High-Resolution Structured Light Range Scanner with Automatic Calibration

ALEXANDER M. BRONSTEIN*  MICHAEL M. BRONSTEIN‡
EYAL GORDON§  RON KIMMEL¶

August 4, 2003

Abstract

In this work, we present a low-cost high-resolution structured light range scanner with automatic calibration, based on temporal stripe encoding. We present a novel single-step automatic active stereo calibration method, based on optimal backprojection estimation. We also present a method for estimating local fidelity of the range image and show how this information can be used to obtain sub-stripe resolution. We describe the scanner architecture and show experimental calibration and shape reconstruction results.

1 Introduction

In this work, we present a low-cost high-resolution structured light scanner with automatic calibration. The term structured light refers to a wide variety of active triangulation range scanners, which exploit emission of a controlled light pattern (or a series of patterns) onto the scanned object. Structured light scanners are similar to passive stereometric scanners, where one camera is replaced by a projector. The advantage of structured light systems is that unlike passive stereo, no correspondence problem has to be solved, since the stripe code can be extracted from the image (or a series of images) acquired by the camera.

In its simplest form, a structured light sensor consists of a laser projector, whose beam illuminates a single point in the scene. Assuming the point is found in the camera field of view, it can be easily located and identified. Since the point location on both the projector and

*Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel
‡Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel
§Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel
¶Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel
the camera is known, reconstruction of shape world coordinates can be carried out by simple triangulation. A more efficient scheme replaces the point laser beam by a line beam, that ”sweeps” the object [1]. Still, sweeping the entire scene by a laser beam is a time consuming operation and therefore limited for static objects.

For better scan time, several stripe illumination-based techniques have been proposed. Single pattern approaches, using spatial encoding of projection planes or rays were introduced e.g. in [2], [3], [4]. Although time-efficient, spatial encoding generally produces a sparse depth map with mediocre spatial resolution.

High reliability identification of projection planes with non-restrictive assumptions on scene contents and illumination conditions can be achieved by temporal encoding, which are based on projection of a sequence of light patterns onto the object. A robust, commonly used structured light system based on binary light patterns encoded with Gray code, was proposed in [4]. This technique was adopted in this work. It is also possible to use color modulation, which allows to save up the number of stripes projected onto the object [2]. This, however, requires a color camera.

An important issue in structured light systems is calibration. Reconstruction is possible only when the camera projection matrix and the projection planes are known in the world coordinate system. Calibration of structured light scanners is usually more complicated than that of a passive stereo pair. A standard approach consists of three steps: estimating the camera intrinsic matrix (camera calibration), estimating the plane equations for each of the projection planes (projector calibration) and finally, estimating the Euclidean transformation between the camera and the projector (projector-camera calibration). This technique was adopted in e.g. [5]). In [6], an alternative approach, based on estimation of the world-to-camera image and world-to-projector coordinate system transformation, is proposed. We adopt this approach and extend it to simultaneous backprojection operator estimation.

This work is organized as follows: in Section 2 we present a projective model of a structured light system and discuss shape reconstruction and automatic calibration. Following [6], we propose a more accurate method for calibration. We also present stability analysis of the reconstruction operator. Section 3 is dedicated to implementation issues. We present the scanner architecture and discuss shape reconstruction from temporally-encoded binary light patterns. We also propose a method for pixel fidelity weighting and show how it can be used for obtaining sub-stripe resolution and demonstrate it in reconstruction of human faces. In Section 3.3, we discuss implementation issues of automatic calibration and show experimental results, which prove superiority of the proposed method. Section 5 concludes the work and discusses further research directions.

2 Projective model of a structured light system

A typical structured light system consists of a camera and a projector. The role of the projector is to light the scanned object in such a way, that from the image (or sequence of images)
acquired by the camera a stripe code can be extracted. The encoding can be done either spatially using a single pattern or temporally using a series of varying patterns (see e.g. [1] and references cited therein). The raw output of a structured light scanner is a stripe code assigned for every pixel in the image. Intersection of a ray in world coordinate system (WCS) with a plane in WCS yields the world coordinates of an object point. Using this triangulation method, the raw sensor data is converted into 3D data in WCS.

We will henceforth assume that both the camera and the projector obey the pin-hole optical model (non-linear distortion correction may be required for lenses that do not obey this model). The transformation from 3D world coordinates to camera image plane coordinates is commonly described by a $3 \times 4$ perspective projection matrix (PPM) [7]. Following [6], we model the projector by a $2 \times 4$ PPM, mapping world coordinates to stripe identification code (id).

