Automatic Facial & Body Behavior Analysis and Recognition

Kang Wang, Rui Zhao, Yue Wu and Qiang Ji

Department of Electrical, Computer & Systems Engineering, Rensselaer Polytechnic Institute

Introduction

In this work, we introduce a computer vision system that performs automatic facial and body behavior analysis and recognition using RGB-D camera. From RGB video, the system performs facial landmark tracking, which automatically determines the locations of fiducial facial points near major facial components. From depth video, the system extracts skeleton joint positions. Given tracking results, the system performs human eye gaze tracking, head pose estimation, facial expression recognition and gesture recognition.

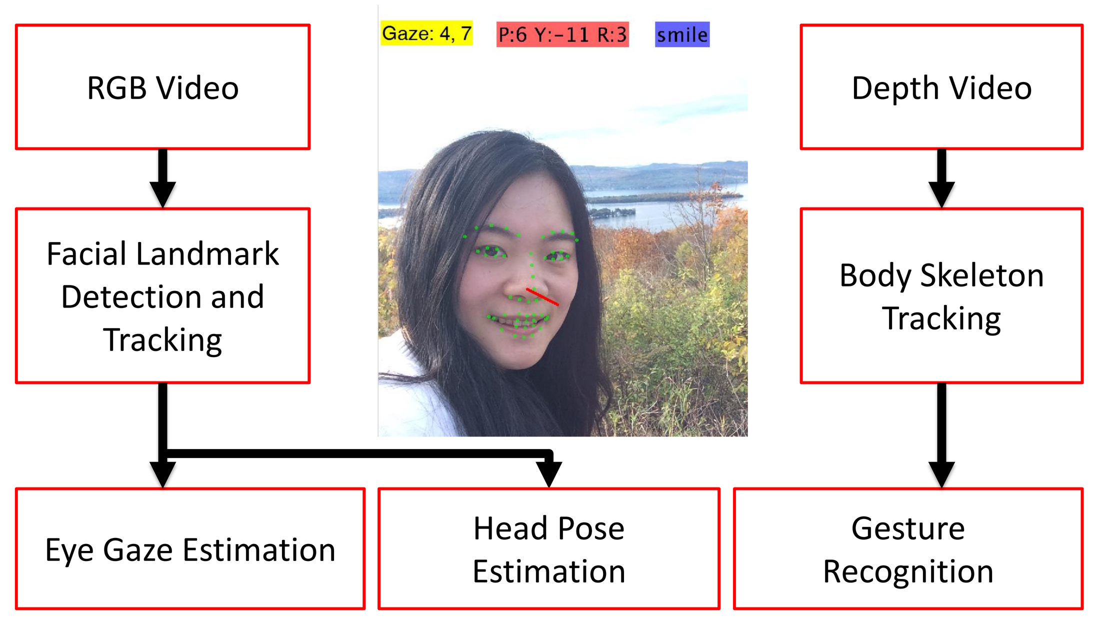

Fig. 1 Automatic facial and body behavior analysis and recognition

Facial Landmark Tracking

Approach [1]:

We proposed the shape augmented regression based facial landmark tracking algorithm that iteratively predicts the face shape updates and estimates the landmark locations. In particular the regression function would automatically change for different face shapes.

