Face Modeling and Recognition in 3-D

Gérard Medioni[1], *Roman Waupotitsch*[2]

[1]*USC-IRIS*
PHE 204 mc 0273
Los Angeles, CA 90089

[2] *Geometrix, Inc.*
1590 The Alameda, Suite 200
San Jose, CA 95126

We demonstrate a complete and automatic system to perform face authentication by analysis of 3-D facial shape. In this live demonstration of the system, the subject is first enrolled, and given a unique identifier. Subsequently, the user's identity is verified by providing the reference identifier. Our approach is to be contrasted with traditional face recognition methods, which compare pictures of faces.

Because we are working in 3-D, we overcome limitations due to viewpoint and lighting variations.

Our 3-D acquisition system consists of two off the shelf, calibrated cameras, and a robust stereo processing, resulting in a dense, accurate 3-D representation, even in the presence of facial hair. The physical set-up of the system is shown on Figure 1 below.

Figure 1: The 3-D shape acquisition system

The system has a large working envelope, both indoors and outdoors, under widely varying conditions.

The acquisition system performs stereo matching on 2 images taken with an angular baseline of a few degrees (typically 8-15). The fundamental algorithmic steps are derived from the description in a paper by Chen and Medioni, "*Building Human Face Models from Two Images*," Journal of VLSI Signal Processing, 2001. Here, however, the cameras are calibrated, both internally and externally, which results in true metric reconstruction. We start by building an image pyramid, and at each level, starting at the coarsest one, match reliable interest points; these are then propagated to produce a dense correspondence map, which is refined at the next level of the pyramid. The final disparity map in image format is transformed into a 3-D mesh of triangles, with an associated texture map.

The cameras produce 2.1 Mpixels color images. Connection to a host computer occurs through a USB-2 interface. The total time, from image capture to 3-D mesh inference is around 9 seconds on a dual Pentium IV processor.

The performance of our system is shown on Figure 2.

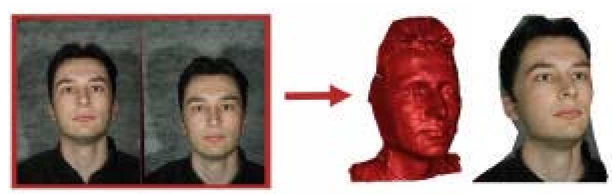

Figure 2: Input images and resulting 3-D shape

To establish the quality of the models we produce, we have compared face models from our system with face models of the same person, the same day, scanned by a Cyberware Head & Face Color 3-D Scanner Model 3030RGB/PS, which is specified to provide accuracy of ~100 microns at 0.25mm image sampling density. We then computed the error between the laser models and the models produced by our system. The models are compared after alignment. The mean and median errors between the two are 0.93mm and 0.92mm respectively.

Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)

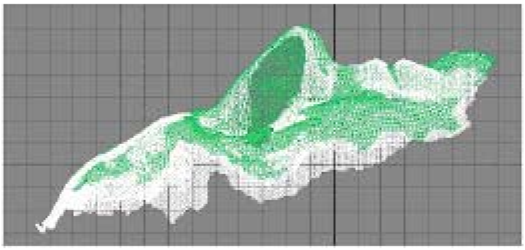

Figure 3: Cyberware(white) and our model after alignment

We also analyzed the image quality requirements in order to generate a good quality 3-D reconstruction from stereo. This is especially relevant under less than ideal conditions. We found that a dynamic range of 32 grey levels is sufficient. This implies that the system should be able to function under severe under and/or over exposed conditions.

We then asked how much light is needed to generate such images. We found that the system can handle wide variations, from 90 to 300 lux, with an exposure time of 80 ms. The exposure was measured with an off-the shelf photographic lightmeter, and the variation accounts for light changing from regular home lighting at night to bright neon lighting. If the light intensity is brighter, the exposure needs to be reduced accordingly. At the other end, the integrated lamps produce enough light even in an otherwise dark room.

We finally asked what range of positioning is allowed for the face with respect to the camera, as it affects both focus and resolution. We found that the face can move up to about 30cm from its optimal distance to the cameras, without noticeable change of quality.

In all experiments, the degradation outside the range is quite graceful.

We have also conducted a number of field experiments, indoors and outdoors, to experimentally validate the above results. We find that, indoors, the system should be useable under normal lighting conditions, even without the integrated lamps. Outdoors under bright sunlight, the illumination reaches over 80,000 lux, and we reduced the exposure time to 1.2 ms. Even under such extreme conditions, we produced good models with the sun behind the subject, behind the camera, and to the side, as shown on Figure 4. The most challenging one comes from side illumination, as the intensity variations become quite extreme.

Figure 4: Outdoors experiment input images

Our 3-D recognition engine performs one-on-one comparison of a probe 3-D model with a reference model for verification, or with *all* existing reference models for identification. We automatically align the two meshes, then compute a distance map between the 2 aligned meshes, and finally perform the classification based on the statistics derived from the distance map.

The recognition results are significantly better than those obtained by 2-D systems, and we expect them to scale up to larger databases.

To validate our 3-D recognition engine, we ran it on all possible pairs from a database of 100 subjects, each acquired in 7 different poses within ±20º of a frontal view.

To compare out results with existing methods, the input images to the stereo algorithm were also directly input to the Identix FaceIt® 2-D recognition engine, through the available SDK interface. This should provide a well-controlled comparison between the methods, as it uses identical image sets. Our 3-D method yielded an equal error rate better than 2%, versus nearly 6% for the Identix FaceIt® system. Note that FaceIt® was rated one of the "top-three" in the FRVT2002 test. Our 3-D system was also able to achieve 0% false acceptance at 3% false reject rate, which suggests the ability to scale to large databases while maintaining high levels of biometric discrimination, under many lighting conditions.

These results are illustrated on the chart in Figure 5. It should be noted that the FRVT 2002 database is much larger, and that our results therefore need to be validated on a larger database.

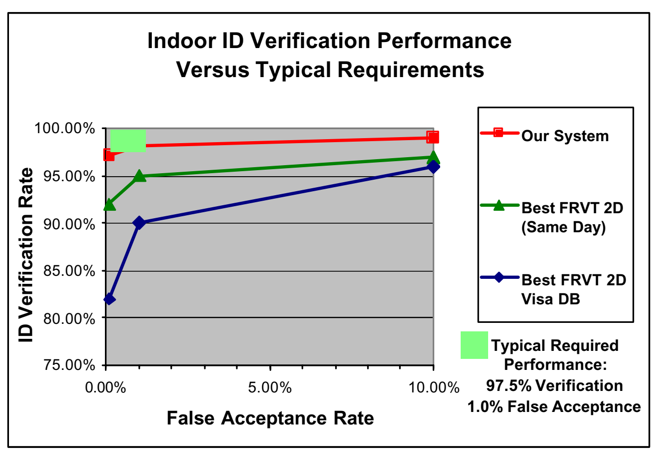

Figure 5: 3-D Face Authentication Performance