Projector-Camera Based System for Fast Object Modeling

Guanghui Wang[a,b], Zhanyi Hu[a], Fuchao Wu[a], Hung-Tat Tsui[b]

a. National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing 100080, P. R. China
b. Department of Electronic Engineering, The Chinese University of Hong Kong
Shatin, N.T. Hong Kong
ghwang@ee.cuhk.edu.hk, huzy@nlpr.ia.ac.cn

Abstract

In this paper, a practical projector-camera based system for fast object modeling and reconstruction is reported. The system consists of an off-the-shelf digital projector, a video camera and a turntable. The system calibration is implemented accurately and automatically by taking a single image of an illuminated calibration template. The multiple color light stripes emitted by the projector adopt only three colors and the color can be modified easily according to the surface texture of the object. This makes the system widely applicable in many situations, and the assumption about color neutrality, which is a limitation to other similar systems, can be largely relaxed. The registration and merging algorithm is greatly simplified by employing a turntable. The validity and high accuracy of the method are demonstrated by extensive real experiments.

1. Introduction

3D reconstruction and modeling of objects is one of the main objectives of computer vision and attracts a lot of attentions in recent years. Examples and applications of this task include robot navigation and obstacle recognition, augmented virtual reality, industrial inspection, reverse engineering, medical diagnosis and many more. Among various modeling and reconstruction techniques, the projector-camera based method (also called structured light) is widely utilized because of its simplicity and high accuracy.

Typically, a projector-camera system consists of a camera and a projector. The projector projects an illuminated light stripes or light patterns on the object surface. The camera captures images of the illuminated object, then the 3D information of the visible surfaces can be recovered by using the calibrated parameters of the system via triangulations. Various light patterns are adopted by previous researchers, such as light strips, light grid/net, light beams/dots, etc. The more complex the light patterns are, the more difficult the processing of the images will be. Most of the existing systems adopt strip pattern, since it is easy to recognize, sample and model [1][2]. Many projector-camera based systems have been proposed in the literature [3][4][5][6]. Most commercial laser scanners use only a single stripe and sweep the stripe across the surface so as to obtain a full range image. A long time is usually required to acquire image data. In order to speed up the sampling speed, some researchers adopt multiple stripes. Such a system must take special care for the determination of stripe correspondence. There are three major ways in doing this, and a good review on the techniques is given by Batlle [1]. The first way is by surface continuity assumption so that adjacent projected stripes are adjacent in the image. This assumption holds true only if the observed surface does not have self-occlusion or disconnected components. The second way is differentiating the stripes by colors [3]. It may fail if the surface is textured rather than having only neutral colors. The third way is to code the stripes by varying their illumination over time [4][5][6]. This needs several frames to decode the light and cannot be applied in the measurement of moving object.

System calibration is an important step for a projector-camera system. The existing calibration methods for structured stripes can be broadly divided into two categories. In the first one, the system is calibrated in two stages: camera calibration and projector calibration. Usually, a standard calibration scheme is adopted to estimate the projection matrix and camera parameters. On the other hand, the projector can be calibrated in several ways. In [7], the light stripe is modeled as a light stripe plane and the calibration is to recover the parameters of the plane. McvIor [8] and Valkenburg [5] model the projector as a reverse pinhole camera with 1D image that can be defined by a 2×4 matrix. Shen and Meng [2] propose a method to calibrate all light beams projected from the digital projector. In the second category, only a 4×3 image to world conversion matrix is calibrated for each stripe. It converts image points on a light stripe into their corresponding 3D coordinates. To obtain the matrix, some world to image point correspondences should be established on each stripe. However, it is in fact very difficult to precisely locate points on the light stripe in the world system. Previous researchers tackle this problem by employing some calibration artifacts to generate the point correspondences using world line to image point [9], world plane to image point [10], and world point to image point correspondences [11]. Most of the above calibration methods assume pinhole camera model, and a high precision moving mechanism is required during calibration.

