Pattern Recognition 43 (2010) 3560–3571

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Nonstructured light-based sensing for 3D reconstruction

Zhan Song [a], Ronald Chung [b,*]

[a] CAS/CUHK Shenzhen Institutes of Advanced Technology, Shenzhen, China
[b] Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 2 October 2009
Received in revised form
23 February 2010
Accepted 4 May 2010

Keywords:
Structured light-based sensing
Surface normal
Orientation map
Depth map

ABSTRACT

Structured light-based sensing (SLS) requires the illumination to be coded either spatially or temporally in the illuminated pattern. However, while the former demands the use of uniquely coded spatial windows whose size grows with the reconstruction resolution and thereby demanding increasing smoothness on the imaged scene, the latter demands the use of multiple image captures. This article presents how the illumination of a very simple pattern plus a single image capture can also achieve 3D reconstruction. The illumination and imaging setting has the configuration of a typical SLS system, comprising a projector and a camera. The difference is, the illumination is not much more than a checkerboard-like pattern – a non-structured pattern in the language of SLS – that does not provide direct correspondence between the camera's image plane and the projector's display panel. The system works from the image progressively, first constructing the orientation map of the target object from the observed grid-lines, then inferring the depth map by the use of a few tricks related to interpolation. The system trades off little accuracy of the traditional SLSs with simplicity of its operation. Compared to temporally coded SLSs, the system has the essence that it requires only one image capture to operate; compared with spatially coded SLSs, it requires no use of spatial windows, and in turn a less degree of smoothness on the object surface; compared with methods like shape from shading and photometric stereo, owing to the use of artificial illumination it is less affected by the surface reflectance property of the target surface and the ambient lighting condition.

1. Introduction

Reconstructing 3D shape from visual data is an important goal of computer vision, whose solution has tremendous mundane and industrial applications. This article addresses how a setup as simple as a projector and a camera, with the illumination of a very simple pattern, can achieve 3D reconstruction.

There has been much effort of emulating the various vision cues of humans in machines for 3D reconstruction. However, monocular vision cues like shading [1,2] and texture [3] determine only surface orientation. Surface orientation description could be adequate for model registration [4], object recognition [5], segmentation [6], etc., but is still behind absolute depth description in the diversity of applications. In addition, the above vision cues generally demand restrictive assumptions on the appearance or reflective property of the target object. Other more artificial cues like shape from polarization [7] and photometric stereo [8] have similar restrictions. Multiocular cues such as stereo vision [9,10] and visual motion [11] can determine absolute depth, but they require solving the correspondence problem – the problem of establishing dense correspondences between multiple image planes – which is widely recognized as a difficult one.

Structured light-based sensing systems (SLSs) [12–17] are systems engineered to avoid the above difficult issues. Their system configuration has much resemblance with that of stereo vision, but with one camera replaced by a projector and the associated image by an illuminated pattern. Like stereo vision, they can determine absolute depth. The replacement of a camera by a projector makes the following differences. With a projector to illuminate pattern onto the scene, they can relieve the demand of texture or certain reflective property of the target object. By embedding into the illuminated pattern a unique code for each pattern position, they can reduce the correspondence problem to a much easier one. Compared with other range sensing systems like laser scanning [18,19], SLSs have the essence that they do not require scanning nor the use of moving parts, meaning that they have the potential of achieving faster operation with simpler hardware.

SLSs have their own deficiencies, however. If the unique label for each pattern position is coded spatially over the pattern domain (for example, in the case of SLSs that project pseudorandom color pattern [20–24]), the labels would manifest as a window of

* Corresponding author.
E-mail addresses: zhan.song@sub.siat.ac.cn (Z. Song), rchung@mae.cuhk.edu.hk (R. Chung).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.05.008

pattern profile about each pattern position. These windows of pattern profile need to be preserved in the image data, or else the labels would be destroyed in the image data. This actually asks for certain smoothness of the target object. The problem is, the finer is the resolution demanded from the reconstruction, the more will be the pattern positions to code, the bigger will be the coding windows, and the more will be the smoothness demanded from the target object. The non-zero size of the coding windows is thus a restriction. The window deficiency can be removed by constructing the code not in the pattern domain but over time, which is the practice of the temporally coded SLSs (for example, those using gray code [13,14]). But such systems require the use of multiple illuminations and multiple image captures, which is undesirable if the target object is in motion.

This article describes a projector-and-camera system that addresses the above deficiencies of SLSs, with a little sacrifice on reconstruction accuracy but with much gain on operation simplicity. It determines not only surface orientation but also absolute depth. It works with only a single illumination and image capture. And it is free of the above window issue. The system projects a very simple pattern, one that is not much more than a checkerboard pattern, which produces in the image a number of grid-lines. Such a projected pattern is considered as *nonstructured* in the language of SLSs, since the pattern does not embed codes to label the pattern positions. Yet it is this nonstructured nature of the projected pattern that allows the system to be window-free, as there is no longer any window of specific pattern profile involved for coding purpose. The precision of 3D reconstruction is generally not as high as that of the traditional SLSs (which have unique codes embedded in the illuminated pattern), but the loss is tolerable (as can be seen from the experimental results presented in this article), and it is traded for the above conveniences.

The nonstructured light-based system operates by the use of two steps: one that determines a surface orientation profile from the observed grid-lines, and another that escalates the orientation description to an absolute depth description by the use of a few tricks related to interpolation. More specifically, we show that the surface orientation of the target object at any projected grid-point **P**, which appears as say point **p** in the image, can be determined from the tangents of the image grid-lines that intersect to form **p**. We further show that, with a certain approximation of the image projection model, we can even remove the need of establishing correspondences between the image plane of the camera and the display panel of the projector. That is why we need project only a checkerboard-like pattern, not a coded pattern. With the mechanism, a discrete orientation map can be constructed for the target object.

In addition, we show that the above orientation map can be escalated to an absolute depth map. In principle, this is not particularly surprising, as we have access to two images or pseudo-images of the target object—the real image captured by the camera, and the pattern constructed on the display panel of the projector. By the triangulation process similar to that of stereo vision, 3D reconstruction should be possible. However, triangulation requires precise point correspondences between the two images, which we do not have. In our system, we tackle the 3D reconstruction task not by triangulation but by a combined use of a few interpolation-related operations plus the above orientation map we recover. The idea goes like this. By embedding into the projected pattern a few very distinct features, like three pattern elements colored in red, green, and blue, respectively, we can easily establish a few (and just a few) correspondences between the image plane and the display panel, and determine absolute depth at these places, which we refer to as the *reference positions*. The depths at the reference positions, and the above orientation map which governs how shape changes on the target object surface, then form the basis of interpolating absolute depth for almost all points on the target surface.

The rest of the paper is organized as follows. In Section 2, we give a brief review of the related works. In Section 3, we describe how we can infer 3D orientation profile from the observed grid-lines in the image. In Section 4, we describe how the orientation description can be used to infer absolute depth. Experimental results on a variety of objects are presented in Section 5. Conclusion and possible future work are offered in Section 6.

2. Previous work

There already exist many reviews in the literature on the vision clues like shading, texture, stereo vision, etc. Since this work is more related to structured light-based sensing, here we only briefly review SLSs and systems that use a projector and a camera for 3D reconstruction.

There are two types of SLSs: the temporally coded ones and the spatially coded ones [12]. The former ones have the code of each pattern position constructed over time. Such systems however require multiple pattern projections and image captures, meaning that they are less suitable for tackling dynamic scenes. The latter ones code each pattern position over the pattern domain, by constructing the pattern in such a way that the window (of a certain size) about each pattern position has a unique pattern profile. Such systems can operate with a single illumination and image capture, though at a price. For a window of pattern profile to be preserved in the image data, the scene must exhibit certain smoothness, which is a restriction. The bigger is the window (whose size generally ought to grow with the number of pattern positions to be coded uniquely), the more demanding is the restriction. One way of reducing the window size is to use more colors in the pattern, so that even a smaller window can contain information rich enough to define a unique code. Unfortunately, there is the color confusion problem on the image side that restricts the number of usable colors. In the method proposed by Adan and Molina [22], a 2D color pattern is designed for dynamic feature tracking. Circular spots with seven different colors are used as the pattern elements. The centroids of circular spots are defined as the pattern feature points. Codeword of each feature point is formed by its six neighboring colored pattern elements. The experimental result shows a spatial resolution of 5 mm at a working distance of 1 m. In [23], a spatially encoded structured light pattern of size 53×38 is created via an iterative process. Every 3×3 window in the pattern is unique. A thresholding method is used to segment the colored pattern elements. A labeling algorithm locates the different regions in the binary images, and the centroids are extracted and encoded subsequently. The accuracy is estimated to be within 2 mm for an object located about 1.5 m from the sensor. In the method proposed by Chen and Li [24], by using seven colors, a pattern of size 82×82 is constructed. The centroid of each rectangular pattern element is defined as the pattern feature. Each feature point is encoded by the colors of its four adjacent (i.e. top-left-bottom-right) neighbors. In the experiment, the relative measurement error is only 2%.

