Institute for Homeland Security Solutions
Applied research • Focused results

Fast 3D Scanning for Biometric Identification and Verification

June 2011

Authors

Anselmo Lastra, University of North Carolina at Chapel Hill
Henry Fuchs, University of North Carolina at Chapel Hill
Peter Stein, University of North Carolina at Chapel Hill
Greg Welch, University of North Carolina at Chapel Hill
Ali Farsaie, SIS, Incorporated
Will Austin, SIS, Incorporated
Ping Fu, SIS, Incorporated
Alex Chen, SIS, Incorporated

Prepared for:

Department of Homeland Security
U.S. Department of Homeland Security
Washington, D.C. 20528

www.dhs.gov

Contract No. HSHQDC-08-C-00100

Prepared by:

RTI International–Institute of Homeland Security Solutions

Research Triangle Park, North Carolina

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Lastra, A., Fuchs, H., Stein, P., Welch, G., Farsaie, A., Austin, W., Fu, P., & Chen, A. (2011). *Fast 3D Scanning for Biometric Identification and Verification*. (Prepared by RTI International–Institute for Homeland Security Solutions under contract HSHQDC-08-C-00100.)

This report is based on research conducted under the Institute for Homeland Security Solutions (IHSS) under contract to the Department of Homeland Security, Washington, DC. (Contract HSHQDC-08-C-00100). The findings and conclusions in this document are those of the author(s), who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Homeland Security. Therefore, no statement in this article should be construed as an official position of the Department of Homeland Security.

Table of Contents

List of Figures

List of Tables

In this research project, which ran from June 1, 2010, to May 31, 2011, we addressed two major issues holding back the adoption of 3D scanning of faces for biometric identification and recognition. This document is the final report, presenting our findings.

Our team's intent under this Institute for Homeland Security Solutions (IHSS) program was to tackle two major problems, the disturbing flashing caused by structured light (the leading 3D acquisition method), and the slow speed of acquisition. We proposed to

- develop practical ways to generate *imperceptible structured light*, a University of North Carolina (UNC) invention that projects patterns and their inverse so rapidly that the subject perceives only white light; and
- employ commodity graphics processors as computational engines to make the 3D model extraction as rapid as conventional 2D digital photography.

The researchers at UNC primarily worked on the first item, while those at SIS worked on the second. The two groups had regular discussions and meetings to enable testing of the imperceptible structured-light projector by SIS to see how it would integrate into the SIS Snapshot 3D scanner. The projector developed at UNC was lent to SIS over the Christmas and New Year holidays (a time when universities generally close) for testing. Feedback from SIS led the UNC team to change the research to look at two problems with the projector: flicker was perceptible when projecting the grey-scale SIS patterns, and the illumination level was low, which reduced the signal-to-noise ratio, thus making it harder to identify the patterns.

In the following sections we describe the work on the projector, detail the computational acceleration, and report on the projector evaluation at SIS. As an appendix, we include a research brief that describes methods of 3D scanning, and contrasts them for the application of face scanning. Readers who are not familiar with the technology should begin by reading the appendix.

Imperceptible Structured-light Projector

Background on Projection

The concept of imperceptible structured light (Fuchs et al., 1999) is to, for a very short period of time, project a pattern from which image correspondences can be extracted, while triggering one or more cameras. Then, very rapidly, the inverse of that pattern is projected. If these patterns are projected rapidly enough, a human will not perceive them, but will instead see the combination.

How rapidly these patterns have to be flashed depends on a number of factors—the rate is related to the *flicker fusion threshold*, the rate at which a flashing picture will appear to a human to be on steadily. Both film-based movies and television take advantage of this perceptual phenomenon. The images are flashing rapidly, but appear solid to a human.

The actual threshold is purely statistical and varies among individuals; some people are more sensitive to flicker. For example, the 50 Hz television refresh rate used in Europe is very objectionable to some. The threshold also depends on illumination level; flicker is more perceptible with brighter imagery. If a computer CRT monitor appears to flicker, the effect will disappear when the brightness is reduced. The portion of the eye used to view the image also determines the threshold—the rods are more sensitive, and the cones less so. Therefore, flicker is more perceptible in the peripheral vision than when looking directly at the source. Motion, either of the image or of the observer, affects perception of flicker. Most of us have experienced this effect with TV when we move our eyes very rapidly (as when we sneeze).

Because the U.S. TV refresh rate is 60 Hz, we expected that we might have to refresh our patterns at a faster rate because of the brightness of the illumination. Furthermore, the patterns can have a very complex form, which also affects how quickly we need to refresh.

Projectors of the type marketed as Digital Light Projectors, or DLP (developed by Texas Instruments [TI], which also holds a trademark on the term) have some characteristics that make them the best choice for generating imperceptible structured light. DLP projectors use micro-electro-mechanical devices known as Digital Micro-mirror Devices. These are microelectronic chips with an array of tiny mirrors on the surface. Each mirror represents a pixel on the final image, and can be pointed either toward the projection screen or away from it. Typical mirror arrays are sized 1024 by 768 and up for computer use, and 1920 by 1080 for television use.

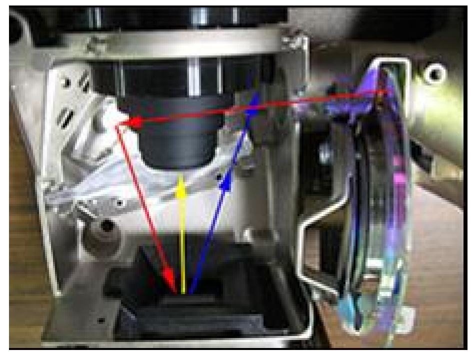
When the chip is illuminated properly and the tiny mirror is pointed toward the screen (through a lens), we see a pixel turned on, white, say. Alternatively, the mirror can be pointed away (toward a black light sink to absorb the light and dissipate any generated heat) to generate a black pixel.

