

A high-resolution and high-speed 3D imaging system and its application on ATR

Thomas T. Lu¹ and Tien-Hsin Chao

Jet Propulsion Laboratory/California Institute of Technology, 4800 Oak Grove Dr. M/S 303-300,
Pasadena CA 91109

ABSTRACT

The paper presents an advanced 3D imaging system based on a combination of stereo vision and light projection methods. A single digital camera is used to take only one shot of the object and reconstruct the 3D model of an object. The stereo vision is achieved by employing a prism and mirror setup to split the views and combine them side by side in the camera. The advantage of this setup is its simple system architecture, easy synchronization, fast 3D imaging speed and high accuracy. The 3D imaging algorithms and potential applications are discussed. For ATR applications, it is critically important to extract maximum information for the potential targets and to separate the targets from the background and clutter noise. The added dimension of a 3D model provides additional features of surface profile, range information of the target. It is capable of removing the false shadow from camouflage and reveal the 3D profile of the object. It also provides arbitrary viewing angles and distances for training the filter bank for invariant ATR. The system architecture can be scaled to take large objects and to perform area 3D modeling onboard a UAV.

Keywords: Stereo vision, color grating projection, 3D imaging, modeling, ATR.

1. INTRODUCTION

We live in a three dimensional (3D) world. But most data we record and display are still in 1D or 2D forms. The 3rd dimension, the depth information is often lost. In many cases, the depth data is critical to the measurement of the object surface shape and profile, the perception of the object distance, and it is important to aid the human judgments. In recent years the advancement of computer technology has made tremendous progress in measuring, modeling, and visualizing 3D objects. This evolution causes an important demand for more complex and realistic 3D models. The problem is that synthesizing realistic models is difficult and time-consuming, and thus expensive. Researchers have been investigating methods to acquire 3D information from objects and scenes for many years¹⁻². There are several techniques for 3D measurements.

1.1 Laser scanning techniques

The laser scanning technique uses a narrow laser beam or point to scan the object surface. A video camera captures the deformed laser lines or dots reflected from the object surface. The computer calculates the displacement of the laser line from the object surface against a straight line. Then the surface shape can be plotted³. The laser scanning method is a robust and accurate 3D modeling technique. It is widely used in industrial inspection and rapid prototyping. However, it is generally slow and not suitable for live or fast moving objects.

1.2 Light projection techniques

The phase shift, Moiré, color or variable grating projection methods, etc. belong to this category^{4,5}. The methods have grid patterns projected on the object, extract the grid or phase deformation from the reflected image and deduce the surface profile of the object. They are generally faster than the laser scanning method, but the accuracy suffers when the surface condition is complex (variable texture color, illumination, or surface discontinuity, etc.)

1.3 Stereo vision

This method is similar to its biological counterpart - human vision. Two cameras focus on the object from two viewing angles⁶⁻⁸. The displacement of the object features in the two images represents the depth changes of the object.

¹ e-mail: Thomas.T.Lu@jpl.nasa.gov , Tel: (818) 354-9513, Fax: (818) 393-4272

Therefore, the object surface can be modeled according to the disparity of the two images. Many stereo vision techniques have been developed that synchronizes multiple cameras to acquire three-dimensional models of objects^{9,10}. One of the stereo vision methods, the photogrammetry¹¹ has been dealing with the extraction of high accuracy measurements from images. These techniques mostly require precise calibration and there is almost no automation. The detailed acquisition of models is therefore time-consuming. The stereo vision techniques have been used widely in the robotic vision applications. One can build a 3D area surface model from a sequence of 2D images taken by digital and video cameras. The user acquires the images from the high resolution cameras. Neither the camera motion nor the camera settings have to be known. The obtained 3D model is an accurate model of the ground surface, and the surface albedo is obtained from the image sequence as well. This approach has been explored over the last few years. However, there were still requirements of complex calibration, user interaction, and extended computations. The main problems associated with these methods are high computational requirement and low resolution and accuracy due to the lack of reliable features on the object surface.

1.4 Time of flight methods

Using short pulse laser to measure the time of flight from the laser source to the object has become popular in remote sensing applications. In the remote sensing community, radar and laser ranging have been used to measure the 3D geometry of the earth¹². It is still expensive to operate.

