3D Parkour Game Manipulated with Real-Time Pose Estimation Technique

111061516 Jin-Yu Young, 111061564 Hao-Jiun Tu, 111061585 Jing-Chun Wang
111061519 Mei-Yen Tsai, 107081028 An-Yan Chang

1. Introduction

1.1. Human Pose Estimation in Video

Human pose estimation is a task in computer vision that focuses on identifying the position of a human body in a specific scene. The essence of the technology lies in detecting points of interest on the limbs, joints, and even faces of a human. These key points are used to produce a 2D or 3D representation of a human body model. The model must identify the fine-grained joint coordinates for the human body. For a multi-frame human pose estimation, it takes a large amount of resources, such as memory bandwidth and computing power, to have a great performance in complicated scenes. It is a tradeoff between the computing resource and the accuracy. In this project, we want to improve this technology, i.e., human pose estimation and video person pose tracking in real-time constraints.

1.2. Application

In the project, we want to combine the aforementioned technology to some video game like Xbox 360 - Kinect Sports in Fig. 1. There is a camera sensor that will capture the human body of players, and the action and movement of characters in the game will be reflected by players. Parkour game seems to be an appropriate application for this technology. It requires real-time feedback and is suitable for body manipulation. Furthermore, parkouring as a personal experience would be such entertainment. Therefore, we want to design a parkour game with the character controlled by the players with their poses. The conceptual graph of the Parkour game we expected is shown in Fig. 2.

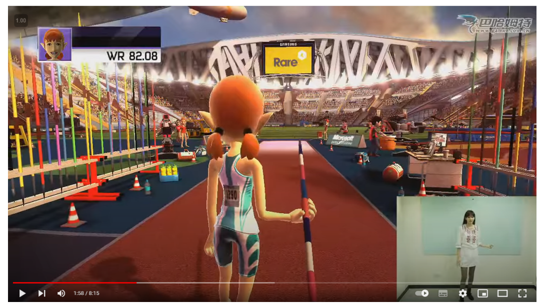

Figure 1. Xbox 360 - Kinect Sports.

Figure 2. Conceptual graph of Parkour game.

2. Technical Part

2.1. 2D Human Pose Estimation

Focusing on this technology 2D Human Pose Estimation, OpenPose model [2] shows great possibilities for achieving real-time application. OpenPose uses a non-parametric representation, which refers to as Part Affinity Fields (PAFs), to learn to associate body parts with individuals in the image. This bottom-up system achieves high accuracy and real-time performance. However, in an immersive virtual game, it would be greatly influenced by 2D display settings in terms of visual experience, movement smoothness and fineness. Therefore, we attempt to exploit 3D human pose estimation.

2.2. 3D Human Pose Estimation

The state-of-the-art work, VNect model [3], will be our first choice to realize his technology, i.e., 3D human pose estimation in video. Fig. 3. provides an overview of VNect method to fulfill real-time 3D human pose estimation. It consists of two primary components. The first is a CNN model to regress 2D and 3D joint positions. The second component combines the regressed joint positions with a kinematic skeleton fitting method to produce a temporally stable, camera-relative, full 3D skeletal pose. Fig. 4. shows the model architecture of VNect, which basically adopts the backbone from ResNet50. It predicts the 2D heatmap H and 3D location-maps X, Y, Z to form the kinematic skele-

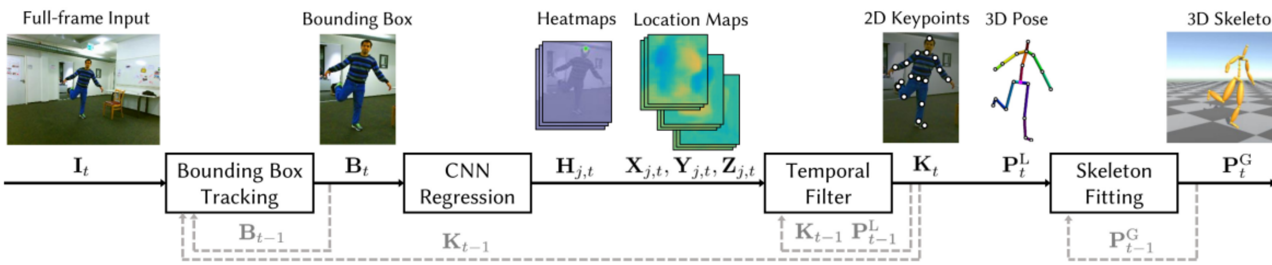

Figure 3. The overview of VNect method.

