An Experimental 3D Digital TV Studio

W. P. Cockshott, S. Hoff, J.-C. Nebel
Imaging Faraday Partnership
Department of Computing Science
University of Glasgow
{wpc, svenja, jc}@dcs.gla.ac.uk

Contents

Abstract

The Michelangelo project at the University of Glasgow has developed an experimental three-dimensional television studio. This uses 24 video cameras and parallel computers to capture moving three-dimensional models of human actors. This allows the capture in real time of the appearance and three-dimensional positions of a human actor. It does this using stereo imaging techniques that have been under development at our laboratory for several years. The development of the studio has thrown up many technical problems which are still to be fully resolved, nonetheless it is already producing convincing animated sequences.

1 Introduction

The Michelangelo project at the Universities of Glasgow and Edinburgh was funded by the Scottish Executive and had as its objective the development of advanced 3D capture facilities for use by medical science [1], ergonomics, digital media and in the production of computer games. It involved the development by the University of Glasgow of two three-dimensional capture systems, one to be installed at each of the two Universities. The systems were intended to be able to capture complete models of the human body. The system to be installed at the University of Edinburgh was to be a still imaging system whose design would emphasise accuracy of measurement and completeness of capture [2]. The system to be installed at the University of Glasgow was to be designed for rapid, repeated high-speed 3D image capture. In this article we concentrate on describing the system installed at the University of Glasgow, which we term the dynamic scanner.

The basic concept of the Michelangelo dynamic scanner was 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. The stereo-pair images collected by the camera pairs are then processed using photogrammetric techniques developed by the Turing Institute [3], Glasgow, to create a spatio-temporal 3D model of this space. This would give a full 3D model of all the action (being up-dated in real-time), which can be viewed from any direction. It would also be possible to build 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 was based on the culmination of over a decade of research into 3D imaging by Turing Institute and subsequently by the 3D-MATIC Faraday Partnership (operating within the Computing Science Department at Glasgow University). Turing developed 3D imaging software [4] [5], called C3D, which 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. This software was to be used as the basis for both the Edinburgh and Glasgow image capture systems. Equipment previously developed by the Turing Institute had included an upper torso scanner developed for Ealing Studios, which used a collection of 15 digital cameras to build models of the upper bodies of subjects. The basic component of this earlier system was termed a Pod and comprised a group of 3 digital cameras, two of which were monochrome and one of which was colour. The monochrome cameras were used to extract range data and the colour camera was used to record skin, clothes and hair texture. The algorithm used to extract range information from the monochrome pair is described in more detail in section 9. The algorithm attempts to find the disparity between the positions of a given real world object as viewed from two cameras.

Given two images $\boldsymbol{A}$ and $\boldsymbol{B}$ it tries to find the maximally correlated local neighbourhood in image $\boldsymbol{B}$ for each neighbourhood in $\boldsymbol{A}$. Given a neighbourhood $\boldsymbol{a}$ in $\boldsymbol{A}$ there exists a correlation image $\boldsymbol{B\acute{}a}$ such that each pixel $\boldsymbol{b} \epsilon \boldsymbol{B\acute{}a}$ is the weighted local correlation coefficient between $\boldsymbol{a}$ and the neighbourhood in $\boldsymbol{B}$ cantered on $\boldsymbol{b}$. For the search algorithm to be reliable the correlation surface $\boldsymbol{B\acute{}a}$ should have a well-defined maximum $\boldsymbol{Tb}$ within the search window used. Around $\boldsymbol{Tb}$ the gradients of the correlation surface in both the horizontal and vertical directions should be steep. This cannot in general be expected for pairs of images of human skin taken under normal illumination conditions. Since skin is relatively featureless at the pixel spacings and digital precisions available using video cameras, the resulting correlation surfaces can lack well-defined maxima. The Turing scanners overcame this problem by projecting a random speckle image onto the subject. This gave a more rugged correlation surface, but potentially introduced artefacts into the pictures taken with the colour cameras. To obviate this flash photography was used for the colour images, with the flashlight being sufficiently bright to make the projected speckle pattern imperceptible. The opening of the shutters of the colour and monochrome cameras was stepped to ensure that the monochrome cameras did not see the flash.

