Impact of Involuntary Subject Movement on 3D Face Scans

Chris Boehnen
University of Notre Dame
Notre Dame, IN
cboehnen@nd.edu

Patrick Flynn
University of Notre Dame
Notre Dame, IN
flynn@nd.edu

Abstract

The impact of natural movement/sway while standing still during the capture of a 3D face model for biometric applications has previously been believed to have a negligible impact on biometric performance [1]. Utilizing a newly captured dataset this paper demonstrates a significant negative impact of standing. A 0.5 improvement in d' [2] (test of correct/incorrect match distribution separation) per 3D face region and noticeable improvement to match distributions are shown to result from eliminating movement during the scanning process. By comparing these match distributions to those in the FRGC dataset this paper presents an argument for improving the accuracy of 3D face models by eliminating motion during the capture process.

1. Introduction

Biometric authentication and identification are of great interest for applications such as border security, identification of criminals/terrorists, bill payment, eligibility for social services, and more. Face recognition, is especially desirable because of its relative societal acceptance, ease of capture in public spaces, and the possibility of verification of results by humans. Three-dimensional face biometrics has the potential to accommodate a greater variance in pose and lighting conditions than 2D face recognition, and thus is potentially useful for deployment applications.

One of the most widely utilized type of 3D scanner today is the laser scanner such as the Konica Minolta Vivid 900/910 [3]. The Face Recognition Grand Challenge (FRGC) 3D face database [4], Colbry et. al [5], and others have utilized this scanner to capture 3D face data of standing subjects. The FRGC 3D face data has been utilized by many works to analyze, compare, and assess the performance of 3D face biometrics [6][7][8][9][10][11][12][13]. While the accuracy of 3D scanners such as the Minolta for face scanning has been previously assessed by the authors [1], it was not assessed using live subjects (it used face masks mounted on tripods). At the time, we noted that the impact of natural movement while standing may affect scan accuracy, but did not assess this impact.

To explore the impact of standing and movement during the Minolta's 2.5 second scanning process we captured two new datasets. Despite the best efforts by a human subject to stand still, some natural movement in the form of swaying will occur while standing. The 'movement dataset' (hereafter referred to as the M dataset) had subjects standing as still as they could. The 'no movement dataset' (the NM dataset) employed sitting subjects with their head braced against a wall to minimize movement as much as possible.

The majority of 3D face biometrics employ a surface distance/ICP [15][14] based match score. The match score in these cases is based upon the distance between two 3D face models, and is an error distance between the two models. We calculate this distance utilizing the 3D signatures approach [15] previously proposed by the authors. The match score, based upon surface error distance, is useful for biometrics, but also for assessing 3D face model accuracy since it is representing the surface difference between scans.

When analyzing the match distribution of the M, NM, and the Fall 2003 FRGC datasets we found a significant difference in match distributions between the NM and M dataset. By minimizing any movement of the subjects' heads, we show a considerable improvement in match/nonmatch score distribution separation (an average per region improvement of 0.5 d'). This suggests that current 3D face biometrics characterized using the FRGC dataset should yield improved performance with a dataset captured under different capture conditions such as the following.

- Decreased acquisition time (e.g. faster stripe movement, stereo capture, etc.)
- Head fixturing or other means to minimize head motion

In this paper, we begin by surveying the accuracy of 3D scanners, the FRGC [4] dataset, and 3D face biometrics in general. Then we present our approach to calculate 3D face match scores based upon a Euclidean surface distance approach [15]. Next, we describe our experiment to capture the M and NM datasets followed by the results from the experiments. Finally, we conclude by discussing the useful information drawn from this experiment.

2. Previous Work

In this section, we discuss issues relating to 3D laser scanners. We focus on the Minolta Vivid 900/910 scanner and discuss the Face Recognition Grand Challenge (FRGC) 2.0 3D face database. The FRGC database is the most widely utilized 3D face database in the field, and since it utilizes the Minolta scanner for data capture is highly relevant to the work here. Finally, we will explore the performance of a few state of the art 3D face biometrics.