To get levels of gray, the mirrors are flashed over time (taking advantage of the human perceptual characteristics described above). The ratio of on to off determines the grayscale level. A typical 8-bit grayscale level can display a total of 256 values. So, for example, a mirror may potentially flip 256 times per frame. If all 256 times point toward the screen, we see white. If the mirror is only pointed to the screen half the time, we can display middle gray, a pixel value of 128. So for this simplistic example, we need to be able to flip the mirror 256 times 60, or 15,360 flips per second.

The display of color images is not relevant for this application, but for completeness we will describe the way it is handled. There are three common approaches. The one used on only the most expensive projectors (those used in movie theaters, for example) is to use three DLP chips, each illuminated by a different color of light—red, green, or blue. The most common approach is to use a color wheel through which the light shines on the DLP, and display each of the three primary colors in sequence over time. This is illustrated in Figure 1. One other approach, currently used on the small, ultraportable projectors known as pico projectors, is to illuminate a single DLP with three colored LEDs (one red, one green, and one white). This works just like the color wheel in that the primary colors are generated one at a time, over the time of a full frame. To support color, the DLP chips must flip tens of thousands of times per second.

Figure 1. DLP Light Path

This image illustrates the light path in a color DLP projector. Light enters from the right, through a color wheel. It then bounces off a front-surface mirror (red path) and hits the DLP chip at the bottom. The mirror on each pixel of the DLP can point either toward the lens (yellow arrow) or the light sink (black arrow). (image from Wikipedia)



System Requirements

Structured-light patterns can be purely black and white, or can be grayscale. The latter has some advantages in accuracy, and is what is used by the SIS Snapshot system. As noted above, we need roughly to be able to flip the micro-mirrors at 15 KHz or more to display grayscale. Our experience with off-the-shelf projectors led us to believe that we needed a higher refresh rate than 60 Hz. Therefore, we needed a projector that can flip the mirrors at 20–40 kHz under program control.

We also need to be able to trigger the stereo camera pair so they will capture an image only when the “positive” structured-light pattern is being displayed, and not when the inverse pattern is being shown. Therefore, a trigger signal needs to be generated by the projector, and the cameras need a trigger input.

System Design

Commercial projectors are designed to display a series of images, typically at a rate of 60 per second. They do not enable the degree of control that we need to display patterns at a high frequency. Also, they do not supply a trigger signal. We needed to use what are essentially projector development kits containing the DLP from TI. Normally these are sold to projector manufacturers for development of typical home and office projectors.

We looked at three models. One is an inexpensive DLP Pico Projector Development Kit manufactured for TI by YoungOptics Corporation in Taiwan. This is a fairly sophisticated product, in a molded plastic case, but we knew that a pico projector would likely not produce enough brightness for our needs, nor have enough resolution (it supports only HVGA). However, it only costs $350 and was available from Digikey immediately, so we ordered one to begin development right away.

The second alternative was the DLP® LightCommander development kit, manufactured for TI by LogicPD of Minnesota. It consists of a light engine (an XGA DLP) and an RGB LED light source, packaged with a lens in an enclosure. The instrument also has a trigger signal for an external camera. It might have been very suitable, with a 500 Hz grayscale refresh rate. However, it was not available for shipment in the summer of 2010, and we knew that we needed to obtain one rapidly to demonstrate a prototype at the IHSS Research Summit in November.

The third alternative, the DLP Discovery 4100 Development Kit, was the most complicated and programmable. It was available with several DLP chips, varying in size and resolution. TI does not sell it directly, but rather through third parties that can configure a kit with an illumination source, optics, and controller board. We contacted one of the vendors, Digital Light Innovations (DLi) of Austin, Texas.

Pico Projector. We purchased a pico projector, as described above, for experimentation (Figure 2). In that way, we could begin experimenting while waiting for the expensive, full-featured kit. A very useful application note on use of the projector for (conventional, not imperceptible) structured-light applications is available from TI [2010]. The DLP Pico Projector has a resolution of only HVGA (480 x 320) and produces only 7 lumens. When displaying imperceptible patterns, the illumination is reduced by one half, so the result is quite dim. We could only run experiments in darkness or near darkness.

Figure 2. DLP Pico Projector

A DLP Pico Projector Development kit, shown beside a ballpoint pen to give an idea of the size of the device.

