Stroboscopic stereo rangefinder

Jean-Christophe Nebel, Francisco J. Rodriguez-Miguel and W. Paul Cockshott
3D-MATIC Research Laboratory
University of Glasgow
Glasgow, Scotland, UK
{jc,francisco,wpc}@dcs.gla.ac.uk

Abstract

The Michelangelo dynamic 3D scanner uses stereo rangefinding along with strobe illumination to capture 3D information at 25 frames per second. The configuration allows rapid motion within the capture volume to be frozen into a series of virtual sculptures. Textured strobe illumination is used for range data and plain strobe illumination for colour data. We present examples of data captured with the system along with measurements of the tolerances attained in the measurements.

1. Introduction

The Faraday Laboratory at the University of Glasgow is currently developing a dynamic 3D whole body scanner.

The basic concept is to equip a studio space such that the "working volume" of the space is imaged from all directions using fixed stereo-pairs of TV cameras [4][8]. The stereo-pair images collected by the camera pairs is then processed using photogrammetric techniques developed by the Turing Institute, Glasgow, to create a spatio-temporal 3D model of this space. So now we have a full 3D model of all the action (being up-dated in real-time), that can be viewed from any direction, are able to track all of the action and also compress this action within a data structure that accommodates information about the objects in this 3D space and how they change over time a true 3D movie.

The concept sounds radical, but is based on the culmination of over a decade of research into 3D imaging by Turing Institute and now in collaboration within the 3D-MATIC Faraday Partnership (operating within the Computing Science Department at Glasgow University). Turing developed 3D imaging software[11][7], called C3D, that processes stereo pair images to generate 3D models. In fact, C3D automates the processing of several stereo-pair views of a subject into a set of 3D views which are then integrated into a single model.

1.1. Related work

The most effective techniques for generating 3D static photo-realistic models of real human are called scanning techniques. Several methods can be used: laser beams [6] and Cyberware^{TM} [1], structured light technique [10] or photogrammetry [4][8]. Their accuracy is usually sufficient for getting very realistic 3D models, whose accuracy is about 1mm. Moreover colour pictures are mapped on these models what ensures photo realistic appearance.

The main difference between the results these full body scanners provide is about the type of data they can capture. Indeed very few of them have short capture time. The commercial scanners, based on laser beams and structured light, have a capture time of about 15 seconds, whereas the ones using photogrammetry only need few milliseconds. Obviously only the later type of scanners has the potential of capturing moving subjects. A part from the team presenting this paper, we know only one other research team whose aim is the generation of true 3D movies using scanning technology: they are part of the British company TCTi^{TM} [5] and have not published any result yet.

Other works are focused on getting sequences of 3D data of moving objects. The Robotics Institute of Carnegie Mellon University has an interest of capturing and analysing 3D human motion. For that purpose they built a "3D room" which is a facility for 4D digitisation: a large number of synchronised video cameras (49 according to [2]) are mounted on the wall and ceiling of the room. However they do not have reported the generation of any 3D models using their "3D room", they seem to be more interested in analysing motion [9]. It is also worth mentioning the work by [3]. They designed a colour encoded structured light range-finder capable of measuring the shape of time-varying or moving surfaces. Their main application was about measuring the shape of the human mouth during continuous speech sampled at 50Hz. Since their system is based on the continuous projection of a colour encoded structured light, their technics has some limitations com-

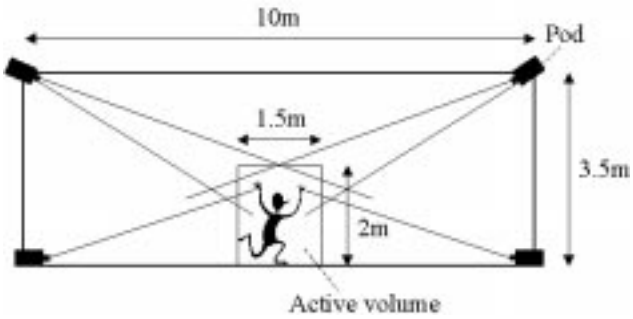


Figure 1. Proposed Layout of the Scanner

Figure 2. A pod

pared to ours. Their structure light can only be projected from a single direction, therefore they cannot get a full coverage of 3D objects. Moreover the capture of the texture of 3D objects is not possible.