The paper describes a 3D imaging technique that performs fast and accurate 3D imaging using a single digital camera to perform stereo vision with an optional color-grating projection.

2. SYSTEM ARCHITECTURE

We present a 3D imaging system that uses a prism optical setup to combine two stereo view pairs into one image in a digital camera. The 3D imaging technology combines the structured light projection and stereo vision techniques in a unique way: it splits a single camera view into stereo vision. It has an option to project a color-coded structure light onto the object using a synchronized flash light source¹³. It achieves 3D imaging in a single flash (< 1/100th second), thus robust to object motion and changing environment. It is also accurate, down to 0.1 mm in depth resolution. We also present advanced 3D modeling algorithms for fast and accurate 3D model generation. The system includes following components, as illustrated in Figure 1.

1. A commercial off-the-shelf digital camera captures 2D color images of object;
2. A mirror and prism based stereo device combines two stereo views into a single image;
3. An optional flash projector projects a color-coded pattern onto the object to improve the 3D resolution and accuracy. The flash projector is synchronized with the camera shutter.
4. A calibration board to calibrate the internal and external parameters of the 3D system.
5. Software creates the 3D model for a stereo image pair.

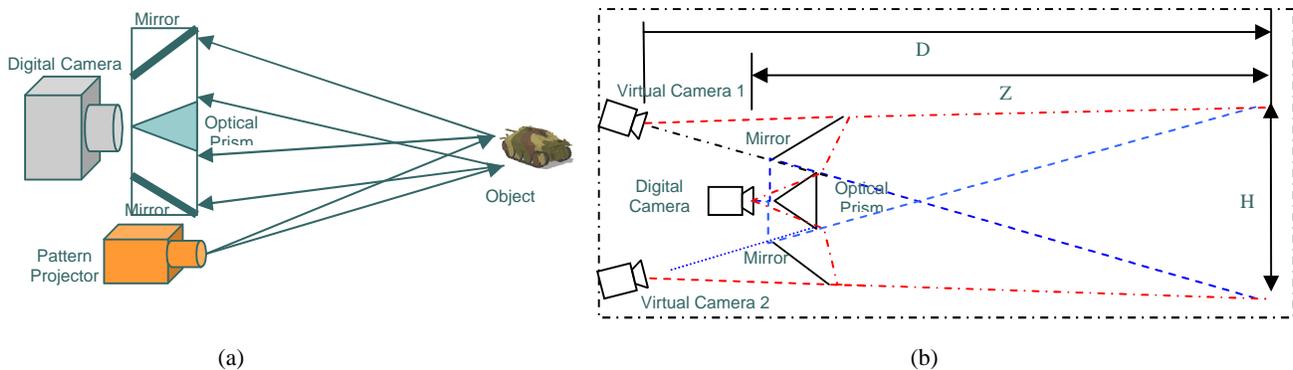


Figure 1: A single digital camera based 3D imaging system for high-resolution and high-speed 3D imaging: (a) Illustration of an optical mirror and prism setup; (b) Schematic diagram of the equivalent 2 virtual cameras stereo vision.

3. 3D IMAGING ALGORITHM

3.1 3D representations

An eye at point C views a point M in the world coordinate. The point M projects a point m on the retinal plane R . For a world point $M = (X, Y, Z)$ and the corresponding image point $m = (x, y)$. Using the homogeneous representation of the points a linear projection equation is obtained:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

This projection is illustrated in Figure 2. The optical axis passes through the center of projection C and is orthogonal to the retinal plane R . Its intersection with the retinal plane is defined as the principal point c .

3.2 Camera model

An actual camera has a focal length f , where f is defined as the distance between the center of projection and the image plane. With a CCD camera the relation between the coordinates in the image and the physical coordinates depends on the size and shape of the pixels and of the position of the CCD chip in the camera.

