Optical Engineering

SPIEDigitalLibrary.org/oe

Dynamic (de)focused projection for three-dimensional reconstruction

Intuon Lertrusdachakul
Yohan D. Fougerolle
Olivier Laligant

Optical Engineering 50(11), 113201 (November 2011)

Dynamic (de)focused projection for three-dimensional reconstruction

Intuon Lertrusdachakul
Yohan D. Fougerolle
Olivier Laligant
Le2i Laboratory
12 Rue de la fonderie
Le Creusot, Burgundy 71200, France
E-mail: intuon.lertrusdachakul@u-bourgogne.fr

Abstract. We present a novel 3-D recovery method based on structured light. This method unifies depth from focus (DFF) and depth from defocus (DFD) techniques with the use of a dynamic (de)focused projection. With this approach, the image acquisition system is specifically constructed to keep a whole object sharp in all the captured images. Therefore, only the projected patterns experience different defocused deformations according to the object's depths. When the projected patterns are out of focus, their point-spread function (PSF) is assumed to follow a Gaussian distribution. The final depth is computed by the analysis of the relationship between the sets of PSFs obtained from different blurs and the variation of the object's depths. Our new depth estimation can be employed as a stand-alone strategy. It has no problem with occlusion and correspondence issues. Moreover, it handles textureless and partially reflective surfaces. The experimental results on real objects demonstrate the effective performance of our approach, providing reliable depth estimation and competitive time consumption. It uses fewer input images than DFF, and unlike DFD, it ensures that the PSF is locally unique. [DOI: 10.1117/1.3644541]

Subject terms: focus; depth from defocus; active illumination pattern; range sensors; blur estimation; 3-D reconstruction.

Paper 110623RRR received Jun. 3, 2011; revised manuscript received Sep. 7, 2011; accepted for publication Sep. 8, 2011; published online Oct. 25, 2011.

1 Introduction

In the extensive field of computer vision, depth recovery approaches have been broadly developed and have attracted substantial attention over recent decades. It is a challenging problem to recover the 3-D information (i.e., depth) that is lost during the projection of a 3-D scene onto a 2-D image plane. Several 3-D reconstruction algorithms have already been proposed; the effectiveness of each method, both qualitative and quantitative, has been studied. However, there is still no unique satisfactory solution that applies to all kind of scenes. Moreover, our new approach, which entails a combination of depth from focus (DFF) and depth from defocus (DFD) with the use of a light-pattern projection, has not yet been explored. Here, we develop a prototype of a range sensor from our new depth-estimation system. Various relevant applications can be found in the real world, starting from large-scale examples, such as feature extraction in video surveillance, to small-scale examples, such as in microbiological analysis.

Depth recovery approaches are generally classified into three categories. The first category consists of time-delay–based approaches, where a transreceiver system computes the delay or any deterioration in the reflected signal after the signal is sent and returns back from the object surfaces. Radar/sonar and laser coherence are examples of this approach. Such methods provide useful rough depth maps for typical distance scenes but can require a very long scanning time. The second class of approaches uses a geometric formulation known as triangulation to infer the depth. The last category is based on the imaging cues. This category is also known as the shape from X approaches, where X can be stereo, texture, shading, motion, or defocus.

Alternatively, we can also define the 3-D recovery approaches into two classes: passive and active techniques. There is a clear distinction between the passive and active techniques in terms of whether an active source/illumination pattern is considered or not. Passive techniques, such as stereo and shape from motion, use at least two images to perform multiple-view correspondence matching.[1,2] The depth is extracted from either the disparity or motion vectors after matching. The main drawbacks of these techniques are that they are computationally expensive to either perform correspondence matching or feature tracking, and there are occlusion problems in scene areas that are visible only by one camera.[3] Other passive techniques include shape from shading and shape from texture. By using only a single image, the depth ambiguities can be retrieved. However, these techniques are only complementary to other strategies. Overall, the common bottleneck shared among all the passive techniques is that the depth cannot be computed accurately in the case of weak texture or textureless scenes.[4] Meanwhile, active techniques use active illumination to solve texture problems and are generally based on the principle of structured light. The most well-known active techniques are the light-striping method, Moiré interferometry, and Fourier-transform profilometry.[5,6] Depth can be extracted from the image deformation of the projected pattern.[7,8,17] Nevertheless, Moiré interferometry and Fourier-transform profilometry return only a relative depth, not an absolute depth. Recently, another prominent technique, known as shape from focus/defocus, has received a remarkable amount of interest. DFF requires several images to be taken, with small incrementing focus settings.[9–11] Depth is estimated by

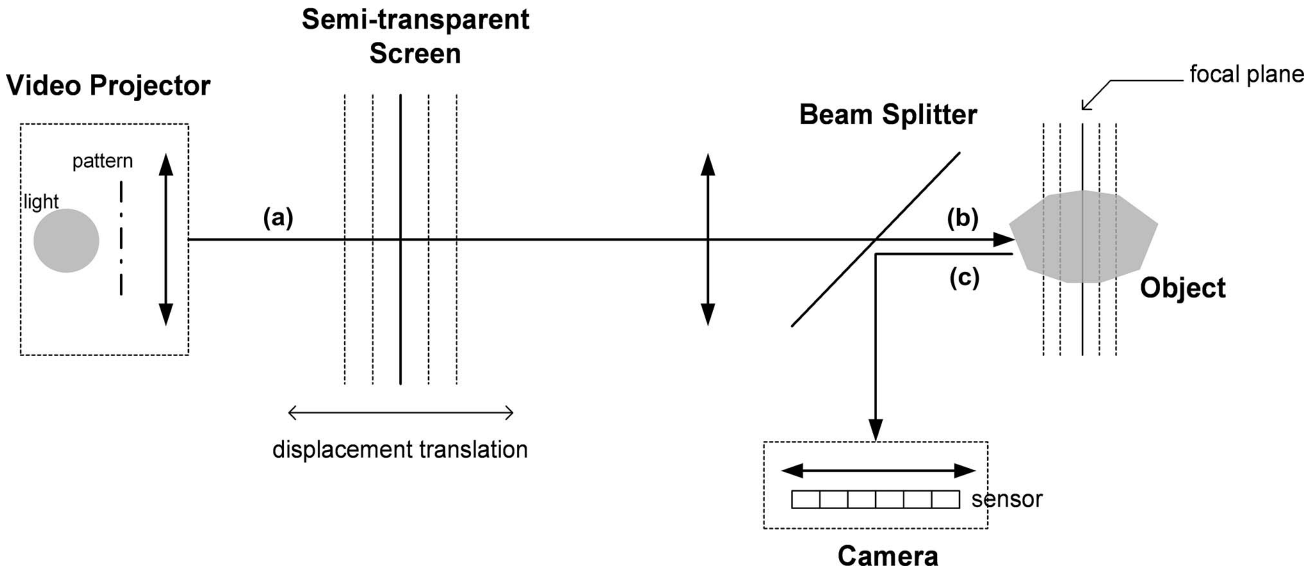

Fig. 1 Model for the proposed approach.

searching for the best focused point through the image stack. Meanwhile, the DFD can use as few as two images with different optical geometric settings to evaluate the difference in the blur level between each point in the defocused images.[12–16] Therefore, DFD has advantages over DFF during the image-acquisition process, when scene objects may change their position, dynamically. However, it is also computationally expensive to return a reliable depth map. Depth from focus/defocus is an example of the case where it can be specified as either a passive or active approach, depending on whether or not it is possible to project a structure of light onto the scene.

