Effects of Eye Position on Eigenface-Based Face Recognition Scoring

Joe Marques
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102
1.703.883.6337
jmarques@mitre.org

Nicholas M. Orlans
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102
1.703.883.7454
norlans@mitre.org

Alan T. Piszcz
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102
1.703.883.7124
apiszcz@mitre.org

ABSTRACT
Eigenface based facial recognition systems rely heavily on pre-determined eye locations to properly orient the input face prior to template generation. Gross errors in the eye detection process can be identified by examining the reconstruction image of the resulting eigenspace representation. Subtle variation in the precision of eye finding that does not prevent subsequent enrollment has not been effectively studied or reported by the biometrics testing community. We quantify the impact of eye locations on face recognition match scores for identical subject images. The scores are analyzed to better understand the consequences and sensitivity of eye finding for more general applications when eye locations must be determined automatically.

Categories and Subject Descriptors
D.2.8 Metrics (performance metrics)

General Terms
Measurement, Performance, Reliability, Experimentation, Verification.

Keywords
Biometrics, eye detection, face recognition, performance evaluation, synthetic imagery.

1. INTRODUCTION
This paper describes the effects of eye finding accuracy on the scores produced by a representative eigenface-based recognition system. Before enrollment, identification, or verification can be made, an input image must be analyzed or reviewed to determine the positions of both the left and right eyes. The difficulty of this task is influenced by a variety of factors including image size, lighting conditions, pose angle, expression, and occultation. The effectiveness of this process is a necessary precondition for successful template generation and proper matching or rejection of input faces.

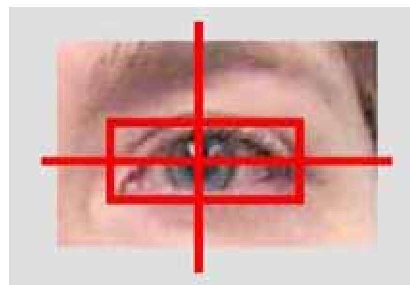

Figure 1 - Eye Position defined by a bounding box[1]

Once eye locations are known, they are used as reference points to normalize the input image to some standardized scale and rotation. The center of the face is then masked to remove non-face regions such as the hair and neckline. Finally, the resulting image is projected into the multidimensional eigenspace[1] . These projections are then subtracted from the original image to produce a residual image that describes the reconstruction error. A high error is generally indicative of poor quality images or highly inaccurate eye locations. In the latter case, eye positions must be specified manually and the process restarted.

A significant amount of effort has been put into making the automatic eye detection process successful. Unfortunately, little is understood about subtle variations in the eye finding process and how these differences affect the ability of eigenface based face recognition scoring.

Many early face recognition evaluations such as the FERET tests contained image metadata specifying eye positions. This assisted in the fair and reproducible testing of face recognition algorithms at a time when face finding, eye detection, and other systems-level concerns were not the primary focus. However, as systems have improved, the variety of tests has increased to encompass subject aging, lighting variation, pose angles, and facial expressions. Now it is standard practice to view the automatic

[1] Image of Eye position bounding box taken from M1/03-0114 standards document.

eye finding algorithms as part of a 'black box' in the overall face recognition process.

Reliance on automatic eye detection has grown considerably since the early days of the FERET tests, when the only reported algorithms used manually specified coordinates [1]. However, we have not located a study on the reproducibility of eye detection and the effects of slight variations in pixel accuracy across samples. This research was motivated in part due to face scoring variations caused by human inaccuracies when manually selecting eye coordinates. Nevertheless, these results are applicable to automated eye detection since those methods are also likely to experience variances that will impact subsequent face recognition performance.

2. COTS SOFTWARE

2.1 Viisage FaceTOOLS™ SDK v2.3

Viisage Technology Inc. is a commercial face recognition product vendor based out of Littleton, MA. The FaceTOOLS™ development kit is based on an eigenface implementation originating at MIT [2]. This software development kit was used to enroll subject images using the programmatic control of varied eye coordinates and to subsequently generate match scores for those enrollments.

2.2 FaceGen Modeller

This software package from Singular Inversions Inc. was used to synthetically generate random faces under a variety of lighting and pose conditions. This experiment includes face data generated using FaceGen version 3.0 to qualify scoring effects for faces that exhibit a highly symmetrical appearance.