2 State of the art

Lead by medical and military researches, 3D body scanners has begun to become a mature technology and many different techniques are used to generate 3D static photo-realistic models of real human [2], [6], [7], [8], [9] and [10]. 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. Most commercial scanners, based on laser beams and structured light, have a capture time of about 15 seconds, whereas the ones using one set of photographs only need few milliseconds. Obviously only the later type of scanners has the potential of capturing moving subjects.

Few research labs have developed such scanners. Monks designed a colour encoded structured light range-finder capable of measuring the shape of time-varying or moving surfaces, where structure light is continuously projected from a single direction [11]. Their main application was about measuring the shape of the human mouth during continuous speech sampled at 50Hz. The Lawrence Livermore National Laboratory has also developed a dynamic scanner using CCD video cameras for medical applications. Their technology dubbed "CyberSight" is also based on the projection of a specific pattern (grid) [12]. Although these 2 laboratories have successfully generated 3D sequences their technologies have some limitations compared to ours because there are based on the projection of a specific known pattern, which cannot be projected on an object from different directions. Therefore they cannot get a full coverage of 3D objects.

The work, which may be the closest to ours, has been developed at the Robotics Institute of Carnegie Mellon University. They are focused on getting sequences of 3D data of moving objects to analyse 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 [13]) are mounted on the wall and ceiling of the room. Although their main interest is motion analysis [14], they also generate crude 3D models. 3D shapes are reconstructed by using a volumetric method called Shape from Silhouette [15] that creates a voxel model that is then converted to surface model by using marching cubes and smoothed.

3 The apparatus

The apparatus developed for the Michelangelo dynamic scanner incorporates many of the design lessons learned from the earlier equipment developed by the Turing Institute. Range information is again obtained from stereo pairs of cameras using speckle projection. Again successive images are taken; a pair of synchronous range images is taken under speckle illumination followed by a texture image under uniform illumination. However the requirements of video capture meant that new challenges had to be met.

One option would have been to use pairs of colour cameras using these to capture both the range and the texture information. At video rates this proved impossible. It is necessary to capture 25 frames per second to obtain the illusion of smooth motion. If colour cameras had been used for both range and texture capture they would have had to have been operating at 50 full frames per second, which was impracticable. Instead the pods were designed with groups of 3 cameras two for range one for texture. The entire rig had 8 pods with the schematic arrangement shown in figure 1.

The earlier Turing scanners had used continuous speckle illumination, which was flooded out by flash when texture was captured. To have operated in this way would

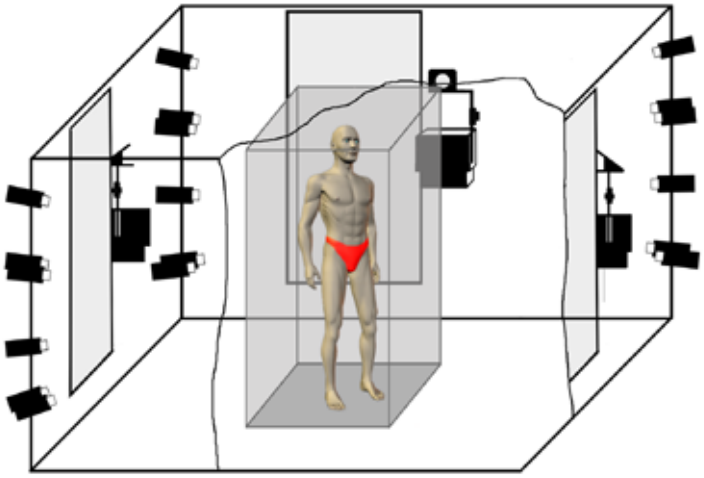

Figure 1: Layout of the cameras in the studio

have required the replacement of the flashes with flood strobes. It proved difficult to obtain strobes sufficiently bright to drown out projected speckles so the inverse arrangement was chosen. We use strobe illumination to provide the speckle and conventional studio lighting for the texture pictures. If the speckle was to be visible to the ranging cameras against the ambient light, their exposure times had to be very short and had to overlap with the strobe flashes. The very brief exposure times, we operate with times of about 400 μs, meant the use of cameras with very good light sensitivity. The whilst colour cameras could have been used throughout, the sensitivity of colour cameras is not as good as that of monochrome ones due to the loss of light in their colour filters. We therefore chose to use two monochrome and one colour cameras per pod.