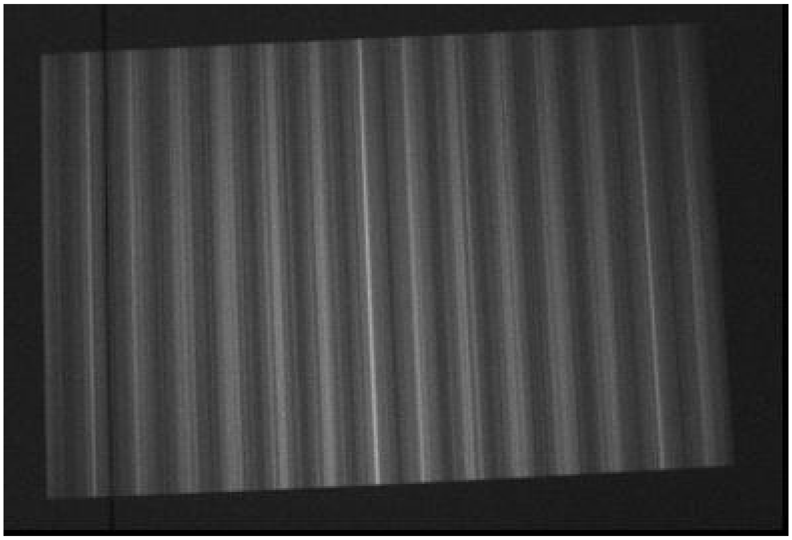
The valuable experience, however, was to be able to begin to experiment with camera triggering. We built a small circuit to take the output trigger of the projector, at 1.8 volts, condition it, and level convert to the required 3.3 volts for the camera. The projector also requires an I^2C connection to program the functions. We used a Beagle Board, sold as an accessory to the projector by Digikey (http://www.digikey.com/pico), to generate the I^2C commands. For a more permanent installation, we would install an I^2C interface board on the PC.

For binary patterns, we could run the projector at a high enough rate to create an imperceptible pattern and image it on the camera. However, the refresh rate was not high enough to imperceptibly generate gray-scale patterns (see snapshot, Figure 3).

DLP Projector Development Kit. We purchased a DLP Discovery 4100 Development Kit from DLi (using other funds because the Department of Homeland Security [DHS] approval process for capital equipment was too elaborate) for approximately $17,000. This kit is not a product, as such. One buys parts, such as the DLP chip, optics, and control electronics, separately to handle particular application scenarios.

Figure 3. Gray-Scale Displayed by Pico Projector

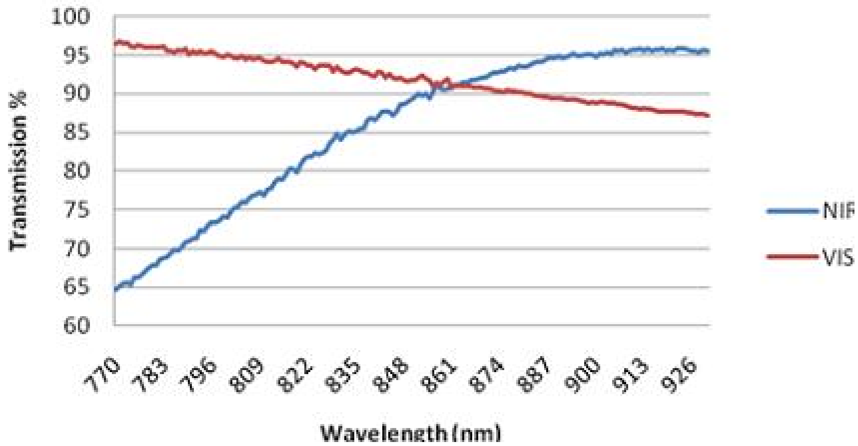
A snapshot of a gray-scale image projected by the DLP Pico Projector. It is low resolution and very noisy (because of low lumens).



We selected a 0.55" visible light, XGA DLP chip. It has some response into the near infrared (usually considered as going from 0.7 to 2 microns). This was important to use because we wanted to be able to experiment with infrared structured-light patterns in the future. Figure 4 shows the transmission of the DLP chip. Other available DLP modules are better for infrared use.

Figure 4. Single Pass Window Transmission

Wavelength-dependent characteristics of the 0.55" visible light, XGA DLP module. Although it is not optimized for infrared, the efficiency is very reasonable.



One of the graduate students (not supported by this project) did run some experiments with an infrared light source. For that project, we bought a machined coupler, as shown in Figure 5, an adapter (see http://www.edmundoptics.com/onlinecatalog/DisplayProduct.cfm?productid=2112), and a fiber-optic LED light source from Edmund Scientific (http://www.edmundoptics.com/onlinecatalog/displayproduct.cfm?productid=3250). Our conclusion is that we needed to invest more time and money to get a bright enough IR instrument.

Figure 5. A Coupler to Enable Fitting of an Infrared LED to the Projector Module

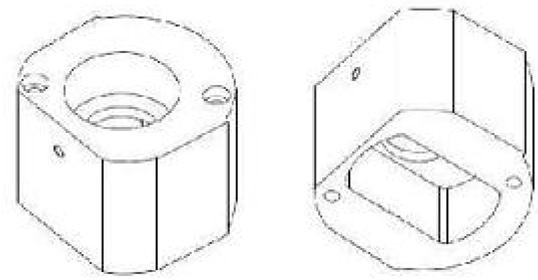
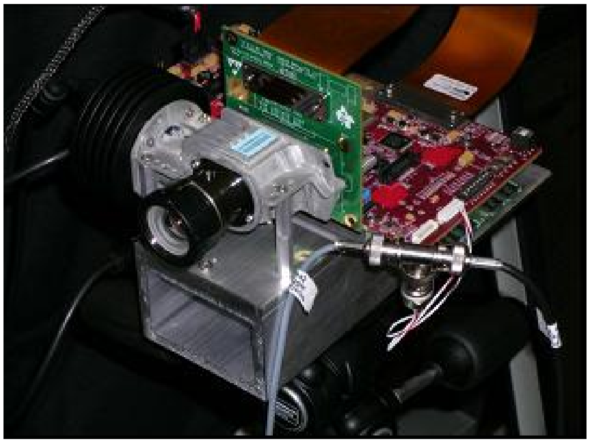


Figure 6. The S3X Optics Module from Digital Light Innovations

Figure 7. Projector Kit Mounting



For visible-light display optics, we selected the S3X Optics Module from DLi (see Figure 6) (http://www.dlinnovations.com/products/s3xomled.html?loco=d4100&name=D4100). It has a small Carl Zeiss projection lens, which focuses from 1.5 to 5 meters. We considered that to be a reasonable range for scanning of human heads. We chose a white LED light source that generated 225 lumens. This was not as bright as we wanted; however, delivery time was a big factor because we wanted to have a prototype ready to display at the IHSS Research Summit.

