3D depth recovery with grayscale structured lighting

S. Savarese  JY. Bouguet  P. Perona

California Institute of Technology 136-93
Pasadena, CA 91125
{savarese,bouguetj,perona}@vision.caltech.edu

Keywords: 3D Shape recovery, Structured lighting

Abstract

Two new structured-lighting methods for extracting the three-dimensional shape of a scene are proposed. The projected patterns are grayscale ramps and sinewaves, rather than the well-known black-white stripes. While black-white structured light methods allow to trade off measurement speed with sample density, our grayscale methods provide constant sample density and trade off measurement speed with accuracy. Another interesting difference is that black-white methods sample depth at unpredictable locations with respect to the pixel grid of the camera, while grayscale methods sample scene depth at each pixel. The accuracy of each method is estimated analytically and assessed experimentally.

1 Introduction and motivation

There exist different methods for recovering the 3D structure of surfaces: either based on mechanical systems, or purely based on images (photogrammetry). The general framework of our system is structured lighting. Structured lighting is based on projecting light patterns on the three dimensional scene, and infer its shape from the images acquired from a camera placed at a different location in space. There are different versions of this class of methods, mainly depending on the choice of the projected patterns. Many techniques rely on binary stripe patterns or succession of vertical black and white stripes across the image [6, 5, 4].

Disadvantages of this approach are that, for instance, it is no possible to build a dense depth map representation of the scene. In other words, the depth information can not be retrieved for all pixels in the image. Moreover, since the response in time of the projector is limited, the time required to project the entire sequence of patterns cannot be reduced arbitrarily. For instance, LCD projectors are popular devices for projecting structured light. Unfortunately LCD's have long settling times and therefore stripe patterns, which require large swings in transparency, require long settling times as well. Finally, there is no direct trade-off between accuracy and speed. That is, increasing the number of projected patterns (therefore increasing the time required to scan the object) leads to increase the density of the depth map but not necessarily the accuracy in the depth estimation.

We propose two new techniques that differ from previous methods mainly in the nature of patterns [3]. Both techniques are based on projecting grayscale patterns. Along the horizontal direction the brightness is a function of the position, whereas along the vertical direction it is uniform. Therefore every physical point in the scene is illuminated by a light intensity depending only on the horizontal coordinate \( x_p \) in the projector reference frame. Then we build a correspondence between the brightness of an observed point and \( x_p \). Finally, using triangulation, we compute 3D shape (3D location of points in space). However in this paper we focus on recovering \( x_p \) and quantifying the noise attached to it. Since triangulation and its sensitivity analyses are well-studied we will not repeat that analysis here[1, 6, 2].

The proposed techniques differ in the kind of correspondence used. The first one is based on projecting only one pattern whose brightness profile is chosen to be a monotonic function of the horizontal position (see figure 3). Thus, every point in the scene is illuminated by a light intensity that is univocally dependent on the horizontal coordinate \( x_p \) in the projector reference frame. The correspondence is solved using two more patterns in order to normalize the observed brightness for every pixel. Furthermore, in order to improve the accuracy in shape estimate it is possible to project a sequence of patterns (such as in figure 5).

The second technique assumes to project a sequence of sinusoidal patterns (see figure 6). In this case every point in the scene is illuminated by a light intensity that is periodic in time with a phase linearly dependent on its horizontal coordinate \( x_p \) in the projector reference frame. Extracting the phase of the intensity function can directly lead to an estimate of \( x_p \).

The approach we are proposing guarantees a dense depth representation of the scene. Besides, there are only small intensity changes among successive patterns. Therefore from an hardware implementation standpoint, such patterns may very well be generated faster than conventional binary, black and white patterns which induce large swings of brightness on the pixels at the transitions. Also,
since we do not need to extract stripe boundaries on the images as other binary stripe techniques typically do, our method is expected to be more tolerant to defocused projector. Finally, while methods based on binary stripe patterns have fixed accuracy and trade off between resolution and number of projected patterns (i.e. acquisition time), the methods that we propose have fixed resolution (one depth measurement per pixel) and trade off between accuracy with number of projected patterns. That is, they have the possibility of either increasing the accuracy of the 3D shape estimate using a larger number of projected patterns or getting a faster shape estimate no matter the quality of the reconstruction.

The rest of the paper is organized as follows. We will describe in section 2 the final 3D reconstruction step that leads to the 3D shape of the scene (geometrical stage). We then explore in sections 3 and 4 the details of the proposed techniques together with an error analysis. In 5 and 6 we compare the two approaches using the results of the sensitivity analysis and a certain number of experiments. Finally, we give final remarks and suggestions for future works.

2 Description of the setup — Depth measurement

The scene will be represented here by an object facing a camera, and a light projector (see figure 1).