1.2. Configuration

The intended configuration of the scanner is shown in figure 1. The subject will be imaged by a total of 24 TV cameras arranged in threes. We term a group of three cameras a pod. Eight pods, arranged at the corners of a parallelipiped will image the active volume.

The process of 3D capture relies upon flash stereo photogrametry. Each pod (see fig 2) has one JAI CV-M70 colour and two Sony XC55 black and white cameras. Associated with each pod are two strobe lamps, one of which is a flood strobe, the other is fitted within a modified overhead projector.

Associated with each pod is a PC with two frame grabber cards: a Correco Viper RBG card for the colour camera, and a Correco Viper Quad card for the black and white cameras. Each pod is supplied with a feed from a master sync signal operating at 25Hz.

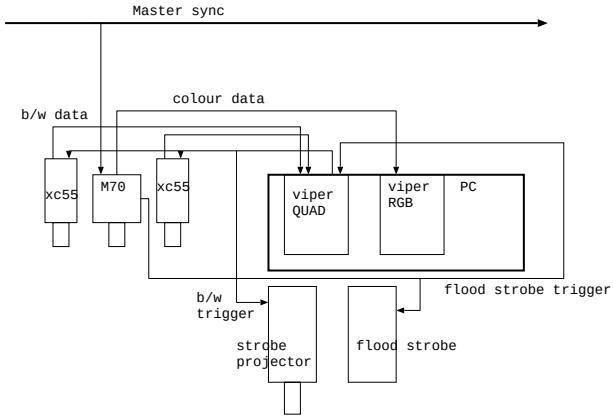


Figure 3. A interconnection of pod components

The sequence of stages required to capture image and range data with a pod is :

1. The M70 accepts a master sync pulse and opens its shutter.

2. The M70 then triggers the flood strobe which provides a $7\mu s$ flash.

3. The flood strobe trigger is input to the Viper Quad framegrabber which, after a delay of $100\mu s$ triggers the monochrome cameras and the strobe projector. The delay is sufficient to allow the colour camera to have closed its shutter.

4. Images are then downloaded from all cameras to the frame grabbers before the next master sync pulse.

5. The frame grabbers DMA the images over the PCI bus into main memory on the PC.

The PCs have 1GB of main memory allowing over 500 frames to be captured in a single sequence. Once captured, the data is written to file for processing by the ranging software.

The monochrome images of the subject illuminated with a random dot pattern are used for stereo range finding. The colour images illuminated with uniform white light are used to capture the surface appearance of the subject. The total capture time for the monochrome + colour images is under $150\mu s$. This is sufficient to ensure that even a subject moving at 10M/s will have moved less than 2mm during the capture process. The equipment is thus suitable for the capture of very dynamic actions, e.g., martial arts katas.

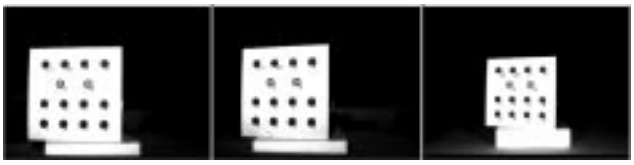


Figure 4. Calibration target viewed by one pod

2. Data capture, processing and results

In order to build 3D models from the data captured by the scanner previously described, the cameras have to be calibrated, e.g. the detailed geometric configuration of all the cameras have to be known. It is done by localising target circles on a calibration target of accurately known dimensions [7] (see fig 4).

Once the capture has been done, the stereo matching process is applied to each stereo-pair images (see Table 1). The stereo matching algorithm we use is a patented algorithm [11] based on multi-resolution image correlation.

