(U) Low-Cost, Stand-off, 2D+3D Face Imaging for Biometric Identification using Fourier Transform Profilometry - Update

September 2010

Brian Redman, John Novotny, Taylor Grow, Van Rudd, Nathan Woody, Michael Hinckley, Paul McCumber, Nathan Rogers, Michael Hoening, Kelli Kubala, Scott Shald, Radek Uberna, and Tiffanie D'Alberto

Lockheed Martin Coherent Technologies
135 S. Taylor Ave., Louisville, CO 80027

Thomas Höft
Tufts University
Department of Mathematics
Bromfield-Pearson Hall
503 Boston Avenue
Medford, MA 02155

Russell Sibell
SIBELLOPTICS
815 Beauprez Avenue
Lafayette, CO 80026

Frederick W. Wheeler
GE Global Research
1 Research Circle
Niskayuna, NY 12309

ABSTRACT

Lockheed Martin Coherent Technologies is developing laser-based technologies for stand-off 2D+3D face imaging for biometric identification. Among other potential industrial, commercial, and governmental users, the Department of Homeland Security (DHS) and the Department of Defense (DoD) desire the ability to capture biometric data from minimally cooperative subjects with a minimally invasive system at stand-off distances. The initial applications are fixed installations for relatively large volume access points such as security check points and transportation gateways for which minimal cooperation, stand-off operation, and real-time operation are desired so that the biometric identification process will have little impact on traffic flow. Last year we presented a paper on the development and testing of a 2D+3D face imager breadboard based on the Fourier Transform Profilometry (FTP) technique. In that paper, we also presented the design for a field testable Engineering Development Unit (EDU) FTP-based 2D+3D imager. In this paper, we report the development and testing of this EDU, including the extension of the mode of operation to include person and head tracking, and the use of the reference-less FTP technique for area surveillance applications with subjects at arbitrary ranges within the operating range of the system. The test results show that the EDU meets the goals for the 2D+3D face imager of Class 1M eye-safe operation, 2D+3D face capture at >20-m stand-off distance, ~1-mm lateral resolution, ~1-mm range precision, and real-time capture at 1-Hz update rate for brisk walking speed (~1.5 m/s), minimally cooperative subjects using technologies and techniques that will be low cost in production.

1.0 Introduction

Among other potential industrial, commercial, and governmental users, the Department of Homeland Security (DHS) and the Department of Defense (DoD) desire the ability to capture biometric data from minimally cooperative subjects with a minimally invasive system at stand-off distances. Lockheed Martin Coherent Technologies is developing laser-based technologies for stand-off 2D+3D face imaging for biometric identification. The initial applications are fixed installations for relatively large volume access points such as security check points and transportation gateways for which minimal cooperation, stand-off operation, and real-time operation are desired so that the biometric identification process will have little impact on traffic flow.

Studies have shown that 2D+3D multi-modal face recognition performs better than either 2D or 3D face recognition alone.[1,2] We, therefore, designed our system to simultaneously capture 2D grayscale and 3D face imagery simultaneously with a single camera so that the 2D and 3D data are inherently pixel registered.

Last year we presented a paper on the development and testing of a 2D+3D face imager breadboard based on the Fourier Transform Profilometry (FTP) technique. In that paper, we also presented the design for a field testable Engineering Development Unit (EDU) FTP-based 2D+3D imager. In this paper, we report the development and testing of this EDU, including the extension of the mode of operation to include person and head tracking, and the use of the reference-less FTP technique for area surveillance applications with subjects at arbitrary ranges within the operating range of the system.

We will first briefly discuss the theory of FTP with references to our previous paper and to the relevant literature. We then describe the improvements over last year's breadboard incorporated in the EDU. We present test results showing that the EDU meets the goals for the 2D+3D face imager of Class 1M eye-safe operation, 2D+3D face capture at >20-m stand-off distance, ~1-mm lateral resolution, ~1-mm range precision, and real-time capture at 1-Hz update rate for brisk walking speed (~1.5 m/s), minimally cooperative subjects using technologies and techniques that will be low cost in production.

2.0 Fourier Transform Profilometry Background and Theory

The method of 3D shape measurement called Fourier transform profilometry (FTP) was invented by M. Takeda, et. al., in 1982-1983.[3,4] FTP is based on projecting an intensity fringe pattern onto an object and imaging the object and fringe pattern with a camera offset laterally from the fringe projector. The FTP processing consists of Fourier transforming the image data, extracting the complex data in the spatial frequency domain near the fringe carrier spatial frequency (bandpass filtering), inverse Fourier transforming the extracted data to form a complex spatial image, extracting the "wrapped" phase image from the complex image, "unwrapping" the "wrapped phase," and scaling the "unwrapped" phase image to the corresponding spatial dimensions. This process in one-dimension is completely analogous to using the analytic (complex) signal representation to find the signal's magnitude and phase, as is often used for processing temporal signals.[5]

Takeda and Mutoh originally performed the forward and inverse Fourier transforms on the data in one-dimension, line-by-line, for which the analytic signal representation applies exactly. However, two-dimensional Fourier transforms have been used for 3D shape measurement with the FTP method,[6] for digital holographic interference phase measurement[7], for interferogram fringe analysis[8], and for 3D surface inspection with an interferometric grating[9]. By 1996, Takeda, et.al., were using the 2D Fourier transform method and spatial frequency multiplexing of fringe patterns to improve the FTP method.[10] There are theoretical issues with the generalization of the analytic signal representation to dimensions higher than one, particularly in regard to its Hilbert transform formulation.[11] However, the straight forward use of the 2D Fourier transform does result in a complex signal in two dimensions for which the magnitude and phase can be separated so that the phase image can be constructed. Using the 2D Fourier

transform in the FTP method eliminates linear artifacts observed in the line-by-line, 1D Fourier transform method. In last year's paper, we discussed the FTP processing algorithms we developed using 2D FFTs.[12]

One of the key advantages of the FTP method is the simplicity of its hardware configuration, which enables low-cost system manufacturing. The theoretical advantages of the FTP technique over Moiré contouring include better accuracy, fully automatic distinction between a depression and an elevation on the object surface[13], and image intensity variation effects on the reconstructed surface shape are eliminated by the FTP method.[14]

The conceptual diagram of the FTP setup is shown in figure 1. As shown in the figure, the transmitter optics of focal length f_{xmt} projects an image of the transmission grating of period $p_{grating}$, and roll orientation angle θ_{go}, onto the subject at range R_{tgt} as a periodic fringe pattern. A digital camera, separated from the transmitter by a distance d_{xmt_rcvr}, captures an image of the subject with the projected fringes.

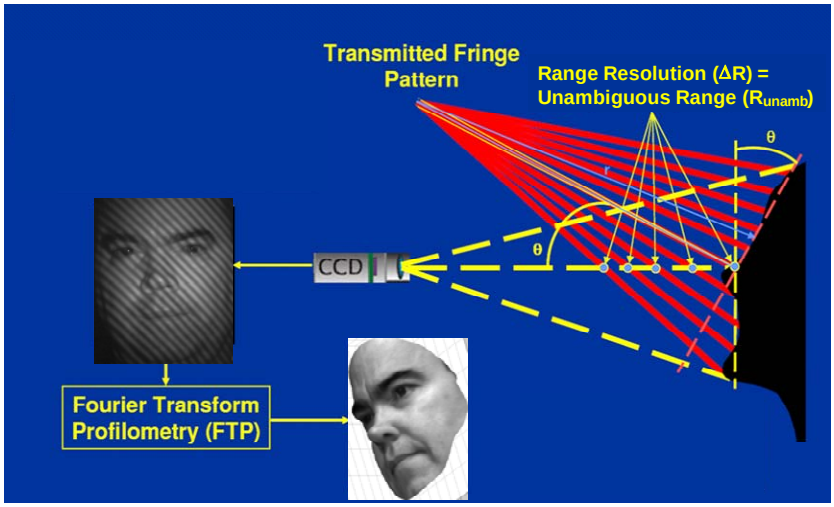


Figure 1. Fourier transform profilometry (FTP) Conceptual Diagram.