Let us suppose that the projector projects a pattern as shown in figure 3. Only the \( x_p \) coordinate in the projector reference frame will be directly observable (i.e., in the projector image, there is no disparity information between points along vertical band). Hence one can only directly indentify which vertical plane \( \Pi \) lit a given point \( \pi_e \) observed on the camera image, not its vertical position along that plane. From the camera however, the full image of the scene is observed, which means that we know that each pixel \( \pi_e = [x_e, y_e]^T \) corresponds to the projection of a point \( P \) in space that lies on the optical ray \( (O_e, \pi_e) \) (see figure 2). Given that we can associate to each pixel \( \pi_e \) its corresponding coordinate in the projector image \( x_p \) (sections 3 and 4 will describe how this stage may be performed), one can recover the 3D location of the observed point in space \( P \) by intersecting the projected plane \( \Pi \) with the optical ray \( (O_e, \pi_e) \) (see figure 2). This 3D recovery stage is called triangulation. In figure 2, the coordinate vector of \( P \) in the camera frame is denoted \( \mathbf{x}_e \).

Within the main description of the method, we will limit ourself to denote the triangulation operation with the symbol \( \Delta \):

\[
\mathbf{x}_e = \Delta(\pi_e, x_p)
\]  

Notice that the only two pieces of information required to perform the triangulation are the pixel coordinate vector \( \pi_e \) in the image plane and the horizontal projector coordinate \( x_p \) associate to that pixel.

---

**Figure 1. The general setup.** The camera and the projector are facing the scene consisting of one or more objects. The projector projects vertical planes defined by their horizontal position (similar to standard vertical stripe methods) each of them associated with a value of brightness.

**Figure 2. Extracting the shape.** Given the observed brightness of a lit point \( P \) in the object (observed from the camera), it is possible to identify the horizontal coordinate of its projection onto the projector image \( (x_p) \). Let us suppose that \( \pi_e \) is its projection on the camera image, one can estimate the 3D location of \( P \) by triangulation. The triangulation process consists of intersecting the optical ray given by \( (O_e, \pi_e) \), in the camera image with the plane defined by the horizontal coordinate only \( x_p \), in the projector image. Given that the relative spatial locations of the two devices, projector and camera, is known, one can naturally compute that intersection, and obtain the coordinate of the point in the camera frame \( \mathbf{x}_e \).
3 Monotonic pattern technique

3.1 Correspondence

Let us suppose a pattern is projected on the scene and let \( B(x_p, y_p) \) be its grayscale intensity map in the projector plane. This pattern is constant along the vertical direction \((y_p)\) and has a monotonic profile along the horizontal direction \((x_p)\). Figure 3 gives one example of such a profile.

Consider a given pixel \( \mathbf{p} = (x, y) \) on the image plane. This pixel corresponds to a unique point \( P \) in the scene. This point \( P \) is also illuminated by a unique vertical stripe defined by one horizontal coordinate \( x_p \) in the projector image. From the projected intensity \( B(x_p, y_p) \) at \( P \), one could identify \( x_p \). However there is not direct access to the projected light \( B(x_p, y_p) \) from the projector, but only to the reflected light leaving the scene and going to the camera sensor. In other words we just know the observed brightness \( I(x, y) \) at pixel \( \mathbf{p} \). The goal is then to estimate for every pixel the corresponding \( x_p \) from \( I(x, y) \).

Let \( I_{max}(x, y) \) be the maximum observed brightness at pixel \( \mathbf{p} = (x, y) \) when it is illuminated by maximum brightness \( B_{max} \), defined in figure 3. Similarly, let \( I_{min}(x, y) \) be the minimum observed brightness at pixel \( \mathbf{p} \) when it is illuminated by the minimum brightness \( B_{min} \).

The first step consists of extracting \( I_{max}(x, y) \) and \( I_{min}(x, y) \) by projecting the two maximum and minimum uniform patterns \( B(x_p, y_p) = B_{max} \) and \( B(x_p, y_p) = B_{min} \).

Then project the pattern shown on figure 3 and let \( I(x, y) \) be the observed brightness at a given pixel \( \mathbf{p} \). Notice that if \( I(x, y) = I_{max}(x, y) \) then the corresponding point \( P \) in the scene is illuminated by the first vertical stripe \((x_p = x_{min})\). Likewise, when \( I(x, y) = I_{min}(x, y) \) then the corresponding point \( P \) is illuminated by the last vertical stripe \((x_p = x_{max})\). In these two special cases the correspondence between \( \mathbf{p} \) and \( x_p \) is solved. The next step consists of solving for correspondence at every pixel in the image. For that purpose, let us introduce the normalized observed brightness \( I_n(x, y) \), defined as follows:

\[
I_n(x, y) = \frac{I(x, y) - I_{min}(x, y)}{I_{max}(x, y) - I_{min}(x, y)}
\]  \(\text{(2)}\)

In addition, let \( B_n(x_p, y_p) \) be normalized projected brightness and define it as follows:

\[
B_n(x_p, y_p) = \frac{B(x_p, y_p) - B_{min}}{B_{max} - B_{min}}
\]  \(\text{(3)}\)

The profile of this function is shown on figure 4.

