A Strategy for 3D Face Analysis and Synthesis

Mark Chan, Chia-Yen Chen, Gareth Barton
Patrice Delmas, Georgy Gimel'farb, Philippe Leclercq, Thomas Fischer
Department of Computer Science, University of Auckland
Private Box 92019, Auckland, New Zealand
ccha196@ec.auckland.ac.nz

Abstract

Using 2D images is one of the most common techniques for the reconstruction of 3D face models. In this paper, we compare the strengths and weaknesses of different image processing techniques for 3D face generation. It is anticipated that the optimal solution will be applied in the future for 3D face analysis and synthesis. This paper presents binocular stereo using stereo correspondence algorithm, binocular stereo using triangulation, orthogonal views, and photometric stereo as approaches to 3D face modeling.

Keywords: 3D Face, binocular views, photometric stereo, orthogonal views, triangulation

1 Introduction

Today, research is actively being conducted for the purpose of creating high performance and reliable human-computer interface systems. As an essential component, face modeling has been a hot topic, recently receiving much attention [1]. Reconstructed face models are required to be reliable, compact, and accurate. Special characteristic face feature areas such as the eyes, mouth, nose, etc, are especially important. Currently, two main approaches are used to create 3D facial models. The first uses a 3D scanner and captures 3D positional values of human head shapes. The second, processes 2D images for 3D model generation. In this paper, we focus on the latter approach.

There are several techniques for creating 3D facial models using 2D images. [2] uses a stereo pair of images to find pixel correspondence between them to generate a disparity map. A depthmap can then be constructed. However, these stereo images has to be rectified beforehand.

Another approach is to calibrate the cameras and to obtain 3D world coordinates of each pair of pixels in the images by using triangulation. Finding correspondence between the stereo pairs can be either performed via automatic pattern localization or by manually finding similarities.

A third approach, using only one camera, is orthogonal views [3]. Two images are taken, one from the front and the other from the side. The front-view image provides the X- and Y-coordinates, while the side-view provides the Z-coordinate of the pixel corresponding to the same feature in both images. [4] detects facial features from colour images and applies the orthogonal views technique to obtain their 3D positional values.

Photometric stereo [5], is based on the way images of 3D objects are formed. Objects can be seen because they reflect light. The surface normal and other characteristics of the surface (e.g. depth) can be obtained using prior knowledge of the scenes' illumination geometry and the nature of surface reflection.

The goal of this paper is to test several 3D face techniques and compare their strengths and weaknesses. In the future, the optimal solution will be applied in 3D face analysis and synthesis. In Section 2 the stereo vision, binocular stereo, orthogonal views, and photometric stereo techniques are explained. Section 3 describes how to apply these techniques in the context of face modeling while Section 4 provides experimental results. The final section summarizes the paper and presents our future work.

2 Facial Reconstruction Techniques

There are many approaches which enable realistic 3D face reconstruction. Using laser scanning [6] or a stripe generator [7] can produce precise results, but both require special hardware and equipment, in addition to their prohibitive cost. A much more economical approach is to reconstruct a 3D face model from 2D-image information. Image processing techniques such as binocular stereo, orthogonal views and photometric stereo allow 3D positional values to be obtained and 3D objects to be reconstructed. In this section, we discuss each of these techniques in detail.

2.1 Binocular Stereo using stereo correspondence algorithms

Two main steps are required in this technique. Firstly, stereo images are required to be rectified to be coplanar. The next step is stereo matching. This procedure is to

find the pixel correspondence between stereo images and subsequently the pixel disparity. A 2D disparity map is then generated for 3D reconstruction.

2.1.1 Rectification

Stereo images are rectified by converting them into an epipolar horizontal stereo pair. The epipolar geometry is recovered by (1) manual selection of a set of corresponding pixels in the both images; (2) computation of the fundamental matrix using an approximate linearised approach [8] (3) refinement of the approximate matrix by using non-linear optimisation with the Levenberg–Marquardt gradient-based algorithm and (4) forming a rectified pair where each initial epipolar line becomes a horizontal scan-line.

2.1.2 Stereo Matching

Two stereo matching algorithms have been studied in this paper:

- SAD: correlation algorithm using the sum of absolute differences,
- P2P: pixel to pixel, a dynamic programming algorithm proposed by Birchfield and Tomasi [9].

$I(P)$ is the intensity at some point P, $I_L(P)$ is the intensity in the left image and $I_R(P)$ in the right image. Correlation functions are evaluated over a 'window' of neighbouring pixels in each image. A window $w(P,r)$ is defined by its centre, P, and its radius r. A radius, r implies a square window of $(2 \cdot r + 1) \times (2 \cdot r + 1)$ pixels.

