Assessing Face Overlay

Mary Theofanos
Brian Stanton
Charles Sheppard
Ross Micheals
Yee-Yin Choong
John Wydler
Kevin Mangold
Michelle Potts Steves
Emile Morse

Information Access Division
Information Technology Laboratory

May 2008

1. INTRODUCTION

In 2003, the Department of Homeland Security deployed cameras to capture facial images of persons passing through the primary and secondary inspection processes for U.S. ports of entry. A quality assessment of the facial images being captured at the airport ports of entry was performed in 2004 and updated in 2008 [1], [2]. This assessment found that the images being captured were not suitable for automated facial recognition, and would not usefully augment the fingerprints for the Visitor and Immigrant Status Indicator Technology's[1] (US-VISIT) identity management system. As the result of this assessment, US-VISIT embarked on an effort to improve the quality of their captured facial images.

One aspect of this effort was the identification of usability and human factors issues that may impact face image capture. The National Institute of Standards and Technology's (NIST) usability and biometrics team was asked to identify any usability and human factors considerations that might improve the capture of face images at the airports. The NIST team reported in [3] targeted usability and human factor enhancements to improve capturing acceptable images.

Implementing these enhancements resulted in:
1. 100 % of the images captured a participant's face, in contrast to the current US-VISIT collection
2. At image capture, all of the participants were facing the camera, so a frontal face image was obtained – this process change resulted in a significant increase in image appropriateness for face matching use.

Further, the study[2] [3] postulated that additional image quality improvement may be realized by using a face overlay guide for the camera operator to help align the camera. The remainder of this report describes the laboratory-based, proof-of-concept study that assessed this feature of image capture and its effect. Particularly the study addressed the question of whether participants (acting as operators) could use the face overlay guide when taking a

---

[1] The U.S. Department of Homeland Security's US-VISIT program provides visa-issuing posts and ports of entry with the biometric technology that enables the U.S. Government to establish and verify the identity of people visiting the United States.

[2] These tests were supported by the Department of Homeland Security. Specific hardware and software products identified in this report were used in order to perform the evaluations described. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

facial photograph to effectively center the face in the image as efficiently as when not using the guide. Image quality, e.g., face centered-ness, efficiency (time to position the camera and capture images), user-satisfaction, and affordance of the overlay are reported.

2. BACKGROUND

2.1 PRIOR WORK

The NIST team reported in [3] on the following five usability and human factors enhancements to improve capturing acceptable images.
1. The camera should resemble a traditional camera.
2. The camera should click when the picture is taken to provide feedback to the traveler of the process.
3. The camera should be used in portrait mode.
4. The camera operator should be facing the traveler and the monitor while positioning the camera.
5. There should be markings on the floor, such as footprints, to indicate to the traveler where to stand for the photograph.

Implementing these enhancements resulted in:
1. 100 % of the images captured a participant's face, in contrast to the current US-VISIT collection where 5 % of the images have some part of the face cropped out of the picture and approximately 70 % of the images had a pose angle of greater than 10° indicating that the subject was frontal to the camera in only about 5 % of images
2. At image capture, all of the participants were facing the camera, so a frontal face image was obtained -- this process change resulted in a significant increase in image appropriateness for face matching use.

Previous analysis of the face image collection held by US-VISIT conducted in [1], showed that geometric problems (in order: pose, size, cropping, etc.) supported the postulation in [3] that additional image quality improvement may be realized by using a camera usability alignment feature. Although the NIST usability and biometrics team had developed a face overlay diagram to assist in analyzing images in [3], they suspected that such a face overlay guide could be used by the camera operator to help align the camera during image capture. By incorporating the overlay into the workstations, the officers could use the guide to center the camera on the participant's face effectively and efficiently. However, a standing requirement of the US-VISIT program was that additional training for station operators was not acceptable. The goals of the study described in this report were to show the effect of the face-overlay during image capture on:

1. image quality,

2. efficiency (time required to capture the image), and
3. training requirements, would additional training be needed to effectively use the overlay.

