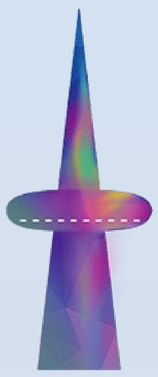

PG23

Daejeon, Korea

Reconstructing Baseball Pitching Motions from Video

Jiwon Kim, Dongkwon Kim, and Ri Yu

Ajou University, South Korea

PROBLEM

Baseball is one of the most loved sports in the world. In baseball game, the pitcher's control ability is a key factor for determining the outcome of the game. There are a lot of video data shooting baseball games, and learning baseball pitching motions from video can be possible thanks to the pose estimation techniques. However, reconstructing pitching motions using pose estimators is challenging (See Figure 1). When we watch a baseball game, motion blur occurs inevitably because the pitcher throws a ball into the strike zone as fast as possible.

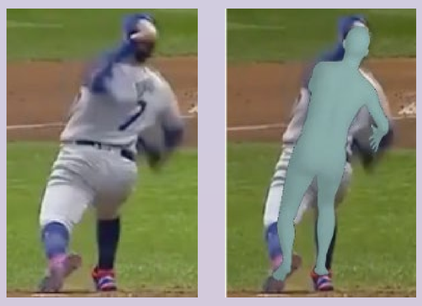

Figure 1

RELATED WORK

Researches reconstructing 3D human motions from 2D video have been developed remarkably [3, 4, 5, 6]. Peng et al. [3] used deep reinforcement learning (DRL) to mimic the dynamic movements such as backflips and cartwheels. Yu et al. [4] simulated figure skating skills. They used key poses and trajectory optimization to recon struct skating motions. Yu et al. [5] used interaction between the character and the environment to reproduce motions with the global trajectory from dynamic-view video. Zhang et al. [6] used two-level imitation systems which used RL and VAE to learned tennis skills.

All these researches used 2D videos as inputs in common. They used various pose estimators [1, 2] to extract human poses from videos. Therefore, the reconstructed motion quality is highly dependent of the accuracy of pose estimation techniques. Although many issues exist for more precise pose estimation from 2D videos, motion blur is one critical issue to reconstruct highly dynamic and fast motions, especially in baseball motions such as pitching or swing.

OVERVIEW

Baseball is a sport loved by people all over the world. There are many video data filming the baseball games, and baseball coaches and players analyze opponents' performance or review their playing watching the videos. Although, there have been many efforts to reconstruct 3D human motions from 2D videos in computer vision and computer graphics fields [3, 4, 5, 6], many issues such as occlusion, motion blur, and pose ambiguities make it difficult to reproduce accurate motions.

We propose a framework to reconstruct baseball pitching motions from 2D video even the pose estimation results are unsatisfactory. Our framework takes advantage of using the reinforcement learning and the physics simulation to reproduce realistic motions according to the given goal. The pitcher must throw a ball into the strike zone. We set a target point signifying the center of the strike zone, and design two reward terms which encourage the character to throw the ball toward the target point as fast as possible.

METHODOLOGY

We adopted the framework proposed by Yu et al. [5] and revised it. When an input video is given, it utilizes OpenPose [1] to obtain 2D joint positions which are used to extract contact information. 3D estimated poses are obtained using VIBE [2]. Finally, using deep reinforcement learning, the motion following the hints earned from the video is reproduced.

- **Environment Setup**

A baseball pitcher starts pitching standing on the mound, which has the highest location in the field. We approximate the pitcher's mound as a box tilted about 0.2 radian to form a down-slope toward the home plate. We set a target point reaching the distance of 18.44 meters from the mound to the home plate. We put a ball as a weld joint of the pitching hand and let the character release the ball when the time comes.

- **Rewards**

The framework proposed by Yu et al. [5] uses five reward terms r_{base} to make the virtual character imitate the high-level hints, such as poses and contact information extracted from the input video.

$$r_{base} = w_q r_q + w_v r_v + w_{contact} r_{contact} + w_{rcg} r_{rcg}$$

The pose and velocity rewards r_q and r_v encourage to mimic the estimated poses. However, in case of the pitching example, estimated arm poses are not reliable because of motion blur. Hence, we reduced the tracking weight of the arm joints 0.1 times.

To give the ability to control the ball, we added two reward terms , $r_{ballPos}$ and $r_{ballDir}$. The first term $r_{ballPos}$ is for reducing the distance between ball and the target point and the second term $r_{ballDir}$ is for matching the throwing direction of the ball toward the target point.

Ball Position. When a ball is released, its trajectory is predictable because it moves along the curve. Therefore, we can calculate the expected position of the ball at every frame after the release point . The ball position reward encourages the character to throw the ball to make a ball reaching the target position as closely as possible.

$$r_{ballPos} = \exp(-\alpha_{ballPos}\|d\|^2),$$

where d is the distance from the ball and the target point.