In figure 1, the intersections of a specific phase point (e.g., fringe maxima) of the projected fringe pattern with the line-of-sight (LOS) of a specific receiver pixel (in this case, the central pixel) are indicated to illustrate both the source of depth information being captured and of the phase/range ambiguity in the captured depth information. As illustrated, a pixel's LOS will intersect along the range or depth dimension different phase points along the projected fringe intensity pattern, and these phases repeat every $2\cdot\pi$ radians since the pattern is periodic in $2\cdot\pi$. The FTP method measures the phase of the fringe pattern, within an ambiguity interval, at the intersection of the pixel's LOS with the surface of the target. Figure 1 shows that the distance corresponding to a 2π radian phase ambiguity interval (the unambiguous range) increases as the angle between the transmitter and receiver optical axes decreases, and is a minimum when this angle is $\pi/2$ radians. The unambiguous range also increases with increasing fringe period. The equation defining the unambiguous range, ΔR, is given by

$$\Delta R = \frac{p_{grating} \cdot R_{tgt}^{\,2}}{p_{grating} \cdot R_{tgt} - f_{xmt} \cdot d_{xmt\_rcvr} \cdot \cos^2\left(\tan^{-1}\left(\frac{d_{xmt\_rcvr}}{R_{tgt}}\right)\right) \cdot \cos(\theta_{go})} \qquad (1)$$

Since the signal modulation that a pixel's line-of-sight sees is equivalent to an amplitude modulation, the equation for the measurement standard deviation for range (σ_{Range}), aka the range measurement precision, for an amplitude modulated signal applies to the FTP method[15]:

$$\sigma_{Range} = \frac{c}{4 \cdot \pi \cdot F_{AM} \cdot m \cdot \sqrt{SNR}} = \frac{\Delta R}{2 \cdot \pi \cdot m \cdot \sqrt{SNR}} \qquad (2)$$

where c = speed of light
F_{AM} = frequency of the amplitude modulation (Note: $\Delta R = c/(2 \cdot F_{AM})$)
m = modulation index of the amplitude modulation ((max – min)/(max + min))
SNR = electrical power signal-to-noise ratio = peak current or voltage squared divided by the variance of the current or voltage noise (usually measured in the power spectrum of the fringe image at the fringe spatial frequency peak).
ΔR is given by equation (1).

For our setup, the equation to convert phase difference mapping, $\Delta\phi(x,y)$, to height difference mapping, $h(x,y)$, is given by[16]

$$h(x,y) = \frac{l_0 \cdot \Delta\phi(x,y)}{\Delta\phi(x,y) - 2 \cdot \pi \cdot f_0 \cdot d} \qquad (3)$$

where l_0 = the distance between the receiver and the target along the receiver's optical axis,

d = the distance between the centers of the transmitter exit aperture and the receiver entrance aperture,

f_0 = the fringe spatial frequency = $1/p_0$, where p_0 is the projected fringe image period on the reference plane.

For most cases of interest with our system, $\Delta\phi << 2 \cdot \pi \cdot f_0 \cdot d$, in which case,

$$h(x,y) = \frac{l_0 \cdot \Delta\phi(x,y)}{\Delta\phi(x,y) - 2 \cdot \pi \cdot f_0 \cdot d} \approx -\frac{l_0 \cdot \Delta\phi(x,y)}{2 \cdot \pi \cdot f_0 \cdot d} = -\frac{p_0 \cdot l_0 \cdot \Delta\phi(x,y)}{2 \cdot \pi \cdot d} \qquad (4)$$

Equation (4) shows that the height sensitivity is better for larger angles between the transmitter and receiver optical axes (i.e., for larger d/l_0), and for higher fringe spatial frequency, f_0.

A counter-balancing constraint that limits the angle between the transmitter and receiver is the maximum measurable slope in the direction normal to the line of a fringe before the phase is over modulated, and this maximum slope is given by the following equation:[17]

$$\left| \frac{\partial h(x,y)}{\partial x} \right|_{\max} < \frac{1}{3}\left(\frac{l_0}{d}\right) \qquad (6)$$

where x is the variable along the direction normal to the line of the fringe.

Chen and Su have shown that the maximum measurable slope can be increased by a factor of three from $(1/3) \cdot (l_0/d)$ to (l_0/d) by subtracting the spectrum of the target image without fringes, which produces only a peak about zero frequency (aka, the DC peak), from the spectrum of the target image with fringes.[18]

In Takeda and Mutoh's original paper, the extraction of the phase map relied on having an image of the grating projected onto a reference plane in addition to an image of the grating projected onto the target. Takeda, Ina, and Kobayashi presented the FTP method without using a reference plane.[19] In this method, the peak spatial frequency of the bandpass filtered data near the grating spatial frequency is shifted down to zero frequency prior to taking the inverse Fourier transform. This reference-less FTP technique can be used for 3D imaging of targets at arbitrary, but known ranges, as long as (1) the transmitter and receiver track the target to aim the crossing points of their optical axes at the target, and (2) the projected and received fringe pattern are kept in focus at the target.

We have implemented automatic target tracking and focusing with processing using the reference-less FTP algorithm in our EDU to enable 2D+3D face capture of minimally cooperative subjects moving at up to brisk walking speed (~1.5 m/s) at any position within the operating envelope of the system which is approximately from 10-m minimum to 25-m maximum range over a field-of-regard full-angle of 38°.

3.0 EDU Description

The EDU design was described in our previous paper. Here we will describe the implementation of that design and enhancements that enable 2D+3D face capture of minimally cooperative subjects moving at up to brisk walking speed (~1.5 m/s) at any position within the operating envelope of the EDU.

Since the FTP technique does not require a temporally or spatially coherent illumination source, many previous implementations have used white light or light emitting diode (LED) sources. We, however, chose to use a near-infrared (NIR) (808 nm wavelength) diode laser as our illumination source in order to produce relatively short pulses (300-500 μs) with sufficient peak power (up to 35 W) in order to capture images of walking speed subjects (≤1.5 m/s) with sub-millimeter image smearing at ranges from ~10 m up to ~25 m. Using a diode laser also enables us to use a 40 nm full width half maximum (FWHM) optical bandpass filter on the receiver so that in conjunction with a short exposure time (~350-550 μs), the receiver is robust to most ambient lighting conditions from full darkness up to about outdoors partly cloudy conditions. In order to extend operation to clear day sunlit conditions, we plan to replace the 40 nm FWHM optical bandpass filter with a 10 nm FWHM optical bandpass filter, shorten the pulse duration to ~100 μs, and reduce the frame integration time to ~150 μs. By operating in the NIR waveband within the response band of silicon (Si) detectors, inexpensive Si CCD/CMOS focal plane array (FPA) based digital video cameras can be used.

Figure 2 shows a photograph of the EDU. The EDU is implemented in a mobile, full-height 19-inch rack with the receiver on its pan and tilt platform at the top, the transmitter on its pan and tilt platform in the middle, and the electronics box at the bottom. On top of the EDU rack is a high resolution, wide field-of-view (WFOV) camera (not shown in figure 2) that captures video images for person tracking by software developed by GE Global Research, under the GE-Shared Vision cooperative development program with Lockheed Martin Corporation. Beside the EDU is a half-height rack containing the operator control station which provides control of the EDU, collection of the data from the EDU, and processing and display of the captured 2D+3D face data. A separate computer or server (not shown in figure 2), is connected via ethernet to the operator station, which transmits the captured data to this computer/server for face recognition matching with the data from subjects enrolled in the database. In our current setup, face matching recognition/identification (ID) is performed using the COTS face recognition software FaceVACS-SDK made by Cognitec Systems Corporation.

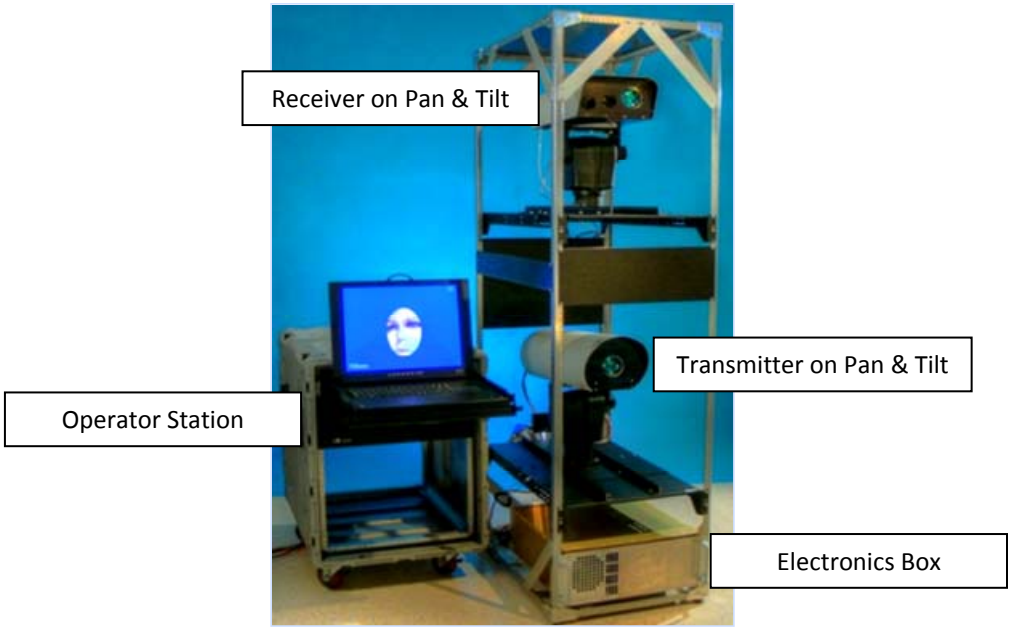


Figure 2. The FTP-based based 2D+3D face capture engineering development unit (EDU) for biometrics.