Our approach uses a novel active range sensor and combines both depth from focus and depth from defocus with the help of light-pattern projection. This method falls into both the imaging-cues approach and the active structured-light–based approach. The aim is to introduce a new and alternative approach to solving some of the specific problems found in classical approaches, such as the weak texture surface and occlusion problem. With this approach, projected light-pattern images are acquired within certain ranges, similar to the DFF approach, but the numbers of captured images are much smaller and the images do not need to be sharp. In traditional DFD, blur estimation is a very difficult problem because a point that represents the defocus information has contributions from several point-spread functions (PSFs) that are induced by different depths. Our method avoids this problem because we can control the deformation by placing an additional semi-transparent screen after the light source. Therefore, by considering the light pattern as a plane, one point on the object representing defocus information corresponds to only one PSF. Moreover, when the projected patterns are out of focus, we assume that their PSFs follow a Gaussian distribution. Eventually, the relationship between the set of PSFs obtained from different blurs and the variation of the object depths can be determined.

The structure of this paper is as follows: We begin with the prototype principle and the system components. On the basis of fundamental background, we derive the depth-estimation model, accordingly. We describe the implementation and obtain experimental results in Secs. 2 and 3, respectively. In the last section, we conclude this paper by summarizing the performance expectations of this new technique and discuss further research perspectives. The contributions of this paper are the demonstration of a novel 3-D retrieval approach that is based on active structured light and the investigation of their benefits over other 3-D recovery methods.

2 Proposed Shape Recovery Method

2.1 *Prototype Principle*

A new prototype of range sensors has been developed. We integrate both depth from focus and depth from defocus with the use of dynamic structured light. A video projector is used as a light source to produce strong projecting light patterns. It is much more powerful compared to normal lamps and much simpler compared to light-emitting diodes (LEDs). Moreover, it moderates the additional pattern modifications that are needed for different types of surface textures.

Figure 1 illustrates the overall design of the system. The main purpose of using a semitransparent screen is to control the defocus level that corresponds to each screen position. It also helps to solve the magnification problem caused by the fact that the projected light patterns from the video projector are originally small compared to the patterns projected on the object without passing through the screen. Reducing the intensity of the powerful light is an additional advantage. Because the normal lamp provides insufficient brightness, the video projector sometimes produces too strong a light, which can be adjusted by a projector setting or by putting in some blocking elements. The beam splitter is mainly used to observe the object on the sensor, and it allows for the projection of a light pattern onto the object. The entire setup can be separated into three systems.

2.1.1 *Light pattern projection system*

In Fig. 2, the video projector projects the elementary light pattern of size $N \times M$ pixels. Consequently, we observe a light pattern of size $H \times L$ mm on the screen. The size of the input pattern from the video projector ($N \times M$) and the size

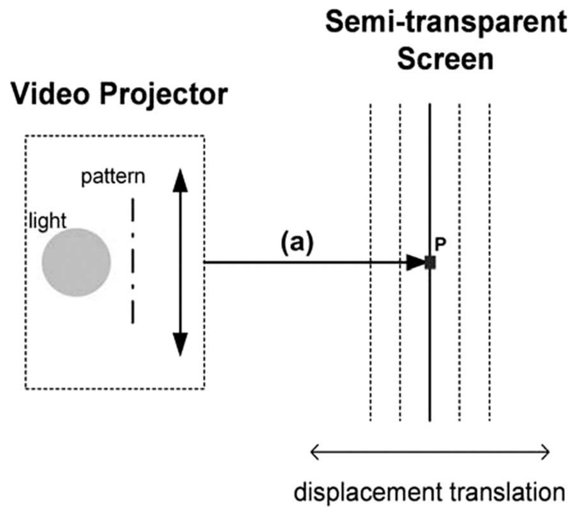

Fig. 2 Light pattern projection system.

of the pattern that appears on the screen ($H \times L$) indicate the resolution of the light pattern. We aim to project the sharp elementary pattern of size P on the screen regardless of the positions. However, at some screen displacements, we may need to adjust the video projector to maintain the pattern sharpness.

2.1.2 *Main optical system*

The system is considered from the light pattern on the screen (size P), projecting through a specific optical system onto the object. The optical components of the system consists of a semitransparent screen, a compound lens, and a beam splitter. An elementary size of the light pattern that appears on the object is denoted as P'. The optical path of the system is illustrated in Fig. 3.

The magnification of this system can be written as

$$\gamma = \frac{P'}{P} = \frac{d'}{d}. \tag{1}$$

Thus,

$$P' = \frac{Pd'}{d}, \tag{2}$$

where γ is the main optical system magnification, d is the distance from the light pattern on the screen to the additional compound lens, and d' is the distance from the additional compound lens to the object.

In general, for the ideal case where the object is placed in or very close to the surface of the best focus, an output image formed on the sensor is sharp or identical to the input. The relationship between the input and output images is the following:

$$I = I_0 * \delta \xrightarrow{\mathcal{F}} \hat{I} = \hat{I}_0. \tag{3}$$

However, our concern is deformation, where a semi-transparent screen's displacements are varied at different depths of field. The blurring function has an influence on the system and therefore must be taken into account. The defocused output image can be rewritten as the convolution between the input image and a blurring function h, as follows:

$$I = I_0 * h_{d,d'}, \tag{4}$$

where $h_{d,d'}$ are the blurring functions (PSF) corresponding to distance d and d'.

For conceptual simplicity, the PSF of the camera is usually assumed to be a two-dimensional Gaussian when paraxial geometric optics are used, and diffraction effects are negligible,

$$\text{PSF} = h(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{5}$$

When we know the input and output images, we could determine the PSF accordingly. The extracted PSF is then used for depth computation.

2.1.3 *Acquisition system*

The acquisition system captures objects with projected light patterns (size P') to the sensor via a beam splitter. The observed pattern size on the sensor is denoted as P''. The optical path of the system is simplified, as shown in Fig. 4. The magnification of this system is

$$\gamma' = \frac{P''}{P'} = \frac{l'}{l}. \tag{6}$$

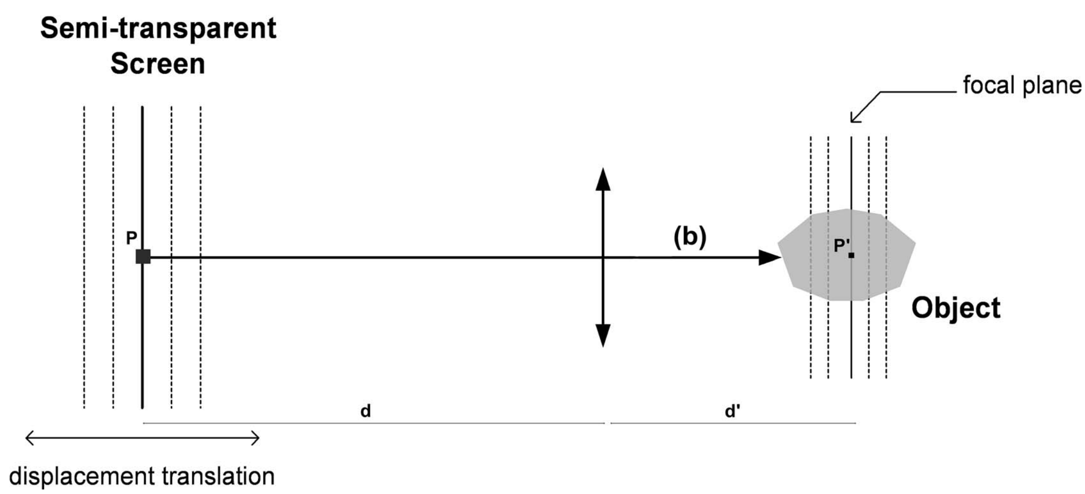

Fig. 3 Main optical system.

