1-1-2006

A system for the 3D reconstruction of the human face using the structured light approach

Dean McGuire
University of Wollongong

Prashan Premaratne
University of Wollongong, prashan@uow.edu.au

A SYSTEM FOR THE 3D RECONSTRUCTION OF THE HUMAN FACE USING THE STRUCTURED LIGHT APPROACH

D. MCGUIRE, P. PREMARATNE
The School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, NSW, Australia

ABSTRACT

This paper will describe a computer vision system under development with the aim of use in robust facial recognition applications. The system employs the structured light approach to solve the correspondence problem. The system has been designed with ease of use in mind with calibration procedures designed to be simple enough to be carried out without the use of precision measuring equipment, but robust enough to provide good quality 3D reconstructions. The hardware in use is of consumer grade and is being used in an unmodified form. Early reconstructions of faces based on the projection of a striped pattern have yielded good results with the hope that higher resolution reconstructions will be achieved in the near future.

1. INTRODUCTION

The field of computer vision is concerned with the ability of computers to perceive the world as humans do.

In essence, this means to perceive the world in three dimensions. The advantages of such a skill are clear. If a computer could rapidly and accurately perceive the depth of objects in its field of vision, it could interact with those objects with accuracy and precision. In addition, a computer has the ability to analyse such objects to a greater degree than humans.

Applications of such a skill include, but are not limited to, medical imaging, manufacturing quality control and robust facial recognition.

There are two main branches of computer vision. The first, known as passive vision, computes the 3D information of a scene via the use of two cameras and “with the help of only the existing ambient illumination [1]”. This is contingent upon knowing the properties and position in space of the cameras.

The process of obtaining these parameters is known as calibration, and is a vital step in any computer vision system. The second, known as active vision, replaces one of the cameras with a projector.

The significance of this arrangement and its advantages over the passive configuration will become apparent in this paper.

The essential properties that this vision system must possess are fast reconstructions (in the order of 5 seconds), and the ability to vision moving objects. This paper will be arranged as follows. Section 2 will describe the process of 3D reconstruction. Section 3 will discuss the current methods available for computer vision and their various advantages and disadvantages. Section 4 will present the details of the system under development and Section 5 will analyse the results of this system has produced thus far.

2. 3D RECONSTUCTION

The process of obtaining the 3D coordinates of a scene is relatively straightforward in principle. [2] notes that “Fusing the pictures recorded by our two eyes and exploiting the difference (or disparity) between them allows us to gain a strong sense of depth”.

To calculate the 3D position of any point in space, it is required that the point in question be visible in both camera images and, more significantly, that the point is able to be located in both images. This criterion is known as the correspondence problem and is the most fundamental problem of computer vision. [1] defines correspondence as detecting "features that are projections of the same physical identity in each view."

Once a correspondence has been obtained, the process of triangulation is performed to calculate the 3D position of the point. The distinction of vision systems into active and passive is based on the way in which each solves the correspondence problem.

In passive vision systems, two cameras are utilised and correspondence is obtained via the comparison of the subsequent images. The techniques to achieve this, as outlined in [1, 2] all suffer from the problem of being computationally intensive and hence not suitable for applications which require fast processing.

In active vision systems, one camera is replaced with a projector which projects a known pattern onto the scene to be surveyed. The purpose of this is to reduce the correspondence problem to that of finding the same pattern in the remaining camera image. As [3] points out "The correspondence problem can be considerably alleviated by an active method."

Specifically, triangulation is performed in the following way.