The EDU transmitter's output beam is formed as an extended source with a diameter and divergence such that the output is Class 1M eye-safe when the system is operated at 1 Hz pulse repetition frequency (PRF). The PRF of 1 Hz is consistent with real-time operation for most access control and check point applications without impeding traffic flow.

3.1. Tracker Description and Results

The EDU performs three types of tracking: whole person tracking, head tracking, and eye-tracking. As part of the tracking process, we also collect range information using a COTS Class 1 laser rangefinder that provides range measurements with +/- 2 cm accuracy at up to 1 kHz range reporting rate.

The whole person acquisition and tracking is performed using a wide field-of-view (WFOV) camera and tracking software developed by GE Global Research. The WFOV tracker provides person acquisition, tracking, and coarse (+/- 1 ft.) head location data. The person tracker displays the track location as an icon consisting of a red outline of two trapezoids around the tracked person as shown in figure 3. The person tracker can acquire and track several people within its FOV without ambiguity as long as they are separated laterally by about 3 feet from each other.

The user selects which of the multiple person tracks to pass to the intermediate field-of-view (IFOV) tracker by a mouse click on the track chosen by the user. The position of the chosen track is passed from the WFOV tracker software to the IFOV tracker software which moves the transmitter and receiver pan and tilt platforms to bring that chosen subject into the FOV of the IFOV tracking cameras (one each on the transmitter and receiver pan and tilt platforms). When the IFOV tracking software for each IFOV camera has acquired the subject, the IFOV tracking software takes control of the pan and tilt platform to provide fine pointing and tracking. The IFOV trackers use a Proportional Integral Derivative (PID) controller to remain pointed at their track point.

The track point is established by using a fusion of two tracking algorithms: an eye tracking algorithm and a head tracking algorithm. For eye tracking, we use Cognitec's eye tracker software included in the FaceVACS SDK. The head tracker uses the eye locations provided by the FaceVACS eye tracker for its initial head location track point. Lockheed Martin Coherent Technologies developed the head tracker algorithm and the tracking fusion algorithm that utilizes both the head tracker and the eye tracker. We did this so that the IFOV tracker would continue to track a non-cooperative subject's head even if the subject turned away so that both eyes were not visible to the tracking cameras as required by the FaceVACS eye tracker. We did not rely on head tracking alone for the following reasons: (1) we want to know when both eyes are visible to capture the face for best face recognition performance, (2) the eye tracker gives more accurate location information for transmitter and receiver pointing and tracking, and (3) the head tracker is more subject to "confusion" by clutter in the scene than is the eye tracker. Thus, the fusion of the eye tracker and head tracker algorithms provides much more accurate and robust face tracking than either eye or head tracking alone when tracking a non-cooperative subject in a cluttered environment. The tracking algorithm includes a Kalman filter.

The transmitter and receiver's motorized focusing stages are controlled based on the laser rangefinder data to keep the subject's face in focus as the subject moves in range. The transmitter's and receiver's image focus positions are adjusted by linear stages to fit the focus position given by the lens maker's equation for the transmitter and receiver's objective lenses based on the laser rangefinder data. A PID servo loop is used to adjust the stages' positions in real-time to keep the subject in focus as the subject moves in range. This autofocusing system is able to keep the subject's face in focus from 10-m to beyond 25-m range with a subject walking briskly up to 1.5 m/s.

Figure 3 shows examples of 3 frames from the WFOV tracking camera with the red tracking icon shown over the person being tracked as the person moved around in the FOV.

Figure 3. Frames from the WFOV tracker showing tracking of a person walking within the FOV.

4.0 Experimental Results

In this section we show experimental results collected with the EDU from resolution targets, a mannequin head, and human subjects. The resolution target results show that we achieved the goals of ~1 mm lateral resolution and range precision at >20-m standoff distance. The mannequin head data illustrates the potential benefits of 2D+3D imagery for extending the operation of face recognition to greater pose angles with respect to the pose angle in the gallery database. The human subject results demonstrate end-to-end operation of the EDU for face capture and recognition with good matching scores for subjects walking at up to 1.5 m/s brisk walking speed from 10-m up to 25-m standoff distances. Most of the data from people were collected at 10-m standoff distance due to the size of the room available for collecting data on a group of volunteers (about 40 individuals), but we also took data on a few individuals at standoff distances up to 30 m in another lab. This latter data showed that the EDU was limited to standoff distances up to and including 25 m as configured for these tests. We also took eye-safety measurements with NIST traceable calibrated power meters in accordance with the ANSI Z136.1(2007) standard and used the Laser Institute of America's (LIA) Advanced Laser Hazard Evaluator software to process the eye-safety measurement data. The results of the eye-safety measurements and LIA software processing results show that the EDU transmitter is class 1M eye-safe in accordance with the ANSI Z136.1(2007) standard.

4.1. Eye-Safety Measurements, "Fail-safes", and Laser Classification

In order to insure safe operation of the system, eye-safety measurements and fail-safe mechanisms were executed in accordance with the ANSI Z136.1 (2007) standard. The goal was to obtain a Class 1M transmitter eye-safety rating in accordance with the ANSI Z136.1 (2007) standard, with an embedded Class 4 diode laser source. This diode laser source is the nLight Pearl (model NL-P4-040-0808-3-A), which provides up to 40 W of continuous wave (cw) output. Since our system only required 1 Hz pulse repetition frequency (PRF), at 300-500 μs pulse duration, we determined that this source could be embedded in a Class 1M transmitter by accounting for the effects of the optical losses and beam shaping through the transmitter optical system, and by implementing the safety interlocks and "fail-safe" mechanisms required by the ANSI Z136.1 (2007) standard when using an embedded Class 4 laser source. Table 1 shows how the peak power drops as it transits the optical system from the laser output to the transmitter aperture resulting in a final peak power of 2.6 W. By measuring the pulse width and the pulse energy with a calibrated oscilloscope and calibrated energy meter, respectively, this maximum peak power output was confirmed for a number of current settings.

Table 1. Attenuation of the optical power at various stages through the transmitter.

Element	Percentage	Peak Power
Laser ouput	100%	40 W
Fiber output	87.5%	35 W
Fiber coupler	81.2%	32.5W
Aperture Plate	66.4%	26.5 W
Beam Shaping Optics	21.9%	8.7 W
Transmitter Telescope	6.6%	2.6 W

To determine the total pulse energy being emitted from the transmitter output, as well as the beam divergence, the first measurement made was the source size and the beam output diameter. The source size measured the apparent laser source seen by a camera looking at the output aperture. An image of the

source, taken from ~1 meter in front of the transmitter is shown in Figure (a). The source has a nearly square, flat-topped super-Gaussian profile. The image in Figure (b), which shows the output beam's intensity distribution projected onto a piece of paper, is used to determine the initial beam size, and to measure the intensity fluctuations on the beam profile for the eye-safety calculations described below.

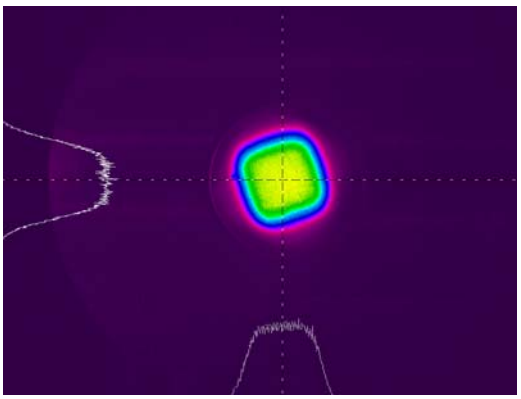
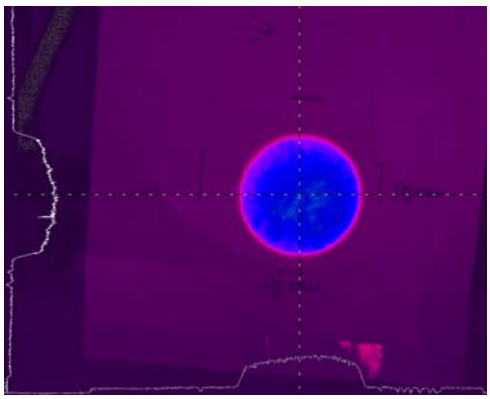
(a)  (b) 

Figure 5. Source (left) and beam images (right) taken at the transmitter output. Source size is 3.7 cm square, while the beam size is 6.7 cm circular, nearly flat top hat.