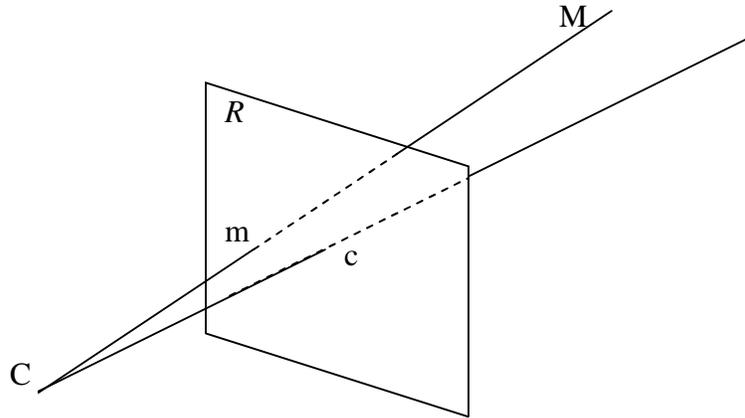


Figure 2: An eye at position C views a point M in the world. The point M projects a point m on the retinal plane R .

The image coordinates are obtained through the following equations:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x_R \\ y_R \\ 1 \end{bmatrix} \quad (2)$$

where f_x and f_y are the focal length measured in width and height of the pixels, $c = [c_x \ c_y \ 1]^T$ is the principal point, and s a factor accounting for the skew due to non-rectangular pixels. The above upper triangular matrix is called the *calibration matrix* K of the camera. So, the following equation describes the transformation from retinal coordinates to image coordinates.

$$m = Km_R \quad (3)$$

3.3 Projection matrix

The following expression is obtained for a camera with specific intrinsic calibration and with a specific position and orientation:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R^T & -R^T t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4)$$

which can be simplified to

$$m \sim PM \quad (5)$$

where R is a rotation matrix and $t = [t_x \ t_y \ t_z]^T$ a translation vector. The 3×4 matrix P is called the *camera projection matrix*. [10]

3.4 Two view geometry

Given an image point in one image it restricts the position of the corresponding image point in another image, as shown in Figure 3. This relationship can be obtained from the calibration or even from a set of prior point correspondences. Although the exact position of the scene point M is not known, it is bound to be on the line of sight of the corresponding image point m . This line can be projected in another image and the corresponding point m' is bound to be on this projected line l' . l and l' are said to be in epipolar correspondence (i.e. the corresponding point of every point on l is located on l' , and vice versa).

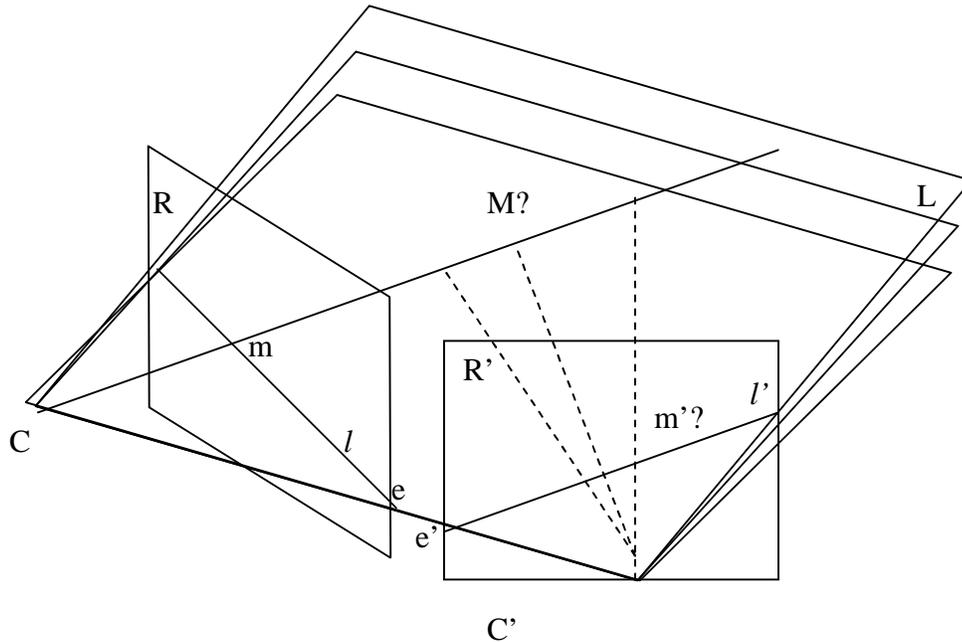


Figure 3: Correspondence between two views. Even when the exact position of the 3D point M corresponding to the image point m is not known, it has to be on the line through C and C' which intersects the image plane in m . Since this line projects to the line l' in the other image, the corresponding point m' should be located on this line. More generally, all the points located on the plane defined by C and C' have their projection l on and l' .

