

## Generating Augmented Lifting Player using Pose Tracking

Jong-In Choi\*, Jong-Hyun Kim\*\*

\*Professor, Dept. of Digital Media, Seoul Women's University, Seoul, Korea

\*\*Professor, Dept. of Software Application, Kangnam University, Yongin, Korea

### [Abstract]

This paper proposes a framework for creating acrobatic scenes such as soccer ball lifting using various users' videos. The proposed method can generate a desired result within a few seconds using a general video of user recorded with a mobile phone. The framework of this paper is largely divided into three parts. The first is to analyze the posture by receiving the user's video. To do this, the user can calculate the pose of the user by analyzing the video using a deep learning technique, and track the movement of a selected body part. The second is to analyze the movement trajectory of the selected body part and calculate the location and time of hitting the object. Finally, the trajectory of the object is generated using the analyzed hitting information. Then, a natural object lifting scenes synchronized with the input user's video can be generated. Physical-based optimization was used to generate a realistic moving object. Using the method of this paper, we can produce various augmented reality applications.

▶ **Key words:** Augmented Reality, Artificial Intelligence, Deep Neural Networks, Posture Tracking, Video Synthesis

### [요 약]

본 논문에서는 다양한 사용자의 영상을 이용하여 축구공 리프팅과 같은 묘기 장면을 만들 수 있는 프레임워크를 제안한다. 제안된 방법은 핸드폰 등으로 촬영된 일반적인 사용자의 영상이라면 수 초 이내에 원하는 결과를 생성할 수 있다. 본 논문의 프레임워크는 크게 세 부분으로 나누어진다. 첫 번째는 사용자의 영상을 입력받아 자세를 분석하는 것이다. 이를 위해서는 딥러닝 기법으로 영상을 분석하여 사용자의 포즈를 계산하고, 원하는 신체 부위의 움직임을 추적할 수 있다. 두 번째는 지정된 신체부위의 이동 궤적을 분석하여 물체를 타격하는 위치와 시간을 계산하는 것이다. 마지막으로 분석된 타격 정보를 이용하여 물체의 이동 궤적을 생성하는 것이다. 그러면 입력된 사용자 영상과 동기화되는 자연스러운 물체 리프팅 장면을 생성할 수 있다. 사실적인 물체의 움직임을 생성하기 위해 물리 기반 최적화를 사용하였다. 본 논문의 프레임워크를 이용하면 다양한 증강현실 어플리케이션을 제작할 수 있다.

▶ **주제어:** 증강 현실, 인공 지능, 심층 신경망, 자세 추적, 영상 합성

- 
- First Author: Jong-In Choi, Corresponding Author: Jong-Hyun Kim
  - \*Jong-In Choi (funtech@swu.ac.kr), Dept. of Digital Media, Seoul Women's University
  - \*\*Jong-Hyun Kim (jonghyunkim@kangnam.ac.kr), Dept. of Software Application, Kangnam University
  - Received: 2020. 04. 03, Revised: 2020. 04. 27, Accepted: 2020. 04. 27.

## I. Introduction

4차 산업혁명시대의 화두는 인공지능을 꼽을 수 있으며, 현재 인공지능은 다양한 영역에 확대 적용되면서 엄청난 속도로 기술의 발전을 이끌고 있다. 인공지능 중에서도 머신러닝 기술이 중요하다 할 수 있고, 머신 러닝 중에서도 딥 러닝(Deep Learning)이 가장 핵심적으로 떠오르고 있는 인공지능 기술이다. 전통적인 컴퓨터 그래픽스 기술에도 점차 이러한 딥 러닝 기법이 광범위하게 적용되고 있다. 캐릭터 애니메이션 분야에도 예외 없이 딥 러닝 기법이 적용되고 있다. 반복되는 이동 모션을 자동으로 생성하는 기법[1]이 있었고, 영상으로부터 인간의 동작을 학습할 수 있는 기법[2]까지 다양하게 등장하고 있다. 모션캡처 기술은 모션 데이터의 기반 캐릭터 애니메이션 분야에서는 필수적으로 필요한 기술이다. 이는 캐릭터 애니메이션에 필요한 다양한 모션들을 모션캡처 기법을 이용하여 획득할 수 있기 때문이다. 그러나 모션캡처는 여전히 비싸고 시간이 많이 걸리는 영역이다. 때문에 많은 비용과 시간이 투자되고 있다. 이러한 단점을 해결하기 위해 영상으로부터 모션 데이터를 추출하는 기법들[3, 4]이 속속 등장하고 있다. 그러나 아직까지는 고가의 모션캡처 장비의 성능을 따라가지는 못하는 것이 현실이다. 그러나 이러한 기법들은 모션 데이터의 추출 이외에도 다양한 방법으로 활용할 수 있는 잠재력을 가지고 있다. 가령 전문가의 동작과 일반 사용자의 동작을 비교하거나, 모션 데이터의 프로토타입으로 빠르게 동작을 확인하는 용도로 활용이 가능하다.

