(공학박사)
2019년 6월 ~ 현재 :
한국전자기술연구원 선임연구원
관심분야 : IMU 기반 저 비용

전신 모션 캡처, RGB/D 손 자세 추적, 전신 모션
합성