

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

September 2009

Brian Redman, Thomas Höft, Taylor Grow, John Novotny, Paul McCumber, Nathan Rogers, Michael Hoening, Kelli Kubala, Russell Sibell, Scott Shald, and Radek Uberna

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

Ronald Havermann

Lockheed Martin Space Systems Company  
PO Box 179, Denver, CO 80102

Dr. Sandalphon

Cinnabar Optics, Scientific Research Consulting  
325 Conrad Drive, Erie, CO 80516

### **ABSTRACT**

Lockheed Martin Coherent Technologies is developing laser-based technologies for stand-off 2D+3D face imaging for biometric identification. 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. This system is also extendable to wide area surveillance applications when coupled with a wide field of view acquisition camera, tracking system, and rangefinder. The goals for the low-cost, stand-off 3D face imager include Class 1M eye-safe operation, 2D+3D face capture at >10-m stand-off distance, <1-mm lateral resolution, <1-mm range precision, and real-time capture and processing for walking speed, minimally cooperative subjects using technologies and techniques that will be low cost in production. Lockheed Martin Coherent Technologies is developing both coherent and direct detection active imagers for biometrics applications. The coherent detection method is digital holography (presented in a companion paper) and the direct detection method is Fourier transform profilometry (FTP). In this paper, we report on the development of a tripod-mounted, breadboard direct detection 2D+3D face imager based on FTP. We discuss the theory of FTP with references to the literature. We then describe our implementation of the FTP breadboard, and present experimental results using this breadboard. We conclude with a discussion of the design for a fieldable engineering demonstration unit (EDU).

## 1.0 Introduction

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. The system is also extendable to wide area surveillance applications when coupled with a wide field of view acquisition camera, tracking system, and rangefinder. The goals for the low-cost, stand-off 2D+3D face imager include Class 1M eye-safe operation, 2D+3D face capture at >10-m stand-off distance, <1-mm lateral resolution, <1-mm range precision, and real-time capture and processing for walking speed, minimally cooperative subjects using technologies and techniques that will be low cost in production. In this paper, we report on the development of a tripod-mounted, breadboard direct detection 2D+3D face imager based on Fourier transform profilometry (FTP). We will briefly discuss the theory of FTP with references to the literature. We then describe our implementation of the FTP breadboard, and present experimental results using this breadboard. We conclude with a discussion of the design for a fieldable engineering demonstration unit (EDU).

## 2.0 Fourier Transform Profilometry Background and Theory

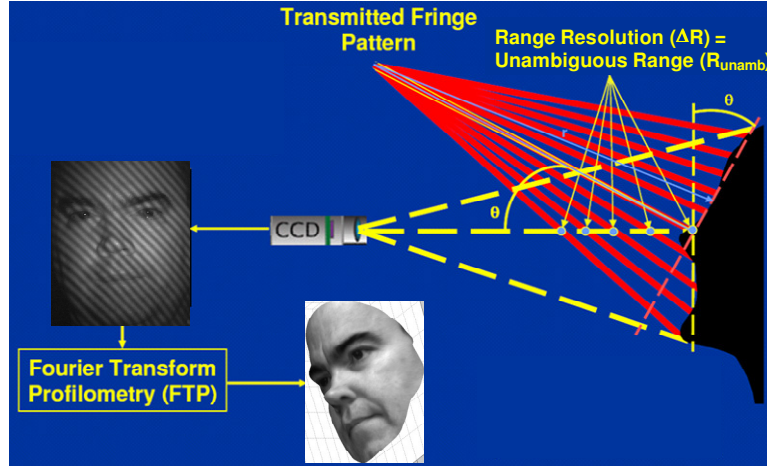
The method of 3D shape measurement called Fourier transform profilometry (FTP) was invented by M. Takeda, et. al., in 1982-1983.<sup>1,2</sup> FTP is based on projecting a sinusoidal intensity fringe pattern onto an object and imaging the object and fringe pattern with a camera offset laterally from 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.<sup>3</sup>

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,<sup>4</sup> for digital holographic interference phase measurement<sup>5</sup>, for interferogram fringe analysis<sup>6</sup>, and for 3D surface inspection with an interferometric grating<sup>7</sup>. By 1996, Takeda, et.al., were using the 2D Fourier transform method and spatial frequency multiplexing of fringe patterns to improve the FTP method.<sup>8</sup> 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.<sup>9</sup> 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. Below, we discuss the FTP processing algorithms we developed using 2D FFTs.

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<sup>10</sup>, and image intensity variation effects on the reconstructed surface shape are eliminated by the FTP method.<sup>11</sup>

The conceptual diagram of the FTP setup is shown in figure 1. As shown in the figure, the transmitter optics of focal length  $f_{\text{xmt}}$  projects an image of the sinusoidal transmission grating of period  $p_{\text{grating}}$ , and

roll orientation angle  $\theta_{go}$ , onto the subject at range  $R_{tgt}$ . A digital camera, separated from the transmitter by a distance  $d_{xmt\_rcvr}$ , captures an image of the subject with the projected fringe pattern.



**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 sinusoidal intensity pattern, and these phases repeat every  $2\pi$  radians since the pattern is sinusoidal. The FTP method measures the phase of the sinusoid, 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  $2\pi$  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,  $\Delta R$ , is given by

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

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 ( $\sigma_{Range}$ ), aka the range measurement precision, for an amplitude modulated signal applies to the FTP method<sup>12</sup>:

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

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.

$\Delta R$  is given by equation (1).

The geometry used in our breadboard FTP imager is referred to by Takeda and Mutoh as the crossed-optical-axes geometry since the optical axes of the transmitter projector and the receiver imager cross at a reference plane in object space.<sup>13</sup>

Our breadboard FTP imager is currently setup such that the transmitter and receiver are approximately coplanar as required in Takeda and Mutoh's original configuration. This coplanar configuration is no longer required since Mao, et.al., extended the FTP method to a more general geometry in 2007 for which the exit pupil of the projector and the entrance pupil of the camera are neither at the same height above the reference plane nor coplanar.<sup>14</sup>

For the original coplanar, crossed-optical-axes geometry of Takeda and Mutoh, the equation to convert phase difference mapping,  $\Delta\phi(x,y)$ , to height difference mapping,  $h(x,y)$ , is given by<sup>15</sup>

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

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,  $\Delta\phi \ll 2\pi f_0 d$ , in which case,

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

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:<sup>16</sup>

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

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  $(1/3)(l_0/d)$  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.<sup>17</sup> This method required taking a separate grayscale, fringeless image of the object in the same geometry as the fringe image of the object. In order to work with moving subjects without an additional camera, we have developed a method of accurately estimating the fringeless image 2D spectrum from the fringe image 2D spectrum so that no additional camera or additional captured image frame is required. Details of this DC subtraction algorithm are proprietary and will not be discussed in this paper.

