Biometric Recognition based on 3D Face Geometry

M.Sc. Thesis

Maarten Boersma

University of Twente
Department of Electrical Engineering,
Mathematics & Computer Science (EEMCS)
Signals & Systems Group (SAS)
P.O. Box 217
7500 AE Enschede
The Netherlands

Report Number:	SAS004-05
Report Date:	June 1, 2005
Period of Work:	01/09/2004 – 07/06/2005
Thesis Committee:	Prof. Dr. ir. C.H. Slump
	Dr. ir. A.M. Bazen
	Dr. ir. R.N.J. Veldhuis

2

Structured Light System

2.1 Introduction

For the 3D acquisition of the facial surface, several approaches are possible. This chapter starts with a brief overview of commercially available 3D scanners, and then continues with describing a self-made low-cost structured light system. First, the principle is explained; then a light coding strategy is selected. Finally, the hardware used for the implementation is chosen and the choices are motivated.

For the computation of a 3D surface, calibration of the acquisition system is necessary. This procedure is postponed to Chapter 3. Further, for a practical implementation, specific image processing is needed, which is the topic of Chapter 4.

2.2 3D Acquisition Alternatives

For the generation of a 3D-model of the face, several alternative approaches are available. First of all, a commercial 3D imaging device can be bought. An overview of commercially available devices, for which a price estimate was available, is found at `http://www.simple3d.com` [15] and reproduced in Table 2.1. An actual and comprehensive list, including links to the sites of the manufacturers, can be found at the web site mentioned. Not all devices listed in the table might be suitable for face shape acquisition.

As an alternative to one of these commercial scanning solutions, a low-cost structured light system is developed in this project. For the design of the system, two boundary conditions need to be fulfilled:

1. Total costs of the acquisition system should not exceed EUR 500 (PC not included),

2. The system should be able to work with infrared light to capture a facial image without hindering the person.

Manufacturer	Device name	Price in k$
3DScanner	Modelmaker	65
Cyberware	Model-15	15
Cyrax	2500	125
Digibotics	Digi-Bot II	60
Eyetronics	Shape Snatcher[a]	7.5
GSI	Inca2 Camera + software	155
Immersion	LightScribe	10
Immersion	Microscribe	4
Inspeck	3D Capturor	13
Interzart AG	3D Commerce 3D Scanbook Pro	9
Interzart AG	3D Commerce 3D Scanstation Pro	24
Kodak	DCS-660M Professional Camera	15
MetricVision	MV200	400
Minolta	Metaflash	4
Minolta	Vivid 300 Laser Scanner	15
Minolta	Vivid 700 Laser Scanner	30
Minolta	Vivid 900 Laser Scanner[b]	40
Polhemus	Fast Scan	30
Realscan	3D Laser Range Camera	35
Roland	LPX-250 Desktop Laser Scanner	10
Surphaser	3D Laser Scanner	30

[a] Available at the UT, CTW faculty, `m.e.toxopeus at utwente.nl`

[b] Used by Lu et. al. [10][11][12] and in the Face Recognition Grand Challenge project [18]

Table 2.1: Incomplete list of available 3D acquisition hardware

These conditions are met by selecting appropriate hardware. For the structured light system, an obsolete overhead projector is used. The lamp in the projector is a high-power version of a common light bulb and thus has a lot of energy in the infrared part of the spectrum. The light of the projector is structured by printing a slide sheet and putting it on the projection plane. The light is filtered by placing a large plastic infrared filter on top of the projected sheet.

For the camera, an USB webcam is preferred because it is easily interfaced with a computer running MATLAB. Essential however is its sensitivity to infrared light. A specific camera type is chosen in Section 2.4.1.

2.3 Structured Light Theory

For range measurements, the structured light principle is a less expensive alternative to laser scanning techniques. An exhaustive overview of various kinds of structured light techniques is listed by Pagès et. al. [16]. A good example of the implementation of the technique for scanning cultural heritage like statues, is given by Rocchini et. al. [19]. An application for face vision was developed by Beumier and Acheroy [4][5].

2.3.1 Coordinate triangulation

Conceptual approach

To support the explanation of the structured light technique, Figure 2.1 is shown. From the left, a well chosen stripe pattern is projected onto the object. The projection of each stripe spans up a three-dimensional plane. When the object is positioned in the projector range, the plane intersects with the object's surface. This results in curved stripes being projected on the surface.

