Abson K, Ugail H and Ipson S (2008): "A Methodology for Feature based 3D Face Modelling from Photographs", ***Eighth IASTED International Conference on Visualization, Imaging and Image Processing (VIIP 2008), pp. 367-372,*** **Palma de Mallorca, Spain.**

A METHODOLOGY FOR FEATURE BASED 3D FACE MODELLING FROM PHOTOGRAPHS

Karl Matthew Abson, Hassan Ugail and Stan Ipson
The School of Informatics
The University of Bradford
United Kingdom
{K.M.Abson, h.ugail, s.s.ipson}@bradford.ac.uk

ABSTRACT

In this paper, a new approach to modelling 3D faces based on 2D images is introduced. Here 3D faces are created using two photographs from which we extract facial features based on image manipulation techniques. Through the image manipulation techniques we extract the crucial feature lines of the face in two views. These are then used in modifying a template base mesh which is created in 3D. This base mesh, which has been designed by keeping facial animation in mind, is then subdivided to provide the level of detail required. The methodology, as it stands, is semi-automatic whereby our goal is to automate this process in order to provide an inexpensive and expedient way of producing realistic face models intended for animation purposes. Thus, we show how image manipulation techniques can be used to create binary images which can in turn be used in manipulating a base mesh that can be adapted to a given facial geometry. In order to explain our approach more clearly we discuss a series of examples where we create 3D facial geometry of individuals given the corresponding image data.

KEY WORDS
Facial modelling, Binary images, Feature extraction, Animation

1. Introduction

In many industries both modelling and animation are extremely important parts of the production pipeline. The games industry for instance is worth USD 58 billion. Furthermore the creative breakthroughs made for leisure in those industries trickle down to benefit others such as the medical industry. Therefore it is essential to develop techniques, which will eliminate challenges, which currently stand in the way. The creation of human faces both by conventional, manual means and using computer aided methods still require a great deal of time and expertise, as well as money. In fact, most computer-aided methods require expensive software as well as lots of hardware resources and time to reach a result.

The major problem in creating automated systems for the purpose of face modeling and animation is the identification of corresponding features in different faces as well as recreating the structure of the face in an efficient and tidy way. This is especially important when it comes to generating faces which are suitable for animation. The human visual system can easily distinguish between different shapes due to the way our minds have developed and the way we have learned to pick out key features. In an automated system however we have much less information to work with, especially if we only have one image and lack depth information [1].

1.2 Previous and related work

Clearly, geometric modelling of human faces as well as geometric modelling as a whole has been a problematic subject since the beginning of computer graphics. Moreover finding ways of automating this process can be equally difficult. Nowadays it is more common to find individuals with more artistic backgrounds rather than individuals with computer science degrees working in the game, film and television industries. Furthermore these industries are constantly pressuring for advances in speed, automation and reuse [2], as it is no longer cost effective or advisable to model using traditional methods.

There has been significant automation in all areas of 3D graphics including consumer games, virtual reality and more academic areas such as geography and landscape simulation. One such example is the creation of systems, which in many ways automate the creation of complex 3D objects, such as trees. For example, a system developed by Makoto Okabe et. al. [3], does just that. This particular system aims to quickly create complex three-dimensional models of trees from hand drawn sketches and free the user from complicated rules and parameters. This sketch-based method is an inductive process where the user specifies the 2D appearance of the model and the system generates a 3D structure by inferring hidden parameters. However these systems are not without faults and a common occurrence is that in order to rapidly construct objects some principles must be ignored. As a result, final models can sometimes include artefacts not seen in conventional modelling systems. Furthermore algorithms exist which make it possible to derive 3D models of high visual quality from single images of arbitrary resolutions. However even though these approaches are robust in general they suffer from

errors depending on the quality of input images and input image complexity [4].

On the other hand there are systems, which work quite effectively using a compromise of automated and interactive techniques. One system called VideoTrace allows users to interactively generate realistic 3D models of objects from video, which could for example be used in video games. To achieve this, the user must trace the shape of the object over one or more frames of video and the system interprets this using computer vision techniques. It is true that a combination of automated and manual techniques could allow certain operations to succeed in cases where just automation approaches would fail. The user's input in many cases can be vital, however doing this does limit who could operate the system [5]. On the other side of this argument interactive methods have been proven to be simple enough for use by non-expert users in such a way that they could work rapidly and intuitively [6].