2.1. Mechanics of Involuntary Subject Movement

When a human stands without leaning or stabilizing themselves some natural involuntary movement sometimes referred to as sway occurs. Bottaro et. al [16] attempted to analyze sway and model how it occurs. They found that sway movement typically occurs in either patterns of four to six small sways or two large sways. An individual subject's sway pattern over time typically consists of a combination of these types of movement. They found that an average sway lasts approximate 0.40 seconds and that the human body typically produces 4-5 muscle bursts per second in order to stay standing. Bottaro et. al [16] propose that this involuntary movement is the result of 'low resolution' balance sensory data combined with a neurological delay in processing response to exterior stimuli.

To back up this claim Bottaro et. al [16] point to the fact that a light finger touch on an exterior non-moving object allows for a significant reduction in sway. Touch is an extremely accurate sense. During sway the feet stay fixed creating a pivot point, and fingers are vertically a far distance from the pivot point. This distance means that sway creates a larger movement farther from the pivot point which is more easily measureable at the fingers than lower down in the body such as the knees. They propose that this is why a light touch on a surface while standing results in a significant reduction in sway. Whatever the cause of sway, non-intentional sway while standing occurs in human beings. This prompts the focus of this paper, which is what the impact of this sway during scanning has on the resulting 3D face model.

2.2. Accuracy of Minolta Laser Scanner

Previous studies of 3D scanners found the Minolta Vivid 910 to be the most accurate 3D scanner with an average RMS error of approximately .08 mm [1]. The experiments were performed utilizing white plastic face models fabricated from face scans and mounted on tripods. This eliminated the possibility of assessing scanner accuracy due to subject sway since live subjects were not involved in the scanning process.

As we noted in [1], while the Minolta was the most accurate scanner reviewed, it requires 2.5 seconds to complete a scan. While the subject stands, some natural sway in movement during the 2.5 seconds is typical which as noted may cause scanning error greater than the .08mm found with static face models. Previous work [1] did not quantify the influence of sway movement with live subjects and hypothesized that it was most likely insignificant.

2.3. 3D FRGC Face Dataset

The 3D Face Recognition Grand Challenge (FRGC) 2.0 dataset [4] was developed for the testing and development of 3D face recognition systems. Images were captured utilizing a Minolta Vivid 900/910 3D laser scanner. The validation portion of the dataset consisted of 4007 images of 466 subjects. Subjects stood for the data acquisition process. Some subjects carried purses, backpacks or other materials on their person during the scanning process.

The FRGC dataset has been and is currently being utilized in the development and testing of a large number of 3D face recognition systems (e.g. [6][7][8][9][10][11][12][13]). Large movement artifacts in 3D scans present in the database have been noted in previous work such as [7] and [13] resulting in 3D scans such as the one in Figure 1. However, small amounts of movement may not be visible to the human eye and is more difficult to detect. The effect of small amounts of movement on face recognition is not yet known and is the focus of this paper.

Figure 1: Sample Movement Artifact in FRGC 3D Database

2.4. 3D Face Recognition

Improvement of 3D face recognition performance is the ultimate goal of our attempt to understand and quantify the impact of subject movement on 3D laser scan data. This makes a basic understanding of 3D biometric methods helpful. We present three 3D face biometric approaches that we believe to be representative of many of the state of the art 3D face biometric approaches.

Kakadiaris et al. [12] produced depth maps utilizing a spin image converted into a wavelet representation utilized for comparison. Their approach achieved a 97.3% rank one recognition rate on the FRGC database. Their approach is unique in that the match score is a result of a wavelet comparison that does not require the use of ICP for each match score computation.

Faltemier et al. [7] used multiple regions around the face determined by spherical cuts to obtain 28 face regions. These regions were aligned via ICP and the surface error distance was used as a match score. The results from each region were fused together utilizing a modified borda count approach. The fusion of these regions achieved a 97.2% rank one recognition rate on the FRGC dataset.

Boehnen et al. [15] presented a method for 3D face biometrics utilizing 3D signatures. 3D signatures are a vector representation based upon an operator pre-defined reference region that allows for an RMS distance comparison between surfaces similar to the match score utilized by Faltemier et al. [7] and others. We present a brief overview of this method in the next section; for a complete discussion refer to [15].

