C3DTM: a Novel Vision-Based 3-D Data Acquisition System

J.P. Siebert and C.W. Urquhart

The Turing Institute, 77-81 Dumbarton Road, Glasgow G11 6PP, Scotland
and
Computing Science Department, University of Glasgow, Scotland

Abstract

This paper reports on the development of C3D, a computer-based product for constructing 3-D surface models of real world 3-D objects, e.g. a model of a human face, from stereo image pairs. C3D is a GUI-based computer program that converts the implicit 3-D information contained within the stereo image pair into an explicit 3-D surface model in a standard CAD file format. Hence, C3D extracts automatically the 3-D information from a stereo pair of images which have been frame-grabbed (digitised) into a computer and then builds a 3-D surface model. This surface model can then be edited using standard CAD packages or manipulated within virtual reality environments in the usual way.

1 Introduction

While 3-D visualisation and virtual reality systems have revolutionised the manipulation and use of 3-D information respectively, users of these systems encounter a major bottle-neck in the construction of surface models of existing objects. In an industrial context, many pre-CAD objects exist and require to be captured in three dimensions to allow design upgrades. C3D was conceived to provide a fast and cost effective route to capturing 3-D surface models of these classes of objects for applications ranging from industrial design, virtual reality and graphic arts to stereo biometrics (human body modelling) for pre-surgical planning.

1.1 Background

The core technology required for C3D was primarily developed with support from the UK Joint Funding for Information Technology scheme under the Active Stereo Probe (ASP) project, IED3/1/2109, which was lead by the Turing Institute and completed in July of 1993. This research work was further developed under the UK Department of Trade and Industry Small firms Merit Award scheme to produce a technology demonstrator. The goal of the current C3D project is to integrate this core technology into a 3-D data conversion product prototype. Hence, it is anticipated that this prototype will culminate eventually in the development of a

Figure 1. The complete C3D processing chain.

complete “shrink-wrapped” product. Currently, the prototype system serves as the basis for further research and custom development work by the Turing Institute.

1.2 Technical Approach Adopted

The technical approach to surface model construction proposed in C3D employs a correlation-based scale-space automatic stereo matching algorithm to determine the displacements, or parallaxes, between stereo image pairs of objects to be modelled. As these parallaxes reveal the 3-D structure of objects within stereo image pairs, photogrammetric analysis can be applied to the parallaxes to compute the 3rd dimension of interest. Hence, this 3-D information can be converted into a format compatible with, for example, standard packages such as the Silicon Graphics “Inventor” graphics information format. Figure 1 outlines the complete C3D processing chain.

Image capture hardware and object illumination have been accomplished by utilising the binocular sensor head developed under the ASP project as described fully in [1], [2].

2 An Overview of C3D

Currently, C3D has been implemented using the TK/TCL [3] GUI development toolkit which produces X-Windows compatible code running under UNIX. C3D is a modular program that enables the user to perform image acquisition, stereo matching, photogrammetry and calibration, 3-D surface construction and also 3-D surface visualisation within a single operating environment. These facilities are described further in the following sections.

Figure 2. Schematic of the C3D stereo matching algorithm.

2.1 Image Capture and Scene Illumination

Image capture is provided by commercially available frame digitisation hardware (DataCell S2200) hosted within a SUN SPARCstation. The binocular sensor head is connected to the red and green channels of the frame grabber. This configuration allows live red-green anaglyph display of objects being digitised and synchronised capture of stereo image pairs. It is also possible to input stereo pairs from existing image files.

In order to ensure that the stereo image pairs of the object being digitised contain a uniform spatial distribution of high frequency *image energy* or “texture”, the scene is bathed in “textured light” [4] during image capture. A second image is captured of the scene using “plain” illumination to provide a surface reflectance map which is later “draped” over the digitised surface during a subsequent photo-realistic rendering stage for 3-D visualisation. Illumination switching is accomplished by means of a computer controlled slide projector.

2.2 Stereo Matching

The Turing Institute’s proprietary signal matching algorithm is used to determine correspondences between the stereo image pairs, the detailed operation of which is discussed in [5]. This algorithm builds a scale-space image pyramid of octave separated image spatial frequency bandpasses. Image matching proceeds at the *top* of the pyramid, at the coarsest image resolution using area correlation to generate a low resolution map of horizontal and vertical inter-image parallaxes. This low resolution map is “expanded” by Gaussian filtering to the size of the next lower (higher resolution) level in the image pyramid and matching then proceeds as before. However, in this case the parallaxes held in the expanded low resolution map are now used as the starting point where the correlation search proceeds. Hence as the correlator is now operating at a higher image resolution than before, the correlation search process is now *refining* the estimates of parallaxes found at the

previous level of resolution in the pyramid. This matching strategy is repeated for each remaining level in the input image pyramid until the pyramid *base* (containing the highest resolution bandpass images) is reached whereupon matching is complete. The C3D matching process is outlined in figure 2.