The above SLSs are generally designed to recover absolute depth. Some structured light-based methods have also been proposed to recover local surface orientations. In the method proposed by Shrikhande and Stockman [25], a grid pattern is used. Surface orientations at the intersections are inferred from the change of lengths of the grid edges in the image data. The result however indicates that there could be large errors at places where the grid-cell edges have large distortion, and such distortion is

often induced by sharp edges in 3D. In the surface normal measurement with ideal plane, errors of about 7–9° are reported. In [26] Sugihara proposed a method, that infers surface orientation using the distortion of a specific and known pattern projected on the surface. The projected textures on the object surface are detected and the distortion from the regular pattern is measured in order to estimate the surface normals. In the work by Davies and Nixon [27], a pattern which consists of a hexagonally tessellated array of colored encoded circular spots is proposed. The coding scheme is based upon Griffin's method [28]. The work assumes that a circular pattern spot will be perceived as an ellipse in the image when it is projected on the surface. A novel formulation of the Hough Transform is proposed to extract the ellipses from the image. The work showed that surface normal at the ellipse's center point can be deduced from the shear and scaling factor of the ellipse in the direction of the epipolar line. However, in calibrating both the camera and projector only the simple pinhole camera model is used, and in the reconstruction process perspective effects are considered negligible. In addition, the author did not provide any experimental data of surface normal or depth for evaluation of the 3D reconstruction quality. Winkelbach and Wahl [29] used a uni-direction strip pattern to compute the surface normal at each edge point. A fringe pattern is projected onto the object twice from two different angles, each time with a separate image captured, so that two surface tangents are available for each image position. Surface normal at every image point is then interpolated from the data in the two images. Accuracy of the measurement is evaluated by the standard deviation within a planar surface, and the result is approximately 0.41°. A simplified version of the method which determines surface normals from the slopes and intervals of the stripes in the image was also proposed in [30]. The method is based on the assumption that the surface patch between two strip edges is planar or very smooth. Thus the tilt angle can be estimated from the deformed widths of the strip by comparing them with the strip-width measured for a reference plane. In the experiment with a sphere, a mean angular error of 1.79° with a standard deviation of 1.22° are reported.

Compared with shape-from-shading and photometric stereo methods, SLS does have appealing advantages. At the cost of involving artificial pattern projection, it relieves the need of having distinct features or certain albedo property exist on the target object, and it is less affected by imaging noise or the lighting condition. Yet those SLS methods that determine absolute depth require the use of either multiple image captures or coding windows of nonzero size. As for those that determine local surface orientations, they generally require to assume parallel projection on the projector side and parallel imaging on the camera side. In other words, the intrinsic parameters of both the camera and projector are not considered, and errors arise from such much simplified projection and imaging models can be significant.

In this work, we pursue a variation of the projector-and-camera system that reconstructs absolute depth, not just orientation map, from a single pattern projection and image capture, and that is achieved without the issue of nonzero window size in processing the image data. The way to avoid the window issue is not to introduce coding to the projected pattern. In other words, we use nonstructured illumination. In fact the projected pattern is not much more than a uniform grid pattern that consists of rhombic pattern elements. We show that it is possible to determine orientation map from a single illumination of the nonstructured pattern and a single image capture, without establishing explicit correspondences between the projector's display panel and the camera's image plane (the projected pattern is not coded anyway). In the process, we also introduce ways to allow the projection models to be escalated from parallel projection and parallel imaging, and become closer to perspective projection and imaging. The orientation map is then converted to a depth map in another step, using a mechanism related to interpolation. The performance of the whole system is also benefited from the use of an accurate projector-and-camera calibration system that was reported in [31].

3. Inference from observed grid-lines to surface orientation

Suppose we program a pattern on the display panel of the projector, which consists of a 2D array of alternate dark and bright rectangular blocks, and project it to the object surface. The edge of the blocks forms the grid-lines, and the intersections of every two grid-lines form the grid-points. The grid-lines and grid-points are generally accessible in both the projection panel and the image plane.

3.1. The underlying principle

Consider any grid-point $\mathbf{p}_p$ in the pattern panel of the projector, and the two accompanying grid-lines that compose it, as illustrated by Fig. 1a. Suppose the grid-point and grid-lines induce an imaged grid-point $\mathbf{p}_c$ and two imaged grid-lines (or more correctly grid-curves, as the original grid-lines are generally modulated by the curvature of the surface in 3D and appear not as lines in the image data) on the image plane via a 3D point $\mathbf{P}$ on the object surface in space. Suppose the tangents to the grid-lines at point $\mathbf{p}_p$ in the projector's projection panel are $\mathbf{t}_{p1}$ and $\mathbf{t}_{p2}$, respectively, and the tangents to the grid-lines at point $\mathbf{p}_c$ in the image plane are $\mathbf{t}_{c1}$ and $\mathbf{t}_{c2}$.

Tangent $\mathbf{t}_{p1}$ and the grid-point $\mathbf{p}_p$ in the pattern projection panel together form a plane $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ of illumination from the projector's light source, which is reflected by the object surface at point $\mathbf{P}$ and becomes the plane of projection $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ to the camera's optical center. The intersection of the two light planes $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ actually defines a tangent $\mathbf{t}_1$ in 3D to the object surface at point $\mathbf{P}$. $\mathbf{p}_p$ and $\mathbf{t}_{p1}$ are fully accessible as they are entities under system design, and so are $\mathbf{p}_c$ and $\mathbf{t}_{c1}$ as they are entities observable from the image data. Thus the two light planes $\Pi(\mathbf{p}_p, \mathbf{t}_{p1})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c1})$ are both constructible, and their intersection $\mathbf{t}_1$ can be determined. In fact the tangent $\mathbf{t}_1$ to the object surface at point $\mathbf{P}$ is merely the cross-product of the surface normals of the two light planes.

Similarly, another tangent $\mathbf{t}_2$ to the object surface at point $\mathbf{P}$ can be determined as the cross-product of the surface normals of two other light planes: $\Pi(\mathbf{p}_p, \mathbf{t}_{p2})$ and $\Pi(\mathbf{p}_c, \mathbf{t}_{c2})$, which are both accessible from design and image observations. In other words, by

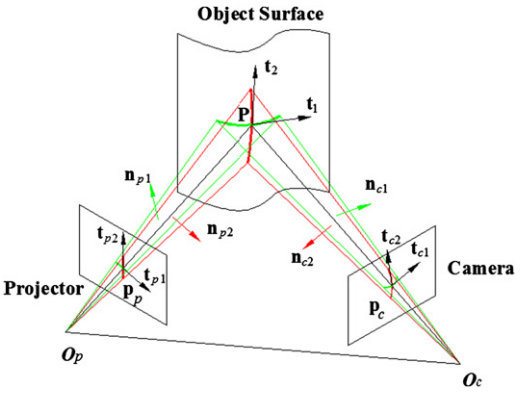

Fig. 1. Determining local surface orientation by the projection and image-capture of grid-lines that meet at a grid-point.

simply taking one image of the object surface that is under projection of a grid-like pattern, for any imaged grid-point $\mathbf{p}_p$ at position (x,y) on the image plane, the surface orientation $\mathbf{n}(x,y)$ of the object surface at the associated 3D point can be determined simply as $\mathbf{n}(x,y)=\mathbf{t}_1\times\mathbf{t}_2$ from image observations $\{\mathbf{p}_c, \mathbf{t}_{c1}, \mathbf{t}_{c2}\}$ and pattern data $\{\mathbf{p}_p, \mathbf{t}_{p1}, \mathbf{t}_{p2}\}$. This is the underlying principle of this work.