Every plane passing through both centers of projection C and C' results in such a set of corresponding epipolar lines, as can be seen in Figure 5. All these lines pass through two specific points e and e' . These points are called the *epipoles*,

and they are the projection of the center of projection in the opposite image. This epipolar geometry can also be expressed mathematically:

$$m^T F m = 0 \quad (6)$$

This matrix F is called the *fundamental matrix*.

When the calibration is not known, equation (6) can be used to compute the fundamental matrix F. Every pair of corresponding points gives one constraint on F. The rank F = 2, therefore 7 point correspondences are sufficient to compute F through nonlinear algorithms¹⁴.

3.5 Camera aberration

Practically, the camera model has some error caused by the optical aberration and misalignment of the CCD chip to the lenses.

The lens aberration can be represented by

$$\begin{aligned} c'_x &= c_x + (c_x - c_{x0}) \left[k_1 (x^2 + y^2) + k_2 (x^2 + y^2)^2 \right] \\ c'_y &= c_y + (c_y - c_{y0}) \left[k_1 (x^2 + y^2) + k_2 (x^2 + y^2)^2 \right] \end{aligned} \quad (7)$$

where c_x and c_y are the ideal point position, and c'_x and c'_y the actual point position. k_1 and k_2 are the camera lens aberration parameters.

3.6 Camera calibration

The camera calibration procedure calculates the external parameters R, t, and the internal parameters K. In order to calculate the parameters, we map the known points $M_i \quad i = 1, 2, \dots, N$ in the 3D space to the corresponding points $m_i \quad i = 1, 2, \dots, N$ in the 2D images. We receive the following equation:

$$m_i = K [R \quad t] M_i, \quad i = 1, 2, \dots, N \quad (8)$$

If we find enough number of point pairs, then we can calculate the internal and external parameters. In general, the points need to be in different planes. That requires building a special calibration object. We used a simplified flat plate with chess board pattern on it. We take 3 – 5 shots of the calibration board with the 3D camera with the board in different planes, as shown in Figure 4. The corner points of the squares on the chess board are detected by the software. The geometrical relationship among the points on the same board is known. Then, we can calculate the parameters from the equation. The parameters are further optimized by applying the Bundle Adjustment algorithm.

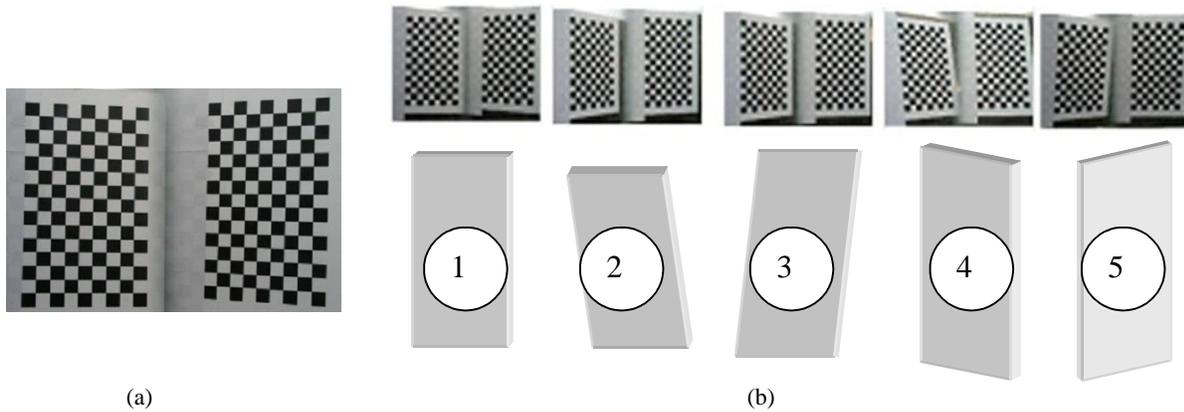


Figure 4: Calibration procedure: (a) a chessboard pattern; (b) 5 positions of the chessboard for the camera calibration.