This paper mainly focuses on developing a practical and automatic system for fast object modeling. There are three main features of our system. First, the calibration of the system is accurate and can be implemented automatically from a single illuminated image. Second, the light stripes are encoded by three different colors that can be modified flexibly according to the surface texture of the object. This can, to some extent, overcome the limitation of other similar systems, which strongly depend on the color neutrality assumption of the object. Third, the registration and merging algorithm is greatly simplified by employing a turntable.

The remaining parts of this paper are organized as follows. Section 2 gives a brief description of the system configuration and characteristic. Section 3 presents a simple and robust method of system calibration and object modeling. Some experiment results and analysis are given in section 4, followed by a conclusion at the end of this paper.

2. System description

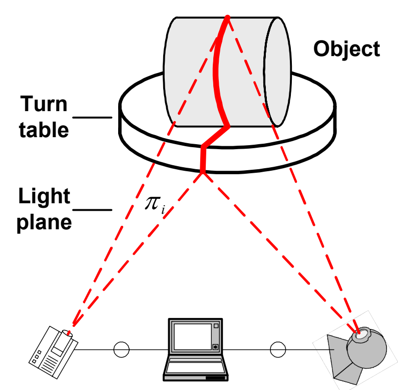

Fig.1 Schematic figure of the system configuration

The system consists of an off-the-shelf digital projector, a video camera and a turntable, as shown in Fig.1. The operation procedure is quite easy: Put the object to be modeled on the turntable, take an illuminated image in addition to a pure texture image, the visible surface can be reconstructed automatically from the illuminated image, while the texture image is used for texture mapping after reconstruction so as to form a more realistic object model. Rotate the turntable to another position and repeat the same procedure until every surface is reconstructed, then the complete model is obtained by merging the reconstructed results together. The turntable here is optional, nevertheless, the registration problem, i.e. transforming measurement from different views into one common coordinate frame, can be greatly simplified by employing the turntable. For example, if we choose the rotation axis of the turntable as the Z-axis of the world coordinate system, the measurements at different viewpoint is simply related with a rotation matrix $\boldsymbol{R}(\alpha)$, where, α is the rotation angle between two views.

$$\boldsymbol{R}(\alpha)=\begin{bmatrix}\cos\alpha, \sin\alpha, \ 0\\ -\sin\alpha, \cos\alpha, 0\\ 0, \quad 0, \quad 1\end{bmatrix}.$$

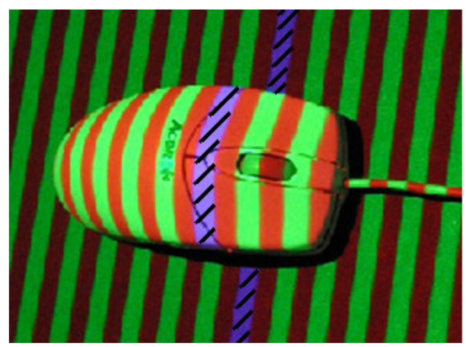

Fig.2 Object illuminated by color light stripes

The reason for adopting a video projector is that it is widely available, and can allow us to experiment different type of light patterns. The light pattern used in the system is multiple color stripes, so the measurement on visible surface of the object can be taken from a single shot, this makes the system fast and capable of modeling moving object. As shown in Fig.2, only three colors are used in the system, where, two colors project alternatively, while the third color (marked with upward diagonals in Fig.2) is only applied to the center stripe. This special stripe is taken as an index mark to solve the correspondence problem between different stripes. Obviously, this is under the assumption of space continuity and hard to deal with the scene with severe self-occlusion. However, this tradeoff makes the system computationally inexpensive and easy to process automatically, and in some cases, even the scene is discontinuous (as the case in Fig.2), the stripe correspondences can also be solved by this special stripe. Furthermore, the assumption about color neutrality, which is a limitation to other color structured light system, can be relaxed, since the three colors can be changed flexibly according to the texture property of the object to be modeled. We use the edge of each stripe to model the light plane in practice, and the edge can be detected in sub-pixel accuracy [5,12].

