Reconstruction of Cultural Artifact Using Structured Lighting with Densified Stereo Correspondence

H.-J. Chien[1], C.-Y. Chen[1], and C.-F. Chen[2]

[1] Computer Science & Information Engineering, National University of Kaohsiung, Taiwan
[2] Electrical Engineering, I-Shou University, Taiwan
ayen@nuk.edu.tw

Abstract. In recent years a variety of 3D scanning technologies has been applied to the field of Heritage Preserve. The paper describes an operational 3D scanning system that has been set up at Kaohsiung Museum of History for digital archival of cultural artifacts. Also, an approach that refines resolution of stereo correspondence to improve model quality obtained by structured lighting is proposed. The reconstructed models are shown to evaluate performance of improved scanner.

Keywords: 3D reconstruction, structured lighting, stereo correspondence, digital archival, heritage preservation.

1 Introduction

Computer vision based shape recovery techniques have been widely applied to preservation of cultural heritages in recent years [1][2][3][4]. These techniques are able to reconstruct 3D models in an accurate, non-tactile, and non-destructive manner without involving high energy projection, making them the suitable for digitalization of fragile objects. Structured lighting is a popular approach because it works on most surface materials, from perfect diffuse to specular and even under water, as long as the camera can sense change of intensity due to light projection. Furthermore, it does not rely on textures characteristics, thus it can be applied to featureless surface.

In a typical stereo vision setup, two or more cameras are placed to look toward the same object. The relationship between pixels associated with same 3D coordinates in different frames is known as the *stereo correspondence* between the images [5], which may be difficult to solve. Computation intensive techniques like block matching or epipolar analysis are required, and in the worst case only degenerated correspondence is available.

Structured lighting simplifies the correspondence problem by using an active projection device to replace one of two cameras in stereo vision [4], [6]. The accuracy of structured lighting is directly influenced by the accuracy of lens parameter and stereo correspondence. Theoretically, the range measurement can be quite accurate if the cameras are well-calibrated and the projector's correspondence is precisely determined. However in real applications, many practical issues exist. One of the most noticeable constraints is the heterogeneity of image dimension. An off-the-shelf projector is often capable to project a screen of up to 1024 by 768 pixels. However, the resolution of an

F. Huang and R.-C. Wang (Eds.): ArtsIT 2009, LNICST 30, pp. 239–246, 2010.

image captured by a typical digital camera could be at least two times the projector's resolution. We have observed that such a mismatch in resolution causes recovered surface to have a saw-toothed profile, as will be discussed later.

This paper describes an operational 3D scanning system that has been set up at Kaohsiung Museum of History for digital archival of cultural artifacts as well as an approach that improves model quality obtained by structured lighting. The rest of the paper is organized as follows. In section 2 an overview to our reconstruction system is stated, and the subsequent section 3 focuses on the establishment, evaluation and improvement of stereo correspondence. The performance of our proposal applied to real cases is studied in section 4, and finally the conclusion is given in section 5.

2 System Setup

This section describes how the various components in the system are placed to triangulate the 3D coordinates, as well as the calibration method adopted.

2.1 Physical Setup

The basic setup of the proposed structured light system consists of one camera and one projector placed in front of the object to be reconstructed. To acquire a complete 3D model the object is placed on a turntable. The position of camera is adjusted so that the rotation axis of the turntable is as close as possible to the camera's XZ plane. In this way, we can obtain not only meshes from different views but also volumetric model by the *Shape from Contour* technique [7]. The intersection of the camera and the projector's fields of view decides the system's effective view. Therefore the projector is arranged to make sure that space covered in the effective view is large enough to fit in the cultural artifact from all viewing directions. The setup of the system is shown in Fig. 1.