3. Method

In this section, we describe the 3D scanner used in this work as well as the method employed to calculate match scores.

3.1. Konica Minolta Vivid 900/910 Laser Scanner

All of the captured data utilizes the same Minolta Vivid 910 scanner [3] used in the FRGC dataset. The Vivid 910 scanner uses a single camera and a moving laser stripe, and acquires 3D data using triangulation. The scanner can take a full detailed scan in 'fine mode', which requires approximately 2.5 seconds to complete. Alternatively, a fast scan at lower quality can be taken in approximately 0.3 seconds utilizing the same process. The data captured for this paper was taken in the fine mode, as was the case when the FRGC dataset was taken. During the 2.5 second scan time, subjects must remain motionless or a poor scan will result.

3.2. 3D Signature Generation

Many traditional 3D face matching approaches such as [7][8] have calculated the Root Mean Square (RMS) error distance between two aligned 3D face surfaces (the gallery and probe faces) as the basis for match score computation. In order to speed up the match score calculation process we have previously proposed a 3D signatures approach [15] that utilizes a vector representation of the 3D surface to calculate the distance between two 3D surfaces. The reference region represents the regions on the face which are being compared by this surface distance calculation. A complete discussion of how to calculate surface distance based match scores is outside of the scope of this paper and we refer the reader to [15] for a detailed explanation. Figure 2 shows the face regions compared in the different match score calculations.

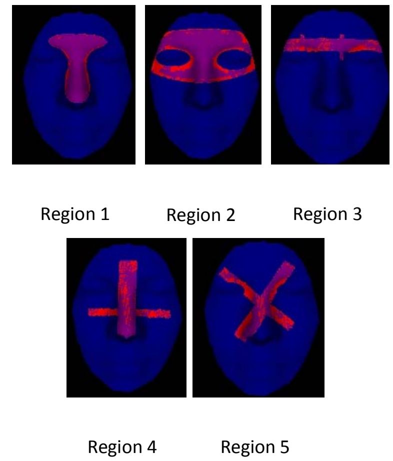
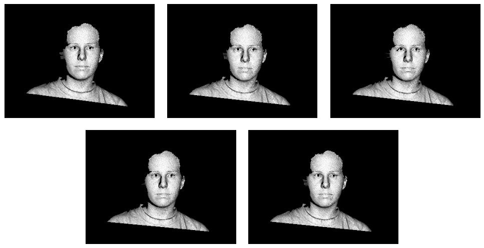


Figure 2: Reference regions (shown in red) utilized

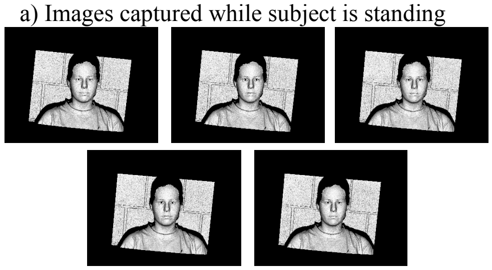
4. Experiment

In this section, we will describe the experimental setup utilized to capture two new datasets. Our goal in this experiment is to assess the impact of a standing subject's involuntary body sway on 3D face models and matching scores. We created two datasets; one with free standing subjects (dataset M), and another with subjects sitting leaning their head against a wall to minimize movement (dataset NM). While other issues such as deformation and lighting conditions exist that may contaminate the 3D scans, we attempted to minimize them as much as possible so that the dominant experimental variable is involuntary subject movement. We also captured subject entries in M and NM during the same acquisition session as close as possible temporally.

We captured data from 34 different subjects (13 females and 21 males). For each subject we captured five 3D scans while standing. Subjects were asked to "hold still" consciously as much as possible and to remain in the same position and pose for each scan. Subjects did not lean on or touch any other objects around them, which has been shown to reduce involuntary sway [16]. The five scans were taken in rapid succession and no changes were made during the acquisition process. We show the 3D data captured for a single subject in Figure 3.a.



a) Images captured while subject is standing



b) Images captured while subject sits leaning against the wall

Figure 3: Images of one subject captured while standing and sitting

We then captured five additional 3D scans of each subject sitting and leaning their head against the wall. A high chair was used so that subjects were approximately the same height as when standing. Subjects were asked to hold as still as possible and to brace their heads against the wall behind them. We show the visual result for a single subject of the sitting and leaning scans in Figure 3.b.