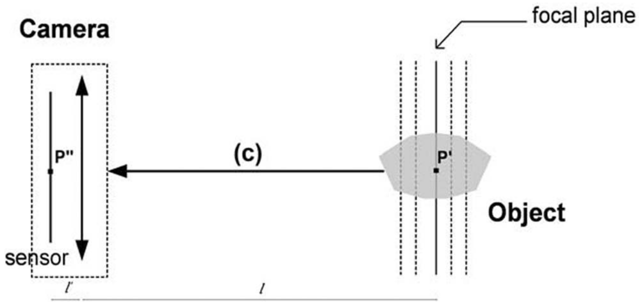

Fig. 4 Acquisition system.

Hence,

$$P'' = \frac{P'l'}{l} = \frac{Pd'l'}{dl}, \tag{7}$$

where γ' is the acquisition system magnification, l is the distance from the light pattern on the object to the camera lens, and l' is the distance from the camera lens to the sensor.

The relationship between the sizes (in pixels) of the pattern P'' on the sensor and of the elementary pattern from the video projector is used for the light-pattern specification (width and density). Only 1 pixel of the light pattern (from the light source) may require several representing pixels on any of the captured images.

This concludes the prototype principles. The corresponding practical implementation is detailed in Sec. 3. We now explain the theoretical relationship between the blur level (the spread parameter) and the depth.

2.2 *Depth Estimation*

The scene can be captured in either a sharp or blurred form, depending on the convergence of all the light rays from a single point on the object. The object appears sharp if each point on the object plane is projected onto the image plane. However, if the sensor plane and image plane are misaligned, then the image is distributed over a circular patch called a circle of confusion (CoC) on the sensing element, resulting in a blurred image.[4,18] Therefore, the blur level can be determined from the diameter of the CoC, which also increases proportionally to the distance from the object in focus. This phenomenon allows us to estimate the geometry of the scene by measuring the amount of blur in the image.

Figure 5 represents the camera geometry of the camera lens (compound lens) with variable camera parameters $(s,\ f,\ D)$. Here, s specifies the sensor distance ($\|\overline{H'C'}\|$), f specifies the focal length ($\|\overline{H'F'}\|$), v is the object distance ($\|\overline{H'A'}\|$), and D is the aperture diameter. According to paraxial geometric optics, to define the radius of the circle of confusion, the knowledge of similar triangles is applied,

$$\tan\alpha = \frac{D/2}{\|\overline{F'A'}\|} = \frac{\phi}{\|\overline{A'C'}\|} = \frac{\phi}{x}. \tag{8}$$

The normalized diameter of the CoC can be rearranged into the following:

$$\phi = x\frac{D/2}{\|\overline{F'A'}\|}, \tag{9}$$

where x is the object distance and ϕ is the diameter of the CoC.

The higher the degree of blur is, the larger the CoC. The CoC is also proportional to the value of spread parameter σ and can be written as

$$\sigma = K_1\phi, \tag{10}$$

where K_1 is a positive constant.

Given a spread parameter σ, aperture size D, and the controllable distance $\|\overline{F'A'}\|$, the only unknown of the system [see Eq. (9)] is x. Eventually, we can derive the object distance x, which directly relates to the real depth of the object as

$$x = \pm\frac{2\sigma\cdot\|\overline{F'A'}\|}{K_1 D}, \tag{11}$$

$$x = \pm K_2\sigma, \tag{12}$$

where K_2 is a positive constant for a given position of the screen.

The aim of the whole system is to achieve the object depth x. The extracted PSFs are used for spread parameter σ computation. However, K_2 is an unknown variable. Another concern is that the pattern on the screen at different displacements may not be constant, even after tuning a video projector for the best sharpness. Moreover, the distance l varies according to the object depths, which is also an unknown variable. To solve these problems, the calibration process is required. The depth is eventually computed by using a derived depth formula with the parameters obtained from calibration.

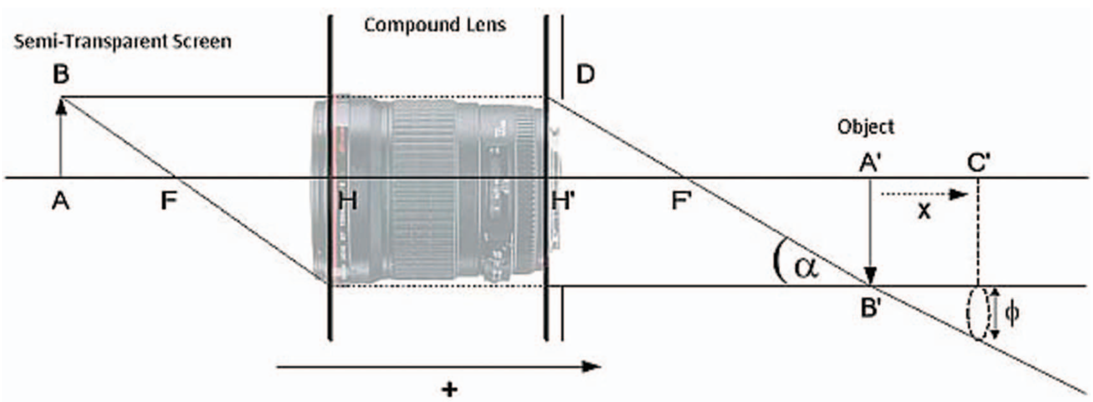

Fig. 5 Camera geometry of compound lens.

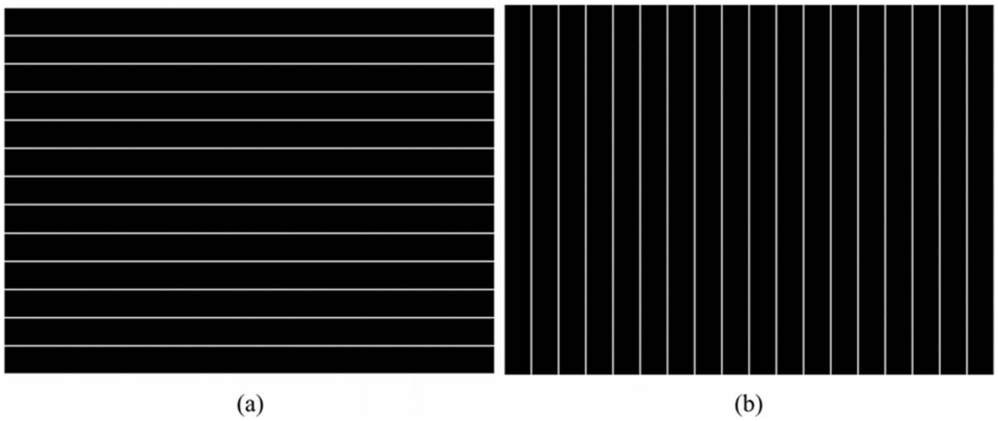

Fig. 6 Example of illumination pattern.

2.3 *Design of Dynamic Light Pattern*

All passive techniques share the same inherent weakness when the nature of an object's texture is poor. A weakly textured surface does not provide sufficient details for depth estimation because both the focus and defocus give the same representation. An effective solution to solve this problem is to employ an active illumination pattern. The structured light source projects a pattern on the scene through a specific optical setting while the camera captures it. Because the original projected light source is known, the defocus blur introduced by the depth in the scene can be measured against the original pattern. The choice of an appropriate light pattern is important to optimize our final reconstruction. Highly textured light patterns are forced onto the object, improving the overall depth recovery system to be reliable and more precise. Moreover, to avoid rotational variance, it should be designed in a symmetrical or semi-symmetrical arrangement. The density of the projected pattern or its spatial frequency should correspond to the frequency of the height variation to be captured. For example, an object with a high level of detail requires a finer texture, whereas an object with smooth structural changes can use a sparse pattern instead. To be specific, for the object with small depth variations, we can reduce the processing time by projecting the sparse pattern instead. There are fewer intensity profiles to be analyzed, and it returns sufficient results that are similar to the results from a denser pattern projection. For our experiment, we employ a set of parallel stripes with regular spacing. Spacing and shifting step sizes are determined from the scale of the texture to be analyzed. The width of the lines and the density of the pattern are selected according to Eq. (7), to cover as much of the reconstructed area as possible. Figure 6 illustrates a sample of a stripe light pattern with a width of 1 pixel and 20 pixels for spacing in between.