The algorithm takes as input a pair $[l, r]$ of monochrome images held as two dimensional arrays of 32 bit floating point numbers. It outputs a tripple $[x, y, p]$ of images again in 32 bit floating point format where x_{ij} specifies the horizontal displacement in pixels of pixel l_{ij} to the matched point in image r, y_{ij} specifies the vertical displacement in pixels of pixel l_{ij} to the matched point in image r, $0 < p_{ij} < 1$ specifies the correlation between the neighbouryhood around l_{ij} and the matched point in image r. We will denote the overall matcher as the function `match(l,r →x,y,p)`.

The matcher is implemented using a difference of gaussian image pyramid, and an inner matching function `innermatch(l,r →x,y,p)` with the same functional form as `match`.

The outer structure of the algorithm is

1. Construct difference of gaussian pyramids for the images l, r call these L, R where L_0 is the base of the l pyramid and L_t is the uppermost plane on the pyramid. The top layer of the pyramid is 16 by 12 pixels in size for a base of 640 by 480. We call c the current level of the pyramid.

2. Set $c \leftarrow t$.

3. Apply the function `innermatch` to L_c, R_c.

4. If $c = 0$ then exit.

5. Use the resulting x, y images to perform a pixel by pixel warp of the image R_{c-1}. Thus if the estimated disparities from matching at layer c were correct, image L_{c-1} would now be identical to R_{c-1}, occlusions permitting. To the extent that the estimated disparities were incorrect there will remain disparities that can be corrected at the next layer of the algorithm, using information from the next higher waveband in the images.

6. Decrement c.

7. Return to step 3.

The function `innermatch` proceeds as:

1. For each pixel position ij in l take a 5 by 5 neighbourhood centered on $l_{i,j}$ and computes its correlation with the corresponding neighbourhood in the image r and also with the neighbourhoods centered on $r_{i-1,j}$, $r_{i+1,j}$, $r_{i,j-1}$, $r_{i,j+1}$.

   (a) A polynomial is fitted through the 5 correlation values and the maximium of the polynomial is detemined.

   (b) If that is within the range of the initial search the relative coordinates of the peak are returned as $x_i j, y_i j$.

   (c) If the coordinates of the imputed correlation peak are outwith the search window, then the relative coordinates of the highest measured correlation are returned.

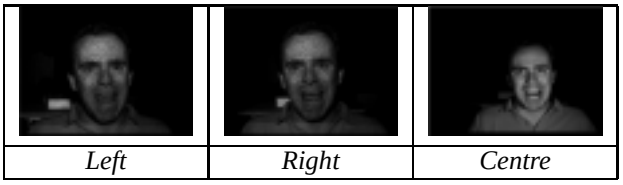
   Note that the coordinates will typically be fractional rather than integral numbers of pixels.

2. Having determined an imputed peak in the correlation function which may be on fractional pixel coordinates, the correlation function is recomputed using these coordinates to give p_{ij}.

3. Perform a smoothing operation on images x, y using a unique unitary kernel for each pixel position, the weights of which are derived from a normalisation of the corresponding neighbourhood in image p. This allows disparities whose values are more certain to bleed into the areas for which the disparity is less certain.

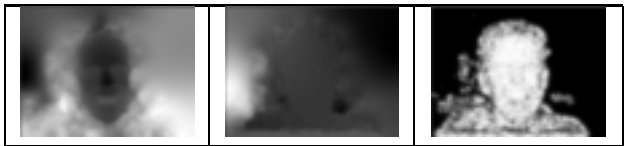
The outputs of the stereo matcher are disparity and correlation maps (see Table 2). These maps combined with the calibration file of the associated pod allow the generation of a range map, i.e. the map of the distances between each pixel and the coordinate system of the pod. The three-dimensional coordinates of those pixels are calculated using a pinhole-based model for the cameras.