Let us define a homogenous world coordinate system $X_w$, in which the object position is specified; a homogenous camera coordinate system $X_c$, in which pixel locations in the image plane are specified, and a homogenous projector coordinate system $X_p$, in which stripe ids are specified. The latter is particular, since it contains only one independent coordinate.

The transformation form world coordinates to camera coordinates is given by

$$X_c = C_c X_w,$$

where $C_c$ is the camera PPM of the form

$$C_c = \alpha \begin{bmatrix} f_x & k f_y & x^0_c \ 0 & f_y & y^0_c \ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_c & t_c \end{bmatrix}.$$  \hfill (2)

The rotation matrix $R_c$ and the translation vector $t_c$ define the transformation between WCS $X_w$ and the camera-centric reference frame $X_c$. The parameters $f_x$ and $f_y$ are the camera focal length scaled to each of the CCD dimensions, and $x^0_c$ and $y^0_c$ are the origin of $X_c$ in image coordinates. The parameter $\alpha$ is a proportion coefficient and $k$ is the shear of the camera coordinate system.

Similarly, the transformation form world coordinates to projector coordinates is given by

$$X_p = C_p X_w,$$

where $C_p$ is the projector PPM of the form

$$C_p = \alpha \begin{bmatrix} f_p & 0 & x^0_p \ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_p & t_p \end{bmatrix}.$$  \hfill (4)

$R_p$ and $t_p$ define the transformation between WCS and $X_p$. The parameter $f_p$ is the projector focal length scaled to LCD dimensions, and $x^0_p$ is the origin of $X_p$ in projector coordinates, which physically is the $x$-coordinate of the intersection of the optical axis and the LCD [6].
Here we implicitly assume that the stripe code varies along the horizontal direction of the LCD. McIvor and Valkenburg [6] show that $C_p$ is a valid camera PPM iff the submatrix formed by its first three columns has full rank. Similarly, $P_p$ is a valid projector PPM iff the submatrix formed by its first three columns is of rank 2.

Equations 1 and 3 define the transformation

$$T : X_w \rightarrow (X_c, X_p), \quad (5)$$

which maps an object point in WCS into pixel location in the camera image plane and a stripe id (coordinate in the projector system of coordinates). We refer to this transformation as forward projection.

The world coordinates of the object point are usually unknown and have to be determined, whereas the pair $(x_c, x_p)$ is what the structured light sensor measures and can be extracted from the raw data. Therefore, given the camera and the projector PPMs and a pair of measurements $(x_c, x_p)$, one can attempt inverting 5 in order to calculate $x_w$. We will term the inverse transformation

$$T^{-1} : (X_c, X_p) \rightarrow X_w, \quad (6)$$

as backprojection and the process of determining world coordinates from measured data as reconstruction.

Reconstruction requires the knowledge of $C_c$ and $C_p$. Therefore, calibration must be performed before, during which the forward projection operator is estimated. This is done by measuring a set of pairs $\{(x_c, x_p)_n\}_{n=1}^N$ corresponding to a set of points with known world coordinates $\{(x_w)_n\}_{n=1}^N$. Physically, a calibration object with a set of fiducial points, whose location is known, is scanned. WCS is then chosen to be some local coordinate system of the calibration object, in which the coordinates of each fiducial point are specified.

Reconstruction will be discussed in Section 2.1 and calibration in Section 2.3.

### 2.1 Reconstruction

In this section we assume that the forward projection operator $T$ is known (i.e. the projective matrices $C_c$ and $C_p$ are given). The reconstruction problem can be stated as follows: given measured $(x_c, x_p)$, calculate $x_w$ according to

$$x_w = T^{-1}(x_c, x_p). \quad (7)$$

Explicitly, $x_w$ has to satisfy the linear system of equations

$$x_c = C_c x_w \quad (8)$$

$$x_p = C_p x_w. \quad (9)$$
However, since all vectors are given in homogenous coordinates, it is possible that no \( x_w \) satisfies equations 8 and 9 simultaneously. Let us denote \( x_c = [w_c x_c, w_c y_c, w_c] \) and \( x_p = [w_p x_p, w_p] \) and let \( c_k, p_k \) be the \( k \)-th row of \( C_c \) and \( C_p \), respectively. Then, the linear system of equations can be rewritten as

\[
\begin{align*}
w_c x_c &= c_1 x_w \\
w_c y_c &= c_2 x_w \\
w_c &= c_3 x_w
\end{align*}
\] (10)

and

\[
\begin{align*}
w_p x_p &= p_1 x_w \\
w_p &= p_2 x_w
\end{align*}
\] (11)