The following property constitutes the key element behind the idea exposed in this paper.

**Property 1:** For every pixel \( \mathbf{p} = (x, y) \) of the image there exists a point \( \mathbf{p}_n = (x_p, y_p) \) on the projector plane such that \( I_n(x, y) = B_n(x_p, y_p) \).

**Proof:** Suppose that the following hypothesis are satisfied:

- The observed brightness at a given pixel \( \mathbf{p} \) is a linear function of the bright intensity projected on the corresponding point \( P \) in the scene. This is true if the gain of the camera is linear and does not introduce saturation effects.
- Every point \( P \) in the scene is illuminated by a light intensity that is only depending upon the horizontal coordinate \( x_p \). In other words two or more different points on the projector plane do not affect the incident light to the same point \( P \) in the scene. This is equivalent to supposing that effects due to internal reflection are negligible.
- The above mentioned linear relationship is not a function of time. This is true, for example, if the gain of the camera does not change in time.

Under these conditions, for every pixel \( \mathbf{p} = (x, y) \) of the image, there exist two coefficients \( \alpha \) and \( \beta \) and a point \( \mathbf{p}_n \) on the projector plane such that:

\[
I(x, y) = \alpha B(x_p, y_p) + \beta
\]  \(\text{(4)}\)

The coefficients \( \alpha \) and \( \beta \) capture scene geometry, surface properties and sensor charateristics. More specifically, \( \alpha \) is function of the relative position of the scene surface with respect to the light source and camera as well as the surface texture properties. In addition, shadowed regions in the scene (self-shadows) will correspond to very small values of \( \alpha \). The coefficient \( \beta \) captures the offset of the image sensor.
Let us write the equation 15 in the following particular cases:

\[ I_{\text{max}}(x, y) = \alpha(x, y)B_{\text{max}} + \beta(x, y) \]
\[ I_{\text{min}}(x, y) = \alpha(x, y)B_{\text{min}} + \beta(x, y) \]

Therefore the normalized brightness can be written as follows:

\[ I_n(x, y) = \frac{I(x, y) - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \]
\[ = \frac{B(x_p, y_p) - B_{\text{min}}}{B_{\text{max}} - B_{\text{min}}} \]
\[ = B_n(x_p, y_p) \]

which completes the proof of Property 1.

Since \( B(x_p, y_p) \) is a monotonic function along the horizontal direction \( (x_p) \), it is invertible in \( x_p \). Therefore so is \( B_n(x_p, y_p) \). Let \( B_n^{-1} \) be its inverse function. Consequently, knowing \( I_n(x, y) \), the horizontal coordinate \( x_p \) associated to any pixel \( \pi \) is given by:

\[ x_p = B_n^{-1}(I_n(x, y)) \]

This correspondence is shown on figure 4.

In conclusion the correspondence between \( \pi \) and \( x_p \) is solved for every pixel.

### 3.2 Noise Sensitivity Analysis

As shown in the previous section, the main idea of this technique is based on extracting the normalized observed brightness for every pixel and computing the corresponding horizontal coordinate \( x_p \). The purpose of the noise sensitivity analysis is to quantify the effect of the noise in the measurement data \( I(x, y) \), \( I_{\text{min}}(x, y) \) and \( I_{\text{max}}(x, y) \) on the \( x_p \) estimation. Let us model such a brightness noise by an additive Gaussian random variable with zero-mean and variance \( \sigma_I^2 \) (uniform across the whole image).

Let \( \sigma_{x_p}^2 \) be the variance of the noise attached to the estimate of \( x_p \). Looking at equation 8, \( \sigma_{x_p}^2 \) can be computed as follows (for clarity the variables \( x \) and \( y \) are omitted):

\[ \sigma_{x_p}^2 = \left( \frac{\partial B_n^{-1}}{\partial I_n} \right)^2 \sigma_I^2 \]

(9)

where (see equation 7):

\[ \sigma_{I_n}^2 = \left( \frac{\partial I_n}{\partial I} \right)^2 \sigma_I^2 + \left( \frac{\partial I_n}{\partial I_{\text{max}}} \right)^2 \sigma_{I_{\text{max}}}^2 + \left( \frac{\partial I_n}{\partial I_{\text{min}}} \right)^2 \sigma_{I_{\text{min}}}^2 \]

