Journal of Industrial Information Technology and Application Vol.2 No.1
ISSN(Online) : 2586-0852

Robust Head Pose Estimation under Facial Shape Deformation

Byunghun Hwang, Sang-Heon Lee, Hyunduk Kim, Yoon-Jib Kim, Youngsun Ahn
Daegu Gyeongbuk Institute of Science & Technology (DGIST), Daegu, Republic of Korea

Abstract. Robust and accurate head pose estimation considering varying facial expression remains challenging issue in the area of research on vision-based human-computer-interaction. In this paper, we discuss how to reduce the influence of face landmark changes due to facial shape deformation to achieve robust head pose estimation. We select only a few landmark points that have lower dependency on facial movement for robustness of the head pose estimation. The results show that the head pose estimation error can be reduced even when the facial expression changes.

Keywords; head pose estimation, shape deformation, constrained local model, shape model

1. Introduction

Accurate and robust head pose estimation is crucial for human-machine interaction and is based on human behavior analysis. This ability is likely to be most used in entertainment systems such as interactive gaming platform for AR/VR/MR contents and vehicle driver monitoring systems[1]. In fact, it would be most accurate and quickest to attach sensors involving gyroscope or accelerometers to determine the exact head orientation in various environments. A vision-based approach would be more appropriate because these applications ultimately should be performed in unrestricted environments and should be user-friendly.

There are various approaches to head pose estimation. The appearance template methods involves finding the most similar pose data by comparing discrete pose data with head pose in new images. The detector array method is to learn detectors to respond

* Corresponding author: {bhhwang, pobbylee, hyunduk00, rladbswlq20, youngsunan}@dgist.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2018. Journal of Industrial Information Technology and Application (JIITA)

to specific poses. The nonlinear regression method uses a nonlinear regression tool to develop a function to map the image or feature data to the head pose. Geometric methods use positions to determine a pose from the relative configuration of features such as eyes, mouth, and nose tip. The tracking method recovers the changing in a global pose of the new head pose from the movement observed between video frames. Hybrid methods can combine one or more of the above methods to produce better results when using one method. Recently, research has been conducted to improve pose estimation accuracy using Deep Learning approaches [2].

There has been much research on head pose estimation, some of which has already been commercialized. However, the various commercially available technologies do not provide similar results. This is because conventional head pose estimation methods often require accurate initialization or facial landmark tracking, and most of all; these methods should cope with the non-rigid shape deformation problem before doing head pose estimation. In this paper, we propose a geometric approach using 3D face landmarks that are relatively less dependent on facial movement.

2. Head Pose Estimation

We found that head pose estimation accuracy and robustness using facial landmarks is dependent on facial movement such as facial expression. Therefore, in this paper, we propose to reduce the effect on head pose estimation caused by the problem of facial landmark deformation by facial expression. The method used is to estimate the head pose by selecting landmarks that are less dependent on non-rigid facial motion. Fig .1 shows the head pose estimation procedure with a Constrained Local Model(CLM) based landmark detection algorithm[3,4]. First, 68 well-fitted 2D and 3D facial landmark points were detected in images of face, and the head pose was estimated using Perspective-n-Point(PnP) and landmark point selection.