The transmitter output has a variable pulse duration between 300 µs and 500 µs. To insure eye-safe operation, the pulse energy output from the laser was taken from lasing threshold up to pulse durations of ~800 µs. The results from these measurements can be seen in Figure where the output energy and peak powers are shown as a function of output pulse duration for drive currents from 2.0 A to 4.0 A. The measurements were made using a 1 cm^2 area, calibrated pulse-energy meter with a calibrated neutral density filter. The apparatus was then put directly at the output plane of the transmitter to obtain the energy/pulse at the peak of the beam profile shown in Figure (b). The resulting measurement was then scaled by the optical density of the filter and the area of the output beam to obtain a total pulse energy. Simultaneously, a calibrated photodiode was used to detect the output pulse in order to obtain a pulse duration measurement. The peak power was calculated by dividing the energy by the pulse duration.

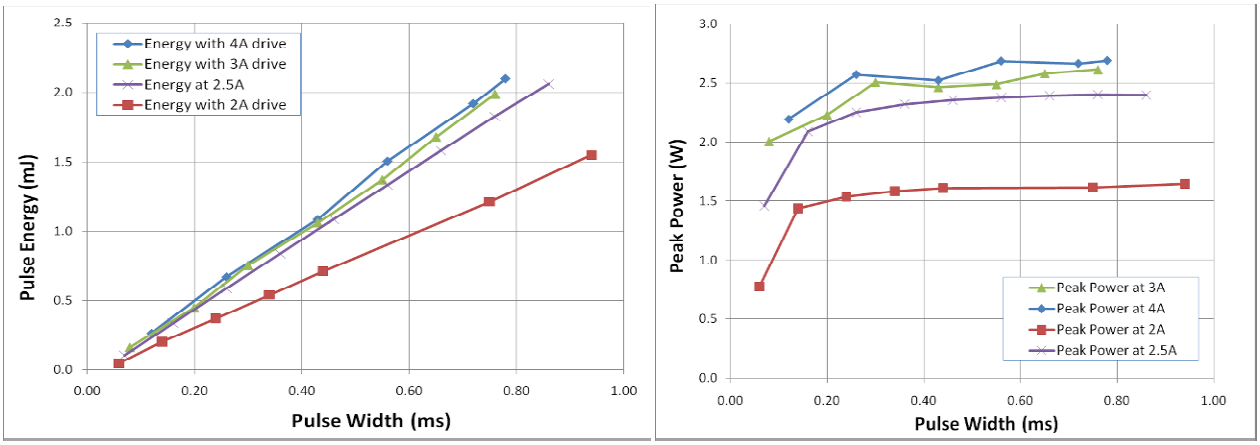


Figure 6. The laser's energy/pulse (left) and peak power (right) saturated just above 2.5 A drive current.

For all drive currents above ~2.5 A, the peak power and thus, the energy/pulse at a given pulse duration, saturated. As the current was turned up above this value, the output pulse became distorted and the energy-per-pulse fluctuated resulting in the decision to run the laser at 2.5 A. The more stable pulse-to-pulse energies obtained at 2.5 A provided more stable transmitter operation.

Once the energy and beam parameter measurements were complete, the Laser Institute of America (LIA) Laser Hazard Evaluation software was used to generate reports classifying the laser output for both aided and unaided viewing conditions. Additional safety factors were needed prior to classifying the laser output since the LIA software does not account for the peak intensity transmitted through the clear spaces of the Ronchi ruling grating, nor for the fluctuations in the beam's intensity profile. The ruling consisted of 54% transmission area through clear spaces and 46% blocked light area by the metal rulings, resulting in a 100/54 = 1.85 peak power enhancement in the bright fringes over the average intensity measured as described above. The intensity fluctuations on the beam profile shown in Figure 5 (b) had a peak value 1.6 times higher than the minima. Taking these together resulted in a factor of three increased peak irradiance compared to the measured average irradiance that had to be taken into consideration when calculating the exposure limits using the LIA software.

Table 2. Calculated Nominal Ocular Hazard Distance (NOHD) and Exposure Limits for the Biometrics transmitter both aided and unaided. The ratio between the Class limit and the measurement is greater than 3x in all cases.

Unaided Viewing Measurements

Measurement Distance: 1.7 m

Pulse width	Output pulse energy	Measured Energy/pulse through the Limiting Aperture (7 mm diam.)	NOHD	Class 1 Energy/pulse Limit through the Limiting Aperture (7 mm diam.)	Class 1 Limit/ Measurement
260 us	0.6 mJ	3.42 μJ	0	16.88 μJ	4.94
450 us	1.1 mJ	6.28 μJ	0	25.86 μJ	4.12

Aided Viewing Measurements w/ 7x50 Binoculars

Measurement Distance: 2.6 m

Pulse width	Output pulse energy	Measured Energy/pulse through the Limiting Aperture (50 mm diam.)	NOHD	Class 3a Energy/pulse Limit through the Limiting Aperture (50 mm diam.)	Class 3a Limit/ Measurement
260 us	0.6 mJ	67.57 μJ	3.7 m	253.82 μJ	3.76
450 us	1.1 mJ	123.88 μJ	4.4 m	389.37 μJ	3.14

A summary of the key results from the LIA software is shown in Table 2. As can be seen, the laser output for the unaided viewing case is a factor of more than four lower than the Class 1 limit for both operational cases, while the aided viewing (with 7x50 binoculars) case is a factor of more than three lower than the Class 3a limit. Taken together, these result in a Class 1M classification for the transmitter after taking into account the additional 3x factor needed due to the peak transmission through the Ronchi ruling grating and due to the peak fluence due to intensity fluctuations in the beam profile. Therefore, the transmitter is Class 1M, and beyond the Nominal Ocular Hazard Distance (NOHD) for aided viewing (with 7x50 binoculars) of 4.4 m, the beam's peak fluence is also below the maximum permissible exposure (MPE) for both aided and unaided viewing. Note that the EDU's nominal minimum operating range of 10 m is

much greater than the NOHD of 4.4 m for aided viewing with 7x50 binoculars.

In order to prevent inadvertent exposure to a beam with a higher power, a number of additional safety mechanisms were incorporated into the transmitter in accordance with the ANSI Z136.1 (2007) standard. These mechanisms prevent the laser from (1) outputting a pulse that is longer than 500 μs, (2) outputting more than 1 pulse per second, (3) operating the laser without the light-tight enclosure cover attached, and (4) outputting a pulse with a peak power higher than 2.6 W. Since the diode laser is incapable of outputting pulses with peak powers of more than 40 W, which corresponds to 2.6 W at the transmitter aperture, this last parameter was not a concern; however, a safety mechanism to prevent exceeding this peak power was still implemented in the interlock software.

The interlock software uses the signal from a photodiode mounted inside the transmitter head to monitor the transmitter output in real-time. This photodiode output is sampled by a National Instruments A/D board, and a LabView program then analyzes the signal for the three optical parameters mentioned: pulse duration, pulse repetition frequency, and peak power. If any of these parameters are out of the safe operating range, the trigger pulse from the computer that initiates laser firing is shut off, thereby stopping the laser's output. As the software must be running in order for the digital trigger to be generated, the system fails in a safe state since the laser output stops if the computer or software crashes.

Since the speed of software's reaction to the various faults was limited to roughly 50 ms, it was unable to shut the laser off in the case of a "pulse-too-long" fault immediately as the pulse reached the duration limit. To solve this problem, an analog pulse-duration interlock circuit was designed and built around a Hall current sensor attached to the diode's current lines. The circuit has three main parts, the Hall sensor, the interlock circuit board, and a high-power FET. The Hall sensor sends a digital signal to the circuit board, which is high when no current is present, and low when a current of more than .5 A is detected. The sensor output is passed through a low-pass filter with an RC time constant of 1 ms and compared to a variable set-point voltage. If the low-passed signal falls below the set-point voltage the interlock is tripped and the diode current is shunted through the FET instead of being allowed to go to the laser diode. Once the interlock is tripped it latches in the interlocked state until an operator manually pushes the reset button. The set-point voltage was manually set to cause the interlock to trip for any pulse durations longer than 500 μs, thus providing an immediate shut-down when the duration of any single pulse reaches this pulse duration limit.

4.2. Lateral Resolution and Range Precision Measurements

We used the EDU to collect 2D+3D images of a flat calibration target with an image of a black filled-in rectangle canted at 45 degrees with respect to the camera's pixel grid to make resolution, fringe modulation, SNR, and range precision measurements at 21-m range while the target was moving at brisk walking speed (~1.5 m/s) diagonally with respect to the receiver line-of-sight.