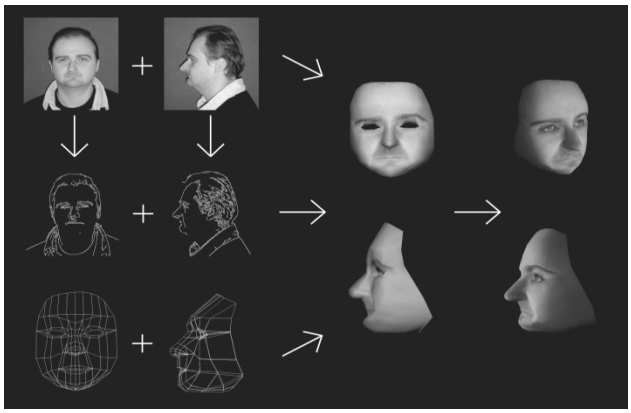
With this in mind more and more ways of automating modelling are coming about. Many however lack the advantage of being fully automated and require human interaction or require expensive resources [7]. Our approach aims to remain affordable and effective as well as requiring little human intervention. Furthermore the idea itself allows the creation of output, which would be ideal for multi-purpose.

Most methods, which do not require great deals of manual assistance, require extensive resources when they are first set up and therefore appear to be designed as a complete package with features, which may or many not be required by some users. Many techniques use a database of laser scans of face models, which are then used to morph a facial model at run time. Such techniques are fundamentally expensive and sparse, meaning that production of software using this method would prove difficult and make finished software expensive for the end user. For example the work of Blanz and Vetter described in [7] explains one such process which has proven to give highly realistic interpretations of the target. However it is unclear as to how long it would take the system to complete its task. Using such techniques would likely require large amounts of storage space as well as processing power.

Furthermore a method known as the golden ratio is also intriguing. The work described in [8] utilizes the golden ratio to three-dimensional scans based on facial mask. The golden ratio is believed to be a blueprint for facial features that conform to beauty. In this work the authors describe how they proceed to morph this mask to the ideal positions to suggest a new arrangement for facial features in cosmetic surgery, resulting in a more beautiful face. According to their claims these techniques could be applied to modify characters to make them look more beautiful or the opposite and make them less beautiful. This technique appears very interesting and the applications of such a feature could prove extremely useful. With further work such a technique could be used to find measurements between points on a face solving our lack of depth information and allowing us to turn one image in a decent interpretation of a 3D mesh. When it comes to animation however it is unclear as to how well the mesh will deform and how natural the animation would appear.

The idea of using an image-based approach to 3D reconstruction is not a new one and we have already discussed numerous methods, which use laser scanning to produce a morphable model. There are however other methods using photographs as well as other special equipment. The disadvantage of those techniques though, such as Quicktime VR [9] or Lightfields [10] is that they use a huge number of photographs.

Systems such as the Realviz Image Modeler [11] have also attempted to make this a reality and even though they succeed they still have the requirement of calibrating cameras. The work of Liebowitz and Criminisi [12] are often works referred to as being the most accurate. However they are also the most labour-intensive as they work on the principle of recovering a metric reconstruction of an architectural scene using geometry constraints [13] in order to compute 3D locations of user-specified points given their distances from a ground plane. Furthermore the user must also specify various other information, making this approach very heavy on the user. Our method aims to give an alternative to this by employing the use of a simple mesh with relatively few points. This is then modified to match target images, which are provided by the user. This method would have the advantages of requiring less storage space, being relatively simple to set up when compared to other methods and less expense in development. This technique could even be designed as a small application used in conjunction with a much larger piece of software, which is already available.



Input **Output**

Fig. 1 – Creating 3D face from photographs.

1.3 Contribution

In this method a base mesh has been developed through looking at various animation models and how effectively they animate. The result is a mesh that posses topology

with very few points, which both animates effectively and is easily modified to match various faces. It is possible to closely and easily map points using binary images of the subject in a short time period. These binary images, produced using Canny image manipulation techniques have the advantage of revealing major features of the face in a simple binary form, which may be advantageous as automation is the primary goal. Furthermore this low polygon mesh can be subdivided in such away to produce further detail without deviating from the profiles or proportions of the image. This procedure is illustrated in Fig. 1.