Each pod (see figure 2) has one JAI CV-M70 colour and two Sony XC55 black and white cameras. Associated with each pod are strobe lamps fitted within a modified overhead projector. Associated with each pod is a PC with two frame grabber cards: a Coreco Viper RBG card for the colour camera, and a Coreco Viper Quad card for the black and white cameras. This allows colour and monochrome 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.

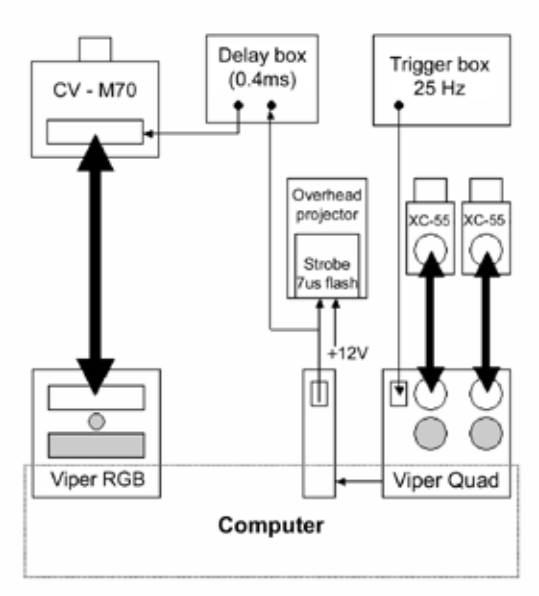

Figure 2: A interconnection of pod components

4 The optical capture process

To be able to extract proper range information from a black and white pair of cameras, some surfaces of an object need to be illuminated with a pattern, as mentioned earlier on, to add features on it. Otherwise, the matching software will fail to give an adequate ranging result. With the current software, the best results were achieved by projecting a binary random dot pattern so that its dots on the imaging chip respect the Nyquist limit.

The pattern is projected on to a three-dimensional object. Because of this the quality of the observed speckle image is dependent on whether the projector's depth of field is sufficient to maintain a sharp pattern on parts of the object that are outside of the focal plane, either because of dimensions or position. The depth of field is dependent on several factors including the size of the aperture, the size and distance of the light source to the lens and the distance the projector should be focused on. Given that it is only possible to focus in one plane, projecting on to an object with a certain depth results in image points with less sharpness, different sizes and loss of contrast for all planes that are not the focal plane. Objects or parts thereof in front of or behind the focal plane of the projector will have a defocused speckle pattern.

By reducing the aperture of a system, one can choose to use only ray bundles near to the optical axis, and thus with a relatively small angle in respect to the axis. This gives smaller angles on the imaging side of the system and so a larger depth of field

around the focal plane. This can be done by introducing a diaphragm to the system or by selecting the diameter of the projection lens mount. The obvious disadvantage is that stopping down the aperture by one stop, results in a halving of the total amount of light.

A lens with a long focal length will also help to increase the depth of field of a projector, since a long focal length lens produces smaller angles with respect to the optical axis than one with a short focal length.

Another important point is the size of the light source and its distance to the projecting lens. The best depth of field can be created by a point like light source. Unfortunately the strobe discharge tubes used are far from being point sources, as they measure about 2 by 3 cm. One could compensate by placing the tube far from the projecting lens so that it appears to the imaging system to be point like. Light losses following the inverse square law make this option infeasible.
With knowledge of all these factors, one has to define an acceptable compromise between depth of field and light output of the projecting system. In the case of the dynamic scanner, an industrial stroboscope as light source built into an ordinary overhead projector was the easiest and most inexpensive possibility. A mirror with an angle of 45° to the optical axis of the following lens system allows creating a reasonable distance to the first lens surface, here the Fresnel lens of the projector. The projecting lens has been replaced with a long focal one to obtain, as well, a larger depth of field.

Because of these constraints the active volume in the head scanner set-up is a 50 cm cube. The focus of the projected pattern is ideally set to the first third of the space in which the object is expected to move, since the acceptable depth of field are from 1/3 before the focus plane to 2/3 behind it.