3.2. The path from image tangents to 3D orientation

Suppose $\mathbf{n}_{c1}$ and $\mathbf{n}_{c2}$ are the surface normals of the light planes $\prod(\mathbf{p}_c, \mathbf{t}_{c1})$ and $\prod(\mathbf{p}_c, \mathbf{t}_{c2})$ on the camera side, and $\mathbf{n}_{p1}$ and $\mathbf{n}_{p2}$ those of the light planes $\prod(\mathbf{p}_p, \mathbf{t}_{p1})$ and $\prod(\mathbf{p}_p, \mathbf{t}_{p2})$ on the projector side. Suppose the intrinsic parameters of the camera and projector (which is modeled as another perspective camera, except that light is coming out of it instead of going into it) have been calibrated, which are focal lengths f_c, f_p, and principal points $\boldsymbol{C}_c(x_{c0}, y_{c0})$, $\boldsymbol{C}_p(x_{p0}, y_{p0})$. Then $\mathbf{n}_{c1}$, $\mathbf{n}_{c2}$, $\mathbf{n}_{p1}$, $\mathbf{n}_{p2}$ can be determined as

$$\mathbf{n}_{ci}=\mathbf{p}_{ci}\times\mathbf{l}_{ci},\quad \mathbf{n}_{pi}=\mathbf{p}_{pi}\times\mathbf{l}_{pi},\quad i=1,2 \tag{1}$$

where $\mathbf{p}_{ci}=[x_{ci},\ y_{ci},\ -f_c]^{\mathrm{T}}$, $\mathbf{p}_{pi}=[x_{pi},\ y_{pi},\ -f_p]^{\mathrm{T}}$ represent the projective points on the image and projection panel planes, respectively, and $\mathbf{l}_{ci}$ and $\mathbf{l}_{pi}$ indicate the projective lines defined by the image tangents $\mathbf{t}_{c1}$, $\mathbf{t}_{p1}$ and projective points $\mathbf{p}_{ci}$ and $\mathbf{p}_{pi}$.

As discussed earlier, the two 3D tangents $\mathbf{t}_1$, $\mathbf{t}_2$ to the object surface at point $\mathbf{P}$ that is associated with image point $\mathbf{p}_p=(x,y)$, if with reference to the camera coordinate system, can be obtained as

$$\mathbf{t}_1=\mathbf{n}_{c1}\times(\mathbf{R}\mathbf{n}_{p1}) \tag{2}$$

$$\mathbf{t}_2=\mathbf{n}_{c2}\times(\mathbf{R}\mathbf{n}_{p2}) \tag{3}$$

where $\mathbf{R}$ represents the rotational relationship between the camera and projector coordinate frames.

Finally, the surface normal to the object surface at point $\mathbf{P}$ that is associated with the image position (x,y) can be determined as

$$\mathbf{n}(x,y)=\mathbf{t}_1\times\mathbf{t}_2 \tag{4}$$

Notice that the determination of the local surface orientation as expressed by Eq. (4) is a deterministic process that asks for only image information local to the specific point. Unlike shape from shading and other image intensity-based methods, it does not require assumption about how local surface orientations at neighboring points are related. More specifically, it requires no process of iterations to determine local surface orientations.

3.3. Avoidance of the correspondence need

Eqs. (1)–(4) lay down a simple mechanism of determining local surface orientations at all grid-points. However, the equations also reveal that at each grid-point pictured in the image data, the associated tangents $\{\mathbf{t}_{c1}, \mathbf{t}_{c2}\}$ have to be paired up with certain $\{\mathbf{t}_{p1}, \mathbf{t}_{p2}\}$ on the illuminated pattern. This is in a way a request of correspondences.

We avoid the request in the following way. Suppose we can assume the parallel projection model on the projector side (and just on the projector side not on the camera side). With that projection model, grid-lines that are parallel on the projection panel will have parallel light planes coming out from the projector. As the illuminated pattern we use is a regular grid, the grid-lines on it have only two orientations on the projection panel; one half of the grid-lines have one orientation, and the other half have the other orientation. In turn, the light planes from the projector have only two surface normals. In other words, it does not matter which pair of $\{\mathbf{t}_{c1}, \mathbf{t}_{c2}\}$ are to be paired with; the associated planes of light from the projector are fixed (and can be predicted from the projector-and-camera system calibration). In fact, we do not stop at the orientation estimates resulted from the assumption of parallel projection on the projector side. In the next subsection we also describe how we escalate the projection model on the projector side to one that is closer to perspective projection, and in so doing improve the quality of the surface orientation estimates.

3.4. Correction for the parallel projection approximation on the projector side

The projector's illumination process cannot be exactly captured by the parallel projection model, and discrepancy from it generally causes bigger error to the determination of n_{p1} and n_{p2} at positions farther away from the center (i.e.., the principal point) of the projection panel. The perspective model is more accurate but it requires the correspondence between p_c and p_p. We use a simple correction mechanism to fix the problem. For each grid-point p_c $(=(x_c, y_c))$ in the image plane, we adopt a linear mapping, as described below, to predict more finely the associated grid-point p_p $(=(x_p, y_p))$ in the projection panel:

$$[x_p,y_p]=\left(\frac{x_c-x_{min}}{x_{max}-x_{min}}W_p, \frac{y_c-y_{min}}{y_{max}-y_{min}}H_p\right) \tag{5}$$

where (x_{min}, y_{min}), (x_{max}, y_{max}) represent the positions of the top-left and bottom-right grid-points in the image plane, W_p and H_p are the width and height of the pattern on the projection panel counted in the projection panel's pixels. This way, an estimation of the correspondence better than that under the parallel projection model (for the projector) is used to determine $\mathbf{n}_{p1}$ and $\mathbf{n}_{p2}$. The determination accuracy is indeed improved, as evidenced by the experimental results presented in the experimental section.

3.5. Calibration of projector–camera system

While in previous works like [25–27,29,30] illumination and imaging are modeled as parallel projections, here they are modeled as perspective projections or models close to them. With this, the proposed method can expectedly reach more precision in 3D reconstruction. However, it is also obvious that the calibration of the intrinsic and extrinsic parameters of the projector-and-camera system is necessary and important, or else the light planes and their intersections cannot be determined from Eqs. (1)–(4). On this, we use a calibration mechanism [31] that makes use of an LCD display panel as the external reference object. The calibration mechanism has shown to be able to reach higher calibration accuracy, allowing more precise surface normal calculation to take place. Details of the calibration mechanism are available in [28].

3.6. Feature extraction

The reconstruction accuracy also depends upon how accurately the grid-lines and grid-points are located in the image, and how precisely 2D tangents to the grid-lines at the grid-points are extracted. There has been a rich body of works in the literature on feature detection, such as LoG [32], Harris [33], or SUSAN [34], etc. All these detectors are image intensity based, making their feature localization accuracy generally sensitive to image noise and surface textures.

In this work, we use a regular grid pattern that consists of rhombic pattern elements as the illuminated pattern, and we designate the grid points – the points of intersection of the grid-lines – as the feature points. As detailed in a separate paper [35]

that focuses on the subject, a local symmetry property of the grid-points can be exploited for their localization in sub-pixel accuracy. Notice that a small circular disc of intensities centered at any grid-point resembles itself when rotated by 180°. In the study of plane symmetry, this is referred to as the two-fold rotation symmetry or *cmm* symmetry. Such a symmetry on the illumination side is quasi-invariant with image noise, image blur, and projective distortions, and is largely preserved in the image data for the reason that the linearity of the segment that divides the circular disc into two symmetrical halves is preserved under the processes. If the circular disc is sufficiently small, even nonzero curvature of the illuminated surface in 3D has little effect to the symmetry. This quasi-invariance of the symmetry, as opposed to direct image intensities, is what we exploit for precise grid-point detection from the image. More details about the grid-point detector, such as the robustness test, comparison with traditional operators, etc., are reported in [35]. As detailed in [35] again, once the grid-points are located in the image data, the image tangents $\mathbf{t}_{c1}$ and $\mathbf{t}_{c2}$ can be calculated from its four neighboring grid-points via a simple linear interpolation process.

4. Inference from orientation map to depth map

The escalation of an orientation profile to an absolute depth map naturally involves interpolation. Any interpolation scheme generally starts from the absolute depth information of a few positions, and propagates the depth values to other parts of the image domain, using the above orientation information as the binary constraint that relates the absolute depths of every two neighboring image positions.