The FaceGen application also allows for fine control of facial shape, texture, and expression. However, these were explicitly left in a 'neutral' position for these experiments.

3. TESTING ENVIRONMENT AND SETUP

Experiments were conducted using a common Dell PC running the Windows XP operating system. Microsoft Excel and Perl were used to compile and display statistical results.

After a preliminary investigation, eight images were chosen to form the experimental data set. Seven of these are high-quality color images of MITRE employees and were originally used as the basis for 3D head model generation. The eighth is a synthetically generated color image of an ideal human face/head. It was included in the sample to better interpret the results in the context of a highly symmetrical image. Thumbnails of these images appear in Figure 2.

Figure 2 - Experimental Face Images

Since all of the images were of rather large size, each was rescaled so that the size of the subject's iris would be approximately 18-20 pixels in diameter. Subsequent experiments varied the eye position by translating eye coordinates in concentric circles around their original positions. Since the eyes are of nearly uniform size, the experimental context does not vary across subjects. The dimensions (width x height) of the original and normalized images are given in Table 1. The numbering of the images is from upper left to lower right.

Table 1 - Image Sizes by Subject

Subject	Original	Normalized
1	529x681	371x477
2	2048x1536	490x552
3	807x931	372x429
4	1055x1421	349x469
5	915x1207	366x483
6	1017x1329	367x479
7	1091x1375	328x413
8	400x400	400x400

Image rescaling was also necessary to facilitate the automatic eye detection of the Viisage FaceTOOLS™ software. The original sizes proved too large and defeated the internal algorithms. After resizing, Viisage properly located the eye positions of all the images[2].

The following image is taken from the synthetically generated face. The area of the highlighted circle in Figure 3 represents the approximate working region for the following experiments. The exact position varies depending on the eye centers selected by the Viisage FaceTOOLS™ software.

[2] The accuracy of these initial eye positions was not evaluated according to any standard. Since they satisfied a visual quality check, they were merely used as a reference.

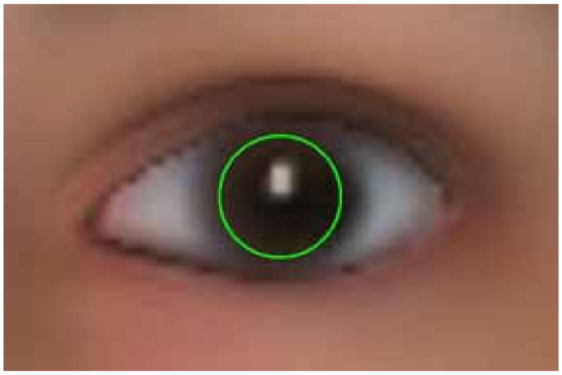

Figure 3 - Synthetic Eye Closeup

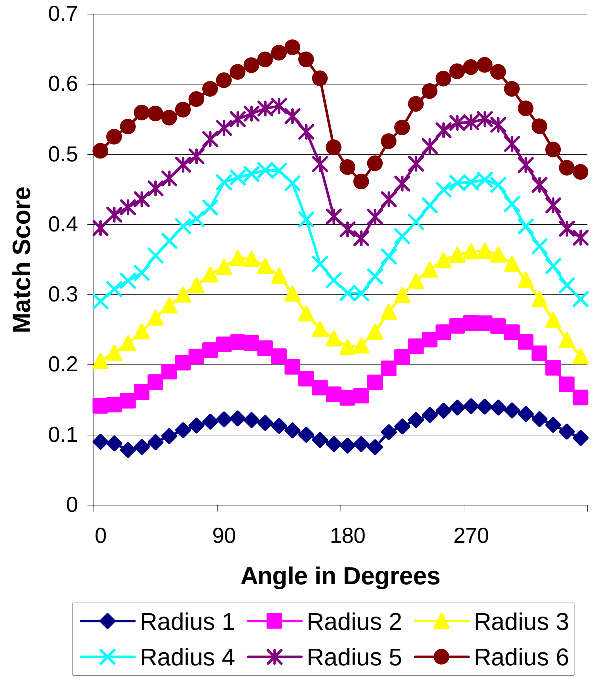

Figure 4 - Score Performance for Subject #1

The enrollment software was specially modified to automatically enroll subject images multiple times with precisely controlled eye positions. The first enrollment for any given subject used the original, reference eye coordinates. Subsequent enrollments used modifications of these positions and were later compared against the first enrollment to assess score variation.

