Research article

Human body 3D imaging by speckle texture projection photogrammetry

J. Paul Siebert and Stephen J. Marshall

The authors

J. Paul Siebert (psiebert@dcs.gla.ac.uk) and **Stephen J. Marshall** (sjm@dcs.gla.ac.uk) are based at 3D-MATIC Faraday Partnership, UK.

Keywords

Vision, Medical, 3D image processing, VR

Abstract

Describes a non-contact optical sensing technology called C3D that is based on speckle texture projection photogrammetry. C3D has been applied to capturing all-round 3D models of the human body of high dimensional accuracy and photorealistic appearance. The essential strengths and limitation of the C3D approach are presented and the basic principles of this stereo-imaging approach are outlined, from image capture and basic 3D model construction to multi-view capture and all-round 3D model integration. A number of law enforcement, medical and commercial applications are described briefly including prisoner 3D face models, maxillofacial and orofacial cleft assessment, breast imaging and foot scanning. Ongoing research in real-time capture and processing, and model construction from naturally illuminated image sources is also outlined.

Electronic access

The research register for this journal is available at
http://www.mcbup.com/research_registers/aa.asp

The current issue and full text archive of this journal is available at
http://www.emerald-library.com

Sensor Review
Volume 20 · Number 3 · 2000 · pp. 218–226
 · ISSN 0260-2288

1 Introduction

3D image sensing[1] is rapidly coming of age, maturing from a laboratory curiosity to being a driving force behind new and exciting applications. 3D imaging devices promise to open a very wide variety of applications, particularly, those involving a need to know the precise 3D shape of the human body, e.g. e-commerce (clothing), medicine (assessment, audit, diagnosis and planning), anthropometry (vehicle design), post-production (virtual actors) and industrial design (workspace design). In short, any situation where you want to get the 3D shape of a person into the computer.

In digital form, all manner of further analysis becomes possible, allowing entire processes to be conducted that start with the human as input, e.g. entire shoe design and custom fitting starting from a digitised last or foot and ending with the machine tool instructions for cutting the upper and sole. Anthropometric data is required in many such areas of manufacture to provide information for the design of products such as clothing, safety equipment, furniture, vehicles and any other objects with which people interact. 3D human body images are also increasingly used in a range of applications within the creative media sector. However, the pliant nature of human flesh, combined with the complex surface form of the human body, dictates that the measuring technique employed is a non-contact one that can produce large amounts of dense 3D data.

A variety of non-contact, optically-based 3D data acquisition techniques have been developed in recent years that can be applied to the imaging of the human body. It is now possible to obtain commercial off-the-shelf devices based on any one of a wide variety of underlying triangulation-based 3D sensing techniques that employ: laser-camera

The authors would like to thank Dr Colin Urquhart for the use of his figures in this paper and acknowledge research funding support from: the Engineering and Physical Sciences Research Council, the Chief Scientist Office (Scotland), the UK National Lottery Fund, the Cleft Lip and Palate Association, the Scottish Higher Education Funding Council, the UK Department of Trade and Industry and the Defence Evaluation Research Agency.
Received 31 March 2000
Accepted 2 May 2000

baselines, camera-projector baselines and/or camera-camera baselines (Figures 1a, 1b and 1c). Such approaches include laser scanning (Addleman and Addleman, 1985),