3.7 Image rectification

After we have obtained the camera internal parameters, we can rectify the image. The rectification process can be carried out from Eq. 8. After rectification, the 3D point pairs are on the same parallel lines between two stereo image pairs, as shown in Fig. 5. Therefore, it saves time for the stereo matching process to shift in one dimension only.

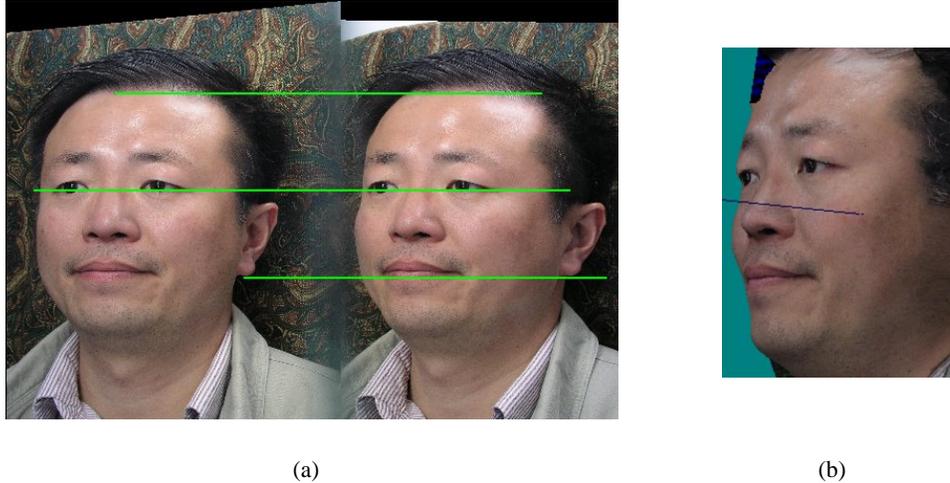


Figure 5: After rectification, the stereo point pairs are on the same epipolar lines; (b) a 3D model with color texture.

4. EXPERIMENTAL RESULTS

We have used the 3D imaging system to take live and static objects. The model sizes vary from ~ 20 cm to 1 meter. A digital camera (Olympus digital camera) with 5 Mega pixels was used. The lens focal length was 100 mm. The camera shuttle speed was 1/125s. Single shot was taken for each image view. Software was implemented to download the images from the digital camera via a USB link, to calibrate the camera parameters, to reconstruct the 3D models, to stitch the models together, and to display the 3D models.

Figure 6 shows the 3D reconstruction of the face of a model. A color-coded grid pattern was flashed onto the surface as the camera captures the stereo image pair. The software reconstructed the 3D model with details of the facial features clearly seen. Smoothing of the surface gives a better visual appearance, but sacrifices the resolution.

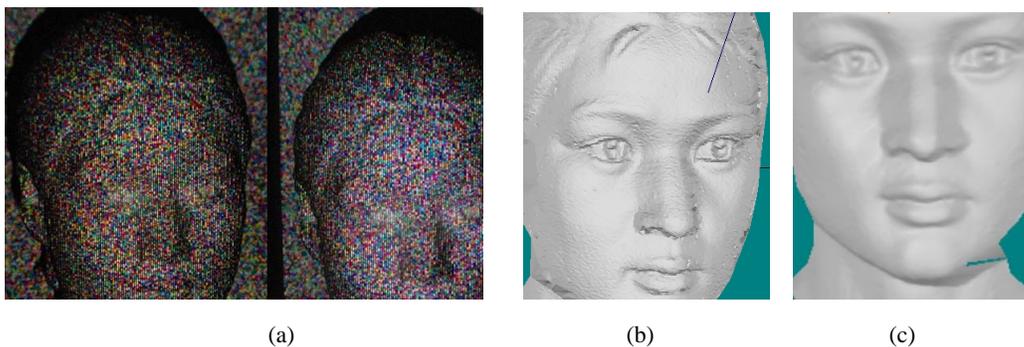


Figure 6: 3D imaging of the sculpture of a girl: (a) camera stereo view pairs under projected light; (b) reconstructed 3D model; (c) 3D model after surface smoothing.