2.2 AFFORDANCE

To address the requirement that no additional training could be imposed to use the face overlay effectively, the NIST usability team turned to the concept of affordance. This concept was originally introduced by psychologist James J. Gibson in his 1977 article "The Theory of Affordances"[6]. Donald Norman applied the term to human computer interaction in his book *The Design of Everyday Things* [7] in 1988. According to Norman an affordance is the design aspect of an object which suggests how the object should be used; a visual clue to its function and use. Norman writes:

> "...the term affordance refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used. [...] Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. Slots are for inserting things into. Balls are for throwing or bouncing. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction needed." (Norman 1988, p.9)

The study design was constructed to allow affordance of the overlay to be examined in this study, as well, as traditional assessments of effectiveness, efficiency, and user satisfaction.

2.3 FACE OVERLAY

A face overlay diagram, shown in Figure 1, was designed according to the ANSI INCITS 385-2004 Standard [4] and ISO/IEC 19794-5:2005 [5]. These standards indicate that the approximate horizontal midpoints of the mouth and of the bridge of the nose shall lie on an imaginary vertical line at the horizontal center of the image. The upper tick-mark represents the ideal height of the crown of the head and the distance from the edge of the picture. The lower tick-mark represents the ideal position for the base of the shoulder-line. A horizontal line passes through the center of both eyes of an individual's face image and a horizontal midpoint of the bridge of the nose with the horizontal center of the image.
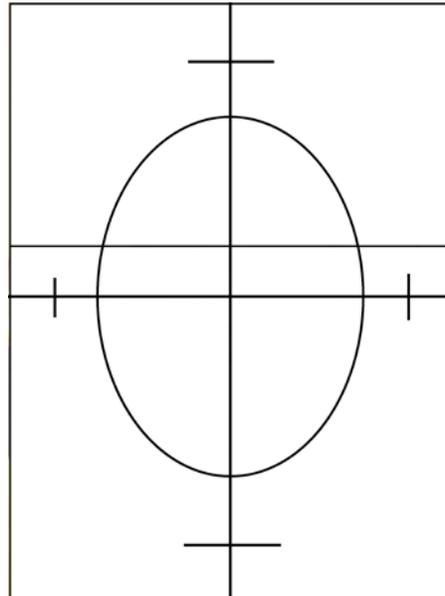
Figure 1: Face overlay

3. METHOD

3.1 SET-UP

A Logitech Quickcam Pro 5000 webcam was mounted on a tripod and placed on a table. The camera could be panned right and left and tilted up and down. The Quickcam captured images at 640 pixels wide by 480 pixels high. The Quickcam images were displayed on the computer monitor to the right of the tripod and camera. An Optimus keyboard model mini three from Art Levedev Studio, consisting of three 4mm X 4mm programmable liquid crystal display (LDC) pushbuttons was positioned in front of the monitor for participants to use to initiate the capture of an image.

The physical layout of the face capture station is illustrated in Figure 2. The tripod was secured to the table 49.5 cm (19.5 in) from the table's back edge. The subjects of the photograph were a mannequin or a NIST researcher posing as a model. The subjects were positioned 45.7 cm (18 in) from the back edge of the table (1/2 the total lane width at a representative POE processing center) and 104.1 cm (3 ft 5 in) left or right from the webcam. Additionally, the photographic subjects were positioned on an adjustable height table such that the photographed heights would be 157.5 cm (5 ft 2 in) or 193 cm (6 ft 4 in). This produced four subject positions. The left and right offset positions provided the extreme representations of presenter positioning at a processing counter. The two heights, the 5th

percentile female and 95th percentile male, respectively, were chosen as they align with the endpoints of the design specification range for traveler height.