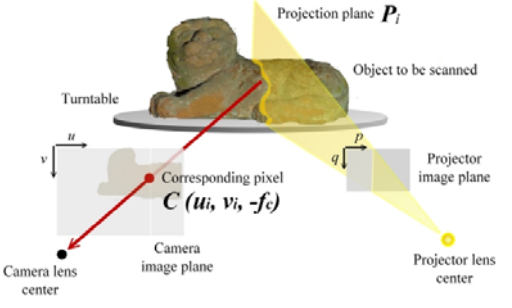


Fig. 1. 3D relationships between the camera, data projector and the object

The distance from the camera's optical center to object surface can be measured from the lens parameters and the correspondence. Assume these two prerequisites are available, the triangulation algorithm gives depth of pixel (u_i,v_i) by finding intersect

point of the back-projected ray L_i through the camera's center and the corresponding vertical projection plane P_i. In equation (1) and (2) the ray in camera-centered space and the plane in projector-centered space are defined respectively regarding to camera focal length f_c and angle between projection plane and optics axis of projector α_i. Note the coordinates (u_i,v_i) are shifted to match center of image plane.

$$L_i = s\left(\frac{\hat{u}_i}{f_c}, \frac{\hat{v}_i}{f_c}, -1\right), s \in \Re \tag{1}$$

$$P_i = \left(p\cos\alpha_i, q, -p\sin\alpha_i\right), \quad p, q \in \Re \tag{2}$$

We obtain the transform from camera space to projector space from the extrinsic parameters . Let a 3 by 3 rotation matrix R and a 3 by 1 translation vector T denote that transform, equations (1) and (2) form a linear system as shown in equation (3).

$$\begin{pmatrix} p\cos\alpha_i \\ q \\ -p\sin\alpha_i \end{pmatrix} = s \cdot R \begin{pmatrix} \hat{u}_i / f_c \\ \hat{v}_i / f_c \\ -1 \end{pmatrix} + T \tag{3}$$

Although the linear system can be solved by computing its inverse matrix, dividing third row by first yields a straightforward solution of s, as shown by equation (4).

$$s = -\frac{T_x \tan(\alpha_i) + T_z}{\left[R_x \tan(\alpha_i) + R_z\right] \cdot \left(\frac{\hat{u}_i}{f_c}, \frac{\hat{v}_i}{f_c}, -1\right)} \tag{4}$$

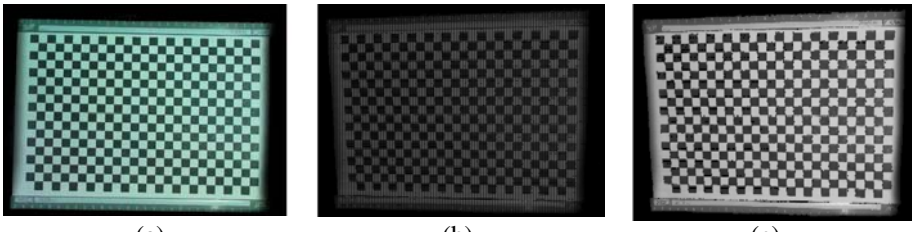
2.2 Lens Calibration

The main task of calibration is to find intrinsic and extrinsic parameters of the len, and the correctness of calibration directly affects system accuracy. Once parameters of both camera and projector are acquired, the linkage between two devices can be constructed for triangulation. Typically a projector is treated as an inversed pinhole device as it uses backward perspective projection [8][9].

There are several mature and practical methods for camera calibration [10]. Our system is calibrated by a variation of Tsai's method. A checkerboard pattern is used as a reference object for Tsai's method. To start the calibration, the board is placed in the scene and an image is captured by the camera. The corners are extracted and used to compute the homography transform. Rectified control points are returned to facilitate the labeling procedure. By solving the constructed over-determined linear system all sixteen initial parameters are established. We then apply the Levenberg-Marquardt method for non-linear optimisation to minimize re-projection error. To preserve orthogonality and normality of rotation matrix, we augmented the objective function by adding significantly weighted penalty terms, since these two characteristics are important and should not be neglected during optimisation.

Calibration for the projector is similar, with the exception of being unable to capture images. Therefore, we need to find the correspondence first. Once found, the mapping is reversed to generate a partial view from the projector, since many pixels

may map to same pixel location in the projector's image plane. Image enhancement and interpolation may be required to generate a complete view, as shown in Fig. 2. The calibrated parameters are shown in table 1. Despite the degeneracy of using planar calibration object instead of volumetric objects (e.g. a cube), multiple calibrations yield consistent result.



(a) (b) (c)

Fig. 2. (a) the camera's view of a reference checkerboard, (b) the projector's view of the same scene, obtained from correspondence, and (c) after image enhancements. Note (b) looks dark due to loss of noninvertible information caused by sparse correspondence.