Finally we deal with the triangulation of those points. We implemented a method which is based on the fact we digitise continuous objects. This can be expressed by the continuity of the disparity values within the disparity map. We



Left	*Right*	*Centre*

Table 1. Input images



Disparity maps (x,y)	*Correlation map*

Table 2. Output from the matcher

proceed as follows: given a point in the disparity map with a disparity d_1 we look for its neighbours within the disparity map and we only triangulate them if they have a disparity d_2 that satisfies :

$$d_1 - d_2 < Threshold \quad (1)$$

A threshold of 1 gives us the results which are shown in Table 3. The generation of photo-realistic models is achieved by mapping the colour pictures taken by the colour cameras to their corresponding triangulated meshes.

3. Accuracy of the system

Whereas there are many systems allowing 3D captures, very few of them provide figures about the accuracy of the data captured. Actually no methodology exists in the UK to test 3-dimensional capture systems in order to determine whether the data they produce is fit for purpose. In order to measure the precision of our dynamic 3D scanner, we captured 20 sets of frames of a 3D object whose dimensions are known with a very high precision: a ping-pong ball bouncing on a table (see Table 5, left column). Its diameter, at rest, was measured as being 37.63mm +-0.02.

Using the process described in the previous section, we built 20 3D models of this bouncing ball. Since we want to

Mesh	*Views of a 3D model*

Table 3. Face model

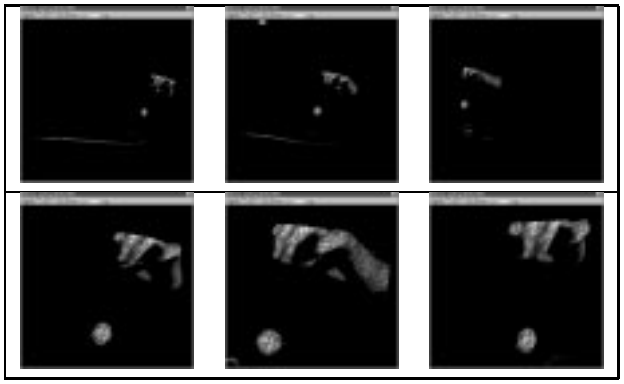


Table 4. Views of the ping-pong ball

fit the model of the ping-pong ball on spheres, we want to analyse only the 3D data of the points belonging to the ball.

Since the method used during the correspondence process generates a continuous disparity map, we need to distinguish between points that are within the object to digitise and points in the background. We have implemented an algorithm based in the variation of the pixel grey values for a given window, if this variation is lower than a given threshold we remove this from the disparity map.

From this filtered disparity map, we compute the three-dimensional coordinates of the points, the results are shown in Table 4, where are shown different views of the digitised scene.

Once the model is generated, we select interactively the volume of interest, i.e. the ping-pong ball, using a 3D box. In average the ball is defined by 2606 3D points.

Finally we deal with the fitting of these 3D data to a sphere using the following method.

Given a set of points (x_i, y_i, z_i) which should satisfy the following equation

$$(x - x_o)^2 + (y - y_o)^2 + (z - z_0)^2 = R^2$$

where (x_0, y_0, z_0) is the centre of the sphere and R is its radius, we compute the sphere variables we need to minimize the energy function:

$$E(x_0, y_0, z_0) = \sum_{i=1}^{m} (L_i - R)^2 \quad (2)$$

where $L_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}$ and m is the number of points we have.