The correlation process is illustrated by figure 1. A point P on the reference image (left for instance) is chosen, all correlations with a sliding window - for all disparity values - in the right image for the whole disparity range are computed and the best value is chosen, defining the matching pixels. SAD is defined as:

$$C(P) = \sum_{P' \in w(P,\beta)} \left| I_R(P') - I_L(P') \right| \quad (1)$$

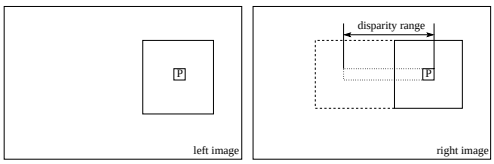

Figure 1: CorrelationWindow

The pixel to pixel algorithm minimises a cost function to find the best path through all possible solutions. The cost function defined as:

$$\gamma(M) = N_{occ} \cdot \kappa_{occ} - N_m \cdot \kappa_r + \sum_{i=1}^{N_m} d(x_i, y_i) \quad (2)$$

where x_i is the index in the left image and y_i the index in the right image (using the epipolar constraint only indices on the same lines are needed), N is the number of occlusions (N_{occ}) or matches (N_m), κ is the cost for an occlusion (κ_{occ}) or reward for a match (κ_r) and $d(x_i, y_i)$ is the dissimilarity measure used to decide whether $I_L(x_i)$ and $I_R(y_i)$ are image pixels of the same scene point.

2.2 Binocular Stereo using Triangulation

Three main steps are involved in binocular stereo. The first step is to calibrate the cameras, thus attaining the physical and optical properties of the acquisition system are attained. The second step is finding correspondence between the stereo-pair images. This operation can be either performed via automatic pattern localization or by manually finding similarities. The last step is to calculate the 3D coordinates of the corresponding points in the images by triangulation technique.

2.2.1 Calibration

Camera calibration is the process of estimating the intrinsic and extrinsic parameters of a camera. These coefficients allow a 3D point from the world reference frame to be transformed into its corresponding point in the image reference frame and vice versa (See Figure 2).

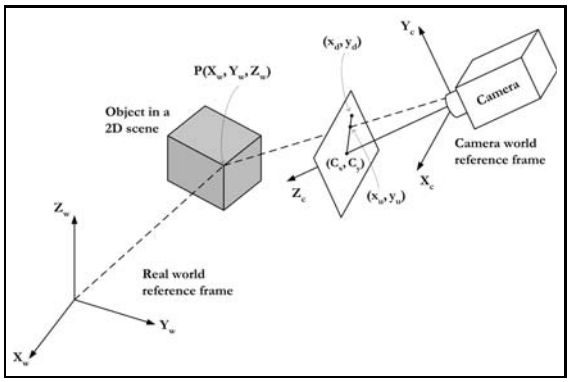

Figure 2: Reference Frame Model

Extrinsic parameters are the parameters that define the location and orientation of the camera axis with respect to a known world reference frame. Intrinsic parameters are the parameters necessary to link the pixel coordinates of an image point with the corresponding points in the camera reference frame. Tsai's calibration is a "two-step" calibration method [10] involving the direct computation of most of the calibration parameters and an iterative approach to estimate the remaining parameters (namely the depth component of the translation vector, the focal length and the radial distortion parameter).

2.2.2 Triangulation

Triangulation is a technique which back-projects a pair of corresponding points in two images into the real world, so that the real 3D positions of these points can be attained. Since the calibration parameters of the cameras are known, a ray in 3D space which a given object point must lie on, may be derived given a set of corresponding image coordinates. With two rays in a calibrated stereo environment, that correspond to the same 3D world point in two camera views, an approximation of the point's three-dimensional location may be computed by finding the point of closest approach of the two rays (an intersection is not guaranteed).

2.3 Photometric Stereo

The theory of Photometric Stereo for Lambertian surfaces was developed by Woodham [5]. It calculates surface normal and other surface information by employing prior knowledge of the illumination geometry and the nature of surface reflection. For Lambertian surfaces, a surface normal can be determined if the considered surface point is illuminated from three or more light sources using the albedo-independent PSM method. Therefore, three consecutive images are taken with light sources being switched on from three different directions in our experiments (See Figure 3).

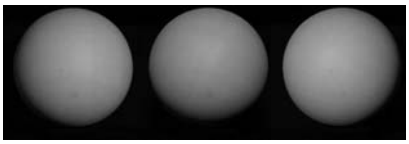

Figure 3: Calibration Spheres

Photometric Stereo allows the reconstruction of a depth map or a 2.5-D model (See Figure 4). The reconstruction accuracy depends on the quality of the generation of the surface normal and the transformation from the surface normal to the depth map. Further details can be found in [11].