Table 1. Calibrated parameters of camera and projector lens

	Camera			**Projector**		
Focal length	3566 pixels			2449 pixels		
Principle	(969.03, 643.93)			(509.29, 384.61)		
Pixel aspect ratio	1.0013			1.0086		
Translation	(12.50, 4.26, -86.87)			(18.22, 1.59, -126.04)		
Rotation matrix	-0.99	0.00	-0.12	-0.99	0.02	-0.11
	-0.01	-1.00	0.05	-0.02	-1.01	0.02
	-0.12	0.04	0.99	0.16	0.03	0.99

3 Stereo Correspondence

Stereo correspondence is a prerequisite to perform 3D measurement for all shape recovery methods based on multiple view geometry [5]. A correspondence states how pixels are related in two image planes. In practical applications, such pixel-wise relationships are too discrete and not dense enough to produce smooth surfaces. This section focuses on an approach that reliably improves density of correspondence.

3.1 Establishing Correspondence

Comparing to passive vision the structured lighting technique obtains the correspondence in relatively straightforward way. Gray-coded pattern projection has been considered a robust and fast technique to establish pixel-wise mapping between image plane of the camera and of the projector [7]. Since it uses a time-multiplexed binary coding scheme, a sequence of light patterns which consist of bi-colour stripes is projected onto the scene over time. The one-dimensional coordinate of each unit strip can be uniquely identified from decoded binary images.

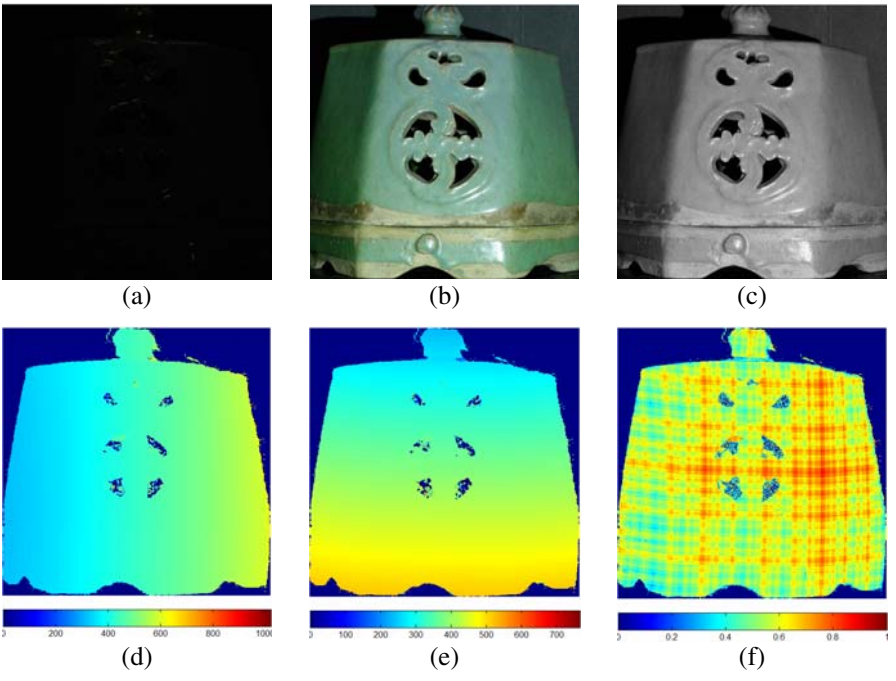


Fig. 3. First row: images of scene illuminated by uniform (a) black and (b) white patterns used to define the interval and (c) the denominator of intensity ratio R_k. Second row: (d) and (e) show stereo correspondences in x- and y-coordinates respectively, (f) shows the confidence for pixel correspondence.