Ball Direction. The pitcher should throw a ball toward the strike zone. We devise the ball direction reward term which minimizes the difference between two vectors $\vec{p}$ and $\vec{q}$, where $\vec{p}$ is the CoM speed of the ball and $\vec{q}$ is a vector from the ball and the target point.

$$r_{ballDir} = \exp(-\alpha_{ballDir}\|1 - \vec{p} \cdot \overrightarrow{q}\|^2)$$

For learning, we use the reward $r = r_{base} + w_{ballPos} r_{ballPos} + w_{ballDir} r_{ballDir}$. Note that $r_{ballPos}$ and $r_{ballDir}$ are zero before releasing the ball.

RESULTS

We used the video clip of the major league game as an input video. Figure 2 shows the reconstruction results of the previous framework [5] and ours. Although the input is overhand throwing pitcher's video, the arm motion reconstructed using [5] is not the same and the character cannot throw the ball as in the video (Figure 2 left). It's because the controller is trained to imitate the reference motion based on incorrectly estimated poses due to the motion blur. On the other hand, our framework successfully reproduced overhand pitching motion even though we used the same reference motion (Figure 2 right).

Our framework shows the potential to reconstruct human motions based on the estimated poses from off-the-shelf pose estimators which suffer from various reasons, such as motion blur and pose ambiguities. Currently, our character doesn't have sufficient control over the speed or position of the ball. In the future work, we will adopt curriculum learning to enhance the control ability of the virtual pitcher. Furthermore, we plan to collect pitching videos and reconstruct pitching motions to built the data set. Based on that, we will train a model to learn a variety of pitching style.

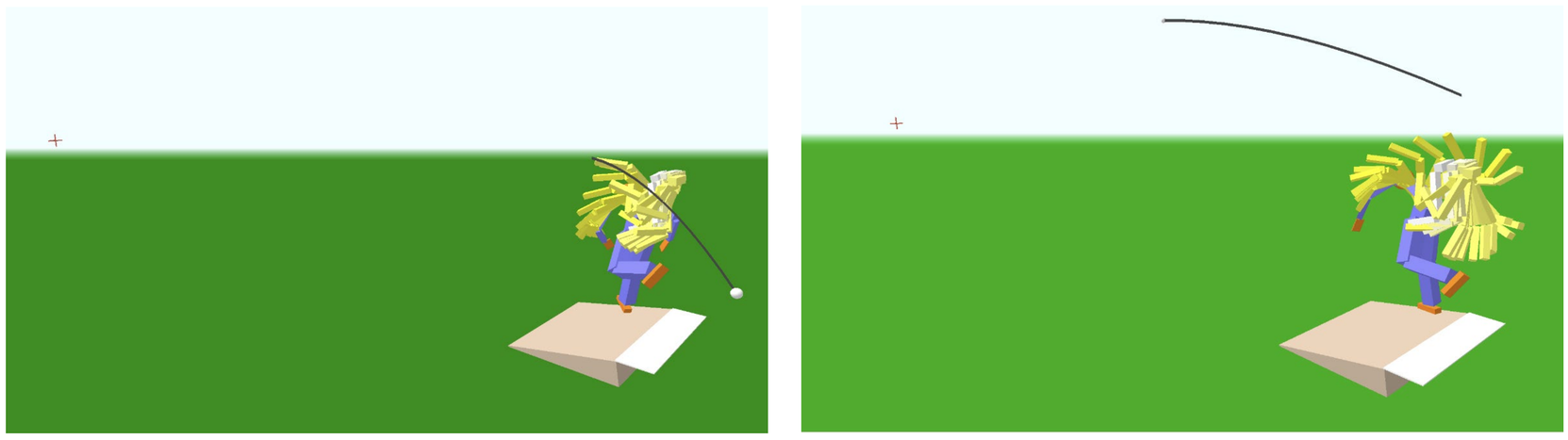

Figure 2 : Reconstructed results of a pitching motion using the previous framework [5] (left) and using our framework (right). You can see the trajectory of the left arm for each frame. In the left subfigure, the motion of the pitching arm is not properly reconstructed because it mimic the reference poses estimated from the video suffering from motion blur. On the other hand, our framework can generate plausible pitching motion as shown in the right figure. We used rewards terms to guide the character to pitch the ball toward the right direction and physics simulation to reconstruct successful pitching motion.

아주대학교 AJOU UNIVERSITY 1973

REFERENCES

[1] CAO Z., HIDALGO MARTINEZ G., SIMON T., WEI S., SHEIKH Y. A.: Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019)

[2] KOCABAS M., ATHANASIOU N., BLACK M. J.: Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 5253–5263.

[3] PENG X. B., KANAZAWA A., MALIK J., ABBEEL P., LEVINE S.: Sfv: Reinforcement learning of physical skills from videos. *ACM TOG* 37, 6 (2018)

[4] YU R., PARK H., LEE J.: Figure skating simulation from video. *Computer Graphics Forum* 38, 7 (2019)

[5] YU R., PARK H., LEE J.: Human dynamics from monocular video with dynamic camera movements. *ACM TOG* 40, 6 (2021)

[6] ZHANG H., YUAN Y., MAKOVIYCHUK V., GUO Y., FIDLERS., PENG X. B., FATAHALIAN K.: Learning physically simulated tennis skills from broadcast videos. *ACM TOG* 42, 4 (2023), 1–14.