From a different view angle, the scene is observed by a camera. For each point in the image plane where a stripe is detected, a line is constructed from the camera pinhole, through the specified point in the image plane, running towards the object. As can be seen from Figure 2.1, the line and the plane intersect in exactly one point in 3D. Provided that the equations for both the line and the plane are known, the 3D surface coordinate can be computed. A complete surface scan is done by following each projected stripe vertically, and by repeating the procedure for every projected plane.

Mathematical approach

For each specific plane i, of which only one is shown in Figure 2.1, the equation is given by

$$a_i x + b_i y + c_i z + d_i = 0 \tag{2.1}$$

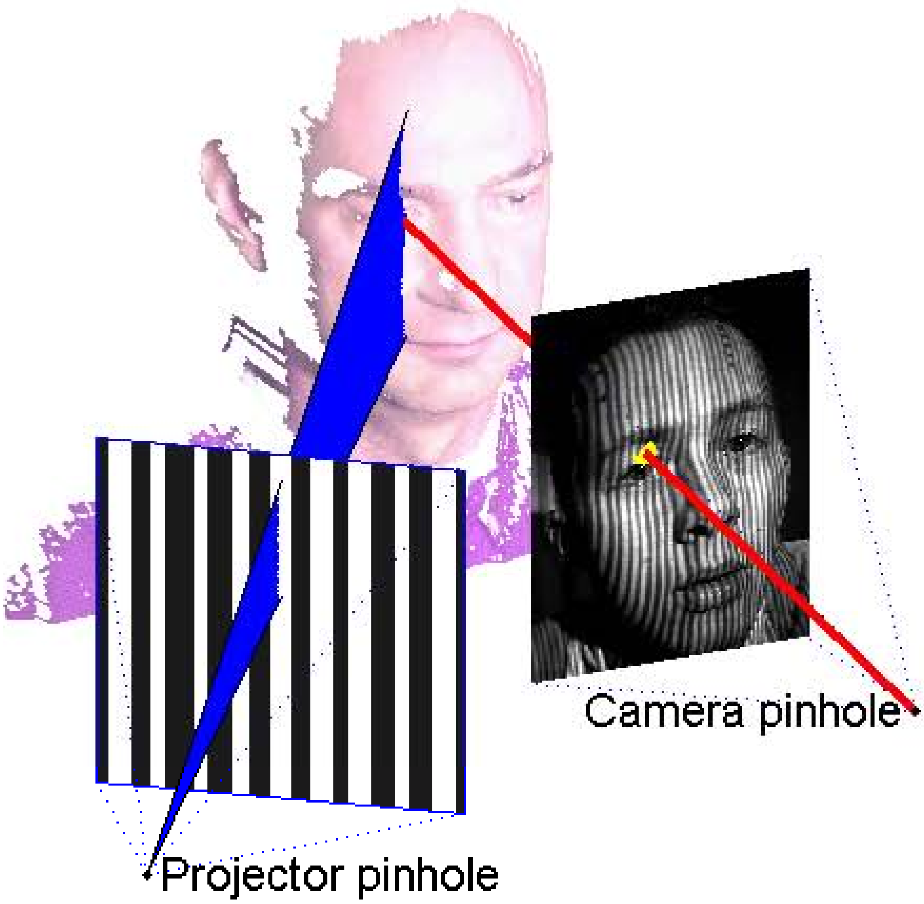


Figure 2.1: 3D coordinate follows from line-plane intersection

where x, y, and z are defined in the camera coordinate system. The coefficients a_i, b_i, c_i and d_i are supposed to be known from calibration.

Finding an equation for the line in Figure 2.1 is accomplished by finding all possible 3D coordinates that project on (x_{im}, y_{im}), where the subscript $_{im}$ denotes a two-dimensional coordinate on the image plane. Because the camera pinhole is chosen as the origin of the coordinate system and the image plane is chosen at $z = D$, the projection equations are found easily. (D is called the focal distance and corresponds to the distance of the camera pinhole to the CCD chip.)

To find the line equation, first a single point $P = (x, y, z)$ is projected onto the image plane:

$$\begin{pmatrix} x_{im} \\ y_{im} \end{pmatrix} = \begin{pmatrix} \frac{xD}{z} \\ \frac{yD}{z} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \frac{D}{z} \tag{2.2}$$

Inverting this relation shows that every point, projected on (x_{im}, y_{im}) in the image plane, must obey

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x_{im}z}{D} \\ \frac{y_{im}z}{D} \end{pmatrix} = \begin{pmatrix} x_{im} \\ y_{im} \end{pmatrix} \frac{z}{D} \tag{2.3}$$

Now, equations for both the plane (Equation 2.1) and the line (Equation 2.3) of Figure 2.1 are known. Combining these equations should now give a single 3D coordinate. Therefore, the assumption is done now that $P = (x, y, z)$ is not only part of the line, but lies in the plane as well.