System calibration is another crucial step for the accuracy of measurement. The calibration is done once for a fixed system. However, when the parameters of some components, such as the focal length of projector or camera, are adjusted or the position changes, the whole system needs to be re-calibrated. So it is important to find a simple, accurate and automatic calibration method. In our system, calibration is accomplished automatically by taking one shot of a calibration template. The template is composed of two orthogonal planar surfaces, with 128 circular black fiducial marks evenly distributed on the two white surfaces. The position of each mark on the template is known with high accuracy, while the corresponding coordinates in the image is estimated in sub-pixel accuracy by computing the centroid of each mark [8,13]. Fig.3 shows a schematic figure of the template, where, only 18 fiducial marks and one light stripe edge are displayed.

3. Method of calibration and modeling

3.1. System calibration

Under perspective projection, a 3D point $\boldsymbol{x} = [x, y, z]^T$ in space is projected to an image point $\boldsymbol{m} = [u, v]^T$ via a 3×4 rank 3 projection matrix $\boldsymbol{P}$ as:

$$s\tilde{\boldsymbol{m}} = \boldsymbol{P}\tilde{\boldsymbol{x}} \tag{1}$$

where, s is an unknown nonzero scalar; $\tilde{\boldsymbol{m}}$ and $\tilde{\boldsymbol{x}}$ are the homogeneous coordinate vectors of $\boldsymbol{m}$ and $\boldsymbol{x}$ in form of $\tilde{\boldsymbol{m}} = [\boldsymbol{m}^T, 1]^T$, $\tilde{\boldsymbol{x}} = [\boldsymbol{x}^T, 1]^T$. The projection matrix has 11 degrees of freedom since it can only be defined meaningfully up to a scale factor, and each world to image correspondence can give rise to two independent linear constraints on the projection matrix as follows:

$$\boldsymbol{A}_1 \boldsymbol{p} = 0 \tag{2}$$

where, $\boldsymbol{A}_1 = \begin{bmatrix} x, y, z, 1, 0, 0, 0, 0, ux, uy, uz, u \\ 0, 0, 0, 0, x, y, z, 1, vx, vy, vz, v \end{bmatrix}$,

$\boldsymbol{p} = [p_{11}, p_{12}, p_{13}, p_{14}, p_{21}, p_{22}, p_{23}, p_{24}, p_{31}, p_{32}, p_{33}, p_{34}]^T$, $p_{i,j}$ is the (i, j) -th element of matrix $\boldsymbol{P}$. Given N ($N \geqslant 6$) point correspondences, a $2N \times 12$ matrix $\boldsymbol{A}$ is formed by stacking equation (2) for each point pair. Theoretically, $\boldsymbol{A}$ is of rank 11, and it may be of full rank (with rank 12) for non-perfect data with noise. However, $\boldsymbol{p}$ can be estimated by Singular Value Decomposition (SVD) of matrix $\boldsymbol{A}$. The solution of $\boldsymbol{p}$ is the unit singular vector corresponding to the smallest singular value. Since the precision of projection matrix is closely related to that of the whole system, we introduce a Maximum Likelihood Estimation (MLE) algorithm to optimize the solution.

$$\min_p \sum_i d^2(\tilde{\boldsymbol{m}}_i, \boldsymbol{P}\tilde{\boldsymbol{x}}_i), \ i = 1, \dots, N \tag{3}$$

The problem can be solved via an iterative method such as Gauss-Newton or Levenberg-Marquardt [14], while the previous SVD estimation is taken as the initial value of the iteration.