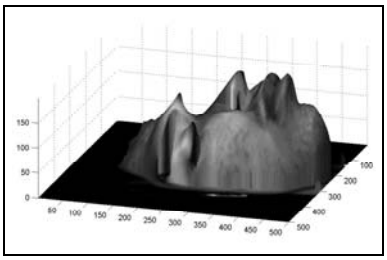

Figure 4: Depthmap

2.4 Orthogonal Views

To reconstruct a 3D face model from orthogonal view images, two images are required, the first from the front of the face, the other from the side. 3D coordinates of the face points, visible in both images, are then captured using the X,Y coordinates of the front view, while their Z values (depth) are attained from the side view.

Changes in illumination between the front and side views make it difficult to automatically find corresponding pixels in both images. Instead, research has been focusing on extracting similar features such as the eyes, eyebrows, lips, nose and mouth which can be extracted using image processing techniques [12]. These features can then be mapped to a 3D generic face model to reconstruct a 3D face [13]. The more feature points used, the more detailed the reconstructed 3D model will be.

3 Experiment

3.1 Binocular Stereo using stereo correspondence algorithms

Firtly, the stereo images are rectified. Then image matching is performed using both SAD and P2P. Studies of these stereo algorithms against noise [14] suggests that a window radius of 4 for the SAD algorithm, an occlusion cost of 5 and a match reward of 40 for the P2P algorithm are most suitable. Since the disparity map is retrieved, a depthmap can be generated by computing with the camera focal length, image pixel sizes and the disparity.

3.2 Binocular Stereo using Triangulation

In this experiment, full camera calibration and triangulation is used to obtain 3D coordinates of feature points on the test subject's face.

3.2.1 Calibration

Two Sony EVI-D100P video cameras, a tripod with a horizontal bench and a calibration box are the main equipment used in this experiment. The video cameras are fixed on the tripod at a distance at 20 cm apart. Two images of the calibration cube with 150 non-coplanar 3D reference points (See Figure 5) are taken simultaneously. 17 calibration parameters are then estimated for each camera by using Tsai's calibration algorithm which is applied to the 2D reference points extracted from the images and the measured 3D world location of each point.

In order to find the optimal distance between the cameras and the calibration object, tests on calibration accuracy at varying distance were performed. Results (See Table 1) indicate that given the current setup, calibration error is minimal at 115 cm.

For the purpose of testing the accuracy of back-projection, using the computed camera parameters, the 3D coordinates of all reference points on the

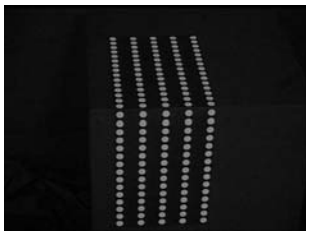

Figure 5: Calibration Box

Distance (cm)	Error(mm)			
	Avg	Min.	Max.	Std Dev.
90	0.85	0.18	2.51	0.43
95	0.85	0.14	2.96	0.50
100	0.79	0.14	2.47	0.42
105	0.74	0.11	2.33	0.37
110	0.78	0.19	2.15	0.38
115	0.72	0.17	1.64	0.33
120	0.93	0.10	3.18	0.56
125	0.95	0.15	3.64	0.59
130	1.04	0.21	3.78	0.67

Table 1: Calibration accuracy at varying distances.

calibration box are calculated. The calibration accuracy is obtained by comparing the calculated 3D world coordinates with the reference points' real coordinates. Experimental results show that 86% of the reference points' calibration error is less than 1.2 mm with maximum error on average 2.2 mm.

3.2.2 Triangulation

After both cameras are calibrated, a stereo pair of images is taken of the test subject. (See Figure 6)

(Left Image) **(Right Image)**

Figure 6: Stereo-pairs

Finding corresponding points between the images is preformed manually in this experiment. 280 dots are placed on the test subject's face. Since the camera calibration parameters are known, these 2D image points are back projected into real world and the real 3D coordinates are obtained by triangulation. An application is developed for this particular purpose.

3.3 Photometric Stereo

The application of Photometric Stereo in our experiment has been developed by [11]. The experiment took place in a dark room where all external light sources are blocked. This is necessary as uncertain illumination can affect the experimental results. The equipment used for this experiment includes a JVC CCD camera, three halogen light bulbs used as light sources and a serial box which connects all the hardware with the computer.