In some ways, the most important part of the instrument for us was the electronics module. It is critical to be able to control the timing of the micro-mirror flips very precisely. We chose the ALP 4.1 High Speed control module, firmware, and software. The hardware consists of a Xilinx Virtex 5 FPGA (field programmable gate array, a configurable block of digital logic), 32 Gbit of onboard DRAM, and a high-speed interface to the DLP. The package includes various types of software. We mostly used a scripting language that enabled us to precisely control the display of images (patterns) that we had stored in the DRAM. The maximum DLP switching rate is over 22 KHz for a 1-bit pattern, and 291 Hz for an 8-bit pattern. This is constrained by the hardware and software. With faster electronic hardware, the DLP chip itself could run at over 30 KHz.

We rigidly mounted the parts on an aluminum extrusion, which we drilled and tapped to fit onto a tripod (Figure 7). One option that we considered was to machine a combination projector mount and camera bar (a rigid metal bar onto which to mount a stereo pair of cameras). However, we decided to opt for flexibility and use separate mounts for the two components. To preserve calibration, the cameras must be very rigidly mounted. However, we did not plan to do any projector-camera calibration, so the units did not have to be mounted together.

Cameras. Finally, we had to choose a pair of suitable cameras for baseline stereo. Some structured-light scanners use more than two cameras (for example, the 3D Snapshot from SIS) to get more complete coverage. However, for our experimental purposes this was not necessary.

We evaluated a number of possible cameras, and selected the Flea cameras from Point Grey (http://www.ptgrey.com/products/flea/flea.pdf), shown in Figure 8. They have an IEEE 1394 (Firewire) interface, and a Sony monochrome CCD with a resolution of 1024 x 768. Another advantage is that we have a great deal of experience with the FlyCapture software development kit that runs the Point Grey cameras. A key feature of the Flea cameras is the flexible external triggering, which includes a number of options. We also considered a fixed stereo pair of cameras, the Bumblebee or Bumblebee 2 from Point Grey (http://www.ptgrey.com/products/bumblebee2/bumblebee2_stereo_camera.asp). An advantage of this camera is that a pair of lens/sensor units is built into a single housing and calibrated at the factory. A problem is that the stereo baseline is fixed at 12cm, and did not match that used by our research partners at SIS. Ultimately, however, we decided not to use it because the delivery time was too long and would not enable us to have a prototype ready for the IHSS Research Summit.

Figure 8. Flea Camera from Point Grey

Figure 9. Camera Bar with Twin Flea Cameras

We built a metal camera bar with a number of holes drilled in it (Figure 9) so we could change the camera baseline over a relatively wide range. The SIS Snapshot system uses a wide baseline, whereas our own software is set up for a smaller baseline. It is also important to use lenses with focus and aperture locking screws, and to secure the cables rigidly because it is crucial that the cameras not move relative to each other after they have been calibrated.

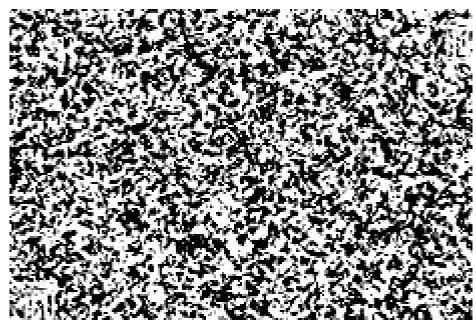
We calibrated the camera pair using an in-house calibration package, which uses the method of Zhang (2000).

Prototype Demonstrated at IHSS Research Summit

On November 4, 2010, we demonstrated a prototype at the IHSS Research Summit. Although we had intended to use SIS Snapshot software to extract a 3D model, the time was too short from when we obtained the projector to November. Therefore, we used some UNC stereo matching software.

Instead of a gray-scale pattern, we used a random pattern (Figure 10) from which the software extracts features, which are then used for triangulation between the left and right frames. This is a more general method, and is from software developed under contract to DARPA for automated 3D modeling of cities by a truck driving the streets, or by a helicopter or airplane over flight.

Figure 10. Pattern Used for Structured Light at IHSS Research Summit



A pair of images captured at the summit is shown in Figure 11. We used external trigger mode 5 on the Flea cameras. This enables us to trigger the cameras (simultaneously) multiple times and accumulate an image over multiple triggers so we could flash the imperceptible pair very rapidly, but capture an image over a longer interval, say 1/30th to 1/60th of a second. Typically we capture tens of flashes, each of a millisecond or less duration, to generate an image. No one at the meeting was able to detect any flashing whatsoever.

Integration Experiment and Subsequent Improvements

After the demonstration at the IHSS Research Summit, we concentrated on changes that were necessary to try the projector prototype with the SIS Snapshot software. Because the projector was in use for multiple projects at UNC, we decided to lend it to SIS over the university Christmas and New Year holidays. To prepare, we built a circuit that enabled us to trigger the cameras for a gray-scale pattern (previously we did not need that feature).