Figure 1a Laser-camera baseline configuration

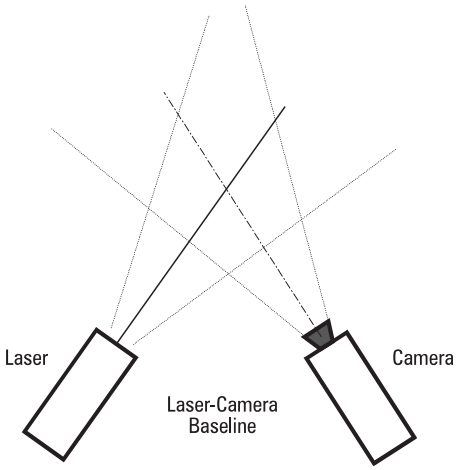


Figure 1b Projector-camera baseline configuration

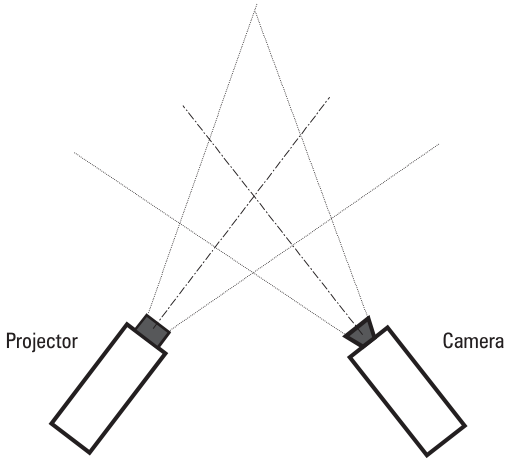
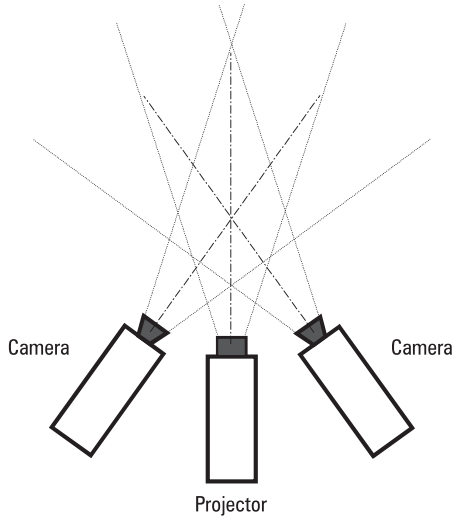


Figure 1c Camera-projector-camera baseline configuration



phase-stepping moiré fringe contouring (Reid *et al.,* 1984), phase measuring profilometry (Halioua and Liu, 1989) and structured light (stripe sequences, or temporal modulation (Sato and Inokuchi, 1987), stereo scene coding (McDonald *et al.,* 1993) and speckle texture projection (Siebert and Urquhart, 1994) respectively. Each method has merit in its application to human body imaging, however, the question remains as to which 3D-imaging device to choose for which particular application. In this paper the authors describe a 3D sensing technique, known as C3D, that has been developed over a ten-year period by the Turing Institute (principally Dr Joseph Jin, Dr Tim Niblett and Dr Colin Urquhart). C3D is based on white light speckle texture projection photogrammetry and is particularly suited to human body imaging in general.

2 Background

C3D was chosen by the 3D-MATIC Faraday Partnership (see[2]) for our research work on human body imaging as it affords a number of advantages over other non-contact optical measurement techniques. The geometric simplicity of the capture hardware and well-understood calibration methods employed ensure that the accuracy can be superior to more optically complex techniques which can be more difficult to engineer. Perhaps more importantly, the technique is also a full field one, producing three-dimensional information from the whole scene without the need for scanning. In applications where data acquisition speed is critical, such as in the measurement of live subjects, this can be fundamental to the success of the data acquisition technique.

Human beings, particularly babies, infants and the elderly, cannot remain perfectly still for more than a few seconds. In photography, we can see the effects of this when we attempt to take photographs of people using a slow shutter speed. The longer the exposure (or data capture time), the more unreliable or blurred the data becomes. This is especially true if we wish to make measurements to sub-millimetre resolutions, as in facial imaging or anthropometry. Because the stereo photogrammetry technique is a full field one, it is, therefore, an inherently faster and consequently more suitable data capture method than non full field techniques such as

laser scanning triangulation. It also possesses an advantage over other full field techniques, such as phase-stepping moiré fringe contouring, phase measuring profilometry and temporal light modulation, in that only a single image is required per camera.

The advantages of stereo photogrammetry can be summarised as follows:

- fast data capture;
- easy to calibrate;
- high measurement accuracy better than $\pm$0.5mm root mean square error (RMSE) on the face and $\pm$2.0mm on the body using TV resolution cameras;
- inexpensive hardware;
- inherently reliable with no moving parts;
- software principal component;

Limitations of the technique are considered to be:

- measurement resolution dependent on cameras;
- large number of cameras required for full coverage difficult to integrate;
- texture illumination required for featureless body surfaces;
- computer intensive data analysis;

The above limitations are largely centred on three factors:

(1) digital camera technology;
(2) computer technology; and
(3) flash projection technology.

Digital camera technology is rapidly maturing to deliver adequately high-resolution cameras to support digital photogrammetry at consumer prices. However, despite the advent of multi-drop serial bus peripheral connection technologies such as universal serial bus (USB) and IEEE1394 (Firewire), integration of many cameras to a single host computer remains an issue. (In the experience of the authors, consumer camera manufacturers have not yet awakened to the possibilities of connecting more than one camera at a time to a PC hence host driver and SDK (software development kit) support remains, for the time being, in short supply for multi-camera integration.) Computer power continues to rise inexorably and today's intensive computation issues will fade with the availability of tomorrow's 64-bit desktop PCs. Unfortunately, projection systems required to generate special illumination for photogrammetric imaging systems are not generally available as off-the-shelf items and require to be specially fabricated, usually by coupling pattern projection optics to a conventional professional flash unit.