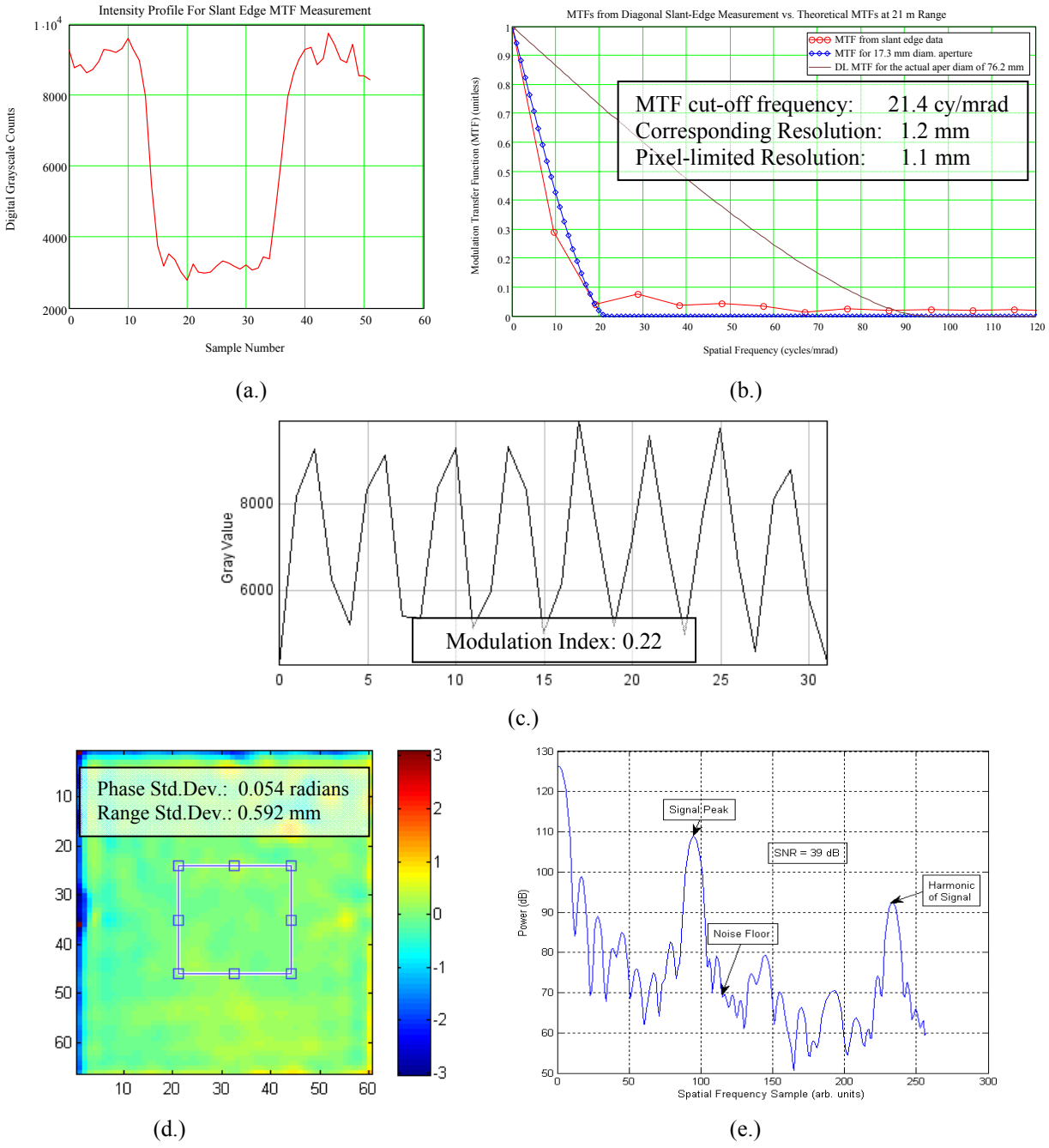


Figure 7. Results for a calibration target at 21-m range, 1.5 m/s diagonal motion (a.) Intensity profile through black rectangle target for slanted edge response, (b.) MTFs: calculated from the measured slanted edge response data shown in (a.) (red line, circles), diffraction-limited for clear aperture diameter of 17.3 mm, for which the MTF has the same cut-off frequency as the MTF calculated from the measured slanted edge response data (blue line, diamonds), and diffraction-limited for the physical clear aperture diameter of the receiver lens (76.2 mm) (brown line, no symbols), (c.) intensity profile through fringes perpendicular to the fringe lines for modulation index measurement, (d.) pseudo-color phase map of a uniform reflectance flat area of the target with the phase and range standard deviations measured from this data over the indicated square region, and (e.) intensity profile through the power spectrum of the fringe image of this area showing the SNR measured from this data.

Figure 7 shows the intensity profile of a cut through the black rectangle which provides the slanted edge response, the modulation transfer function (MTF) plot computed from this slanted edge response plotted

along with the diffraction-limited circular aperture MTFs for the receiver's physical clear aperture diameter of 76.2 mm and for the diffraction-limited circular aperture diameter that gives the same cut-off frequency as the slanted edge response MTF, the intensity profile of a cut perpendicular to the fringe lines to measure the fringe modulation index, the pseudo-color phase map of a uniform reflectance flat area of the target with the phase and range standard deviations measured from this data over the indicated square region, and the intensity profile through the power spectrum of the fringe image of this area showing the SNR measured from this data.

The equation for the lateral resolution, d_{lat_res}, in terms of the cut-off spatial frequency of the MTF in cycles per milliradian, ν_{cut_off}, for target range R is

$$d_{lat\_res} = 1.22 \cdot \frac{R}{\nu_{cut\_off}} \cdot 10^{-3} \tag{7.}$$

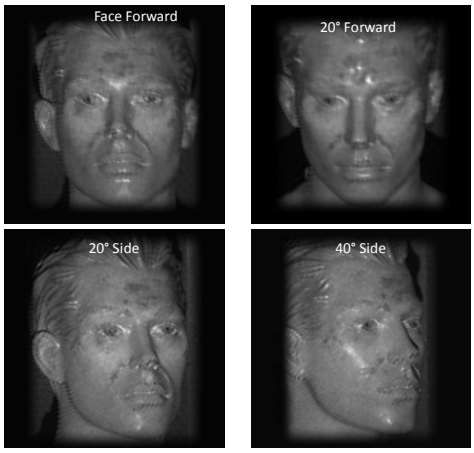
For the data shown in figure 7 (b.), ν_{cut_off} = 21.4 cy/mrad at 21-m range, which yields 1.2-mm lateral resolution. For comparison, the pixel-limited lateral resolution is 1.1 mm at 21-m standoff distance. This measured lateral resolution differs from the pixel-limited resolution by 9.1% for this data. (Note that the diffraction-limited resolution is much smaller than the pixel-limited resolution for the EDU receiver. The relatively large receiver aperture is needed for return signal light collection to insure sufficient SNR to provide the ~1-mm range precision.)

With the EDU transceiver parameters, the unambiguous range is 7 cm. The measured fringe modulation is 0.22 and the measured SNR is 39 dB at 21-m standoff distance. With these measured results, equation (2) predicts a range precision of 0.559 mm for 21-m standoff distance. The range standard deviation measured from the data shown in figure 7 (d.) collected at 21-m range is 0.592 mm. The measured range precision differs from the prediction of equation (2) by 5.9%.

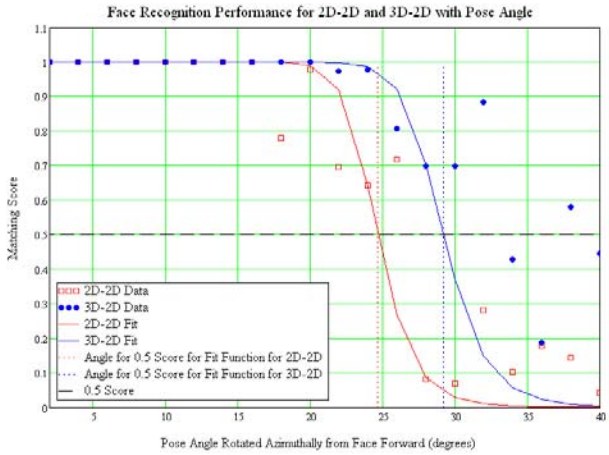
These results demonstrate that the EDU achieves the goals of ~1-mm lateral resolution and range precision for targets moving up to 1.5 m/s at greater than 20-m standoff distance.