본 논문에서는 제안된 프레임워크를 구축하기 위해 딥러닝을 이용한 모션 분석 기법을 활용하였다. 본 논문에서는 영상으로부터 사람의 이차원 포즈를 추적하는 기법[5]을 활용하여 사용자의 동작을 실시간으로 분석하였다. 이는 영상 속에 있는 여러 사람들의 스켈레톤 구조를 분석하여 추출할 수 있는 기법이다. 삼차원 위치 정보까지는 제공해주지는 않지만, 이차원 이미지에서는 거의 정확하게 사람의 스켈레톤 구조를 추출할 수 있다. 제안된 프레임워크는 영상 속 인물에게 가상의 물체를 동기화시켜 묘기 장면을 생성하기 때문에 캐릭터의 삼차원 정보는 필요하지 않다. 또한 정밀한 이차원 위치나 정보를 요구하지 않기 때문에 해당 논문의 기법을 이용하면 충분히 원하는 목적을 달성할 수 있다. 본 프레임워크에서는 원하는 신체 부위의 타격 지점을 추출하여 물체의 움직임을 생성하여 동기화 시킨다. 그러면 누구나 다양한 물체를 다루는 묘기 장면을 쉽고 빠르게 생성할 수 있다. 제안된 방법은 사용자의 영상 위에 가상의 물체를 덧입혀 묘기 장면을 생성하

는 방식이므로 증강현실 분야에도 포함된다. 본 논문의 구성은 다음과 같다. 먼저 다음 장에서는 제안한 프레임워크에 관련된 기존 연구들을 다룬다. 3장에서는 사용자의 영상을 이용하여 묘기장면을 생성할 수 있는 프레임워크 설계와 구현에 대한 자세한 내용을 다룬다. 4장에서는 제안된 기법에 대한 실험 방법과 결과에 대한 내용을 기술한다. 마지막 장에서는 논문의 결론 및 향후 진행할 연구의 방향을 제시한다.

## II. Related Works

### 1. Human Pose Estimation

이미지에서 인간의 포즈를 추정하는 것은 개인의 신체 부위를 찾는 데 중점을 두고 있다[6-9]. 특히, 연관성 없는 여러 사람들이 동시에 등장하는 장면에서 각 사람의 포즈를 추정하는 것은 실로 어려운 문제이다. 이는 첫째로 각 이미지에 임의의 위치와 크기로 등장하는 알 수 없는 사람들이 있기 때문이다. 둘째로 사람 사이에 신체가 서로 접촉하는 상호작용을 하거나 타인의 신체 일부를 가리는 것으로 인해 복잡한 공간 간섭을 유발하여 정확한 각 사람의 포즈 탐색을 어렵게 만든다. 셋째로 등장하는 사람의 숫자가 늘어남에 따라 복잡성이 증가하는 경향이 있어서 실시간으로 모든 사람들의 포즈를 추출하기 어렵게 된다. 일반적인 접근법들[10-13]은 사람을 추출하는 탐지 알고리즘을 사용하고, 각각의 탐지에 대해 해당하는 각 사람의 포즈 추정을 수행하는 것이다. 이러한 하향식 접근 방식은 일인칭 포즈 추정을 위해 기존의 기법들[14-17]을 직접 활용하지만, 포즈의 빠른 탐지에 어려움을 겪는다. 특히 사람들이 서로 가까이 붙어 있을 때 탐지가 실패하는 경향이 있다. 또한 이러한 하향식 접근 방식의 실행 시간은 인원수에 비례한다. 각 탐지에 대해 한 사람의 포즈 추정기가 실행되고 사람이 많을수록 계산 비용이 증가한다. 반대로 상향식 접근 방식은 빠른 탐지에 대한 안정성을 제공하고 이미지에 있는 사람들의 수와 실시간 복잡성을 분리할 수 있는 가능성이 있으므로 매력적이다. 그러나 상향식 접근법은 다른 신체 부위 및 다른 사람들의 전역 상황 신호를 직접 사용하지 않는다. 실제로, 이전의 상향식 기법들[18, 19]은 최종 구문 분석에 시간이 많이 소요되는 전역 추론이 필요하기 때문에 효율이 떨어진다. Insafutdinov et al.[19]은 ResNet[20] 및 이미지 종속 페어 와이즈 점수를 기반으로 하는 강력한 신체 부위 검출 알고리즘을 사용하

여 Pishchulin et al.[18]의 방법을 기반으로 실시간 성능을 대폭 개선했지만, 이 방법은 부위 당 제안 수에 제한을 두어 이미지 마다 몇 분이 소요된다. Cao et al.[5]은 하나의 이미지에 등장하는 여러 사람의 스켈레톤 구조를 실시간으로 빠르고 정확하게 찾을 수 있는 기법을 제시했다. 본 논문에서는 해당 기법을 이용하여 영상 속의 사용자 포즈를 빠르게 분석하여 묘기 동작의 생성에 활용하였다.