3 How it works

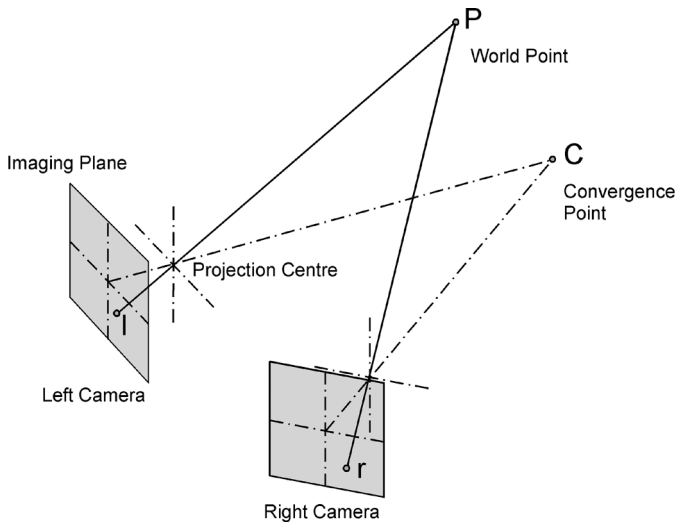
3.1 Triangulation and stereo

C3D relies on camera-camera base line triangulation (Figure 1c) to perform depth sensing. Most people have two eyes and can perceive depth cues from the slight parallaxes (or disparities) present between the views observed in each eye. C3D takes these same parallaxes, as captured by a stereo pair of cameras, and decodes them to produce an explicit depth map. Convergent geometry has been adopted to maximise the depth coverage obtained over the field of view of the cameras. As can be seen from Figure 2, a point P in space will project to two slightly different locations on the imaging plane of each camera and the difference in location is termed parallax or disparity. This disparity increases as the imaged point P in the world is translated further in the depth axis from the convergence point C of the camera stereo-pair (the sign of the disparity is reversed depending on whether the imaged world point lies in front or behind C). Accordingly, the magnitude and sign of the disparity values can be decoded to produce depth values if the geometry of the camera configuration is known (this is found by means of calibration).

3.2 Stereo matching

The tricky part is to determine for each point imaged in the left camera, the corresponding

Figure 2 Converged stereo imaging geometry



point in the right camera. C3D adopts a patented algorithm (Zhengping, 1988) based on multi-resolution image correlation-based image matching. In outline, this algorithm deconstructs the input images into a pair of spatially band-pass filtered "image pyramids". It then proceeds by stepping a small "reference" image window centred over each pixel of the left pyramid at the coarsest resolution (Figure 3) and using image correlation to find the best match location in the coarsest resolution image of the right pyramid. For each reference window location in the left image a search is carried out around the same location in the right image by computing the correlation score (goodness of match metric) between the reference window and a "test" window placed over neighbouring pixel locations. This local search is shown schematically in Figure 4. In fact, since the search process only steps the correlation search window in integer (pixel) steps, the coordinates and correlation scores of the best-matching locations are processed to compute an "ideal" best-match location to sub-pixel accuracy by interpolation.

Figure 3 Scale-space stereo-matching using image pyramids

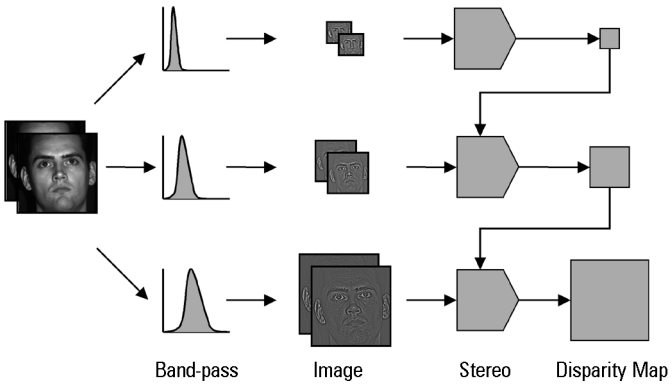
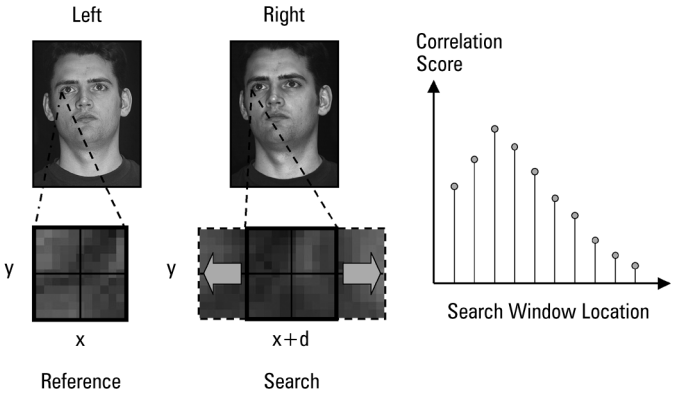