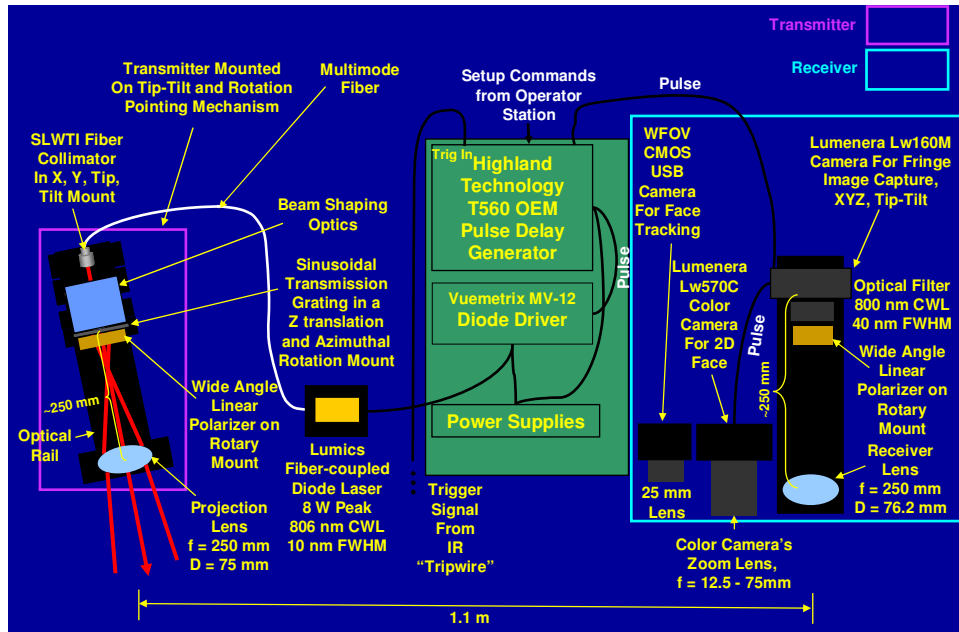
In Takeda and Mutoh's original paper, the extraction of the phase map relied on having an image of the grating projected onto a reference plane in addition to an image of the grating projected onto the target. Most of the results presented herein utilized an image of the grating projected onto a flat reference target which was collected before collecting target data, stored in the computer, and used in processing the target data as the target data was collected. Takeda, Ina, and Kobayashi presented the FTP method without using a reference plane.<sup>18</sup> 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 recently implemented the reference-less FTP algorithm in our breadboard system.

### 3.0 Breadboard Description

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 ( $\leq 500 \mu\text{s}$ ) with sufficient peak power (8 W) in order to capture images of walking speed subjects ( $\leq 1.5 \text{ m/s}$ ) with sub-millimeter image smearing at ranges from 5 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 ( $\sim 600 \mu\text{s}$ ), the receiver is immune to most ambient lighting conditions from full darkness to clear day indirect sunlight. By operating in the NIR waveband within the response band of silicon (Si) detectors, inexpensive Si CCD or CMOS focal plane array (FPA) based digital video cameras can be used in the receiver.

As shown in figure 2, the output of the diode laser's fiber pigtail is coupled to a fiber collimator. Additional free-space optics are used to shape the laser beam and illuminate a sinusoidal transmission grating which is imaged by the transmitter's objective lens onto the subject. The transmitter's objective lens is a 75 mm diameter, 250 mm focal length doublet with a NIR antireflective coating. The transmitter and receiver optical trains each include a linear polarizer crossed with each other to greatly reduce the intensity of specular reflections (aka glints), such as from eyeballs, eye-glasses, and saliva on lips and teeth, in the captured fringe image.



**Figure 2.** Fourier Transform Profilometry (FTP) breadboard layout.

The 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.

The projected sinusoidal fringe pattern and subject's face are imaged by the receiver's objective lens through the receiver's polarizer and narrowband optical filter onto the NIR responsive Si CCD camera's FPA. The NIR receiver's objective lens is a 76.2 mm diameter, 250 mm focal length doublet with a NIR antireflective coating.

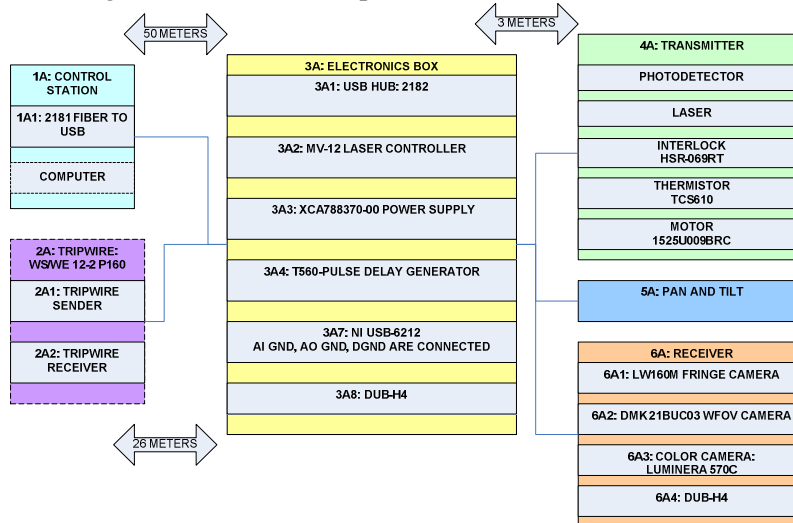
In addition to the NIR channel, the receiver includes a color Si CMOS camera connected to a manual zoom c-mount video lens with a focal length setting from 12.5 mm to 75 mm. This camera captures a color image of the subject's face for draping onto the 3D face surface reconstructed from the NIR fringe image data, thereby providing multi-modal 2D+3D face imagery for more robust face recognition.

The NIR monochrome and the color cameras each operate in externally triggered snapshot mode. A pulse delay generator sends trigger pulses to the laser diode controller and to the NIR camera to trigger a fringe image snapshot. The NIR camera outputs a trigger pulse to the color camera to trigger a nearly simultaneous color image snapshot. The exposure time on the NIR camera is set to 600  $\mu$ s to capture the entire 300  $\mu$ s to 500  $\mu$ s laser pulse with allowance for some timing delays and jitter. Currently, the pulse delay generator is externally triggered by an input pulse output by an optical "tripwire" triggered by the subject crossing the LED beam between the "tripwire" transmitter and receiver. The pulse delay generator can also be software triggered so that the fringe and color images can be snapped by man-in-the-loop control or in future configurations, automatically based on face-tracker and rangefinder tracking feedback.

The breadboard also includes a wide field-of-view (WFOV) monochrome camera for face tracking using the FaceVACS SDK eye-finder and eye-tracker software modules to control pan and tilt platforms for pointing the transmitter and receiver at the subject's face.

### 3.1. Electronics Description

As shown in figure 3, the breadboard consists of six electronics functional blocks: the control station, electronics box, receiver, transmitter, tripwire, and pan and tilt (P&T). Although each of these blocks is unique in function, the design considerations for power and communication are the same for each.



**Figure 3.** Breadboard electronics functional block diagram.

The system's requirements allow for the control station to be located up to 50 meters away. This distance is beyond a USB's functional communications length. Therefore, the communication design centers on USB to fiber conversion. The fiber to USB unit is located at the control station, and the USB to fiber conversion occurs in the electronics box. The USB to fiber unit has four USB inputs. Because there are six components with which to communicate, two four-to-one USB hubs have been integrated into the system: one in the electronics box and one in the receiver box. The hub in the electronics box connects to the laser controller, the National Instruments Digital Acquisition Card (NI DAC), the P&T, and the pulse delay generator, through RS-232 to USB converters. The hub in the receiver box connects the NIR fringe imaging camera, the WFOV face-tracking camera, and the color camera. By positioning the USB Hub in the receiver, the number of connections between the receiver and electronics box has been reduced, thereby, reducing cable clutter.