(10)

where \( \sigma_I \), \( \sigma_{I_{\text{max}}} \), and \( \sigma_{I_{\text{min}}} \) are the standard deviations of the noise attached respectively to the current observed brightness, to the maximum observed brightness and to minimum observed brightness at pixel \( \pi = (x, y) \). Assuming such quantities to be equal, we obtain:

\[ \sigma_{I_n}^2 = \left[ \left( \frac{\partial I_n}{\partial I} \right)^2 + \left( \frac{\partial I_n}{\partial I_{\text{max}}} \right)^2 + \left( \frac{\partial I_n}{\partial I_{\text{min}}} \right)^2 \right] \sigma_I^2 = K \sigma_I^2 \]

(11)

Since \( \frac{\partial B_n^{-1}}{\partial I_n} = 1/\frac{\partial B_n}{\partial x_p} \), \( \sigma_{x_p}^2 \) becomes:

\[ \sigma_{x_p}^2 = \frac{1}{\left( \frac{\partial B_n}{\partial x_p} \right)^2} K \sigma_I^2 \]

(12)

Now it is easy to find an expression \( K \):

\[ K = \frac{1}{(I_{\text{max}} - I_{\text{min}})^2} + \frac{(I_{\text{max}} - I_{\text{min}})^2}{(I_{\text{max}} - I_{\text{min}})^4} + (I_{\text{max}} - I_{\text{min}})^2 \]

(13)

Looking at \( K \) as function of \( I \), we find out that \( K \) ranges between \( 0 \) and \( \frac{3}{2(I_{\text{max}} - I_{\text{min}})^2} \) when \( I = I_{\text{min}} \) and \( \frac{2}{(I_{\text{max}} - I_{\text{min}})^2} \) when \( I = I_{\text{max}} \) or \( I = I_{\text{min}} \).

Therefore:

\[ \sigma_{x_p}^2 = \frac{\gamma \sigma_I^2}{\left( \frac{\partial B_n}{\partial x_p} \right)^2 (I_{\text{max}} - I_{\text{min}})^2} \leq \frac{2 \sigma_I^2}{\left( \frac{\partial B_n}{\partial x_p} \right)^2 (I_{\text{max}} - I_{\text{min}})^2} \]

(14)

with \( \gamma \in [2/3, 2] \).

From equation 14 we can derive two interesting remarks.
First, $\sigma_{xp}$ is inversely proportional to $\left( \frac{\partial B}{\partial x_p} \right)$. This confirms an intuitive consideration: the greater is the slope of the brightness intensity profile $B(x_p,y_p)$ at $x_p$ (see figure 4) the less sensitive is the estimate of $x_p$. Such an observation suggests us a possible way to design the projected pattern. We may increase the slope of the projected brightness profile where we need to have a more reliable estimate of $x_p$. Of course we cannot increase the slope for every pixel. However in the next section we will show that using multiple patterns we can solve this problem.

Second, $\sigma_{xp}$ is inversely proportional to the difference between $I_{\max}$ and $I_{\min}$. Therefore for every pixels where $I_{\max} - I_{\min}$ is large, a better accuracy in the $x_p$ estimate is guaranteed. This condition is not satisfied, for instance, in areas where the object to scan looks dark. Furthermore, in order to improve the accuracy, we may choose $B_{\min}$ small and $B_{\max}$ large (see figure 3). However we need to pay attention that saturation effects do not occur (see the first hypothesis in property 1).

### 3.3 How to improve the accuracy of $x_p$

The simplest way to improve the accuracy of the $x_p$ estimate is to increase the derivative of the intensity profile $B(x_p,y_p)$. For instance, one possibility is to project the double ramp pattern shown in figure 5 (left). Of course, now the observed brightness at a given pixel $\mathbf{P}$ will not univocally correspond any longer to the coordinate $x_p$ in the projector plane. Hence, in order to solve correctly the correspondence, we need to use two patterns. A first pattern (such as the single ramp in figure 3 will be necessary to estimate (even roughly but univocally) the corresponding $x_p$. With a second pattern (e.g. the double ramp in figure 5 (left)) we can improve the accuracy of the estimate solving the ambiguity using the previous result. Using multiple ramp patterns (e.g as in figure 5(right)) better and better results can be achieved. For instance a sequence of 8 patterns with 1,2,4,8,16,32 ramps can be used to univocally estimate $x_p$ with an accuracy related with the slope of the 32-ramps profile.

### 4 Sinusoidal pattern technique

This technique is based on projecting a succession of horizontally shifted grayscale sinusoidal patterns (see section 4.1). Then, from the temporal brightness information collected from the images, we recover, at every pixel $\mathbf{P}$, the projector coordinate $x_p$. We will first present the projected patterns themselves in section 4.1, and then derive the necessary machinery to extract the projector coordinate $x_p$ in section 4.2.

