Generation of True 3D Films

Jean-Christophe Nebel

3D-MATIC Research Laboratory, Department of Computing Science,
University of Glasgow, Glasgow, Scotland, UK
`jc@dcs.gla.ac.uk`

Abstract. We define a true 3D film as a film that can be viewed from any point in space. In order to generate true 3D films, the 3D-MATIC Research Laboratory of the University of Glasgow has been developing a capture 3D studio based on photogrammetry technology. The idea is simply to generate 25 photo-realistic 3D models of a scene per second of film. After the presentation of the state of the art in the domain, the core technology of our dynamic 3D scanner is detailed. Finally first results, based on a 12 camera system, are shown and the potential applications of this new technology for virtual storytelling are investigated.

1 What is a true 3D film?

From the 30s onwards cinema has tried to create the illusion of the 3D image by artificially reproducing binocular vision. The cinema aims to film separately the same object from two angles corresponding to our binocular vision and to project them onto the same screen. It only remains for each eye to select the image, it is meant to receive, in order to recreate a 3D illusion. Special glasses are used for this purpose. This technology has been quite successful considering that nowadays such 3D films are offered every day in cinemas such as IMAX^{TM}.

Since these films are filmed from a specific viewpoint, spectators can only see 3D images from that viewpoint. For some people, it may be a bit frustrating, they would like to see these 3D images from other view points: they would like to see true 3D film. We define a "true" 3D film as a film that can be viewed from any point in space. Ideally spectators should have the ability to choose interactively the position of the camera while watching these films; they should be able to fly over scenes of a film, as it is now possible to navigate through 3D virtual reality environment using VRML browsers. In order to generate these true 3D films, the 3D-MATIC Research Laboratory of the University of Glasgow has been developing a capture 3D studio based on photogrammetry technology.

In this paper, we present the prototype of the 3D studio we are currently developing. First we relate our research to previous work in the fields of 3D capture, modelling and animation of virtual humans. Then we describe the technology we developed and finally we show some experimental results and offer some applications for virtual storytelling.

2 Related work

In spite of being a very active research topic, convincing animations of realistic virtual humans have been demonstrated only in very few short films such as "Tightrope" by Digital Domain^{TM} and "L'Opus Lounge" with "Eve Solal" by Attitude Studio^{TM}. Two main challenges have to be addressed in order to generate such films: the creation of photo-realistic 3D models of real humans and the realistic animation of these models.

2.1 Creation of photo-realistic 3D models of real humans

The first challenge is the creation of photo-realistic 3D models of real humans. On one hand skilled designers are able to make human models using software such as 3D Studio Max^{TM}. However since few months are necessary for the creation of a convincing model, they generally represent average humans, specific film stars or athletes. On the other hand specific human models can be generated using automatic or semi-automatic techniques. There are mainly two main methods: the deformation of generic 3D models and the generation of 3D models using scanners. In the first case, pictures are mapped on a generic 3D model of an average character, which has been scaled and deformed in order to match the pictures. Blanz et al. generate faces from a single picture using a morphable face model [1], which was built using statistics acquired from a large dataset of 3D scans. Hilton et al. map pictures of the full body of a person, taken from 4 orthogonal views, on a generic 3D human model representing both shape and kinematic joint structure [6]. These techniques produce very realistic 3D models. However since they are based on only few pictures and a generic model, the similarity between the human model and the generated model depends on the viewpoint. The other way of generating automatically realistic humans is by using scanners. Several techniques can be used to scan a human body: laser beams [17] and Cyberware^{TM}, structured light technique [21] or photogrammetry [15] and [18]. Their accuracy (about 1mm) is usually sufficient for getting very realistic 3D models. Moreover colour pictures are mapped on these models what ensures photo realistic appearance.