Figure 4 Correlation-based image matching



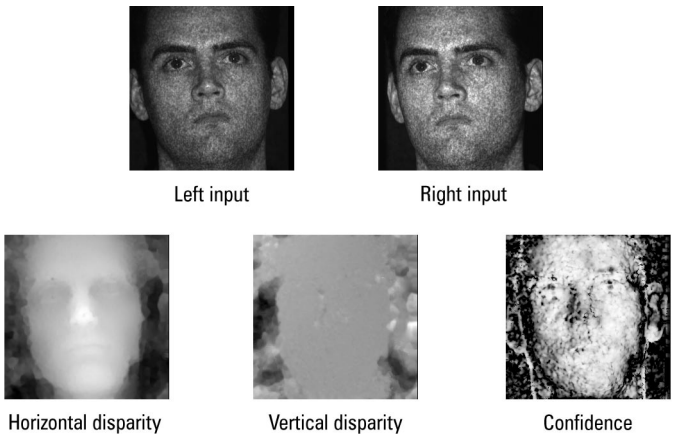
The above process is repeated for every pixel in the left image until a coarse disparity (x,y) offset map has been computed. This coarse map is expanded (also by interpolation) to match the size of the next higher level resolution images of the pyramids and matching proceeds once more. In this case, however, searching begins at the location pointed to by the disparity map (also scaled in magnitude to take account of the expansion of the disparity map). As the images are becoming finer and finer in resolution as the match process proceeds down to the base of the pyramids, the disparity values are also refined. Hence this is termed a coarse-to-fine matching strategy that employs scale-space tracing, *scale-space* referring to the search over multiple image scales.

The output of the above process (Figure 5) is an (x,y) disparity map whose values map each pixel coordinate in the left image of the input stereo-pair to the coordinate of the corresponding pixel in the right image of the stereo-pair. In addition, a confidence map is also computed that gives an indication of the reliability of each match value. There are unavoidable instances where a point could be observed in the left image but be occluded in the right image (and vice versa). Such occlusions lead to bad matches, but can be detected by means of the confidence map in conjunction with other image processing techniques.

3.3 Image capture

A further obstacle must be overcome to ensure that reliable stereo-matches are

Figure 5 Input stereo-pairs (speckle texture projection illumination) and output disparity and confidence maps

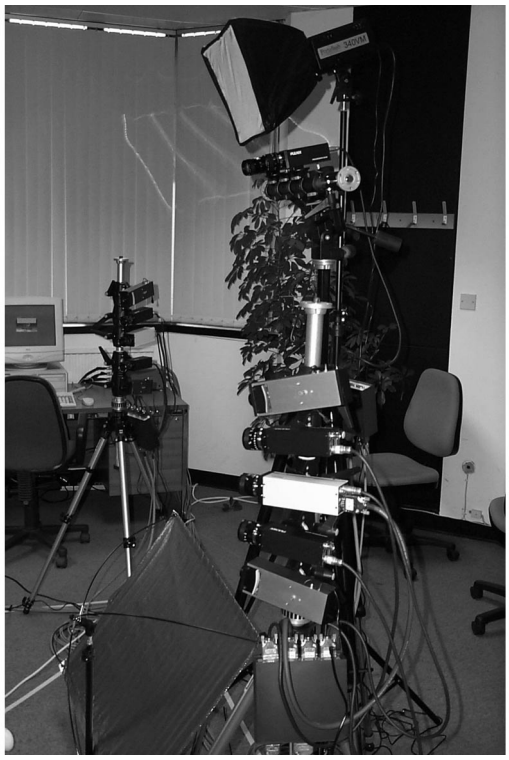


Sensor Review
Volume 20 · Number 3 · 2000 · 218–226

produced where the imaged scene contains bland untextured surfaces. Under such conditions, when imaging a plain white wall for example, the pixels in both left and right images will appear almost uniformly white and hence indistinguishable from each other. The correlation scores computed for each "trial" location will appear similar and the true correspondences will not be apparent. This situation can easily be avoided by simply projecting a random texture pattern (random such that it does not repeat and produce false matches) onto the scene, thereby ensuring that there is always a unique pattern on which to correlate in the field of view of the cameras.