During the enrollment process, the locations of the eyes were stored as additional attributes in the database. This information was reported by the identification engine when matches were subsequently performed. Post processing of the results data was done using Perl and the results displayed using Microsoft Office Web Components.

4. DEVIATIONS OF BOTH EYES

For this experiment, the positions of both eyes were simultaneously modified to study the effects of translation on the eigenface algorithms. The Viisage database was cleared of all prior enrollments before testing.

As each subject was enrolled into the database, their left and right eye positions were translated in concentric circles about the original coordinates. The translation radius varied from 1 to 6 pixels and the angle from 0 to 350 degrees in 10 degree increments. Hence, for each subject, there were 217 different templates created ((36 * 6) + 1), with the first representing the original eye positions and used as a reference for subsequent match score calculation. The actual image data for each subject remained constant during the enrollment process.

The images were rescaled so that the iris dimensions were larger than the controlled variance. Therefore, the new eye coordinates remained within the pupil/iris region.

Figure 4 shows how progressively larger radii affected score performance for Subject #1. Each point represents the score using modified eye coordinates versus the unmodified enrollment. The score trend within each radius appears to be due to the natural left/right symmetry of the face. Scores at 0 degrees (subject's left) and 180 degrees (subject's right) are at local minima. This suggests that the eigenface algorithm is more sensitive to eye positions that deviate above or below the enrolled reference.

The synthetically generated face was specifically chosen to assess the effects of symmetry on scoring results. The rendering of this image was made using ideal frontal lighting and no texture model was applied to the skin. Although it was not possible to explicitly specify a perfectly facing pose angle, one was approximated by trial and error.

Figure 5 is a radar plot of scores with 0 degrees at the top. Note the relative symmetry of the plot about the 90 and 270 degree positions. The same pattern seen earlier with the Subject #1 is now more apparent. When the new eye positions are above the reference position, the match score increases considerably even for smaller radii. Eye positions below the reference also score higher. Positions to the left and right (0 and 270 degrees) have a less significant impact on scoring.

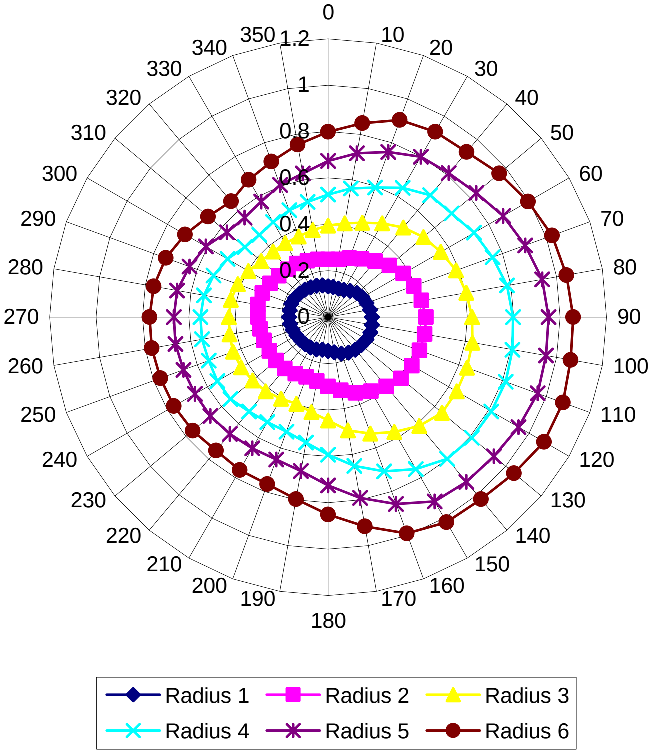

Figure 5 - Radar Plot of Scores for Synthetic Face

The eigenface implementation used by Viisage uses two points to align the input face prior to processing. Therefore, only translation, uniform scaling, and rotation can be compensated. If the input subject's facial shape is not adequately predicted from eye locations alone, there may not be a good alignment with lower facial features such as the lips.