4.3. Mannequin Face at Various Poses Results

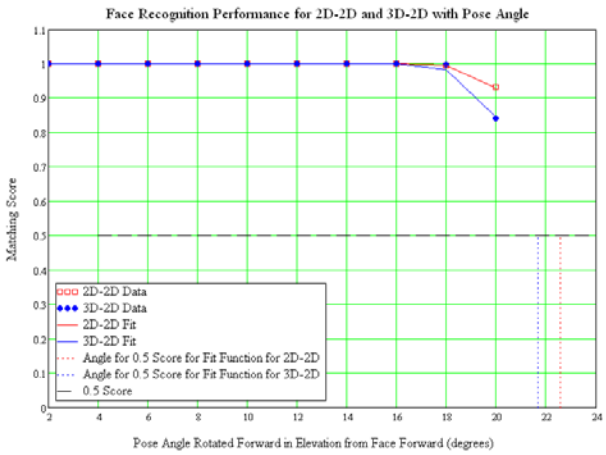
We used the EDU to collect mannequin 2D+3D face data with the mannequin's face turned in azimuth and elevation in increments of 2 degrees from zero degrees (face forward pose) to 40 degrees in azimuth and to 20 degrees in elevation. We used the mannequin for these measurements so that we could do the measurements early in development prior to human subject testing approval by the internal review board (IRB), and so that we could precisely control and measure the pose angles. Figure 8 shows a subset of the 2D+3D images of the mannequin's face at various poses. Also shown are the plots of matching scores from Cognitec's FaceVACs software versus pose angles for pose angles varying over azimuth and elevation angles for 2D-2D and 3D-2D matching against a 2D gallery image at face forward pose. 2D-2D matching means that the 2D grayscale image from the EDU is compared to the 2D grayscale image in the gallery database (in this case, Cognitec's FaceVACS software performs pose rotation transformations on the 2D image data internally to provide the best match with the gallery image). 3D-2D matching means that the EDU's 2D grayscale image overlaid onto the 3D image is rotated in our software to the pose of the 2D grayscale image in the gallery database, the 2D grayscale image from the EDU is projected to a 2D grayscale image at this pose, and the resulting rotated 2D grayscale image is compared to the 2D grayscale image in the gallery database.



(a.)



(b.)



(c.)

Figure 8. (a.) Sample 2D+3D images of a mannequin face at various poses, (b.) matching scores versus pose angle in azimuth for 2D-2D and 3D-2D matching against the 2D face forward gallery image, and matching scores versus pose angle in elevation (rotated forward) for 2D-2D and 3D-2D matching against the 2D face forward gallery image. The fitting function shown in the graphs has the form of the logistics function used in the Target Transfer Probability Function (TTPF) in the U.S. Army NVESD's models of human observer target recognition performance.

The fitting function for the data in the graphs for figures 8 (b.) and (c.) has the form of the logistics function used in the Target Transfer Probability Function (TTPF) in the U.S. Army NVESD's models of human observer target recognition performance, and has the following form:

$$\text{Score}_{\text{Fit}}(\text{Angle}, A_{50}) = \frac{\left(\frac{A_{50}}{\text{Angle}}\right)^{10\cdot\left(1+\frac{A_{50}}{\text{Angle}}\right)}}{1+\left(\frac{A_{50}}{\text{Angle}}\right)^{10\cdot\left(1+\frac{A_{50}}{\text{Angle}}\right)}}$$

where Angle = the pose angle from face forward (zero degrees) in degrees

A_{50} = the angle in degrees at which the score fitting function is 0.5

This score fitting function fits the data in figure 8 with the values for A_{50} given in Table 3.

Table 3: Values for the logistics function fitting parameter A_{50} for the data shown in figure 8.

Type of Matching	A_{50} for Azimuthal Rotation	A_{50} for Elevation Rotation
2D-2D	24.7^{o}	22.6^{o}
3D-2D	29.2^{o}	21.7^{o}

Note that the scatter in the score data shown in figure 8 indicates that the experimental results are not reliable beyond the A_{50} angle.

The results shown in figure 8 and table 3, indicate that the 3D data used in the 3D-2D matching significantly extends the angular range over which good matching scores (i.e., scores > 0.5) are achieved compared to 2D-2D matching for azimuthal pose angles, but that there is little difference between the two methods for elevation pose angles for this data set.

Due to time and funding limitations, we could not take a statistically significant set of data on many subjects and many pose angles to provide error rate and matching statistics over pose angles. However, the results shown in figure 8 are suggestive that the 3D-2D matching has the potential to provide an improvement in the angle out to which matching against a single pose, 2D gallery image can be accomplished compared to the angle out to which 2D-2D matching can be accomplished for azimuthal pose variations. This indicates the improvement that our 2D+3D face capture system may provide even in applications for which only 2D image galleries are available for matching.

4.4. Walking Human Subject Results

We used the EDU to collect data on people at brisk walking speed (~1.5 m/s) at standoff distances from 10 m to 30 m. Most of the data from people were collected at 10-m standoff distance due to the size of the room available for collecting data on a group of volunteers (about 40 individuals), but we also took data on a few individuals at standoff distances up to 30 m. This latter data showed that the EDU was limited to standoff distances up to and including 25 m as configured for these tests. Some subjects wore glasses and/or jewelry, and some did not. The data with glints from glasses and jewelry were used to examine the robustness of Cognitec's FaceVACS software against glints in the 2D+3D imagery. We also had one subject wear various disguises to test the system's robustness to such countermeasures.

Cognitec's FaceVACS software had shown very good immunity to glint points in our previous tests with glasses on a mannequin target and with synthetic glints added to imagery. If a large area glint obscures one or both eyes, FaceVACS cannot determine the eye-locations as needed in its image registration processing. In these cases, however, we can manually input the eye locations to FaceVACS which then successfully processes the input imagery with eye-obscuring glints.

Figure 9 shows samples of the fringe images of subjects' faces captured while they walked at brisk walking speed (~1.5 m/s) at 10-m standoff distance, and the resulting 3D-2D matching score normalized histograms. The subjects were not told to constrain their poses or facial expressions during the walking speed data collections, but most people looked forward and had mostly neutral expressions. FaceVACS matching scores range from 0 to 1, and Cognitec recommends a threshold score of 0.5 for a match in the FaceVACS manual. For the walking speed normalized histogram data shown on the left in figure 9, the matching is to a passive color photograph as the enrolled gallery face image, whereas the normalize histogram on the right shows the performance when matching to a grayscale image captured by the EDU as the enrolled gallery face image. The improved performance in the latter case, is likely due to lack of differences in illumination which can adversely affect performance in the former case. However, performance is still good when using a passive color photograph for the enrolled gallery image. Note that in all cases, for subjects who wear glasses, the enrolled gallery images were taken without glasses, but the probe images were collected with the subjects wearing their glasses.

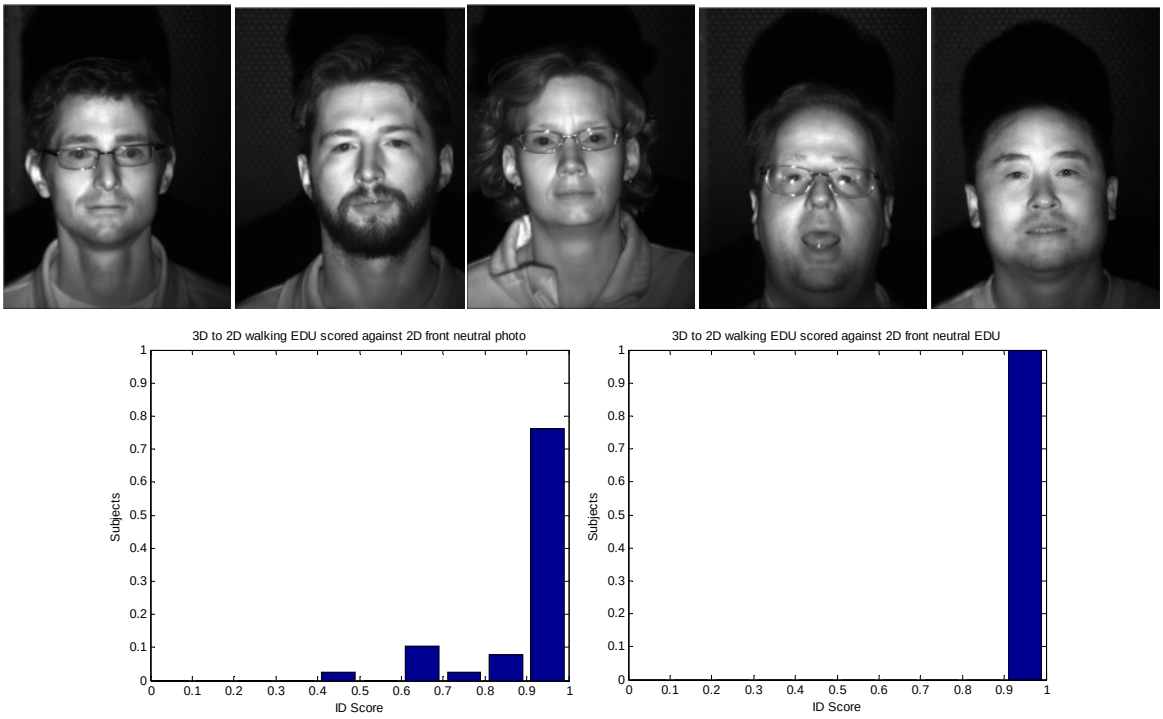


Figure 9. Samples of the processed 2D+3D face images of people captured while they walked briskly (~1.5 m/s) at 10-m standoff distance, and the resulting 3D-2D matching score normalized histograms. FaceVACS matching scores range from 0 to 1, and Cognitec recommends a threshold score of 0.5 for a match in the FaceVACS manual. For the walking speed normalized histogram data shown on the left, the matching is to a passive color photograph as the enrolled gallery face image, whereas the normalize histogram on the right shows the performance when matching to a grayscale image captured by the EDU as the enrolled gallery face image.