Figure 7 shows the 3D imaging of a cruise ship model Queen Mary 2. The model is ~ 1.2 meter long. Approximately 8 shots were taken from 360 degree views. The 3D models are stitched together to form a complete 3D model.

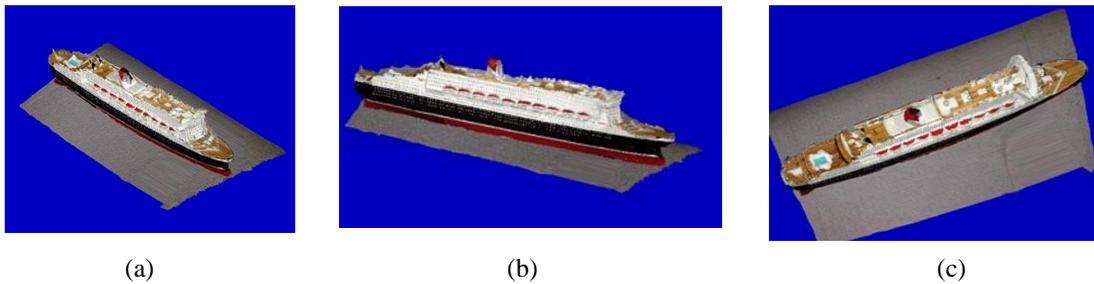


Figure 7: 3D imaging of models: cruise ship Queen Mary 2 model was imaged in 360 degrees; (a) – (c) multiple views were stitched together to form a complete 3D model;

Figure 8 illustrates the 3D imaging and modeling of several military models, a Patriot tank, and a Thunderwing fighter jet. From Figures 8(a) – (c), we can see, the false shadow of the camouflage can be removed and the true surface structure can be revealed by using the 3D imaging method. It provides valuable 3D information of an object for ATR applications. The 3D model can be rotated and scaled to arbitrary viewing angle and size to match the target orientation and size in the image. It is also an excellent model for building training database for precision target recognition¹⁵⁻¹⁶.

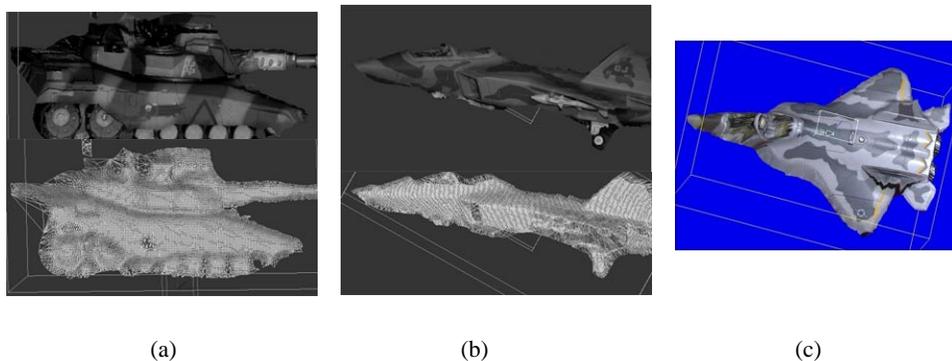


Figure 8: 3D imaging of models: (a) a Patriot tanks side view (top) and its underlying 3D structure (bottom); (b) the side portion of a Thunderwing fighter jet (top) and its 3D model (bottom), and (c) the top 3D view of the Thunderwing with color texture map.