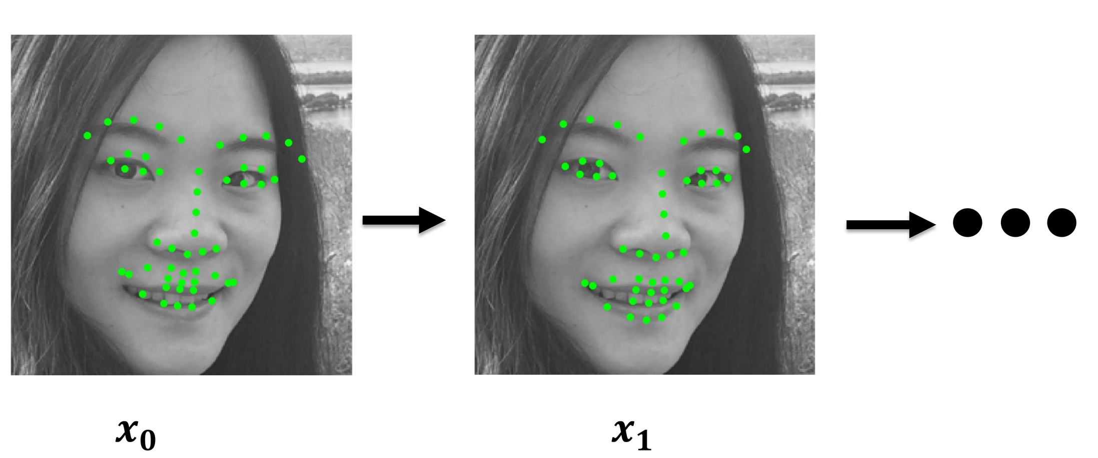

Fig. 2 Shape augmented regression based facial landmark detection method

Algorithm 1: Shape augmented regression method

Initialize the landmark locations $\mathbf{x}^0$ using the mean face

for *t=1, 2, ..., T or convergence* **do**

Predict the landmark location update

$$\Delta\mathbf{x}^t = \mathbf{R}^t\Phi(\mathbf{x}^{t-1}, \mathbf{I}) + \mathbf{b}^t + \mathbf{Q}^t\Psi(\mathbf{x}^{t-1})$$

Update the face shape

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t$$

end

Output the estimated landmark locations $\mathbf{x}^T$

Fig. 3 Facial landmark tracking results

Facial Landmark based Eye Gaze Estimation

Approach:

1) Relate gaze features (eyeball center) with rigid facial landmarks, and obtain their relative position information from offline training.
2) During online gaze tracking, given the detected 2D facial landmark positions, recover the 3D eyeball center and 3D pupil center.
3) Obtain gaze direction or PoR given eyeball center, pupil center and personal eye parameters.

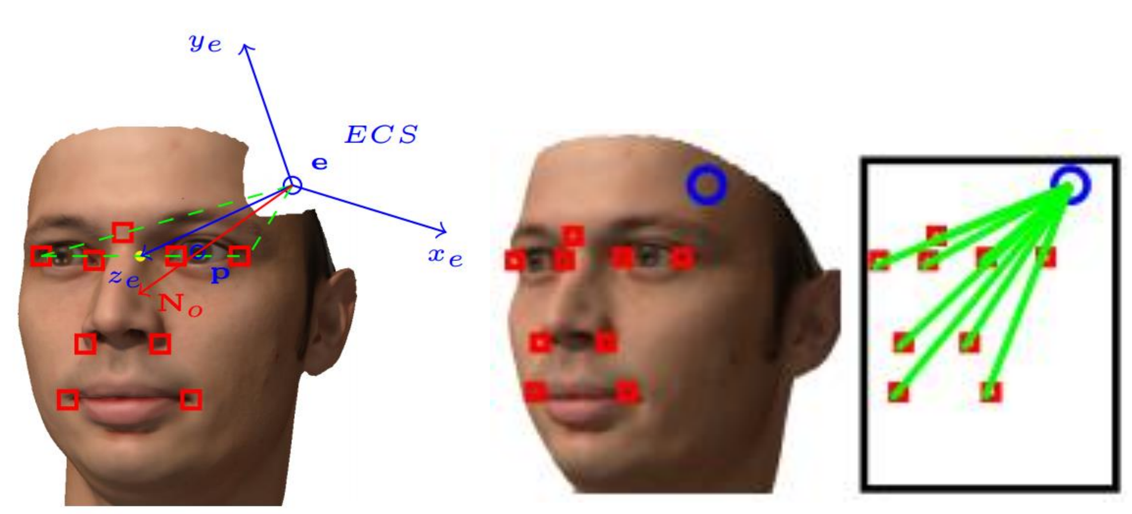

Fig. 4 Relationship between rigid facial landmarks and eye ball center and extracted relative position information. Images are from [2].