For a camera using the pinhole model, the world coordinates (X, Y, Z) are projected onto the pixel coordinates (u, v) by

$$\left[u\ v\ 1\right]^T = KK\left[x_n\ y_n\ 1\right]^T, \quad \text{- (1)}$$

where $\left[x_n\ \mathrm{y_n}\right]^T = \left[X_c / Z_c\ \mathrm{Y_c} / \mathrm{Z_c}\right]^T$ - (2)

and (x_n, y_n) is the normalised image projection and where (X_c ,Y_c , Z_c) is the coordinates of the point (X,Y,Z) in the camera's coordinate frame which is given by

$$\left[X_c\ Y_c\ Z_c\right]^T = \left[R\ \mathrm{t}\right]\left[X\ \ Y\ \ Z\ \ 1\right]^T, \quad \text{- (3).}$$

KK is the camera's 3x3 intrinsic calibration matrix and includes the focal length, principal point, and skew of the camera. It should be noted that this model does not incorporate radial and tangential distortion which has been left out here for simplicity. The matrices R_{3x3} and t_{3x1} are the rigid motion rotation and translation matrices respectively.

The projector is modelled as an inverse camera and as such can be represented by the same model above.

Assuming the correspondence problem has been solved and two points (u_1, v_1), (u_2, v_2) are known to be corresponding, the problem is to determine the values of (X, Y, Z).

To solve this, the normalised image projections (x_{n1}, y_{n1}) and (x_{n2}, y_{n2}) are computed by performing the operation $\left[x_n\ y_n\ 1\right]^T = (KK)^{-1}\left[u\ v\ 1\right]^T$.

As such, there is now 9 unknowns $(X, Y, Z, X_{c1}, Y_{c1}, Z_{c1}, X_{c2}, Y_{c2}, Z_{c2})$ and 10 constraining equations. To clarify, there are 2 equations each for the camera and projector from equation (2), and 3 equations from equation (3). Therefore, in total there are 10 constraining equations. As such, this system is overdetermined and can be solved using a linear least squares approach.

As the system is overdetermined, the correspondence can be reduced to simply finding x_n or y_n in the projector image. This is the reason that triangulation can be performed via a stripe only in the projector pattern.

3. AVAILABLE METHODS

The main thread of research in active vision systems is on the pattern which is projected onto a scene. The so called 'Codification Strategy' determines how a pattern encodes the information pertinent to locating each point in both the camera image and projector pattern.

[4] defines the codification strategy of a pattern such that "a set of pixels are easily distinguishable by means of a local coding strategy."

There are three basic ways in which to code a pattern in a structured light system. These are Time-Multiplexing, Spatial Neighbourhood and Direct Coding [4].

In Time-Multiplexing based approaches, a "set of patterns are successively projected onto the measuring surface [4]." Specific techniques for this include Binary Coding [5] and Gray Codes [6]. The main drawback of these patterns are that they do not allow for a 'one-shot' approach, as multiple patterns are projected. As such, they are inadequate for this application where the vision of moving objects is required.

In Spatial Neighbourhood based approaches, the identification of any point is "obtained from a neighbourhood of the points around it [4]. The detection of these patterns is based on the premise that every point is uniquely characterised by the points that surround it. Of the proposed patterns in this approach, those based on De Bruijn sequences, as implemented, for example, by [7], are the most popular due to the ease with which these patterns can be produced from precise mathematical formulae. Although this approach can be used on moving objects, problems can arise when neighbouring pixels are obscured from view.

The final approach, Direct Coding, codes the pattern so that each pixel contains all the information required to identify it [6]. This is almost always achieved through the use of different shades of grey, as in [8], or through the use of different colours, as in [9]. The main problem of this approach is that although each point depends on no other, the use of a large number of colours or shades means that identifying each colour or shade is very difficult.

4. THE VISION SYSTEM

The essential dilemma when developing a 3D vision system of any kind is to decide on a method which provides the right mix of speed and model density.

Speed refers to the speed with which the subject is acquired by the computer and the model subsequently determined. Model density refers to the number of points which are reconstructed to determine the model.

Obviously, a tradeoff exists. The more points that are required to be reconstructed, the longer it will take to compute those points.

In the system currently in development, two main criteria are of most importance. Firstly, the system requires that a 'one-shot' method be employed, so that moving scenes can be modelled in 3D. Secondly, the reconstruction must be fast enough to be used in real time. The choices that have been made for the system are made in the context of these requirements.

