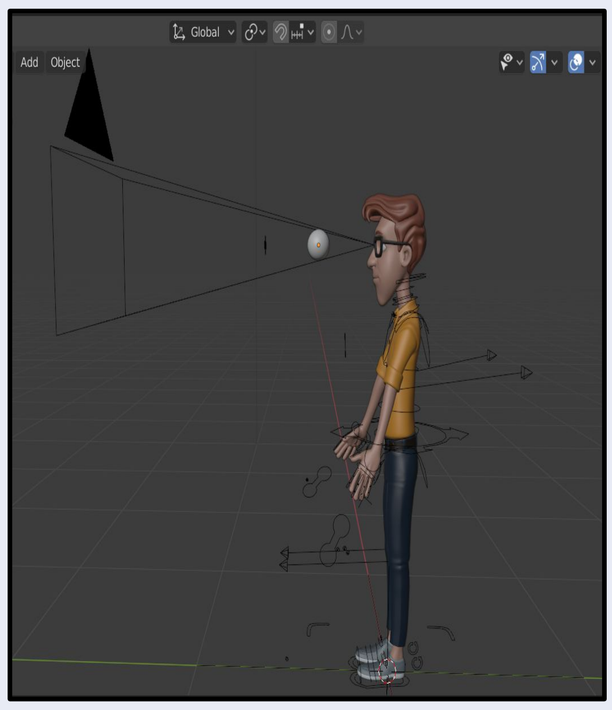
VR MOTION CAPTURE

IIT KANPUR

Abstract/Introduction

MeCap is a low cost VR extension to capture the body and hand movement of the wearer from the rear camera of a cellphone through reflection from metal sphere ,without requirement of any external infrastructure or other sensors.

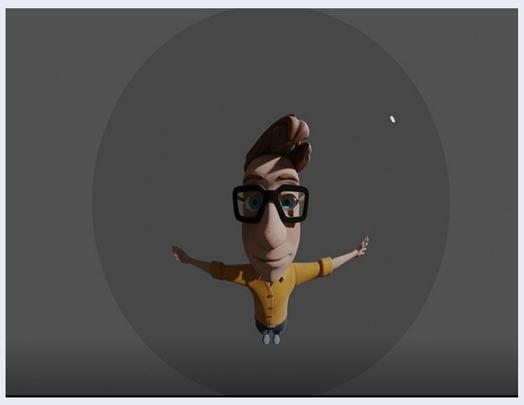


Possible application:

1) Can be used in Mobile Gaming as an extended controllers.
2) Can change the use of convention smartphone lens with alternatives like spherical or Fisheye lens

Theory

The body of the user is reflected on the spherical mirror. This reflection is captured by the camera present in the mobile phone. This captured reflection is useful in getting user body position.