2. Image Manipulation

Since this system will be completely automated and aimed at users with varying levels of experience the only input the system will have to work with are simple colour images. It cannot be expected for the user to know image manipulation techniques so any system must include these and carry them out automatically with a minimum of user intervention. It can however be quite difficult when trying to extract information, such as the location of facial features from colour images. All that information can prove to be very confusing especially in an automated system. To overcome this problem we need to extract only the required information and simplify what we give to the system. We do this by turning those same images into binary images, which through manual testing prove to be a principle way of helping to produce more accurate faces.

There are two major steps involved in converting the images to the binary form which we use at the mesh creation stage. Firstly we extract an 8bit intensity image from a 24bit colour image. Secondly a Canny edge detection algorithm [14], which uses Gaussian smoothing, non-maximum suppression of the intensity gradient and hysteresis tracking, as shown in Fig. 2, is applied to the result. Gaussian smoothing is necessary due to the fact that the Canny algorithm uses a first derivative type filter and is therefore susceptible to noise. The gradient magnitude G and direction angle Θ

$$G = \sqrt{G_x^2 + G_y^2} \quad \Theta = arctan\left(\frac{G_y}{G_x}\right)$$

of the smoothed result are calculated from the *x* and *y* components of the gradient. The direction angle is rounded to one of four values representing vertical, horizontal and the two diagonals (0, 45, 90, 135 degrees). Gradient magnitude points are then set to zero if a neighboring value in the direction of maximum gradient is greater. This process, which is referred to as non-maximum suppression reduces the gradient magnitude image to a set of edge lines of strengths equal to the local gradient maxima. Intensity gradients, which are large, are more likely to correspond to significant edges than intensity gradients, which are small. A binary edge image is created using hysteresis tracking to set edge points with values above an upper threshold and continue to set contiguous edge points providing they are above a lower threshold. This causes fainter sections of contiguous lines to be detected while ignoring unconnected faint sections and noise.

We use this form of edge detection due to the fact that the Canny algorithm aims to be an optimal edge detection algorithm if the assumptions of the underlying model are valid. Optimal means firstly good detection – as the algorithm attempts to mark as many real edges in the image as possible. Secondly, good localization, as the edges marked appear as close as possible to the true edges. Thirdly minimal response – as edges in the image are only marked once, and as such creates little image noise and do not create false edges.

There are other edge detection algorithms, which can be used, such as Laplacian, Sobel, Prewitt and Roberts, etc. and these have all been tested with varying degrees of success. Canny edge detection achieved best overall results as well as producing a complete binary image without half tones or distortions, which could cause difficulties. This may have advantages over other methods, as the binary result could be extremely welcome when it comes to automation. Using this image in an automated face creation system could be more successful due to its lack of complexity and unwanted detail.

Fig. 2 – Canny edge detection.

Currently this technique has been tested with a handful of test subjects all producing good results. However it is possible that the binary image could be susceptible to the quality of the input image. As such any automated system would need to be robust and able to accommodate difference in quality. The image shown in Fig. 2 was created using Gaussian smoothing with standard deviation of 1.0 and hysteresis thresholds of 25 and 0. Depending on the image the settings may need to be altered to achieve a good result.

3. Mesh creation

The creation of the initial mesh used in this technique evolved from lots of experimentation with various facial models and research into various animation techniques. The aim of creating a mesh which deforms naturally, while at the same time using as few points as possible can be rather conflicting as numerous animation methods require a good level of detail to morph correctly. By looking at Fig. 3 one can clearly see that even a low

polygon model requires many points in order to represent the various features of the face as well as the changes in height and shape, like contours on a map.

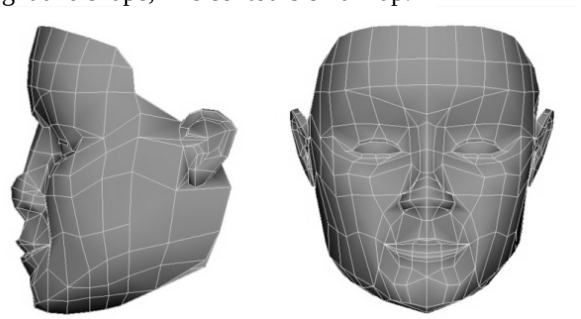


Fig. 3 – Original base mesh.

