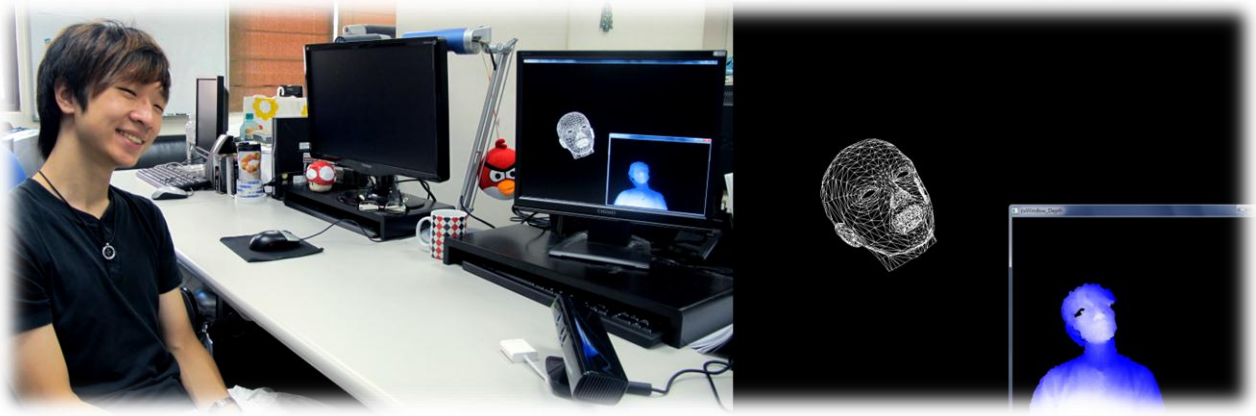
REAL-TIME HEAD POSE ESTIMATION USING DEPTH MAP FOR AVATAR CONTROL

[1]*Yu Tu* , [1]*Chih-Lin Zeng* , [1]*Che-Hua Yeh* , [2]*Sheng-Yen Huang*, [2]*Ting-Xin Cheng*,[1]*Ming Ouhyoung*
[1]*Dept. of Computer Science and Information Engineering*, National Taiwan University
[2]*Reallusion Inc*, Taipei, Taiwan.
{tantofish, yesazcl, chyei}@cmlab.csie.ntu.edu.tw , {elvis, opal}@reallusion.com.tw , ming@csie.ntu.edu.tw



ABSTRACT

In this paper, we propose a system to estimate head poses only using depth information in real-time, thus does not even need illumination. We first track the user's nose, and sample an amount of 3D points around the nose. Then we use a plane to fit the point cloud by least square error method, and the normal vector of the plane yields yaw and pitch angles of the user's head orientation. On the other hand, the rolling angles are estimated by fitting the head contours with an ellipse. Our system adopt depth cameras such as Microsoft Kinect Sensor which has the capability to estimates the depth information in real-time. We also propose algorithms to address the problem of noises in the depth maps. Our system is not affected by the illumination conditions in environments since only depth information is required in our system. We demonstrate that 3D head pose estimation in the range of $\pm 45°$ *in pitch, yaw and* $\pm 70°$ *in roll angle can be achieved at 30 fps with noisy depth data and without the user calibration.*

Keywords *Head Pose Estimation; Depth Map; Kinect; Least Square Error Plane; Calibration Free; Nose Tracking*

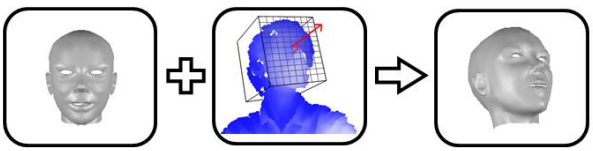
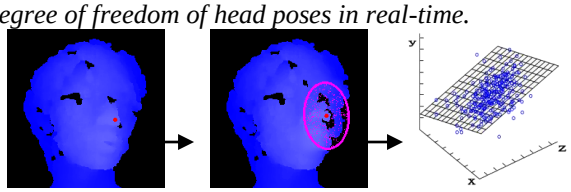
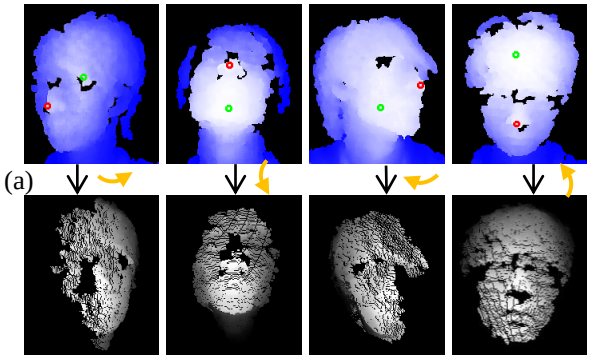


Fig.1: The goal of this system is to estimate the threes degree of freedom of head poses in real-time.



(a) (b) (c)

Fig.3: (a) Detect and track user's nose and (b) sample several points from the nose's neighboring area, (c) then apply least square error algorithm to fit a plane to the sample points.