Figure 6 shows a reconstruction image taken from the enrollment of a blank (white) image[3]. This provides an interpretation of the eigenspaces used by the Viisage FaceTOOLS™ SDK. Note that the eye and nose regions are relatively detailed, but that the mouth is rather indistinct. The interplay of the input image with this mouth region can account for much of the asymmetry in the radar plot.

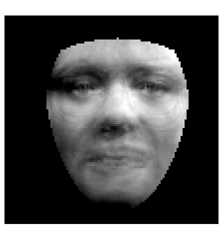

Figure 6 - Viisage Reconstruction Image

[3] A wide range of reasonably spaced eye positions, selected near the vertical center of the white image, produced this reconstruction.

Figure 7 shows the results of averaging the scores from all eight subjects. The symmetry of the plot shows the relative sensitivity of horizontal and vertical position changes. Within a given radius, scores are similar for eye locations that have been varied left or right. Score sensitivity is greater for lower eye locations, but is greatest for eye locations above the reference enrollment.

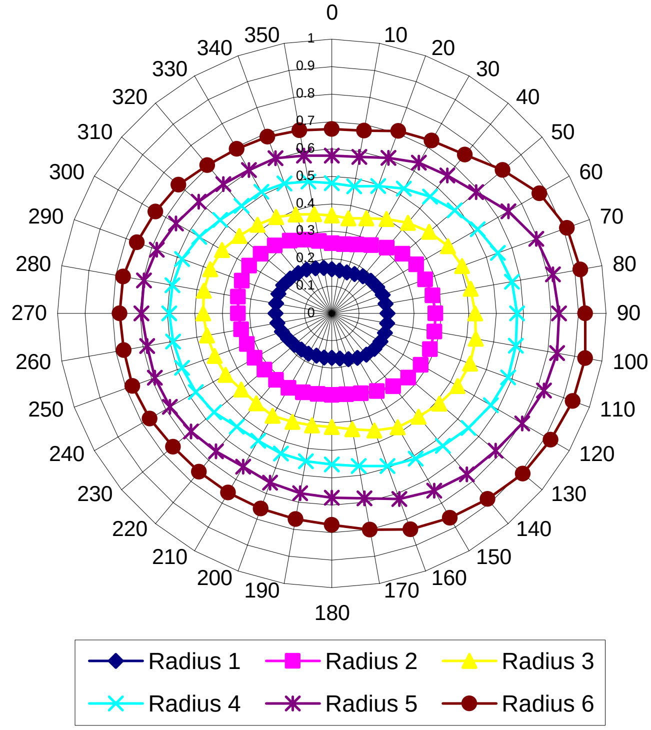

Figure 7 - Averages of all 8 Images

Mean and standard deviation values for a given pixel radius are shown in Figure 8. There is a linear trend in score values within the tested range. Standard deviations increase along a more limited slope. Previous experiments using the Viisage FaceTOOLS™ SDK have revealed that a score of 1.4 is an effective upper value subject mismatch. It is reasonable to conclude that the standard deviation values will ultimately reach a limit as our scores near this bounded 1.4 value. Eventually, the high scores produced by the face recognition engine cease to grow worse in magnitude as the radius increases.

Mean and standard deviation values for each subject *and* pixel radius are shown in Figure 9. Surprisingly, the poorly illuminated Subject #1 exhibits the best scores of the group. Subject #4, though well illuminated and oriented, shows the highest scores and standard deviations. The synthetic image also has slightly higher standard deviations.

A visual review of the subjects suggests that images of uniform lighting and greater symmetry can generate higher score variations.