However it was found that a base mesh with very few points could be constructed using Maya software environment when based on essential feature points [15]. The creation of the base mesh was actually accomplished by using an image of the target face in both side and front views. With these images it is then possible to use the various tools in the Maya environment to create and deform a primitive cube. By focusing on feature points, such as the nose and using tools to split polygons and merge points it is possible to match points to their correct position in the images in one view and then move it in the next. The end result is something similar to the image shown in Fig. 4.

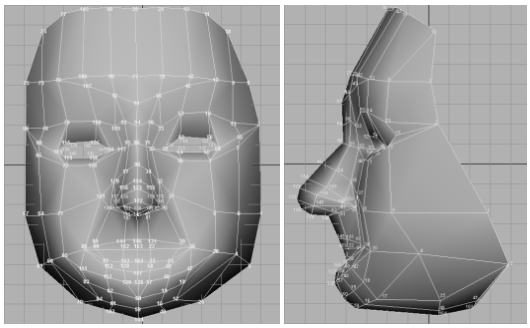


Fig. 4 – Base mesh with compact points.

These feature points represent the face in a very compact form and could be sufficient as well as readily animated, as illustrated in Fig. 4. In comparison to the topology used in techniques involving laser scanning where points can range from 700 to 70,000 vertices, the base mesh in this technique uses less than 200 vertices to describe a face. This means the face can be made to match the position of feature lines in binary images relatively easily. Furthermore by producing a mesh in combination with proven animation techniques and by studying the muscle structure of the face the topology of the model is of such detail for use in applications such as computer games. This base mesh is good for mapping out low-level detail and effectively maps the feature lines of the face due to the fact that the majority of the points are clustered around key features.

Obviously this mesh is not always going to be perfect for all applications. In some cases further detail would be a necessity. The mesh can be further subdivided for applications requiring further detail such as animated film. Due to the way the topology has been developed it is possible to add further detail. This is only possible using subdivision techniques developed by us, which do not alter the models current feature points. As such models still match feature lines in source images.

This could be implemented in many hundreds of ways and could be possibly built into an automated system, in much the same way, as you would go about it by hand. One would simply tell the system what sort of detail you were looking for, at the beginning and watch it firstly move the base mesh into position and then subdivide and move following levels till completion. One way we have found is in the form of a smoothing algorithm. However it is not just a case of smoothing the whole mesh at a certain level of intensity and arriving at the end result. The result of doing this would be a complex mesh which will certainly not conform to the profiles specified in the binary images as the original points are moved as well as the new ones which are created. There are two ways to avoid this unwanted feature. One can either subdivide and then move all the new points into place individually or reduce the intensity of the smoothing algorithm and reach the end result in a step-by-step process. By doing this one can subdivide and smooth in smaller increments, creating new points and moving them into position slightly and carrying out the procedure again and again until the end result is achieved. This technique has been tested and as can be seen the end result is a complex mesh, which conforms to the original feature lines as shown in Fig. 5.

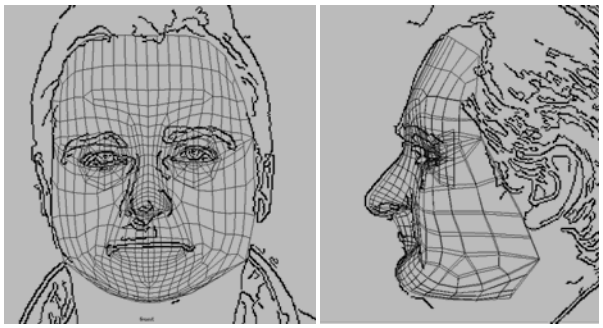


Fig. 5 – Base mesh with compact points.