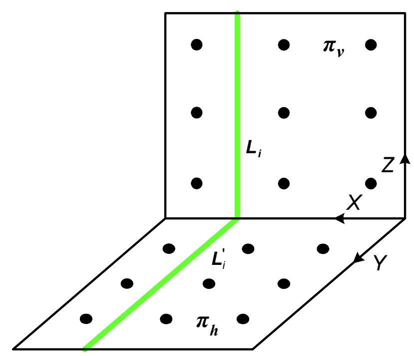

Fig.3 Schematic figure for calibration template

Next, we will present the method of light stripe calibration. Select the world coordinate system as shown in Fig.3, then the coordinates of the two orthogonal planes of the calibration template are $\boldsymbol{\pi}_v = [0,1,0,0]^T$ and $\boldsymbol{\pi}_h = [0,0,1,0]^T$. Suppose the i -th light stripe plane $\boldsymbol{\pi}_i$ intersects the plane $\boldsymbol{\pi}_v$ and $\boldsymbol{\pi}_h$ at L_i and $L_i^{'}$ respectively, their corresponding image can be extracted by edge detection in sub-pixel precision. Let $\boldsymbol{m}_{ij}$ ($j = 1, \dots, N_i$) be the detected edge points of the image of L_i, $\boldsymbol{m}_{ik}$ ($k = 1, \dots, N_i^{'}$) be that of $L_i^{'}$, then the corresponding space point $\boldsymbol{x}_{ij}$ with respect to $\boldsymbol{m}_{ij}$ can be determined by the intersection of the back-projected ray of $\boldsymbol{m}_{ij}$ with plane π_v, i.e.

$$\begin{cases} s\,\tilde{\boldsymbol{m}}_{ij} = \boldsymbol{P}\,\tilde{\boldsymbol{x}}_{ij} \\ \boldsymbol{\pi}_v^T \tilde{\boldsymbol{x}}_{ij} = 0 \end{cases} \tag{4}$$

It is easy to compute the coordinates of $\boldsymbol{x}_{ij}$ from equation (4), and so does the coordinates of point $\boldsymbol{x}_{ik}$ corresponding to $\boldsymbol{m}_{ik}$. Since points $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{ik}$ are nonlinear and lie on the same optical plane $\boldsymbol{\pi}_i$, so we have:

$$\begin{bmatrix} \tilde{\boldsymbol{x}}_{ij}^T \\ \tilde{\boldsymbol{x}}_{ik}^T \end{bmatrix} \boldsymbol{\pi}_i = 0 \,,\ j = 1,\ldots,N_i \,,\ k = 1,\ldots,N_i^{'}$$

rewrite the equation into a linear system as:

$$\boldsymbol{A}\boldsymbol{\pi}_i = 0 \tag{5}$$

where, $\boldsymbol{A}$ is a matrix of dimension $(N_i + N_i^{'}) \times 4$ formed by stacking all the homogeneous vectors of $\tilde{\boldsymbol{x}}_{ij}$ and $\tilde{\boldsymbol{x}}_{ik}$. Thus, the coordinates of plane $\boldsymbol{\pi}_i$ is computed by SVD decomposition, and a more faithful estimation can be obtained via some nonlinear iterative algorithm by minimizing the following cost function:

$$f(\boldsymbol{\pi}_i) = \sum_j d^2(\boldsymbol{x}_{ij}, \boldsymbol{\pi}_i) + \sum_k d^2(\boldsymbol{x}_{ik}, \boldsymbol{\pi}_i) \tag{6}$$

In the same way, we can calibrate all the light stripe planes illuminated on the template.

3.2 Object modeling

The computation of the 3D model of the object is straightforward once the system is calibrated. The algorithm can be summarized as follows.

i). Extract the edge points of each stripe from the image of the illuminated object;
ii). Establish the correspondences between the light planes and the extracted edges by virtue of the special stripe;
iii). 3D reconstruction by computing the intersection of the back-projected ray of each edge point with its corresponding light plane;
iv). Obtain the complete model by merging the reconstruction results of each side of the object.