4.1 Variation of camera field-of-view to accommodate object sizes and distances

The above experiments were carried out in the laboratory using a standard digital camera with a lens of 100 mm focal length. It is calibrated to image an object of ~ 30 x 30 cm² in a distance ~ 1 meter. The distance between the virtual camera pair is ~ 20 cm. If an object is larger than 30 cm in size, several partial images may be taken and then stitched together. The advantage of this approach is the high resolution of the reconstructed models. But the stitching of the models by the software could introduce systematic error in the model. If the object is much larger than the current one, then we need to pull the camera further away from the object, and separate the cameras further. The role of thumb of the viewing angle between two cameras is ~ 15 degrees. For example, the cameras can be placed 20 meters away from an object. The camera separation would be ~ 4 meters. The object size can be in the order of 6 x 6 m².

In the UAV application of imaging the scene and reconstruct the 3D world view, we may use only one camera on-board the UAV to continuously image the scene as the UAV moves in its course. The neighboring images in the sequence form stereo pairs. The camera internal and external parameters can be estimated based on fixed feature points on the

ground. Each stereo image pair can form a 3D model of the scene it sees. The computation is very intensive in this application. Optimization and digital FPGA parallel processing board may be used to achieve real-time processing.

4.2 Accuracy test experiment

This experiment tests the error and noise effect of the 3D algorithm on a flat surface. The flatness of the plane also represents the resolution of the 3D object. When we smooth the surface, the flatness is increased, but the resolution is reduced.

The 3D camera takes a picture of a flat surface. The 3D model is reconstructed. A Least-mean-square (LMS) algorithm is used to construct a flat plain and calculate the distance from the points of the 3D model to the flat plane. The square root error represents the flatness measure.

Assume N points in space; $P_i = [x_i, y_i, z_i]^T, i = (1, 2, \dots, N)$; assume they are on a plane

$$z = ax + by + c \quad (9)$$

For a point on the 3D model P_i' , the error ε_i is

$$z_i = ax_i + by_i + c + \varepsilon_i, i = 1, 2, \dots, N \quad (10)$$

In matrix form:

$$Z = AX + E \quad (12)$$

Use LMS method, we can calculate the parameters a , b , and c :

$$\begin{aligned} A &= (X^T X)^{-1} X^T Z \\ E^T E &= Z^T Z - (X^T Z)^T A \end{aligned} \quad (13)$$

The mean error in z direction is

$$\varepsilon = \sqrt{E^T E / N} \quad (14)$$

The surface area error is

$$\varepsilon_p = \varepsilon / (1 + a^2 + b^2) \quad (15)$$

We use samples in the center, and the 4 corners to represent a large surface.

According to the stereo imaging theory, if the camera distance is d , the distance from the camera to the object is z , the camera focal length is f , and then the 3D resolution is

$$\frac{dz}{du} = -\frac{z^2}{f \cdot d} \quad (16)$$

If the camera parameters are $z = 1250.00\text{mm}$, $d = 197.9\text{mm}$, $f = 5107$, then the theoretical resolution is 1.546 (mm/pixel). A sub-pixel matching algorithm can be used in the 3D calculation; the minimum resolution is increased between $1/3$ to $1/10$ of a pixel. Therefore the resolution is increased to $1.546/3 = 0.515$ (mm), and the mean error is decreased to $0.515/3 = 0.17\text{mm}$.

The test result is shown in table I. Testing points were sampled from five areas of the plate: the upper left, upper right, center, lower left, and lower right regions. The average error was calculated and it shows the error was as low as 0.14 mm in an area of 30 x 30 cm² plate.

3D Model Error (mm)	Upper Left	Upper Right	Center	Lower Left	Lower Right	Average
Unsmoothed Flat Plane	0.16	0.13	0.16	0.15	0.12	0.14
Smoothed Flat Plane	0.12	0.12	0.13	0.08	0.06	0.10

Table I: Experiment results of the accuracy of a 3D model of a flat plane. The average mean error is ~ 0.14 mm.

4.3 Potential applications

3D vision can be used to improve the robustness of automatic target recognition. The 3D imaging technique enables real-time 3D acquisition of 3D features and profiles that a regular camera can not achieve. A 3D model of a target gives the complete information, e.g., the physical sizes, surface profile, illumination/shading, the viewing angles and the distance information that can not be given by the conventional imaging modality. These are critical information that can tremendously help target detection, recognition and identification. It is also a valuable tool to digitize a model for the construction of a filter bank of targets for target ID. It performs quantitative measurement of target features and 3D visualization of objects.