The orientation figures we recover from the above process are discrete, in the sense they are available only at the image positions where grid-points are observed. Yet the interpolation process requires dense orientation information to proceed. Thus we ought to, in the first place, interpolate a dense set of orientations, in the form of an orientation map, from the above discrete set of orientations.

On the absolute depth information of a few positions, we can always make use of the extreme corners of the entire illuminated pattern or introduce a few distinct feature points into the pattern itself. Upon recognition of such distinct feature points or corners in the image data, which we call the reference points, triangulation can proceed and a few absolute depth values can be attained. That is the simple mechanism we employ in this work.

As for the use of the orientation information in the interpolation process, we have a more challenging issue to address. We have to transform the orientation map, which is over the image domain, to an orientation map over 3D and in metric units, as the interpolation process is about inferring depth in space and in metric units. Suppose we designate $\mathbf{x}=(x, y)$ as the image coordinates of an image position, and $\mathbf{X}=(X, Y, Z)$ as the 3D coordinates of the associated object point with reference to the coordinate frame of the camera, where the X–Y plane (that contains the camera's optical center) is parallel to the x–y plane (the image plane). If we express the interpolated orientation map as $\nabla_{x,y}Z(x,y)=(\partial Z/\partial x,\partial Z/\partial y)^{\mathrm{T}}$ over a dense collection of image positions (x, y), the above issue is about transforming $\nabla_{x,y}Z$ to $\nabla_{X,Y}Z$. The challenge is, generally speaking, that (X,Y) are related to (x, y) through Z, and so are $\nabla_{X,Y}Z$ and $\nabla_{x,y}Z$, yet the depth measure Z is yet to be made known at this point. We have to find a way to arrive at a reasonable approximation of $\nabla_{X,Y}Z$ from $\nabla_{x,y}Z$ without knowing Z. Below we address the above points one by one.

4.1. Turning discrete orientations to an orientation map

Given a mesh of grid data, we have at each grid-point (x, y) in the image data a surface normal $\mathbf{n}(x, y)$. In our implementation, a simple median filter is first used to remove the outliers within the collection of discrete orientations, which exist mostly around the object silhouette. Then the discrete data go through a bilinear interpolation to form a regular and dense surface orientation map. However, such an orientation map is not guaranteed to be integrable. Below we describe how we adjust the orientation map to make it an integrable orientation map N. Our ultimate task is to reconstruct a depth map in space from the surface normal field N.

To ensure the integrability of the orientation map, we do the following. Denote p and q to be the partial derivatives of scene depth Z with respect to x and y, respectively:

$$p=\partial Z/\partial x, \quad q=\partial Z/\partial y \tag{6}$$

The surface normal at any given image point (x, y) can be represented by

$$\mathbf{n}(x,y)=(p,q,-1)^{\mathrm{T}} \tag{7}$$

To generate an integrable gradient map $(\tilde{p},\tilde{q})$ from the above interpolated orientation map(p,q), we conduct a minimization of a distance measure that has the following form:

$$E(\{(\tilde{p},\tilde{q})\})=\iint(|\tilde{p}-p|^2+|\tilde{q}-q|^2)\,dx\,dy+\lambda\iint(p_x^2+p_y^2+q_x^2+q_y^2)\,dx\,dy \tag{8}$$

where the second term indicates a smoothness constraint with a weight λ of importance, and p_x,p_y,q_x,q_y are the gradients of p and q along the x and y directions, respectively. Notice that the minimization process should also satisfy:

$$\partial p/\partial y=\partial q/\partial x \tag{9}$$

at every position (x,y), which is known as the integrability constraint. In our implementation, Frankot and Chellappa's method [36], which uses a collection of Fourier basis functions to represent the surface slope profile, is adopted for the minimization process.

4.2. Transforming orientation map from over image coordinates to over 3D coordinates

Fig. 2 shows three projection models: the full perspective, orthographic, and weak perspective (also named scaled orthographic) projections. The full perspective model is accurate, but then under it (x, y) and (X, Y) are related by Z, which is unknown yet and is generally different at different (x, y). The orthographic model is simple but has a very restrictive limited field of view. The weak perspective model is a compromise between the full perspective and the orthographic models on accuracy versus complexity. It will be a suitable projection model under the conditions that: (1) the depth range ΔZ of the imaged part of the scene is small, with respective to the average distance $\overline{Z}$ of the scene from the camera; (2) the imaged part of the scene is close to the optical axis, again with respective to the average distance $\overline{Z}$ from the camera. As we target our projector-and-camera system at applications that satisfy the conditions, we adopt the weak perspective model for the integration process.

Under the weak perspective model, (X, Y) is a certain scaling of (x, y), though the scaling could be different along the x- and y-directions. More precisely, the transformation from the image plane x–y to the object plane X–Y under weak perspective

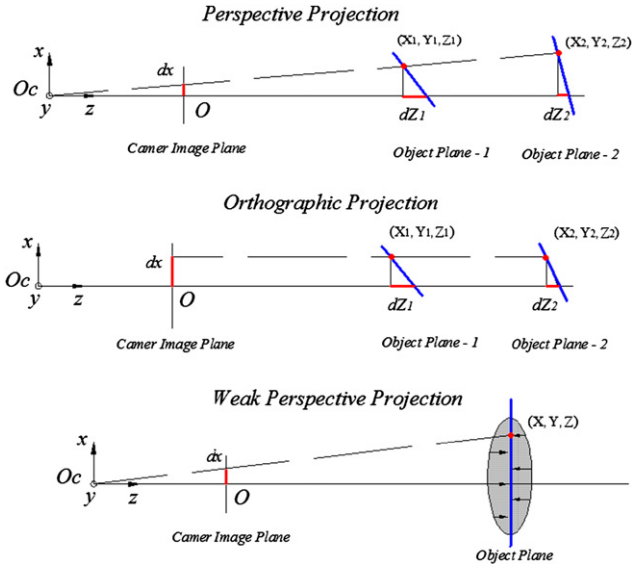

Fig. 2. Illustrations of three projection models: full perspective, orthographic and weak perspective models.

projection can be formulated as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} k_{\bar{Z}} & 0 & 0 & 0 \\ 0 & k'_{\bar{Z}} & 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

or simply

$$X = \frac{x}{k_{\bar{Z}}}, \quad Y = \frac{y}{k'_{\bar{Z}}} \tag{10}$$

for certain scalars $k_{\bar{Z}}$, $k'_{\bar{Z}}$ that depend upon the average depth $\bar{Z}$ of the imaged scene.

The above equation shows that only one correspondence (over a reference point as we call it) between x–y and X–Y is sufficient to determine $k_{\bar{Z}}$ and $k'_{\bar{Z}}$. However, to boost the accuracy, multiple correspondences (over more than one reference point) are generally used. Suppose we have m correspondences

$$\{ (x_i, y_i) \quad (X_i, Y_i, Z_i) \}$$

where $i = 0,1,\ldots,m-1$

Then the simple average measures of $k_{\bar{Z}}$ and $k'_{\bar{Z}}$ can be determined as

$$\begin{cases} k_{\bar{Z}} = \dfrac{1}{m} \displaystyle\sum_{i=0}^{m-1} \frac{x_i}{X_i} \\ k'_{\bar{Z}} = \dfrac{1}{m} \displaystyle\sum_{i=0}^{m-1} \frac{y_i}{Y_i} \end{cases} \tag{11}$$

Once we have knowledge of $k_{\bar{Z}}$ and $k'_{\bar{Z}}$, the transformation from (x, y) to (X, Y) can be conducted through Eq. (10), i.e.,

$$X = \left(\frac{1}{k_{\bar{Z}}}\right) x, \quad Y = \left(\frac{1}{k'_{\bar{Z}}}\right) y \tag{12}$$

The orientation information over the image domain (x, y) can also be transformed to the equivalent over the (X, Y) domain:

$$\frac{\partial Z}{\partial X} = k_{\bar{Z}} \frac{\partial Z}{\partial x}, \quad \frac{\partial Z}{\partial Y} = k'_{\bar{Z}} \frac{\partial Z}{\partial y} \tag{13}$$

4.3. Escalating orientation map to a depth map

Assume that the imaged scene consists of an integrable surface, and that the 3D information of a reference point $(X_0, Y_0, Z(X_0, Y_0))$ of the surface in space is available. Given an image position (x, y), the depth Z of the associated object point (X, Y, Z) is captured by the following integration along any smooth path from $(X_0, Y_0, Z(X_0, Y_0))$ to (X,Y,Z) on the imaged surface:

$$Z(X,Y) = \underbrace{\int_{(X_0,Y_0)}^{(X,Y)} \left(\frac{\partial Z}{\partial X} dX + \frac{\partial Z}{\partial Y} dY \right)}_{Z'(X,Y)} + Z(X_0, Y_0) \tag{14}$$

Here we designate $Z'(X,Y)$ as the first term (i.e., the integration term) of the above equation. Geometrically, it refers to the difference of depth (from the camera) between the object point (X, Y, Z) and the reference point $(X_0, Y_0, Z(X_0, Y_0))$.

