Accurate Single View Model Based Head Pose Estimation

Pedro Martins ♦ Jorge Batista
Institute of Systems and Robotics
Department of Electrical and Computer Engineering, University of Coimbra, Portugal

Abstract

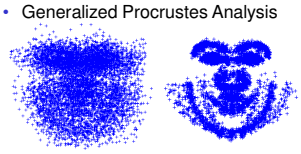
- Head pose estimation from single view images.
- The 6DOF was estimated using Pose from Orthography and Scaling with ITerations (POSIT) where a statistical anthropometric 3D rigid model is used as an approximation of the human head, combined with Active Appearance Models (AAM) for facial features extraction and tracking.
- The results show that orientations and head location were, on average, found within 2º or 1cm error standard deviations respectively.

Active Appearance Models

Shape Model

$$x = (x_1, y_1, ..., x_n, y_n)^T$$

- Generalized Procrustes Analysis



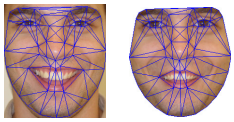
- Principal Components Analysis (PCA)

$$x = \bar{x} + \Phi_s b_s$$

Texture Model

$$g = (g_1, g_2, ..., g_{m-1}, g_m)^T$$

- Piecewise Affine Warp



- Low Memory PCA

$$g = \bar{g} + \Phi_g b_g$$

Combined Model

- Remove correlations between shape and texture model parameters

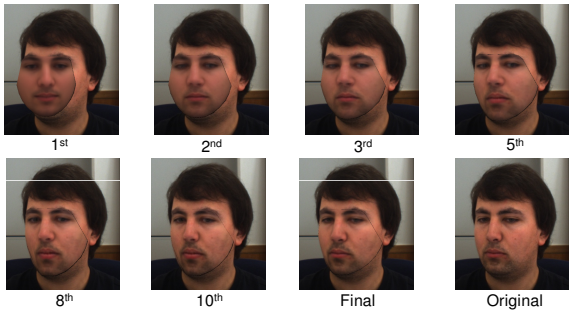
$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s^{-1}\Phi_s^T(x - \bar{x}) \\ \Phi_g^T(x - \bar{x}) \end{pmatrix} \quad \Phi_c = \begin{pmatrix} \Phi_{cs} \\ \Phi_{cg} \end{pmatrix} \Rightarrow \begin{aligned} x &= \bar{x} + \Phi_s W_s^{-1} \Phi_{cs} c \\ g &= \bar{g} + \Phi_g \Phi_{cg} c \end{aligned}$$

- **c** is a vector of appearance controlling shape and texture

Model Fitting

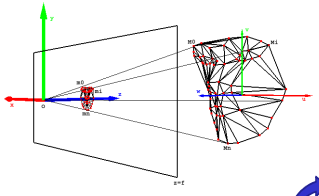
$$\arg\min_c \sum_{pixels} \left[\text{image} - \text{model} \right]^2$$

- Updating the appearance parameters, **c**, and pose



POSIT - Pose from Orthography and Scaling with Iterations

- POSIT is a fast and accurate iterative algorithm for finding the 6DOF of a 3D model given a set of 2D image projections and 3D points correspondences



$$\begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \\ \vdots & \vdots \\ u_{n-1} & v_{n-1} \\ u_n & v_n \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 \\ X_2 & Y_2 & Z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ X_{n-1} & Y_{n-1} & Z_{n-1} & 1 \\ X_n & Y_n & Z_n & 1 \end{bmatrix} \begin{bmatrix} r_1/T_z & r_2/T_z \\ T_x/T_z & T_y/T_z \end{bmatrix}$$

M

Until Pose Converge

Normalize Image Coordinates $u_i = u_i - \frac{c_x}{f}, v_i = v_i - \frac{c_y}{f}$

Compute Model Inverse M^{-1}

Assume $w_i = 1$

Get Scaled Orthographic coordinates $(u_i, v_i) = w_i(u_i, v_i)$

Compute $\begin{bmatrix} r_1/T_z & r_2/T_z \\ T_x/T_z & T_y/T_z \end{bmatrix} = M^{-1} \begin{bmatrix} u_1 & v_1 \\ \vdots & \vdots \\ u_n & v_n \end{bmatrix}$

Find **Tz**, **Tx**, **Ty**, **r1** and **r2**

Compute **r3** by the cross product $r_3 = r_1 \times r_2$

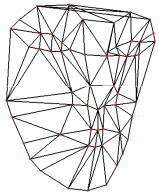
Update $w_i = 1 + \frac{r_3}{T_z}(X_i, Y_i, Z_i)$

Anthropometric 3D Model

- Suitable rigid body model that describes the 3D face of several individuals

- Physical model

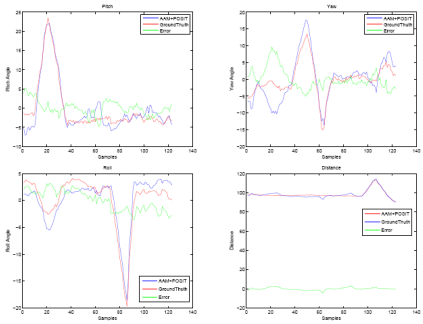
- 3D laser scan



- Sparse model

One-to-One 2D/3D Correspondences

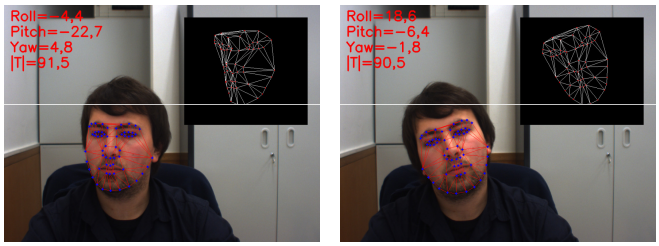
Pose Evaluation



Parameters	Error Avg std
Roll	1.94 deg
Pitch	2.57 deg
Yaw	1.7 deg
Distance	1.33cm

- Comparison between the estimated pose (AAM+POSIT) with the one estimated from a planar checkerboard

Examples of Head Pose Estimation



- The application with AAM fitting plus POSIT pose estimation runs at 5 fps on 1024x768 images using a Intel 3.4GHz P4 under Linux OS. AAM is based on 58 landmarks sampling 48178 pixels with color information (m=144534)

3D glasses Augmentation

- A 3D model of glasses is backprojected on image with the estimated 6DOF

This work was funded by FCT project POSC/EEA-SRI/61150/2004, AI_FI - Advanced Interation using Facial Information
http://aifi.isr.uc.pt/

Institute of Systems and Robotics
University of Coimbra