These two different sets, standing (M) and sitting while leaning (NM), represent our attempt to evaluate the impact of natural movement on the 3D scanning process. Since the amount of time between scans was minimized as much as possible, and subjects were asked to keep the same expression throughout the scans are as similar as we can make them with the dominant experimental variable being standing vs. sitting and leaning. Other datasets such as the FRGC may have more extreme movement that might be the result of wearing backpacks or other heavy items. We did not examine these types of large movement artifacts in our experiment.

In addition, our subjects were specifically asked to “consciously” stand still during the scanning process. While subjects in the FRGC dataset also stood still, this was not stressed. These two influences suggest that the amplitude and perhaps impact of movement on the FRGC database and match scores are similar to, or greater than the impacts observed here.

5. Results

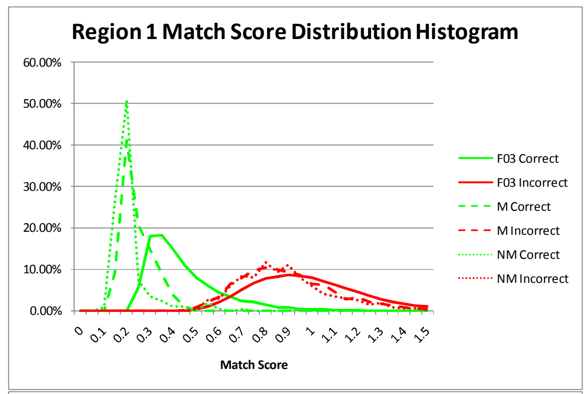
We generated correct and incorrect match score distribution histograms for the M, NM, and Fall 2003 FRGC dataset for each of the 5 reference regions. These results can be seen in Figure 4. Then, we examined the impact of missing data on the match score distributions.

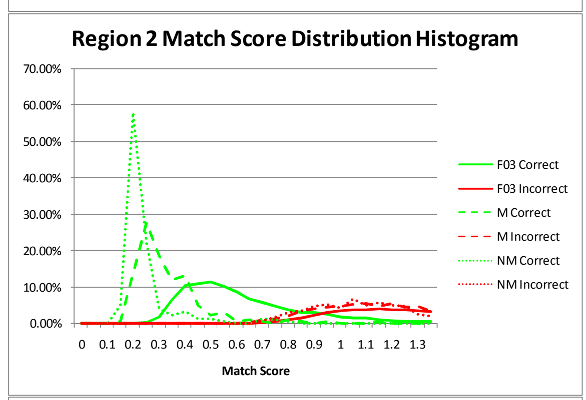
5.1.1 *Understanding match distributions*

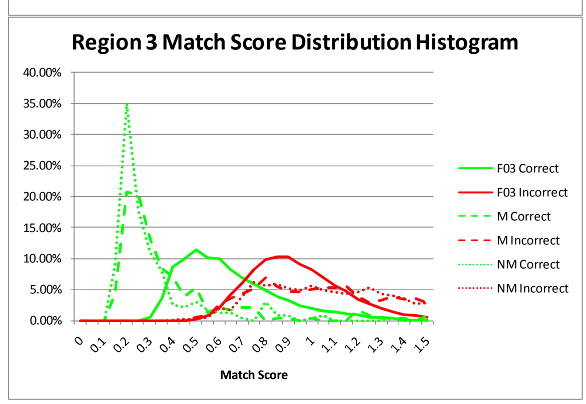
Better score distributions for biometrics are those that exhibit a greater separation between correct and incorrect matches. This is also representative of 3D face model accuracy, scans of the same face should be more similar than those of different faces. Improved separation of the match and nonmatch means, and decreased standard deviation of the distributions both correspond to higher matching performance.

5.1.2 *Non-Match distribution*

The non-match distribution is consistent within a region and does not appear to be dataset dependent. This is expected, since the non-match distribution is relatively unaffected by time, or movement and is instead primarily affected by the difference in facial geometry between different subjects.