The other necessary signals communicate with the control station through the NI DAC. These signals include the tripwire stimulus, the laser and camera triggers, laser power supply inhibit, and various status signals. The selected NI DAC has the following capability: 16 Analog Inputs, 2 Analog Outputs, 32 Digital I/O, and two 32-bit counters. The major selection criterion for choosing this NI DAC was USB communication. With use of the USB to Fiber system, this eliminated the need to run signal wires 50 meters between the electronics box and the control station.

The other major components were selected for their functionality. The essential requirement for the laser controller was that it provides 9.5 A for the 8W laser. The breadboard's laser controller exceeds this requirement by sourcing up to 10 A. Additional considerations include small size, stand-alone operation, and an easy to use interface. The breadboard uses a very compact pulse delay generator with four outputs having 10 ps delay and width resolution. The optical tripwire provides both a positive and negative triggered response, from which to choose. The electrical design incorporates the positive 24 V trigger signal, which is divided down to 5V to be within the analog input range on the DAC. Lastly, the power supply has interchangeable cards, allowing for easy exchange of voltages as components change. The power supply also has individual inhibit/enable signals. Through this capability the system can autonomously shut down the laser should the pulse width, repetition rate, and/or peak power exceed eye-safe limits. The operator station's computer and fiber to USB unit receive direct wall power.

Due to the systems' range of signal and power levels, combination D-subminiature connectors are used at the box levels to minimize the number of cables and connectors, while still providing pins that can handle the current levels and signal requirements.

#### **4.0 Image Processing Algorithms with Experimental Results**

In this section we give an overview of the image processing algorithms, from preprocessing to final output, for implementing the FTP method described in section 2.0 above. Each of the major processing steps is illustrated with results from experimental data collected with the FTP breadboard. Note that for the data shown in this section, the breadboard configuration was slightly different than the current configuration described above, in that the transmitter and receiver objective lenses' focal lengths were each 160 mm instead of the current 250 mm, the separation between the transmitter and receiver pupils was 1.25 m, instead of 1.1 m, the diode laser output power was 3 W instead of 8 W, and no polarizers were used in the transmitter and receiver.

After the NIR camera and the visible camera simultaneously capture the fringe image and the visible image, respectively, the pre-processing steps are glint removal, region of interest (ROI) extraction, and apodization. Glints are frequently observed from specular surfaces on the subject, e.g., from eyeglasses, eyeballs, and saliva on the lips and teeth. While the polarizers in the transmitter and receiver reject the vast majority of these glints, in some cases, residual glints may remain to degrade the final surface reconstruction. In order to remove these residual glints in software, we identify high-slope regions of the data with a thresholded Laplacian filter, resulting in a "glint mask" indicating edges of the glints. We apply morphological filling operations to this mask so that large glints do not have empty interiors which go unprocessed. The glints are filled in with a median filter of the original data which is chosen to be large enough to smoothly connect the removed glint region with the data. We also use anisotropic diffusion methods, also known as "inpainting," to fill in the glints based on neighboring data.<sup>19</sup>

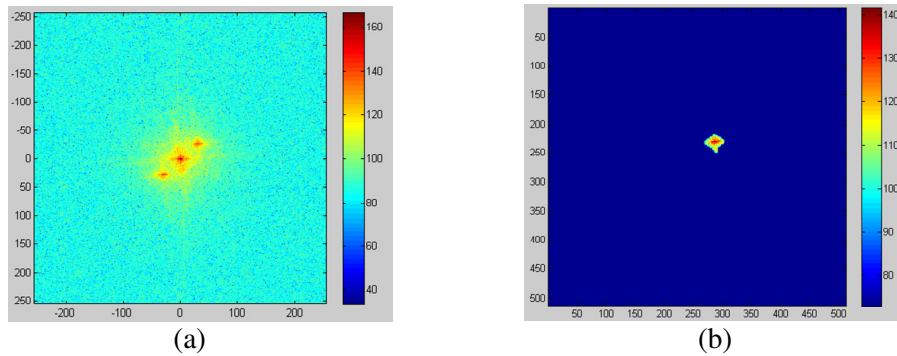
After glint removal, a ROI on the face image is selected based on the eye locations found by FaceVACS SDK. The size of the ROI is chosen so that high slope areas at the edge of the face are excluded, but the relevant face features for identification such as the eyes, nose, cheeks, and mouth are included. This ROI is extracted from the original data, and, after apodizing with a Tukey window, is placed into a  $2^N$  by  $2^N$  array, where N is typically 8 or 9. Figure 4 shows an example of pre-processed data collected by our breadboard with a live human subject at 10-m range.



**Figure 4.** Pre-processed fringe data of a live human subject at 10-m range.

From the pre-processed fringe image, we extract the frequencies near the fundamental frequency of the fringe pattern as explained in section 2.0. Using the 2D FFT we compute the complex 2D discrete Fourier transform of the fringe image; an example is given in Figure 5 (a), which shows the dB-scale power spectrum. Note that the spectrum is Hermitian symmetric about the origin since we have transformed real data. The large DC peak near the origin contains received irradiance information about the face image, while the two regions about the peaks at  $\pm$  the fringe fundamental frequency (first harmonic) along the 45-degree line contain the 3D surface information produced by the distortions in the fringe pattern.

The only portion of the spectrum which is needed for the FTP method is the region around the fringe fundamental frequency. We arbitrarily choose the upper-right (positive) frequency. Using a gradient-zero-crossing scheme we identify the peak of the spectrum away from the DC peak, and use this as the fundamental fringe frequency. The location of this peak can also be computed from geometric optics, grating frequency, and range-to-target, but we adaptively find it so that future hardware can leave the problem of determining frequency to the data-processing software if needed. From this peak we determine the distance (in frequency units) to the minimum value of the power spectrum located between the peak and DC. If we extract too little of the region about the peak, then maximal slope suffers. If we extract data that overlaps data from the region about the DC peak, then corruption of the surface results. The region extracted is a circle of this radius centered at the fringe fundamental frequency peak. Figure 5 (b) illustrates the extracted region about the fundamental frequency. This peak extraction process is equivalent to a 2D spatial frequency bandpass filter.