Clearly, for this coarse-to-fine matching strategy to succeed, image energy must be present in each of the spatial frequency bandpasses of the input stereo pair. Hence the need for texture projection when the inherent image resolution prohibits capturing fine surface texture of objects (e.g. low resolution images of human skin) or where the objects are man-made and lack surface texture, i.e. are smooth. In these cases, the low spatial frequency nature of bland naturally illuminated surfaces would not provide any image information to enable the correlator to identify unique correspondences between the pixels of the input stereo pair images. Hence, by projecting texture onto these surfaces, it is possible to ensure that there is adequate image matching energy at all levels of the matching pyramid and thereby provide accurate parallax recovery over the interiors of smooth inherently featureless surfaces. A match confidence map is also generated as a result of the matching process and gives a pixel-by-pixel map of the estimated confidence of the recovered parallax maps.

It is possible for match errors to occur at *occlusions*, where a local image feature is visible in one of the images of the stereo pair, but *occluded* in the other. Although these errors could be corrected manually after model construction, we have found that the range image filtering operation makes this, at least superficially, unnecessary. Accordingly, the results presented in section 3 have been achieved without such manual intervention. The most satisfactory method for dealing with occlusions is to fuse together several 2.5-D surfaces recovered by the system in such a manner as to omit the portions of these depth maps which contain occlusions and merge those areas which are unoccluded. This could be achieved by utilising the match confidence map to label those recovered surface areas suffering from the effects of occlusion and thereby omit these areas (since occluded regions return poor match confidence scores).

2.3 Photogrammetry and Calibration

The DLT (Direct Linear Transform) [6] method of system calibration has been integrated into the C3D program to allow the operator to configure the system to the geometry of the cameras, i.e. vergence angle, baseline separation, relative orientation, focal length of lenses etc., used for stereo capture. A grid comprising an 8×8 matrix of circular rods of varying length (measured to 50 microns tolerance) serves to calibrate the system. The centre-points of these rods are located automatically within stereo image pairs captured of the grid using an implementation of Cumani's method [7] of circle centre finding. The current overall accuracy of the calibration of the order 0.2 mm in planimetry and 0.5mm in elevation for a camera baseline separation of 0.3m and object distance of approximately 2m. However, we are currently researching more advanced photogrammetric techniques to improve this result considerably, as described in section 4.

2.4 3-D Representation

The basic 3-D representation takes the form of an elevation image, where each pixel in the image represents a height value. Image processing operations have been included such as Gaussian smoothing and rank order sifting to allow the recovered elevation field to be filtered. It is then possible to convert the elevation field into a polygon model.

Two approaches are possible to achieve polygonisation. A surface polygon grid can be produced by simply treating pixels as the vertices of polygons. This is a convenient representation since the number of polygons that comprise a surface can be controlled by building a Gaussian pyramid from the height field map. The desired resolution can then be selected by extracting the appropriate level in the pyramid for direct polygonisation.

A better approach to polygonisation has been investigated which comprises selecting a matching confidence threshold above which height field pixels are accepted and below which they are rejected. Those accepted pixels are then Delaunay triangulated to deliver a mesh of variable polygon density depending on the local match confidence.

Polygon complexity control has been developed but not yet fully integrated into the system and could be applied to either the regular or triangulated grids. This comprises a standard polygon relaxation algorithm to reduce automatically the polygon density based on merging adjacent co-planar polygons (or near-coplanar polygons) into single polygons. Polygon merging takes place if the merged polygon is within a surface displacement error threshold when compared to the un-merged polygons. Clearly, it would be desirable to combine the merging and triangulation algorithms together to exploit directly the confidence information associated with the height field in a single algorithm that outputs a triangulated polygon grid of the desired complexity based on the best (most confident) set of vertices.

2.5 Visualisation

In order to visualise polygon surface models, a fast surface rendering algorithm has been incorporated into the system to allow this data to be displayed as either a shaded surface model or rendered with one of the input stereo images “draped” over the surface to give a photo-realistic perspective view. Conversion programs have been developed to allow this data to be exported in either DXF format or Silicon Graphics Inventor format.

3 Results

Figure 3 shows two stereo pairs of images acquired using the Active Stereo Probe: one captured while the subject was illuminated with textured light, and the other, captured immediately afterwards, under normal non-textured illumination. The images shown are a 512×320 pixel region extracted from the original full 576×768 pixel images. The surface model recovered using the C3D package is shown in figure 4 as a wire frame, a Gouraud shaded surface and a surface rendered (surface “draped”) with the left image from the normally illuminated stereo pair.

Figure 3. Input images. Top: stereo pair captured under textured illumination. Bottom: stereo pair captured under normal illumination.

Figure 4. Recovered surface model. Top: wire frame and Gouraud shaded. Bottom: rendered.

Figure 5. A conventional *film* camera equipped with a commercial stereo adapter attachment

4 Future Extensions

Currently, we are extending C3D both in terms of the image sources that it can accommodate and also the photogrammetric operations it provides.