The first procedure is to calibrate the camera and calculate the light source direction. A sphere has been chosen as the calibration object (See Figure 3) due to its reflecting properties as well as its concave shape. Three images of the test subject (Figure 7) are then acquired and processed to reconstruct the face depth map (See Figure 4 for an example).The application also allows the mapping of the test subjects' texture on to the depth map which is then presented in VRML format.

Figure 7: Test subject images taken with different light sources orientation

3.4 Pseudo-Orthogonal Views

In our experiment, the test subjects are required to sit at an angle θ ($\theta < 90^\circ$) to acquire the side-view image (See Figure 8 b). This approach allows feature points from the side-view image to be easily extracted and also decreases error in their detection.

a) Front View **b) Side View with angle**

Figure 8: Orthogonal views of test subject with markers on the face

The real coordinates of the test subject's face can be calculated from the image using the following formula:

$$(X_{world}, Y_{world}) = \frac{(X_{image}, Y_{image})}{\cos\theta}$$

In this experiment, 280 markers (small white dots) have been distributed over the test subject's face. The dots are mostly scattered around face features and locations have been detected manually.

4 Experimental Result

4.1 Binocular Stereo using stereo correspondence algorithms

Figure 10 presents the depthmap obtained from binocular stereo using SAD. The result gives a very clear outline of the 3D face. However, the 3D surface appears to be rough and a small amount of data is lost.

Figure 9: Result - Binocular Stereo using SAD

This may be due to the lack of brightness in the stereo images.

4.2 Binocular Stereo using Triangulation

Figure 10 shows the result obtained from binocular stereo by manual pixel-correspondence. Binocular Stereo requires a high level of accuracy to obtain quality results. The inaccuracy of the corresponding point extraction between stereo-pairs, often caused by the change in illumination on the face, may be a major factor for the reconstruction error. Although feature points, represented by white dots, are extracted manually during this experiment, the inherent error in manual reading is large enough to dramatically affect the reconstruction results.

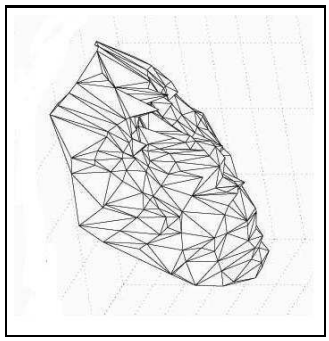

Figure 10: Result - Binocular Stereo

4.3 Photometric Stereo

Figure 11 shows results obtained using photometric stereo. This technique provides detailed face maps that are time-efficient to create and not dependent on any generic model. However, the inconsistency and unreliability of the system, are its greatest drawbacks. In (Figure 11), the first two sets of 3D face models are of the same test subject and with the same equipment settings taken on consecutive days. As reconstruction accuracy is dependent on the estimation of the light source direction and strength, the system is extremely sensitive to noise and measurement error [15], causing the obtained results to be different. The consistency of the results is affected by changes in the camera calibration parameter estimation due to the unstable experimental environment.

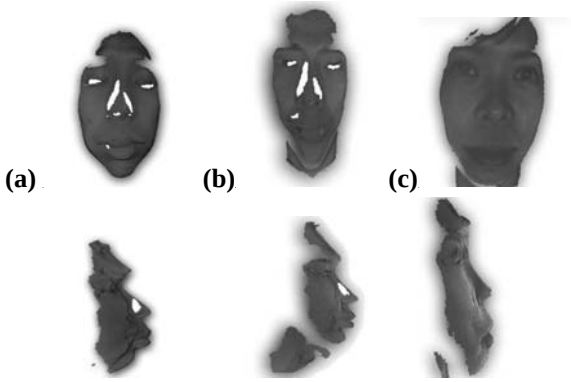

Figure 11: Photometric Stereo results

The accuracy of the surface orientation determination is also dependent on the surface reflection. Errors and data loss may occur on dark surfaces. Results in Figure 11 show that test subjects with light skin color (Figure 11 c) are better constructed than test subjects with darker skin color (Figure 11 b). Data is lost in the eye area, nose bridge, lips, and eyebrows, which are crucial areas in 3D face analysis and synthesis.