The mannequin was positioned on the table so that the eyes were always facing the camera. The NIST model was positioned on the table and was instructed not to look at the camera.
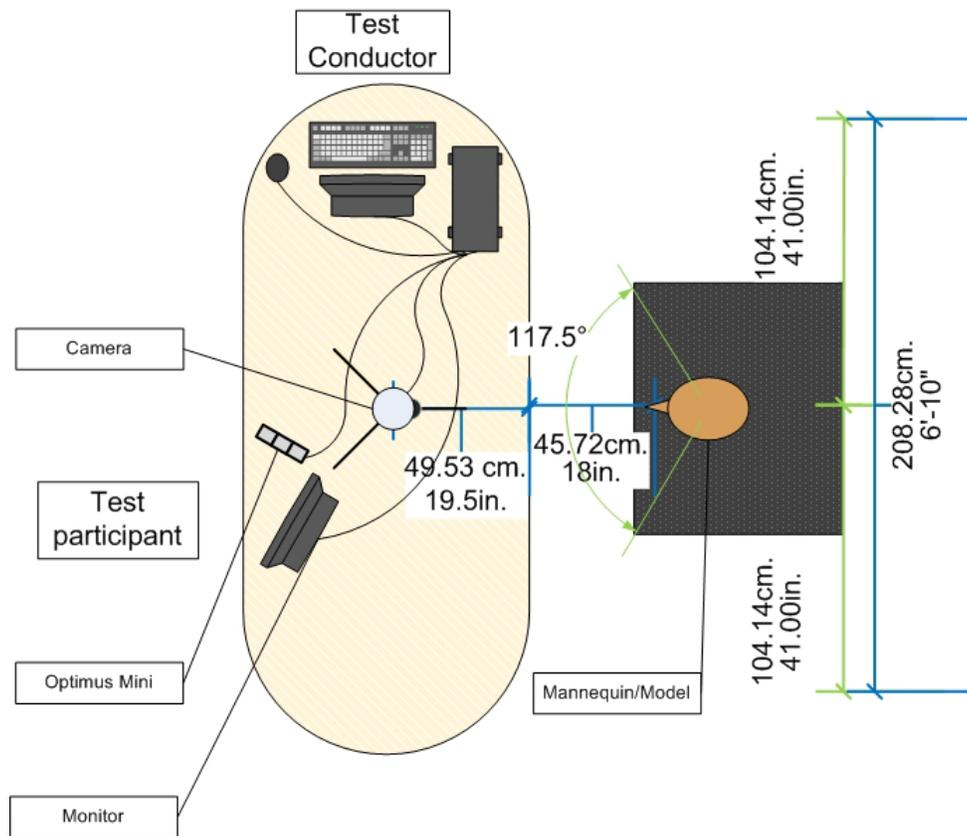
Figure 2: Face overlay test layout

3.2 PROCEDURE

Forty-one NIST employees participated in the study. Employees who characterized themselves as photographers did not participate. Each participant was asked to take four pictures of a subject. Participants were instructed to "take the best passport picture in the shortest amount of time". They were informed that they could swivel the camera right and left and tilt it up and down, but not move its location. They could also request that the subject of the picture face the camera. For each of the four pictures the participants were told when they could start taking the picture.

Twenty of the participants took pictures of a mannequin, the remaining 21 took pictures of a NIST researcher as a model. Within each of these conditions half were provided the face overlay (Figure 1) within the displayed image and half did not see the overlay. There was no mention of the overlay to the participants. The presentation order of the four positions (right of camera at the two heights and left of the camera at the two heights) were counterbalanced to address order affects.

For each photograph, the facilitator performed the following:
1. Moved the table to the right or left position
2. Set the camera to the starting position (centered)
3. Adjusted the table height
4. Asked the participant if he/she were ready
5. Upon confirmation, started the software to record the session
6. Immediately after the picture was taken, stopped the session.

4. RESULTS

4.1 AFFORDANCE

As indicated in the previous section, none of the participants received any explanation or instructions about the overlay, yet all of the participants who saw the overlay knew exactly how to use it. Each positioned the camera such that the overlay framed the subject's face and used the horizontal and vertical lines to align the eyes and the nose as in Figure 3. None of the participants asked questions of the facilitators concerning the overlay or were confused by the overlay. All used the tool that was provided to assist in positioning the camera.

In a survey after the participants had completed capturing the images, participants were asked:
1. How did you decide when to take the picture?

2. How did you decide when the picture was good enough to take?

Responses included "when the eyes lined up with the overlay" or "centered within the overlay" and "head was completely inside the oval". All the participants who used the oval made some comment about the head or face within the oval.