The monochrome cameras in each pod, observing the same field of view, are mounted one on top of the other roughly 45 cm apart. This distance, called the camera baseline, is the basis of a triangulation based 3D capture method. To get reasonable results the baseline should be in a specific range. It shouldn't be to small otherwise small changes in distance to the camera couldn't be detected by camera and software. If the baseline becomes too big, occlusions will occur due to the fact that one of the cameras, observing a 3D object, could not see the same details than the other does. This commonly occurs around the nose, the chin and cheeks for a front view. The effect of such occlusions are poor stereo matching forcing interpolations, or in the worse cases holes in the model.

The colour camera is mounted in between the black and white cameras to ensure that it sees the same region of the active volume. The incandescent white lights providing the illumination for the colour cameras are set up behind white diffusion screens to

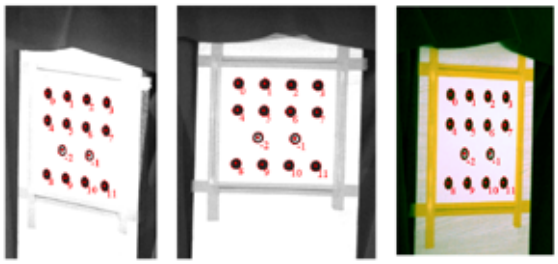

Figure 3: The calibration target with the circles detected by the software

reduce specular highlights. These have to be avoided as far as possible since the reconstructed 3D models may subsequently be placed in a virtual environment with different lighting conditions. The colour camera itself is set internally to a colour temperature of 5600 Kelvin (daylight), but the lights used create a temperature of 3600 Kelvin (tungsten) so colour correction by software has to take place to create a convincing colour image.

5 Calibration

Before the system can be used to capture 3D models, it has to be calibrated to extract the absolute intrinsic and extrinsic parameters of every camera that is used. That is needed because of the use of non-metric cameras. To do so a calibration target (see figure 3), consisting of 12 black circles and 2 black rings of accurately known dimensions and locations on a white background, has to be presented to all the cameras.

At least two cameras should see the target within every capture of the calibration process; otherwise, the position and orientation of some of the cameras may not be determined with respect to the others. Software then automatically locates the contour, centre and location on the target for every black circle. The algorithm detects zero-crossings to determine the contour. When all the contours are found an ellipse computation takes place to distinguish and delete all non-ellipsoid shapes, as well as compute the centre of the detected ellipsoids. To determine the orientation of the target itself the distances and positions of the circles to the black rings are calculated. With the outcome of this computation, it is possible to generate an approximate geometric model of each camera and its orientation to the target by using the direct linear transform (DLT) [16]. From this approximate model a much more accurate model can be computed. It is termed space resection by photogrammeterists and the general algorithmic approach is contained within a process known as bundle adjustment [17]. Calibration experiments have indicated the ranging accuracy to be 1/160 of the distance from the pod to the subject [18].

6 The synchronisation of images

Each pod has a feed from a *master sync* signal operating at 25 Hz. The sequence of stages required to capture image and range data with a pod is:

1. The operator activates a grab on the console of one of the computers.
2. The control software tells the frame grabbers for all cameras on all machines currently being used to wait for trigger signals before capturing a sequence of frames.
3. The operator throws a switch on the trigger box to start pulses.
4. The trigger box generates a pulse.
5. The trigger is input to the Viper Quad frame grabbers, which trigger the monochrome cameras and the strobe projectors. The shutters are open for 400 μs.
6. The strobe trigger from one of the frame grabbers is also passed through a Thurlby Thandar TGP 110 digital pulse generator operating in triggered delay mode. This allows a controllable delay to be inserted into the signal. We currently use a delay of 400 μs.
7. The output from the TGP110 is sent to all of the colour cameras to trigger their shutters to open. These open for 4 ms (1/250 of a second).
8. Images are then downloaded from all cameras to the frame grabbers before the next master sync pulse.
9. The frame grabbers perform a Direct Memory Access transfer of the images over the PCI bus into the system area of main memory on the PC.

7 The control software

A software package, called XYZT, has been developed to control the display and the recording of images captured by the 24 video cameras of the system. XYZT allows as well the control of the processing of the data either for calibration purpose or the building of 3D models. The software runs under Windows NT, the operating system used for the capture machines.