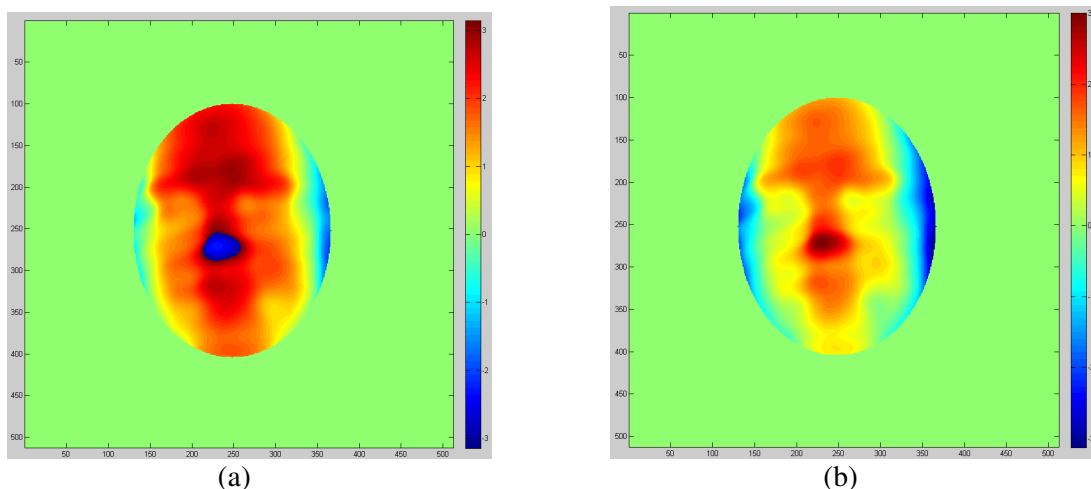


**Figure 5.** (a) Two-dimensional power spectrum of the fringe data in Figure 4, and (b) Extracted portion of the power spectrum of (a); the color bars are on a dB scale.

The complex spectrum in this circle is placed in a  $2^N$  by  $2^N$  array which is zero padded. Note that we have now broken the symmetry of the spectrum. We compute the inverse 2D FFT to obtain a complex data array, the phases of which correspond to ranges, and the magnitudes of which are not used in the 3D image reconstruction. All of the previous steps have also been performed on a reference flat which is now

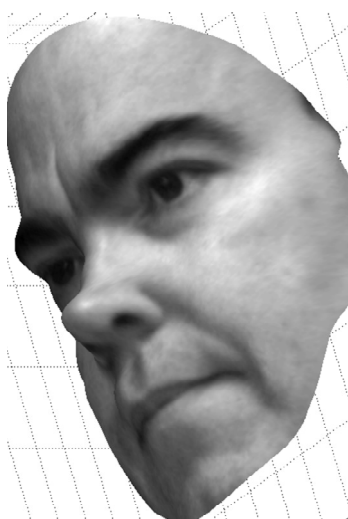


used as a phase reference. We compute the phase difference between the fringe image and the calibration flat to obtain a modulo- $2\pi$  “wrapped” phase (i.e. taking values on the interval  $[-\pi, \pi]$  repeatedly) surface. Figure 6 (a) shows this phase surface, where portions outside the face image have been set to zero. Using the  $L_p$ -norm minimization approach of Ghiglia and Pritt<sup>20</sup>, we “unwrap” the phase to a continuous surface. Example results are shown in Figure 6 (b). The values are still in radians, but are no longer constrained to the  $[-\pi, \pi]$  interval.



**Figure 6.** (a) Wrapped phase surface computed from the data in Figure 5, (b) Unwrapped phase surface computed from the wrapped phase in (a).

The unwrapped phase values are converted from phase to distance by equation (4). The digital grayscale image acquired simultaneously with the fringe image is registered to the 3D distance image, and the two are plotted together, with the grayscale image draped over the surface. The reconstructed 2D+3D image at a rotated pose is shown in Figure 7.



**Figure 7.** Reconstructed and rotated 3D image with the 2D grayscale image overlaid onto the 3D surface.

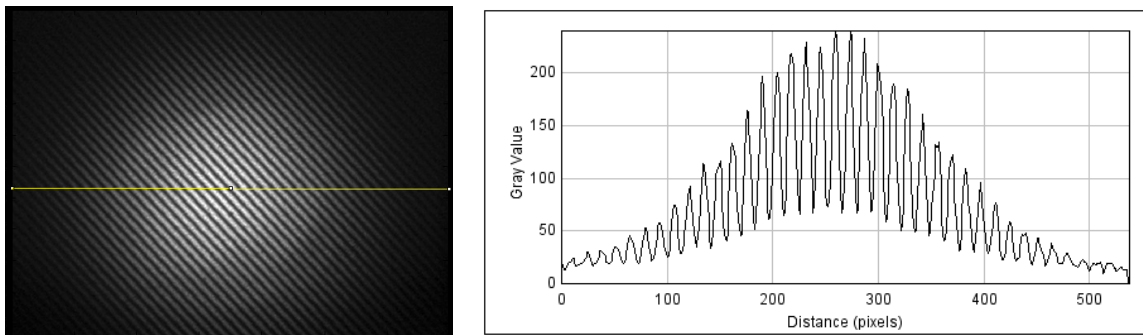
This 2D intensity + 3D surface image is the final data product output of our capture sensor for input to face recognition software. For automated 2D and/or 3D face recognition, the 2D+3D image data aids pose estimation and rotation of the 2D and 3D data to the pose of the face images in the gallery (i.e., in the face image database). After 3D rotation to the correct pose, the 2D face data overlaid onto the 3D surface data can be projected back to a flat 2D image at the correct pose. This image can be used for 2D only face recognition for backward compatibility to existing 2D only face image databases. If 3D data is available

in the gallery, full 3D face recognition matching can be performed in addition to the 2D matching, or combined 2D+3D multi-modal face recognition can be performed. Studies have shown that 2D+3D multi-modal face recognition performs better than either 2D or 3D face recognition alone.<sup>21,22</sup>

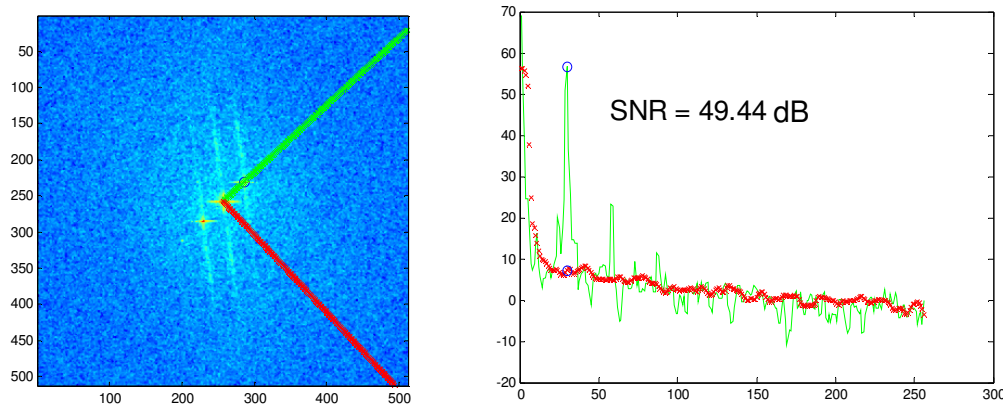
#### 4.1. Range Precision Measurements

We illuminated a flat target with the sinusoidal fringe pattern to measure the modulation index  $((\max - \min)/(\max + \min))$ , SNR, and range precision. The image of the flat target and a plot of the cross-section profile, marked with a horizontal line across the flat target image, are shown in figure 8. The modulation index computed from this profile varies across the flat target from about 0.3 to 0.5 in the useful area of the illumination spot. For calculations, we used an average modulation index of 0.4 for this data. The fringe period of the grating in the transmitter for this data was 100  $\mu\text{m}$ . The grating roll orientation angle was  $45^\circ$ . The focal lengths of transmitter and receiver objective lenses' used for this data collection were each 160 mm. The lateral distance between the transmitter and receiver pupils was 1.25 m. The target distance was 10 m. The unambiguous range calculated from equation (1) for this setup was 7.234 cm.