## 2. Physics-Based Animation

본 논문은 수동적인 다이나믹 시스템의 제어에 관한 선행 기술을 기반으로 한다. 기존 관련 연구들을 보면 강체 [21, 22], 연기 및 물[23, 24] 및 변형 가능한 연체[25, 26]를 제어하기 위해 다양한 기술을 적용했다. 제안한 방법은 강체 시뮬레이션에서 강체의 이동을 제어하기 위한 대화식 인터페이스에서 영향을 받았다[22]. 해당 프레임워크는 수동적인 강체의 시뮬레이션을 제어하기 위한 소수의 물리적 파라미터를 사용했다. 또한 물체를 다루는 캐릭터를 생성하기 위해 강체의 이동 경로와 캐릭터의 포즈를 일일이 수작업으로 제어하는 방법[27]도 있었다. 물리 기반 캐릭터를 이용하여 물체를 다루는 방식의 논문이 있었다[29, 30]. Chemin et al.[29]은 2D 물리 캐릭터의 상반신을 이용하여 저글링 모션을 생성하였다. 이를 위해 강화학습을 이용하였으며, 다양한 저글링 동작을 생성하였으나 강화학습으로 생성된 모션이기 때문에 실제 인간의 모션과는 다소 거리가 있다. Hong et al.[30]은 MPC (Model Predictive Control) 기법을 이용하여 비교적 사실적인 축구공 드리블 모션을 생성하였다. 물리 기반 캐릭터로 축구공을 타격할 때 정확한 방향으로 보내기 위해 발목의 움직임을 세밀하게 조절할 수 있도록 하였다. 이와 달리 본 논문에서 제안한 방법은 모든 작업이 자동으로 이루어진다. 본 논문에서는 사용자의 영상에서 물체를 다루는 부위의 위치에 맞춰 강체의 이동 경로를 생성하기 위해 강체 시뮬레이션의 경로 제어를 사용하였다.

## III. The Proposed Scheme

본 연구에서는 사용자의 영상을 분석하여 가상의 물체를 다루는 묘기 동작을 생성하는 방법을 제안한다. Fig. 1은 본 연구에서 제안한 방법에 대한 개요이다. 사용된 영상은 사용자가 발 안쪽으로 제기차기 하듯이 물체를 다루는 동작이다. 좌측 그림은 사용자의 영상에서 스켈레톤 구조를 추출한 결

과이다. 굵은 회색 선이 사용자의 스켈레톤 구조이다. 가운데 그림은 스켈레톤에서 양 발의 위치를 추적한 결과이다. 빨간 선은 좌측 발의 이동 궤적이고, 파란 선은 우측 발의 이동 궤적이다. 우측 그림은 양 발의 이동 궤적을 분석하여 물체의 이동 궤적을 생성한 결과이다. 검은 포물선들이 생성된 물체의 이동 궤적들이다. 요약하자면 제안한 프레임워크에 사용자의 영상을 입력하면 최종적으로 물체의 이동 궤적이 출력되는 것이다. 본 연구에서는 이를 위해 크게 세 가지로 작업을 수행한다. 첫 번째는 딥러닝 기법을 이용하여 영상으로부터 사용자의 움직임을 추적한다. 두 번째는 사용자의 움직임을 분석하여 물체를 타격하는 지점의 정보를 찾는다. 마지막으로 찾아낸 타격 지점의 정보를 이용하여 자연스러운 물체의 이동 궤적을 생성한다. 이렇게 생성된 물체의 움직임과 사용자의 영상을 합성하면 가상의 물체를 자유롭게 다루는 묘기 영상이 완성된다.

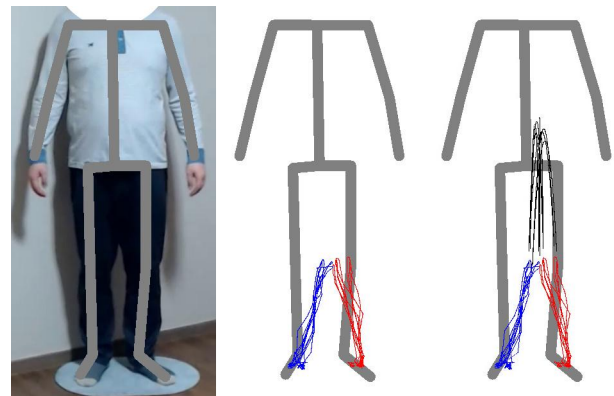


Fig. 1. Process Overview

### 1. Pose Tracking

본 연구에서는 사용자의 동작을 입력받아 사용하기 때문에 사용자 동작의 품질이 매우 중요하다. 사용자의 동작이 물체와 효과적으로 동기화되기 위해서는 사용자의 동작이 묘기 동작과 유사도가 높을수록 보다 자연스러운 결과 영상을 만들 수 있기 때문이다. 보다 사실적인 장면의 생성을 위해 사용자가 묘기 동작 영상을 보면서 연습한 후에 영상을 촬영하면, 생성되는 묘기 장면의 품질을 높일 수 있다. 본 연구에서는 발이나 무릎 등을 이용해 공 모양의 물체를 다루는 리프팅 장면들을 생성하였다. 이러한 장면의 생성을 위해 사용자는 축구 리프팅 영상을 보면서 발과 무릎의 움직임을 여러 차례 연습하였다. 본 논문에서는 Cao et al.[5]와 유사한 기법을 이용하여 사용자 영상으로부터 신체 부위의 위치를 추출하였다. 이는 딥러닝을 이용하여 영상 속의 많은 사람들에 대해 실시간으로 자세를 추적할 수 있는 기법이다.