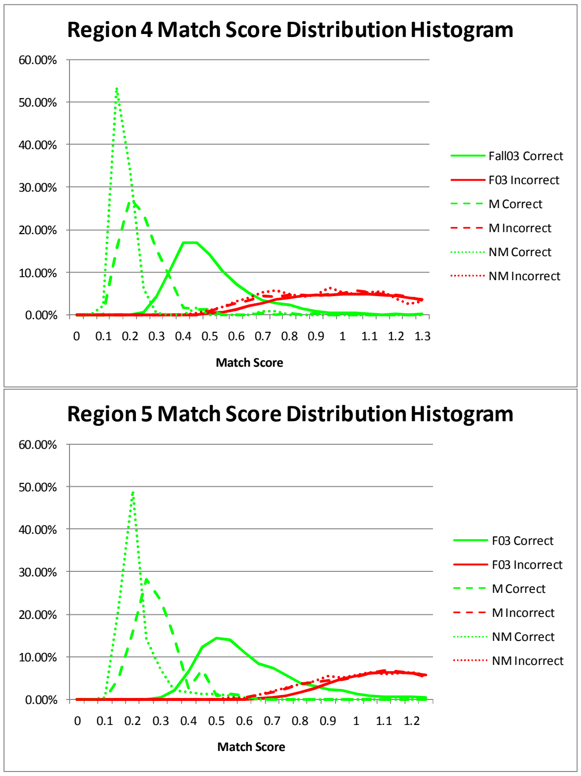


Figure 4: Histogram of match distribution by region (NM = No Movement, M = Movement, and F03 = the FRGC Fall 2003 dataset)

5.1.3 Fall 2003 Correct Match Distribution

The correct match distribution shows a clear and consistent trend across all 3 datasets for each region. The Fall 2003 dataset yielded a larger standard deviation and smaller separation between the correct and incorrect match score distribution means. We expected this because the Fall 2003 data was captured over a period of several months resulting in a large temporal influence on the data as well as due to expression change and sway.

5.1.4 Movement and no movement match distributions

We expected data captured on the same day to outperform data captured over several months, and that the incorrect match distribution would be affected more by facial geometry than anything else and hence stay relatively consistent. However, the impact of movement on the scanning process has traditionally been considered fairly low and not something to be concerned with. However, the results here show a large and consistent difference in separation between the M and NM match distributions with the NM dataset having a superior match distribution. This infers that the 3D face models from the NM dataset are more accurate and consistent to the true 3D face surface of the subjects since the match scores are based upon surface distance error.

5.1.5 Numerical Comparison

d□' [2] is calculated based upon the means and standard deviations of the correct and incorrect distribution as shown in Equation (1). Since the match scores are based upon error in surface distance, d' is also expressing 3D face model repeatability, a form of 3D face accuracy [1], of the resulting scans.

$$d' = \frac{\left|\overline{correct} - \overline{incorrect}\right|}{\sqrt{.5*(((\sigma(correct))^2 + (\sigma(incorrect))^2)}} \qquad (1)$$

d' values for each region and experiment are shown in Table 1. The NM datasets shows an average per region improvement over the M dataset of 0.5 in d'.

TABLE 1: D' VALUES FOR MOVEMENT AND NO MOVEMENT DATASETS

	Movement	No Movement
1	4.35	4.79
2	2.62	2.95
3	2.26	2.82
4	2.27	3.122
5	3.68	4.03
Average	3.036	3.5424

6. Conclusion

We have created two new datasets. The first was captured under conditions to allow for natural movement, which has occurred in widely used databases such as the FRGC 3D face data set [4]. The second dataset minimized natural human movement by having the subjects lean their head against a wall to eliminate any small amounts of sway that naturally occurs. By comparing those results, we showed that natural human movement has an adverse affect on match distribution and as a result 3D face biometric performance.

By examining the match distributions in the 'movement' and 'no movement' datasets we discovered a significant difference. The 'no movement' dataset showed significantly improved match distributions and and a d' improvement of approximately 0.5 per region for each of five regions examined. When considered with the FRGC data, reducing the impact of movement on the 3D data should result in a similarly large improvement. A 0.5 improvement in d' for all 3D face biometrics is significant.