The affordance of the overlay was excellent – each user knew how to use it without any instruction.   Participant comments included "it was clear what to do" and "it explained everything by itself".
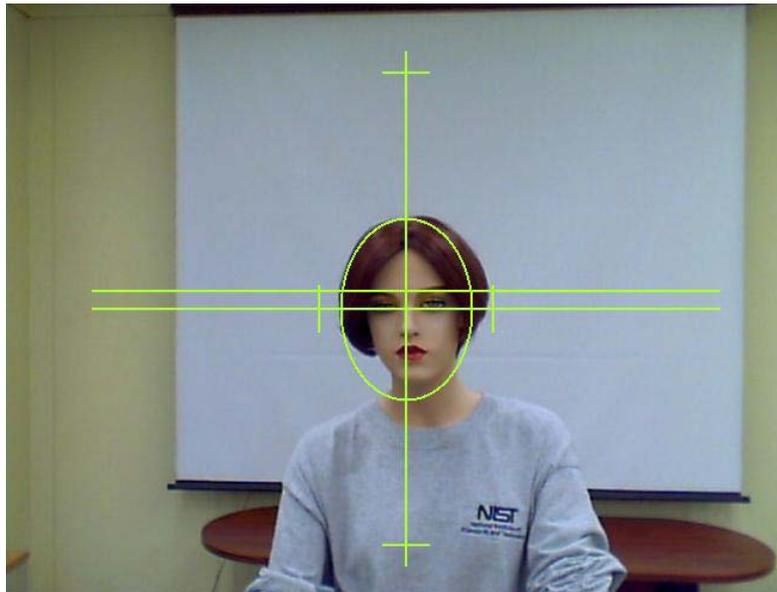
**Figure 3: Use of the overlay to frame the face in the image**

4.2 QUALITY

We analyzed quality by dividing the photographs into quadrants using the overlay. For each photograph we identified whether the face image was centered in the x and y axes, which quadrant (1 to 4) the image appeared or if the image appeared on one of the axes (A, B, C, or D). Figure 4 illustrates the positions that were identified.

Four judges were used to rate each of the 164 collected images. (We report on the analysis of 160 images since one participant's data (4 images) was eliminated because they received incomplete instructions.) The judges were instructed to use the following rules to assign codes to each image:

1) Is the subject (either mannequin or NIST model) facing the camera (both eyes are visible)? If not, code as 'Non-frontal view'.

2) Are the eyes touching any part of the space enclosed by the two parallel horizontal lines?
3) Is any part of the nose on the vertical axis?
4) If answer to (2) and (3) are 'yes', then code image as 'centered'. Figure 3 is an example.
5) If the answer to (2) is 'yes' and to (3) is 'no', code the direction along the horizontal as 'B' or 'D' as appropriate. Figure 5 is an example.
6) If the answer to (2) is 'no' and to (3) is 'yes', then code the displacement along the AC axis as appropriate.
7) All remaining images can be categorized into quadrants 1-4 depending on the shift directions noted in (5) and (6). Figure 6 is an example.

.
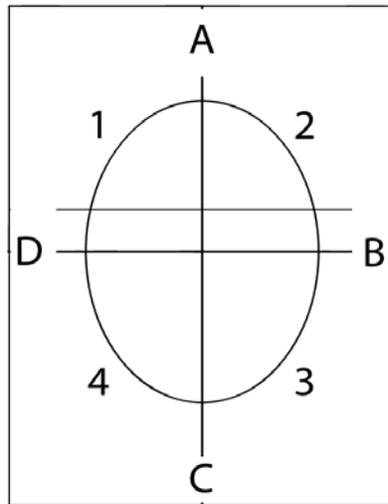
Figure 4: Positions for Measuring Face Placement