Substituting the line equation into the plane equation is called triangulation and yields

$$a_i \frac{x_{im}z}{D} + b_i \frac{y_{im}z}{D} + c_i z + d_i = 0 \tag{2.4}$$

Factoring out z gives

$$z \left(\frac{a_i x_{im}}{D} + \frac{b_i y_{im}}{D} + c_i \right) = -d_i \tag{2.5}$$

Now, the solution for the depth coordinate of point P is

$$z = -\frac{d_i D}{a_i x_{im} + b_i y_{im} + c_i D} \tag{2.6}$$

Substituting this solution into the inverse projection relations in Equation 2.3 gives the complete solution for point P:

$$P = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_{im} \\ y_{im} \\ D \end{pmatrix} \cdot \frac{-d_i}{a_i x_{im} + b_i y_{im} + c_i D} \tag{2.7}$$

Now, all three components of P are expressed in its projection (x_{im}, y_{im}), the structured light plane coefficients a_i, b_i, c_i and d_i, and the camera parameter D.

2.3.2 Camera distortion model

The coordinates x_{im} and y_{im} are not directly available from the camera image. The image plane used so far is an ideal image plane, where no lens distortion is introduced yet. To model the distortion of the lens, the coordinates of the image plane are mapped into a distorted image plane by using the following radial distortion model:

$$x_{distorted} = \left(1 + kr^2\right) x_{im} \tag{2.8}$$

$$y_{distorted} = \left(1 + kr^2\right) y_{im} \tag{2.9}$$

where $r^2 = x_{im}^2 + y_{im}^2$ and k is the radial distortion coefficient. k is found during camera calibration. Although a more complex distortion model is possible (with tangential distortion and higher order radial terms), the model proposed is accurate enough for the camera selected in Section 2.4.1.

The final step is to map the coordinates on the distorted image plane to pixel coordinates. This step consists of just translating the origin. In the distorted image plane, the origin is in the center of the image. In the pixel image plane, the origin is in the upper left corner of the image. The center coordinates are found by simply dividing the camera resolution by 2.

If 3D triangulation is done, the inverse mapping must be performed. Every pixel where a stripe center is located, must be mapped backed onto the ideal image plane, so a line through the camera pinhole can be defined correctly. In the camera calibration toolbox, the necessary inverse mapping function is provided.

For detailed information about distortion models, see the website of the camera calibration toolbox by Bouguet [6], where additional references are listed. A useful article on calibration of a structured light system is written by Beumier [3].

2.3.3 Light coding

In the previous section, the computation was carried out for a single known plane. In a structured light system, useful for face shape acquisition, multiple planes must be used, all of which have different coefficients. To know to intersect which line with which plane, it is necessary to distinguish between the different planes in the two-dimensional image. In case the surface is flat and all projected stripes are in the camera image, this is easy: just count the stripes from the left to the right. For general surfaces, this is more difficult because occlusions may occur. A structured light stripe may be present on the surface, but may be invisible from the camera point of view. A simple counting scheme would fail in that case.

In a robust system, it should be possible to determine the plane number (called *index* from now on) by itself or by a small environment in the 2D image. Therefore, an appropriate coding strategy has to be chosen. Again by

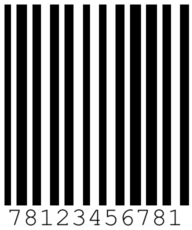


Figure 2.2: Stripe coding scheme (stripes are white)

Left Neighbor	Stripe itself	Right Neighbor	Index mod 8
thin	thick	thin	1
thick	thin	thick	2
thin	thick	thick	3
thick	thick	thick	4
thick	thick	thin	5
thick	thin	thin	6
thin	thin	thin	7
thin	thin	thick	8

Table 2.2: Stripe decoding scheme

Pagès [16], a detailed overview of possible coding strategies is given. Because an implementation with infrared light should work as well, a monochromatic coding scheme is obligatory. Furthermore, time-multiplexing schemes are out of scope because only a static overhead projector is used.