Figure 9 shows the 2D power spectrum (PS) of the flat target image data shown in figure 8, and a plot of the PS cross-section from zero frequency along a  $45^\circ$  diagonal line segment in the positive x – positive y frequency quadrant. The PS noise floor profile shown in the same plot of figure 9 is taken from zero frequency along a  $45^\circ$  diagonal in the positive x – negative y frequency quadrant. These PS profiles are plotted in dB. The SNR in dB is given by the peak value of the PS profile containing the fringe fundamental frequency peak minus the value of the PS noise floor profile at the same absolute frequency. The SNR measured from this data is  $49.44 \text{ dB} = 10^{4.944} = 87902$ .



**Figure 8.** Flat target illuminated with the sinusoidal fringe pattern (left) and the plot of grayscale level profile for a horizontal cross-section across the flat target image (right).



**Figure 9.** 2D Power Spectrum of the flat target data showing the signal profile line (green) and noise floor profile line (red) (left), and the signal (green) and noise floor (red) Power Spectrum profile plots (right) for measuring the SNR. The fringe fundamental frequency peak and corresponding point on the noise floor are indicated by blue circles in the plot on the right.

From the data in figure 9, using equation (2), we predict a range standard deviation,  $\sigma_{\text{Range}}$ , of 0.097 mm. We reconstructed the unwrapped phase image of this flat target using the algorithm described in section 4.0. We then measured the standard deviation of phase,  $\sigma_\phi$ , for this data to be 0.038 radians. The equation for converting the phase standard deviation to range standard deviation is

$$\sigma_{\text{Range}} = \frac{R_{\text{tgt}} \cdot \sigma_\phi}{2 \cdot \pi \cdot d_{\text{xmt-rcvr}}} \cdot \left[ \frac{p_{\text{grating}} \cdot R_{\text{tgt}}}{f_{\text{xmt}} \cdot \cos\left(\tan^{-1}\left(\frac{d_{\text{xmt-rcvr}}}{R_{\text{tgt}}}\right)\right) \cdot \cos(\theta_{\text{go}})} \right] \quad (7)$$

(Note: The term in square brackets [...] in equation (7) is the projected fringe period at the target along the dimension parallel to the direction of the line segment connecting the transmitter and receiver pupil centers, for the flat target's surface oriented perpendicular to the receiver's line-of-sight, i.e.,  $p_0$  (see equations (3) and (4)).)

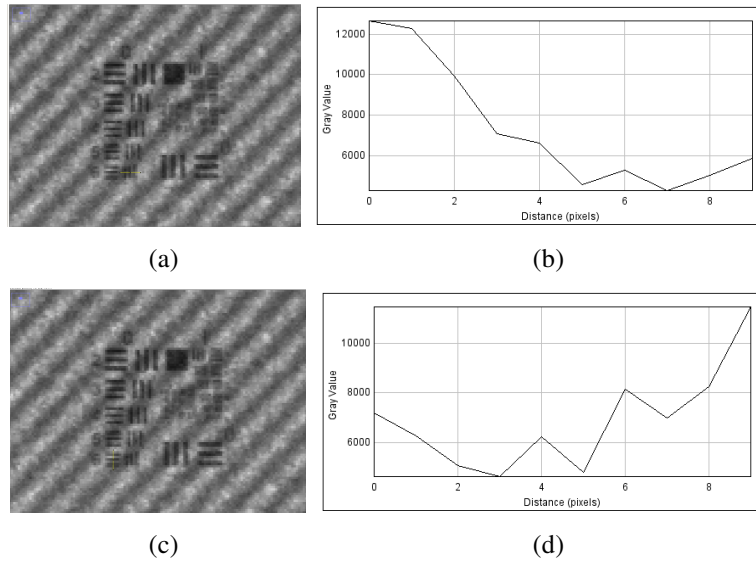
For the flat target data shown in figures 8 and 9, and the measured standard deviation of phase,  $\sigma_\phi = 0.038$  radians, equation (7) yields a range standard deviation of  $\sigma_{\text{Range}} = 0.431$  mm. The theoretically achievable range standard deviation as calculated using equation (2) of 0.097 mm is 4.44 times lower than this measured range standard deviation. Clearly, the range precision is not limited by the ratio of the unambiguous range to the square root of the SNR for this data set. Other errors not accounted for by equation (2) must dominate the range precision for this high SNR data. These other errors may be due to the roll-off in modulation index from the center of the illuminating spot, to deviations of the grating pattern from a true sinusoid, to errors caused by the overlap of the tails of the DC peak with the fundamental fringe frequency peak, and/or to errors in the 3D reconstruction processing, such as phase unwrapping errors.

For the human subject face data shown in section 4.0, taken with the same configuration as the flat target data shown in this section, the measured SNR is 37.25 dB and the fringe modulation index is about 0.2 on average, yielding a predicted range standard deviation of 0.527 mm for this data.

## 4.2. Lateral Resolution Measurements

In order to measure the FTP breadboard's lateral resolution, we imaged a flat 1951 U.S. Air Force Resolution bar target pattern at 10-m range. The resolution target's bar spacing at Group 0, Element 1 is 2 mm. The images of the resolution target, and vertical and horizontal grayscale profiles through the limiting resolution bar pattern are shown in figure 10. We define the limiting resolution to be the bar spacing for which the modulation index for the profile through a bar pattern is between 0.05 and 0.15. We specify a modulation index range since the bar patterns represent discrete rather than continuous resolution measurements. Under this definition, the limiting resolution corresponds to Group 0, Element 6 for which the bar spacing is 1.12 mm. The data shown here was taken with a 160 mm focal length lens on the FTP breadboard receiver. With this focal length, 12.9  $\mu\text{m}$  pitch CCD array pixels, and 10-m range, the predicted pixel-limited resolution is 0.806 mm. This predicted resolution is 28% lower than measured.

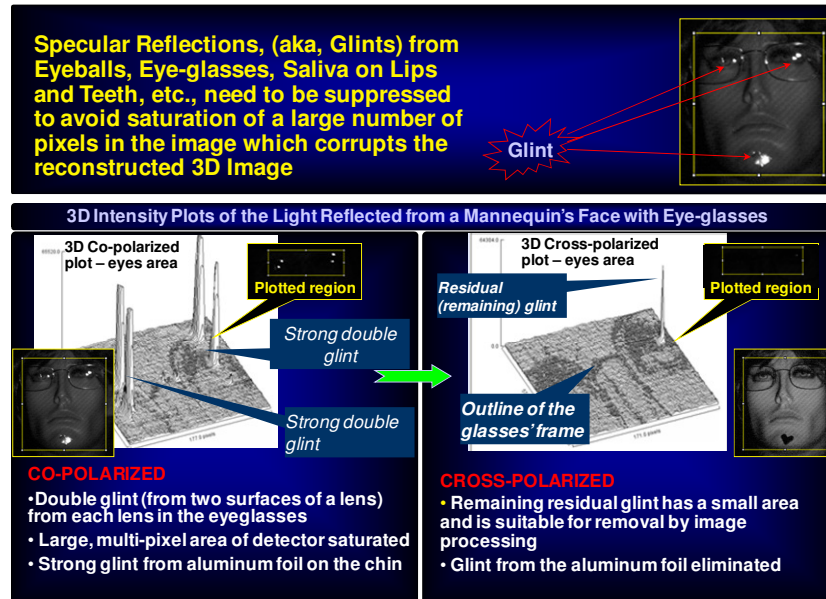
Since the data shown in sections 4.0 – 4.2 was taken, we have increased the transmitter and receiver object lenses' focal lengths from 160 mm to 250 mm in order to improve the range precision and the lateral resolution for operation at longer ranges. We also increased the output power of the diode laser source from 3 W to 8 W to compensate for losses due to the addition of crossed linear polarizers in the transmitter and receiver, which are used for glint suppression. These changes have resulted in qualitatively improved images. Since the design of the Engineering Demonstration Unit (EDU) described in section 6 is based on the new FTP breadboard design, measured performance results for this new configuration will be reported at a later date when the EDU has been built and characterized.