본 논문에서는 사람의 수를 한 명으로 제한하고 영상의 화질을 낮추어 자세 추적의 성능을 향상시켰다.

### 2. Hit Spot Search

물체의 움직임을 생성하기 위해서는 물체를 타격하는 지점에 대한 정보가 필요하다. 타격 지점의 정보는 크게 타격 위치와 타격 시점으로 구성된다. 실제로 선수가 축구공 리프팅을 하는 영상에서 발의 위치 및 속도와 타격 시점의 관계를 분석한 결과, 발의 속도가 최대로 높아진 직후 속도가 최소가 되면서 발의 높이가 최대가 될 때, 축구공을 타격하는 것으로 분석되었다. 발의 속도가 최대가 되는 지점은 축구공을 타격하기 위해 발을 올리는 과정에서 나타나는 것이고, 발이 축구공과 접촉하는 순간에는 발이 거의 정지 상태가 된다. 본 연구에서는 Choi et al.[28]의 방법과 유사하게 신체 부위의 속도를 타격 지점을 추출하는 기준으로 사용하고자 했으나, 영상에서 추출된 신체 부위의 속도를 분석한 결과, 속도 정보를 사용하기가 매우 곤란하다는 사실을 알게 되었다. 영상으로부터 추출된 속도에는 상당히 많은 노이즈가 포함되어 있어서 속도가 불규칙하게 변화하고 있었다. 이는 영상에서 자세를 추적할 때 정확한 위치가 아닌 근사적인 위치를 출력하기 때문이다. 이를 해결하기 위해 노이즈 필터링을 포함하여 여러 방법을 시도했지만 크게 개선되지는 않았다.

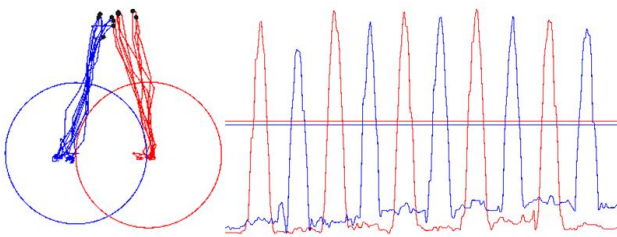


Fig. 2. Search Hit Spots

이에 본 연구에서는 영상에서 추출된 스켈레톤의 위치 정보만을 가지고 타격 지점의 정보를 생성하였다. Fig. 2는 사용자의 스켈레톤으로부터 양 발의 움직임을 분석하여 타격 지점의 위치와 시간을 추출한 결과이다. 빨간 색은 좌측 발, 파란 색은 우측 발에 대한 정보를 의미한다. Fig. 2의 좌측 그림은 추출된 타격 지점의 위치를 보여주고 있다. 검은 점들이 타격 지점의 위치이다. 불규칙한 선들은 양 발의 이동 궤적이다. 원은 타격 제한 거리를 의미한다. 즉, 해당 원의 안쪽에 있으면 물체를 타격할 수 없는 것으로 간주하여 타격 지점에서 제외된다. 이는 정지 상태의 지점들을 제거하기 위한 것이다. 원의 중심은 발의 거

리를 측정하는 기준 위치로 영상의 초반 0.5초 정도의 구간에서 측정된 발의 평균 위치이다. 원의 반지름은 기준점에서 가장 멀리 떨어져 있는 발의 위치까지의 거리의 절반으로 지정하였다. Fig. 2의 우측 그림은 기준점에서 발까지의 거리를 시간별로 나타낸 그래프이고, 수평선은 타격 제한 거리를 의미한다. 그래프의 가로축은 시간, 세로축은 기준점으로부터의 거리이다. Fig. 2에서 검은 점으로 표시된 타격 지점들은 타격 제한 거리 보다 멀리 있는 지역 최대값들이 되고, 이는 거리 우측 그래프에서 타격 제한 거리선 위쪽에서 뾰족한 지점들에 정확히 대응한다.

### 3. Object Synchronization

본 연구에서는 자연스러운 물체의 움직임을 생성하기 위해 Popovic et al.[22]의 방법과 유사하게 물체의 위치와 방향, 물체의 위치와 속도 등의 정보를 계산하여 사용하였다. 두 타격 지점 사이에서 자유비행 하는 물체의 이동 궤적은 물리 법칙을 적용하여 경사하강법으로 최적화 시켜서 생성하였다. 물체의 전체 이동 궤적은 생성된 모든 이동 궤적을 하나로 연결해 주면 된다. 사용자의 동작과 동기화되는 물체의 움직임을 생성하기 위해서는 타격 지점의 위치를 미리 알고 있어야 한다. 본 연구에서는 앞에서 설명한 방법으로 타격 지점 정보를 미리 생성하여 물체의 이동 궤적을 생성하였다. 서로 이웃하는 두 타격 지점을 시작과 도착 위치로 하고, 두 타격 지점 사이의 시간을 계산하면, 두 타격 지점 사이에서 물체의 움직임을 생성할 수 있다. 이런 방식으로 모든 타격 지점들 사이에서 물체의 이동 궤적을 생성하여 하나로 연결시켜주면 물체의 전체 이동 궤적을 생성할 수 있다.