#### 4.1 Sequence of patterns

The projector projects a succession of $N$ grayscale patterns which are translated from one to another along the horizontal direction (in the $x_p$ direction). The figure 6 shows four samples of patterns. In that particular experiment, we chose a sinusoidal waveform, and $N = 32$ patterns. Fewer patterns might be sufficient.

Figure 7 shows the horizontal profile of two patterns, the first one and the ninth one (therefore as a function of $x_p$). Notice how pattern #9 is shifted to the right by a quarter of period ($\pi/4$ phase) with respect to pattern #0.
4.2 Correspondence

As each pattern is projected onto the scene, one camera image is acquired. Throughout the sequence, one given pixel $\mathbf{p}_c$ corresponds to a unique point $P$ in the scene. This point $P$ is also illuminated by a unique vertical stripe defined by one horizontal coordinate $x_p$ in the projector image. We know from section 2 that we can recover the 3D location of $P$ given that we know the coordinate $x_p$ associated to the pixel $\mathbf{p}_c$. We will show here that given the temporal brightness signal at a pixel $\mathbf{p}_c$, we can infer its corresponding projector coordinate $x_p$, hence infer the 3D shape.

If we observe the temporal patterns of the incident light at two points $P_1$ and $P_2$ in the 3D scene, they only differ from the phase. Indeed, they are both sinusoidal signals, but one is shifted with respect to the other by an amount that is directly related to the difference of their projector coordinates $x_p$. For example, the temporal signal attached to a point illuminated by the middle projector stripe $x_p = 320$ will be $\pi$ phase shifted with respect to that of a point illuminated by the first stripe $x_p = 1$. There is therefore a linear one-to-one map between the temporal phase shift, and the projector coordinate. Extracting the phase shift of the incident light signal corresponds to estimating the coordinate in the projector image of the vertical stripe that lit the observed point in the scene.

However, we don’t have direct access to the incident light going in the scene, but only the reflected light leaving the object and going to the camera sensor. If we assume that the material reflection function in the scene and the imaging sensor (the camera) have significantly linear behaviors, we can still make the same phase statement on the temporal brightness waveform collected in the images for every pixel. Therefore, the problem of extracting projector coordinate at a given pixel in the image directly translates into estimating the phase of the temporal brightness signal at that pixel.

Figure 8 shows the temporal brightness at 5 different pixels located on the same row in the image, as a function of time (or patterns index). Notice that the waveforms are all sinusoidal, and differ one from the other in amplitude ($A$), offset ($B$) and phase ($\Phi$). They all have the same frequency: $\omega_0 = 2\pi/N$, where $N = 32$ is the number of patterns.

Let us define $I(n)$ to be the observed temporal brightness function at a given pixel $\mathbf{p}_c = (x_c, y_c)$ in the image, as a function of $n$, the pattern number ($n$ goes from 0 to $N - 1$ and is sometimes associated to time). For clarity reasons, we will not index $I(n)$ with the pixel location $\mathbf{p}_c$. However the reader needs to keep in mind that this function is different from pixel to pixel.

We can model $I(n)$ in a similar fashion as we did in section 3:

$$I(n) = A \sin(\omega_0 n - \Phi) + B$$

Given the type of pattern we use in that particular case (a single period sinusoidal waveform), the phase shift $\Phi$ can be shown to be linearly related to the projector coordinate $x_p$ through the following one-to-one equation:

$$x_p = \frac{N_p \Phi}{2\pi}$$

if $\Phi$ is assumed to be expressed in the $0 - 2\pi$ range, and $N_p$ is the width (in pixel) of the projector image ($N_p = 640$ pixels here).

Therefore, estimating $x_p$ at every pixel is equivalent to estimating the phase shift $\Phi$ of the associate temporal brightness function.

Define now two quantities $a$ and $b$ as follows:

$$a = \langle I(n), \sin(\omega_0 n) \rangle = \frac{2}{N} \sum_{n=0}^{N-1} I(n) \sin(\omega_0 n)$$

$$b = \langle I(n), \cos(\omega_0 n) \rangle = \frac{2}{N} \sum_{n=0}^{N-1} I(n) \cos(\omega_0 n)$$

Given the model equation 15 for the temporal brightness function $I(n)$, it is relatively straightforward to the following properties for $a$ and $b$:

$$\begin{cases}
    a = A \cos(\Phi) \\
    b = A \sin(\Phi)
\end{cases}$$

The most interesting feature of those relations is that neither $a$ nor $b$ contain the offset $B$. That
allows to isolate and extract independently the amplitude $A$ and the phase $\Phi$:

$$\begin{align*}
\Phi &= \text{arg}(a + ib) = \arctan(b/a) \\
A &= \|a + ib\| = \sqrt{a^2 + b^2}
\end{align*}$$

Notice that the arctan function here is assumed to return the argument in the $0-2\pi$ range without any $\pi$ ambiguity. In other words, we have access here to both values $a$ and $b$ not only the ratio $b/a$ between them. There is therefore no sign ambiguity in the two terms, which means that the phase is extracted with no $\pi$ ambiguity.