From the above equation, we can see that any absolute depth value $Z(X,Y)$ can be represented by an integrated term $Z'(X,Y)$ and the depth $Z(X_0,Y_0)$ at a given reference point (X_0,Y_0). In other words, if we have $Z(X_0,Y_0)$, we can determine absolute depth $Z(X,Y)$ at any object point (X,Y).

Now suppose we have depth $Z(X_i,Y_i)$ at not one but m reference points (X_i,Y_i), where $i = 0,1,\ldots,m-1$. We have the following with reference to the ith reference point:

$$Z(X,Y) = \int_{(X_i,Y_i)}^{(X,Y)} \left(\frac{\partial Z}{\partial X} dX + \frac{\partial Z}{\partial Y} dY \right) + Z(X_i, Y_i) \tag{15}$$

The above equation can be expressed as the following:

$$Z(X,Y) = \underbrace{\int_{(X_0,Y_0)}^{(X,Y)} \left(\frac{\partial Z}{\partial X} dX + \frac{\partial Z}{\partial Y} dY \right)}_{Z'(X,Y)} - \underbrace{\int_{(X_0,Y_0)}^{(X_i,Y_i)} \left(\frac{\partial Z}{\partial X} dX + \frac{\partial Z}{\partial Y} dY \right)}_{Z'(X_i,Y_i)} + Z(X_i, Y_i) \tag{16}$$

Putting together the above equation of all m reference points, we have the following for any (X, Y):

$$\begin{cases} w_0 Z(X,Y) = w_0 Z'(X,Y) - w_0 Z'(X_0,Y_0) + w_0 Z(X_0,Y_0) \\ w_1 Z(X,Y) = w_1 Z'(X,Y) - w_1 Z'(X_1,Y_1) + w_1 Z(X_1,Y_1) \\ \vdots \\ w_{(m-1)} Z(X,Y) = w_{(m-1)} Z'(X,Y) - w_{(m-1)} Z'(X_{m-1},Y_{m-1}) + w_{(m-1)} Z(X_{m-1},Y_{m-1}) \end{cases} \tag{17}$$

where w_i represents the weight of relevance of the ith reference point to (X,Y) according to how far it is away from (X,Y), in the form:

$$w_i = \frac{1/d_i(X,Y)}{\sum_{i=0}^{m-1} 1/d_i(X,Y)} \tag{18}$$

with $d_i(X,Y)$ being the distance between the reference point (X_i,Y_i) and the considered point (X,Y). Notice that $\sum_{i=0}^{(m-1)} w_i = 1$. Summing all the component equations of (17) together, we have a measure of depth $Z(X,Y)$ that takes into account the information of all m reference points:

$$Z(X,Y) = Z'(X,Y) + \sum_{i=0}^{m-1} w_i [Z(x_i,y_i) - Z'(x_i,y_i)] \tag{19}$$

Obviously, in the above formula, $Z'(X_i,Y_i)$ for all i, as the difference in depth between one reference point and the other, can be determined without relating to any arbitrary point (X,Y). They can be read directly from the depth data of the reference points. The only term that need be processed is the first term of the right hand side, which involves an integration (which is laid down explicitly as the first term of the right hand side of Eq. (14)) from the 0th reference point to the point (X,Y) along any integrable path on the target surface.

In our implementation, we embed three colorful pattern elements (for ease of their detection in the image data, we use red, green, and blue) into the illuminated pattern, as shown in Fig. 3, and use them as the reference points. The absolute depths at their centroids can be determined via triangulation. Suppose their depth values are

$$\{Z(X_r,Y_r),Z(X_g,Y_g),Z(X_b,Y_b)\}$$

respectively. Then Eq. (19) can be expressed as

$$Z(X,Y)=Z'(X,Y)+\begin{bmatrix} w_r, & w_g, & w_b \end{bmatrix}\left(\begin{bmatrix} Z(X_r,Y_r) \\ Z(X_g,Y_g) \\ Z(X_b,Y_b) \end{bmatrix}-\begin{bmatrix} Z'(X_r,Y_r) \\ Z'(X_g,Y_g) \\ Z'(X_b,Y_b) \end{bmatrix}\right) \tag{20}$$

where

$$w_r=\frac{1/d_r(X,Y)}{[(1/d_r(X,Y))+(1/d_g(X,Y))+(1/d_b(X,Y))]}$$
$$w_g=\frac{1/d_g(X,Y)}{[(1/d_r(X,Y))+(1/d_g(X,Y))+(1/d_b(X,Y))]} \tag{21}$$
$$w_b=\frac{1/d_b(X,Y)}{[(1/d_r(X,Y))+(1/d_g(X,Y))+(1/d_b(X,Y))]}$$

Indeed the method also demands a degree of smoothness in the imaged scene, as reflected by the use of integration on the orientation map. However, unlike the spatially coded SLSs, the method does not involve coding windows, and thus does not generally demand more smoothness in the imaged scene for more resolution in the 3D reconstruction. As the method does not require to assure that each pattern position comes with a unique window code, it is also simpler to operate; for example, a black-and-white pattern in its illumination is already sufficient.

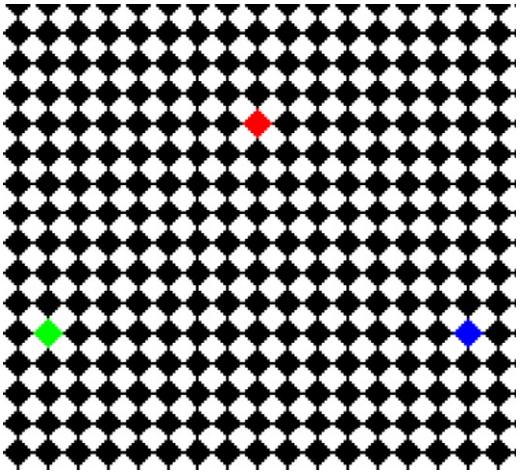

Fig. 3. Partial view of the proposed non-structured pattern. Three colorful elements are embedded into the illuminated pattern as the reference points, and the absolute 3D depths at their centroids are determined by triangulation.

5. Experimental results

The projector–camera system we used for experimentation consisted of a DLP projector of resolution 1024×768 (NEC LT25 projector) and a camera of resolution 1500×1200 (Pentax K100D DSLR camera with Pentax DA 18–55 mm lens), both being off-the-shelf equipments. The focal length of the projector was of the range 25–31 mm. The setup was configured for a working distance of about 850 mm. The system was first calibrated using the LCD panel-based method described in [31].

Our system is designed for the reconstruction of objects with size and depth range similar to human face. For such target objects, the depth range is about 5–10 cm and is relatively small with respect to the projection and imaging distance. The view angle between the camera and projector is about 25–30°, and it does not produce too big disparity in the captured image either. Such configuration can benefit the linear mapping procedure and control the mapping errors within a small range. Although such linear mapping is not general tenable, but it does give major improvement to surface normal estimation in comparison with the previous parallel model-based methods. In the pattern design, denser grid-points are expected to boost the reconstruction precision. However, higher grid-point density is subject to the pixilation effect and limited resolutions of both the projector and the camera. In this work, we use rhombic pattern elements of 5 pixels radius. We also tried pattern elements of smaller sizes, such as those of 3 pixels radius, but the pattern showed strong pixilation and that also aroused difficulty in grid-point detection and image tangent estimation.