4.1 System Overview

The process involved in creating a 3D model of a scene proceeds in three main steps. Firstly, the camera and projector must be calibrated to determine their intrinsic properties, such as focal length, and their extrinsic properties, which describe how the device is positioned relative to a world coordinate system.

Secondly, the pattern must be projected and correspondences between the projector and camera obtained. Lastly, these correspondences are used, in combination with the calibration results, to obtain the 3D coordinates of the points in the scene.

Figure 1 shows the system setup in its current state. The use of two cameras and a projector is to enable the system to view a larger area of the scene than is possible with only a single camera. Note however that the reconstruction is still done using each camera and the projector.

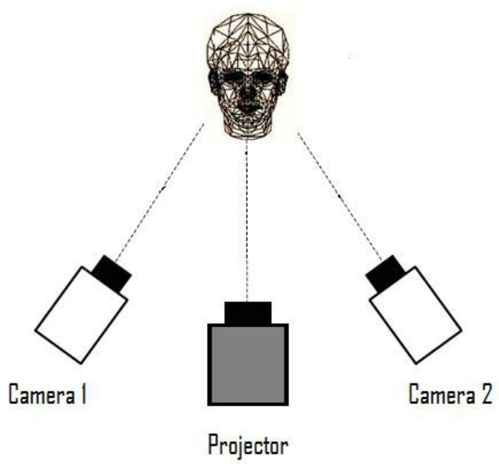

Figure 1: The System Setup

4.2 Camera Calibration

The purpose of the camera calibration procedure is to determine the intrinsic and extrinsic properties of the cameras [1]. These will determine the camera intrinsic calibration matrix KK present in equation (1) and the rigid motion matrices R and T present in equation (3). The essential task of the calibration procedure is to determine a set of 3D-2D mappings. This involves locating a set of points in an image that correspond to known points in the 3D world space.

Most methods for achieving this use a calibration object and precision measuring equipment to determine the coordinates of points of interest in the 3D space. This technique, while accurate, is not appropriate for users with minimal training. It is also expensive. Given that this system needs to be easy to use and cheap, methods such as this are not appropriate.

For this reason, this system uses a method of calibration proposed by [10]. This method requires only that the user present a number (approximately 20) of different poses of a planar checkerboard (Figure 2) to the cameras.

The system then computes the homography for each plane and uses this to produce the calibration results. Any of these plane poses can be used as the basis for the world coordinate system.

In this system, the first pose is used as the reference plane for the coordinate system. This calibration is done only once for any stationary setup of the system.

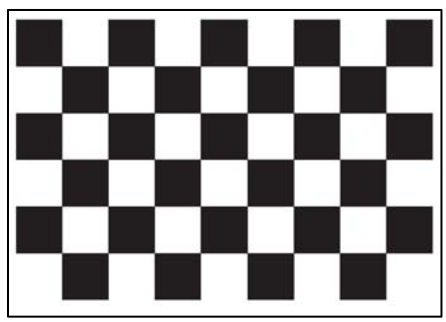

Figure 2: The Calibration Object

4.3 Projector Calibration

Projector calibration is analogous to camera calibration in that it determines the intrinsic and extrinsic parameters of the projector, which is modelled as an inverse camera. The important task of projector calibration is to obtain a set of 2D-3D mappings from the projector space to the same world space as the cameras were calibrated in.

Research into projector calibration has been much less prevalent than for cameras. As such, there are not many methods available for use.

Of the ones that are published, many rely on a point from the projector striking a known point on a surface, and using these for the 2D-3D mappings [12]. This method can be tedious, time consuming and requires extra processing to detect correspondences. As such, this method was not used in this system.

The method that has been utilised is a version of that presented by [11]. This method uses the two previously calibrated cameras to aid in the calibration of the projector. The projector projects the same pattern as in Figure 2 onto a planar surface. This is shown in Figure 3. The cameras then image this surface and extract the corner points.