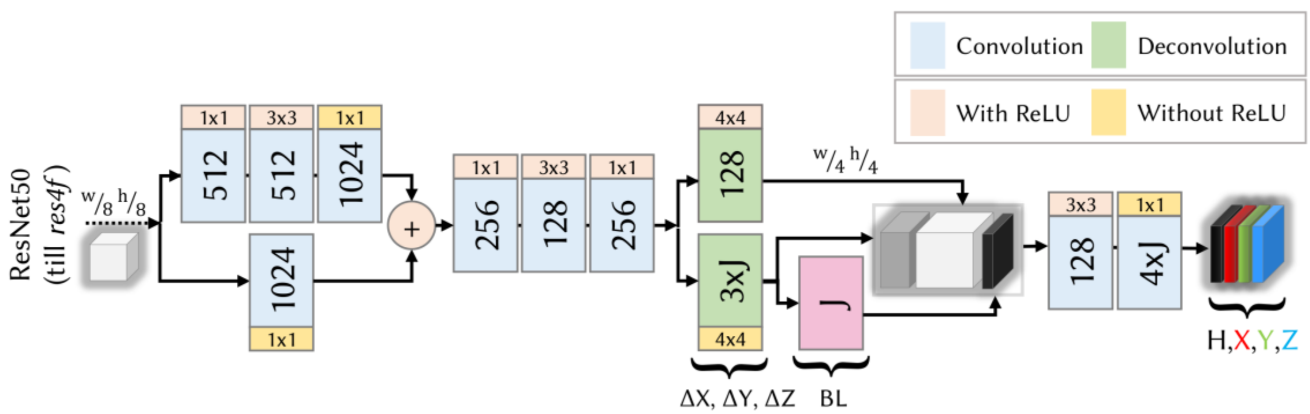

Figure 4. Network structure for VNect. The structure above is preceded by ResNet50 till level 4. The network predicts 2D location heatmaps H and root relative 3D joint locations X, Y, Z.

ton. Further, we will use the information as our Unity input to design the Parkour game.

2.3. Game Designing

Unity is a great tool for prototyping everything from games to interactive visualizations. We will use Unity to design our Parkour game, further plugging in our pretrained VNect model to integrate the human pose estimation technology.

2.4. Some Issues

If our model is too complex, we might encounter low-frame-rate issues, making the video laggy or not smooth. In this case, we need to survey some model compression techniques like pruning to decrease our model complexity without degrading too much model accuracy.

3. Milestones

Currently, the scene arrangement for our Parkour game had been completed with Unity. Additionally, we have plugged the predicting information from VNect into Unity, as shown in Fig. 5. We also have filmed a simple demo video [1] to display our current progress. At present, 3D human pose estimation utilized in our Parkour game can detect any big movement well; while it has some difficulties in handling the finest action. Besides, the range of movement in left-right directions is smaller than that in up-down directions. Thus, we need to adjust some parameters or exploit another way to balance this problem.

4. Schedule

The following lists our remaining milestones and the schedule for the project:

- Week 10: Midterm project report (due: 11/23)
- Week 11-12: Detect the finest movement successfully
- Week 13-14: Improve the range of movement between left-right and up-down directions
- Week 15-16: Final presentation
- Week 17: Final project report (due: 1/11).

5. References

Figure 5. Our Parkour game (current progress).

References

[1] Group08: Unity parkour game design. https://www.youtube.com/watch?v=PTCPzhxxQnU. 2

[2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1

[3] Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. Vnect: Real-time 3d human pose estimation with a single rgb camera. *Acm transactions on graphics (tog)*, 36(4):1–14, 2017. 1