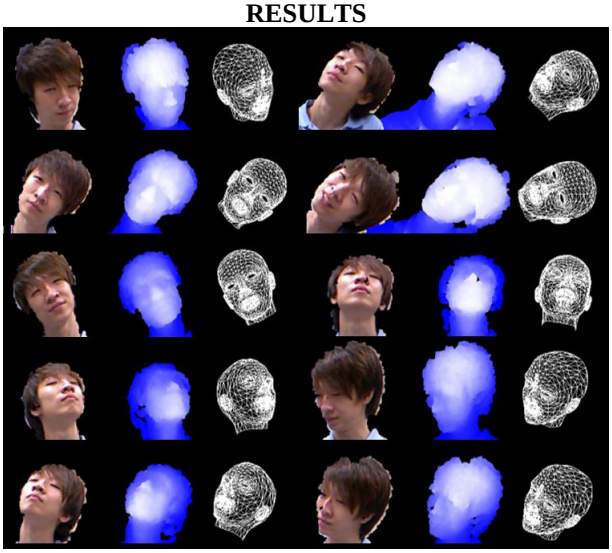


(a)

(b)

Fig.2: In order to locate the nose correctly, a reverse rotation transform is applied to the whole point cloud (a) to rotate the head back to the normal pose (b). Green circle indicates the shallowest point while red circle indicates the shallowest point after reverse rotation.

RESULTS



The results of a user who makes ten arbitrary pose. The first and fourth columns show what pose users make. Note that our system doesn't take any advantage from these color images so that it still performs well when the there is no lighting in the environment.

REFERENCES

[1] R. Yang, and Z. Zhang, "Model-based head pose tracking with stereovision," *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 255-260, 2002.

[2] L.-P. Morency, P. Sundberg, and T. Darrell, "Pose estimation using 3d view-based eigenspaces, " *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 45-52, 2003.

[3] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister, "Real-time face pose estimation from single range images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[4] V. N. Balasubramanian, J. Ye, and S. Panchanathan, "Biased manifold embedding: A framework for person-independent head pose estimation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-7, 2007.

[5] M. Osadchy, M. L. Miller, and Y. LeCun, "Synergistic face detection and pose estimation with energy-based models," *Journal of Machine Learning Research*, Vol. 8, pp.1197 – 1215, 2007

[6] M. Storer, M. Urschler, and H. Bischof, "3d-mam: 3d morphable appearance model for efficient fine head pose estimation from still images," *Proceedings of IEEE International Conference on Computer Vision,Workshop on Subspace Methods*, pp. 192 – 199, 2009.

[7] M. Jones, and P. Viola, "Fast multi-view face detection," *Technical Report TR2003-096, Mitsubishi Electric Research Laboratories*, 2003.

[8] T. Vatahska, M. Bennewitz, and S. Behnke, "Feature-based head pose estimation from images," *Proceedings of IEEE-RAS International Conference on Humanoid Robots*, pp. 330 - 335, 2007.

[9] S. Malassiotis, and M. G. Strintzis, "Robust real-time 3d head pose estimation from range data," *Pattern Recognition*, vol.38, no. 8, pp.1153 – 1165, 2005.

[10] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 499 - 504, 2000.

[11] J. Yao, and W. K. Cham, "Efficient Model-Based Linear Head Motion Recovery from Movies," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. II-414-II-421, 2004.

[12] T. Weise, B. Leibe, and L. Van Gool, "Fast 3D Scanning with Automatic Motion Compensation," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[13] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang, "3D Deformable Face Tracking with A Commodity Depth camera," *European Conference on Computer Vision*, pp. 229-242, 2010.

[14] L.-P. Morency, P. Sundberg, and T. Darrell, "Pose Estimation Using 3D View-Based Eigenspaces," *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 45-52, 2003.

[15] E. Murphy-Chutorian, and M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 31, No. 4, pp. 607–626, 2009.

[16] G. Fanelli, J. Gall, and L. V. Gool, "Real Time Head Pose Estimation with Random Regression Forests," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011.

[17] L. Chen, L. Zhang, Y. Hu, M. Li, and H. Zhang, "Head Pose Estimation Using Fisher Manifold Learning," *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 203-207, 2003.

[18] J. Whitehill, and J. R. Movellan, "A Discriminative Approach to Frame-by-Frame Head Pose Tracking," *Proceedings of the IEEE International Conference on Analysis and Modeling of Faces and Gestures*, pp. 1-7, 2008.

[19] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Real time Performance-Based Facial Animation," *ACM Transactions on Graphics, Proceedings of the 38th ACM SIGGRAPH Conference and Exhibition*, 2011.

[20] E. Seemann, K. Nickel, and R. Stiefelhagen, "Head Pose Estimation Using Stereo Vision for Human-Robot Interaction," *Proceedings of the IEEE International Conference on Analysis and Modeling of Faces and Gestures*, pp. 626-631, 2004.