Figure 11. Stereo Pair with Random Dot Structured-Light Pattern

A pair of images with a random dot pattern. To a human observer, these looked like white light. The cameras were triggered to capture a single pattern of the positive/inverse pair.

Ping Zhuang from SIS used the projector for almost a month, and adapted it to the Snapshot setup. His report is included below. A short summary follows:

- Four types of objects were scanned to test results under various conditions. The results showed that the integration of imperceptible light with SIS's structured-light pattern and 3D imaging system is possible, yet additional enhancements will be necessary to optimize performance and achieve the same quality 3D imaging that SIS's current system yields. Major findings include the following:
  - The structured-light pattern was still slightly perceptible to the human eye, suggesting that the system and frequency will need to be fine-tuned further.
  - A brighter lamp will be necessary in the future to increase the dynamic range of SIS's structured-light pattern to also increase the accuracy of the captured 3D data.
  - Ambient light thus had a strong influence on the capture data; thus, the system had to be used as configured in a dark room with limited depth of field (200mm).

In response to the January findings, when we encountered perceptible flicker when projecting the grey-scale structured-light patterns, we looked for solutions. Flicker is perceptible mainly during saccade, and depends on the frequency at which the light is displayed. In February, we were able to reorder the DLP mirror flips to increase the frequency above the perceptible threshold. We took advantage of trigger mode 5 of the Point Grey cameras to generate our own gray-scale pattern using binary patterns (as a typical projector might do) but reordered to show a single binary pattern followed by an inverse pattern. The camera essentially integrated the gray scale over 256 captures of the positive pattern.

Doing this prevented us from using the camera synchronization scheme that we had devised, however, so we would randomly get either the positive pattern or the negative pattern. When integrated, this would yield white. To fix this, we designed a camera trigger circuit using a field programmable gate array (FPGA). Although we needed very few gates, an FPGA gave us an easy way to deal with the fact that the projector supplies 2.5 volt signals, while the Flea camera trigger input uses 3.3 volt logic.

Once we had the projector image back to being imperceptible, we spent the final period of this project on ways to get a better image in low light. For that, we only have very preliminary results. We plan to continue working on this in the future.

We could, of course, use a brighter lamp on the projector prototype, which is what we would do if we were scanning objects. However, for humans this is not the best option because it is uncomfortable to have an extremely bright light shining on people. It is not only disturbing, but it can also feel very warm.

Figure 12. Large Sensor Camera

So we instead used a Canon CMOS camera sensor with a larger area per pixel. The Flea cameras that we used earlier have a pixel size of 4.65 um per side. The image shown in Figure 12 was made with an imager with a pixel size of 8.2 µm, almost 4 times the sensor area. The pattern was on for only 14 milliseconds, so we could improve the result by integrating for a longer period. However, because we are ultimately interested in identifying people as they are walking, the exposure time needs to be short.

SIS's Final Report to UNC Chapel Hill

Summary and Findings for GPU-Accelerated Data Processing

SIS installed the 3D Snapshot software code, which is written in C code, on an Intel dual core T7400 with 2.16 GHz CPU. We conducted data capture of several test objects and recorded the processing time required to generate 3D models.

Next, SIS used the GPU parallel programming technique (e.g., NVIDIA's CUDA technology) to convert the code for the image processing algorithm. GPU parallel programming has succeeded in increasing processing speeds significantly in other tests. For this experiment, SIS used NVIDIA's GeForce GTX 295 video card, with CUDA compute capability of 1.3, running the CUDA version of the code on the same PC.

During conversion of the source code from C to CUDA, we noticed that portions of the C program could not be converted to parallel computing because of the nature of the original algorithm. Ultimately, we decided to leave those portions in C code (i.e., sequential computing, in the CUDA version software).

Table 1 shows results of the processing speed comparison for algorithmic functions associated with 3D data processing. For some functions, we experienced a gain of 40 to 65 times processing speed using CUDA. Overall, we were able to increase processing speed using CUDA by 29.5 times versus C code.

Table 1. CUDA versus C Speed Comparison

CUDA & C Speed Comparison	C (sec)	CUDA (sec)
Gamma correction	0.3096	0
Normalization	0.0686	0
Gaussian filtering	0.4028	0.0155
Stereo matching	2.5158	0.0626
First data cleaner	0.0207	0.0207
First time seeds expand preparation (CUDA)	0	0.0158
Second time seeds expand preparation (CUDA)	0	0.0122
Second data cleaner	0.0257	0
Column and median filter	1.3156	0.02
Average filter	0.066	0.006
Total	4.7603	0.161

We believe the 3D Snapshot software can be further optimized by rewriting original codes to work specifically on GPUs, thus increasing processing speed significantly. Also, there is a new generation of CUDA cards that became available recently (with compute capability of 2.0) that can provide faster processing.

Summary and Findings for Integration of Imperceptible Structured Lighting and SIS's 3D Snapshot

Using the DLP projector kit provided by UNC, SIS set up the hardware identical to the hardware setup for SIS's existing 3D Snapshot system. We used 12.25 inches baseline shown in Figure 13. We also replaced the camera lenses to 12.5 mm focal length lenses, so that the field of view of cameras and lighting would be aligned. Although this configuration would lose some accuracy of image matching, it greatly increases the accuracy of triangulations so that it could maintain the final 3D results at a high level of accuracy.