5.1. Accuracy in planar reconstruction

In one experiment, we used the proposed method to reconstruct the orientation map of an ideal plane and examined how planar the reconstruction was. We used an LCD panel as the test object, whose planarity was of industrial grade. Shape reconstruction from the proposed method showed only small

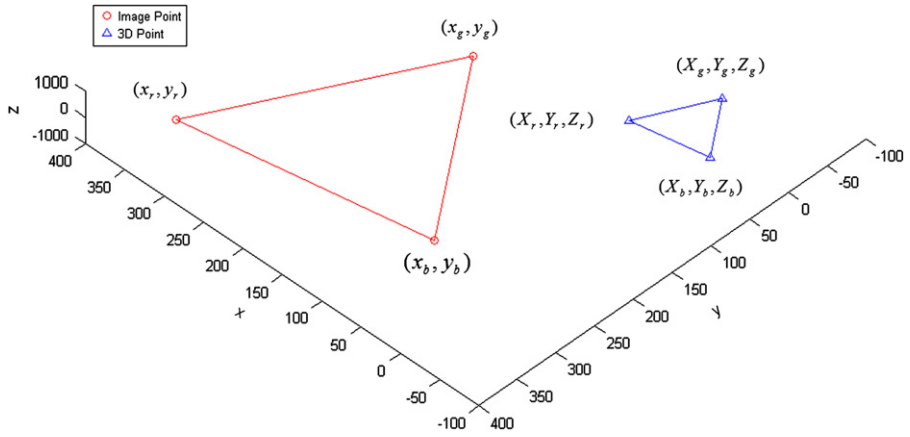

Fig. 4. Illustration of the reference points in an experiment, that are essential for the determination of the scalar factors $k_{\bar{Z}}$ and $k'_{\bar{Z}}$ (which allow orientation information over the image domain x–y to be transformed to those over the object plane X–Y) and for the subsequent integration process.

deviation from planarity: of all the measured points, only an average discrepancy of 0.82° (from an ideal plane) and a standard deviation of 0.15° (within the orientation distribution) were shown in the recovered orientation profile.

To compare the surface normal reconstruction accuracy with that of a spatially coded SLS system, we replaced the black-and-white grid pattern with a colored one (consisting of three colors: red, green, and blue) which used a pseudorandom pattern coding strategy to assure a unique coding window at each pattern position; the details can be found in an earlier report [21]. Without the system configuration ever adjusted, result that had an average discrepancy of 0.69° and a standard deviation of 0.13° was acquired. The comparison shows that, the sacrifice of the proposed method in the reconstruction quality was tolerable.

Note that in methods like [25–27,29,30] and other earlier works, discrepancies in the range of 2°–8° and higher were

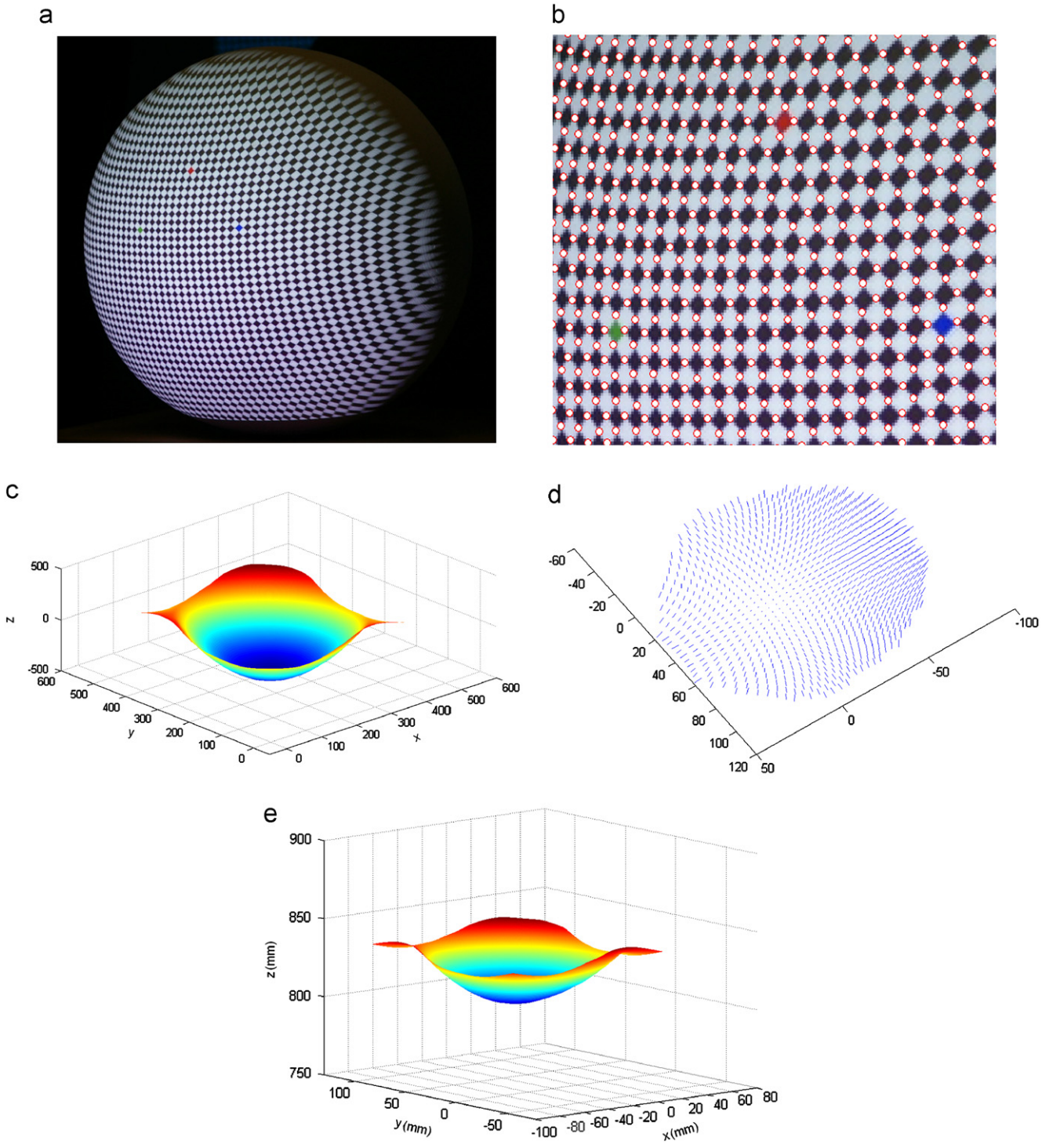

Fig. 5. Reconstruction of a spherical surface by the proposed method: (a) object surface under pattern illumination; (b) grid-points detection result; (c) local surface orientations determined at the grid-points; (d) relative depth map in pixel units acquired without using reference points; and (e) absolute depth map reconstructed by the proposed method.

reported. Certainly, the experiments were conducted with different equipments and under different setups, and plain comparison of the figures would not be sensible. Nonetheless, one can observe that the reconstruction quality of the proposed method could possibly achieve figures that were reported in the literature.

We attribute the promising reconstruction quality to the use of a linear mapping to refine the correspondences inferred from parallel projection, making them closer to what one would get under perspective projection and establishment of point-to-point correspondences. We have done experiments to see what reconstruction quality would be obtained when the linear mapping is not used. When the linear mapping was not in place, an average deviation of 1.65° and standard deviation of 0.48° of the surface orientations were obtained, which were quite distant from the figures obtained under the use of the linear mapping.

5.2. Accuracy in spherical reconstruction

In another experiment, we used the proposed method to reconstruct a curved object, more specifically a spherical object, of radius about 97 mm. Before the integration process, the scaling factors $k_{\bar{Z}}$ and $k'_{\bar{Z}}$, which allow orientations over the image plane to be transformed to those over the object plane, were first determined. In the determination of $k_{\bar{Z}}$ and $k'_{\bar{Z}}$, and in the subsequent integration process, we need reference points. Fig. 4 shows $(x_r,y_r),(x_g,y_g),(x_b,y_b)$: the centroids of the red, green, and blue reference pattern elements. The associated positions in space are referred to as $(X_r,Y_r,Z_r),(X_g,Y_g,Z_g),(X_b,Y_b,Z_b)$, respectively.

Using Eq. (18), three values of $k_{\bar{Z}}$ and $k'_{\bar{Z}}$ were calculated. The results are: $k_{\bar{Z}}=[0.316, 0.318, 0.311]$, $k'_{\bar{Z}}=[0.324,0.327,0.321]$. We can see that the deviations in each group were relatively small, showing that the weak perspective model was a reasonable one in the experiment.