.
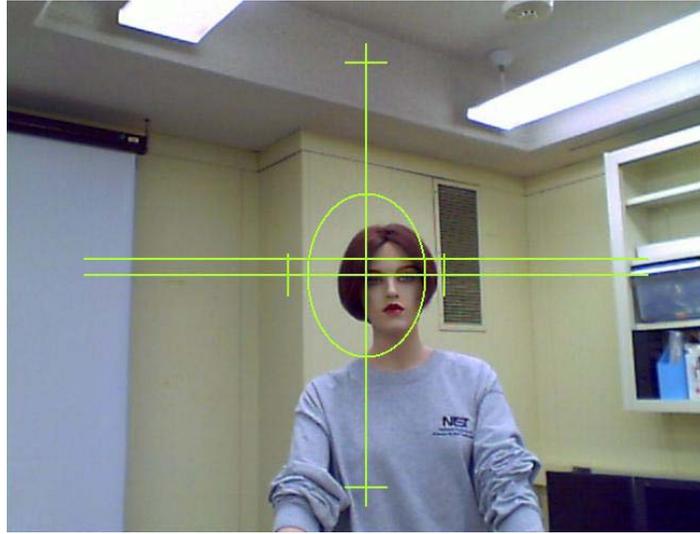
Figure 5: Example of displacement on axis "B"
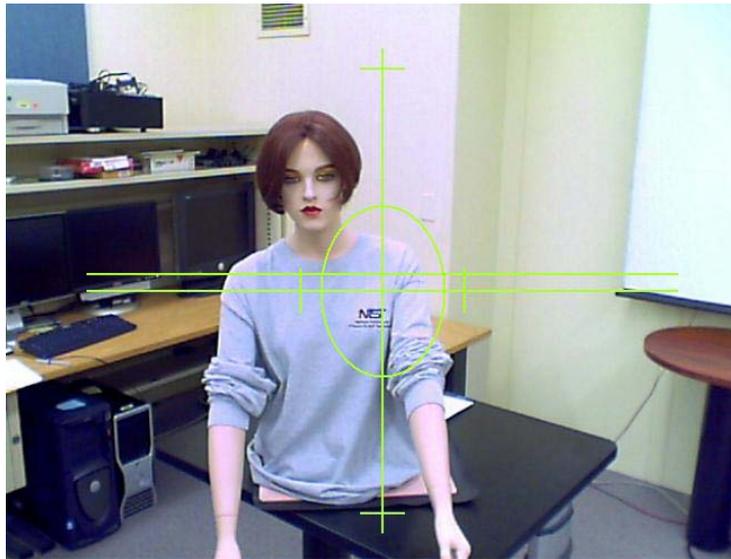
Figure 6: Example of displacement in Quadrant 1

We used judges' ratings since it was not clear before the judging that any quantitative method could be used. However, after the coding was complete, there was a consensus among the judges that the coding scheme did not capture the fact that some images were obviously more off target than others even though they ended up with the same code. Judges

agreed that a point on the center line of the face (i.e. middle of the nose) defined the horizontal center and a point fairly equidistant from the bridge of the nose and the tip of the nose defined the vertical center for both the mannequin and the NIST model. For those images that were frontal (i.e. 141 out of 160), the images were reanalyzed by a single person in order to measure the displacement in pixels of each image's center point from the standard measurement overlay center. Images were viewed using GIMP (Gnu Image Manipulation Program) and the caliper measurement tool was applied. Values for straight line distance were recorded as were the vertical and horizontal components.

Table 1: Displacement of image center from target center (mean $\pm$ SEM)

	No overlay n = 71	With Overlay n = 70
Displacement (pixels)	85 $\pm$ 5	12 $\pm$ 1*
Horizontal distance (pixels)	-7 $\pm$ 5	2.4 $\pm$ 1
Vertical distance (pixels)	70 $\pm$ 6	-4 $\pm$ 1*