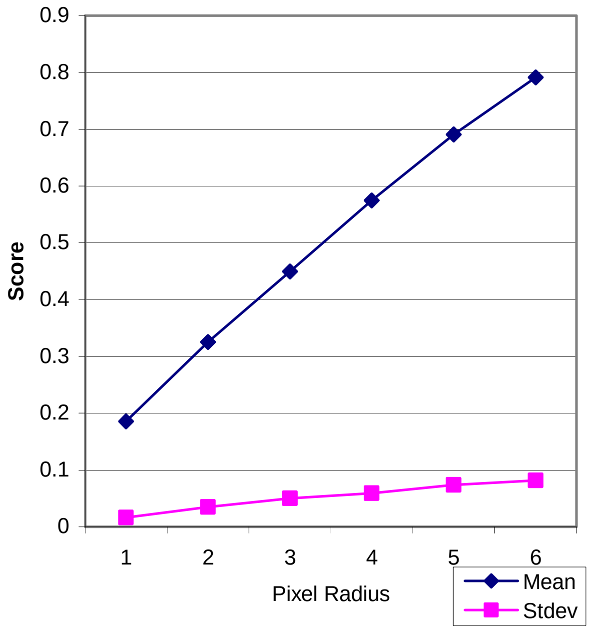

Figure 8 - Mean and Standard Deviation for Figure 7 Data

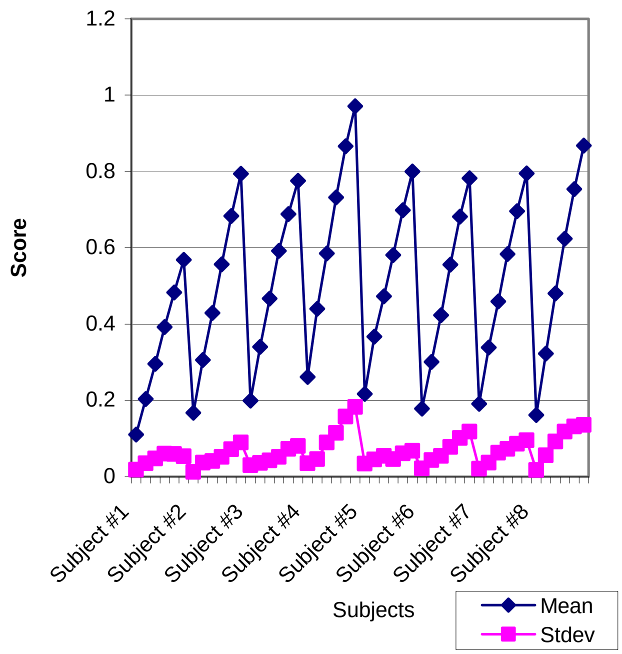

Figure 9 - Mean and Standard Deviation by Subject and Pixel Radius

5. DEVIATIONS OF THE LEFT EYE

This experiment is substantially similar to the dual eye experiment, but in this case, only the position of the left eye was modified. By keeping the right eye coordinates constant, we introduce rotation and scaling effects into the image alignment process. Again, the Viisage database was cleared of all prior enrollments before testing.

Figure 10 shows the alignment and reconstruction images from the Viisage FaceTOOLS™ user interface. The top row corresponds to the primary enrollment using centered eye coordinates. The alignment image (right) shows the normalized image after scaling and rotation. The quality of the eye positions is apparent in the un-rotated orientation of the image. The reconstruction image (left) shows the eigenspace representation of the face. The quality of the reconstruction can be ascertained by visual comparison to the alignment image. The bottom row illustrates the effects of moving the left eye coordinate down by 6 pixels (270 degrees). The alignment image is now noticeably rotated, and the reconstruction image has sufficient error to be visibly different.

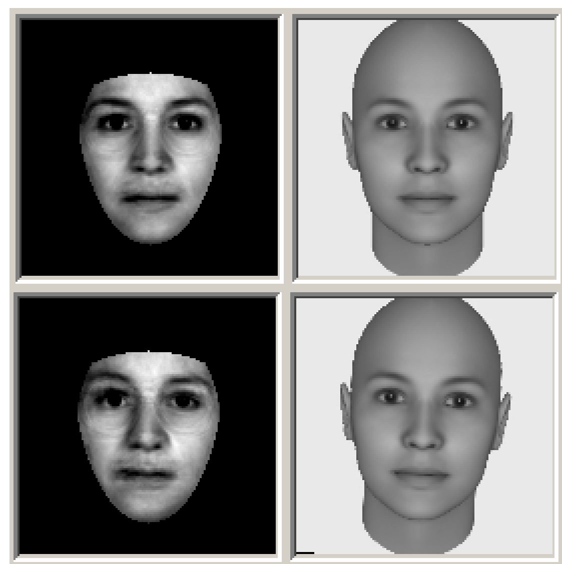

Figure 10 - Reconstruction and Alignment Images

Figure 11 shows the resulting average scores by pixel radius and angle. Scores improved slightly as compared with the dual eye experiment. Although scaling and rotation have been introduced, the constancy of the right eye has mitigated those factors. The up/down and left/right symmetry effects observed in the dual eye experiment have also decreased in magnitude.