Substituting \( w_c \) into 10 and \( w_p \) into 11 yields

\[
\begin{align*}
x_c c_3 x_w &= c_1 x_w \\
y_p c_3 x_w &= c_2 x_w \\
x_p p_2 x_w &= p_1 x_w
\end{align*}
\] (12)

which can be written in matrix notation as \( Q x_w = 0 \), where

\[
Q = \begin{bmatrix}
x_c c_3 - c_1 \\
y_p c_3 - c_2 \\
x_p p_2 - p_1
\end{bmatrix}
\] (13)

The matrix \( Q \) can be split into a \( 3 \times 3 \) matrix \( R \) and a \( 3 \times 1 \) vector \( s \): \( Q = [R, s] \). Substituting \( x_w = [w_u x_w, w_u y_w, w_u z_w, w_u] \) yields

\[
[R, s] \begin{bmatrix}
w_u x_w \\
w_u y_w \\
w_u z_w
\end{bmatrix} = R \begin{bmatrix}
w_u x_w \\
w_u y_w \\
w_u z_w
\end{bmatrix} + w_u s = 0.
\] (14)

Therefore, the object point in non-homogenous world coordinates \( x_w = [x_w, y_w, z_w] \) is a solution of the linear system

\[
R x_w = -s.
\] (15)

Backprojection is therefore given by

\[
x_w = -R^{-1} s.
\] (16)
We remind that both $R$ and $s$ are functions of $x_c$, $y_c$ and $x_p$.

If $C_c$ and $C_p$ are valid camera and projector PPMs, $R$ is invertible except of cases where the ray originating from the camera focal point to the object point is parallel to the plane originating at the projector focal point and passing through the object point. The latter case is possible either when the object point is located at infinity, or when the camera and the projector optical axes are parallel (this happens when $R_c = R_p$). This gives a constraint on camera and projector mutual location: in order to make triangulation possible, the camera should not have its optical axis parallel to that of the projector.

## 2.2 Reconstruction stability

We have seen that the matrix $R$ in Equation 15 becomes singular when the ray in the camera coordinate system and the plane in the projector coordinates system are parallel. A reasonable question that may arise is how stable is the solution under random perturbations of $x_c$ and $x_p$. In this work, we will address only perturbations in $x_p$, since they are the most problematic ones in structured light systems.

For simplicity, let us assume that WCS coincides with the camera coordinate system and the transformation to the projector coordinate system is given by

$$ x_p = R_p + t_p. $$

(17)

Without loss of generality, we assume that the center of the camera and projector coordinate system coincides with their optical axes, i.e. $x_c^0 = y_c^0 = x_p^0 = 0$.

Let us assume that the object point is found on some ray in $x_c = \alpha v_c$; the ray is uniquely defined by the camera image plane coordinates $x_c$ and the point location is uniquely defined by the parameter $\alpha$. Let us denote by $x_p$ the stripe id corresponding to the given object point. Then, the following system of linear equations

$$ n^T x_p = 0 $n^T(R_p x_c + t_p) = 0, $$

(18)

must hold simultaneously; $n$ denotes the normal to the plane defined by the stripe id $x_p$. Substituting $x_c = \alpha v_c$ yields

$$ n^T x_p = n^T(\alpha R_p v_c + t_p), $$

(19)

hence

$$ \alpha = \frac{n^T x_p}{n^T R_p v_c}. $$

(20)

However, in practice, the stripe id $x_p$ is estimated using structured light, and therefore it is especially sensitive to noise. Let us assume that instead of the real stripe id $x_p$, a perturbed
stripe id \( \tilde{x}_p = x_p + \delta x_p \) was measured. This, in turn, means that \( \hat{x}_p = x_p + [\delta x_p, 0, f_p]^T \), which yields

\[
\tilde{\alpha} = \frac{n^T \tilde{x}_p}{n^T R_p v_c}.
\]  

(21)

Hence, the perturbation in \( x_p \) causes a perturbation in the location of the object point along the ray \( x_c = \alpha v_c \) by

\[
\delta \alpha = \frac{n_1 \delta x_p}{\|n\|_2 \|v\|_2 \sin \Theta_{nv}},
\]  

(22)

where \( \Theta_{nv} \) is the angle between the plane defined by the normal \( n \) and the ray defined by the direction \( v_c \). Therefore,

\[
\delta \|x_w\|_2 = |\delta \alpha| \|v_c\|_2 = \left| \frac{n_1}{\|n\|_2 \sin \Theta_{nv}} \right| |\delta x_p|.
\]  

(23)

The ratio \( \cos \theta_P = \frac{n_1}{\|n\|_2} \) has a geometrical interpretation of cosine of the projection angle; substituting it into Equation 23 yields the sensitivity of the reconstructed object point to perturbations in the stripe id:

\[
\frac{\delta \|x_w\|_2}{\delta x_p} = \left| \frac{\cos \theta_P}{\sin \Theta_{nv}} \right|.
\]  