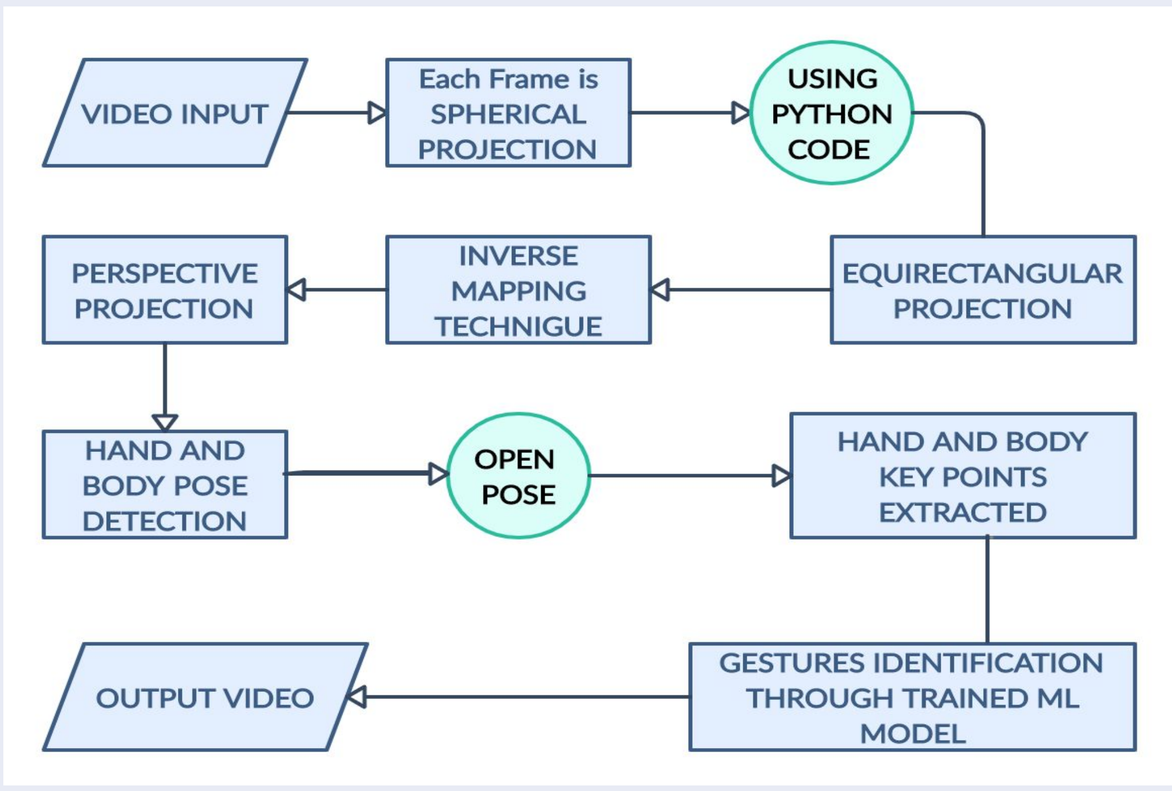


SPHERICAL PROJECTION

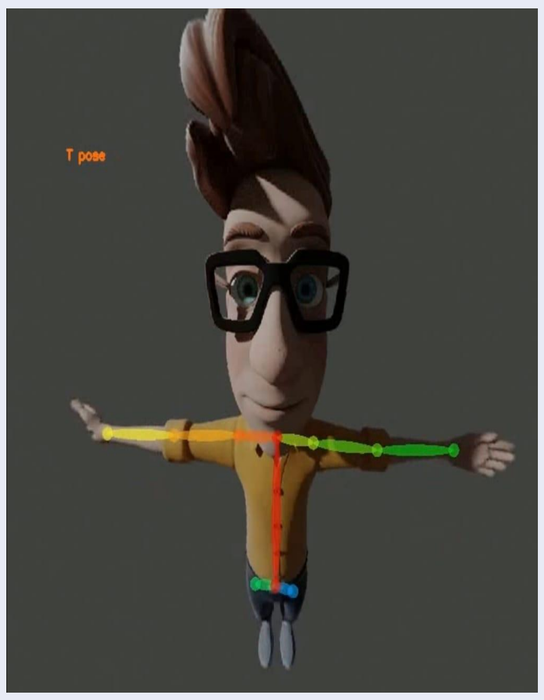
PERSPECTIVE PROJECTION

After converting the each frame into perspective projection , the further task was body and hand key point detection and their gesture identification.For the keypoints detection openpose was used and then for hand gesture identification, CNN which uses trained VGG 16 model on our own database was used.Pose classification was done through our own python code.

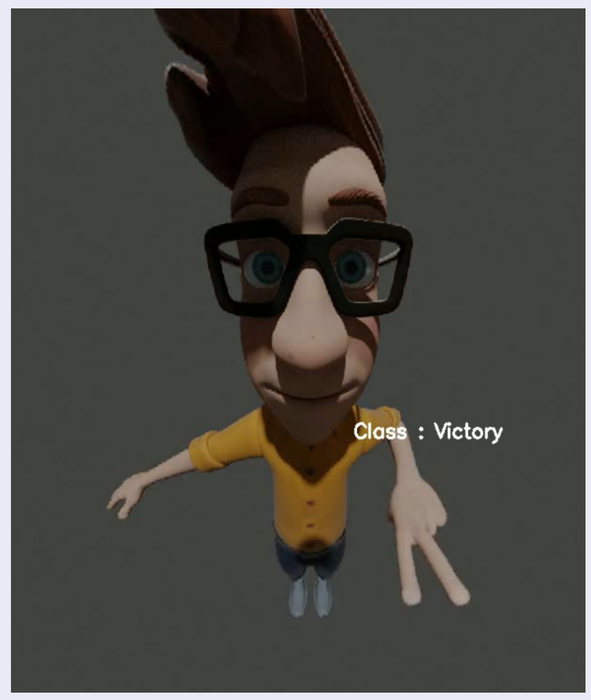
Workflow



Step 1: Conversion of spherical image into equirectangular image.
Step 2: Converting equirectangular image into perspective image.
Step 3: Using OpenPose to get keypoints of various joints.
Step 4: Using a dynamic window of keypoint to extract image of that particular gesture. For example: wrist keypoint for image of fist.
Step 5: Hand gesture classification through CNN which uses trained VGG 16 model.
Step 6: Body pose classification done through geometry of nodes. Nodes are returned by openpose and processed using our own algorithm.



BODY POSE PREDICTION



HAND GESTURE PREDICTION

Future Work

- **Implementation on actual hardware using smartphones and mirrors.**
- **Extract other visual appearances including beards, hairstyle, color and category for further avatar personalization.**
- **The wide angle view can used for keypoint tracking the other people in the view, without accessories**

References

- **Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multiperson 2d pose estimation using part affinity fields.**
- **Karen Simonyan, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).**
- **Paul Debevec. 2008. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography.**