Computational Anatomy for Generating 3D Avatars

M. Vaillant, G. Zang, J. Aliperti, N. Santhanam, S. Doucette, B. Hoffman, M.I. Miller
Animetrics Inc., Jackson NH, 03846

1 Introduction

Face ID systems and face trackers have been in use for several years. Typically, face ID systems are comprised of detection and identification systems based on the manipulation of 2D likenesses of faces which represent photometric and geometric variation robustly as manifest in the 2D likeness. The main problem with conventional face ID systems are that they are limited to the manipulation of 2D geometric variations of in plane geometric variation, so that while they can be used for tracking and or IDing faces for in plane variations (as is currently done in all systems) their power and efficiencies degrades as the target subjects are viewed out of plane. Since most systems rely on 2D likeness and manipulation of facial features representing the individual they cannot be robust to changes in both photometric and geometric variation which is not directly manifest in the 2D image plane representation of the face being IDed.

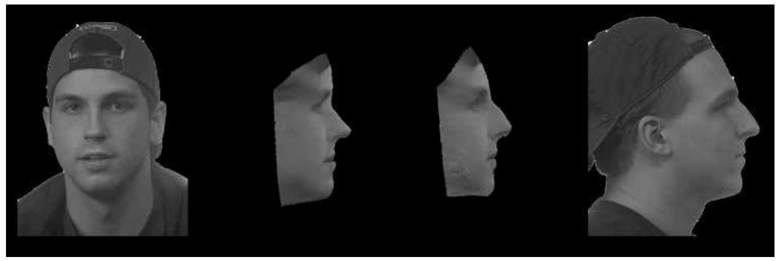

Figure 1. Shown in far left and far right are front and profile view photographs respectively. Second panel shows avatar generated from front view only. Third panel shows avatar generated from both front and profile views.

The avatar based 3D facial ID systems which we are engaged in are robust to geometric variation of the pose of individual faces and efficiently detects and identifies faces from projective imagery such as measured by conventional video cameras. The key technological advance required for their application lies in a system to automatically determine the 3D geometry of a person's face with a finite set of images or photographs or through the analysis of a persistent set of images from video data. Knowledge of the 3D geometry of a subject allows for automated understanding of variables such as photometric and chromatic characteristics of the environment or particular video, occlusion compensation, and pose. The use of a structured 3D model also allows for more finite analysis of human faces for expression analysis and direction of gaze for any pose.

2 Avatar Generation

2.1 One and Two-View Avatar Generation

All of our work in tracking 3D objects has been fundamentally in 3& 4D,i.e. tracking the full Euclidean motions in space and time [1–3]. Our work in Computational Anatomy for tracking and deformation of objects has all been in 3D as well, rather than manipulating motions in the image plane [4]. The power of such 3D approaches is that they are not limited to estimated translation and rotational motion in the image plane, therefore they imply complete invariance to full 3D pose, scale and translation. Herein we extend these methods to accommodate the projective geometry

$$P : (x, y, z) \to p(x, y, z) = \left(\frac{\alpha_1 x}{z}, \frac{\alpha_2 y}{z}\right),$$

associated with video imagery. The core problem, at least for deformable objects, is to infer the flow of the dense volume representations of the 3D geometric objects from 2D projective imagery. Specifically, we shall merge our work on estimating low-dimensional matrix group for tracking associated with rigid body dynamics, as well as our work on high-dimensional transformations for tracking deformable motions (such as facial expressions), to incorporate information obtained from projective geometry. Shown in Figure 1 are results from generating a 3D geometry from a front view as well as a single merged geometry from both front and profile views. Input photographs taken from the FERET [5] database are shown in far left/far right panels. The second panel shows an avatar generated from the frontal view only. The third panel shows an avatar generated from both front and profile views. Notice the clear improvement of the profile view of the geometry generated from both views.

3 Anthropometric Validation of 3D Avatar Generation

We have quantified the geometric accuracy of the 1 and 2 view avatar generation in the projective plane on 95 frontal and profile images from the FRGC_EXP_1.0.1 database [6]. We have generated extensive hand featured data bases in each of the experiments with trained manual raters. Manual raters were trained and retrained over a period of about 1 week, and then were asked to delimit on the order of 20 highly recognizable features in the projective imagery for front and 10 features for side view data. For anthropometric validation, on each 3D model 140 fiducially identifiable points were defined on the surface of the volume. These points became the anatomical markers from which our accuracy analysis is calculated. To determine the geometric accuracy of the 3D avatar generation, each 3D avatar generated from each of the photographs was placed into the 3D space associated with the rigid motion generated for the solution of each variational problem. Using the projective equation, the 140 fiducial markers were projected into the image plane (if not occluded) and were compared to each of a subset of the 40 identifiable features which were manually marked in the respective photograph. The root-mean-square error (RMSE) was calculated for each of the manually labelled photographs for which the experiment was performed on.