**Figure 10.** Lateral Resolution Measurement Results using an Air Force Resolution Target. (a) Zoomed image of the resolution target showing the horizontal line through element 6, (b) grayscale profile along the horizontal line shown in (a) (the modulation index measured from pixels 5 and 6 in this profile is 0.071), (c) Zoomed image of the resolution target showing the vertical line through element 6, (d) grayscale profile along the vertical line shown in (c) (the modulation index measured from pixels 4 and 5 in this profile is 0.13).

### 4.3. Glint Suppression Results

In order to suppress specular reflections (glints) from eyeballs, eye-glasses, saliva on lips and teeth, etc., we placed a linear polarizer in the transmitter optical path and a crossed linear polarizer in the receiver optical path. Figure 11 shows the results of experiments comparing the glints from eye-glasses and aluminum foil on a mannequin head with the transmitter and receiver polarizers co- and cross-polarized. The excellent glint suppression provided by crossed polarization is evident in these images.



**Figure 11.** Experimental results for glint suppression using crossed linear polarizers.

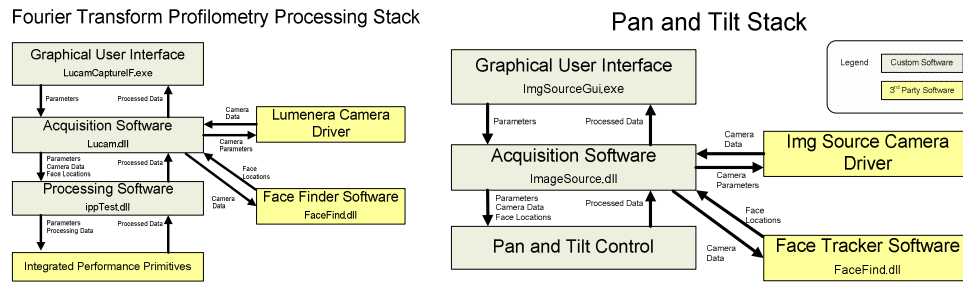
## 5.0 C++ Software

In order to speed up the processing of the data collected with the FTP breadboard, we ported the algorithms originally developed in Matlab to C++, and utilized the Intel Integrated Performance Primitives (IPP) and Intel Math Kernel Library (MKL) to implement particular functions.

In addition to the FTP processing software, we also developed software for hardware control, including controlling a pan and tilt (P&T) platform to track a subject's face based on the eye-finding and tracking software modules in the FaceVACS SDK.

### 5.1. Software Organization

The 3D face construction software is organized into software component stacks. Figure 12 shows the Fourier Transform Profilometry stack, including the processing, and the camera control software. Figure 12 also shows the pan and tilt software component stack, including the Imaging Source WFOV tracking camera control software.



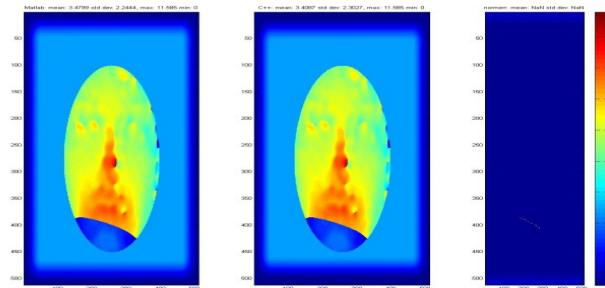
**Figure 12.** Fourier Transform Processing Stack, including camera control for image acquisition (left), and Pan and Tilt (P&T) Software Stack (right).

#### 5.1.1 Implementation of the FTP Algorithms

In order to reduce the 2D+3D image reconstruction and display time to 1-2 seconds, we implemented the FTP algorithm in C++. Using C++ improved memory allocation and reduced the overhead of running the algorithm in Matlab. By using C++ we also took advantage of the C++ Intel Integrated Performance Primitives (IPP) and the Intel Math Kernel Library (MKL). These libraries provide best in class execution times for basic image processing operations.<sup>23</sup> Finally, C++ allowed easy integration of different components, including the USB cameras, the pan and tilt device, and the face tracker software (FaceVacs SDK 8.0.1).

#### Verification of FTP results

To compare the C++ results with the Matlab results, we used Matlab to find the squared differences between C++ and Matlab results. The differences were on the order of floating point precision error. As shown in figure 13, we compared the image of the differences between the Matlab and C++ results in order to visualize the distribution of errors over the image area.



**Figure 13.** Comparison of Matlab and C++ results for the unwrapped phase image. The left image is the Matlab result, the middle image is the C++ result, and the right image is the squared difference between the two images.

### ***FTP Performance Analysis***

The total time for the FTP processing is 1.23 seconds. One of the longest duration processes was the phase unwrapping. To reduce the number of phase wrapping ambiguities near the edges of the face, and thus, reduce the total processing time and number of phase unwrapping errors, the software applies a tight elliptical ROI mask to condition the input to the phase unwrapping algorithm. This masking has reduced the duration of the phase unwrapping process to 0.04 seconds. Another long duration (730 ms) process is the “Face find” process for the color and grayscale images performed by the FaceVACS SDK eye-finder module. Table 1 shows the execution times for the most time consuming major processing components.

**Table 1.** Execution Times for Major Processing Components

Processing Component	Time (s)
Time to unwrap	0.04
Time to drape	0.37
DC removal	0.03
Glint Removal	0.01
2 Face Finds	0.73
<b>Total estimate</b>	<b>1.18</b>

### ***Software Opportunities for Improvement***

There are several opportunities for improvement of the software. “Face find” time could be reduced by applying tighter cropping to the color image. Currently, the entire color and fringe images are searched for the face, but the face location data from the WFOV face tracking camera should enable us to narrow the search in both the color and fringe images. Draping time could also be reduced by using the eye locations already calculated to limit the domain over which image registration between the color and NIR image is performed. This may also increase the accuracy of the registration because multiple peaks currently occur in the correlation image, and limiting the registration domain would constrain the algorithm to only searching over a narrower window with fewer correlation peaks. Utilizing an Nvidia Compute Unified Device Architecture (CUDA) card, such as the GeForce 6600, to parallelize parts of the algorithm might dramatically reduce processing times, especially with the sometimes long and variable execution times observed for phase unwrapping. Using a different phase unwrapping algorithm might also reduce phase unwrapping execution time and its variability compared to the current implementation using the Giglia and Pritt  $L_p$ -norm minimization algorithm. Another phase unwrapping algorithm we are investigating is PUMA.<sup>24</sup>

### **5.1.2 Implementation of the Pan and Tilt Control**

For the breadboard, we used a Quickset Model Number QPT130IC pan and tilt (P&T) platform with the following relevant specifications:

- 150 ft. lb. lift capacity fixed mount (130 ft.lbs. mobile)
- .25° Pointing resolution/repeatability
- 435° ( $\pm 217.5^\circ$ ) Pan movement range
- .3° to 8° /sec. pan speed
- 90° ( $\pm 45^\circ$ ) Tilt movement range

- .1° to 4.5°/sec. tilt speed
- RS232/422 embedded control
- 24 VDC motors-variable speed

The P&T needs to be controlled such that the subject's face is centered in the field of view of the camera. The P&T controller has no software API, but has a well documented serial protocol with a rich feature set. The P&T control software implements the serial protocol, and wraps the protocol with some higher level commands.