We are also developing a variety of sensor heads ranging from high resolution DFCs (Digital Frame Camera) to low cost single chip cameras. High resolution cameras offer the possibility of capturing whole human torsos in a single shot process, since skin pore texture can be resolved and matched at an appropriately high image sampling density. Accordingly, automatic matching of high resolution stereo image pairs obviates the need for textured illumination of the subject. To this end we have recently investigated stereo-pair image capture by means of a conventional *film* camera equipped with a commercial stereo adapter attachment, as shown in figure 5. Kodak's PhotoCD digitisation service is used to obtain $3k \times 2k$ digital images of each film frame, providing stereo pairs of sufficient ($1.5k \times 2k$) resolution to allow images of human face subjects to be matched *without* texture projected illumination. Figure 6 shows a typical stereo-pair captured using this method and figure 7 shows two rendered views of the resultant 3-D surface models constructed by C3D.

Single chip cameras can be utilised in multiple-stereo pairs to provide cost-effective all-round views of objects/subjects. We also intend to provide a GUI based

Figure 6. A stereo-pair of images captured using the equipment shown in figure 5 and digitised via Kodak's PhotoCD service.

Figure 7. Two rendered views of the resultant 3-D surface model constructed by C3D from the stereo-pair images in figure 6.

interface to allow input of *ground control* points in order to accommodate stereo image pairs which have been captured from uncalibrated camera sources.

We have developed, but not yet integrated into C3D, a *bundle adjustment* [8], [9] photogrammetry package. Using bundle adjustment we expect to improve considerably our calibration and model accuracy. We shall also use bundle

adjustment to fuse multiple stereo image pairs into a single 3-D surface model. In this case, we shall investigate both manual and automatic selection of tie-points between the stereo image pairs, using scale-space signal matching in the automated case. It may also be possible to use bundle adjustment in conjunction with scale-space matching to recover 3-D surface models from digitised high resolution cine sequences such as 70mm film.

5 Conclusions

The essential innovation of C3D is to provide a cost effective software tool for generating 3-D models from stereo images using automatic stereo matching. While manual methods for model construction do exist, these manual methods are laborious, time consuming and are dependent on the skill of the operator for the final model quality. Precision automatic stereo matching employed within C3D provides for far greater model complexity and accuracy than could possibly be achieved by hand entered digitisation methods.

The principal technical challenge of the project has resided in devising and implementing a software architecture that both efficiently and seamlessly integrates stereo image capture, stereo image matching, photogrammetry, surface model construction and visualisation of this model. The human-computer interface challenge has been to devise a system that can harness the sophisticated algorithms and techniques developed under the ASP project within a simple-to-use yet powerful and productive software product.

In conclusion, the results presented demonstrate the technical viability of utilising stereo images for 3-D digitisation of human subjects. Two exciting implications stem from this development. Firstly, automated 3-D digitisation need no longer be highly costly as is the case for mechanically scanned systems such as laser stripers, since our system need only comprise a stereo pair of cameras (we have demonstrated C3D using only a conventional *film* camera) and a workstation to process these images. Secondly, data capture times are now restricted only by shutter speeds (electronic or mechanical) opening the possibility of capturing 3-D surface models of subjects *in motion*.

6 References

1. Urquhart C.W. and Siebert J.P. Development of a Precision Active Stereo System. In: *Proceedings of the IEEE International Symposium on Intelligent Control*, Glasgow, Scotland, August 1992, pp. 354-359.
2. Urquhart C.W. and Siebert J.P. Towards real-time dynamic close range photogrammetry. In: *Proceedings of SPIE Videometrics II*, Boston, U.S.A., September 1993.
3. Ousterhout J.K. *Tcl and the Tk Toolkit*. Addison-Wesley Publishing Company, Reading, Massachusetts, U.S.A., 1994.
4. Siebert J.P. and Urquhart C.W. Active Stereo: Texture Enhanced Reconstruction. *Electronics Letters*, 26, 7 (March 1990), pp. 427-430.

5. Jin Z.P. and Mowforth P.H. *A Discrete Approach to Signal Matching*. Research Memo TIRM-89-036, The Turing Institute, Glasgow, Scotland, January 1989.
6. Abdel-Aziz Y.F. and Karara N.M. Direct Linear Transformation from Comparator Coordinates into Object Coordinates in Close-Range Photogrammetry. In: *Proceedings of the ASP Symposium on Close-Range Photogrammetry*, Illinois, USA, January 1971, pp. 1-18.
7. Cumani A. *et. al.* High Accuracy Localisation of Calibration Points for Dimensional Measurements by Image Processing Techniques. In: *Proceedings of the Fifth International Conference on Advanced Robotics*, Pisa, Italy, 18-22 June 1991, pp. 1761-1765.
8. Karara H.M. (ed.). *Handbook of Non-Topographic Photogrammetry*, 2nd edition. American Society for Photogrammetry and Remote Sensing, Falls Church, VA., 1989.
9. Slama C.C. (ed.). *Manual of Photogrammetry*, 4th edition. American Society of Photogrammetry, Falls Church, VA., 1980.