2.2 Realistic animation of human models

Once these photo-realistic 3D models are available, the second challenge needs to be addressed: their animation. There are many software allowing the animation of human like figures using key frames such as Maya^{TM} and Poser4^{TM} and a lot of work has been done in order to ease the way of generating poses using techniques such as emotional posturing [3], [1] and genetic algorithms [4]. However, it is still a very long task and requires highly skilled animators to generate realistic motion. Therefore research has focused more recently on high level techniques such as adapting reference movements obtained either by keyframing or motion capture. The higher level of control provided reduces the animator's load of direct specifications for the desired movement. Many approaches have been followed such as the interpolation between keyframes of reference motions [5] and [2], the generation of collision free motions [12] and the derivation of a motion from a reference motion by adding emotions or behaviours to keyframes [20] and [3]. In conclusion, a lot of work has been done in order to speed up the process of generating realistic animations, but ultimately an animator is still needed to set the fine tunings.

2.3 3D Capture of human motion and shape

For the time being it seems that the only practical way of generating quickly convincing 3D animations of human beings is to use real people as much as possible: the actors should be scanned and their motions should be captured. Therefore what is needed is a 3D scanner which would be able to scan a full moving body in a very short capture time, in order to freeze the motion, and would be able to scan this body, ideally, at a cinema or TV frame rate. Very few of the scanners presented previously have a short enough capture time. The commercial scanners, based on laser beams and structured light, have a capture time of about 10 seconds, whereas the ones using photogrammetry only need few milliseconds. Obviously only the later type of scanners has the potential of capturing moving subjects. People of the research team of the British company TCTi^{TM} [16] work on the generation of true 3D movies using photogrammetry based scanning technology, however no result has been published yet.

The Robotics Institute of Carnegie Mellon University has also 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 [8] are mounted on the walls and ceiling of the room. Since their main interest is the analysis of human motion [19], they have not developed any advanced method in order to generate accurate 3D models. However using silhouette carving, they managed to generate sequences of crude 3D models which have allowed them to create amazing true 3D films (see [14]).

It is also worth mentioning the work by Monks [11]. 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 technique has some limitations compared 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.

3 Principle

Since realistically believable true 3D films cannot be generated easily using animation techniques, we offer a totally new method: the capture of 3D data using scanning techniques at a frame rate of 25 scans per second. The idea is simply to generate 25 3D models of the scene per second of film. Since the capture time of the scanner has to be very fast, we use a scanner based on photogrammetry which has a capture time of few milliseconds [15]. Therefore that gives us the ability to capture subjects at a frame rate of 25 scans per second. The prototype of the 3D studio we are currently developing allows the 3D capture of a scene fitting a 2 metres side cube, typically we can capture the motion of a single actor. The configuration of this scanner is the following: the scene 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 parallelepiped will image the active volume (see Fig. 1). For the time being only 12 cameras have been installed. A more detailed presentation of our dynamic 3D capture system is given in [13]

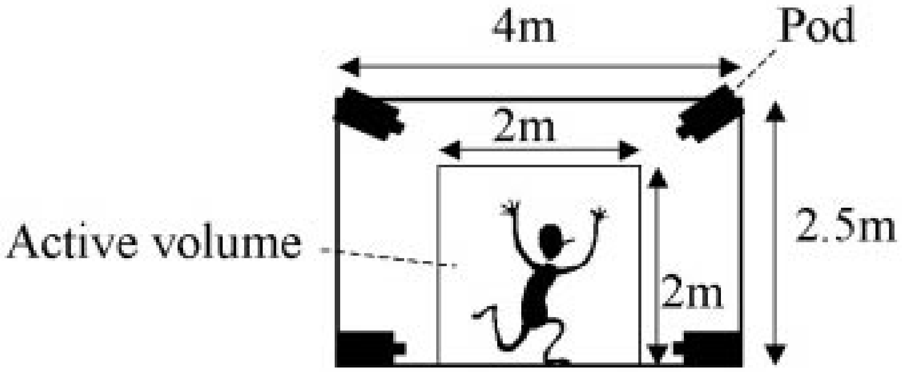

Fig. 1. Configuration of the 3D studio

4 Dynamic 3D data capture

The process of 3D capture relies upon flash stereo photogrametry. Each pod has one colour and two 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 which illuminates the scene with a random dot pattern. At the rate of 25Hz successively, the colour cameras capture the scene illuminated with uniform white light, and then the mono cameras capture the scene illuminated with the texture. The total capture time is under $150\mu s$. The monochrome images are used for stereo range finding and the colour images are used to capture the surface appearance of the subject.