These corner points, whose correspondence is easily determined, are used as the basis for a 3D reconstruction via the two camera system. These 3D points are then used as the 2D-3D mappings with the projector pattern. Having computed these mapping enables calibration to be carried out in the same manner as for the cameras.

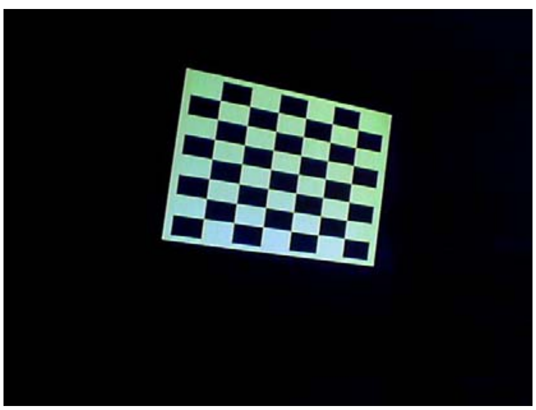

Figure 3: A Projector Calibration Image

4.4 Pattern Codification

The variety of pattern choices was reviewed in Section 3. After testing a number of methods including De Bruijn patterns and a colour coded grid, a pattern proposed by [13] was chosen (Figure 4).

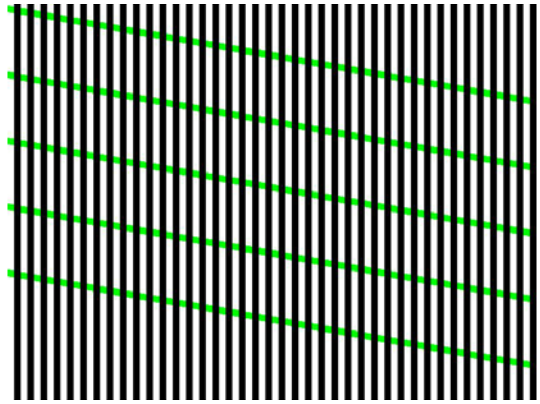

Figure 4: The Projector Pattern

This pattern was selected as it consists only of black and white stripes in the pattern, with coloured 'coding lines' beneath this pattern. This means that the pattern is easily detectable but also does not rely on any strict neighbourhood relationship.

The way in which each black and white edge is coded is through using the epipolar geometry which relates the projector to the cameras. Consider any point in the camera image where the coding line intersects with the base pattern. This point will produce a unique epipolar line in the projector image. This epipolar line is a function of the intrinsic and extrinsic parameters of both the camera and projector. This epipolar line will intersect one or more of the coloured coding lines.

By detecting these intersections, and determining the correct one via a clustering algorithm, a correspondence between the camera and projector for that point can be obtained. Obviously, having a correspondence for any one point on a stripe edge is equivalent to having the correspondence for that complete stripe edge provided that edge is continuous in the camera image.

Performing this for all stripe edges in the image results in correspondences which are then used for the 3D reconstruction. All this is obtained without using complex coloured patterns which are difficult to detect.

5. RESULTS

Results for the camera and projector calibration are evaluated by determining the reprojection error. This is the error, in pixels, that is determined by taking the mean of the difference between the feature locations in the image and as predicted by the camera or projector model. Tests have indicated that the reprojection error in the model is in the order of 0.1 to 0.2 pixels. This is very small and more than satisfactory for this application.

Figure 5 shows the projector pattern striking the face of a test subject. The stripes have been deliberately spaced at large intervals to allow the reader to see the pattern easily.

Figure 5: Pattern Striking Face

Figure 6 shows the 3D reconstruction of the face based on approximately 4000 points.