3 Implementation

In our experiment, a Canon SX80 Mark II video projector with a resolution of 1400×1050 is used as the light source. The horizontal stripe illumination patterns with a width of 1 pixel and 20 pixels for spacing are applied. The beam of the projecting light pattern then reaches the semi-transparent screen and an additional lens (Canon telephoto lens 135 mm). The light rays passing through the lens are split into two directions by the beam splitter. One beam is projected onto the object, and another is transmitted from the object to the sensor. The scene object is captured using a Canon EOS-1Ds camera with an attached 50-mm lens. The data flowchart illustrated in Fig. 7 describes the main operations.

3.1 *Image Acquisition*

The camera settings (e.g., F, ISO, shutter speed) are carefully tuned such that the system keeps the whole object sharp in all the images and only the defocused patterns experience varied deformation according to the object's depth. All the optical components in this setup are fixed. Only the semi-transparent screen is moved, which results in several scene images with different blur levels. With this specific setting, we can analyze the defocus of the light pattern projecting on the object. We calibrate the system using a planar surface at five different depths ($D1-D5$). Each plane is captured through the projection of six screen displacements ($Pl1-Pl6$), resulting in a total of 30 calibrated images. Similarly, we apply the same procedure to the test object and obtain six images. Before starting the main algorithm, the acquired images are preprocessed. First, we convert image from the RGB to the gray-scale level. Next, we crop the images, selecting only the effective areas, which are object-projecting regions within the beam splitter.

3.2 *Image Profile Analysis*

From a particular viewpoint, a 3-D object can be thought of as the variation in depth over the object. However, these depth variations are missing during the process of imaging, and what remains are the intensity variations that are induced by the shape and the lighting. The profile analysis is performed to extract the intensity values along multiline paths of the images. The algorithm computes equally spaced points along the specified path and uses interpolation to determine the image intensities for each point. This operation is performed along an orthogonal direction to the axis of the pattern projection. To be precise, when the projected pattern is the horizontal stripes, the vertical profile analysis will be applied column by column, whereas for the vertical stripe pattern, the horizontal profile will be analyzed row by row. The output is stored in the profile stacks regarding their intensities and pixel coordinates. In either column- or rowwise approaches, profile analysis will provide numbers of peaks corresponding to the numbers of stripe patterns. Each peak occurs at the center of its pattern and decays along both sides with a different speed. This scenario is based on the same

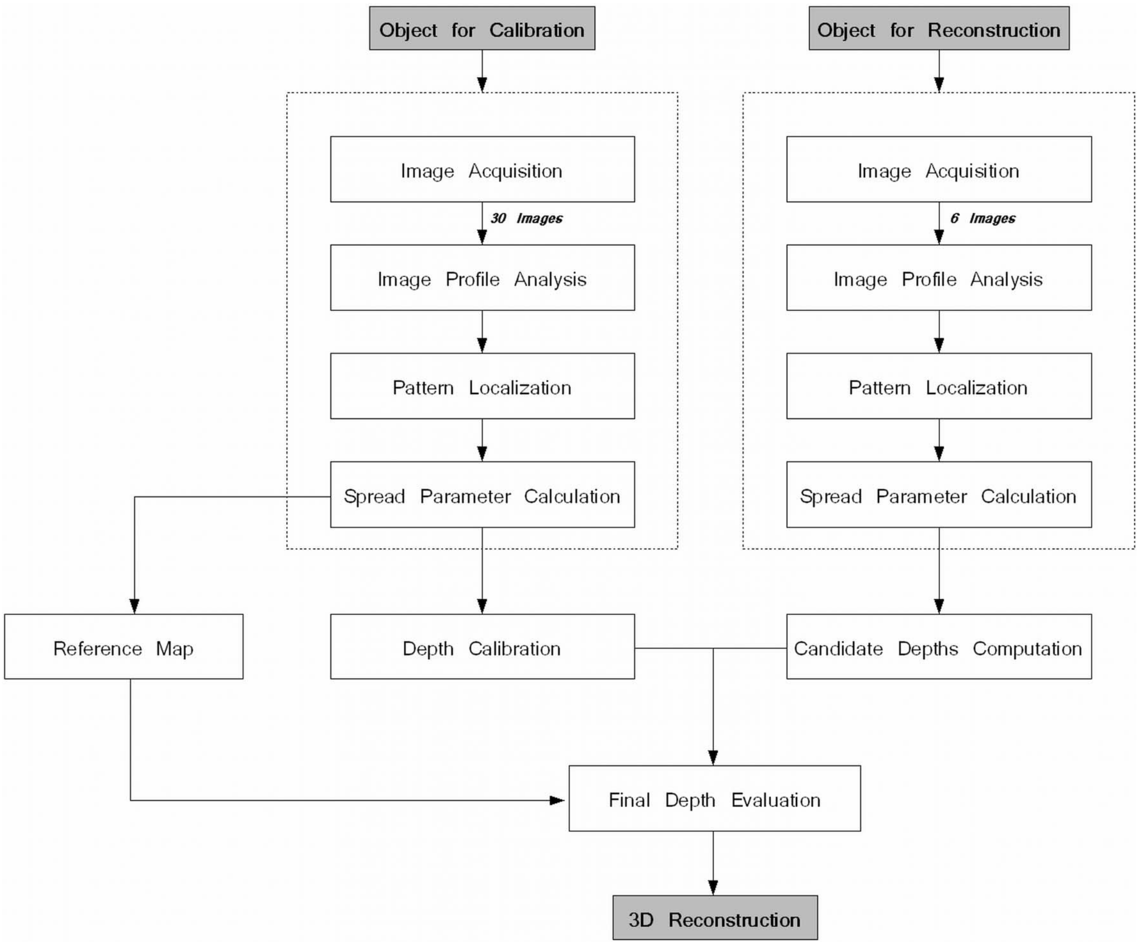

Fig. 7 Flowchart.

concept that explains why the focused or sharp pattern gives a smaller width and higher profile intensity than the blurred pattern.

3.3 *Pattern Localization*

Each single profile obtained previously contains either important data or noise. The difficult task is to differentiate the noise from the important signal before localizing the intensity pattern. The aim is to smooth the noisy part and maintain the important intensity details, simultaneously. We selected the Savitzky–Golay (Sgolay) filter,[20] which is a smoothing polynomial or a least-squares smoothing filter. It is very effective and works well for our type of signal, unlike typical Finite Impulse Response (FIR) filters, which tend to filter out a significant portion of the signal's high-frequency content along with the noise. To define a pattern cutoff coordinate, we extracted the local maximum and minimum by an absolute peak detection algorithm. We prefer a non-derivative method because finding the zero-crossing of first derivative can yield false results in the presence of noise. Moreover, we compute another controlled parameter (δ) regarding the highest slope between the local maxima and local minima on both spans. This approach is applied to guarantee that we optimize the cutoff portion of the significant information as much as possible (see Fig. 8). In addition, when the object sharpness is varied over a wide range of depths, the blur levels in the far focus region are also considered noise. To eliminate these high defocused profiles, a contrast criterion is also applied. By setting the significant contrast, we can discard the defocused profile, in which the ratio between the maximum and minimum is lower than the contrast. We obtain proper cutoffs to isolate the light pattern in each intensity profile.