Figure 6 shows a typical image "Pod" that combines image capture and scene illumination functions. Notice that three cameras have been employed. Two monochrome cameras serve to form a stereo baseline and are synchronised to capture images illuminated by special texture flash projectors. A third central colour camera is synchronised to capture the natural photographic appearance of the subject under normal white-light flash, a few tens of milliseconds after the flash texture stereo capture.

Figure 6 Detail of a capture pod (foreground) – monochrome cameras in black, colour camera in white, texture flashes in blue and white-light flash (black diffuser hood)

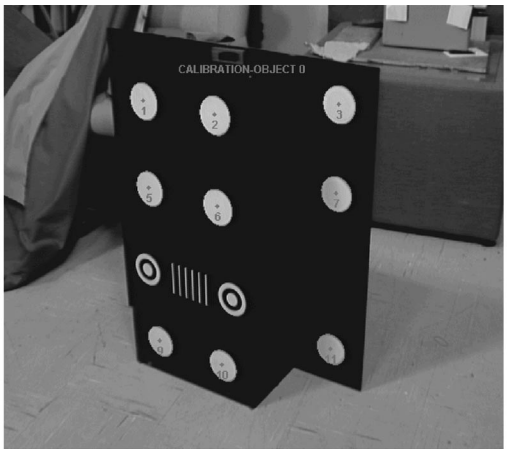


3.4 Calibration

In order that the detailed geometric configuration of all of the cameras can be determined (monochrome stereo-pair and single colour camera), a calibration process has been developed based on photogrammeteric techniques. A calibration target (Figure 7) comprising discs on contrasting background and of accurately known dimensions and location is presented and captured by the cameras for a variety of target poses. Images of the target from all the cameras are processed to find the central location of the discs and these coordinates are used to fit an approximate geometric model of each camera and its respective relative orientation to the target. The direct linear transform (DLT) (Abdel-Aziz and Karara, 1971) is used for this purpose. This approximate model is used to bootstrap a much more accurate model of the cameras. Essentially a full projection model is adopted that contains N parameters corresponding to N physical quantities such as: sensor pixel pitch, lens focal length, camera baseline and, importantly, the *principal point* on each imaging plane (where the projective centre of each camera projects on to each respective imaging plane). The act of computing these intrinsic camera parameters and also the extrinsic (relative camera orientation) parameters is termed *space resection* by photogrammeterists and the general algorithmic approach is contained within a process known as *bundle adjustment* (Karara, 1989).

Since the full geometry of the camera system has been deduced by calibration, the disparities computed through stereo matching

Figure 7 A calibration target with target circles located automatically



can be used to project a notional ray from each corresponding pair of pixels in the left and right stereo-pairs and their intersection in 3D space can be computed, i.e. range values recovered from disparities. This process is termed *space intersection* and results in the computation of a *point cloud* in X, Y, Z space.

3.5 Model construction and integration

The point cloud captured by a single stereo-pair of cameras comprises only 2.5D information, i.e. no undercuts. This point cloud can easily be re-sampled or warped (*ortho-rectified*) to fit a regularly spaced grid in X,Y and a triangulated mesh constructed from the Z values (Figure 8). The entire process for single-pod capture is shown in Figure 9.

For true 3D information capture, for example all round a human head or even body, the 2.5D point clouds recovered from multiple pods imaging the subject can be integrated together. Figure 10 shows the configuration of a notional whole body imager design and Figure 11 shows the output of a system used for capturing the human head that integrates five pods together (four equally spaced around the head viewing it side on and one viewing the top of the head). Integration is achieved by discovering through calibration the relative orientation of each pod with respect to a calibration target. This process enables the point cloud captured from each pod to be transformed into the same coordinate frame. Each point cloud recovered is stored in an efficient volumetric data structure (an octree).

A implicit surface is computed that merges together the point clouds into a single triangulated polygon mesh using a variant of the well known marching cubes algorithm (Lorensen and Cline, 1987). This mesh is then further decimated to any arbitrarily low resolution for display purposes.