Due to budget and schedule limitations, we were unable to collect and process a statistically significant human subject data set over various ranges from 10 m to 25 m, expressions, poses, and lighting conditions with 2D-2D, 3D-2D, and 3D-3D matching. Therefore, we have not yet determined such statistical performance measures as the rank 1 matching rates, receiver operating characteristic (ROC) curves, and equal error rates for 2D-2D, 3D-2D, and 3D-3D matching. We did collect data with the subjects told to look to the right, up, and down at ~20-30 degrees from front view. The resulting matching score normalized histograms for these cases are shown in figure 10.

We had one subject wear various disguises to test the system's robustness to such countermeasures. Figure 11 shows the fringe images of the subject undisguised and with various disguises captured at brisk walking speed. Also shown in figure 11 are the resulting scores for 3D-2D and 3D-3D matching of the undisguised subject's enrollment image against the disguised subject and against 10 other undisguised subjects. The results show good robustness against disguises and against false cross matching for this limited data set, except for the case of the subject wearing a burqa (set 4 in the graph) for which the occlusion of the face was too great for good matching, so that a matching score very near zero was returned for both types of matching in this case.

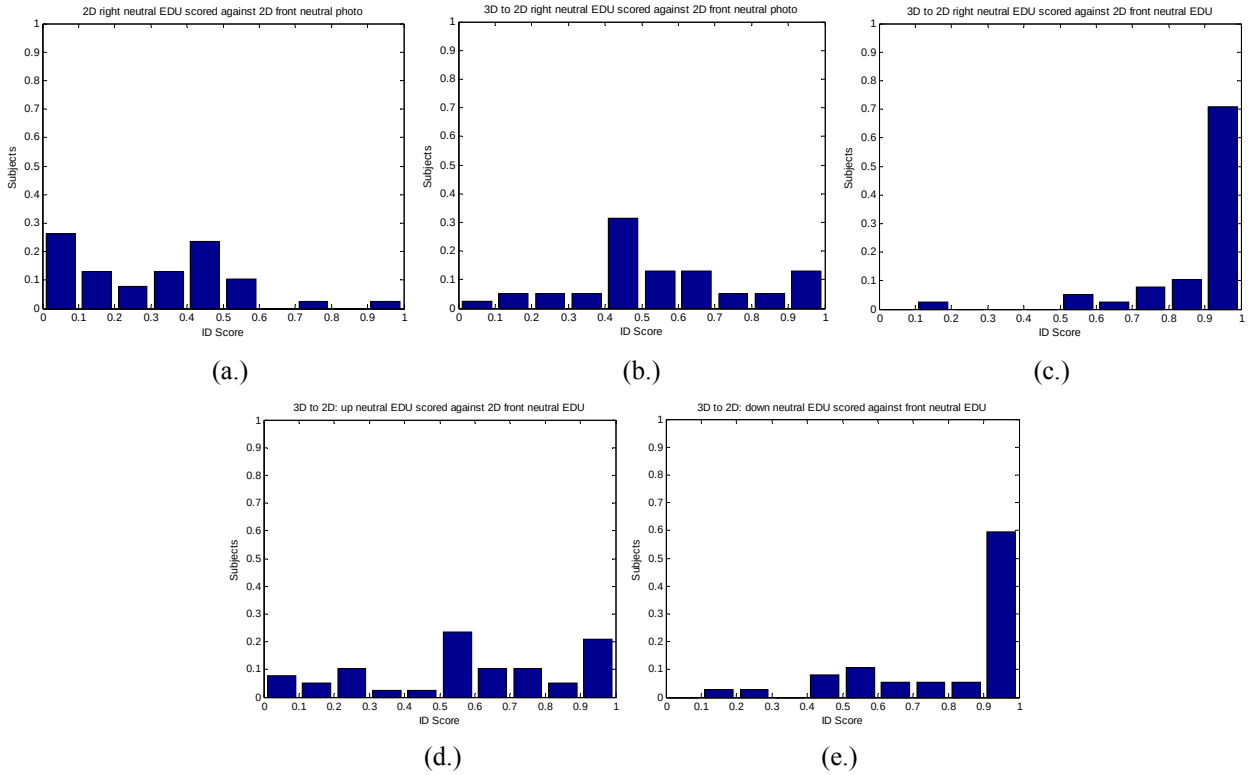


Figure 10. Matching score normalized histograms for various poses matched to a forward looking 2D gallery image. The pose angles were ~ 20-30 degrees from forward looking. (a.) pose looking right, 2D-2D matching to front neutral 2D color photograph gallery image, (b.) pose looking right, 3D-2D matching to front neutral 2D color photograph gallery image, (c.) pose looking right, 3D-2D matching to front neutral EDU 2D grayscale gallery image, (d.) pose looking up, 3D-2D matching to front neutral EDU 2D grayscale gallery image, and (e.) pose looking down, 3D-2D matching to front neutral EDU 2D grayscale gallery image.

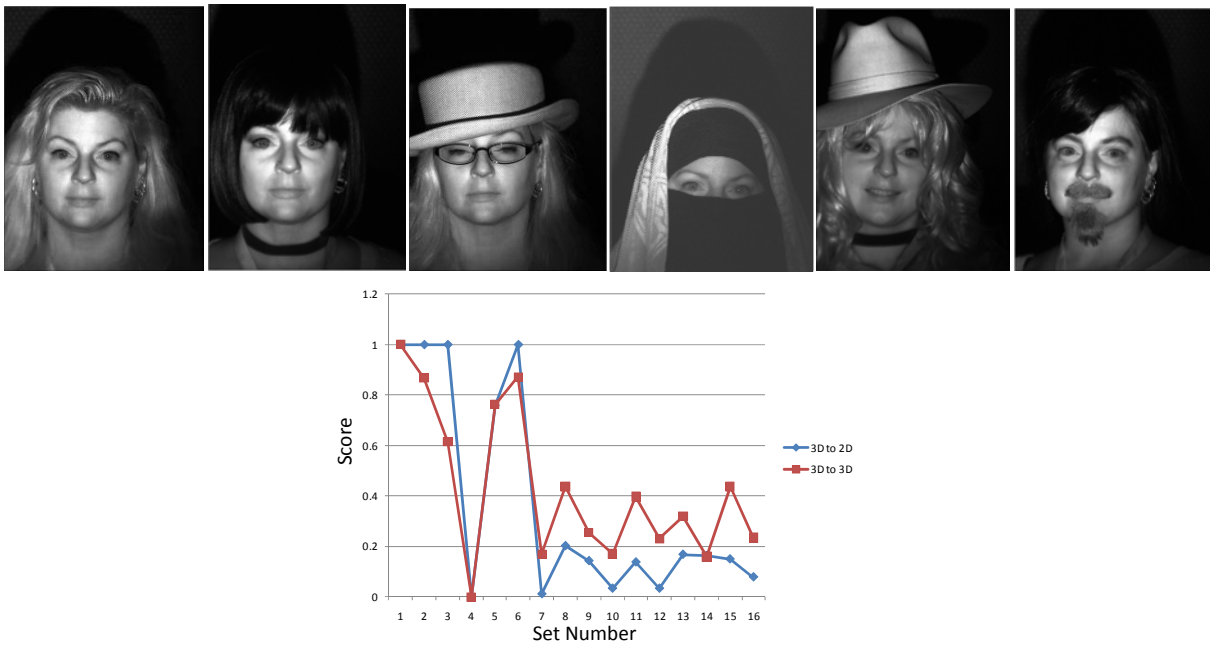


Figure 11. EDU performance for matching undisguised (set 1) against disguises (sets 2-6) and against other undisguised people (sets 7-16). For this set, only the burqa disguise (set 4) "fooled" the system. In this case, too much of the face was occluded for a good match.

While not statistically conclusive, the results presented in this section indicate that Lockheed Martin's biometric 2D+3D face image capture EDU may have the potential to provide more robust performance for minimally cooperative, walking subjects in uncontrolled conditions at extended standoff distances up to 25 m, while maintaining backward compatibility to existing 2D face image databases using 3D-2D matching. We are funded to collect more human subject data during the remainder of this year in field test demonstrations, but more work is needed to systematically collect and process a statistically significant set of data to determine the statistical performance measures for the system over a variety of poses, ambient illumination levels, and subject expressions at standoff distances from 10 m to 25 m.