(24)

### 2.3 Calibration

In this section we assume that the forward projection operator \( T \) is unknown and has to be estimated from a given set of measured \( \{(x_c, x_p)_n\}_{n=1}^N \) and corresponding known \( \{x_w\}_{n=1}^N \). Explicitly, it is desired to find such \( C_c \) and \( C_p \) that obey

\[
(x_c)_k = C_c(x_w)_k
\]  

(25)

\[
(x_p)_k = C_p(x_w)_k,
\]  

(26)

for \( k = 1, \ldots, N \). Since data measurement is not perfect (e.g., both the camera and the projector resolution is finite), no projection operator will fit the data perfectly. Our goal is therefore to find such a \( T^{-1} \) that will relate the measured and the known data in an optimal way. It is thus important to address the optimality criterion.

McIvor and Valkenburg [6] study the possibility of optimizing separately the camera and the projector forward projections in the sense of the \( L_2 \) norm. Mathematically, this can be
Figure 1: Sensitivity of the reconstructed object point to perturbations in $x_p$.

formulated as

$$C_c = \arg\min_{C_c} \sum_{k=1}^{N} \|C_c(x_w)_k - (x_c)_k\|_2^2 \quad \text{s.t.} \quad C_c \in \text{PPM}$$

$$C_p = \arg\min_{C_p} \sum_{k=1}^{N} \|C_p(x_w)_k - (x_p)_k\|_2^2 \quad \text{s.t.} \quad C_p \in \text{PPM}. \quad (27)$$

Let us define

$$B_k = \begin{bmatrix} (x_w)_k & 0 \\ 0 & (x_w)_k \\ -(x_c)_k(x_w)_k & -(y_c)_k(x_w)_k \end{bmatrix}^T$$

$$l = [c_1, c_2, c_3]^T,$$  

(28)

where $c_k$ is the $k$-th row of $C_c$. Using this notation, the set of $N$ equations 25 can be rewritten as

$$B_k l = 0,$$  

(29)

for $k = 1, \ldots, N$, which in turn can be expressed as a single homogenous linear equation

$$Al = 0,$$  

(30)
where \( A = [B_1^T, ..., B_N^T]^T \). The vector of variables \( l \) is the camera projection matrix \( C_c \) needed to be determined. Since the camera PPM is defined up to a scaling factor, we will demand \( \|l\|_2 = 1 \) in order to avoid the trivial solution. With physically measured data, the matrix \( A \) will usually have full rank and therefore, no \( l \) will be an exact solution of equation 30. However, one can find the best least-squares solution by solving

\[
\begin{align*}
l &= \text{argmin} \|Al\|_2^2 \quad \text{s.t.} \quad \|l\|_2 = 1,
\end{align*}
\]

and ensuring that the obtained \( C_c \) is a valid PPM [6]. The problem 31 is equivalent to problem 27 for the camera matrix, and its solution minimizes the square error between the measured image plane coordinates of the set of fiducial points and those obtained by projecting the set of the corresponding points in WCS onto the camera image plane.

Similarly, replacing \( B_k \) and \( l \) in 28 with

\[
\begin{align*}
B_k &= \begin{bmatrix} (x_w)_k \\ -(x_p)_k(x_w)_k \end{bmatrix}^T \\
l &= [p_1, p_2]^T
\end{align*}
\]

yields the \( L_2 \) minimization problem 27 for the projector matrix.

Optimization problem 31 is a minimum eigenvalue problem and it can be shown that \( l \) minimizing \( \|Al\|_2 \) is the eigenvector corresponding to the minimum eigenvalue of \( A^T A \). It must be noted, however, that since usually the minimum eigenvalue of \( A^T A \) is very small, numerical inaccuracies are liable to rise.

Solution to the problem 27 finds two PPMs that minimize the squared error between the measured data and the forward projection of the known fiducial points in WCS into the camera and the plane coordinate systems. However, what is actually needed is to minimize the squared error between the known fiducial points in WCS and the backward-projected measurements. Mathematically, this can be formulated as

\[
\begin{align*}
T &= \text{argmin} \sum_{k=1}^{N} \left\| T^{-1}(x_c, x_p)_k - (x_w)_k \right\|_2^2 \\& \quad \text{s.t.} \quad C_c, C_p \in \text{textPPM}.
\end{align*}
\]

The above problem is no more separable and is non-convex; therefore, it has to be solved by numerical global optimization methods. Still, an efficient solution in few iterations is possible using the Newton method, since the number of variables in the problem is small ( \( 3 \times 4 + 2 \times 4 = 20 \) ) and both the cost function, its gradient, and the Hessian can be computed analytically. As the starting point for iterative optimization, a solution of problem 27 can be used.