Figure 13. Projector and Cameras

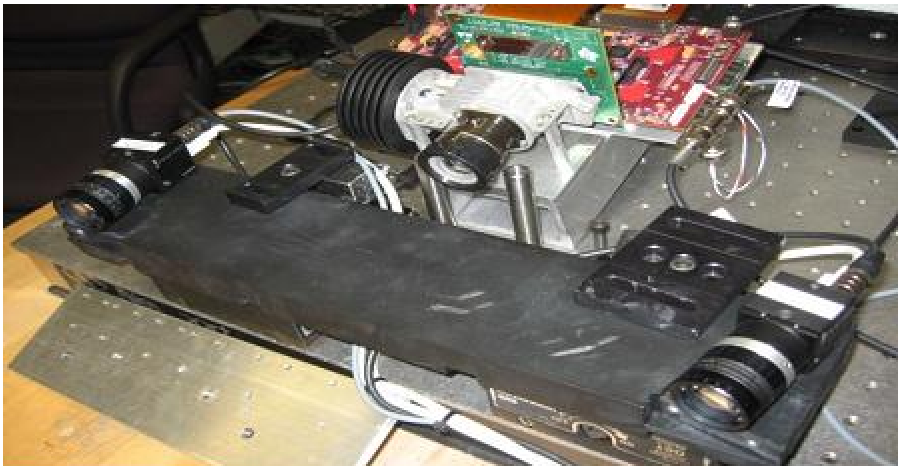


Figure 14 shows the full view of our setup with calibration station.

Figure 14. Calibration Station

With this configuration, the structured-light pattern was noticed during image capture. Future work could involve fine tuning the system so that the light pattern is completely invisible to the human eye. However, a major issue was the low light source, which required the use of a large aperture, resulting in a small depth of field and high influence of ambient light. Even using a large aperture, the captured images were still dark, which resulted in a decrease in signal-to-noise ratio because of the high percentage of dark current noise. The low lighting also resulted in a decrease of the dynamic range of SIS's structured lighting, which will greatly affect the digital accuracy of 3D data.

References

Fuchs, H., Livingston, M. A., et al. (1999). *Dynamic generation of imperceptible structured light for tracking and acquisition of three dimensional scene geometry and surface characteristics in interactive three dimensional computer graphics applications*. U.S.P. Office. University of North Carolina at Chapel Hill.

Texas Instruments Corporation. (2010). *Using the DLP Pico 2.0 kit for structured light applications*.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 1330–1334.

Appendix

Institute for Homeland Security Solutions
Applied research • Focused results

June 2010

3D Scanning for Biometric Identification and Verification

Project Leads

Anselmo Lastra, PhD, University of North Carolina at Chapel Hill

Henry Fuchs, PhD, University of North Carolina at Chapel Hill

Greg Welch, PhD, University of North Carolina at Chapel Hill

Ali Farsaie, PhD, SIS, Inc.

Statement of Problem

Reliable and robust identification and verification of individuals is critical to homeland security applications such as surveillance, authorization for entry to secure areas, and passport identity verification. Traditional biometrics, such as mug shots, fingerprints, and voice recognition, have been used with some success. However, they exhibit serious disadvantages for some tasks. These three biometrics, for example, are problematic for surveillance (identification); even the traditional mug shot is difficult to use in automated surveillance applications because many factors, such as lighting and frontal visibility, cannot be controlled.

A relatively new biometric, 3D facial recognition, holds great promise. Although the technology is nascent, in a comprehensive 2006 study (Phillips et al., 2007) recognition performance using 3D shape and texture matched that of the much more mature technologies of high-resolution image recognition (which featured controlled lighting) and iris recognition. Additionally, 3D modeling promises to enhance recognition performance because it can be used to recognize people in profile as opposed to a typical forward-looking, mug-shot pose. Even when using 3D to match to a mug shot, an advantage is that a 3D model allows one to render a view of the person from any desired perspective—the pose, distance, and even lighting can be factored into the rendering to match any photos.

Scenarios in which 3D recognition could be profitably used include (a) verification of identity at an airport (for example the subject's face could be rapidly scanned while his or her smart-card ID is being examined, and the system could then match the scan with data on the ID); (b) identification at a secure site or even at an airport while people are walking down the hallways or standing in line; or (c) 3D pose extraction of a moving subject, thereby potentially enhancing recognition performance and enabling intent analysis.

This brief presents the technical background of the 3D scanning technologies, briefly surveys related biometrics that may be combined with 3D recognition, provides an overview of the major technical issues, and highlights research opportunities to overcome those issues.

Background

Probably the most studied technology for 3D modeling is baseline stereo vision. Two or more cameras image the scene, and corresponding points are selected in the images. If the cameras are calibrated (camera position and orientation, as well as lens and imager characteristics), the correspondences can be used via triangulation to determine the distance, and thus the geometry, of the visible structures in the scene.

A major problem is to determine the correspondences. A scene with very uniform color, such as white walls, is clearly problematic. If the scene is highly textured, however, then correspondence points may be extracted automatically. Stereo imagining techniques fall into two categories, intensity matching and feature detection (Faugeras, 1993; Forsyth & Ponce, 2002), with the latter having proven more reliable. Stereo reconstruction may also be performed from video sequences (Pollefeys & Gool, 2002; Pollefeys et al., 2004). The problem of accurately finding correspondences, however, has proven to be difficult and not always robust, leading researchers to investigate active approaches, primarily laser scanning and structured light.

Laser Scanning

When used for faces, human bodies, or other objects at short distances, triangulation is typical employed. A laser stripe scanned across the subject essentially provides correspondences for the camera(s). Cyberware makes a well known laser scanner of this type that has been extensively used in the movie industry. Unfortunately, this scanning technique takes anywhere from seconds to minutes—not a problem for scanning a seated and supported actor's face, but prohibitively long for identification purposes. Some laser techniques project complex patterns using interference of two beams, essentially the structured light technique described at the end of this section.