Since the cameras are connected to 8 different computers, communications through a network are needed in order to control the whole system. Once XYZT is running on each computer connected to cameras, any running copy can be used to control all the frame grabbers. When the users select an action using the XYZT user interface, a

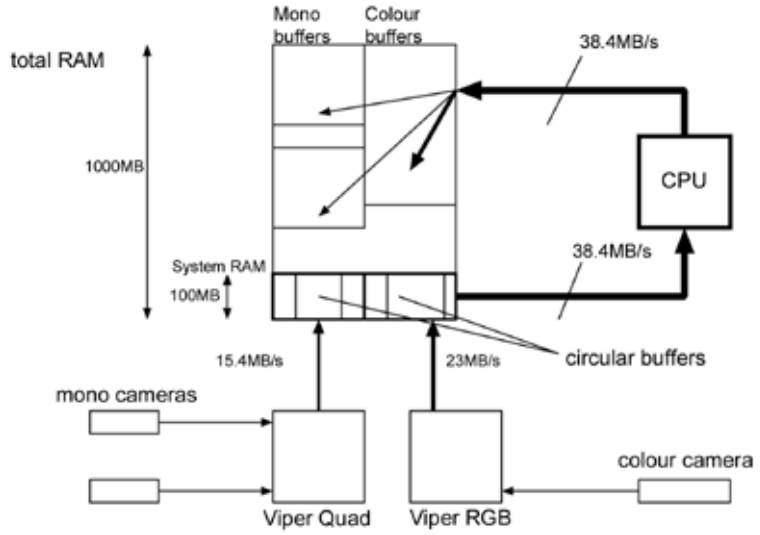

Figure 4: The dataflow during combined monochrome and colour capture.

message is broadcasted to the other copies of XYZT, which then execute the action.

Based on the $Sapera^{TM}$ library, XYZT sends commands to the different frame grabbers to display capture data in life windows, to stop the capture or to record the data on disks. Many parameters can be set such as the capture rate (up to 25 Hz), the capture time (up to 25 s because of RAM size limitation) and the output file format. When synchronisation is critical, i.e. for the recording of data generated through several computers, software commands have to be completed by a hardware synchronisation device (see section 6).

Once the data have been saved, that network of 8 computers can be used as a parallel machine to process the data. When the scanner is used in calibration mode, XYZT can launch in parallel the search of the features of the calibration target from the images captured by each computer. Then the calibration itself can be started from XYZT. For building 3D models from the scan data, first XYZT can launch the 3D data extraction routine (see section 9) for each stereo pair of images. Then XYZT can dispatch the 3D mesh generation tasks (see Section 9) over the network.

8 The data flow

The process of 3D data capture and model building involves very large data flows.

With 8 colour and 16 monochrome cameras operating, data is arriving from the frame grabbers at a combined data rate of 307 MB/s. A twenty-second "take" thus generates over 6 GB of data. This data rate is in excess of the sustained throughput of the disks attached to the processors. It is thus necessary to buffer the data in RAM

as it is being captured.

The input rate of 38.4 MB/s per machine is sustainable over the PCI bus. Unfortunately the maximum allocatable buffer under Windows is 100 MB in system RAM, which restricts full colour and monochrome capture to 2.5 s of continuous filming. For longer sequences data has to be transferred from system to user RAM by the CPU. This configuration is illustrated in figure 4. Whilst the total bandwidth to and from memory of 115 MB/s is within the theoretical capacity of a PCI bus, experiments with 800 Mhz Pentium processors indicate that it is not in practice sustainable. For long capture sequences it is necessary to partition the 8 computers into two sets, one of which is responsible for capturing monochrome range data and the other for capturing colour data.

Once the data has been captured to RAM, it is transferred to a half terabyte fileserver. Connection to the fileserver from the capture machines is via 100 Mb/s Ethernet from each computer into a concentrator and from there via a 1000 Mb/s optical fibre link to the fileserver. The fileserver, a dual processor PIII, runs Linux with Samba daemons handling communications to the NT machines.

During the process of extracting the 3D information, the input image files and results are held on the fileserver, with the processing done by the capture machines. We have found that despite the relatively high specification of the network connection to the server, network congestion is a major factor in determining throughput.

9 The extraction of 3D data

The process of extracting 3D information from a stereo-pair of images is termed stereo matching. The matching algorithm used was developed by Jin, and an earlier version is reported in [4].