4. Test results

The projector used in our system is Sony VPL-PX11, with a resolution of 1024×768; the digital camera is Nikon Coolpix 990 with an USB link, whose resolution is set to 2048×1536; the turntable is just a manually operated one. The working volume depends on the focusing capability and the field of camera's view, usually, the object is placed at 100~150mm from the system. During the experiment, totally 128 light stripes are projected, among which only 112 stripes are calibrated, since other stripes fall outside the calibration template.

After calibration, we reconstruct the template according to the detected edges of each stripe, as shown in Fig.4, a total of 80426 points are reconstructed, where, 40519 points belong to the plane $\boldsymbol{\pi}_v$ of the template, and other points belong to $\boldsymbol{\pi}_h$. We also calculate the distance of each reconstructed point to the corresponding plane, the average distance to $\boldsymbol{\pi}_v$ is -3.3×10^{-5} mm, standard deviation is 0.153, while the average distance and standard deviation to $\boldsymbol{\pi}_h$ are -3.6×10^{-5} mm and 0.198 respectively. The results indicate that the reconstructed model is of higher accuracy.

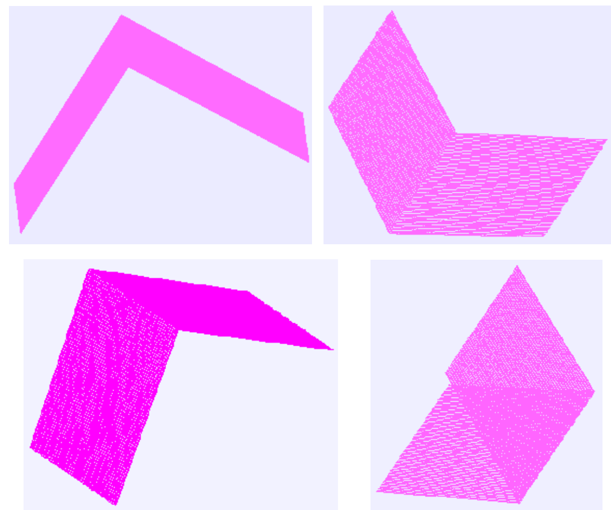

Fig.4 The reconstructed dense 3D points of the calibration template under different viewpoint

Fig.5 shows a reconstruction result of a volleyball, where, the upper left figure is a texture image taken by the camera, the upper right one is the reconstructed 3D mesh, while the lower two figures are the reconstructed surface with texture mapping. We fit the reconstructed points into a ball equation via a least squares technique as $(x-8.2)^2 + (y-24.6)^2 + (z-102.5)^2 = 104.1^2$, this result is very close to the ground truth that the radius of the volleyball is 103.8mm. Note that only a single image is used here, thus the equation is estimated only from the points on the visible part of the surface, and the true value is taken manually on the spot.

Fig.5 The texture image of a volleyball (upper left), reconstructed 3D mesh (upper right) and the reconstruction result with texture mapping (lower two)

Fig.6 Three pairs of statue images with texture and illuminated stripes

Fig.7 Visible surface reconstruction corresponding to Fig.6

Fig.8 Complete model of plaster statue in VRML format (left two) and 3D points (right)

We also use the system to model a plaster statue. Put the statue on the turntable and rotate the table to four different positions, a texture image and an image illuminated by light stripes are taken at each position, so we have four pairs of images altogether. Fig.6 shows three pairs of the images. Fig.7 gives the corresponding reconstructed 3D points and the results with texture mapping. Fig.8 shows the complete statue model by merging four side results, a total of 288504 points are reconstructed for the complete model, but only one third of the points are shown in Fig.8. The reconstructed model is satisfactory and looks similar to the original one. However, it is very difficult to obtain a representative ground truth for the model due to its curved shape. Besides, some small parts cannot be reconstructed by the system due to occlusion.