3.4 *Spread Parameter Calculation*

The distribution of light energy within the blur circle is referred to as the PSF. Because of the lens aberrations and diffraction effects, the PSF will be a circular blob, with its brightness falling off gradually rather than sharply. Thus, most algorithms use a two-dimensional Gaussian function instead of the Pillbox function. From the pattern patches isolated by pattern localization, we determine their PSF individually. The spread parameters σ can be extracted from the fitting between PSF and Gaussian model as exemplified in Fig. 9. The spread parameter is used to indicate the blur level in defocused images. Consequently, the depth can be deduced and assigned back to the pixel coordinates or the local maxima defined earlier. We iterate this algorithm for all the light patterns that cover the whole object.

3.5 *Depth Calibration*

Depth calibration is performed only once by using five planar surfaces ($D1-D5$). The aim is to build the model defining the depth according to the value of the spread parameter σ at specific screen positions (Pl). In an ideal case using

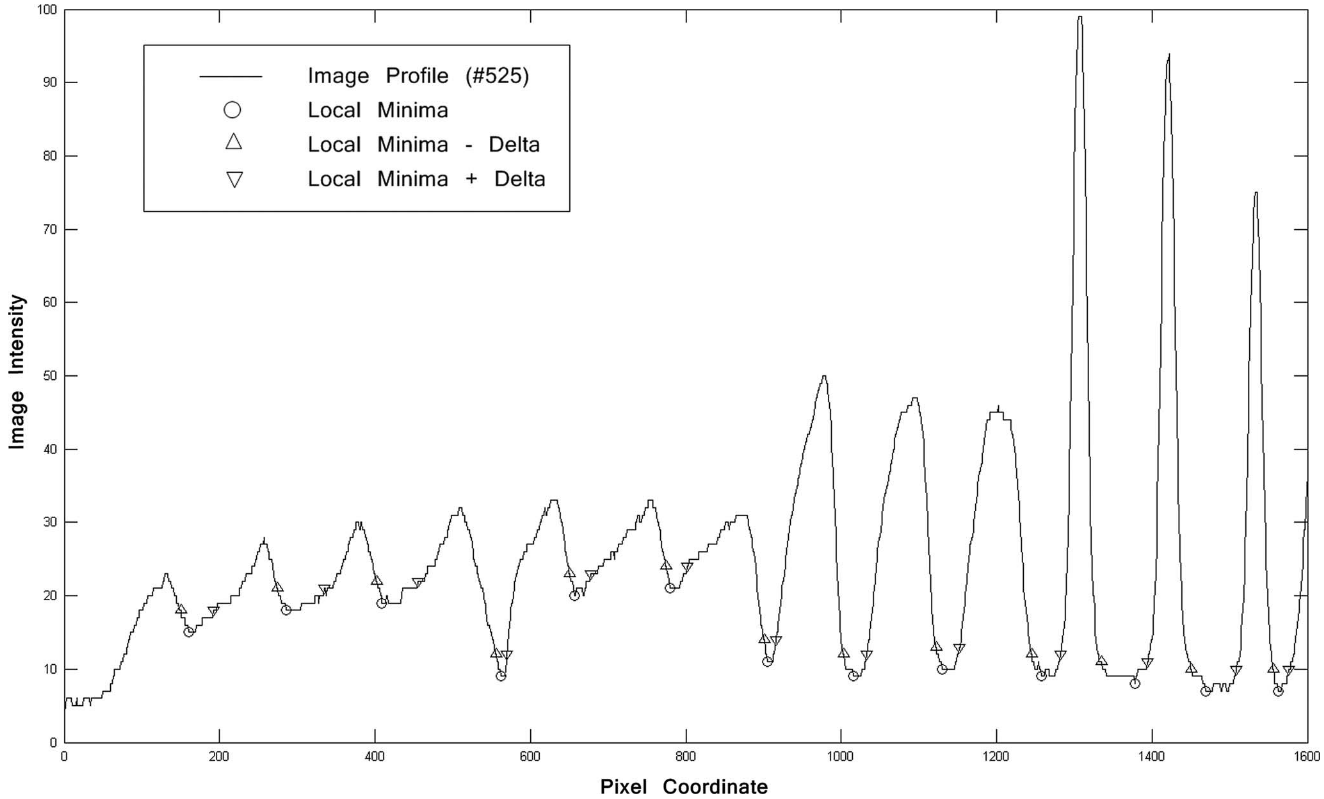

Fig. 8 Image profile in pattern localization.

Eq. (12), we can plot the relationship of the depth x against the spread parameter σ, as illustrated in Fig. 10. This plot consists of two tangent lines having zero as a minimum at the center. Nonetheless, in practice, the elementary pattern has a minimum size on the captured image (see Eq. (7)), causing a nonproportional law. To be specific, the spread parameter σ is not proportional to the object depth x for a small blurred interval around the center. Moreover, the optical system has certain acceptable sharp ranges related to the depth of field (DOF), resulting in a smooth valley instead of a sharp cut at the nadir. Therefore, we assume that the closed-form model follows the parabolic function as follows:

$$\sigma = ax^2 + bx + c. \tag{13}$$

Generally, if we have three unknowns and three equations, we can solve it as a linear problem. However, our system is dealing with more constraints. For each screen position, we calibrate five depths, which lead up to five equations. Therefore, with three unknowns (a, b, c) and five equations, we obtain an overdetermined system, which is solved in the

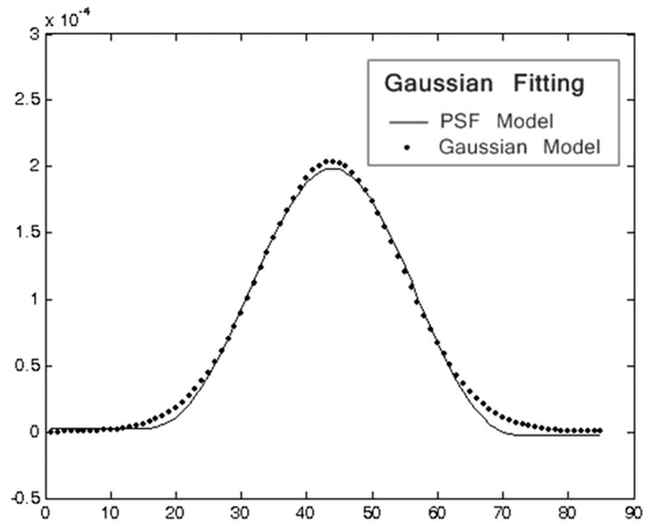

Fig. 9 Example of Gaussian fitting.

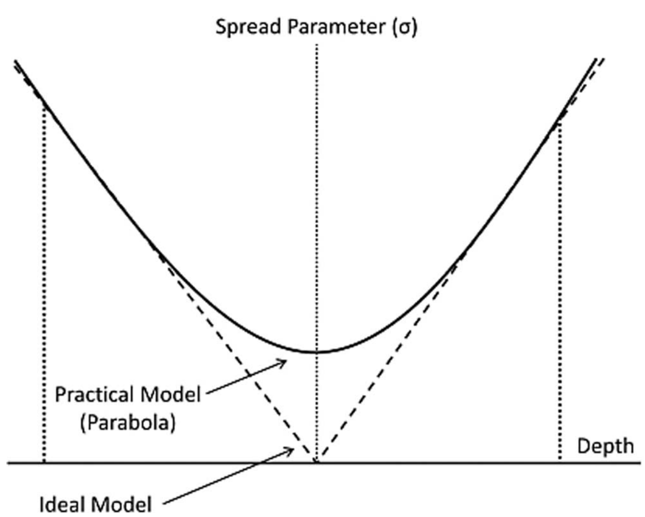

Fig. 10 Ideal depth model and the practical conic model.