* Difference between overlay and no overlay is significant: $p < 0.01$

The data in Table 1 shows that when an overlay was not used the overall displacement of the image center was significantly greater than when an overlay was used. The data also shows that users were very likely to get the vertical displacement very wrong, but they were equally likely to move the image off-center with respect to the horizontal. Figure 7 provides a visual representation of the data in Table 1. Figure 7 clearly shows that the use of the overlay for both the photographs of the mannequin and those of the NIST model, resulted in images that were centered.
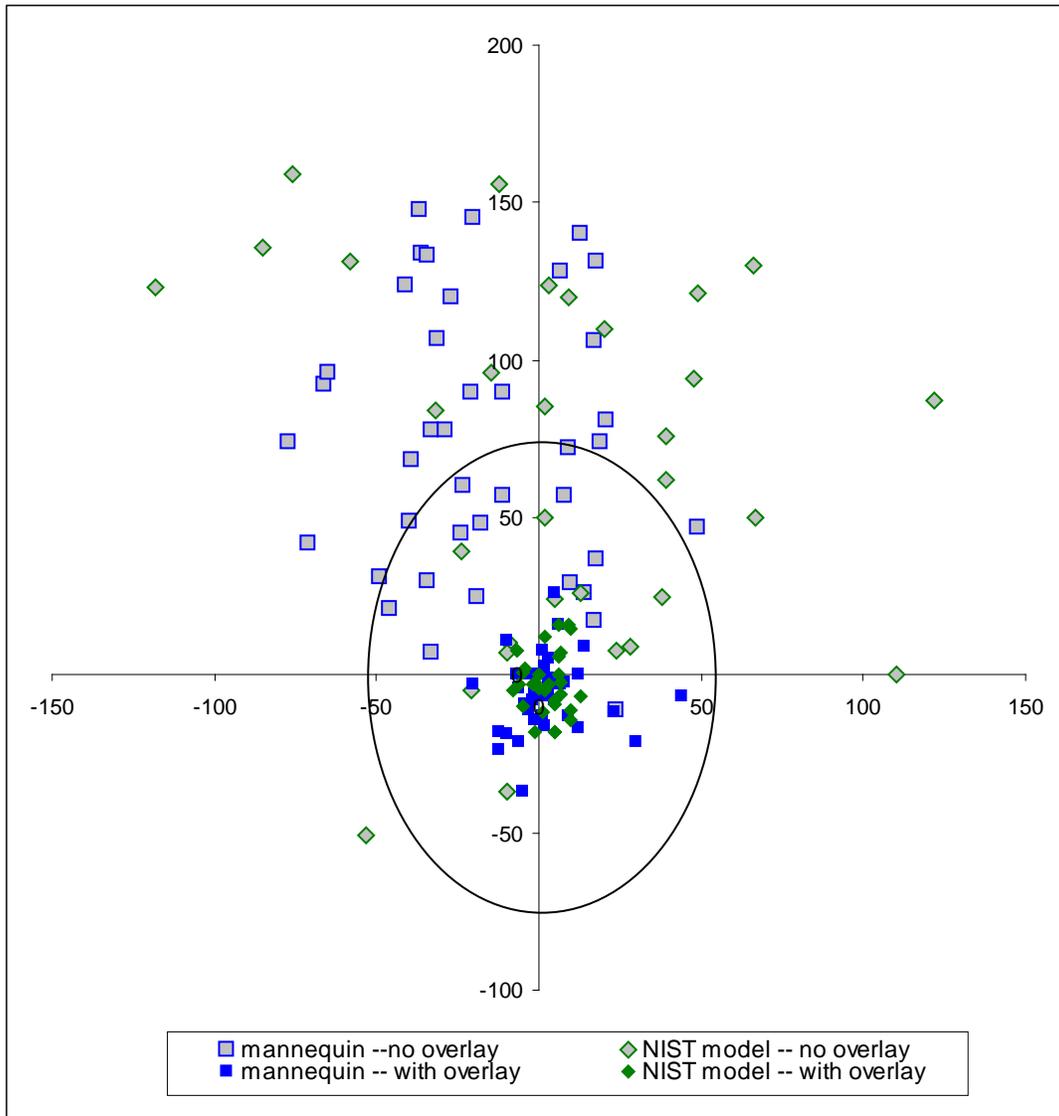
Figure 7: Displacement of image center from target center

The results of this analysis for the mannequin and the NIST model are presented in Table 2 and Table 3. The inter-rater reliability of the judges' ratings was measured using the Fleiss kappa statistic. For the full set of images, the kappa was 0.75. For images that were frontal (i.e. 141 out of 160), agreement achieved a kappa of 0.81. These values according to published criteria show substantial and 'almost perfect' agreement, respectively.