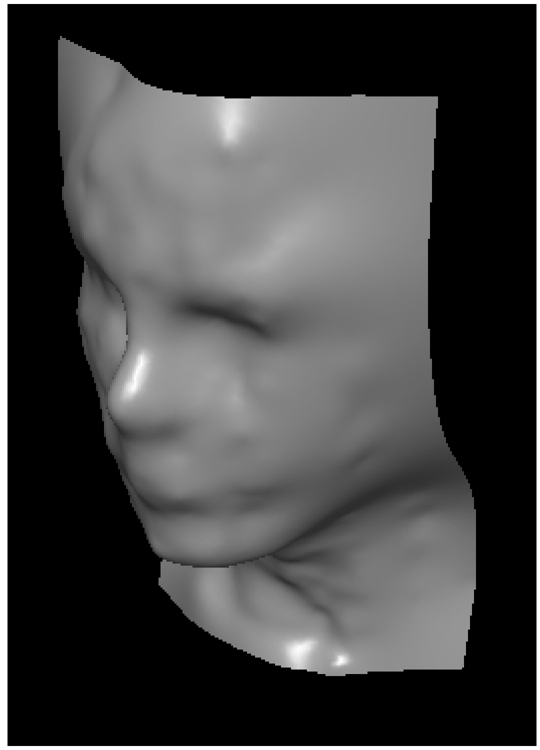

Figure 6: 3D Reconstruction of Face

6. CONCLUSION

This paper has described a computer vision system under development. The calibration procedures are complete and provide robust and accurate calibration information for both the cameras and the projector. The projector pattern consists of a base pattern of black and white stripes, with coloured coding lines placed beneath the base pattern. These coding lines make absolute correspondence between the cameras and the projector possible. This system has been used to generate sparse 3D reconstructions with promising results.

Future work on this system will consist of increasing the resolution of the reconstruction to provide more accurate models of the scanned scene. This will involve projecting a larger number of stripes from the projector. This implies smaller width stripes and hence makes detection more difficult. This will be the main challenge in future work.

REFERENCES

[1] U. R. Dhond, J.K. Aggarwal, "Structure from Stereo – A Review”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp.1489-1510, 1989.

[2] D. Forsyth, J. Ponce *Computer vision: a modern approach*, London : Prentice Hall, 2003.

[3] J. Batlle, E. Mouaddib, J. Salvi, “Recent Progress in Coded Structured Light as a Technique to Solve the Correspondence Problem: A Survey”, *Pattern Recognition*, vol. 31, no. 7, pp. 963-982, 1998.

[4] J. Salvi, J. Pages, J. Batlle, “Pattern codification strategies in structured light systems”, Pattern Recognition vol. 37, no. 4, pp. 827-849, 2004.

[5] J. L. Posdamer, M. D. Altschuler, “Surface measurement by space-encoded projected beam systems”, *Computer Graphics and Image Processing*, vol. 18, no. 1, pp. 1-17, 1982.

[6] D. Bergmann, “New approach for automatic surface reconstruction with coded light”, in *Proceedings of SPIE: Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes*, vol. 2572, SPIE, 1995, pp. 2-9.

[7] J. Salvi, J. Batlle, E. Mouaddib, “A robust-coded pattern projection for dynamic 3D scene measurement”, *International Journal of Pattern Recognition Letters*, vol. 19, no. 11, pp. 1055-1065, 1998.

[8] B. Carrihill, R. Hummel, “Experiments with the intensity ratio depth sensor, in: Computer Vision”, *Computer Vision, Graphics, and Image Processing*, vol. 32, pp. 337-358, 1985.

[9] J. Tajima, M. Iwakawa, “3-D data acquisition by rainbow range finder”, in *International Conference on Pattern Recognition*, 1990, pp. 309–313.

[10] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations”, in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 666-673, 1999.

[11] R. Raskar et al, “Multi-Projector Displays Using Camera-Based Registration”, in *Proceedings of the conference on Visualization ’99*, 1999, pp. 161-168.

[12] T. Shen, C. Menq, “Digital Projector Calibration for 3-D Active Vision Systems”, *Journal of Manufacturing Science and Engineering*, vol. 124, no. 1, pp. 126-134, 2002.

[13] Koninckx, T., Griesser, A, Van Gool, L,
“Real-time Range Scanning of Deformable Surfaces by Adaptively Coded Structured Light”, in *Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling*, 2003