Figure 8 Face mesh recovered by C3D. Decimated polygon mesh (left) and photorealistic rendered model (right)

Figure 9 Basic single-pod C3D process

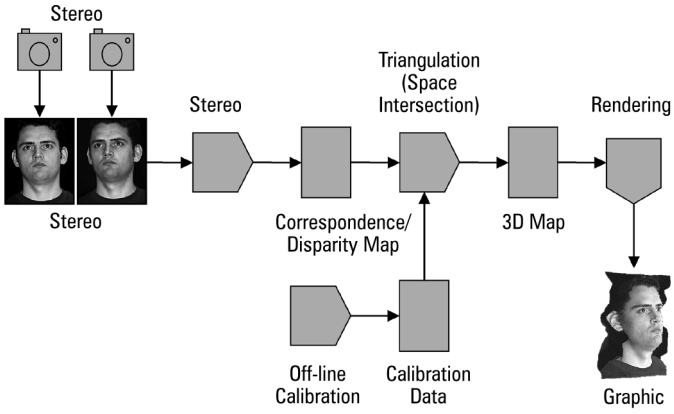
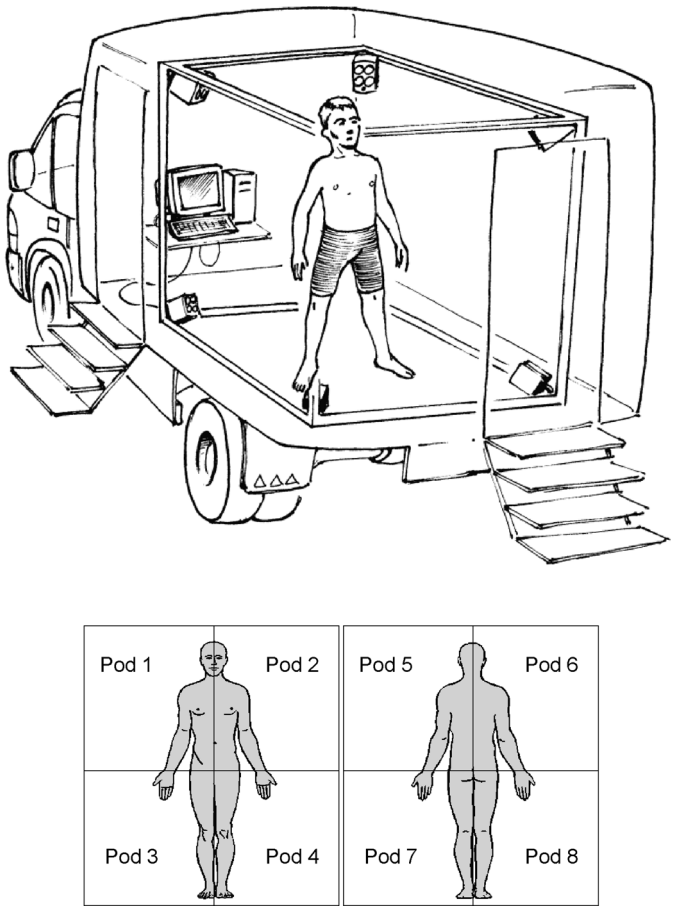


Figure 10 An all-round whole-body 3D imager configuration (above) with corresponding fields-of-view for each pod (below)



Sensor Review
Volume 20 · Number 3 · 2000 · 218–226

Figure 11 Head models generated by a prototype five pod all-round 3D imager

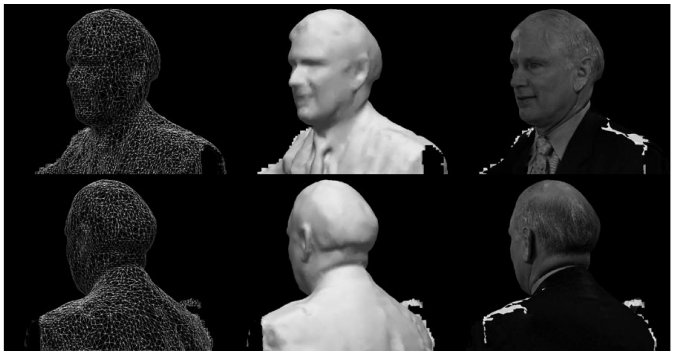
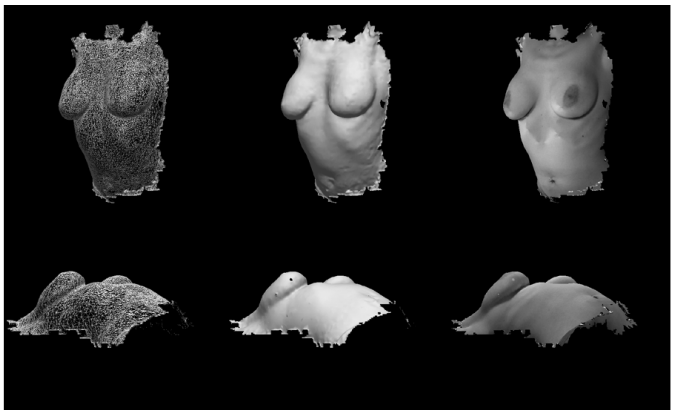