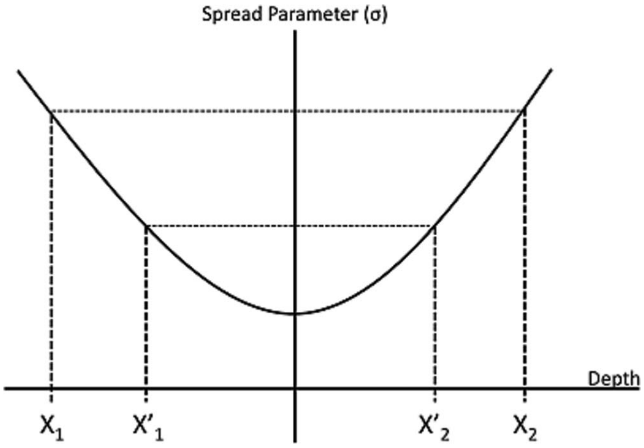

Fig. 11 Candidate depths.

least-squares sense using the well-known Levenberg–Marquardt (LM) optimization algorithm.[19]

3.6 *Candidate Depths Computation*

For the real object used in the experiment, we follow the same manipulation for image acquisition, image profile analysis, pattern localization, and spread parameter calculation. Once we obtain σ, we can determine the candidate depths. By using the already-extracted parabola parameters (a, b, c) from the depth calibration, we solve Eq. (13) for the final depth x. As seen in Fig. 11, it returns two possible solutions for x_1 and x_2, as follows:

$$x_1 = \frac{-b - \sqrt{b^2 - 4a(c - \sigma)}}{2a}, \tag{14}$$

$$x_2 = \frac{-b + \sqrt{b^2 - 4a(c - \sigma)}}{2a}. \tag{15}$$

Therefore, we need an additional clue called the reference map to make the decision.

3.7 *Final Depth Evaluation*

One spread parameter σ corresponds to two candidate depths. To select the correct depth value, we employ the reference map obtained from the spread parameter calculation during the calibration process; it is the mapping of the spread parameter at each screen position for all of the calibrated depths. Consider the reference map that is shown in Fig. 12. We plot the spread parameters of our object (to be reconstructed) according to their six observed screen positions. Then, we compute the minimum global distance, comparing them to five calibrated depths. This step is to roughly define to which calibrated depth our object belongs. Eventually, the nearest candidate depth value closest to the calibrated depth will be selected as the final depth.

4 Experimental Results

We conducted experiments using an acrylonitrile butadiene styrene (ABS) plastic built by a 3-D printer as the test objects. The model of the 3-D printer is a Dimension Elite (Stratasys Inc., Eden Prairie, MN). Without coating, the material has some reflectivity. Two types of surfaces have been tested: a planar and a pyramidal structure (see Fig. 13). A sequence of light-pattern images is shown in Fig. 14. They are acquired at different projecting distances (physical displacement of 1 cm

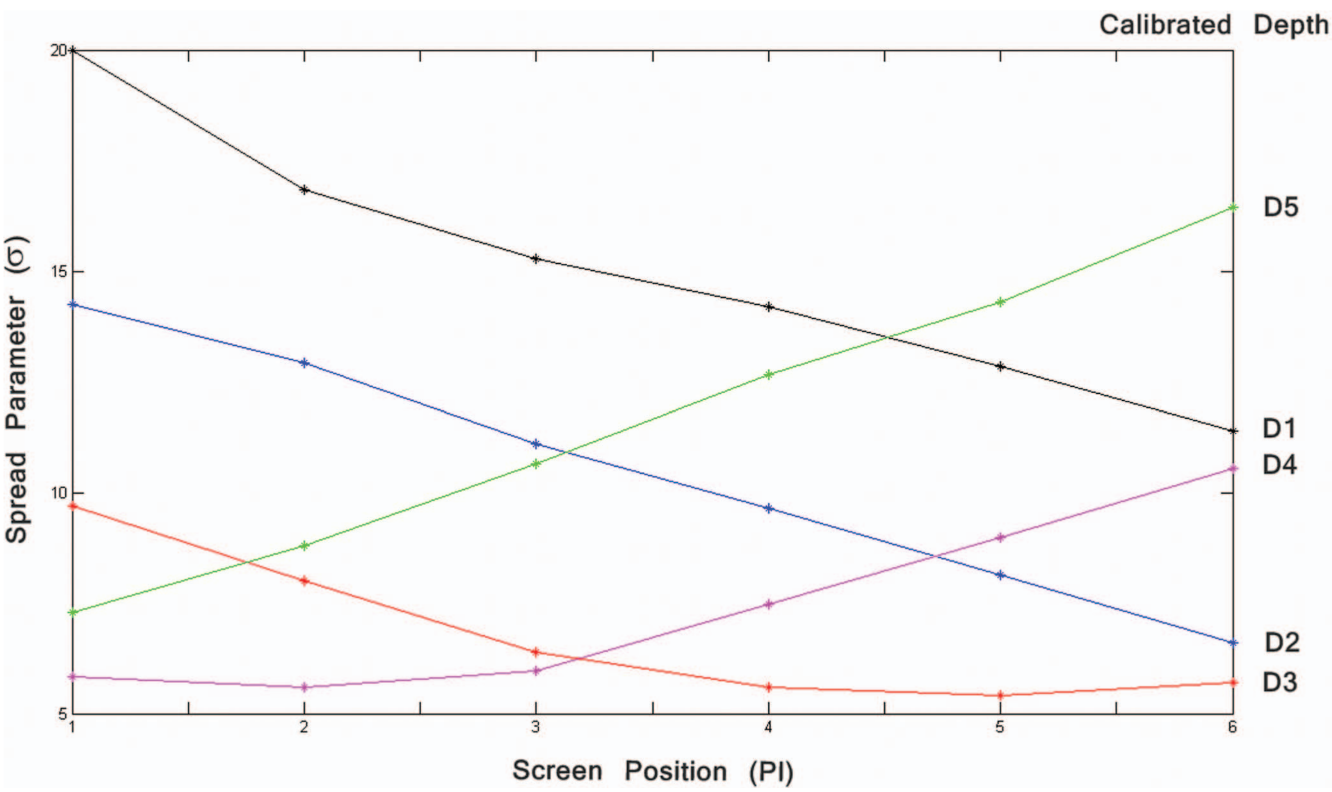

Fig. 12 Reference map.

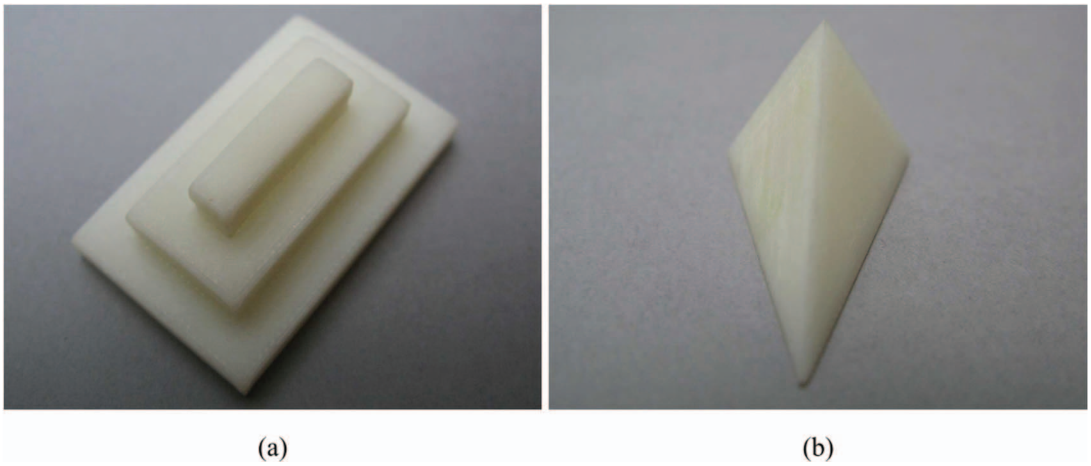

Fig. 13 Scene objects: (a) staircase and (b) pyramid.

between each screen position). Because of some constraints of our optical setup (e.g., telephoto lens distortion and beam-splitter size), the effective reconstruction areas are limited only at the center of the beam splitter. The input images are then put into the stack for profile analysis and pattern localization. For each isolated light pattern, the spread parameter is extracted by fitting the PSF to the Gaussian model. The candidate depths are then calculated by Eqs. (14) and (15), with the already known parabolic parameters from the calibration process. In the final step, the final depth is determined by taking the reference map into consideration.