The shape reconstruction result of the proposed method is shown in Fig. 5. There the grid-point extraction result, the orientation estimates at the grid-points, and the final depth map are all displayed. In Fig. 5(d) we show a depth map determined by integrating over the recovered orientation map, without the use of any reference point. The map is in pixel units because in the absence of reference points the orientation map had to stay with respect to the image domain not the object plane domain. It is also only a relative depth map because the image position where the integration started was in a way assigned with depth 0. Three colorful elements were then designated as the reference points, extracted, and matched to get absolute 3D coordinates at their centroids. The resultant absolute depth

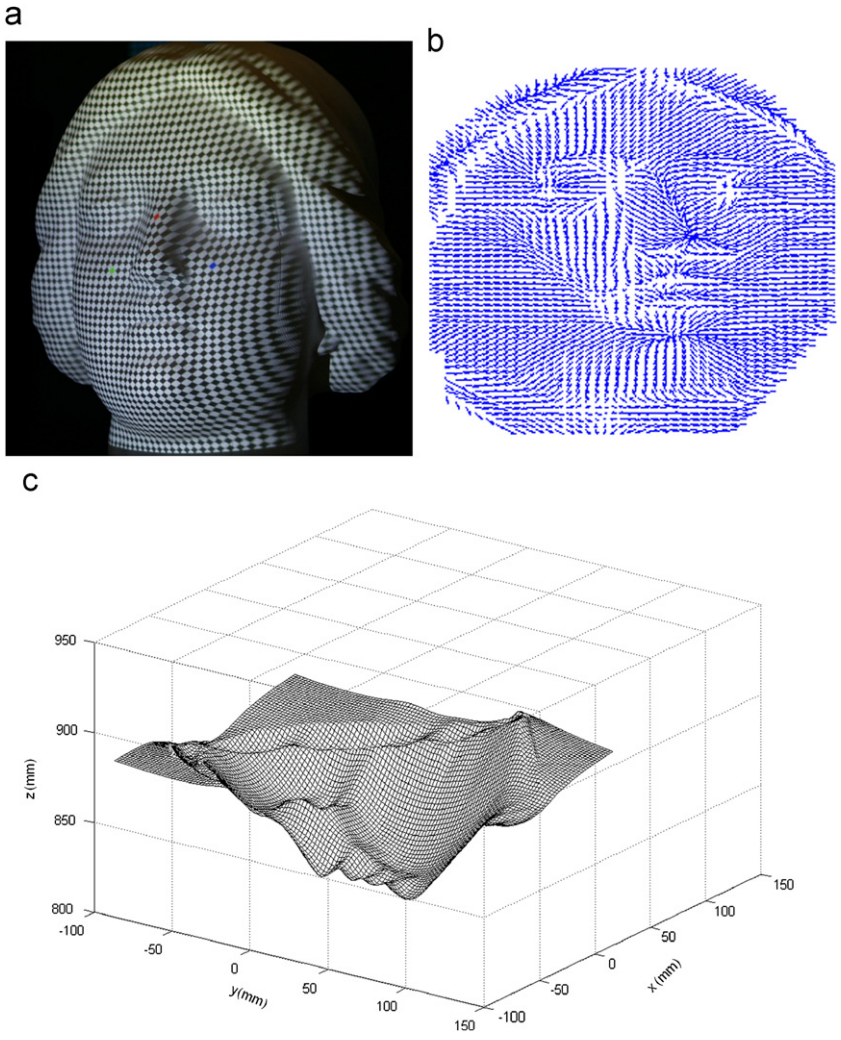

Fig. 6. Reconstruction of a human face bust model by the proposed method: (a) object surface under pattern illumination; (b) recovered orientation map; and (c) recovered absolute depth map viewed from a particular perspective.

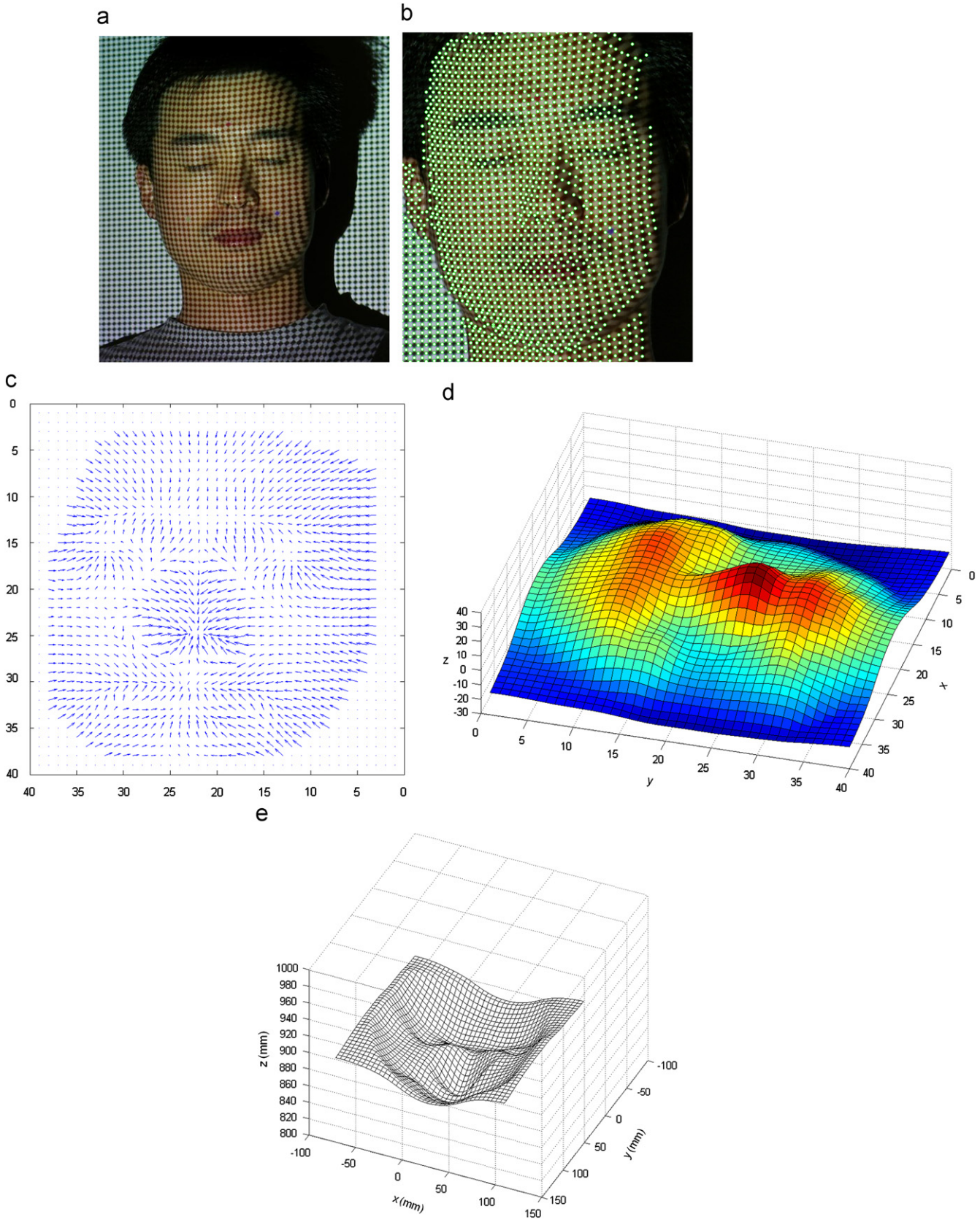

Fig. 7. Reconstruction of a real human face by the proposed method: (a) human face under pattern illumination; (b) grid-point detection result; (c) recovered orientation map; (d) relative depth map in pixel units acquired without using reference points; and (e) reconstructed absolute depth map viewed from a particular perspective.

map that took advantage of the three reference points is shown in Fig. 5(e). A sphere was fitted in the least square error sense to such a depth map. The radius of the fitted sphere was 94.95 mm with a standard deviation of 1.82 mm or 0.23% with respect to the imaging distance. The reconstruction accuracy also outperforms many spatially encoded structured light systems like [22–24].

5.3. Reconstruction of generic objects

Human faces are among the most popular targets in various vision applications. Human eyes are so sensitive to them that they can tell even subtle discrepancies in their reconstruction quality. We have also used the proposed method to reconstruct a variety of human faces.