Figure 12 Experimental breast data captured using a four pod C3D system



4 Systems and applications

Many applications have been investigated and trialled over the past decade using this technology. The first established area was that of 3D facial imaging. A significant fraction of the development work of C3D has been funded by the UK Home Office with a view to developing 3D prisoner face capture to investigate replacing standard police station "mug shots" with 3D face models. The rationale being that a 3D face model implicitly encodes many 2D face poses and may better assist a witness in the identification of a suspect. Facial imaging continues to be an important 3D imaging application, particularly with regard to craniofacial, maxillofacial and plastic surgery applications. A twin pod C3D facial imaging system has been in operation at Canniesburn Hospital Maxillofacial Unit in Glasgow since 1996 to support research in this area (Ayoub *et al.*, 1996; Ayoub *et al.*, 1998). Currently, twin pod C3D systems have been deployed in Glasgow Dental Hospital and School and Yorkhill Hopital for Sick Children. These systems provide 3D facial data capture for projects investigating methods for assessment and surgical planning related to oro-facial clefts in a collaboration between Glasgow Dental Hospital, the Statistics Department at Glasgow University and 3D-MATIC.

A five pod experimental system was constructed in 1998 in collaboration with the UK National Film and Television School's CREATEC research group. This system has been deployed at the Ealing Studios, London, and is used to capture whole heads for virtual actor research. During commissioning in Glasgow, this system was configured to capture 3D breast data (Figure 12) to support work on the development of breast surgical planning tools in collaborations with Ninewells Hospital, Dundee, and also Glasgow Dental Hospital and School.

The first fully-fledged commercial application incorporating C3D is a foot scanner incorporating five pods currently being developed by Shoemaster International Ltd. An earlier prototype 3D foot scanner was constructed and demonstrated by the Turing Institute and Dundee Royal Infirmary Footpressure Laboratory.

Currently, a Glasgow University spin-out company, C3D Limited, is commercialising the C3D technology and further details can be obtained from: www.c3d.com. At least two other companies have developed similar speckle texture projection photogrammetry systems in the UK. An early system was demonstrated by Thorn EMI, though this does not appear to be commercially available. In London, Tricorder Technology plc have developed a range of systems and back-end application suites for specific medical applications, such as breast surgery. More details of Tricorder's systems and products can be found at: www.tricorder.com. The 3D-MATIC Faraday Partnership exists to promote 3D sensing technology and applications in general and further details of current and ongoing research, development, services and publications are available at: www.faraday.org.

5 Continuing research and development

The principal thrust of ongoing research at 3D-MATIC is in real-time 3D capture and

processing, and also texture free model recovery (natural illumination) from stereo-pair images and movie sequences. In a joint research infrastructure development project with Edinburgh Virtual Environment Centre (EdVEC), Edinburgh University, we are investigating high speed/high resolution all-round body capture techniques and applications. Work is currently progressing (Dr Paul Cockshott and Dr Ming Zhou) to construct a whole body all-round 3D imager that will comprise six to ten pods and be capable of collecting up to 20 second bursts of real-time (30 frames/second) stereo-pair and natural colour image capture. This 2D data will be processed to produce a 30 frames/second "3D model movie" sequence, i.e. capture 600 3D whole body models! A key challenge is to be able to represent, store, manipulate and display the huge quantities of data collected by all-round real-time 3D sequence capture. The prize is the ability to gather accurate 3D data that describes dynamic effects such as visco-elastic soft tissue deformations, e.g. muscle flexure or deformations through the gait cycle.

An even more considerable challenge is to be able to extract 3D models from stereo-pair images, or movie sequences, without resorting to bathing the scene in speckle illumination. A number of approaches are being investigated by various research groups. For example, direct fitting of explicit body models to captured imagery has been investigated at Surrey University (Hilton *et al.*, 1999) and the integration of shading cues with stereo matching and also generalised space-carving approaches are being investigated here in Glasgow (Dr Joseph Jin and Dr John Patterson).