Since the calibration process is performed once, we invest additional computational complexity in order to obtain better projection estimation and better reconstruction results.
3 Implementation considerations

3.1 Hardware

In our implementation, we used a FireWire DragonFly black and white CCD camera and a computer-controlled SVGA LCD projector (see Figure 3). The camera and the projector specification are given in Table 1; additional information can be found in [8] and [9]. Acquisition control and reconstruction was carried out on an IBM ThinkPad A portable computer with Pentium IV 2.0 GHz with 1024MB running Microsoft Windows XP.

The camera and the projector were fixed to separate adjustable tripods, which allowed almost arbitrary mutual location of both of them. From Equation 24 in Section 2.2, it follows that large angle disparities between the projector and the camera optical axes are highly desirable in order to reduce the noise gain of the backprojection operator. However, practically there is a limitation on the angle disparity, since due to the Lambertian law, less light is reflected by the scanned surface to the camera and due to occlusions, which are more significant at large angle disparities. In all test presented in this work, angle disparity did not exceed $20^\circ$ in each axis.

Figure 2 shows the scanner architecture. The acquisition controller was implemented in software using Microsoft Visual C++ 6.0. The camera was operated at 30 FPS with shutter time set in integer units of $1/50 \text{ sec}$ in order to avoid flickering due to neon illumination. IEEE-1394 bus served for both data transfer and control. The projector was fed with controllable images through the VGA interface using the IBM ThinkPad dual video output capabilities.

Data processing modules responsible for reconstruction and calibration were implemented in MathWorks MATLAB 6.5 and equipped with a graphical user interface.

<table>
<thead>
<tr>
<th>Camera</th>
<th>Projector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>ViewSonic</td>
</tr>
<tr>
<td>Model</td>
<td>DragonFly</td>
</tr>
<tr>
<td>Lens</td>
<td>17.7 - 21.2 mm</td>
</tr>
<tr>
<td></td>
<td>manual focus and zoom</td>
</tr>
<tr>
<td>Resolution</td>
<td>640 × 480 pixels</td>
</tr>
<tr>
<td></td>
<td>8 bit</td>
</tr>
<tr>
<td>Interface</td>
<td>IEEE-1394</td>
</tr>
<tr>
<td>CCD/LCD</td>
<td>1/3” Sony CCD (BW)</td>
</tr>
<tr>
<td>Frame rate</td>
<td>30 Hz</td>
</tr>
<tr>
<td></td>
<td>0.7” TFT (color)</td>
</tr>
<tr>
<td></td>
<td>60 Hz</td>
</tr>
</tbody>
</table>

Table 1: Camera and projector specifications
Figure 2: Schematic description of scanner architecture.
3.2 Reconstruction

In our implementation, we used temporal encoding, which allowed to obtain a dense \( z \)-map of the scanned objects. Eight binary patterns, encoding the stripe id using the Gray code were projected onto the object and 8 corresponding images \( I_k \) were acquired (\( I_1 \) corresponding to the most significant bit). In addition, we acquired a full-darkness image \( I_L \) (the projector LCD was blackened) and a full-illumination image \( I_H \) (the projector LCD was set to maximum intensity). These two images served for compensation of ambient light and the non-constant albedo of the object.

The quantity \( I_L(x, y) \) reflects the illumination of the object in pixel \( (x, y) \) at darkness and differs from zero only due to presence of ambient illumination. Since the reflectance of objects at illumination levels used in normal conditions obeys linear superposition law, subtracting \( I_L \) from the rest of the images compensates for the ambience light. The quantity \( I_H(x, y) - I_L(x, y) \) is proportional to the object albedo at the pixel \( (x, y) \).

We define a set of 8 normalized intensity images

\[
J_k(x, y) = \frac{I_k(x, y) - I_L(x, y)}{I_H(x, y) - I_L(x, y)}. \tag{34}
\]

A normalized intensity image \( J_k(x, y) \) has the values in the range \([0, 1]\) and reflects the amount of light irradiated onto the object surface at pixel \( (x, y) \). The value of 1 stands for full illumination, whereas the value of 0 stands for no illumination. Theoretically, \( J_k \) should be binary images: 1 where a light stripe is present and 0 in places where there is a dark stripe.
In practice, however, $J_k$ are not binary and we therefore define

$$B_k(x, y) = \begin{cases} 1 : & J_k(x, y) > 0.5 \\ 0 : & J_k(x, y) \leq 0.5 \end{cases}. \quad (35)$$

Better results were obtained when $J_k(x, y)$ was smoothed with a Gaussian filter prior to binarization. Figure 4 depicts $I_H$ and $I_L$ as well as the LSB and the MSB patterns. Figure 5 presents the normalized intensity image $J_3(x, y)$, the corresponding binary image $B_3(x, y)$ and a profile of a vertical line from these two images.