Finally, from relations 16 and 20 we obtain expressions for both the projector coordinate $x_p$ and the sinewave amplitude $A$:

$$x_p = \frac{N_p}{2\pi} \arctan \left( \frac{\langle I(n), \cos(\omega_0 n) \rangle}{\langle I(n), \sin(\omega_0 n) \rangle} \right)$$

$$A = \sqrt{\langle I(n), \sin(\omega_0 n) \rangle^2 + \langle I(n), \cos(\omega_0 n) \rangle^2}$$

Typically, pixels with large corresponding amplitudes $A$ will be more reliable for phase estimate than pixels with small amplitudes. This is used to help rejecting noisy regions in the image. If a pixel falls in a shadow region of the scene, it will not be lit by the projected sinewave pattern. And therefore, its associated temporal brightness signal will almost not change across time (except within the pixel noise). A similar situation occurs for dark regions of the scene, or regions poorly reflective. In these cases, any phase extraction is hopeless from the start. Fortunately, those situations nicely translate into significantly small amplitude estimates $A$.

Therefore, one can simply reject regions of the image that have a corresponding amplitude $A$ less than the pre-chosen threshold $A_T$.

### 4.3 Noise Sensitivity Analysis

The goal of a noise sensitivity analysis is to compute the effect of this input brightness noise onto the $x_p$ estimate. As we did in section 3.2 the brightness noise can be modeled by an additive Gaussian random variable with zero-mean and variance $\sigma_I^2$.

Let $\sigma_{x_p}^2$ be the variance of the error on the projector coordinate $x_p$. Equation 21 provides an expression for the projector coordinate $x_p$ (in pixel units) as a function of the $N$ brightness measurements at pixel $\tau_c$ across time $I(0), I(1), \ldots, I(N)$:

$$x_p = \frac{N_p}{2\pi} \arctan \left( \frac{\langle I(n), \cos(\omega_0 n) \rangle}{\langle I(n), \sin(\omega_0 n) \rangle} \right)$$

That is, $x_p = F(I(0), I(1), \ldots, I(N))$. Since all the brightness measurements $I(0), I(1), \ldots, I(N)$ are supposed to carry the same noise term of variance $\sigma_I^2$, the final projector coordinate variance $\sigma_{x_p}^2$ may be approximated by the following expression:

$$\sigma_{x_p}^2 = \sum_{n=0}^{N} \left( \frac{\partial F}{\partial I(n)} \right)^2 \sigma_I^2$$

After straightforward derivation, we may show that:

$$\frac{\partial F}{\partial I(n)} = \frac{N_p}{\pi AN} \cos(\omega_0 n + \Phi)$$

where $\Phi$ and $A$ are defined in equation 15. Therefore, the variance of the error on projector coordinate $x_p$ may be written:

$$\sigma_{x_p}^2 = \left( \frac{N_p \sigma_I}{\pi AN} \right)^2 \sum_{n=0}^{N} \cos^2(\omega_0 n + \Phi)$$

Recalling that $\omega_0 = 2\pi/N$, one may show that the following relation holds:

$$\forall \Phi \in \mathbb{R}, \quad \sum_{n=0}^{N} \cos^2(\omega_0 n + \Phi) = \frac{N}{2}$$

Therefore, the variance of the error attached to the projector coordinate $x_p$ takes the following compact form (in pixel units):