To get the radius we can derive this expression with respect R to obtain:

$$\frac{\partial E}{\partial R} = -2 \sum_{i=1}^{m} (L_i - R)$$

and setting equal to zero

$$R = \frac{1}{m}\sum_{i=1} L_i$$

Proceeding the same way deriving equation 2 with respect x_0, y_0 and z_0 we get:

$$x_0 = \frac{1}{m}\sum_{i=1}^{m} x_i + (\frac{1}{m}\sum_{i=1}^{m} L_i) \times (\frac{1}{m}\sum_{i=1}^{m} \frac{x_0 - x_i}{L_i})$$

$$y_0 = \frac{1}{m}\sum_{i=1}^{m} y_i + (\frac{1}{m}\sum_{i=1}^{m} L_i) \times (\frac{1}{m}\sum_{i=1}^{m} \frac{y_0 - y_i}{L_i})$$

$$z_0 = \frac{1}{m}\sum_{i=1}^{m} z_i + (\frac{1}{m}\sum_{i=1}^{m} L_i) \times (\frac{1}{m}\sum_{i=1}^{m} \frac{z_0 - z_i}{L_i})$$

On Table 5, we show the 20 frames of the left camera which were used for building the 20 3D models. The central column shows the value of the radius and its standard deviation for each model (in mm). The last point represents the value of the average radius and its standard deviation for the full sequence of 20 frames. The black thick line shows the value of the radius of the ping-pong at rest (18.81mm).

If we analyse the different models separately, we see that the data is concistent since the standard deviation is less than 0.4mm, which is very close to the accuracy of the calibration target (0.3mm). Moreover most of the radius values are in a range of 1mm from the expected value, three values are out of range by more than 1mm, we will speak about them later. If we study the sequence as a whole, we see that the average value of the radius (19.03mm) is very close to the expected one (18.81mm), the standard deviation being of 0.5mm.

On Table 5, right column, we show the measured altitude of the centre of the ball (in cm). The curve corresponds to the expected motion of a ball bouncing.

If we look at the three values, which seem to be out of range, on the trajectory curve we notice that they correspond to positions just after bouncing. We suspect that the difference between the radius measured by the scanner and the radius of the ball at rest comes from the fact that the shape of the ball is quite distorted after bouncing! Further studies should allow us to validate this hypothesis.

4. Conclusion and future work

In this paper we described the dynamic 3D whole body scanner we are developing and we presented some results obtained with one pod. Moreover we measured the accuracy of the system by using a bouncing ping-pong ball. The accuracy of the system was measured as being of 0.2mm with a standard deviation of 0.5mm. That accuracy corresponds to the accuracy of the calibration target we use. Finally our scanner seemed to be sensitive enough to detect the deformation of a ping-pong ball after bouncing.

In the future, apart from building the whole body scanner, we will investigate more precisely this deformation effect by capturing balls made of different materials. Moreover we will work with more accurate calibration targets to improve the accuracy of the system.

References

[1] Cyberware. http://www.cyberware.com/.

[2] T. Kanade, H. Saito, and S. Vedula. The 3d room: Digitizing time-varying 3d events by synchronized multiple video streams. Technical Report CMU-RI-TR-98-34, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1998.

[3] T. P. Monks. Measuring the shape of time-varying objecs. In *PhD thesis, University of Southampton*, 1994.

[4] J. P. Siebert and S. J. Marshall. Human body 3d imaging by speckle texture projection photogrammetry. In *Sensor Review, 20(3),pp 218-226*, 2000.

[5] TCTi. http://www.tcti.com/.

[6] R. Trieb. 3d-body scanning for mass customized products - solutions and applications. In *International conference of numerisation 3D - Scanning*, 2000.

[7] C. W. Urquhart. The active stereo probe: the design and implementation of an active videometrics system. In *PhD thesis, The Turing Institute and The University of Glasgow*, 1997.

[8] G. Vareille. Full body 3d digitizer. In *International conference of numerisation 3D - Scanning*, 2000.

[9] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene flow. In *Proceedings of the 7th International Conference on Computer Vision*, September 1999.

[10] S. Winsborough. An insight into the design, manufacture and practical use of a 3d-body scanning system. In *International conference of numerisation 3D - Scanning*, 2000.

[11] J. Zhengping. On the multi-scale iconic representation for low-level computer vision systems. In *PhD thesis, The Turing Institute and The University of Strathclyde*, 1988.