6 Conclusions

3-D human body imaging applications are an important sector of the 3-D image capture market. By affording quasi-instantaneous capture to deliver accurate 3D photorealistic body models, the speckle texture projection stereo-photogrammetry approach is uniquely suited to imaging the human form. Furthermore, as the principal ingredients of this approach are simply cameras and software, it also has the potential to become a very low-cost technology. Mainstream status will likely be achieved when low-cost digital cameras can be integrated to capture in synchronism and communicate with a single host computer to allow affordable whole body 3D imaging systems to be constructed. Availability of off-the-shelf pattern projection systems remains an outstanding issue, although technologies such as holographic projection gratings illuminated by pulsed laser diodes might provide a compact and reasonably inexpensive solution. The consumer LC projector could also serve as the pattern projection source, but the cost of these devices is prohibitive for many whole-body all-round 3D imaging applications, requiring perhaps up to ten projectors.

With commercial systems based on speckle texture projection stereophotogrammetry now available and ongoing R&D promising real-time, low-cost and flexible 3D imaging (in terms of imagery sources), there is an exciting future for new and creative applications based on 3D human body data and models.

Notes

1 i.e. a means of reconstructing the three-dimensional structure of the world as imaged within the field of view, a non-contact optics-based sensing device.

2 3D-MATIC (3D Multi-Media And Technology Integration Centre) is an Engineering and Physical Sciences Research Council (EPSRC) sponsored Faraday Partnership based at the Department of Computing Science, University of Glasgow. It is one of four such partnerships in the UK focusing on improving the interaction between the UK research base and industry. The purpose of 3D-MATIC is to address the needs of small to medium-sized enterprises (SMEs) wishing to gain access to advanced 3D non-contact digitisation techniques by connecting industrialists with research groups and researchers working in that field.

References

Abdel-Aziz, Y.F. and Karara, N.M. (1971), "Direct linear transformation from comparator coordinates into object coordinates in close-range photogrammetry", *Proc. ASP Symposium on Close-Range Photogrammetry*, Illinois, January, pp. 1-18.

Addleman, D. and Addleman, L. (1985), "Rapid 3D digitising: an innovative technique holds promise for a variety of applications", *Comp. Graph. World*, Vol. 11, pp. 41-4.

Ayoub, A.F., Siebert, P., Moos, K. and Wray, D. (1998), "A vision-based three dimensional capture system for maxillofacial assessment and surgical planning", *Brit. J. Oral Maxillofacial Surg.*, Vol. 36, pp. 353-7.

Ayoub, A.F., Wray, D., Moos, K.F., Siebert, J.P., Jin, J., Niblett, T.B., Urquhart, C.W. and Mowforth, P.H. (1996), "Three-dimensional modelling for modern diagnosis and planning in maxillofacial surgery", *J. Orthognathic and Orthodontic Surgery*, Vol. 11, pp. 225-33.

Halioua, M. and Liu, H. (1989), "Optical three-dimensional sensing by phase measuring profilometry", *Opt. and Lasers in Eng.*, Vol. 11, pp. 185-215.

Hilton, A., Beresford, D., Gentils, T., Smith, R. and Sun, W. (1999), "Virtual people: capturing human models to populate virtual worlds", *Proc. IEEE Computer Animation*, 1999.

Karara, H.M. (1989) (Ed.), *Handbook of Non-Topographic Photogrammetry*, 2nd ed., American Society for Photogrammetry and Remote Sensing, Falls Church, VA.

Lorensen, W.E. and Cline, H.E. (1987), "Marching cubes: a high resolution 3D surface construction algorithm", *ACM Computer Graphics*, Vol. 21.

McDonald, J.P., Siebert, J.P. and Fryer, R.J. (1993), "Stereo scene coding using SLM active illumination", *Proc. 26th International Symposium on Automotive Technology and Automation*, Mechatronics Conference, Aachen, Germany, September 1993, pp. 169-76.

Reid, G.T., Rixon, R.C. and Messer, H.I. (1984), "Absolute and comparative measurements of three-dimensional shape by phase measuring moiré topography", *Opt. and Laser Technology*, Vol. 16, pp. 315-19.

Sato, K. and Inokuchi, S. (1987), "Range-imaging system utilizing nematic liquid crystal mask", *Proc. IEEE International Conference on Computer Vision*, London, June, pp. 657-61.

Siebert, J.P. and Urquhart, C.W. (1994), "C3D: a novel vision-based 3D data acquisition system", in the *Proceedings of the Mona Lisa European Workshop*, Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, Germany, 23-24 August, 1994.

Zhengping, J. (1988), "On the multi-scale iconic representation for low-level computer vision systems", PhD thesis, The Turing Institute and The University of Strathclyde.