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 has to be known. Then once the capture has been done, the stereo matching process is applied to each stereo-pair images. The algorithm we use is based on multi-resolution image correlation [22].

The algorithm takes as input a pair of monochrome images and outputs a pair of images specifying the horizontal and the vertical displacements of each pixel of the left image compared to the matched point in the right image (see Fig. 2). The matcher is implemented using a difference of gaussian image pyramid: the top layer of the pyramid is 16 by 12 pixels in size for a base of 640 by 480. Starting from the top of the pyramid, the matching between the 2 pictures is computed. Then using the displacements, the right image of the next layer of the pyramid is warped in order to fit the left image. Thus if the estimated disparities from matching at the previous layer were correct, the two images would now be identical, occlusions permitting. To the extent that the estimated disparities were incorrect there will remain disparities that can be corrected at the next step of the algorithm, using information from the next higher waveband in the images. Since at each layer, the two images are supposed to match more or less, thanks to the warping step, only a neighbourhood of five by five pixels is needed for each pixel in order to find the matching pixel in the other image.

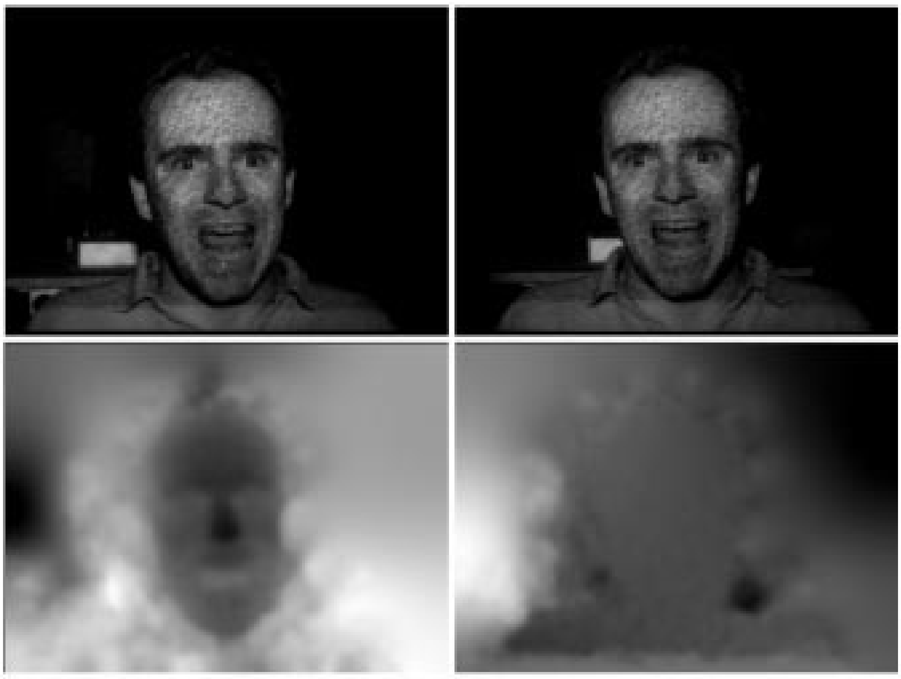

Fig. 2. Input images and final disparity maps (x,y)

Once the stereo matching process is completed, the final displacement files 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. Since the pods have been calibrated together, the 8 range maps of a given time step can be integrated in a single coordinate frame. A implicit surface is computed that merges together the point clouds into a single triangulated polygon mesh using a variant of the marching cubes algorithm [10]. This mesh is then further decimated to any arbitrary lower resolution for display purposes.