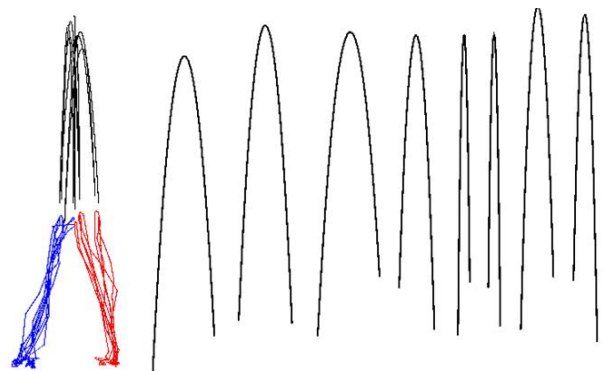


Fig. 3. Object Trajectories

Fig. 3은 타격 지점들의 정보로부터 생성된 물체의 이동 궤적을 보여준다. 좌측의 이미지는 타격 지점으로부터 생성된 이동 궤적들의 실제 위치이고, 우측의 포물선들은 각각의

궤적들을 하나씩 순서대로 나열해 놓은 것이다. 각각의 이동 궤적은 시작 위치, 도착 위치, 이동 시간 등이 상이하다. 그러므로 서로 다른 위치와 모양을 갖는 궤적들이 생성되었다.

### IV. Experiments

본 연구의 실험 결과를 생성하기 위해 Intel Core i7-7700K CPU, 64GB RAM, NVIDIA GeForce GTX 1080Ti GPU가 탑재된 컴퓨터와 보급형 웹캠을 이용하여 제안한 프레임워크를 구현하였다. 실험을 위해 사용된 영상은 모두 720 × 720 픽셀의 해상도이며, 초당 30프레임의 속도로 촬영되었다. 모든 사용자는 좌우측 발을 교대로 사용하면서 총 10회의 타격 모션을 취하였다. 사용자 스켈레톤으로부터 10개의 타격 지점이 추출되었고, 물체에 대한 9개의 이동 궤적이 생성되었다.

Fig. 4, 5, 6은 본 논문에서 제안한 프레임워크를 이용하여 사용자 영상을 분석한 결과와 그로부터 생성된 축구공 리프팅 영상의 일부 구간에 대한 연속 이미지들이다. 연속된 이미지들 사이의 시간 간격은 0.1초이다. 각각의 그림에서 좌측 상단의 이미지는 물체를 다루는 부위의 이동 경로이다. 회색 스켈레톤은 영상으로부터 추출된 사용자의 포즈이다. 여러 검은 포물선들은 물체의 이동 경로이고, 파란 곡선은 우측 부위의 이동 경로, 빨간 곡선은 좌측 부위의 이동 경로이다. 우측 상단의 이미지는 기준 위치로부터 물체를 다루는 부위까지의 거리에 대한 그래프이다. 물체를 다루는데 사용된 부위의 개수만큼 기준점이 존재하게 된다. 가령 양발을 이용했다면 기준점도 두 개가 된다. 수평 선분은 물체를 타격하는 프레임들 찾기 위한 기준 거리이다. 빨간 색은 좌측 부위, 파란색은 우측 부위의 거리 그래프이다. 하단의 이미지는 생성된 결과 영상의 연속된 이미지들이다. 검은 포물선은 해당 타격 구간에서 계산된 물체의 이동 궤적이다.

Fig. 4는 발등을 이용하여 축구공 리프팅 장면을 생성한 결과이다. 입력 영상의 전체 길이는 9.4초이고, 총 283프레임에 대한 포즈를 분석하였다. 영상 생성을 위해 소요된 시간은 7초였다. Fig. 5는 허벅지 부분을 이용하여 축구공 리프팅을 장면을 생성한 결과이다. 입력 영상의 전체 길이는 9.1초이고, 총 273프레임에 대한 포즈를 분석하였다. 영상 생성을 위해 소요된 시간은 6초였다. Fig. 6은 발 안쪽을 이용하여 제기 차듯이 축구공 리프팅 장면을 생성한 결과이다. 입력 영상의 전체 길이는 8.3초이고, 총 249프레임에 대한 포즈를 분석하였다. 영상 생성을 위해 소요된

계산 시간은 5초였다. 참고로 Fig. 4, 5는 성인 남성의 영상을 이용한 결과이고, Fig. 6은 초등학교 어린이의 영상을 이용한 결과이다. 이를 통해 제안된 프레임 워크는 사용자의 신체 사이즈에 상관없이 잘 동작하는 것을 확인할 수 있었다. 특히, 어린 사용자의 미숙한 동작으로 그래프가 다소 불규칙하게 나타났으나, 좌우측 부위를 따로 계산하기 때문에 크게 영향 받지 않고 제대로 된 결과 영상을 생성할 수 있었다.