$$\sigma_{x_p}^2 = \frac{N_p^2}{2\pi^2 A^2 N} \sigma_I^2$$

A set of three fundamental observations may be drawn from that relation. First, notice that $\sigma_{x_p}^2$ is inversely proportional to the number of projected
patterns $N$. This is quite intuitive: as the number of patterns increases, the accuracy of phase estimate $\Phi$ increases, which in turns makes the projector coordinate $x_p$ more accurate. Second, notice that $\sigma^2_{x_p}$ is inversely proportional to the square of the brightness amplitude $A$. This supports the fact that pixels with larger temporal brightness variations (larger contrasts) are more reliable for phase estimation than ones with smaller brightness variations. This also adds a supportive argument in favor of the thresholding technique for rejecting too unreliable pixels. Finally, observe that as the projecting image width $N_p$ increases, the accuracy in estimating $x_p$ decreases. That is also quite intuitive: for a given error in phase estimate $\Phi$ (in radians), the corresponding error in estimating $x_p$ (in pixels) will be larger on wider projecting images. In consequence, for a given pattern width, it would be beneficial to project patterns with more than one sinusoidal period. Doing so, the effective width of a period is smaller ($N_p \rightarrow N_p/k$ where $k$ is the number of periods) decreasing the transferred error onto the projector pixel coordinate $x_p$ (the proportionality factor between $\Phi$ and $x_p$ is smaller - see equation 16). However if the projecting patterns contain more than a single period, an ambiguity in corresponding phase information to pixel coordinates is introduced (if $k = 2$, there are two valid pixel coordinates associated to any given phase value $\Phi$). In order to solve for that ambiguity, one could use a combination of high and low frequency patterns. The low frequency patterns (that could be either grayscale, or strict black and white stripes) would help disambiguate the high frequency ones providing the high local resolution. That is a subject for future work.

\section{An error comparison between both techniques.}

In section 3.2 and 4.3 we computed two expressions of the variance of noise attached to the horizontal projector coordinate $x_p$ (see equations 14 and 28 respectively). Let us call $\sigma^2_{x_p}$ the variance achieved with the monotonic pattern technique and $\sigma^2_{x_p}$ the variance achieved with sinusoidal pattern technique. We want now to find a more convenient expression for $\sigma^2_{x_p}$. Let us suppose the monotone brightness profile to be simply a linear profile. That is, $\frac{\partial B}{\partial x_p}$ is constant along $x_p$ and equal to $1/N_P$, where $N_P$ is the width (in pixels) of the image. Moreover it easy to recognize that $I_{max}(x,y) - I_{min}(x,y)$ is the amplitude $A$ of the sinusoid (see equations 28 and 15). Hence, $\sigma^2_{x_p}$ becomes:

$$\sigma^2_{x_p} = \frac{2N_P^2\sigma^2_A}{A^2}$$

(29)

Therefore:

$$\sigma^2_{x_p} = \frac{\sigma^2_{x_p}}{4N_P^2}$$

(30)

This simple relationship quantifies how better the sinusoidal technique estimates $x_p$ compared with the linear pattern technique. But let us see what happens if we try to estimate $x_p$ using the linear pattern technique with a multiple ramp pattern as suggested in section 3.3. Now $\frac{\partial B}{\partial x_p}$ becomes equal to $k/N_P$ where $k$ is the number of linear ramps in the pattern (e.g. for the pattern of figure 5 (left) $k = 8$). Hence, in order to make $\sigma^2_{x_p}$ the need $k = \sqrt{4N_P}\pi^2$ and therefore a pattern with $k$ ramps. An numerical example will show the advantages of the multiple ramp approach. As we will see in the experimental section, if we want to achieve a small error in the $x_p$ estimate (no more than 2 pixels) with the sinusoidal pattern technique, we need to project no less than 30 sinusoidal patterns. So let us suppose $N = 30$. Almost the same result can be obtained with the multi-ramp linear pattern technique with only 8 patterns. Indeed using a sequence of patterns with 1, 2, 4, 8, 16, 32 ramps we are able to build the $x_p$ map univocally (as we described in section 3.3). Besides we guarantees almost the same accuracy of the sinusoidal method through the last pattern (with 32 ramps, so with $k = 32 \approx \sqrt{4N_P}\pi^2 = 34.41$). Two more patterns are needed to compute the normalized brightness.

\section{Experimental results}

In this section we first estimate experimentally the accuracy of the horizontal projector coordinate $x_p$ in order to validate the theoretical analysis. We then compare the performances of two methods and demonstrate the method on a complex scene.

\subsection{Sinusoidal pattern method: experiments}

In order to measure in a convenient way the uncertainty attached to the parameters computed in the theoretical analysis, we started with a simple object (e.g. the plane of figure 9 (top left)). We first acquired a set of 20 images in order to compute experimentally the image brightness noise $\sigma_I$. We found an estimate for $\sigma_I$ between 2 and 3 brightness units. Then, after projecting 30 sinusoidal patterns
as described in section 4, we computed the horizontal projector coordinate \( x_p \) for all pixels satisfying the threshold condition on the brightness amplitude. Figure 9 (bottom left) shows the values of \( x_p \) in the horizontal image coordinate scanline \( y_c = 200 \). We can notice how linearly the projector coordinate increases while going from the left to the right side of the scene. This is naturally expected since the observed object is planar.