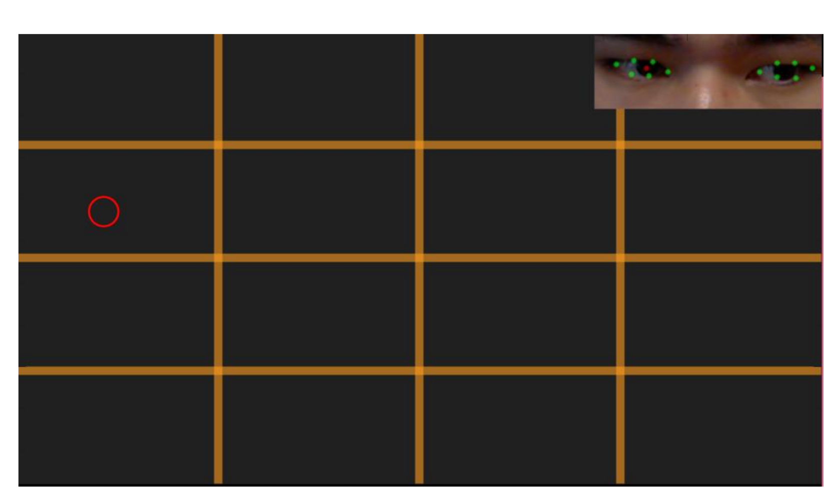

Fig. 5 Screenshots of gaze estimation demo. The red circles represent the current positions the subject looks at. Top right shows the detected facial landmarks and pupil positions.

Facial Landmark based Head Pose Estimation

Approach:

1) Estimate the head pose based on the predicted 2D landmark locations (51 landmarks) and a general 3D deformable face model.

2) The method estimates the projection matrix (depending on the head pose angles) by minimizing the projection error that measures the differences between the projects 2D points (based on 3D deformable model and head pose) and the detected 2D landmarks.

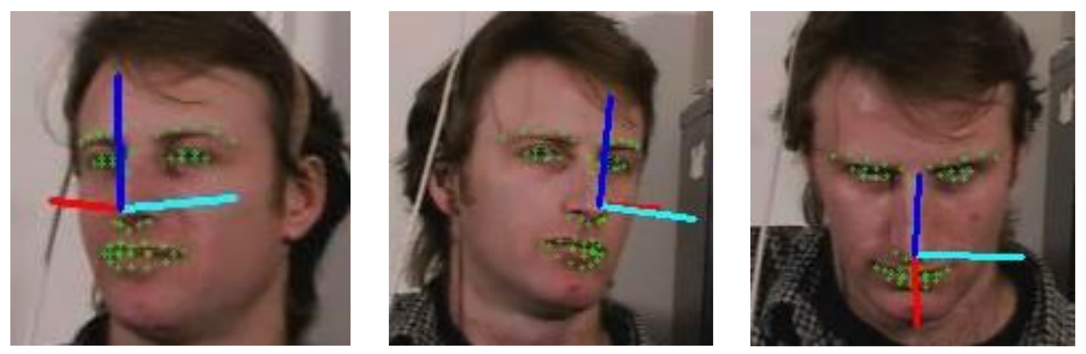

Fig. 6 Pose estimation results

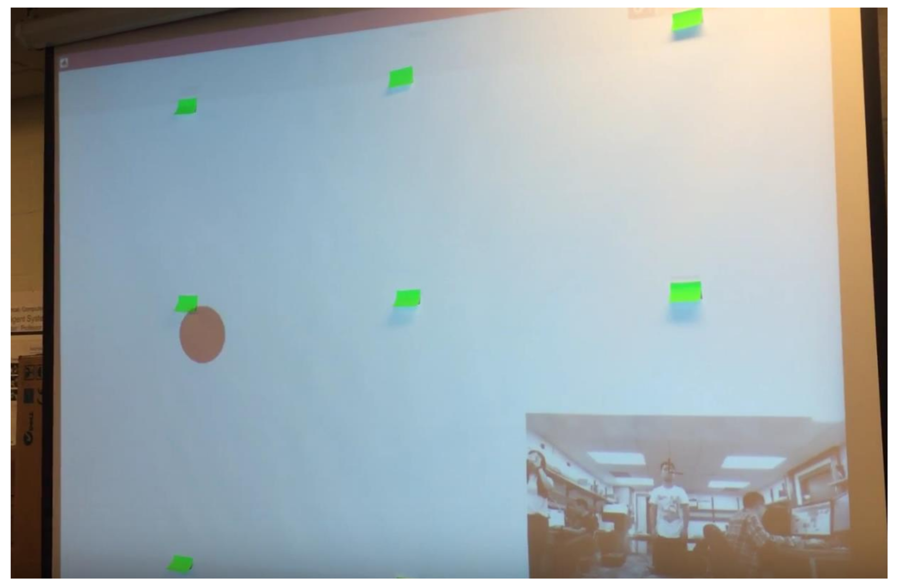

Fig. 7 Cursor control based on head pose estimation

Robot Facial Behavior Mirroring

Approach:

1) Capture human's facial behavior/motion from facial landmark tracking results.