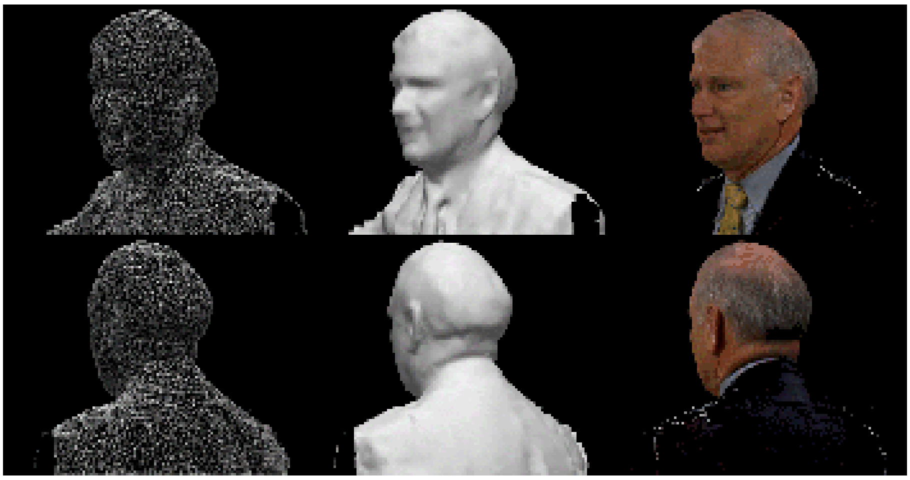

Fig. 3. Photo-realistic 3D model captured using 4 pods

The generation of photo-realistic models is achieved by mapping the colour pictures taken by the colour cameras to the 3D geometry. On Fig. 3, a photo-realistic 3D model, generated from four pods, is presented.

Imaging systems can offer a coverage of 90-95% of the full human body. Therefore the 3D models which are generated will not be complete, what is not acceptable for most applications. However we have recently worked on the conformation from generic models to 3D scanned data [7]. Consequently by conforming a complete generic model to our incomplete scanned models we can get an approximation for the missing pieces of mesh. Regarding the missing pieces of the texture, we will investigate the interpolation of the texture available and the utilisation of the texture generated at other time steps when there is no occlusion of the area of interest.

5 Results and applications

5.1 True 3D film

We present our first results, a true 3D film captured from four pods (head scanner). In total this film is composed of 25 frames (1 second), which represents 150 MB of raw data and 80 MB of VRML and JPEG files for a mesh resolution of 3mm. The data were processed fully automatically using a network of 4 PCs (PIII 803MHZ, 1GB RAM). The total computation time was of 38mn (25mn for the matching process and 13mn for the generation of the 25 3D models).

Fig. 4. 4 frames of a true 3D film captured by one pod

On the Fig. 4 we show the four first frames captured by one of the four pods and the corresponding 3D model generated from these pictures. The models are shown from different viewpoints. Obviously we do not think that pictures are the best supports for showing our results (the film can be downloaded from our web site [9]).

We think that this first true 3D film has achieved its goals. It demonstrates that our technology is reliable and up to the challenge. First, one second of a 3D film can be generated automatically in a bit more than half an hour. And secondly and more importantly the film is realistically believable, it looks like a real film instead of a great achievement of computer graphics.

Obviously the main limitation of our 3D studio is that, since there will be only 8 pods, it will not allow the capture of more than one actor at a time. However that should not prevent the generation of film involving several actors. Our 3D studio could be used as a 3D "blue screen" studio, where actors would be filmed separately in 3D. Then their 3D models could be integrated in any 3D environment (realistic or not) using classical 3D graphics techniques. In the future, the 3D studio could be fitted by much more cameras which would allow the generation of 3D films with several actors: the Robotics Institute of Carnegie Mellon University has demonstrated that using 49 cameras placed at different viewpoints, it is possible to capture two characters interacting with each other (see [14]).

5.2 Applications for virtual storytelling

The technology we have been developing gives the opportunity of telling virtual stories with many new ways. We could classify these applications in two closely related and often mixed categories: virtual storytelling based on virtual reality and animation techniques and virtual storytelling based on cinema and special effects.

Since the animation of virtual actors is still a very difficult task, real actors could be filmed in 3D using our studio and then these models would be integrated in virtual environments. Obviously these data could be edited and modified. Some interesting opportunities come from the fact that as we generate models using a marching cubes algorithm, a full sequence may be defined as a collection of 4D voxels associated to texture maps. Therefore when a model is built for a specific moment of the virtual story, it is possible to do it combining voxels created at different time steps. For example, we could visualise easily a scene where the speed of light would have been slowed down to few metres per second: the position of the extremities of the different limbs of a character would be few frames late compared to the position of the centre of the body.