In one experiment, a human face bust was reconstructed. The result is shown in Fig. 6. Ground truth for such arbitrary object was difficult to obtain in adequate accuracy for performance evaluation. Nonetheless, visual check could still tell the quality of the reconstruction. As displayed in Fig. 6, the reconstruction quality was reasonable.

Fig. 7 shows the result of a more challenging experiment, in which a real human face, where both texture and occlusion exist, was reconstructed using the proposed method. Feature detection and surface normal determination were still successful despite the challenge, leading to a quality orientation map as presented in Fig. 7. Again, for such arbitrary shape it was difficult to acquire ground truth of the depth map for performance evaluation. Nonetheless, visual check shows that the reconstruction was close to the real subject.

6. Conclusion and future work

We have shown that even a setup as economical as that of an off-the-shelf projector and camera, and an operation as simple as that of the illumination of a checker board-like pattern and a single image capture, could achieve recovery of absolute depth map in metric units. The basis of the system lies in the possibility that surface orientations at the illuminated grid-points are decidable by image tangents in the image data deterministically. The absence of coding in the illuminated pattern inevitably incurs certain loss in the reconstruction accuracy, but the loss is not dramatic at least upon some reasonable working settings. Experiments on a variety of objects show that even with such simplicity of the system the reconstruction could be of a promising precision. Future work will address how the use of a finer checker-board pattern could boost the reconstruction precision and 3D data density further.

Acknowledgments

The work described in this article was supported partially by the grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4195/04E), and partially by the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant no. KGCX2-YW-156).

References

[1] B.K.P. Horn, Height and gradient from shading, International Journal of Computer Vision 5 (1) (1990) 37–75.
[2] J.Y. Chang, K.M. Lee, S.U. Lee, Shape from shading using graph cuts, Pattern Recognition 41 (12) (2008) 3749–3757.
[3] J. Garding, Direct estimation of shape from texture, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (11) (1993) 1202–1208.
[4] L. Reyes, G. Medioni, E. Bayro, Registration of 3D points using geometric algebra and tensor voting, International Journal of Computer Vision 75 (3) (2007) 351–369.
[5] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Three-dimensional face recognition, International Journal of Computer Vision 64 (1) (2005) 5–30.
[6] M. Christopher, B. Benjamin, A similarity-based aspect-graph approach to 3D object recognition, International Journal of Computer Vision 57 (1) (2004) 5–22.
[7] D. Miyazaki, M. Saito, Y. Sato, K. Ikeuchi, Determining surface orientations of transparent objects based on polarization degrees in visible and infrared wavelengths, Journal of the Optical Society of America A 19 (2002) 687–694.
[8] G. McGunnigle, M.J. Chantler, Resolving handwriting from background printing using photometric stereo, Pattern Recognition 36 (8) (2003) 1869–1879.
[9] C.J. Taylor, Surface reconstruction from feature based stereo, in: International Conference on Computer Vision, 2003, pp. 184–190.
[10] D. Scharstein, R. Szeliski, Stereo matching with nonlinear diffusion, International Journal of Computer Vision 28 (2) (1998) 155–174.
[11] J.S. Ku, K.M. Lee, S.U. Lee, Multi-image matching for a general motion stereo camera model, Pattern Recognition 34 (9) (2001) 1701–1712.
[12] J. Salvi, J. Pages, J. Batlle, Pattern codification strategies in structured light systems, Pattern Recognition 37 (4) (2004) 827–849.
[13] J. Gühring, Dense 3-D surface acquisition by structured light using off-the-shelf components, Proceedings of the SPIE 4309 (2000) 220–231.
[14] Z. Song, R. Chung, Off-the-shelf structured light-based system for accurate 3D reconstruction, HKIE Transactions 15 (4) (2008) 44–51.
[15] L. Zhang, B. Curless, S. Seitz, Rapid shape acquisition using color structured light and multi-pass dynamic programming, 3DPVT (2002) 22–36.
[16] J. Pages, J. Salvi, J. Forest, Optimized de-Bruijn patterns for one-shot shape acquisition, Image and Vision Computing 23 (8) (2005) 707–720.
[17] A. Dipanda, S. Woo, Towards a real-time 3D shape reconstruction using a structured light system, Pattern Recognition 38 (10) (2005) 1632–1650.
[18] H. Kawasaki, R. Furukawa, Y. Nakamura, 3D acquisition system using uncalibrated line-laser projector, International Conference on Pattern Recognition 1 (2006) 1071–1075.
[19] M. Demeyere, D. Rurimunzu, C. Eugène, Diameter measurement of spherical objects by laser triangulation in an ambulatory context, IEEE Transactions on Instrumentation and Measurement 56 (3) (2007) 867–872.
[20] Y.C. Hsieh, Decoding structured light patterns for three-dimensional imaging systems, Pattern Recognition 34 (2) (2001) 343–349.
[21] Z. Song, R. Chung, Determining both surface position and orientation in structured-light based sensing, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 〈http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.192〉.
[22] A. Adan, F. Molina, 3D feature tracking using a dynamic structured light system, in: 2nd Canadian Conference on Computer and Robot Vision, 2005, pp. 168–175.
[23] D. Desjardins, P. Payeur, Dense stereo range sensing with marching pseudorandom patterns, in: Canadian Conference on Computer and Robot Vision, 2007, pp. 216–226.
[24] S.Y. Chen, Y.F. Li, Vision processing for realtime 3-D data acquisition based on coded structured light, IEEE Transactions on Image Processing 17 (2) (2008) 167–176.
[25] N. Shrikhande, G. Stockman, Surface orientation from a projected grid, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (5) (1988) 749–754.
[26] K. Sugihara, K. Okazaki, F. Kaihua, N. Sugie, Regular pattern projection for surface measurement, in: 2nd International Symposium on Robotics Research, 1984, pp. 17–24.
[27] C.J. Davies, M.S. Nixon, A Hough transform for detecting the location and orientation of three-dimensional surfaces via color encoded spots, IEEE Transactions on Systems, Man and Cybernetics – Part B 28 (1) (1998) 90–95.
[28] P. Griffin, L. Narasimhan, S. Yee, Generation of uniquely encoded light patterns for range data acquisition, Pattern Recognition 25 (6) (1992) 609–616.
[29] S. Winkelbach, F.M. Wahl, Shape from 2D edge gradients, in: Proceedings of the 23rd DAGM-Symposium on Pattern Recognition, vol. 2191, 2001, pp. 377–384.
[30] S. Winkelbach, F.M. Wahl, Shape from single stripe pattern illumination, in: Proceedings of the 24th DAGM Symposium on Pattern Recognition, vol. 2449, 2002, pp. 240–247.
[31] Z. Song, R. Chung, Use of LCD panel for calibrating structured light-based range sensing system, IEEE Transactions on Instrumentation and Measurement 57 (11) (2008) 2623–2630.
[32] D. Marr, E. Hildreth, Theory of edge detection, Proceedings of the Royal Society of London 207 (1980) 187–217.
[33] C. Harris, M.J. Stephens, A combined corner and edge detector, in: 4th Alvey Vision Conference, 1998, pp. 147–152.
[34] S.M. Smith, M. Brady, SUSAN—a new approach to low level image processing, International Journal of Computer Vision 23 (1) (1997) 45–78.

[35] Z. Song, R. Chung, Grid-point extraction exploiting point symmetry in a pseudorandom color pattern, in: International Conference on Image Processing, 2008, pp. 1956–1959.

[36] R.T. Frankot, R. Chellappa, A method of enforcing integrability in shape from shading algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (4) (1988) 439–451.

About the Author—ZHAN SONG was born in shandong, China, in 1978. He received Ph.D. in Mechanical and Automation Engineering from the Chinese University of Hong Kong, Hong Kong, in 2008. He is currently with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), as an assistant researcher. His current research interests include structured-light based sensing, image processing, 3-D face recognition, and human–computer interaction.

About the Author—CHI-KIT RONALD CHUNG received BSEE from the University of Hong Kong, Hong Kong, and Ph.D. in computer engineering from University of Southern California, Los Angeles. He is currently with the Chinese University of Hong Kong, Hong Kong, as Director of the Computer Vision Laboratory and Professor in the Department of Mechanical and Automation Engineering. His research interests include computer vision and robotics. He is a senior member of IEEE and a member of MENSA. He was the Chairman of the IEEE Hong Kong Section Joint Chapter on Robotics & Automation Society and Control Systems Society in the years 2001–2003.