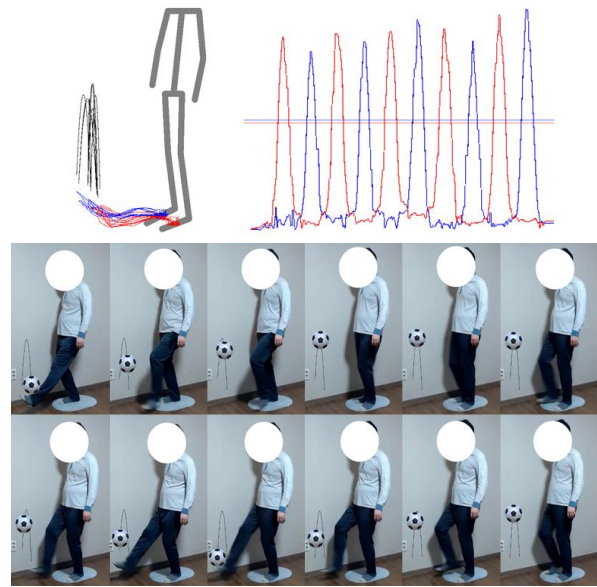


Fig. 4. Instep Lifting

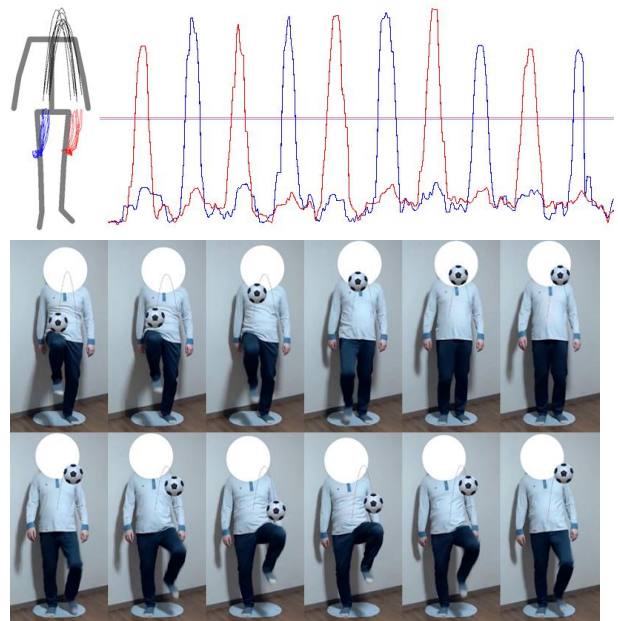


Fig. 5. Thigh Lifting



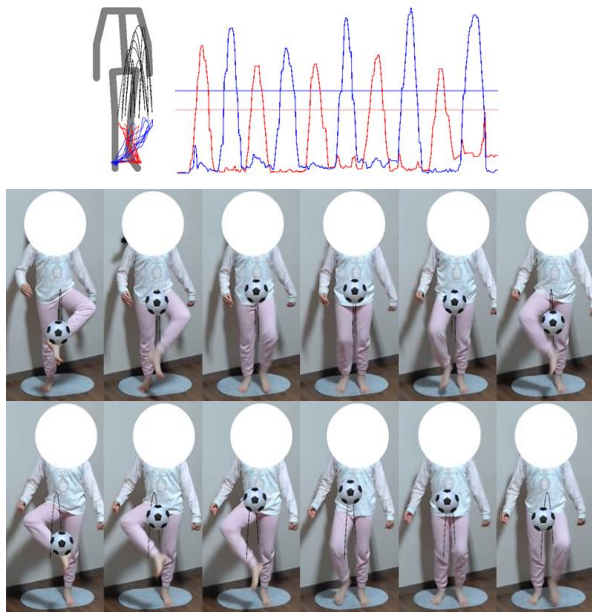


Fig. 6. Inside Lifting

## V. Conclusions

본 논문에서는 사용자의 영상을 분석하여 포즈 추적 및 물체의 움직임을 생성하여 다양한 동작으로 가상의 물체를 다루는 묘기 장면을 생성하는 프레임워크를 제안하였다. 이것을 사용자가 수작업으로 만들려고 하면 상당한 시간과 노력이 소요될 것이다. 그러나 본 논문의 방법을 이용하면 누구나 수 초 이내에 묘기 동작 영상을 제작할 수 있다. 실험 결과에서는 단순하고 반복적인 동작만을 촬영하여 사용하였는데 이는 사용자가 쉽게 익힐 수 있는 동작을 사용하려다 보니 그렇게 되었다. 제안한 프레임워크는 보다 복잡하고 불규칙한 동작의 사용도 가능하다. 예를 들면, 발과 무릎을 번갈아서 사용할 수도 있고, 발뒤꿈치로 리프팅하는 것도 가능하다. 사용자가 보다 익숙해지면, 좌측발로 볼을 다루다가 우측발로 볼을 다루다가 양발로 볼을 다루는 식으로 불규칙한 동작도 얼마든지 가능하다.