Other applications include 3D facial ID for security surveillance. The 3D system can be adapted by the existing setup of a single digital camera to capture the 3D face model in stead of a 2D mug shot. It can reduce the error caused by the ambient light, viewing angles and the make up for robust biometric ID. In medical field, the 3D imaging system can be used to aid the real-time surgical guidance. It can be used to perform 3D modeling, simulation and tracking of live human body. It can be used to register the patient's body with ultrasound, MRI and CT images. It has potential applications in plastic and reconstructive surgery - 3D modeling, simulation and visualization and quantitative measurement of body parts; it has potential application in non-contact burn patients facial mask generation in burn centers. In industrial applications, it can be used in machine vision, real-time 3D imaging and inspection, rapid prototyping, and 3D imaging and modeling of machine parts.

5. CONCLUSIONS

We have demonstrated a fast and accurate 3D imaging system using a single digital camera. The system combines the stereo vision and the color-coded grating projection methods to achieve fast and robust 3D imaging. The accuracy of the 3D imaging system can be 0.1 mm in an area of 30 x 30 cm area. The system can be scaled to image large objects. The system algorithm can also be customized to process image sequences from a video camera onboard a UAV. The system has potential applications in a broad field.

6. ACKNOWLEDGEMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors wish to thank Frank Wu and John J. Zhang for valuable input and software implementation.

7. REFERENCES

1. J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital wavefront measuring interferometer for testing optical surface and lenses." *Appl. Opt.*, 13: 2693–2703, 1974.
2. B. Horn, *Robot Vision*, MIT Press, 1986.
3. M. Levoy, J. Ginsberg, J. Shade, D. Fulk, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, "The Digital Michelangelo Project: 3D Scanning of Large Statues," Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp.131-144, 2000.
4. G. J. Agin and P. T. Highnam, "Movable Light-Stripe Sensor for Obtaining Three-Dimensional Coordinate Measurements," *SPIE Proc.* Vol. 360, p. 326, 1983.
5. K. Harding, "Moire Interferometry for Industrial Inspection," *Lasers and Applications*, p. 73, Nov. 1983.
6. O. Faugeras and G. Toscani, "Camera Calibration for 3D Computer Vision", *International Workshop on Machine Vision and Machine Intelligence*, pp. 240-247, Tokyo, 1987.
7. R. Hartley, "Estimation of relative camera positions for uncalibrated cameras", *Computer Vision -ECCV'92*, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 579-587, 1992.
8. O. Faugeras, *Three-Dimensional Computer Vision: a Geometric Viewpoint*, MIT press, 1993.
9. P. Beardsley, P. Torr and A. Zisserman "3D Model Acquisition from Extended Image Sequences", *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. 1065, Springer-Verlag, pp. 683-695, 1996.
10. R. Koch, M. Pollefeys and L. Van Gool, Automatic 3D Model Acquisition from Uncalibrated Image Sequences, Proceedings Computer Graphics International, pp.597-604, Hannover, 1998.
11. C. Slama, *Manual of Photogrammetry*, American Society of Photogrammetry, Falls Church, VA, USA, 4th edition, 1980.
12. N. Haala, and C. Brenner, "Generation of 3D city models from airborne laser scanning data," *Proceedings of EARSEL workshop on LIDAR remotesensing on land and sea*, Tallin, Estonia, pp. 105-112, 1997.
13. T. Lu, J. Zhang, "3 Dimensional Imaging System," U.S. patent No. 6,252,623, 2001.
14. Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", *Artificial Intelligence Journal*, Vol.78, pp.87-119, October 1995.
15. J. C. Hanan, T.-H. Chao, C. Assad, C. L. Hughlett, H. Zhou, T. Lu, "Closed-Loop Automatic Target Recognition and Monitoring System," p. 244-251, *Optical Pattern Recognition XVI; SPIE 5816*, 2005.
16. T. Lu, C. L. Hughlett, H. Zhou, T.-H. Chao, J. C. Hanan, "Neural network post-processing of grayscale optical correlator," *Proc. SPIE 5908, Optical Information Processing III*, 2005.