### ***Software Description***

**Protocol** – The messaging protocol was implemented initially with a “keep-alive” message and the processing of the standard status response.

#### **Messages:**

- Pan/Tilt Speed – Set the pan speed and the tilt speed
- Move Relative – Move in the pan and tilt axes the commanded degrees relative to current
- Move Absolute – Move in the pan and tilt axes to the commanded absolute position.

**Status Handling** – A callback was implemented where all the errors are identified and handled.

### ***Pan and Tilt Control***

To control the P&T, a linear control loop is implemented. If a face is found in the WFOV image, the commanded pan and tilt speeds are linearly related to  $dx$  ( $x$  distance of the face from the center of field of view) and  $dy$  ( $y$  distance of the face from the center of field of view). In this way, the P&T is controlled to center the face in the camera's field of view. Note that  $dx$  and  $dy$  must be greater than a parameterized threshold to trigger any movement. If a face is not found in the image in a single iteration, the last commanded velocity is not updated. If no face is found in the image for a parameterized number of tries, then the P&T is commanded to stop. The commanded speed of the P&T can be updated at up to an 8 Hz rate. The WFOV camera captures images at up to 30 Hz, and the face finder can find faces in these images at an execution time of 3-10 ms.

### ***Pan and Tilt Analysis***

This algorithm centers slow moving faces at ranges greater than 5 m without issue. However, the tracker will lose lock for faces that change lateral direction suddenly while walking, or that rotate to a pose for which the both eyes are not visible in the WFOV tracking camera's captured image.

### ***Pan and Tilt Opportunities for Improvement***

More complex control algorithms might aid in smoother and faster control of the P&T. For example, to better estimate future coordinates and the motion of the face, a history of the recorded positions could be kept and used to estimate velocity and future position for implementing a Kalman filter tracker. The estimated positions and velocities from the Kalman filter would reduce tracking errors and smooth out “glitches” when a face is not found for several frames in the WFOV camera.

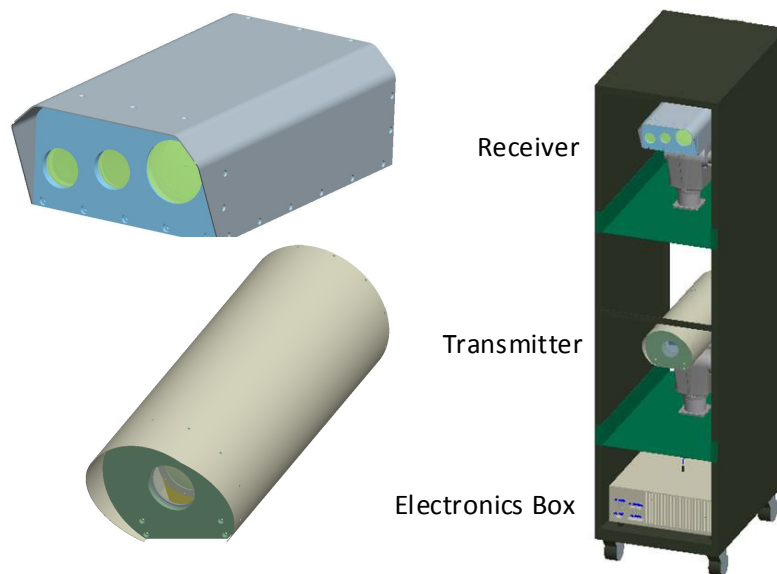
## **6.0 Engineering Demonstration Unit (EDU)**

The design of the Engineering Demonstration Unit (EDU) is derived from that of the breadboard with more robust mechanical packaging, and added fault detection systems and safety interlocks. These enhancements enable us to transport the EDU for safe operation at other locations for demonstrations and tests. Only those aspects of the EDU that differ from the breadboard will be discussed herein.



## 6.1. EDU Mechanical Design

The EDU mechanical design is concerned with three modules: the Transmitter, the Receiver, and the Electronics Control Chassis (aka, Electronics Box). The Transmitter and Receiver may be mounted to various commercial pan and tilt (P&T) platforms through interface plates. The Electronics Control Chassis (ECC) contains power and signal control electronics, and is located within a few meters of the Transmitter and Receiver. A computer in the operator station supplies command and control signals to the ECC, and receives image data from the Receiver. The operator station may be in a remote location up to 50 m from the transmitter, receiver, and electronics box. Figure 14 shows the outer mechanical packaging designs for the transmitter and receiver, and their mounting to Directed Perception DTU-100 pan and tilt (P&T) platforms on shelves in an aluminum frame on casters. The receiver aperture is about 2 m above ground level and about 1 m above the transmitter aperture. The electronics box is mounted on the bottom shelf of the aluminum rack.



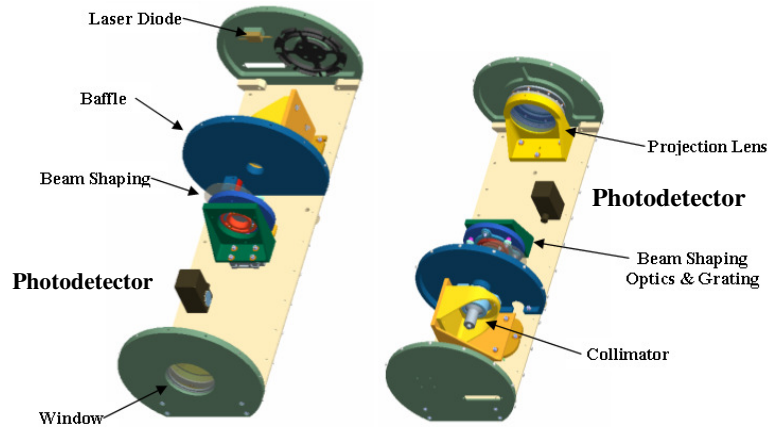
**Figure 14.** The outer mechanical packaging designs for the receiver (upper left) and the transmitter (lower left), and their mounting on P&Ts with the electronics box in an aluminum rack (right).

The two P&T's are controlled by software so that both the transmitted beam and receiver cameras' fields of view are centered on the subject's face as the system tracks the subject. Note that in production installations, the transmitter and receiver would each be separately installed on their P&Ts in fixed positions on walls, ceilings, posts, or other structures depending on the particular application and site. The electronics box would likely be hidden from view in a wall or ceiling for a production installation.