We then quantified the noise attached to \( x_p \). Equation ?? lets us compute a theoretical estimate of \( \sigma_{x_p} \). Each dot in figure 9 (top right) is the value of \( \sigma_{x_p} \) predicted by such an equation in the horizontal image scanline \( y_c = 200 \). Notice that \( \sigma_{x_p} \) ranges between 0.7 and 1.6 pixels. Similar results have been achieved experimentally. We fitted a line to a neighbor neighborhood (100 pixels) of the \( x_p \)-map and we then computed the residual deviations of all the points of the neighborhood to the line. In particular figure 9 (bottom right) shows the residuals computed for pixels between \( x_p = 200 \) and 300 along \( y_c = 200 \). In other words the figure shows the \( x_p \) values once the linear signal component is removed, that is, it shows the reconstruction noise. The experimental estimate of \( \sigma_{x_p} \) has been obtained computing the standard deviation of such a reconstruction noise, getting a numerical value of 1.47 (for the specific neighborhood). Measurements repeated for different neighborhoods yield consistent results.

In figure 9 (top right) we compare the theoretical and experimental \( \sigma_{x_p} \) estimate (dotted values and bold line respectively). Therefore the experimental value of 1.47 validates the theoretical equation we used to estimate of \( \sigma_{x_p} \).

After triangulation, we obtained the 3D coordinates of the observed points in the scene in the camera reference frame. Figure 11 shows a synthetic view of the final set of 3D reconstructed points from a given position of the camera. The error on the depth reconstruction has been experimentally estimated to be about 3mm (dimensions of the observed scene are about 60x50x30cm³). Indeed, as above, we can fit a plane across the points in space and then look at the residual deviations of the points to the plane.

We conclude this section showing a 3D shape estimate of a more complex scene (see figure 11).

6.2 linear pattern method: experiments

We used again the object of figure 9 for our evaluation. We first projected two patterns with maximum and minimum brightness in order to compute the normalized brightness \( I_n \). Figure 10 (top left) shows the values of \( A = I_{\text{max}} - I_{\text{min}} \) in the horizontal image coordinate scanline \( y_c = 200 \). We projected a pattern with a linear brightness profile (such as in figure 3 but assuming \( B \) to be a linear function of \( x_p \)). We then computed \( x_p \) using equation 8. Figure 10 (bottom left) shows the values of \( x_p \) in the same horizontal scanline \( y_c = 200 \). As we can easily see from the figure, the noise attached to \( x_p \) is fairly larger than that in the sinusoidal case.

We repeated the same procedure seen in the previous section. We first estimated \( \sigma_{x_p} \) theoretically using equation 14. We plotted such values of \( \sigma_{x_p} \) in figure 10 (top right) along \( y_c = 200 \). Notice that \( \sigma_{x_p} \) is larger (> 40 pixels) where \( I_{\text{max}} - I_{\text{min}} \) is smaller. This theoretical conclusion can be experimentally validated looking at the dot values of \( x_p \) in figure 10 (bottom left): the noise attached to \( x_p \) is larger for small values of \( A \) (that is, for darker pixels). Then we computed an experimental value (17.63 pixels) for \( \sigma_{x_p} \) fitting a line across the \( x_p \)-0 values for pixel \( x_c \) between 250 and 350 (with \( y_c = 200 \)). The corresponding construction noise attached to \( x_p \) is shown in figure 10 (bottom right). In figure 10 (top right) we compare the theoretical and experimental \( \sigma_{x_p} \) estimate (dotted values and bold line respectively). Notice that, since the \( \sigma_{x_p} \) values plotted in figure 10 (bottom right) have been computed using the worst case equation 14 with \( \gamma = 2 \), the comparison confirms our calculation of the theoretical upper bound on the standard deviation.
6.3 Final remarks on experimental results

In conclusion with the linear pattern technique we are able to achieve a dense 3D shape reconstruction using only 3 patterns. However the large uncertainty in the \( x_p \) estimate does not allow us to obtain a reliable depth estimate. On the contrary, using patterns with multiple ramps we can improve the accuracy of \( x_p \). As we showed in section 5, only 8 patterns are enough to get the same performance of the method with sinusoidal brightness.

7 Discussion and conclusion

We presented two methods for reconstructing 3D depth using structured lighting. Unlike previous structured lighting methods based on projecting black/white stripes on the scene our methods are based on projecting grayscale patterns. While stripe-based methods only allow a sparse recovery of scene depth our methods compute a dense depth map with one depth value per pixel. Our gray-scale methods allow to trade off acquisition time for accuracy, while stripe-based methods trade off acquisition time for sampling density.

We investigated the performance of the two methods both theoretically and experimentally. Theoretically predicted accuracy and measured accuracy were found to be in agreement. Our experiments demonstrate that a coarse assessment of depth is available if only 3 patterns are projected (corresponding to acquisition time of 1/10 of a second), while higher accuracies are available if more patterns are used.

References