In the near future we think that true 3D films will be used mainly at the production level in order to generate films that could not have been generated without our technology. At first, stories could be told by setting the position of the camera without any physical restriction. For example the films could be watched from the viewpoint of any character, human or not: we could imagine a second release of the successful 1999 fantasy comedy "Being John Malkovich", where the film would be shown from the eyes of John Malkovich! Secondly, stories could be also told with different sensitivities according to which pre-set path of viewpoints is chosen: the focus could be set on different characters such as the victim or the criminal, or a film could be suitable for an adult public under a specific viewing path and suitable for a younger public under a different one. In few years we should be able to offer to the public a unique experience: a full 3D immersion inside a film. By projecting the film using polarized light stereoscopy, the spectator, equipped with glasses and a joystick, will have the possibility of watching the film in 3D from any viewpoint they will navigate to.

Our technology is only at its prototype phase, but already many applications have been foreseen. There is no doubt that we will have to wait for its development and the involvement of more creative people in order to have a better idea of its full potential.

6 Conclusion and future work

In this paper, after having presented the state of the art in the fields of creation and animation of virtual humans, we described the 3D studio we are currently developing. Then we demonstrated the validity of the concept by generating a true 3D film using a system configured for head scanning. Finally we offered many applications of this technology for virtual storytelling.

In the future we will complete the system and optimise the computation, specially the matching process which is the most time consuming. Moreover since the amount of data that our 3D studio generates is quite important, we will need to address, as well, the compression of our data and new representations for 3D models.

References

1. V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH'99 Conference Proceedings, pp 187-194*, 1999.
2. M. F. Cohen C. Rose and B. Bodenheimer. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Comp. Graphics and application*, 18(5):32–40, 1998.
3. D. J. Densley and P. J. Willis. Emotional posturing: A method towards achieving emotional figure animation. In *Computer Animation 97*, pages 8–14, 1997.
4. S. Etienne. *Positioning articulated figures.* PhD thesis, University of Glasgow, 1998.
5. S. Guo and J. Roberge. A high-level control mechanism for human locomotion based on parametric frame space interpolation. In *Computer animation and simulation'96*, pages 95–107. Springer Computer Science, 1996.
6. A. Hilton, D. Beresford, T. Gentils, R. Smith, and W. Sun. Virtual people: Capturing human models to populate virtual worlds. In *CA'99 Conference, 26-29 May 1999, Geneva, Switzerland*, 1999.
7. X. Ju and J. P. Siebert. Conformation from generic animatable models to 3d scanned data. In *International conference of numerisation 3D - Scanning*, 2001.
8. Takeo Kanade, Hideo Saito, and Sundar 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.
9. 3D-MATIC Research Lab. http://www.dcs.gla.ac.uk/ jc/films/theawakening.html.
10. W. E. Lorensen and H.e Cline. Marching cubes: a high resolution 3d surface construction algorithm. In *ACM Computer Graphics, 21*, 1987.
11. T. P. Monks. Measuring the shape of time-varying objecs. In *PhD thesis, University of Southampton*, 1994.
12. J.-C. Nebel. Realistic collision avoidance of upper limbs based on neuroscience models. In *Computer Graphics Forum, volume 19(3), August 2000*, 2000.

13. J.-C. Nebel, F. J. Rodriguez-Miguel, and W. P. Cockshott. Stroboscopic stereo rangefinder. In *3DIM2001: Third International Conference on 3D Digital Imaging and Modeling*, 2001.
14. Virtualized Reality. http://www.cs.cmu.edu/ virtualized-reality/.
15. 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.
16. TCTi. http://www.tcti.com/.
17. R. Trieb. 3d-body scanning for mass customized products - solutions and applications. In *International conference of numerisation 3D - Scanning*, 2000.
18. G. Vareille. Full body 3d digitizer. In *International conference of numerisation 3D - Scanning*, 2000.
19. Sundar Vedula, Simon Baker, Peter Rander, Robert Collins, and Takeo Kanade. Three-dimensional scene flow. In *Proceedings of the 7th International Conference on Computer Vision*, September 1999.
20. D. J. Wiley and J.K. Hahn. Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and application*, 17(6):39–45, 1997.
21. 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.
22. 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.