Another laser-scanning technique uses time of flight (the time for illumination to travel to and from a surface, divided by the speed of light) to determine distance. This is also known as LIDAR. Typically this method is used for longer ranges. Some new devices, such as the Swiss Ranger (from Mesa Imaging, http://www.mesa-imaging.ch) and Canesta (http://canesta.com) cameras work at ranges of a few meters and at video rates. However, their low resolution—160 x 120 pixels (Canesta) and 176 x 144 pixels (Swiss Ranger)—makes them unsuitable for biometrics. The marketing focus for these devices seems to be in vehicle safety applications (backup alarms for cars, for example) and human-computer interaction (potentially for video games).

Structured Light

The second general approach, *structured light*, is very similar to laser triangulation except that a light projector is typically used to project a pattern onto the subject. This provides a rich field of correspondences across the subject that can be used to extract a 3D model from the camera images. The use of time-multiplexed coded structured light patterns was first proposed by Posdamer and Altschuler (1982) and has sparked a great deal of research. Typically a small number of patterns are projected in sequence and the result imaged. Monochrome cameras can be used to capture geometry, and a color camera to add texture. This is the technology used by the 3D Snapshot system from SIS, Inc. The following sections focus on structured light (since it is the most suitable for human-subject scanning) and examine the challenges as well as possible research directions.

Issues

- The process should not disturb the subject. A major problem with conventional structured light approaches is that the rapidly flashing patterns are uncomfortable for the people being scanned. There may also be situations in which it would be important to scan a subject without his or her knowledge.
- Speed of capture is critical for any moving subject, especially for human biometrics. Many systems take less than a second (0.3 seconds for 3D Snapshot) to scan, but humans can move significantly in that time. An ideal scan duration would be from 1/10th to 1/30th of a second.
- Speed of processing is also important. The result must be available within a second or two. Ideally, the processing could be done at real-time rates in order to generate 3D at video rates.
- Accuracy is a major issue, of course, especially under less-than-ideal lighting and environmental conditions.
- The scanner should have a reasonably wide field of view so the subject does not have to be in a very precise location. Analogously, the scanning device should have reasonable depth of field.

- Eyeglasses are a problem because of reflections from, and refraction through, the lenses.
- Geometry of hair can be difficult to capture, and a beard can also be used to hide features.

Research Directions

Research directions in this section are proposed in priority order, based on the importance of the problem to be solved, as well as the amount of time expected to develop a technical solution.

Imperceptible Scanning

The authors see two fruitful technical directions to make the scanning process invisible to the subject. The first, *imperceptible structured light*, was invented at the University of North Carolina at Chapel Hill (Raskar et al., 1998) to enable 3D modeling of persons for 3D video conferencing applications. The key idea of imperceptible structured light is to flash a pattern and its inverse rapidly enough that it will appear to the subject as white light. A fast camera can be synchronized to the projector and will capture an image of the pattern. Most of the work in this area has been to calibrate projector systems shining on non-planar environments (Cotting, Naef, Gross, & Fuchs, 2004; Cotting, Ziegler, Gross, & Fuchs, 2005; Zollmann & Bimber, 2007). Although the authors have demonstrated the concept, many challenges remain with the hardware implementation.

The other potential approach is to use infrared illumination. Infrared may be imaged directly (essentially to detect skin temperature) (Abayowa, 2009; Colantonio & Benvenuti, 2007), or infrared patterns can be projected, much as with visible light. There has been little work on infrared structured light. The authors know only of a bench prototype tested in Japan (Akasak, Sagawa, & Yagi, 2007).

Speed

Two factors account for the time required for a scan: acquisition and processing. Carefully synchronizing the camera with the projector, such as the authors have done with their prototypes (Cotting et al., 2004; Cotting et al., 2005; Raskar et al., 1998), can make the image acquisition process faster. However, imaging in a shorter amount of time, or with less light, tends to make sensor noise more problematic, and this should be combated using techniques such as those of Bennett and McMillan (2005). To make the processing faster, the authors can use the graphics processing unit, an approach they pioneered (Harris, Coombe, Scheuermann, & Lastra, 2002) that is now becoming popular. Speed increases of 20 to 40 times are possible.

Improved Biometric Accuracy

It is possible to combine multiple biometrics, with the resulting biometric fusion potentially increasing accuracy. A promising approach may be to combine iris/retinal scanning with 3D scanning. The texture of the human iris forms during the gestational period, and it exhibits a great deal of detail, including furrows, freckles, and other features (Daugman, 2004). The iris can be imaged unobtrusively, and the near-infrared modality used brings out patterns even in persons with dark pigmentation. Because imaging of the iris requires cooperation from the subject, however, it may be less useful for identification from surveillance imagery (Abayowa, 2009). A survey of techniques is presented in Bowyer, Hollingsworth, & Flynn (2008).

Field of View and Depth of Field

The ability to capture 3D models of people over a wide working area will provide a very powerful biometric tool. This is a very difficult problem, however. For the hardware part of the solution, the authors propose overlapping, synchronized structured light projectors and a set of cameras. The prices are dropping rapidly for both of these devices, so cost is not the primary barrier.

This net of projectors and cameras could be coupled with software algorithms for a progressive refinement of the biometric over time. For example, the scanning might occur as people are standing at the line waiting for the TSA screening. Even if there is no wait, just the walk through the cordoned area could serve.