![Figure 4: Raw data acquired by the scanner. From left to right: full illumination image, full darkness image, MSB pattern and LSB pattern. All images are normalized.](image)

For every pixel $(x, y)$, we define the stripe code as the Gray code sequence

$$S(x, y) = [B_1(x, y), ..., B_8(x, y)]. \quad (36)$$

Decoding $S(x, y)$ yields a number $T(x, y) \in [0, 1]$, which will be referred to as stripe id. Note that $T(x, y)$ is not really continuous but rather has the values $T(x, y) \in \{2^{-N}n : n = 0, ..., 2^N - 1\}$ (in our implementation $N = 8$).

For every pixel, $x_p(x, y) = T(x, y)$ defines the projector coordinate of an unknown object point corresponding to the pixel $(x, y)$, transformed by the projector PPM. Similarly, the pixel indices $(x, y)$ define the camera image plane coordinates $x_c = x$, $y_c = y$ of the object point projected onto the camera coordinate system. Given the camera and the projector PPMs, world coordinates of the object point can be calculated according to equation 21 in Section 2.1.

### 3.2.1 Pixel fidelity estimation

It is obvious that although both $J_k(x, y) = 0.95$ and $J_k(x, y) = 0.55$ will be binarized as $B_k(x, y) = 1$, they should definitely be treated differently. In the first case, one may say that
Figure 5: Top row (from left to right): raw image corresponding to pattern 3, normalized intensity image, binary image and fidelity image. All images are normalized. Bottom: profile of a vertical line from the normalized intensity image (dashed blue) and the binary image (solid red).
the pixel \((x, y)\) in image \(k\) is indeed illuminated with high probability, whereas in the second case the probability of that pixel to be non-illuminated is almost equal to the probability of it being illuminated.

In order to give a quantitative measure of the pixel fidelity, let us assume that the measured normalized intensity image \(J_k(x, y)\) is obtained from some ideal binary image \(B_0^k(x, y)\) contaminated by zero-mean Gaussian noise \(\xi(x, y)\) with variance \(\sigma^2\). In case \(J_k(x, y) > 0.5\), the pixel fidelity can be defined as

\[
F_k(x, y) = \Phi \left( \frac{0.5 - J_k(x, y)}{\sigma} \right),
\]

where \(\Phi\) denotes the c.d.f. of the normal distribution. The lower value of \(F_k\) is 0.5 and it is obtained when \(J_k(x, y) = 0.5\). Similarly, for \(J_k(x, y) < 0.5\), the fidelity is

\[
F_k(x, y) = 1 - \Phi \left( \frac{0.5 - J_k(x, y)}{\sigma} \right) = \Phi \left( \frac{J_k(x, y) - 0.5}{\sigma} \right).
\]

Binarization errors in images corresponding to the most significant bit of the stripe code affect more the resulting stripe id \(T\) than errors in the least significant bit. Therefore, pixel fidelity in each stripe should be weighted by the stripe significance. We define the pixel fidelity as the weighted sum of the pixel fidelities in all stripes

\[
F(x, y) = \sum_{k=1}^{N} w_k F_k(x, y) = \sum_{k=1}^{N} w_k \Phi \left( \frac{0.5 - J_k(x, y)}{\sigma} \right),
\]

where \(w_k = 2^{-k}\) is the significance of stripe \(k\), and \(N = 8\) in our implementation. The variance \(\sigma^2\) was set empirically to 1. Figure 6 shows the decoded stripe code \(T\) of a scanned object and the fidelity image \(F\).

Pixels fidelity is an important side information and can be used, for example, for weighted mesh smoothing or decimation. In the next subsection we discuss an additional use of the pixel fidelity map for obtaining sub-stripe resolutions.

### 3.2.2 Sub-stripe resolution

One of important concerns in structured light systems is sub-stripe resolution. The stripe code \(T\), which usually has granular nature due to quantization can be approximated by a continuous smooth surface, taking into account the fidelity map. We used a separable cubic spline, which was fitted to the image \(T\) with weighting inverse proportional to pixel fidelity. As the result, a smooth stripe id image \(\tilde{T}\) with sub-stripe resolution was obtained.