4.5. 2D+3D Image Processing Times

We have measured the times for various steps in the end-to-end capture-to-scoring process. The image capture time is 350 μs. The time from image capture to a processed 2D+3D image varies because of large variations in the time to unwrap the wrapped phase image produced by the FTP method. The phase unwrapping time varies due to variations in the 3D structure of each face. Typically the time from capture to processed 2D+3D image is from one to six seconds on a Pentium processor based computer.

Currently, using Cognitec's FaceVACS-SDK, the time from input of the image data to determination of the matching score in one-to-one matching varies with the type of matching being performed and details of the imagery. Typically, for a Pentium processor based computer, the time for 2D-2D matching is ~5 seconds, for 3D-2D matching is ~9 seconds, and 3D-3D matching is ~15 seconds. Note that the time for 3D-2D matching also includes the time for Lockheed Martin Coherent Technologies developed code to determine the pose, rotate the 2D+3D probe image to the correct pose of the gallery image, and project the rotated 2D+3D image back to a 2D probe image prior to 2D-2D matching with the 2D gallery image. These FaceVACS matching times are currently longer than needed for real-time end-to-end face capture and one-to-one identification verification at 1-Hz rate.

A recommended next step is to move parts of the image processing from the Intel-Pentium-based CPU to an array of NVIDIA graphical processing units (GPUs) in the CUDA architecture to provide 2D and 3D processed images for input to FaceVACS in much less than 1 second. In addition, speeding up the face recognition matching times would require a cooperative program with Cognitec or with another face recognition software developer.

5.0 Future Work

The next step is to transport the EDU to test sites for data collections and demonstrations. We would like to collect a statistically significant data set and analyze this data to produce statistical performance measures such as ROC curves, equal error rates, and rank-1 identification performance over various conditions of pose, ambient illumination, expressions, genders, ethnicities, and ages at various standoff distances and subject motions. This data collection and analysis is necessary to demonstrate conclusively that the 2D+3D face imager EDU provides robust face identification with minimally cooperative, moving subjects under various conditions while maintaining backward compatibility with existing 2D face image databases.

We would also like to perform data collections with the EDU with subjects outdoors and in cars in order to statistically quantify performance under these conditions. We would need to modify the EDU to replace the 40 nm FWHM optical filter with a 10 nm FWHM optical filter, to reduce the pulse duration to 100 μs, and to reduce the frame integration time to 150 μs, in order to operate outdoors in full clear day indirect sunlight (sun not within the FOV). The EDU's fringe imaging system itself operates at low ambient light levels down to complete darkness since it is an active sensor with its own illumination source. The tracking cameras in the current EDU, however, are standard CCD/CMOS based cameras so that we would need to either add LED illuminators or replace the tracking cameras with low light level or

thermal infrared cameras to provide subject acquisition and tracking at low ambient light levels or in complete darkness (either when providing LED illumination or when using thermal infrared cameras). GE Global Research's person acquisition and tracking algorithms will work with visible color imagery or with thermal infrared imagery, but have not been developed for grayscale imagery as provided by most low light level, near-infrared (NIR) sensitive cameras. Thus, there would need to be development of grayscale person acquisition and tracking algorithms and software if we use either the NIR LED illuminators or low light level NIR cameras for the tracking cameras for operation in low light levels.

We would like to move parts of the image processing from the Intel-based CPU to an array of NVIDIA graphical processing units (GPUs) in the CUDA architecture to provide 2D and 3D processed images for input to the face recognition software in much less than 1 second. Speeding up the face recognition matching times would also require a cooperative program with Cognitec or with another face recognition software developer.

In parallel with EDU demonstrations, we would like to perform engineering design studies for the follow-on production prototype to reduce the size, weight, power, and cost of the transmitter, receiver, and electronics box, and to make the system amenable to low-cost mass production.

6.0 Summary

Lockheed Martin Coherent Technologies is developing a potentially low-cost, Fourier transform profilometry (FTP) based active 2D+3D face imaging system for standoff biometric identification of minimally cooperative, moving subjects in uncontrolled lighting situations. The initial applications are fixed installations for relatively large volume access points such as security check points and transportation gateways for which minimal cooperation, standoff operation, and real-time operation are desired so that the biometric identification process will have little impact on traffic flow. The technique is also applicable to area surveillance applications when using a wide field of view acquisition camera, tracking system, and rangefinder, which are all also incorporated in the engineering development unit (EDU).

The FTP-based EDU's experimental results presented in this paper show that the EDU has met the system goals of 3D face capture at >20-m standoff distance, ~1-mm lateral resolution, ~1-mm range precision, and real-time capture at 1-Hz update rates for brisk walking speed (up to ~ 1.5 m/s), minimally cooperative subjects using technologies and techniques that will be low cost in production. The EDU is now ready to be transported to test sites for data collections and demonstrations.

7.0 Acknowledgements

The work reported herein was funded by Lockheed Martin Corporation under the Ladar Growth Initiative internal research and development (IRAD) program. The authors thank Edward Kogan, Derek Hardy, and Michael Lamke of Lockheed Martin Information Systems and Global Security (IS&GS), for helpful suggestions on applications, deployment configurations, modes of operation, and desired system specifications and capabilities. Special thanks go to Tiffanie D'Alberto of Lockheed Martin Coherent Technologies, who managed this project, keeping it on schedule and within budget, and making the data collection arrangements, so we could focus on building and testing the system, and analyzing the data.

8.0 References

[1] Chang, K.I., K.W. Bowyer, P.J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on, Volume 27, Issue 4*, April 2005, pp. 619-624.

[2] Bowyer, K.W., K.I. Chang, P.J. Flynn, and Xin Chen, "Face Recognition Using 2-D, 3-D, and Infrared: Is Multimodal Better Than Multisample?" *Proceedings of the IEEE, Volume 94, Issue 11*, Nov. 2006, pp. 2000-2012.

[3] Takeda, M., H. Ina, and S. Kobayashi, "Fourier-transform method for fringe-pattern analysis for computer-based topography and interferometry," *JOSA, 72*, 1982, pp. 156-160.

[4] Takeda, Mitsuo and Kazuhiro Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Applied Optics, Vol. 22, No. 24*, 15 December 1983, pp. 3977-3982.

[5] Bedrosian, Edward, "The Analytic Signal Representation of Modulated Waveforms," *Proceedings of the IRE*, October 1962, pp. 2071-2076.

[6] Lin, Jin-Feng and Xian-Yu Su, "Two-dimensional Fourier transform profilometry for the automatic measurement of three-dimensional object shapes," *Opt. Eng., Vol. 34*, No. 11, 1995, pp. 3297-3302.

[7] Kreis, Thomas, "Digital holographic interference-phase measurement using the Fourier-transform method," *JOSAA, Vol. 3, No. 6*, June 1986, pp. 847-855.

[8] Bone, D.J., H.A. Bachor, and R.J. Sandeman, "Fringe-pattern analysis using a 2-D Fourier transform," *Appl. Opt. 25(10)*, 1986, pp. 1653-1660.

[9] Hung, Y.Y., S.H. Tang, and Q. Zhu, "3-D surface inspection using interferometric grating and 2-D FFT based technique," *Proc. SPIE 954*, 1988, pp. 32-35.

[10] Takeda, M., Quan Gu, Masaya Kinoshita, Hideaki Takai, and Yousuke Takahashi, "Spatially frequency-multiplexed number-theoretic phase unwrapping technique for the Fourier-transform profilometry of objects with height discontinuities and/or spatial isolations," *SPIE Vol. 2860*, 1996, pp. 46-53.

[11] Felsberg, Michael and Gerald Sommer, "The Monogenic Signal," *IEEE Trans. Signal Proc., Vol. 49*, No. 12, December, 2001, pp. 3136-3144.

[12] Redman, et. al., "Low-Cost, Stand-off, 2D+3D Face Imaging for Biometric Identification using Fourier Transform Profilometry," *MSS Proceedings of the 2009 Meeting of the Military Sensing Symposia (MSS) Specialty Group on Active E-O Systems, Vol.1*, October, 2009.

[13] Takeda and Mutoh, op. cit., p. 3977.

[14] Goodman, Joseph W., Speckle Phenomena in Optics, Theory and Applications, Englewood, CO: Ben Roberts & Company, 2007, p. 298.

[15] Jelalian, Albert V., Laser Radar Systems, Boston: Artech House, 1992, p. 45.

[16] Takeda and Mutoh, op. cit., p.3980.

[17] Ibid.

[18] Chen, Wenjing and Xianyu Su, "A new method of improving the measuring precision of 3D shape," *Proc. SPIE 4231*, 2000, pp. 412-415.

[19] Takeda, Ina, and Kobayashi, op. cit., pp. 156-157.