Reflectance-based adaptive binarization is a commonly used method to decide whether a pixel is illuminated by a certain colour of a light pattern. Non-uniform thresholds are computed according to per-pixel intensity ratio given in equation (5). Two additional patterns are required to define range [I_{min} I_{max}]. Fig. 3(a-c) shows an artifact lit by patterns of different colour (black and white) and resultant intensity range.

$$R_k(u,v) = \frac{I_k(u,v) - I_{\min}(u,v)}{I_{\max}(u,v) - I_{\min}(u,v)} \tag{5}$$

The number of patterns needed to establish dense correspondence is logarithmically bounded by number of unit strips. For a projector that has 1024 pixels in the horizontal direction, ten projections of light pattern are required to label all 1024 unit strips. However in practice, dense correspondence is often not feasible due to limits of heterogeneous image dimensions, placement of devices, boundary uncertainty, light interference, noises and difficulty in extraction of frequently interleaved stripe. Fig. 3(d-f) visualizes established vertical and horizontal 512-stripe correspondences, both increases monotonically due to strict ordering property.

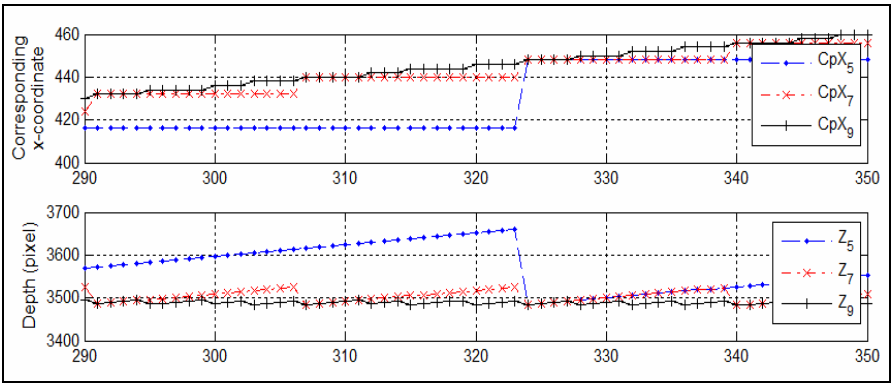


Fig. 4. Upper plot displays x-coordinate correspondences obtained using 5, 7, and 9 Gray-coded patterns respectively. The lower are triangulated depths with respect to each correspondence.

Although the visualized results may look smooth, the correspondences are discretely stepping, causing sudden changes in recovered surfaces, as shown by the profile in Fig. 4. By reducing number of unit stripes the saw-toothed effect becomes more significant. Our approach to solve the problem is to adjust the correspondence in each stepping interval such that the correspondence grows smoothly. An intuitive approach would be to use linear interpolation for correspondence adjustment.

3.2 Quantitative Confidence

A dual-peaked function is used to evaluate the reliability of decoded information. In particular, the value of this function depends on intensity ratio and is peaked at zero and one. As the ratio moves away from these peaks the confidence indicator decays gradually and approaches zero where the ratio becomes insignificant.

The modeling of the confidence indicator is reasonable as each surface patch is assumed to behave as being illuminated by either a white or a black stripe. Factors such as light inter-reflection and noise may cause the ratio to shift away from expected position, hence a tolerance is given. We model the function by summing a zero-mean and a one-mean Gaussian function. The deviations are adjustable to control the tolerable uncertainty. Fig. 3(c) shows the evaluated pixel confidence.

3.3 Correspondence Adjustment

More flexible adjustment on correspondence can be achieved by taking the quantitative confidence into account. The idea is to assign new values to unsure elements by referencing their reliable neighbors. Such model is actualized by equation (6), where f is the original correspondence, w is the confidence, d is a function for spatial proximity, and g is the new correspondence. In our implementation, d is a radial basis function such that the correspondence can be adjusted using convolution.

$$g(u,v) = \frac{\sum_{i=0}^{n}\sum_{j=0}^{m} d_{u,v}(i,j)\cdot w(i,j)\cdot f(i,j)}{\sum_{i=0}^{n}\sum_{j=0}^{m} d_{u,v}(i,j)\cdot w(i,j)} \tag{6}$$

4 Experiment

We have implemented the proposed approach and applied it several objects. We choose the cap of the cultural object (in Fig. 3) to demonstrate the quality of the reconstructed model. The system has a projector with 512 vertical stripes. Therefore each stripe occupied averagely 800/512 (1.5625) pixel width in the projector's image plane. The 512-level correspondence is further expanded to sub-pixel accuracy using the proposed method. Fig. 5 shows primordial-decoded, convoluted, as well as spline interpolated correspondences of a profile. The recovered surface using correspondence in sub-stripe resolution is shown in Fig. 6(a). In contrast, Fig. 6(b) looks rugged due to sparse correspondence.