2) Transfer human's facial motion to robot's facial motion by controlling robot's facial joints.

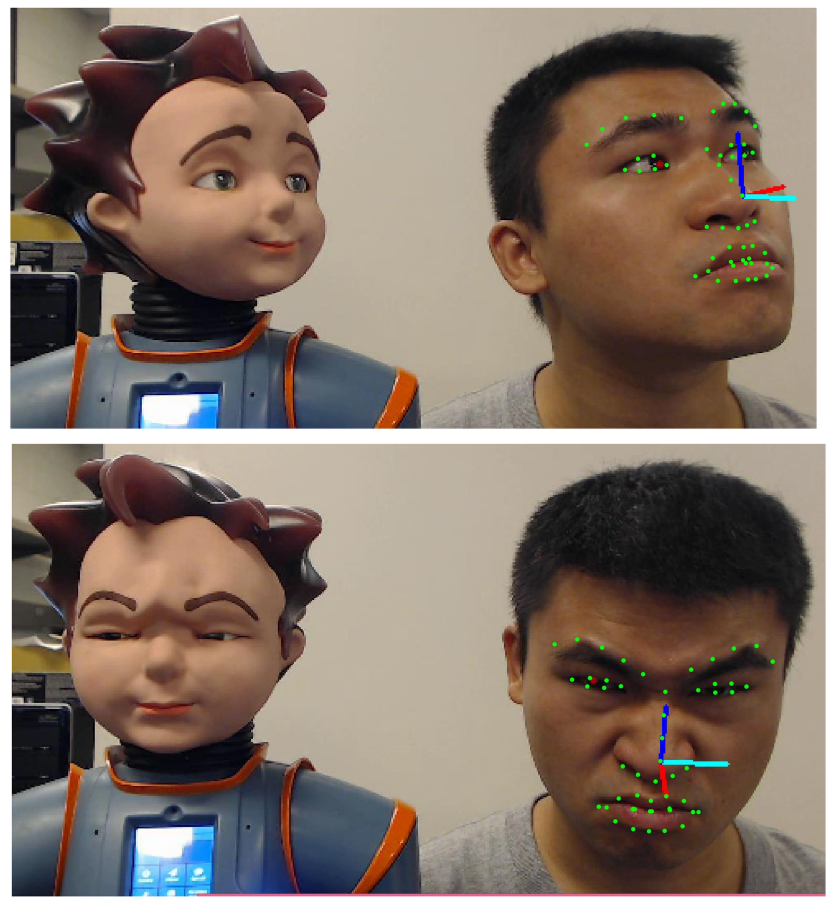

Fig 8. Robot facial behavior mirroring based on facial landmark and head pose estimation

Skeleton based Body Gesture Recognition

Approach:

1) Extract skeleton position data from depth camera.
2) Normalize skeleton data for view and body size invariance.
3) Extract features from skeleton data including joint location, distance, motion and angle.
4) Perform camera calibration.
5) Perform gesture recognition using either geometry based method or statistical model based method, which uses SVM classifier trained offline using publicly available datasets.