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 triple [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 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 neighbourhood around l_{ij} and the matched point in image r. We will denote the overall matcher as the function `match(l,r` $\rightarrow$ `x,y,p)`.
The matcher is implemented using a difference of Gaussian image pyramid, and an inner matching function `innermatch(l,r` $\rightarrow$ `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.
2. Set c $\rightarrow$ 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. Starting on the top layer.

The function `innermatch` proceeds as:

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

   (a) A polynomial is fitted through the 5 correlation values and the maximum of the polynomial is determined.
   (b) If that is within the range of the initial search the relative coordinates of the peak are returned as x_{ij}, y_{ij}.
   (c) If the coordinates of the imputed correlation peak are out with 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 where the disparity is less certain.

10 Model building

Once 3D data have been extracted for each pod, these data need to be merged to generate a complete 3D model. Since the pods have been calibrated together, the 8 sets of 3D data captured at a given time step can be integrated in a single coordinate frame. An implicit surface is then computed that merges together the point clouds into a single triangulated polygon mesh.

First the volume of interest is divided in voxels using an octree structure. Then a process of volume carving starts to generate the implicit surface:

```
Set all voxels to the value UNSEEN
For each pod
 For each pixel of the left image
  Cast a ray from the camera to the associated 3D point
  Voxels which are crossed by the ray are set to EMPTY
  If the voxel containing the 3D point had the value UNSEEN
    Set the value FULL to the voxel
    Store the position of the 3D point relative to the voxel
```

The voxels with a value FULL are kept. They define the implicit surface of the 3D model.

Finally a single triangulated mesh is generated from this implicit surface using a variant of the marching cubes algorithm [19]. This mesh can be further decimated to any arbitrary lower resolution for display purposes.

The generation of photo-realistic models is achieved by mapping the colour pictures taken by the colour cameras to the 3D geometry. Since the colour cameras have been calibrated in the same coordinate frame as their corresponding black and white stereo pair, each 3D point can be associated to a value on the colour image. However since the colour balance and the settings of the 8 cameras are not identical, it is necessary to process the colour images to get a seamless texture for the 3D mesh. That is achieved in 2 steps: first the colour images are processed globally in order to get similar colour balances. Then a local processing is applied for pixels which appear as being in contact on the 3D models, but belong to different colour images. Finally the 3D mesh and the processed colour images are exported as a VRML file and a set of 8 colour texture files.

Figure 5: Derivation of individual range data from images

11 AVI generation

The output of our system is a sequence of VRML files with their associated JPEG textures files. In order to visualise 3D films, we have developed a 3D Film Editor which allows the rendering of a sequence of 3D models at the capture frame rate (up to 25 Hz), while the viewer can set interactively using their mouse the position of the camera. Since our software requires all the data to be loaded in RAM, we are limited so far to the visualisation of few seconds of data depending on the resolution of the meshes.

Depending on the application, it can be more convenient to visualise the 3D film through a series of Avi files, each of them showing the film under a different camera path. These Avi files can be generated as well using our 3D Film Editor software which has the ability to record the camera path while a user watches the 3D film interactively.

12 Sample Results

The sample images are of the actor Jeremy Killick playing the part of Napoleon in an experimental film. Figure 5 shows the process by which a set of three captured images from a pod are converted into distance information. The images labelled top and bottom come from the two monochrome cameras of a pod. The speckle pattern can be observed. The images labelled horizontal and vertical disparity map encode the distances in the horizontal and vertical directions between corresponding pixels in the top and bottom images. The confidence map image records the correlation coefficients over 5 by 5 pixel windows centred on pixels in the bottom image and the corresponding pixel position in the top image. In conjunction with camera calibration data, a depth image is produced.