5. Conclusions

We reported a practical modeling and reconstruction system based on color light stripes. The system calibration is carried out automatically by taking a single image of a calibration template. The structured light stripes have only three colors and the color can be modified easily according to the surface texture of the object. This makes the method widely applicable in many situations. The registration and merging algorithm is greatly simplified by employing a turntable. The validity and high accuracy of the method are demonstrated by extensive real experiments.

Generally speaking, the accuracy of the system depends on several factors: the resolution and optical quality of the projector and camera, the baseline length between the projector and camera with respect to the object, the reflective properties of object surface, etc. In practice, a tradeoff has to be made between the accuracy and complexity of data processing. The process of calibration and reconstruction of our system is basically automatic. Thus it is very easy to operate. However, it is difficult for the system to deal with complicated object with severe self-occlusions at present stage. Another thing to be noted is that lens distortion is not considered here since the distortion of the camera and projector used in our system is negligible. However, it must be taken into account during system calibration if the lens distortion does play a significant role in system error.

Finally, some of the test results and reconstructed 3D models can be downloaded from the author's homepage: http://www.ee.cuhk.edu.hk/~ghwang/Projects.html.

Acknowledgements: The work is partially supported by the National Natural Science Foundation of China under grant No. 60175009, the National High Technology R&D Program under grant No. 2002AA135110 and the Hong Kong RGC Grant CUHK 4378/02E.

References

[1]. J. Batlle, E. Mouaddib and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," *Pattern Recognition*, Vol.31(7), 963-982, (1998).

[2]. T. S. Shen and C. H. Menq, "Digital projector calibration for 3-D active vision systems," *Journal of Manufacturing Science and Engineering*, Vol.124(2), 126-134, (2002).

[3]. C. Rocchini, P. Cignoni, C. Montani, P. Pingi and R. Scopigno, "A low cost 3D scanner based on structured light," *Computer Graphics forum*. Vol.20(3), (2001).

[4]. S. Rusinkiewicz, O. Hall-Holt, M. Levoy. "Real-time 3D model acquisition," In: *Proc. Siggraph 2002*, pp. 438-446, (2002).

[5]. R. J. Valkenburg and A. M. McIvor, "Accurate 3D measurement using a structured Light System," *Image and Vision Computing*, Vol.16(2), 99-110, (1998).

[6]. J. Guehring, "Dense 3-D surface acquisition by structured light using off-the-shelf components," In: *Proceedings of SPIE*, Vol. 4309, pp. 220-231, (2000).

[7]. T. Stahs and F. Wahl, "Fast and Versatile Range Data Acquisition in a Robot Work Cell," In. *Proc. IEEE International Conference on Intelligent Robots and Systems*, Raleigh, N.C., pp. 1169-1174, (1992).

[8]. A. M. McIvor, "Calibration of a laser stripe profiler," In: Proc. *Second International Conference on 3-D Imaging and Modeling*, Ottawa, Canada, pp. 92-98, (1999).

[9]. C. H. Chen and A. C. Kak, "Modeling and Calibration of a Structured Light Scanner for 3-D Robot Vision," In: Proc. *IEEE International Conference on Robotics and Automation*, Raleigh, NC, USA, pp. 807-815, (1987).

[10]. I. D. Reid, "Projective Calibration of a Laser-stripe Range Finder," *Image and Vision Computing*, Vol. 14(9), 659-666, (1996).

[11]. D. Q. Huynh, R. A. Owens and P. E. Hartmann, "Calibrating a Structured Light Stripe System: A Novel Approach," *International Journal of Computer Vision*, Vol. 33(1), 73-86, (1999).

[12]. M. Trobina, "Error model of a coded-light range sensor," *Technical Report BIWI-TR164*, ETH Zurich, Switzerland, (1995).

[13]. M.A.G. Izquierdo, M.T. Sanchez, A. Ibanez and L.G. Ullate, "Sub-pixel measurement of 3D surfaces by laser scanning," *Sensors and Actuators A: Physical*, Vol.76(1-3), 1-8, (1999).

[14]. R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, (2000).