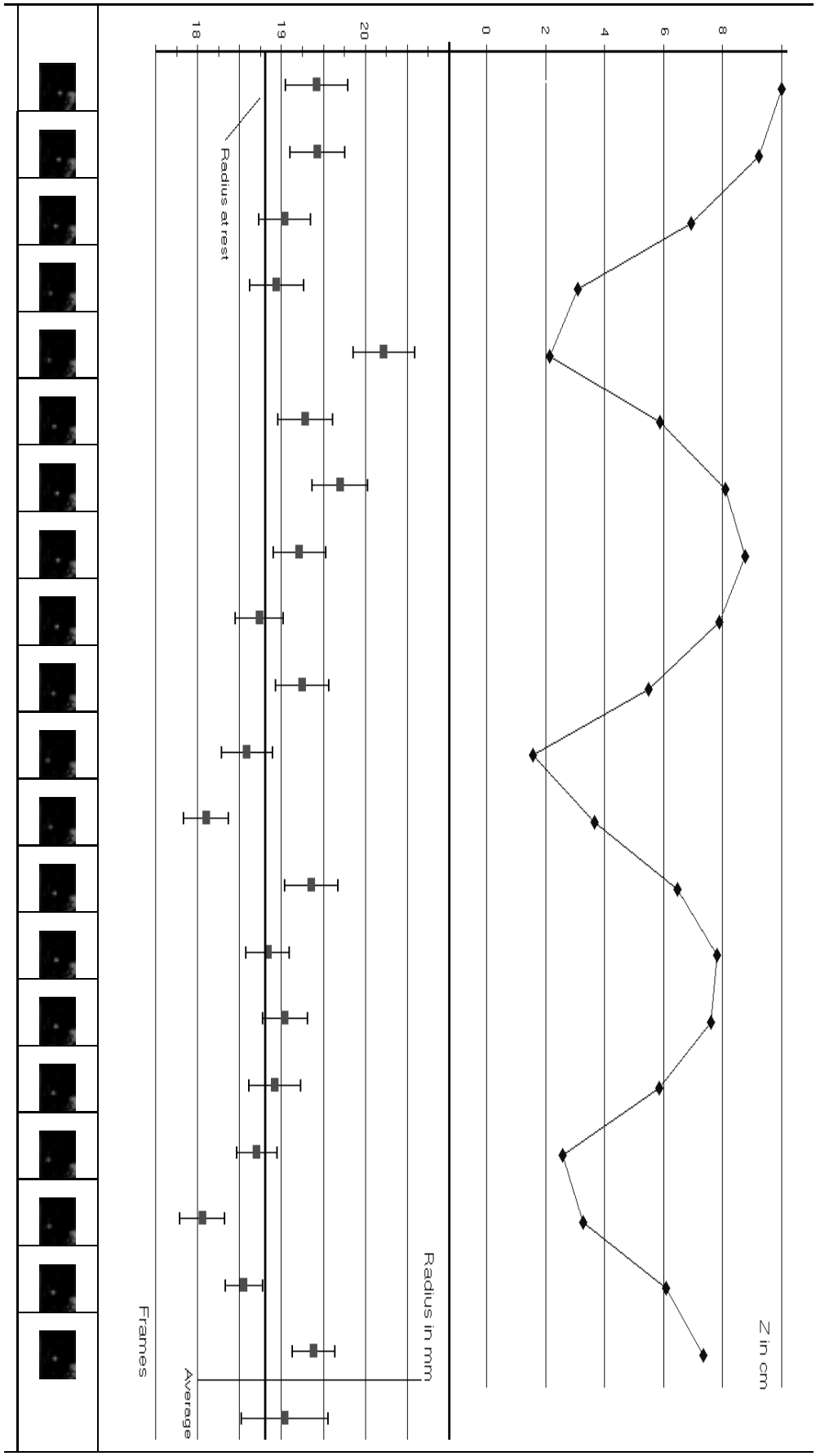


Table 5. Frames, ball radius and altitude