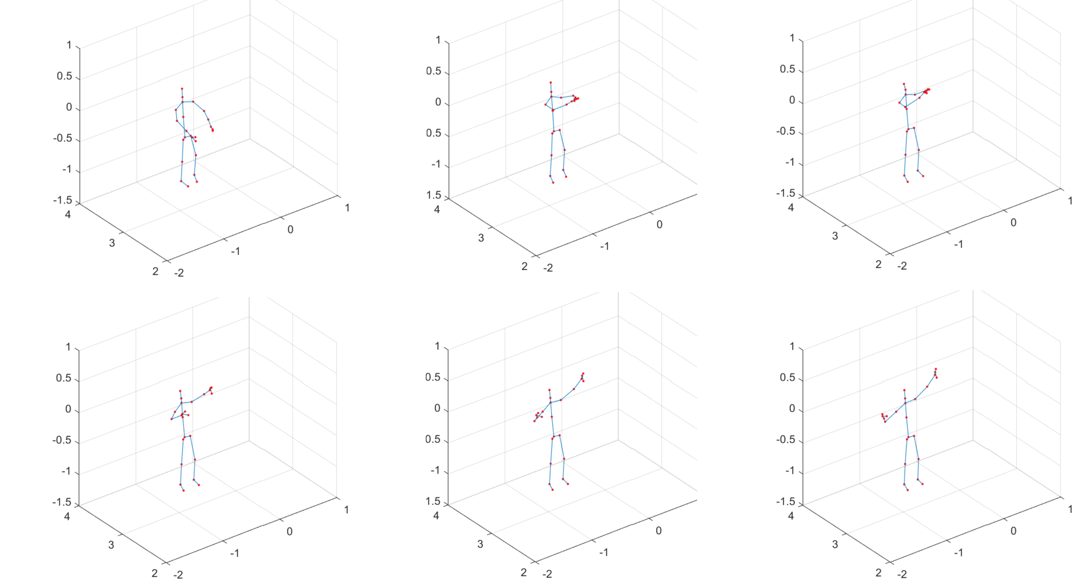

Fig. 9 Visualize skeleton data of zoom gesture

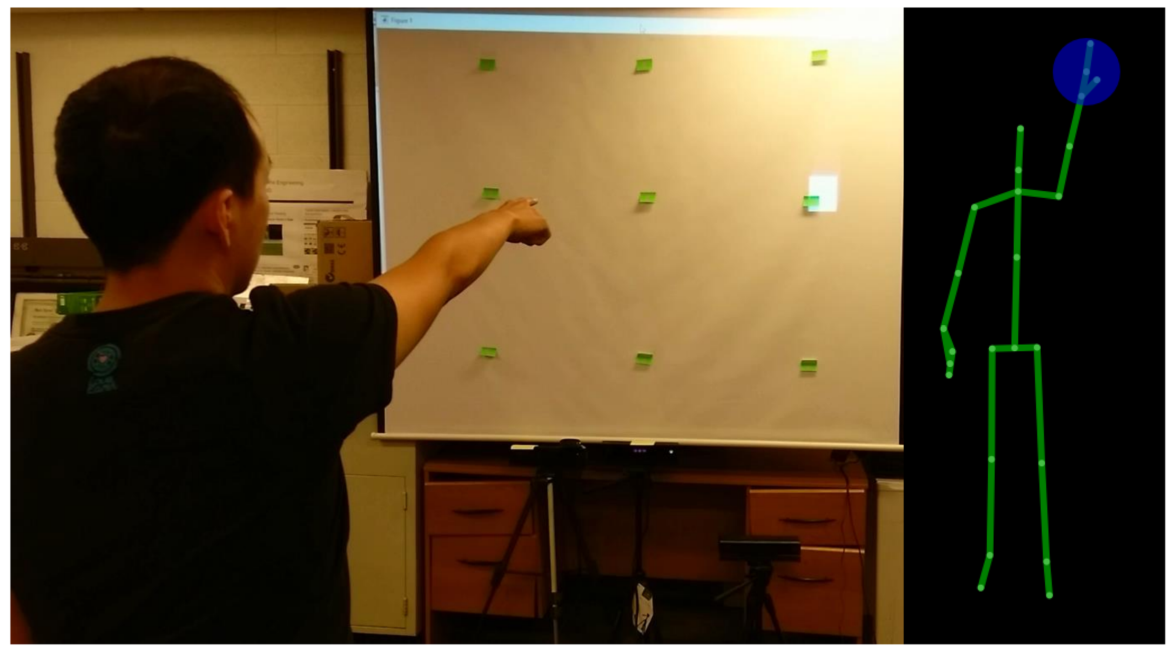

Fig. 10 User controls cursor through pointing direction tracking

References

[1] Yue Wu and Qiang Ji, "Shape Augmented Regression Method for Face Alignment", 300 Videos in the Wild (300-VW) Challenge and Workshop, International Conference on Computer Vision Workshop(ICCVW), 2015.

[2] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. *AVSS*, 2009.

Contact

Intelligent Systems Lab, Dept. of ECSE, Rensselaer Polytechnic Institute
JEC 6219, ECSE Department, Troy, NY, 12180, USA
Prof. Qiang Ji
Tel: + 1-518-276-6440
Email: jiq@rpi.edu
Website: http://www.ecse.rpi.edu/~cvrl/index.html