### 6.1.1 Transmitter Mechanical Design

The interior mechanical design for the EDU transmitter is shown in figure 15. As for the breadboard, an 8-W, 808-nm wavelength diode laser is fiber-coupled to a fiber collimator whose output illuminates a sinusoidal transmission grating through beam shaping optics. The grating is mounted to a high precision translation stage for focus adjustment. The objective lens images the laser light transmitted through the grating onto the subject to be imaged by the receiver. A window protects internal components from dirt and damage. A baffle prevents non-eye-safe light from a broken fiber from reaching the output window. The photodetector detects scattered laser light to measure the laser pulse duration, pulse repetition rate, and pulse amplitude, so that if any of these laser pulse parameters exceeds eye-safety limits, the laser will be powered off.

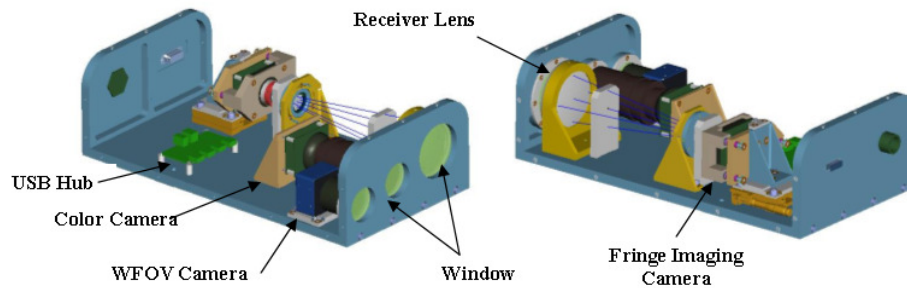




**Figure 15.** Two views of the interior mechanical packaging design for the EDU transmitter.

### 6.1.2 Receiver Mechanical Design

The interior mechanical design for the EDU receiver is shown in figure 16. There are three cameras in the Receiver: a wide field of view camera, a color camera with zoom lens, and a NIR fringe imaging camera. The fringe imaging camera uses a single lens to focus the image of the subject illuminated by the sinusoidal fringe pattern onto the CCD array. This camera is mounted to a high precision translation stage for focus adjustment. A USB hub reduces system wiring. Windows protect internal components.



**Figure 16.** Two views of the interior mechanical packaging design for the EDU receiver.

## 7.0 Future Work

After the EDU assembly is completed, the EDU's performance will be measured in laboratory testing. Once the EDU's performance has been verified to meet the program goals, the EDU will be transported to test sites for data collections and demonstrations. The 2D+3D face imagery data collected using the EDU will be entered into pose estimation, pose rotation, and face recognition software to demonstrate the ability to rotate the 2D+3D face imagery to the corresponding pose (usually frontal and/or profile) in 2D face image databases for improved robustness of 2D face recognition with minimally cooperative subjects for backward compatibility with existing 2D face image databases. The 2D+3D face imagery data collected using the EDU will also be entered into a database for testing 3D only and multi-modal 2D+3D face recognition software, with minimally cooperative subject data under uncontrolled lighting conditions. If sufficient funding is available, in parallel with EDU demonstrations, engineering design studies for the follow-on production prototype will be undertaken 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.

## 8.0 Summary

Lockheed Martin Coherent Technologies is developing a potentially low-cost, Fourier transform profilometry (FTP) based active 2D+3D face imaging system for stand-off biometric identification of minimally cooperative subjects. 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. The technique is extendable to wide area surveillance applications when coupled with a wide field of view acquisition camera, tracking system, and rangefinder.

The FTP breadboard experimental results presented above show that the breadboard is close to meeting the system goals of 3D face capture at >10-m stand-off distance, <1-mm lateral resolution, <1-mm range precision, and real-time capture and processing for walking speed, minimally cooperative subjects using technologies and techniques that will be low cost in production. The Engineering Development Unit (EDU) is based on the FTP breadboard design revised so that the EDU will meet or exceed the system goals, and will be easily transported to test sites for data collection and demonstrations.

## 9.0 Acknowledgements

The work reported herein was funded by Lockheed Martin Corporation under the Ladar Strategic Technology Thread (STT) internal research and development (IRAD) program.

## 10.0 References

- <sup>1</sup> 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.
- <sup>2</sup> 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.
- <sup>3</sup> Bedrosian, Edward, "The Analytic Signal Representation of Modulated Waveforms," *Proceedings of the IRE*, October 1962, pp. 2071-2076.
- <sup>4</sup> 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.
- <sup>5</sup> Kreis, Thomas, "Digital holographic interference-phase measurement using the Fourier-transform method," *JOSAA*, Vol. 3, No. 6, June 1986, pp. 847-855.
- <sup>6</sup> 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.
- <sup>7</sup> 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.
- <sup>8</sup> 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.
- <sup>9</sup> Felsberg, Michael and Gerald Sommer, "The Monogenic Signal," *IEEE Trans. Signal Proc.*, Vol. 49, No. 12, December, 2001, pp. 3136-3144.
- <sup>10</sup> Takeda and Mutoh, op. cit., p. 3977.
- <sup>11</sup> Goodman, Joseph W., Speckle Phenomena in Optics, Theory and Applications, Englewood, CO: Ben Roberts & Company, 2007, p. 298.

- 
- <sup>12</sup> Jelalian, Albert V., Laser Radar Systems, Boston: Artech House, 1992, p. 45.
- <sup>13</sup> Takeda and Mutoh, op. cit., p. 3978, figure 1.
- <sup>14</sup> Mao, Xianfu, Wenjing Chen, and Xianyu Su, "Improved Fourier-transform profilometry," *Applied Optics*, Vol. 46, No. 5, 10 February 2007, pp. 664-668.
- <sup>15</sup> Takeda and Mutoh, op. cit., p.3980.
- <sup>16</sup> Ibid.
- <sup>17</sup> Chen, Wenjing and Xianyu Su, "A new method of improving the measuring precision of 3D shape," *Proc. SPIE* 4231, 2000, pp. 412-415.
- <sup>18</sup> Takeda, Ina, and Kobayashi, op. cit., pp. 156-157.
- <sup>19</sup> Bertalmio, M., G. Sapiro, V. Caselles, C. Ballester, "Image Inpainting," *SIGGRAPH 2000* (Proceedings of the 27<sup>th</sup> annual conference on computer graphics and interactive techniques), ACM 2000, pp. 417-424.
- <sup>20</sup> Ghiglia, D. and M. Pritt, Two-dimensional phase unwrapping: theory, algorithms, and software, Wiley, 1998.
- <sup>21</sup> 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.
- <sup>22</sup> 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.
- <sup>23</sup> Intel Corporation. Performance Benchmarks for Intel® Integrated Performance Primitives, 2009, [http://cache-www.intel.com/cd/00/00/21/93/219360\\_wp\\_ipp\\_benchmark.pdf](http://cache-www.intel.com/cd/00/00/21/93/219360_wp_ipp_benchmark.pdf).
- <sup>24</sup> Bioucas-Dias, J.M., and G. Valadao, "Phase Unwrapping via Graph Cuts," *Image Processing*, *IEEE Transactions on*, Volume 16, Issue 3, March 2007, pp. 698-709.