Table 2: Positions of Photographs of NIST Model

Condition	Centered	A	B	C	D	Quad 1	Quad 2	Quad 3	Quad 4
Non overlay	1.0	4.5	1.0	0.5	2.3	11.8	17.3	1.0	0.8
Overlay	20.3	2.5	3.0	6.0	3.5	0.5	2.8	0.8	0.8

Table 3: Positions of the Photographs of Mannequin

Condition	Centered	A	B	C	D	Quad 1	Quad 2	Quad 3	Quad 4
Non overlay	0.0	1.3	0.8	0.0	1.0	24.5	12.0	0.5	0.0
Overlay	19.3	0.8	6.0	4.5	2.5	0.3	0.3	3.0	3.5

Table 4 summarizes the data presented above, without consideration for the photographic subject, as this factor did not affect the outcome of centered-ness. The majority of the participants asked the model to face the camera. Of the 20 participants two did not ask the model to face the camera for any of the four images taken. Only 1 other image was captured without asking the model to face the camera. Thus, nine of the images were categorized as non-frontal. The results show that use of the overlay was much more effective. The use of the overlay for both conditions resulted in images that were centered. Without the overlay only one image was centered, a success rate of 1.4 %. Using the overlay resulted in 100 % of the images appearing within the oval, 53.2 % of the images were perfectly centered.

Table 4: Centeredness - Data Summary

Condition	Images collected with frontal orientation	Resulting centered-ness	Result (count)	Results (%)
No overlay	71	Not centered	70	98.6 %
		Centered	1	1.4 %
Overlay	71	Not centered	32.3	45.4 %
		Centered	37.8	53.2 %

4.3 EFFICIENCY

We measured efficiency as the time required to complete a task, where a task is defined as taking a photograph. When the participant confirmed he was ready, the facilitator initiated the task by starting the software and a timestamp was automatically recorded. Immediately, the participant's monitor displayed the live image from the camera. When the participant pushed the button to take the picture a timestamp was recorded and the session was ended indicating the end of the task.

Table 5 provides the times in seconds when the subject of the photograph was the NIST model for both the overlay and the non-overlay condition. Table 6 provides the times in seconds when the subject of the photograph was the mannequin for both the overlay and the non-overlay condition. A Kruskal-Wallis test was used to test the null hypothesis that the medians of the overlay and non-overlay conditions are the same. The P-values were greater than 0.05, for both the NIST model (p= 0.82) and the mannequin (p= 0.17). Thus, we detected no statistically significant difference in time at the 95 % confidence level.

Table 5: Summary Statistics for Time (Seconds) (NIST Model)

Overlay	Count	Median	Standard deviation	Range
Not present	40	9.3	3.3	4.5 to 15.9
Present	44	9.3	5.4	4.3 to 33.5
Total	84	9.3	4.5	4.3 to 33.5

Table 6: Summary Statistics for Time (Seconds) (Mannequin)

Overlay	Count	Median	Standard deviation	Range
Not present	40	6.2	3.1	3.7 to 22.2
Present	40	5.5	2.7	3.2 to 16.7
Total	80	6.1	2.9	3.2 to 22.2

The difference in the times for the mannequin condition and the model condition is likely due to the interaction of the participant with the model. The majority of the participants asked the model to face the camera. Of the 20 participants two did not ask the model to face the camera for any of the four images taken. Only one other image was captured without asking the model to face the camera.

4.4 USER SATISFACTION

Each participant was given a satisfaction survey after completing the test. The questions included:
1. How did you decide when to take the picture?
2. How did you decide when the picture was good enough to take?
3. What did you think about the process of taking the picture?
4. Do you have any suggestions on how we can improve the process?

All participants who had the overlay used it for subject framing. Participants who did not have the overlay had various strategies for deciding when to take a picture, e.g., the picture was framed appropriately. Strategies included the following:
- "when it was framed/centered and face slightly above centered"
- "head was centered reasonably and in top 1/4 of frame"
- "when it was centered (horizontally)"
- "facing camera and body squared"
- "when his whole body was in frame"

Participants, whether they had the overlay or not, typically felt the process of taking the picture was easy, although, the mechanics of moving the camera to frame the subject took some acclimation.