제안된 방법은 기존 영상에 가상의 오브젝트를 추가하는 방식이기 때문에 다소 어색해 부분이 존재한다. 빛의 방향에 따른 색감 차이, 물체의 그림자 부재, 물체가 항상 영상 앞에 존재 하는 등의 문제가 있다. 또한 2D 좌표에서 사용자의 포즈를 추적하다보니 오차가 발생하여 발과 공이 닿는 부분이 부자연스러워 보이는 문제도 존재한다. 본 논문의 방법은 물체를 타격하면서 다루는 방식에는 유용하지만, 물체와 접촉한 상태를 유지하는 것은 불가능하다. 즉, 물체를 발등이나 허벅지 부분에 올려놓는 상태를 유지

하는 식의 동작은 사용할 수 없다는 것이다. 또한 같은 다리 발과 허벅지를 번갈아가면서 사용하는 장면은 자동으로 생성할 수 없다. 이를 위해서는 수작업을 통해 발을 사용했는지 허벅지를 사용했는지 구분해 주어야 한다.

본 논문의 방법은 인간이 물체를 다루는 동작에 대한 흥미로운 문제를 풀어나가는데 도움을 줄 것이다. 주변에서 흔히 볼 수 있는 인간과 물체 사이의 상호작용에 대한 규칙성을 찾아내고, 이를 이용하여 다양한 응용 장면을 생성할 수 있을 것이다. 인간과 주변 환경과의 상호작용에 대한 것은 매우 중요한 문제이다. 지금까지는 고정된 주변 환경에 맞춰 캐릭터의 동작을 생성해 왔지만, 본 논문에서는 이와 반대로 인간의 동작에 맞춰 상호작용하는 물체의 움직임을 생성하였다. 이는 인간과 물체의 상호작용에 대한 하나의 대안을 제시한 것이다. 이는 다양한 엔터테인먼트 산업에 적용될 수 있다. 축구공 리프팅과 같은 기술을 일반인이 연마하기 위해서는 많은 시간의 노력이 필요하다. 그러나 본 논문의 방법을 이용한다면 누구나 수 초 이내에 증강현실 환경에서 축구공 리프팅과 같은 어려운 동작을 할 수 있게 되는 것이다. 이것이 핸드폰을 통한 증강현실 앱으로 개발된다면, 이는 사용자에게 기존에는 느낄 수 없었던 새로운 경험을 선사할 것이다. 앞으로 우리는 이러한 문제점들을 해결하고 제안한 방법을 확장하여 다양한 형태로 발전시킬 것이다. 우선 웹캠을 이용하여 사용자의 동작을 실시간으로 입력받아 묘기 영상을 생성할 수 있도록 할 것이고, 여러 사용자가 협력적으로 묘기 동작을 할 수 있도록 만들 것이다. 여러 장애물 오브젝트들을 이용하여 사용자가 가상의 물체를 더욱 절묘하게 다룰 수 있는 수단을 제공할 것이며, 물체를 다루는 부위를 손이나 머리 등의 다양한 신체 부위로 확대하여 보다 역동적인 묘기 장면을 만들 수 있도록 할 예정이다.

## ACKNOWLEDGEMENT

This work was supported by Research Grant from Seoul Women's University(2020).

## REFERENCES

- [1] D. Holden, T. Komura, and J. Saito, "Phase-functioned neural networks for character control," *ACM Transactions on Graphics*, vol. 36, pp. 1-13, 07 2017. doi: 10.1145/3072959.3073663.
- [2] X. Peng, G. Berseth, K. Yin, and M. Panne, "Deeploco: dynamic