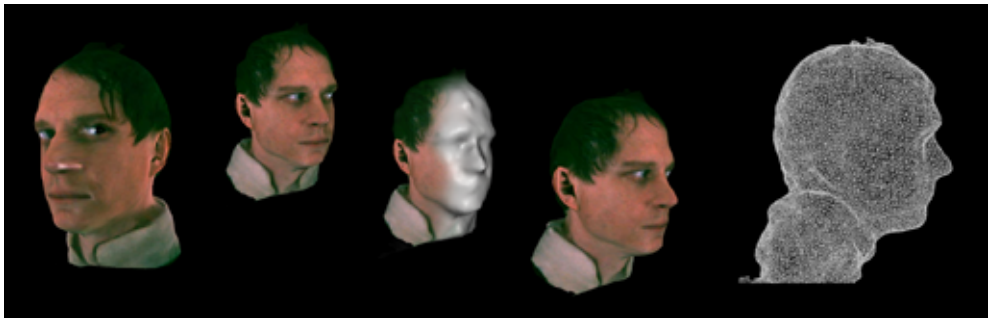

Figure 6: Different views of an individual model

Figure 7: Sequence of models captured over time

Figure 6 shows several different views of Napoleon's head all taken from the same 3D model. The centre image has had the skin texture partly removed to show the underlying model. The rightmost image shows the triangulated mesh of the model. Rotation is use to illustrate the genuinely 3D nature of the model. Figure 7 shows 5 models selected from one take of 625 frames.

13 Conclusion

In this paper we gave a detailed description of the dynamic 3D whole body scanner we are currently developing, where original software and hardware systems have been investigated. Then we demonstrated the validity of the concept by generating a true 3D film using a configuration for head scanning.

In the future, we will address the two main limitations of our system: the illumination of the subject and the data flow generated. First we plan to change the illumination to a continuous pattern illumination based on a shadow throw to get over the depth of field problems. Moreover, we plan to use powerful stroboscopes to over flash the pattern illumination and to provide a better colour images. The strobes

should flash with 100 Hz to ensure a better comfort to the scanned subject. Secondly we intend to tackle our data flow issues by researching better transfer protocols and increasing the size of our computing power using a bigger network of computers

14 Acknowledgments

This work was supported by the SHEFC project "Michelangelo". We would like to thank Paul Siebert, Don N. Whiteford and Ming Zhou. We also would like to thank Kim Bour (Director), Philip Lloyd (Producer), Andrew Bampfield (Writer) and Madeleine Bowyer and Jeremy Killick (Actors) for giving us the opportunity of scanning the first 3D film.

References

[1] Siebert J. P. Wray D. Ayoub, A. F. and K. F Moos. A vision-based three dimensional capture system for maxillofacial assessment and surgical planning. *Brit. J. Oral Maxillofacial Surg.*, 36:353–357, 1998.

[2] J. Paul Siebert and Stephen J. Marshall. Human body 3d imaging by speckle texture projection photogrammetry. *Sensor Review*, 20(3):218–226, 2000.

[3] J. P. Siebert and C. W Urquhart. C3d: a novel vision-based 3-d data acquisition system. *Proc. Mona Lisa European Workshop, Combined Real and Synthetic Image Processing for Broadcast and Video Production*, 1994. Hamburg, Germany.

[4] Zhengping. *On the multiscale iconic representation for low-level computer vision systems*. PhD thesis, The Turing Institute and the University of Strathclyde, 1988.

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

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

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

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

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

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

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

[12] Lawrence Livermore National Laboratory. http://www.llnl.gov/automation-robotics/cyber.html.

[13] Sundar Vedula Takeo Kanade, Hideo Saito. The 3d room: Digitizing time-varying 3d events by synchronized multiple video streams. Technical report, Robotics Institute, Carnegie Mellon University, 1998. Pittsburgh, PA.

[14] Peter Rander Robert Collins Takeo Kanade Sundar Vedula, Simon Baker. Three-dimensional scene flow. *Proceedings of the 7th International Conference on Computer Vision*, 1999.

[15] H. Saito T. Kanade S. Vedula, P. Rander. Modeling, combining, and rendering dynamic real-world events from image sequences. *Proceedings of Fourth International Conference on Virtual Systems and Multimedia*, 1998. Gifu, Japan.

[16] Y.F. Abdel-Aziz and N.M. Karara. Direct linear transformation from comparator coordinates into object coordinates in close-range photogrammetry. *Proc. ASP Symposium on Close-range Photogrammetry*, pages 1–18, 1971. Illinois.

[17] H. M. Karara. Handbook of non-topographic photogrammetry. 2nd ed 1989. American Society for Photogrammetry and remote sensing, Falls Church, VA.

[18] W. P. Cockshott J.-C. Nebel, F. J. Rodriguez-Miguel. Stroboscopic stereo rangefinder. *3DIM2001*, 2001. , Qubec City, Canada.

[19] W. E. Lorensen and H.E Cline. Marching cubes: a high resolution 3d surface construction algorithm. *ACM Computer Graphics*, 1987.