4.4 Orthogonal Views

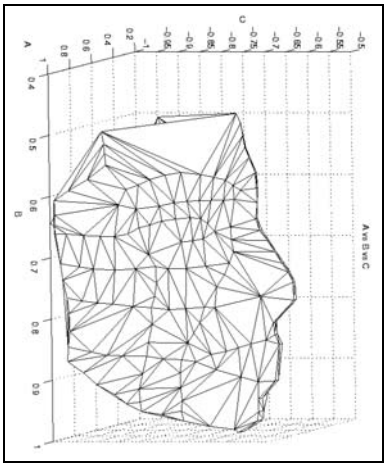

Figure 12: Result – Orthogonal Views

Figure 12 shows an example result obtained from orthogonal views. The outline of the eyes, nose, and mouth can be seen. The pitfall of using orthogonal views is its requirement for deformation. In this experiment, 280 white dots were placed on the test subject's face as feature markers. These feature points were then detected manually. However, current research focuses more on automatic face feature extraction. Due to the processing time constraint, the number of feature points is minimized and mapped to a pre-defined generic face model. This model must then be deformed to fit the input face feature points so that a face depth map can be generated, hence longer processing time is implied.

4.5 Discussion

Experimental results obtained using binocular stereo methods are acceptable with some minor reconstruction errors. Photometric stereo gives as well a time-efficient result. However, the lack of consistency and reliability becomes the main drawback of the system. Orthogonal views are reliable, consistent, and precise. The only disadvantage of this technique is its inability to produce a dense depth map unless deformation to a pre-defined face mesh is performed, which requires additional processing time. At this stage, orthogonal views appears to be the simplest solution for 3D face generation while photometric stereo and stereo-correspondence techniques automatically generate dense depth maps.

5 Conclusion

In this paper, we compared the image processing techniques, stereo vision, binocular stereo, photometric stereo, and orthogonal views for the purpose of 3D face analysis and synthesis. The comparison is made based on the strengths and the weaknesses between the systems. In the future, further comparisons based on the accuracy and processing times of the models will be made. A proper method to perform a face model comparison of accuracy, is to obtain a laser scan of a test subject, and use it as a benchmark with which to compare the results of the techniques described in this paper. Normalization and scaling of the results (depth map) is required. This will provide a means by which to evaluate the precision of each face model. Processing time of the systems can also be compared after both automatic feature extraction and face model deformation is developed.

References

[1] Q. Wang, H. Zhang, T. Riegeland, E. Hundt, G. Xu, and Z. Zhu. Creating animatable MPEG4 face. In *International Conference on Augmented Virtual Environments and Three Dimensional Imaging*, Mykonos, Greece, 2001.

[2] P. Fua and Y.G.Leclerc. Taking advantage of image-based and geometry-based constraints to recover 3-D surfaces. In *Computer Vision and Image Understanding*, volume 34, pages 111–127, 1996.

[3] H.H.S. Ip and L. Yin. Constructing a 3D individualized head model from two orthogonal views. In *The Visual Computer*, volume 12, pages 254–266, 1996.

[4] T. Kurihara and K. Arai. A transformation method for modeling and animation of the human face from photographs. In *Proceedings of Computer Animation*, pages 45–58, Tokyo, Japan, 1991.

[5] R.J. Woodham. Photometic method for determining surface orientation from multiple images. In *Optimal Engineering*, volume 19, pages 139–144, 1980.

[6] http://www.viewpoint/com/freestuff/cyberscan.

[7] Marc Proesmans and Luc Van Gool. Reading between the lines – a method for extracting dynamic 3D with texture. In *Virtual Reality Software and Technology Conference*, pages 95–102, Lausanne, Switzerland, 1997.

[8] O.D.Faugeras and Q.-T.Luong. The fundamental matrix: theory, algorithms, and stability analysis. In *International Journal of Computer Journal Vision*, volume 17, pages 143–75, 1996.

[9] Stan Birchfield and Carlo Tomasi. Depth discontinuities by Pixel-to-Pixel stereo. In *International Conference on Computer Vision*, pages 1073–1080, New Delhi, India, 1998.

[10] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. In *In IEEE Journal of Robotics and Automation*, pages 323–344, 1987.

[11] C.Y. Chen, R. Klette, and C.F. Chen. Improved fusion of photometric stereo and shape from contours. In *Proc. Image and Vision Computing*, pages 103–108, New Zealand, 2001.

[12] Taro Goto, Won-Sook Lee, and Nadia Magnenat-Thalmann. Facial feature extraction for quick 3D face modeling. In *Signal Processing: Image Communication*, volume 17, pages 243–259, March 2002.

[13] M. Escher, I. S. Pandzic, and N. Magnenat-Thalmann. Facial deformations for MPEG4. In *Proc. Computer Animation*, pages 138–145, Philadelphia, USA, 1998.

[14] Philippe Leclercq and John Morris. Robustness to noise of stereo matching. In *International Conference on Image Analysis and Processing*, pages 606–611, Mantova, Italy, 2003.

[15] Angela Ng and Karsten Schlöns. Towards 3D model reconstruction from photometric stereo. In *Image and Vision Computing New Zealand*, Auckland, New Zealand, 1998.