4. Results

We have tested this technique on a number of individuals and found that it holds true when subjected to a number of different possibilities. We started by taking a simple image of each subject, both front and side on before applying Canny edge detection on each of them. The result each time was an image with all the information needed to correctly match the points of the mesh to the equivalent feature points of the image, when imported into the Maya package and used as a guide. It was also

found that it was quite simple to move feature points to the correct location in one view following by the next due to the low number of points. Furthermore the process of generating the 3D mesh was quite short due to the simplicity of the mesh. As can be seen in Fig. 7, Fig. 8, and Fig. 9, the number of feature points used, in most cases is ideal which results in lines which match the binary images. In one or two cases, for example the forehead in Fig. 9, the number of feature points can be insufficient. However this maybe due to specific faces and can be corrected using smoothing as shown in Fig. 6. This results in a mesh which corresponds to the original proportions of the face.

Fig. 6 – Base mesh subdivided twice but still matches feature lines.

5. Conclusions and further work

We have successfully built and tested a low polygon mesh which can be morphed to match the feature lines taken from binary images. The mesh which is designed for animation purposes can be taken as it is or subdivided further, using the smoothing technique developed by us. Using a mesh such as this one could have numerous applications in all aspects of computer animation as its topology and polygon count is ideal for real time games where high counts can be an issue. Furthermore once the points of the base mesh are in place it can gradually be subdivided using our technique to add further detail for applications such as film where high detail is more of a requirement as shown in Fig. 6. We feel that this technique could be implemented in the form of a script in Maya rather than producing a large piece of standalone software. This method of deployment would prove to be very advantageous to animators in the field as all they would need would be the script and their images. This would prove to be very cost effective and relatively inexpensive in comparison to methods which produce a full software package with unneeded features.

Currently the mesh has been tested on three faces and matches feature lines which are extracted from photographs using our Canny edge detection technique. The advantage of using the edge detection algorithm allows the identification of features to become much less difficult as extracting information from full colour images can be overwhelming, especially in the automated system we intend to develop.

Currently certain parts of the face need slightly more attention as they have an inadequate number of points for some faces. This is simply a matter of further testing and modifying the mesh as a result.

Our next step is to automate the current steps in regard to the image manipulation and facial modelling. Our aim is to get to the stage where the user simply supplies one front image or a front and side image and in return the system produces a mesh matching the image(s).

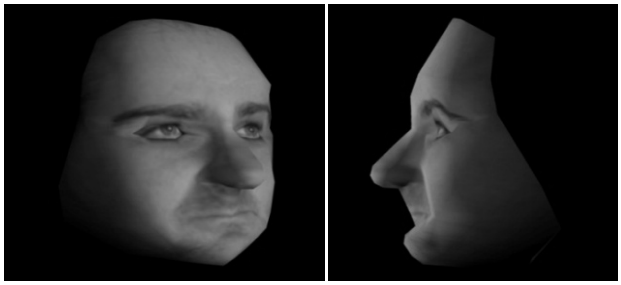


Fig. 7 – Base mesh subdivided twice but still matches feature lines.

Currently the front image is used to texture the face in a planner fashion. This is adequate for testing purposes but slightly lacking for actual use. This technique has the problem of stretching down the side of the nose and the far sides of the face. It may be possible to create a model where two images are combined from different angles to extrapolate all parts of the face in greater detail.