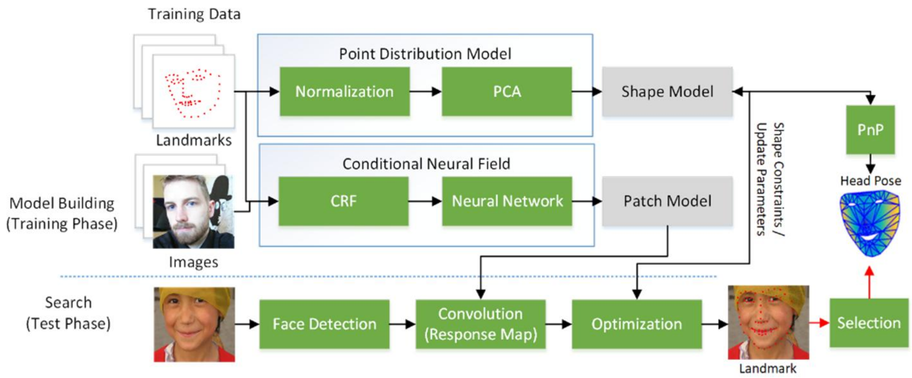

Fig. 1. Procedure of head pose estimation

Among the less-sensitive landmarks near the muscles related to facial expression, it would be best to use relatively stationary landmark points to accommodate non-rigid facial movements. After detecting the facial landmarks with CLM from a dataset of various synthetic facial expressions (Fig. 2), we selected the less sensitive landmark points based on the lowest standard deviation of each landmark. The landmarks around the eyes are relatively insensitive to changes in facial expression, and the left and right inner eyes have the lowest standard deviation. Fig. 2 shows the head pose estimation when using the inner eye and center position. The roll angle of the head pose is determined by the angle between the left and right inner eye, and the pitch angle can be determined by the angle between the center of the all landmarks and the center of the inner eye.

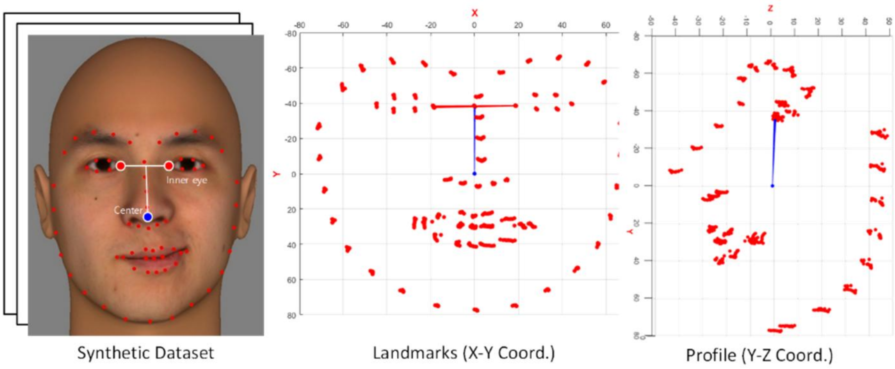

Fig. 2. Low variance landmark selection and head pose (Left/Right inner eye)

3. Results

We experimented to evaluate the method by using only two landmarks with low variance. This requires head pose datasets with various facial expression to verify effectiveness. However, because we could not find an appropriate dataset available, we constructed images in which the facial expression intensity changed gradually (11 images per expression) for an arbitrary facial expression using FaceGen software. Fig. 3 shows the comparison between conventional head pose estimation based on CLM with iterative PnP method, and our result for the BigAah expression dataset. Fig. 3 shows the results from processing with the same dataset as commercial APIs. Results show that the method using the inner eye landmark, was relatively less affected by facial motion than others did.

Fig. 3. Experimental results (Ours, Conventional, and Commercial APIs)

4. Conclusion

We experimented to evaluate the method by using only two landmarks with low variance. This requires head pose datasets with various facial expression to verify effectiveness. However, because

Acknowledgment

This research project was supported by the DGIST R&D Program of the Ministry of Science and ICT(17-IT-01). It was also supported the Sports Promotion Fund of Seoul Olympic Sports Promotion Foundation from Ministry of Culture, Sports and Tourism(s072015r2112015A0).

References

[1] Zhibo Guo, Huajun Liu, Qiong Wang and Jingyu Yang, "A Fast Algorithm Face Detection and Head Pose Estimation for Driver Assistant System", International Conference on Signal Processing, Vol. 3, 2006.

[2] Massimiliano Patacchiola and Angelo Cangelosi, "Head Pose Estimation in the wild using Convolutional Neural Networks and Adaptive Gradient Methods", Pattern Recognition, Vol. 71, pp. 132-143. 2017

[3] D. Cristinacce and T.F.Cootes, "Feature Detection and Tracking with Constrained Local Models", Proc. British Machine Vision Conference, Vol. 3, pp.929-938, 2006.

[4] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. "Constrained Local Neural Fields for robust facial landmark detection in the wild." in IEEE Int. Conference on Computer Vision Workshops, 300 Faces in-the-Wild Challenge, 2013.