The rough 3-D model presents some preliminary results that are obtained from our implementation. The depth map, illustrated in Fig. 15(a), demonstrates the effective performance of the method in the case of the planar structure. A staircase object has a minimum depth at 2 cm, increasing on both sides by 1 cm until the maximum depth of 4 cm is reached. The result shows that both the real object depth and our estimated depth lie within close proximity. Figure 15(b) illustrates the depth map of the pyramid object, in which we can also retrieve the intermediate depths that do not exist during the calibration process. We performed the 3-D modeling as shown in Fig. 16. The result is compared to the actual depth provided for the 3-D printer. The quantitative evaluations are presented in Table 1. Both test objects have an average error of <0.3 mm.

Denser point clouds and higher quality 3-D reconstruction can be obtained once the variation and the number of projected light patterns increase. The total process can be time consuming. With non-optimized MATLAB code, the program takes <1 min of computational time (on the machine equipped with a core 2 duo 2.2 GHz), excluding profile analysis and calibration. However, the benefit of our approach is that most of the processes that require a long time

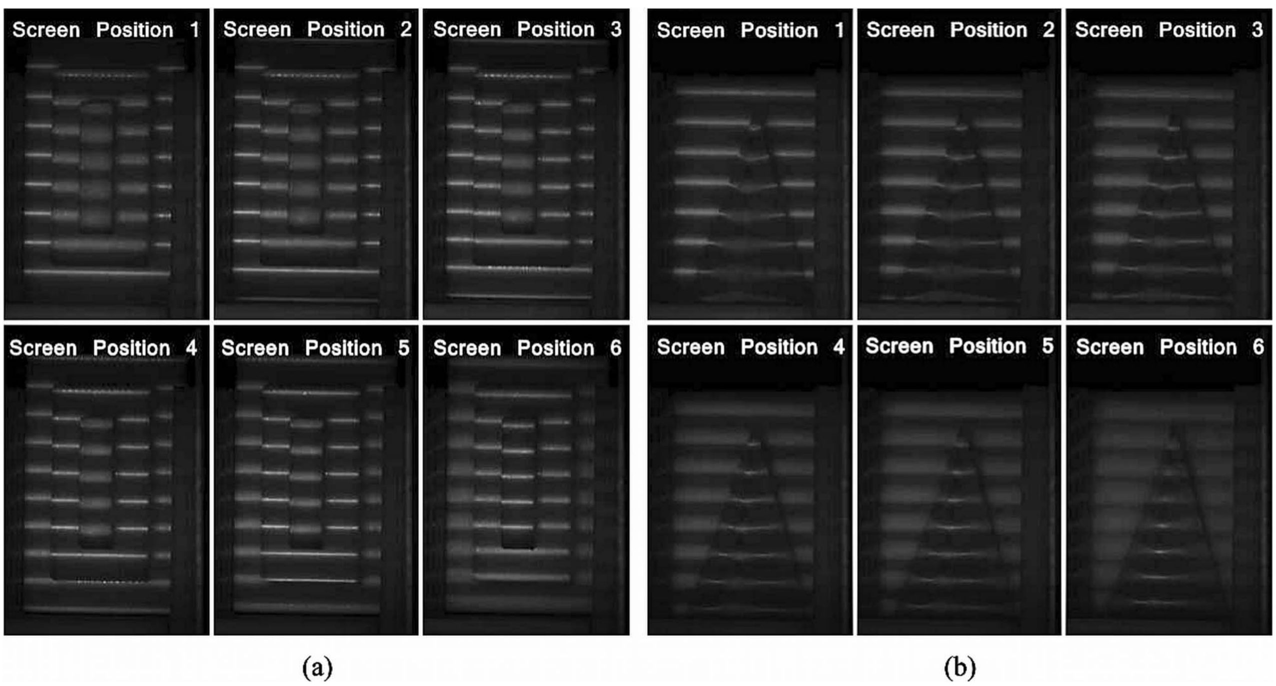

Fig. 14 Captured images: (a) staircase and (b) pyramid.

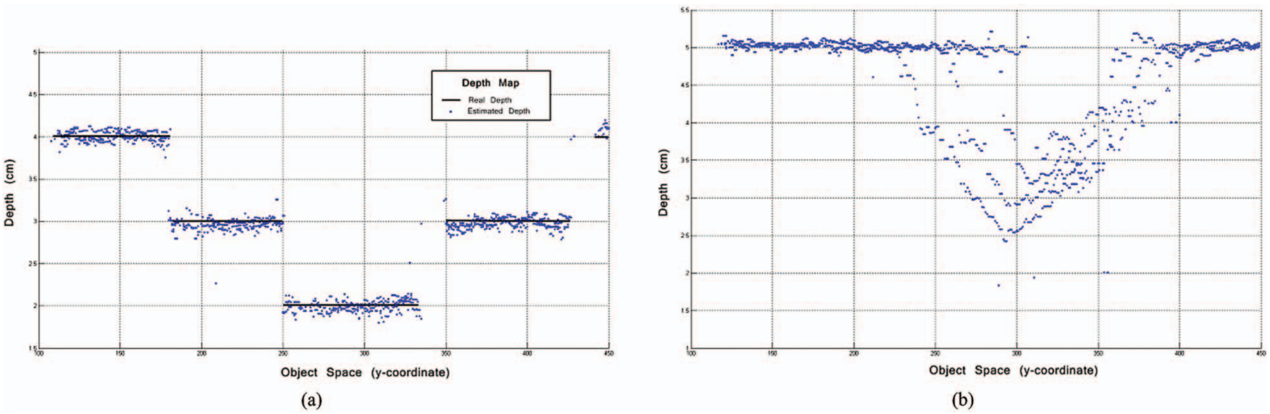

Fig. 15 Depth map: (a) staircase and (b) pyramid.

computation are offline processes. The calibration is required only once for a certain object material. Therefore, if we already have calibration data for a set of object materials, we will be able to create a 3-D model quite quickly, as reported in Table 1.

Table 1 Experimental results.

	Staircase	Pyramid
No. point cloud	8784	6375
Average error (mm)	0.17780	0.28429
Standard deviation	0.28609	0.34347
Computational time[a] (s)	31.901	19.150
Computational Time[b] (s)	208.099	203.790

[a]Excluding offline profile analysis and calibration.
[b]Offline profile analysis and calibration.

5 Conclusions

We have introduced a new 3-D reconstruction method that merge depth from focus and depth from defocus. It can be employed as a stand-alone strategy that returns reliable dense depth maps. The method overcomes the problem of weak textures by projecting an illumination pattern. Moreover, it suffers neither from the correspondence problem nor the occlusion problem found in traditional approaches. Examples of applications can be found in biological specimen analysis and in defect metallic component detection.