A potentially powerful strategy is to combine structured light approaches with extraction of correspondences for a combined modeling approach. The longer observation time allowed in the screening-while-walking scenario can be used to improve the models by predicting the subject's motion and tailoring the imperceptible structured-light patterns to improve the model.

Extraction of Subject Pose and Posture

The way a person walks is a very characteristic identifier for recognizing someone. Furthermore, pose and posture analysis could be used to analyze intent in certain situations. The authors have been working with the Navy to estimate the posture of Marines during training and using the posture to analyze their performance. Because a multitude of views is necessary, this work is using multiple video cameras. Structured light would be a very useful enhancement that is not possible for the outdoor Marine-training scenario.

Contact Information

Anselmo Lastra
Sitterson Hall, CB3175, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3175
919-962-1958
lastra@cs.unc.edu

Anselmo Lastra, PhD, is a Professor and Chairman of Computer Science at the University of North Carolina at Chapel Hill. He received a BS in electrical engineering from the Georgia Institute of Technology and MS and PhD degrees in computer science from Duke University. He was co-chair of I3D 2005 and Graphics Hardware 2000 and 2004 and is an associate editor of *IEEE Computer Graphics and Applications*. His research interests are in the areas of image-based 3D modeling and rendering and graphics hardware architectures.

Henry Fuchs, PhD, is the Federico Gil Professor of Computer Science at the University of North Carolina at Chapel Hill. He has been active in computer graphics since the early 1970s, with rendering algorithms (BSP Trees), hardware (Pixel-Planes and PixelFlow), virtual environments, tele-immersion systems, and medical applications. He is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, and the recipient of the 1992 ACM-SIGGRAPH Achievement Award, the 1992 Academic Award of the National Computer Graphics Association, and the 1997 Satava Award of the Medicine Meets Virtual Reality Conference.

Greg Welch, PhD, is a research associate professor of computer science at the University of North Carolina at Chapel Hill. In 1986 he received a degree in Electrical Technology from Purdue University (with Highest Distinction), and in 1996 he received a PhD in computer science from the University of North Carolina at Chapel Hill. Previously he has worked at NASA's Jet Propulsion Laboratory and Northrop-Grumman's Defense Systems Division. His research interests include human-tracking systems, 3D telepresence, projector-based graphics, computer vision and view synthesis, and medical applications of computers. He has co-authored over 50 peer-reviewed publications in these areas, is a co-inventor on multiple patents, and maintains an internationally recognized Web site dedicated to the Kalman filter. He is a member of the IEEE Computer Society and ACM.

Ali Farsaie, PhD, is president and CEO of Spatial Integrated Systems, Inc. (SIS). He provides strategic guidance and development for new program activities within defense, aerospace, Federal agencies, and commercial activities. Governor Perdue recently appointed him to North Carolina’s Economic Development Board. Dr. Farsaie received MS and PhD degrees in 1979 from North Carolina State University. Prior to SIS, Dr Farsaie was a chief engineer at the Naval Surface Warfare Center Dahlgren Division. He conducted, formulated, and managed research and development of novel approaches in advanced information technology, sensors and robotics, training, and system engineering to meet long-term Navy mission requirements.

References

Abayowa, B. O. (2009). Thermal infrared exploitation for 3D face reconstruction. *Proceedings of SPIE - The International Society for Optical Engineering, SPIE-7347*.

Akasak, K., Sagawa, R., & Yagi, Y. (2007). *A sensor for simultaneously capturing texture and shape by projecting structured infrared light.* Paper presented at the 3DIM.

Bennett, E. P., & McMillan, L. (2005). *Video enhancement using per-pixel virtual exposures*. Paper presented at the ACM SIGGRAPH.

Bowyer, K. W., Hollingsworth, K., & Flynn, P. J. (2008). Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding, 110*(2), 281-307.

Colantonio, S., & Benvenuti, M. (2007). Object tracking in a stereo and infrared vision system. *Infrared Physics & Technology, 49*(3), 266-271.

Cotting, D., Naef, M., Gross, M., & Fuchs, H. (2004). *Embedding imperceptible patterns into projected images for simultaneous acquisition and display.* Paper presented at the ISMAR.

Cotting, D., Ziegler, R., Gross, M., & Fuchs, H. (2005). *Adaptive instant displays: Continuously calibrated projections using per-pixel light control.* Paper presented at Eurographics.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(1), 21-30.

Faugeras, O. (1993). *Three-dimensional computer vision: A geometric viewpoint*. Cambridge: The MIT Press.

Forsyth, D. A., & Ponce, J. (2002). *Computer vision: A modern approach*: Prentice Hall.

Harris, M. J., Coombe, G., Scheuermann, T., & Lastra, A. (2002). *Physically-based visual simulation on graphics hardware.* Paper presented at the SIGGRAPH / Eurographics Workshop on Graphics Hardware.

Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., et al. (2007). *FRVT 2006 and ICE 2006 large-scale results* (No. NISTIR 7408). Gaithersburg, MD: National Institute of Standards and Technology.

Pollefeys, M., & Gool, L. V. (2002). From images to 3D models. *Communications of the ACM, 45*(7), 50-55.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., et al. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision, 59*(3), 207-232.

Posdamer, J. L., & Altschuler, M. D. (1982). Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*.

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., & Fuchs, H. (1998). *The office of the future: A unified approach to image-based modeling and spatially immersive displays.* Paper presented at the SIGGRAPH98, Orlando, FL.

Zollmann, S., & Bimber, O. (2007). *Imperceptible calibration for radiometric compensation.* Paper presented at Eurographics.