These boundary conditions are the same as for the structured light system developed by Vuylsteke et. al. [23]. An appropriate coding scheme could be adopted from this article, where a projection pattern is carefully chosen. For time reasons however, a more simple coding strategy is used.

In the system, vertical monochrome stripes are projected. A De Bruijn sequence is used to code the subsequent stripes. A very simple alphabet of two symbols is used: there exist thick lines (symbol 1) and thin lines (symbol 2). By building words of length n, up to 2^n projected stripes can be identified uniquely.

In this project, a cyclic De Bruijn sequence of word length three is chosen, allowing $2^3 = 8$ unique stripes. A word length larger than three is not preferred because it would take a large number of correctly measured stripe thicknesses to identify a single stripe.

Part of the selected sequence is shown in Figure 2.2. In Table 2.2, the corresponding conversion from subsequence to index is shown. The sequence is generated using a MATLAB routine by Burkardt [8].

If a series of three adjacent stripes can be found in the camera image, and

Figure 2.3: Camera used in the structured light system

if their thicknesses can be determined correctly, the index of the structured light stripe is known. However, because the sequence is repeated, it is only known up to a modulo 8. It is, for example, impossible to see the difference between stripe index 21 and stripe index 13. Both stripes will be mapped on the fifth subsequence (thick-thick-thin). This introduces the need for an additional 'demodulizing' technique, described in Section 4.3.4.

2.4 System implementation

2.4.1 Camera

The major components of a structured light system are the camera and the projector. For the camera, a Philips ToUCam Pro II Webcam is used (type PCVC840K/00, see Figure 2.3). It is chosen because it contains a CCD that is sensitive to near-infrared light. Moreover, according to the manufacturer specifications, it is capable of acquiring high quality images under low light conditions. It is recommended by astronomers to make infrared recordings of the solar system (for example in [14]). The camera is easily interfaced to MATLAB using the Image Acquisition Toolbox. The camera is cheap, compared to professional digital cameras.[1]

To make the camera sensitive to the infrared domain only, the visible part of the spectrum is filtered out by a small plastic disc of IR-passing material. The filter is mounted in front of the camera lens.

Although the claimed maximal resolution is 1280x960, this is only achieved in software by interpolating between pixels. The real maximal camera resolution is 640x480 pixels. The technical specifications are found at the manufacturers web site [17]. A relevant selection is listed in Table 2.3.

By rotating the camera 90° and rotating each acquired image 90° in the opposite direction, a resolution of 480x640 is achieved. This portrait

[1] In April 2005, the webcam was available for EUR 70 [22].

Sensor type	CCD
Maximal (interpolated) photo resolution	1280 x 960
Maximal (real) photo resolution	640 x 480
Maximal video resolution	640 x 480
Maximal framerate	60 fps
Lens	6 mm, f=2.0, H33°
White balance	2500 - 7500 K

Table 2.3: Philips ToUCam Pro II specifications

orientation is better suited to the aspect ratio of the average face and hence allows a higher resolution, expressed in pixels per mm.

2.4.2 Projector and slide

For the structured light projector, a common overhead projector is used in this project. The structured light planes are generated by projecting a slide. On the slide, a cyclic De Bruijn-sequence as described in Section 2.3.3 is printed. The stripe centers are equidistant with a distance of 2.0 mm. For the thick stripes, $\frac{5}{9}$ of the stripe width is transparent. For the thin stripes, $\frac{3}{9}$ is transparent.

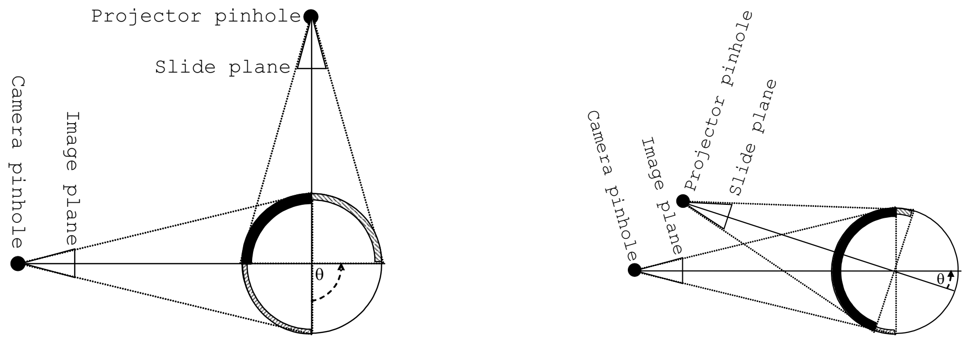
In order to achieve maximal contrast between stripe and non-stripe regions, the projector is put as close to the face as possible. The distance to the face is limited by the lenses in the projector. If the projector is placed too close to the face, the projected stripes are out of focus and become blurred. The resulting minimal distance between upper projector mirror and the face is about 1 meter.