- locomotion skills using hierarchical deep reinforcement learning,” *ACM Transactions on Graphics*, vol. 36, pp. 1–13, 07 2017. doi: 10.1145/3072959.3073602.
- [3] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei Rezvani Nezhad, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, “Vnect: Real-time 3d human pose estimation with a single rgb camera,” *ACM Transactions on Graphics*, vol. 36, 05 2017. doi:10.1145/3072959.3073596
- [4] D. Mehta, O. Sotnychenko, F. Mueller, W. Xu, M. Elgharib, P. Fua, H. Seidel, H. Rhodin, G. Pons-Moll, and C. Theobalt, “Xnect: Real-time multiperson 3d human pose estimation with a single RGB camera,” *CoRR*, vol. abs/1907.00837, 2019. arXiv:1907.00837.
- [5] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, “Openpose: Realtime multi-person 2d pose estimation using part affinity fields,” *CoRR*, vol. abs/1812.08008, 2018. arXiv:1611.08050.
- [6] M. Andriluka, S. Roth, and B. Schiele, “Pictorial structures revisited: People detection and articulated pose estimation,” pp. 1014 – 1021, 07 2009. doi: 10.1109/CVPR.2009.5206754.
- [7] S. Johnson and M. Everingham, “Clustered pose and nonlinear appearance models for human pose estimation,” pp. 1–11, 01 2010. doi:10.5244/C.24.12.
- [8] A. Bulat and G. Tzimiropoulos, “Human pose estimation via convolutional part heatmap regression,” vol. 9911, 10 2016. doi:10.1007/978-3-319-46478-7\_44.
- [9] V. Ramakrishna, D. Munoz, M. Hebert, J. Bagnell, and Y. Sheikh, “Pose machines: Articulated pose estimation via inference machines,” *Conference Paper, Proceedings of European Conference on Computer Vision*. pp. 33–47, 09 2014. doi:10.1007/978-3-319-10605-2\_3.
- [10] M. Sun and S. Savarese, “Articulated part-based model for joint object detection and pose estimation,” pp. 723–730, 11 2011. DOI:10.1109/ICCV.2011.6126309.
- [11] A. Jain, “Articulated people detection and pose estimation: Reshaping the future,” pp. 3178–3185, 06 2012. DOI: 10.1109/CVPR.2012.6248052.
- [12] U. Iqbal and J. Gall, “Multi-person pose estimation with local joint-toperson associations,” vol. 9914, 10 2016. doi: 10.1007/978-3-319-48881-344.
- [13] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, “Towards accurate multi-person pose estimation in the wild,” pp. 3711–3719, 07 2017. doi: 10.1109/CVPR.2017.395.
- [14] A. Toshev and C. Szegedy, “Deeppose: Human pose estimation via deep neural networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 12 2013. DOI: 10.1109/CVPR.2014.214.
- [15] W. Ouyang, X. Chu, and X. Wang, “Multi-source deep learning for human pose estimation,” pp. 2337–2344, 06 2014. doi: 10.1109/CVPR.2014.299.
- [16] J. Tompson, A. Jain, Y. Lecun, and C. Bregler, “Joint training of a convolutional network and a graphical model for human pose estimation,” 06 2014. arXiv:1406.2984.
- [17] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” vol. 9912, pp. 483–499, 10 2016. doi: 10.1007/978-3-319-46484-829.
- [18] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, “Deepcut: Joint subset partition and labeling for multi person pose estimation,” pp. 4929–4937, 06 2016. arXiv:1511.06645.
- [19] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, “Deepcut: A deeper, stronger, and faster multi-person pose estimation model,” 05 2016. arXiv: 1605.03170.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” vol. 7, 12 2015. DOI: 10.1109/CVPR.2016.90.
- [21] C. Twigg and D. James, “L.: Backward steps in rigid body simulation,” *ACM Trans. Graph.*, vol. 27, 08 2008. doi: 10.1145/1399504.1360624.
- [22] J. Popovic, S. Seitz, M. Erdmann, Z. Popovic, and A. Witkin, “Interactive manipulation of rigid body simulations,” *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 10 2001. doi:10.1145/344779.344880.
- [23] R. Fattal and D. Lischinski, “Target-driven smoke animation,” *ACM Transaction on Graphics*, vol. 23, 06 2004. doi:10.1145/1015706.1015743.
- [24] A. Treuille, A. McNamara, Z. Popovic, and J. Stam, “Keyframe control of smoke simulations,” *ACM Trans. Graph.*, vol. 22, pp. 716–723, 07 2003. doi: 10.1145/882262.882337.
- [25] C. Wojtan, P. Mucha, and G. Turk, “Keyframe control of complex particle systems using the adjoint method,” pp. 15–23, 01 2006. DOI:10.1145/1218064.1218067.
- [26] J. Barb<sup>ı</sup> and J. Popov<sup>ı</sup>, “Real-time control of physically based simulations using gentle forces,” *ACM Transactions on Graphics*, vol. 27, 12 2008. doi: 10.1145/1409060.1409116.
- [27] S. Jain and C. Liu, “Interactive synthesis of human-object interaction,” pp. 47–53, 01 2009. doi: 10.1145/1599470.1599476.
- [28] J. I. Choi, S. J. Kang, C. H. Kim, and J. Lee, “Virtual ball player,” *The Visual Computer*, vol. 31, 05 2015. doi: 10.1007/s00371-015-1116-9.
- [29] J. Chemin and J. Lee. “A physics-based juggling simulation using reinforcement learning.” In *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games (MIG ’18)*, Article 3, 1–7. DOI: 10.1145/3274247.3274516.
- [30] S. Hong, D. Han, K. Cho, J. S. Shin, and J. Noh. 2019. Physics-based full-body soccer motion control for dribbling and shooting. *ACM Trans. Graph.* vol 38, no 4, Article 74 (July 2019), pp. 1-12. DOI:10.1145/3306346.3322963.

## Authors



Jong-In Choi received the Ph.D. in 2016 from the Department of Computer Science and Engineering from Korea University. He joined Nexon Korea as a lead client programmer. And he also has worked at

NCSOFT Korea R&D Center and AI Center. Now he is a professor at the department of digital media design and application in Seoul Women's University.



Jong-Hyun Kim received the B.A. degree in the department of digital contents at Sejong University in 2008. He received M.S. and Ph.D. degrees in the department of computer science and engineering at Korea University,

in 2010 and 2016. Prof. Kim is an assistant professor in the department of software application in Kangnam University. His current research interests include fluid animation and virtual reality.