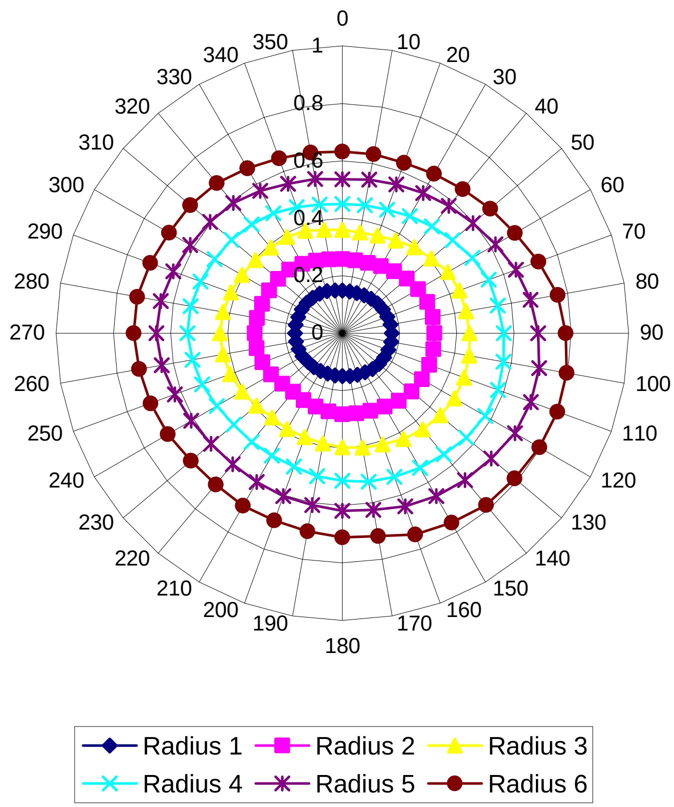

Figure 11 – Averages for all 8 Images (Left Eye)

Figure 12 shows the means and standard deviations by pixel radius. Matches have uniformly improved, with a maximum score of 0.7 now replacing the 0.8 value observed in the dual eye experiment. Standard deviations have also decreased, thus confirming the greater uniformity of score values across angle variations.

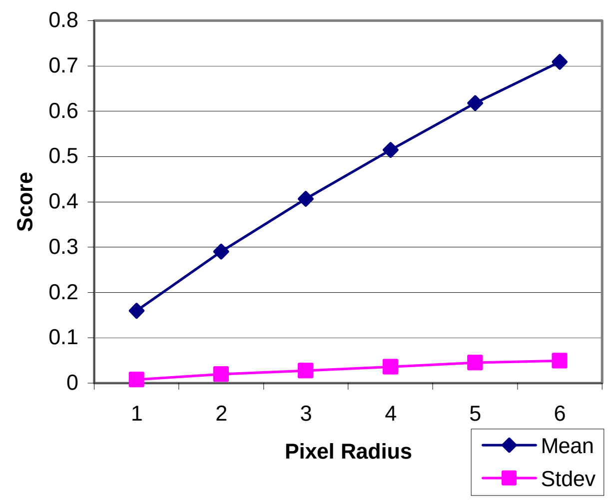

Figure 12 - Mean and Standard Deviation for Figure 11 Data

6. ANALYSIS

The scoring variations observed in the preceding experiments clearly illustrate the sensitivity of eigenface-based recognition to eye location deviations. During previous experiments conducted at MITRE, Viisage recommended that the scoring threshold be reset to 0.75 from the 0.5 default setting. Increasing the threshold lowers the false rejection rate (FRR) but only at the expense of increasing the potential false acceptance rate (FAR). Figure 8 shows that the 0.75 threshold may yield false rejections for eye coordinate deviations as small as 5 pixels. This is a significant influence for a single, parametrically controlled variable.