2.4.3 System geometry

The light structuring slide is placed on the projector such that the resulting planes are vertical. The camera is placed such that the projection of the planes on the surface result in vertical stripes in the acquired image.

The optical axis of the camera is not allowed to run parallel to the structured light planes. On the one hand, when the camera axis and the structured light planes coincide, the stripe locations in the image are known a priori and thus contain no range data. To estimate the range as accurate as possible, the angle should be as close to 90° as possible.

On the other hand however, when the angle θ between the camera axis and the structured light planes becomes too large, there is only a small region on the surface that is both lighted and observed. Suppose the structured light planes come from the North, and the camera is located in the West (see Figure 2.4(a)). When a cylindrical object is present in the scene, no more



(a) Angle θ too big

(b) Better angle θ

Figure 2.4: Different angles (top view)

than 25% of the surface is both lighted and observed. Hence, the surface coverage is unacceptable low.

Therefore, a tradeoff between accuracy and coverage is needed. In Figure 2.4(b), the angle θ is reduced. In the realized system, the angle is about 18°, allowing complete coverage of a facial surface.

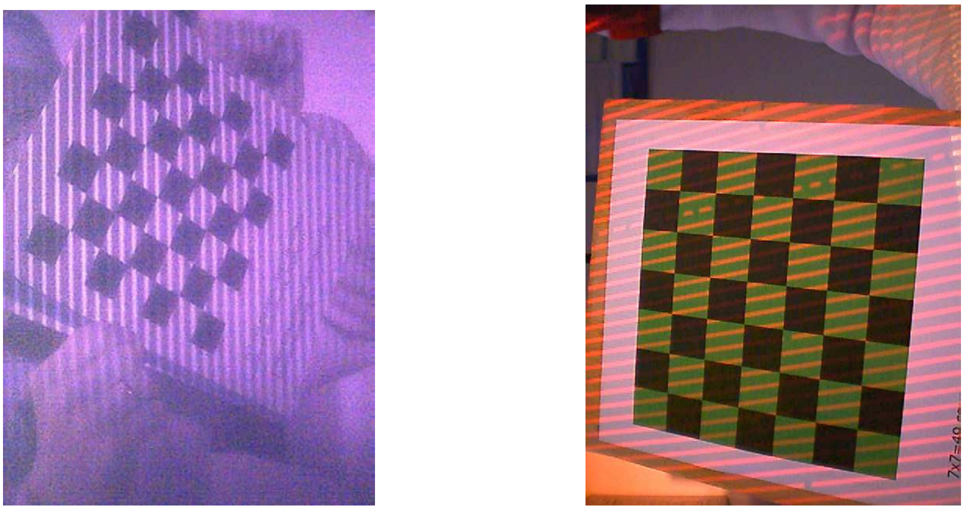
2.5 Results

Unfortunately, the combination of projector and camera does not give good results when only the infrared domain is considered. In Figure 2.5(a), a sample image is shown. For comparison, a visible light image is shown in Figure 2.5(b). In the latter, the infrared filters are removed from both the camera and the projector.

Detailed results for the overall structured light system are not possible without calibration and a detailed description of the acquisition steps. Hence they are postponed to Chapter 4.

2.6 Conclusion

The structured light principle allows for partial reconstruction of 3D-surfaces. For triangulation, the surface must be both lighted and observed. With low-cost hardware (an overhead projector, a slide and a webcam), a 3D acquisition device is built. However, other hardware is needed to obtain images with infrared colored light. Instead of replacing the camera by another one, the infrared requirement is dropped because of the experimental nature of the setup. Moreover, it is always possible to change the hardware



(a) Infrared camera image

(b) Visible light camera image

Figure 2.5: Infrared image quality is poor

afterwards, because the structured light principle is independent of the light color.

Before the results of the 3D-scanner can be presented, a calibration procedure is required; this is covered in the next chapter. Additionally, the camera image needs to be processed. Results of the 3D acquisition system are therefore postponed to the end of Chapter 4.