Several components in the setup limit the size of the object itself and the maximum change of the object depths. However, this is due to the scale of the system. This issue can be fixed by adjusting the smaller or larger optical components for the smaller or larger objects, respectively, while the algorithm remains unchanged. Therefore, these restrictions are not concerned with the approach of the methodology. Another disadvantage of the system is that no dense data can be obtained from a single illumination pattern because of the limitation of the profile analysis method. For further improvement, we will try to determine the depth for every single screen estimate and run the experiment systematically by robot, to significantly minimize human error and to increase overall precision of the system. Another future project is to use several mini video projectors and beam splitters to

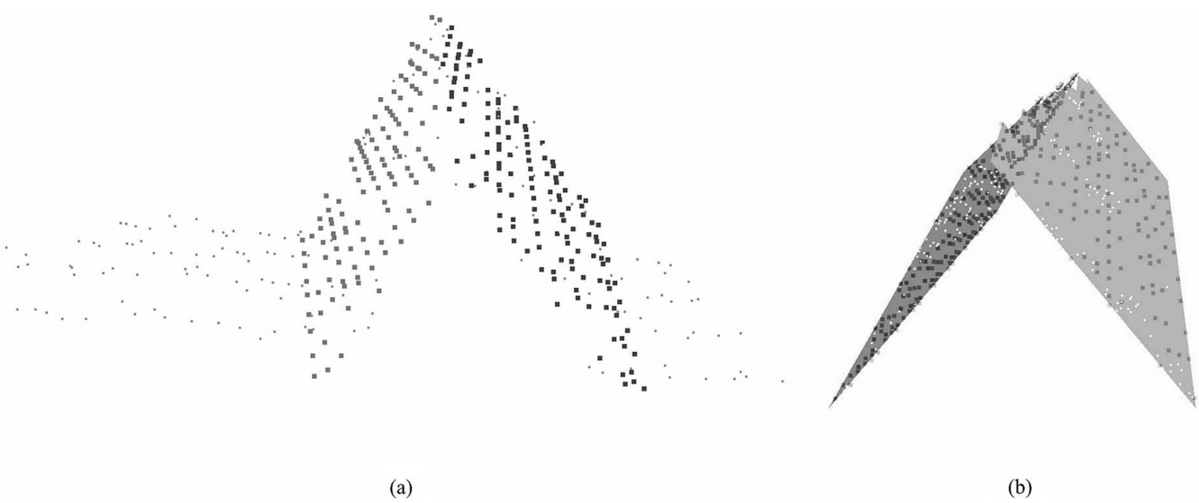

Fig. 16 Rough 3-D reconstruction of pyramid structure: (a) pointcloud and (b) surface fitting.

develop a 3-D progressive feedback system by controlling the light patterns. Only useful defocused input images would be selected iteratively based on a rough 3-D model. The first perspective of this work is to address the problem of the depth and reflectance discontinuities of the object. The relevant concept is to perform deep analysis of the Gaussian model. To be precise, the Gaussian model at these discontinuities will not be close to the reference model (Gaussian distribution). Therefore, these asymmetrical models will be discarded. To determine the depth of this problematic region, we plan to employ another type of light pattern within the 3-D progressive system. For example, if this issue occurs from using a horizontal stripe pattern, then it would not be a problem when we are using a vertical stripe pattern. Another perspective is to address the scenes with multiple objects with different types of textures and texture variations. The possible solution is to collect sufficient numbers of calibration databases and/or reflectance indexes for all the object materials that exist in the scene. The key idea is that certain materials provide different reflectances, different widths of blur (CoC), and different spread parameters, accordingly. For instance, some types of plastic have larger blur levels (larger spread parameters) than can be observed from metallic material at the same depth distance. Therefore, it should be possible to set the order of reflectance and/or blur level with respect to the material types. Then, by matching them with the test objects, we could identify material types and use the current algorithm. Eventually, our last perspective will focus on the estimation of the surface normal by analyzing the deformation of an elementary projecting pattern.

References

1. R. Szeliski, "3d reconstruction," *Comput. Vis.* 505–541 (2011).
2. Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: how different really are they?" *Int. J. Comput. Vis.* **39**, 141–162 (2000).
3. A. Sharma, "Projected texture for 3D object reconstruction," PhD Thesis, Int. Institute of Information Technology, Hyderabad, India (2008).
4. O. Ghita, P. F. Whelan, and J. Mallon, "Computational approach for depth from defocus," *J. Electron. Imaging* **14**, 023021 (2005).
5. M. Watanabe, S. K. Nayar, and M. N. Noguchi, "Real-time computation of depth from defocus," *Proc. SPIE* **2599**, 14–25 (1996).
6. M. Kusanagi, K. Terabayashi, K. Umeda, G. Godin, and M. Rioux, "Construction of a 3D model of real-world object using range intensity images," in *Proc. of IEEE 2010 Canadian Conf. on Computer and Robot Vision*, pp. 317–323 (2010).
7. M. Noguchi and S. K. Nayar, "Microscopic shape from focus using a projected illumination pattern," *Math. comput. modell.* **24**, 31–48 (1996).
8. P. Graebling, C. Boucher, C. Daul, and E. Hirsch, "3D sculptured surface analysis using a structured-light approach," *Proc. SPIE* **2598**, 128–139 (1995).
9. J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 97–108 (1993).
10. S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 824–831 (1994).
11. I. Lertrusdachakul, Y. D. Fougerolle, and O. Laligant, "A novel 3D reconstruction approach by dynamic (de)focused light," *Proc. SPIE* **7538**, 75380L (2010).
12. A. N. Rajagopalan and S. Chaudhuri, "A variational approach to recovering depth from defocused images," *IEEE Trans. Pattern Anal. and Mach. Intell.* **19**, 1158–1164 (1997).
13. M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach," *Int. J. Comput. Vis.* **13**, 271–294 (1994).
14. Y. Xiong and S. A. Shafer, "Depth from focusing and defocusing," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 68–68 (1993).
15. P. Favaro, A. Duci, and G. Rostock, "A theory of defocus via Fourier analysis," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* (2008) pp. 1–8
16. S. O. Shim, A. S. Malik, and T. S. Choi, "Accurate shape from focus based on focus adjustment in optical microscopy," *Microsc. Res. Tech.* **72**, 362–370 (2009).
17. P. J. Besl, "Active, optical range imaging sensors," *Mach. Vis. Appl.* **1**, 127–152 (1988).
18. A. Pentland, T. Darrell, M. Turk, and W. Huang, "A simple, real-time range camera," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 256–261 (1989).
19. M. Hanke, "The regularizing levenberg-marquardt scheme is of optimal order," *J. Integral Equations Appl.* **22**, 259–283 (2010).
20. A. Savitzky, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. chem.* **36**, 1627–1639 (1964).

Intuon Lertrusdachakul received her Bachelor in telecommunication engineering from Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2006, and a joined Erasmus Mundus Master in computer vision and robotics (ViBot) from Heriot-Watt University, United Kingdom, Universitat de Girona, Spain, and Université de Bourgogne, France, in 2008. She is currently a PhD student at Université de Bourgogne. Her research interests include 3-D reconstruction, image processing, and computer vision.

Yohan D. Fougerolle received his MS in electrical engineering in 2002 from the University of Burgundy, Dijon, France, where he earned his PhD in 2005. Since 2007, he has been an assistant professor in the Department of Electrical Engineering at the University of Burgundy, Le Creusot, France. His research interests include 3-D digitization, solid modeling, surface reconstruction, and image processing.

Olivier Laligant received his PhD from the Université de Bourgogne, France, in 1995. He is a full professor in the Computing, Electronic, Imaging Department (Le2i) at the Université de Bourgogne, France. His research interests are focused on multiscale edge detection, nonlinear regularization, and noise estimation.