In a real world situation, not only would eye coordinates differ by some distribution, but pose angle, lighting variations, image size, et. al. would each contribute to the match accuracy.

Emerging standards by Griffin [4] suggest that computer generated eye positions be reviewed by a human operator, but simultaneously recommend manual alignment regardless of the computer's choice. Both Griffin and Viisage define the eye location as the geometric center of the entire eye as bounded by the eyelids and corners. Unfortunately, this point does not necessarily correspond to the pupil center.

Relying on a human operator to consistently select the same region under varying conditions of gaze, squint, and eyelash occlusion is a risky. There may be a strong tendency for the operator to select the pupil center, since this is an easier task. Regardless, human factors studies such as [5] have shown that relatively simple mouse tasks can have high spatial error rates. So, even assuming that the meaning of "eye position" is fully explained to an operator, the repetitive nature of the task and the inherent usability characteristics of a mouse and display terminal will result in an unwanted distribution of selected points.

Some operational scenarios may even mix both human and computer methods. Subject enrollment may be conducted under ideal conditions of lighting and pose with a human operator to manually select the eyes. Subsequent identification or verification could be entirely computer controlled, with automatic eye detection algorithms calculating the eye positions. Any disagreement between methods would impact performance.

7. FUTURE DIRECTIONS

Reproducible eye detection and location is critical for consistent face recognition, yet this step remains poorly qualified and in need of further study. In this paper, we have demonstrated that eigenface approaches suffer degradation when eye locations cannot be precisely determined. Other face recognition techniques may be more or less susceptible to variation.

Controlling and quantifying the eye detection process is a critical first step in understanding the relative merits of differing face recognition systems. Some approaches may be dependent on eye position, while others may be tuned to tolerate variation. As long as the eye detection step is treated as an inseparable component in the recognition process, it will not be clear which algorithms exhibit robust performance for eye location variance. A face recognition engine marketed as “using an advanced, patented neural net approach”, may simply be superior due to repeatable eye detection front-end. Face recognition evaluations such as [6] have studied the effects of age, gender, and lighting on performance, but there may be subtle, undiscovered relationships between those factors and eye detection accuracy.

Synthetically generated imagery permits the parametric testing of varied lighting, pose, scale, rotation, and detail with well-known eye positions. Since the face and eyes are rendered from 3D representations, the exact centers can be consistently expressed independently of other variations.

In addition, the synthetic corpus permits a qualitative assessment of human accuracy when the eye detection process is performed manually. There are no known studies detailing the ability of human subjects to perform complex, repetitive image tasks requiring consistent object selection.

Eye detection algorithms can also be improved by using the corpus as a series of input images. The actual eye locations known for each image facilitate a relative comparison of method effectiveness. The corpus will also contain metadata regarding lighting conditions and pose angles so the limits of each algorithm can be studied.

8. ACKNOWLEDGMENTS

We would like to thank Rick Chavez at the MITRE Corporation for his earlier work with the FaceTOOLS™ SDK scoring utilities. We also express our thanks to Viisage Technology Inc. for their support for this and other experiments.

9. REFERENCES

[1] Viisage FaceTOOLS SDK Users Guide Version 2.3, October 2002

[2] Rizvi, S., Phillips, P. & Moon, H., “The FERET Verification Testing Protocol for Face Recognition Algorithms”, NIST Technical Report 6281

[3] http://www.viisage.com/facepassbrochure.pdf

[4] Griffin, P (ed.). “Face Recognition Format for Data Interchange M1/01-0114”, March 12, 2003, p. 23

[5] Akamatsu, A., and MacKenzie, I. S., “Movement characteristics using a mouse with tactile and force feedback”, International Journal of Human-Computer Studies, Vol 45 (1996), 483-493

[6] D. Blackburn, M. Bone, and Dr. P.J. Phillips, 2000, “Face Recognition Vendor Test 2000 Evaluation Report”, DoD Counterdrug, DARPA, and NAVSEA, February 16, 2001.