HEAD POSE ESTIMATION USING MACHINE LEARNING

Mudam Keerthana[1], Pamballa Ajay[2], Swamy Goud[3],

[1] *Student, Department of Information Technology, B V Raju Institute of Technology, Medak, Telangana, India.*
[2] *Student, Department of Information Technology, B V Raju Institute of Technology, Medak, Telangana, India.*
[3] *Assistant Professor, Department of Information Technology, B V Raju Institute of Technology, Medak, Telangana, India.*

ABSTRACT

Head pose estimation is an computer vision task. In computer vision the pose of an object refers to its relative orientation and position with respect to a camera. You can change the pose by either moving the object with respect to the camera, or the camera with respect to the object. In many applications, we need to know how the head is tilted with respect to a camera. In a virtual reality application, for example, one can use the pose of the head to render the right view of the scene. In a driver assistance system, a camera looking at a driver's face in a vehicle can use head pose estimation to see if the driver is paying attention to the road. And of course, one can use head pose based gestures to control a hands-free application / game. In this Project, we want to know the object's pose from its translation and rotation. We will use Python as the programming language. Also, we use the media pipe library for detecting face key points and the OpenCV library for estimating the pose of a head.

Keywords: Machine Learning, Python, OpenCV

---

1. INTRODUCTION

Head pose estimation (HPE) is the task of estimating the orientation of heads from images or video and has seen considerable research. Applications of HPE are wide-ranging and include (but are not limited to) virtual & augmented-reality , driver assistance, markerless motion capture or as an integral component of gaze-estimation since gaze and head pose are tightly linked .

The importance of HPE is well described in. They describe wide-ranging social interactions such as mutual gaze, driver-pedestrian & driver-driver interaction. It is also important in providing visual cues for the targets of conversation, to indicate appropriate times for speaker/listener role switches as well as to indicate agreement. For systems to interact with people naturally, it is important to be sensitive to head pose. Most HPE methods target frontal to profile poses since applications are plentiful, the face is featurerich and training datasets are widely available. However, covering full range is useful in many areas including driver assistance , motioncapture and to generate attention maps for advertising and retail .

2. LITERATURE SURVEY

The estimate the head pose of another person is a common human ability that presents a unique challenge for computer vision systems. Compared to face detection and recognition, which have been the primary foci of facerelated vision research, identity-invariant head pose estimation has fewer rigorously evaluated systems or generic solutions.

In this project, the method of head pose regression based on machine learning is used. This method requires no pre-calibration and complex modeling procedures, which is robust and able to process a larger scene image. The general process of head pose estimation based on machine-learning regression is to extract some significant features from the captured image first, and then a rotation regression model is built based on the selected features and the corresponding rotation ground truth values. The head pose is at 3 degrees that contain the pitch, yaw, and roll angles of the head. Therefore three regression models are usually being modeled to estimate each angle rotation.

2.1 EXISTING SYSTEM

Existing regression-based methods can be classified into 2D and 3D categories. 2D means that the extracted features only contain 2D features that are extracted from the image while 3D estimation takes depth information into consideration and estimates the rotation with both 2D and 3D features. In a feature extraction method based on scale-invariant feature transform (SIFT) and histogram of oriented gradients (HOG) was used to construct the feature vector of the pose estimation. The integrated features along with depth image captured with Kinect were fed into a random forest algorithm and the final results achieved 8degree errors on average. Eye gaze and head pose estimation method for driver eye-off-road detection was proposed in 2D estimation methods can be less robust and have an identity issue that causes a bad generalization. Therefore depth information is augmented into the feature vectors to enhance the system performance and robustness property. A combination of RGB images and depth images for human head pose detection is proposed RGB and depth images were captured with a Kinect sensor. The head position was first located with a Viola–Jones face detector, which was applied twice to find a rough and precise face location, respectively. Then, the grayscale and depth face features were extracted with multiple feature extractors and concatenated together for pose regression. Finally, an SVR is applied to continuously estimate the head pose frame by frame.

2.2 DISADVANTAGES OF THE EXISTING SYSTEM

1). For training the model, the weights are calculated for every image and it requires lots of time.
2). It also requires lots of computational time.
3). The accuracy of the model is very less.

2.3 PROPOSED SYSTEM

3D estimation takes depth information into consideration and estimates the rotation with both 2D and 3D features. A feature extraction method based on scale-invariant feature transform (sift) and histogram of oriented gradients (hog) was used to construct the feature vector of the pose estimation.

2.3.1 ADVANTAGES OF PROPOSED SYSTEM

However, pose estimation has much potential, even beyond face recognition purposes. In [11], it was shown that pose estimation is of great importance for systems that require human interaction. For instance, in the automotive industry, pose estimation has been employed for driver safety in recent decades.

3 TESTING & VALIDATION

3.1 Testing

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. Finally, we perform several experiments where we deliberately ignore the detection of certain facial regions, to identify the facial regions that are more significant for accurate camera pose estimation than others. In particular, we selectively exclude the following facial regions: left and right ear, left and right eye, left and right side of the face, mouth and nose. The experiment contains recordings of all six participants from three different camera angles (90 , 45∘ and 0∘) relative to the test subject.

∘

3.2 Validation

Validation means observing the behavior of the system. The verification and validation mean, will ensure that the output of a phase is consistent with its input and that the output of the phase is consistent with the overall requirements of the system. This is done to ensure that it is consistent with the required output. If not, apply certain mechanisms for repairing and thereby achieved the requirements.

4. CONCLUSIONS

It addressed the problem of head pose estimation under target motion when the scene is captured by multiple, wide-angle cameras. Challenges in this set-up include low resolution of faces and change in facial appearance with motion due to varying camera perspective and scale. We described two different frameworks to address HPE under target motion, namely, transfer learning and multitask learning. We also presented exhaustive experimental results obtained using both approaches and demonstrating the effectiveness of the described solutions. We believe that our methods represent effective solutions for addressing different high-level tasks in the areas of human behavior analysis, social signal processing, and social robotics.

5. REFERENCES

https://www.sciencedirect.com/science/article/pii/S0262885603002610
https://www.sciencedirect.com/science/article/pii/S0262885603002610
https://www.sciencedirect.com/science/article/pii/S0925231217308706
https://www.sciencedirect.com/science/article/pii/S0031320300000911