Let us denote by \(B_v\) and \(B_u\) the orthonormal spline bases corresponding to the rows and the columns of \(T\), respectively. Decomposition of the \(T\) in the two-dimensional separable
basis obtained as the tensor product of $B_v$ and $B_u$ can be expressed as

$$C = B_u^T T B_v,$$

or, alternatively, as

$$c = B^T t,$$

where $t$ is the column-stack representation of $T$, $c$ is a vector of spline coefficients and $B = B_v \otimes B_u$ is the Kronecker product of the row and the column bases. Weighted spline fitting constitutes to finding such spline coefficients $c$ that minimize

$$\sum_k \frac{1}{f_k} ((Bc)_k - t_k)^2,$$

where $f_k$ denotes the fidelity of the pixel $t_k$. We also add a controllable penalty on irregularity of the smoothed image $\tilde{t} = Bc$. In matrix notation, the weighted spline fitting problem reads

$$c = \operatorname{argmin} \left\{ \|WBc - Wt\|_2^2 + \lambda \|DBc\|_2^2 \right\},$$

where $W = \operatorname{diag} \left\{ \frac{1}{f_k} \right\}$ is the weighting matrix, $D$ is the matrix defining the irregularity penalty and $\lambda$ is the smoothness parameter, controlling the tradeoff between smoothness of $t$ and faith to the original data.

Figure 6: Decoded stripe ids with level sets displayed as a contour plot (left) and pixels fidelity image (right).
Problem 40 has an analytic solution

\[ c = \left[ B^T B + \lambda (D B)^T (D B) \right]^t, \tag{41} \]

where \( A^\dagger = (A^T D)^{-1} A \) denotes the Moore-Penrose pseudoinverse.

Since different amounts of smoothing should be usually applied in the horizontal and the vertical directions of \( T \), it is reasonable to use two penalty factors with two separate smoothness parameters, which control the smoothness in each direction. We used the \( L_2 \) norm of a finite difference operator as the penalty factor, yielding

\[ c = \left[ B^T B + \lambda_x (D_x B)^T (D_x B) + \lambda_y (D_y B)^T (D_y B) \right]^t, \tag{42} \]

where \( D_x \) and \( D_y \) are the row-wise and the column-wise discrete derivative operators and \( \lambda_x \) and \( \lambda_y \) are smoothness parameters controlling the smoothness of \( \tilde{T} \) in the \( x \)- and \( y \)-direction, respectively. The resulting smooth stripe id image \( \tilde{T} \) is given (in column stack representation) by \( \tilde{t} = Bc \).

Figure 7 presents the \( T \) image of a human face obtained by directly decoding the stripe ids and by using fidelity-weighted smooth cubic spline fitting. Note that quantization noise is less significant in the latter case. Figure 8 depicts the reconstructed surface with and without using sub-stripe resolution.

Figure 7: Stipe id image \( T \) of a human face shown as a surface, without using sub-stripe resolution (left) and using sub-stripe resolution (right).
Figure 8: Top: reconstructed $Z$-map of a human face shown as a surface, without using sub-stripe resolution (left) and using sub-stripe resolution (right). Bottom: vertical profile of the $Z$-map without using sub-stripe resolution (solid black) and using sub-stripe resolution (dashed red).
3.2.3 Reconstruction of human faces

The main use of our scanner was for three-dimensional face recognition [10]. Figure 9 presents an example of reconstruction of a human face. The full illumination image $I_H$ served as the texture, which was mapped onto the triangulated mesh, reduced to 10% of the original resolution for computational tractability.

Texture mapping is straightforward in the described setup, since there is one-to-one correspondence between texture pixels and mesh vertices.

![Reconstruction of a human face](image)

Figure 9: Reconstruction of a human face rendered using the Phong lighting model (top) and with mapped texture (bottom) shown at four different view angles.

3.3 Calibration

To find the camera and the projector matrices, we used an automatic calibration procedure described in Section 2.3. As the calibration object, a precisely built $15 \times 15 \times 15cm$ wooden pyramid attached with 3 mutually perpendicular planes attached to a "background" plane, was used. Object’s surfaces were made nearly Lambertian using white mate finishing.
228 circular fiducial points were marked on the calibration object surfaces. The points were grouped into sets of collinear equally-spaced marks, 3 sets of 33 points each on the background plane, and two sets of 22 points each on each of the pyramid sides (Figure 10). WCS was defined as the local coordinate system of the calibration object.

At the first stage, the calibration object was scanned and a sub-stripe resolution stripe id image $\tilde{T}$ was calculated from the raw data. At the second stage, the full-illumination image $I_H$ was binarized, the fiducial points were automatically located and their centroid coordinates were calculated. A set of 228 stripe ids at locations corresponding to the fiducial points centroids, together with the centroid locations themselves was used as the input for the calibration algorithm. For numerical stability, WCS coordinates of the calibration objects, the image plane coordinates of the camera and the projector stripe ids were normalized to $[-1, 1]$.