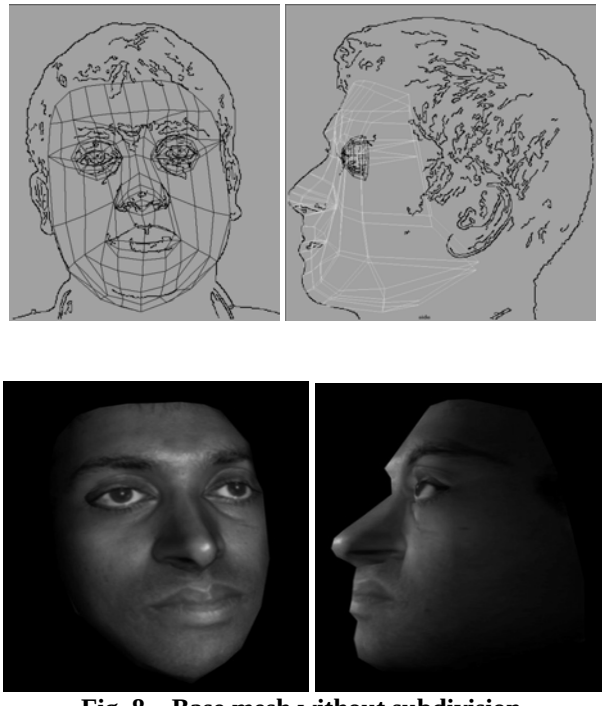


Fig. 8 – Base mesh without subdivision

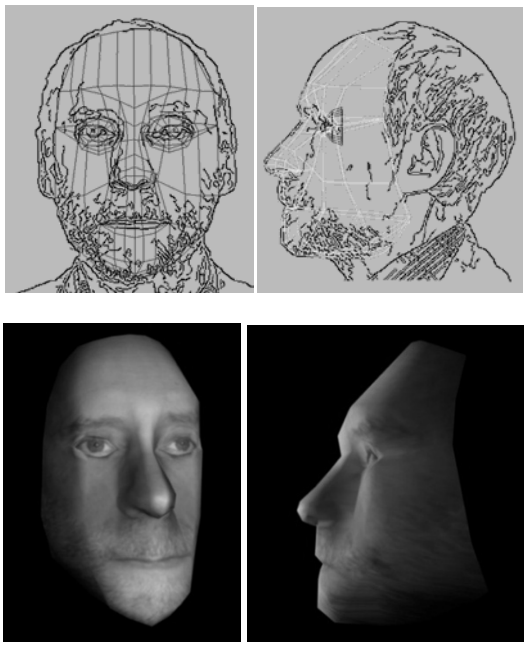


Fig. 9 – Base mesh without subdivision

References

[1] Z. Li, J. Liu and X. Tang, Shape from regularities for interactive 3D reconstruction of piecewise planar objects from single images.*MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia,*Santa Barbara, CA, 2006, 85-88.

[2] P.J. Birch, S.P. Browne, V.J. Jennings, A.M .Day and D.B. Arnold, Rapid procedural-modelling of architectural structures.*VAST '01: Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage, Glyfada,* Greece, 2001, 187-371.

[3] M. Okabea, S. Owada and T. Igarashi, Interactive design of botanical trees using freehand sketches and example-based editing.*SIGGRAPH '07: ACM SIGGRAPH 2007 courses*, San Diego, California, 2007, 487- 498.

[4] P. Müller, G. Zeng, P. Wonka and L.V. Gool, Image-based procedural modeling of facades. *ACM SIGGRAPH 2007 papers*, San Diego, California, 2007, 187-371.

[5] A.V.D. Hengel, A. Dick, T. Thormählen, B. Ward and P.H.S. Torr, VideoTrace: rapid interactive scene modelling from video. *ACM SIGGRAPH 2007 papers*, San Diego, California, 2007, 861– 865.

[6] A.V.D. Hengel, A. Dick, T. Thormählen, B. Ward and P.H.S. Torr, A shape hierarchy for 3D modelling from video. *GRAPHITE '07: Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, Perth, Western Australia, 2007, 63 – 70.

[7] V. Blanz and T. Vetter, A morphable model for the synthesis of 3d faces. *Siggraph '99: proceedings of the 26th annual conference on computer graphics and interactive techniques*, Los Angeles, LA, 1999, 187-194.

[8] R. McDonnell and A. McNamara, Application of the golden ratio to 3d facial models. *Proceedings Eurographics*, Ireland, 2003, 85-83.

[9] E. Chen, QuickTime VR - an image-based approach to virtual environment navigation. *ACM SIGGRAPH 95*, New York, NY, 1995, 29–38.

[10] M. Levoy and P. Hanrahan, Light field rendering.*ACM SIGGRAPH 96*, New Orleans, Louisiana, 1996, 31–42.

[11] D. Holiem, A. A. Efros and M. Hebert, Automatic photo pop-up. *ACM Transactions on Graphics (TOG) 2005*.

[12] D. Liebowitz, A. Criminisi and A. Zisserman, Creating architectural models from images.*In Proc. Eurographics*, 1999, vol. 18, 39–50.

[13] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. (Cambridge University Press, 2004).

[14] J. Canny, A computational approach to edge detection. *IEEE Trans. Pattern analysis and machine intelligence*, 1986, 8:679-714.

[15] Autodesk Maya Press, *Learning Autodesk Maya 2008: The Modeling and Animation Handbook.* (Autodesk Maya Press 2007).