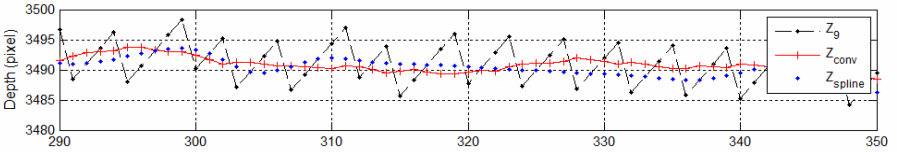
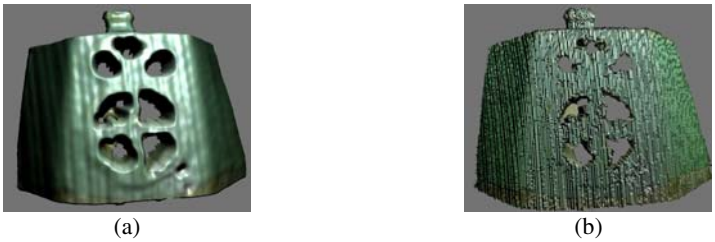


Fig. 5. Comparison between original, convoluted, and interpolated correspondences. It shows a significant reduction of depth ramping. The interpolation doesn't take confidence indicator into account, may result in loss of shape features.



(a) (b)

Fig. 6. (a) shows the surface recovered from densified correspondence. Comparing to (b) the reconstruction using original correspondence, the saw-toothed effect is quite unobvious.

5 Conclusion and Future Work

This paper proposes an approach to improve structured light system accuracy. Results have been provided to demonstrate the improvement of the proposed approach. By applying confidence-based convolution on decoded correspondence the resolution is increased. As the result, the saw-toothed phenomenon is removed from surface. Our

method does not require smoothing and down sampling operation as a conventional structured lighting scanner may do.

Future research will focus on further improvement of the 3D scanning system. As observed during experiments, the calibration errors of projector's lens are much significant than camera's. Therefore, it is reasonable to design a projector-specific calibration method. Also, the approach of not calculating the parameters explicitly but reconstructing the projector from calibrated camera with established correspondence is considered for future work.

Acknowledgments. This research is sponsored by National Digital Archives Program, grant number NSC 97-2631-H-390 -001. Cultural artifacts, working place, and permission to access restricted warehouse are kindly provided by Kaohsiung Museum of History.

References

1. Guidi, G., Beraldin, J.A., Atzeni, C.: High-accuracy 3D modeling of cultural heritage: the digitizing of Donatello's Maddalena. 13th IEEE Transactions on image processing 13(3), 370–380 (2004)
2. Anestis, K., Fotis, A., Christodoulos, C.: On 3D reconstruction of the old city of Xanthi A minimum budget approach to virtual touring based on photogrammetry. Journal of Cultural Heritage 8(1), 26–31 (2007)
3. 3D Modelling of Castles, http://www.gim-international.com/issues/articles/id1073-D_Modelling_of_Castles.html
4. Rocchini, C., Cignoni, P., Montani, C., Pingi, P., Scopigno, R.: A Low Cost 3D Scanner based on Structured Light. In: Eurographics 2001, vol. 20(3), pp. 299–308 (2001)
5. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
6. Salvi, J., Pags, J., Batlle, J.: Codification Strategies in Structured Light Systems. J. PR 37, 827–849 (2004)
7. Chen, C.-Y., Klette, R., Chen, C.-F.: Shape from Photometric Stereo and Contours. Springer, Berlin (2003)
8. Batlle, J., Mouaddib, E., Salvi, J.: Recent Progress in Coded Structured Light as a Technique to Solve the Correspondence Problem: A Survey. J. PR 31, 963–982 (1998)
9. Gühring, J.: Dense 3-D Surface Acquisition by Structured Light using Off-The-Shelf Components. In: SPIE Photonics West, Videometrics VII, vol. 4309, pp. 22–23 (2001)
10. Klette, R., Schlüns, K., Koschan, A.: Computer Vision - Three-dimensional Data from Images. Springer, Singapore (1998)