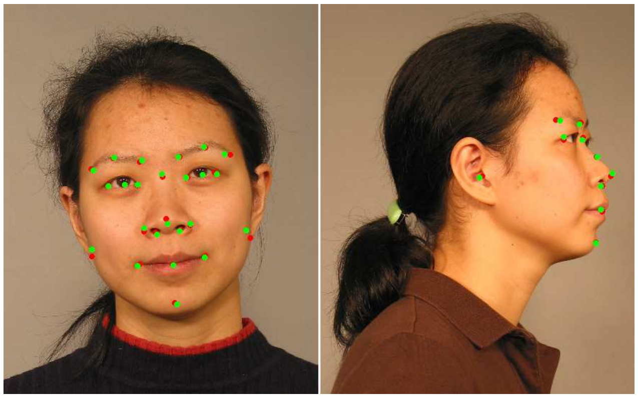

Figure 2. Projected anatomical fiducial features (green) from merged geometry and manually defined features (red) superimposed on front and side views.

Shown in Figure 2 are results depicting the comparison of a subset of the anatomically defined fiducial marks superimposed over the manually defined features. Green points are taken from the single merged geometry, projected into the projective plane. Red points show the manually defined features.

Database	# used	RMSE X	RMSE Y	RMSE
Frontal	95	2.4	1.84	3.07
Profile	95	2.77	2.72	3.91
Multi-Frontal	95	2.38	1.94	3.14
Multi-Profile	95	2.7	2.74	3.8

Figure 3. Table of RMSE for X, Y and sum of the square-root of the sum of squares for the avatar generation accuracy.

The average RMSE between projected and manual features for FRGC frontal, profile, and merged frontal/profile (multiview) are show in Figure 3. All RMSE are relative to a standard face with 64 pixels between the eyes. For the frontal data, the average RMSE shown in column 5 (square root of the sum of the squares in columns 3 and 4) averaged over the landmarked features is about 3.07 pixels corresponding to about 1/20 of an eye distance. For the profile data, the RMSE degrades about 1 pixel to 3.9. For the avatar generation from both profile and frontal view we find slightly degraded RMSE errors for frontal and slightly improved errors for profile.

3.1 Rigid Motion Accuracy

Clearly, the errors in avatar accuracy are determined by our quantitative ability to estimate the rigid motions accurately. We have studied the rigid motion estimation by generating simulated data bases with known rigid motions of up to 30 degrees in X, Y, and Z directions. We observe on the order of 10^{-2} degree bias in Y and Z rotations throughout the full range of simulated rotations, and sub degree bias in X rotations, with under 2 degree standard deviation in X and under .25 degree standard deviation in Y and Z.

References

[1] M. I. Miller, U. Grenander, J. A. O'Sullivan and D. L. Snyder, *Automatic target recognition organized via jump-diffusion algorithms*, IEEE Transactions on Image Processing, 6 (1997), pp. 157-174.

[2] A. Srivastava, M. I. Miller and U. Grenander, *Ergodic Algorithms on Special Euclidean Groups for ATR, Progress in Systems and Control: Systems and Control in the Twenty-First Century*, Birkhauser, (1997), pp. 327-350.

[3] U. Grenander, M. I. Miller and A. Srivastava, *Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20 (1998), pp. 790-802.

[4] U. Grenander and M. I. Miller, *Computational anatomy: An emerging discipline*, Quarterly of Applied Mathematics, 56 (1998), pp. 617-694.

[5] P. J. Phillips and H. Moon and S. A. Rizvi and P. J. Rauss, *The FERET Evaluation Methodology for Face Recognition Algorithms,* IEEE Trans. Pattern Analysis and Machine Intelligence, 22, October (2000), pp. 1090-1104.

[6] P. J. Phillips and P. J. Flynn and T. Scruggs and K. W. Bowyer and J. Chang and K. Hoffman, J. Marques, J. Min, and W. Worek, *Overview of the Face Recognition Grand Challenge,* In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (2005).