Some participants who did not have the overlay expressed a desire to have more guidance on how to frame the picture. Comments included:
- [provide an] "overlay to center on like automatic photobooths for passports (France)."
- "Training on what is a good picture"

For those who received the overlay, two additional questions were included:
1. What did you think about the overlay?
2. Did the overlay help or hinder you taking the picture?

Every participant in the overlay condition believed the overlay was helpful in taking the picture and did not in anyway hinder the process.

4.5 DISCUSSION

Of the participants who had the overlay, all used it to frame the model's face. All of the pictures captured by participants using the overlay appeared within the oval and 53.2 % were perfectly centered within the frame. Conversely, participants who did not have the overlay had a variety of strategies of how to achieve an appropriate framing and only 1 image or

1.4 %, had a centered facial image capture. Additionally, these participants were less satisfied with the process, as they asked for training and guidance on how to frame the model appropriately, while the participants with the overlay were satisfied with the process and knew how to frame the model effectively without instruction.

Use of the overlay in this study showed that it could be used to frame a face with a camera without additional instruction with an effectiveness rate that was clearly superior to the images captured without benefit of the overlay. Additionally, there was no effect on task efficiency, and users who used the overlay were more satisfied with the image framing process than those who did not use the overlay.

5. CONCLUSIONS

This report describes a follow-up study to NIST-IR 7540 Assessing Face Acquisition [3] that incorporates the face overlay into the face image capture process. This study tested if participants could use the face overlay guide when taking a face photograph to center the camera on the face. We found four main results.

1. Affordance: The face overlay had excellent affordance. It was easy to use without instruction or training. Its use was obvious to the participants in this study.
2. Efficiency: There was no significant difference in the time required to capture the face image between those participants that used the guide and those that did not use the guide. Therefore, use of the face overlay was not shown to impact efficiency of the image framing and capture task.
3. Effectiveness: 53.2 % of the images that were taken with the overlay were perfectly centered in the frame and the remaining 45.4 % were at least partially within the oval. Those taken without the benefit of the overlay had a 1.4 % success rate of being centered within the frame.
4. User satisfaction: Users who had the benefit of the overlay expressed satisfaction with knowing when the framing was satisfactory and ease of use. Users who did not have the benefit of the overlay expressed satisfaction with ease of use, but some dissatisfaction with knowing when the image was framed appropriately.

This study indicates that the face overlay guide can be used proactively to improve the quality of captured face images. Incorporating the overlay into the officers' workstations could assist in centering the camera on the subject's face with minimal cost to the process. We expect no additional training requirements and no impact on the time required to capture the face image.

6. REFERENCES

[1] L. Nadel, "Approaches to Face Image Capture at US-VISIT Ports of Entry," NIST Biometric Quality Workshop II, Nov. 2007, retrieved from http://www.itl.nist.gov/iad/894.03/quality/workshop07/presentations.html .

[2] P. Grother and G. Quinn, "Baseline Quality of US VISIT POE Facial Images", NIST Deliverable to DHS US-VISIT Face Image Quality Improvement Project, April 20, 2008

[3] M.F. Theofanos, B. C. Stanton, C. Sheppard, R. Micheals, J. Libert, and S. Orandi, Assessing Face Acquisition (NIST IR 7540), 2008, retrieved from http://zing.ncsl.nist.gov/biousa/

[4] ANSI INCITS 385-2004, Face Recognition Format for Data Interchange. American National Standards Institute, Inc.

[5] ISO/IEC 19794-5:2005 Information Technology - Biometric Data Interchange Formats - Part 5: Face image data. JTC1 : SC37, International Standard Edition, 2005. http://isotc.iso.org/isotcportal.

[6] Gibson, James L. (1977), The Theory of Affordances. In Perceiving, Acting, and Knowing, Eds. Robert Shaw and John Bransford, ISBN 0-470-99014-7

[7] Norman, Donald A. (1988): *The Design of Everyday Things.* New York, Doubleday.