First, camera and projector PPMs that minimize the forward projection error were found by solving Problem 31. These matrices were used as the initial point of the Newton algorithm, which converged to the solution of Problem 33, yielding camera and projector matrices, which minimize the backprojection error. Figure 11 depicts the forward projection of the theoretically calculated calibration object fiducial points, onto the camera and the projector coordinate system. The measured fiducial points centroids and projector stripe ids are shown as a reference.

Table 2 shows the RMS and the maximum reconstruction errors of the calibration object in five tests with random camera and projector locations. Two cases are studied: 1) when the camera and the projector matrices are obtained by minimizing the forward projection error and 2) when the camera and the projector matrices are obtained by minimizing the backward projection error. The errors were calculated on a set of 100,000 points, with the analytical plane equations of the calibration object serving as a reference. Table 3 shows the improvement of the RMS and the maximum reconstruction error when using the optimal backprojection instead of the optimal forward projection. RMS was improved in all tests (improvement ranging from 1.44% to 45.66%, 12% in average). Maximum error was improved in all tests except Test 2, where the maximum error obtained using the optimal backprojection worsened by 6.58% (the fact that improvement in the RMS error was observed in Test 2, suggest that the degradation in the maximum error might have been caused by a spurious pixel). Maximum improvement of about 50% was obtained in Test 5.

The RMS error was about 0.5 mm in Tests 1-3 and about 1.6 mm in Tests 4-5. The latter degradation of the reconstruction quality was due to the highly oblique position of the calibration object with respect to the camera, which resulted in lower SNR, since less light was reflected from the object planes.

Figure 12 shows two profiles of the reconstructed calibration object. Analytical object profiles are shown as a reference.
<table>
<thead>
<tr>
<th>Test</th>
<th>Optimal projection</th>
<th>Optimal backprojection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMS (mm)</td>
<td>Max. (mm)</td>
</tr>
<tr>
<td>1</td>
<td>0.0624</td>
<td>0.2790</td>
</tr>
<tr>
<td>2</td>
<td>0.0626</td>
<td>0.2881</td>
</tr>
<tr>
<td>3</td>
<td>0.0423</td>
<td>0.2054</td>
</tr>
<tr>
<td>4</td>
<td>0.1599</td>
<td>0.6823</td>
</tr>
<tr>
<td>5</td>
<td>0.1579</td>
<td>0.6170</td>
</tr>
</tbody>
</table>

Table 2: RMS and maximum errors for reconstruction of the calibration object using the optimal forward projection and the optimal backward projection.

<table>
<thead>
<tr>
<th>Test</th>
<th>RMS</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.15%</td>
<td>21.57%</td>
</tr>
<tr>
<td>2</td>
<td>1.62%</td>
<td>-6.58%</td>
</tr>
<tr>
<td>3</td>
<td>1.44%</td>
<td>19.21%</td>
</tr>
<tr>
<td>4</td>
<td>2.76%</td>
<td>7.67%</td>
</tr>
<tr>
<td>5</td>
<td>45.66%</td>
<td>49.68%</td>
</tr>
</tbody>
</table>

Table 3: Improvement in the RMS and the maximum error in reconstruction of the calibration object when using the optimal backprojection instead of the optimal forward projection.
Figure 10: Calibration object with marked fiducial points. Axes units are cm.
Figure 11: Reprojected data using the optimal backward projection. Top: object points reprojected to the camera image plane (red circles) and the measured fiducial points centroids (black crosses). Bottom: object points reprojected to the projector coordinate system (dashed red) and the measured stripe ids (solid black). Coordinates are normalized.
Figure 12: Two profiles of the calibration object reconstructed using the optimal forward projection (solid red) and the optimal backprojection (dash-dotted blue). The real profile is shown in dashed black. All units are given in cm.
4 Conclusions

In this work, we presented a structured light range scanner. Our scanner was built of off-the-shelf parts and it provides a competitive alternative to high-resolution commercial-grade structured light scanners at significantly lower costs. Our scanner architecture is flexible and allows, for example, an easy replacement of the stripe encoding method.

We presented a novel single-step automatic active stereo calibration method, based on optimal backprojection estimation. Experimental results showed that our method improves the results provided by previous methods, like the one in [6]. We also presented a method for estimating local fidelity of the range image and showed how this information can be used to obtain sub-stripe resolution.

5 Acknowledgement

The system was built in the Geometric Image Processing Laboratory (GIP), Department of Computer Science, Technion - Israel Institute of Technology. Kind STL support is acknowledged.

References