The movement that can occur during the 2.5 second scanning process means that smaller locally connected regions of the face are likely to be more accurate than larger regions because the movement error is cumulative. It is possible that many of the current state of the art techniques such as those by Faltemier et. al [7], Kakariaris et. al. [12], and Boehenen et al. [15] which utilize smaller regions of the face for comparison may be benefiting from this affect. By focusing on smaller regions, the movement error introduced is smaller which should yield improved biometric performance.

Knowing that movement with laser scanners has an impact on biometric accuracy prompts a change in data collection to improve the models. While this information does not allow us to improve existing datasets such as the FRGC it does allow for several improvements for future dataset collection which should yield improved biometric performance.

- Decrease scanning time for laser scan method
- Capture data while subject's head is as still as possible
- Alternate capture methods that acquire data instantly such as stereo

These experiments demonstrate that improvements in scanning technology and collection methods to decrease the impact of movement can improve the performance of 3D face biometrics. These experiments suggest that improved performance in 3D face biometrics can still come from the acquisition stage to the match score calculation steps. The creation of new 3D face datasets is warranted and should result in increased biometric performance for those datasets compared to existing datasets.

7. Acknowledgements

Biometrics research at the University of Notre Dame is supported by the Central Intelligence Agency, the National Science Foundation under grant CNS01-30839, by the US Department of Justice/National Institute for Justice under grants 2005-DD-CX-K078 and 2006-IJ-CX-K041, by the National Geo-spatial Intelligence Agency, by UNISYS Corp, by the Intelligence Advanced Research Projects Activity, and by the Biometrics Task Force and by the US Army under contract W91CRB-08-C-0093.

[1] C. Boehnen, P. J. Flynn, "Accuracy of 3D Scanning Technologies in a Face Scanning Scenario," 3D Imaging Conference, pp. 310-317, 2005.

[2] John Daugman, "How iris recognition works," IEEE Transactions on Circuits and Systems for Video Technology, vol 14, pp. 21–30, 2004

[3] http://www.minolta-3d.com/

[4] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," IEEE Conf. on Computer Vision and Pattern Recognition, pp. 947–954, 2005.

[5] D. Colbry, O. Folarin and G. Stockman, "3D Face Identification - Experiments Towards a Large Gallery" *SPIE Defense & Security*, 2008

[6] Blanz, V., Scherbaum, K., and Seidel, H.-P., "Fitting a Morphable Model to 3D Scans of Faces," *Proceedings Of the International Conference on Computer Vision,* 2007

[7] T. Faltemier, K.W. Bowyer and P. J. Flynn. "A region ensemble for 3D face recognition." *IEEE Transactions on Information Forensics and Security*, pp.62-73, 2008.

[8] K.I. Chang, K.W. Bowyer, and P. J. Flynn. "Multiple nose region matching for 3D face recognition under varying facial expression." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695-1700, 2006.

[9] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[10] Boehnen, C. and Russ, T. "A Fast Multi-Modal Approach to Facial Feature Detection," *IEEE Workshops on Application of Computer Vision*, pp. 135-142, 2005.

[11] Russ, T., Boehnen, C., and Peters, T. 2006. "3D Face Recognition Using 3D Alignment for PCA," *Computer Vision and Pattern Recognition*, pp. 1391-1398, 2006.

[12] I.A. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, and T. Theoharis. "3D Face Recognition". *Proceedings of the British Machine Vision Conferenc*e, pp. 200-208, 2006.

[13] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, "Performance of Geometrix ActiveID 3D Face Recognition Engine on the FRGC Data," *Face Recognition Grand Challenge Workshop*, pp. 154–154, 2005.

[14] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 239– 256, 1992.

[15] C. Boehnen, T. Peters, and P. J. Flynn, "3D Signatures", International Conference on Biometrics 2009.

[16] A. Bottaro, Y. Yasutake, T. Nomura, M. Casadio, and P. Morasso "Bounded stability of the quiet standing posture: An intermittent control model" *Human Movement Science*, vol. 27, iss. 3, pp. 473-495, 2008.