Comparative Study of 3D Face Acquisition Techniques

Mark Chan, Patrice Delmas, Georgy Gimel'farb, and Philippe Leclercq

Department of Computer Science, University of Auckland, New Zealand
patrice@cs.auckland.ac.nz

Abstract. friendly–user interactivity while permanently eyeing towards 3D display technologies. As such, 3D face generation, modelling and animation techniques are in the frontline to design realistic animated 3D talking faces. Simple, reliable and economic, 2D image processing techniques have been widely used to reconstruct 3D faces. This paper focuses on the comparison of different 2D imaging techniques for 3D face generation. Stereo Vision techniques, using either automatic stereo correspondence algorithm or manual feature points location, Orthogonal Views and Photometric Stereo approaches are introduced and applied to acquire face 3D data. In addition, generated reconstruction results are compared qualitatively and quantitatively.

1 Introduction

Nowadays, research is actively conducted to create highly performant and reliable human-computer interface systems. As an essential component, face modelling has been a hot topic, recently receiving much attention [1]. Special characteristic face feature areas such as the eyes, mouth, nose, etc, are especially important as they carry most of the audiovisual information expressed by humans. Although many approaches (such as laser range scanner devices) may be used to generate 3D faces, 2D imaging techniques have been the most widely researched as they do not require extensive budget or special hardware equipment. For all these reasons, this paper solely focuses on the study of 2D imaging technologies for 3D face generation.

As widely acknowledged to provide satisfactory results while maintaining low complexity computation, Stereo Vision, Orthogonal Views, and Photometric Stereo methods are studied in this paper.

Stereo vision can be either automatic or interactive. Automatic stereo vision requires stereo images placed parallel in a line wise correspondent position (also called epipolar position). Corresponding pixels between both images are then searched automatically along the same lines in both images to generate a dense disparity map (or a depth map for display purpose) [2].

The interactive approach requires to manually (or automatically) chose a subset of corresponding pixels in the stereo images pair. If cameras are calibrated, the pixel 3D world coordinates are obtained using back-projection techniques to provide a sparse depth map of the stereo system common field of view.

A. Gagalowicz and W. Philips (Eds.): CAIP 2005, LNCS 3691, pp. 740–747, 2005.

Orthogonal views have already been used to detect facial features and infer their 3D positional values [3]. Using either one or two camera(s), two images are taken, one from the front and the other from the side of the face. 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. This provides 3D information for all the pixels present in both front and side images.

Photometric stereo [4], 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.

In this paper we test the above introduced 3D face techniques and compare their strengths and weaknesses introducing a new 3D surface comparison approach using Radial Basis Function (RBF) interpolation to normalize 3D faces.

In Section 2, four image-processing techniques are described in the context of 3D face generation. In section 3, 3D surface comparison and results are presented. The final section summarizes the paper and presents our future work.

2 Facial Reconstruction Techniques

In this section, Image processing techniques such as binocular stereo, orthogonal views and photometric stereo are discussed in detail.

2.1 Binocular Stereo Using Automatic Stereo Correspondence Algorithms

Binocular Stereo is the process of obtaining **dense** depth information from a pair of images. Often these two images (stereo images) are related by the epipolar geometry. First, stereo images are rectified to be placed in epipolar position [2]. Next, stereo matching finds the correspondence between stereo images (usually using Pixel to Pixel, correlation windows, surface constraint or Dynamic Programming matching techniques) and produces a dense disparity map.

Stereo Matching. Previous studies proved that for faces simpler stereo algorithms tend to produce marginally lesser results while being much faster than more complex algorithms in favour today [5]. For this reason, SAD has been used in this paper. SAD, is a correlation algorithm, which uses the sum of absolute difference to find the correspondences between stereo images. Correlation functions are evaluated over a 'window' of neighbouring pixels in each image. For each point on the reference image (left for instance), 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.

Experiment. Firstly, the stereo images are rectified. Then, image matching is performed using SAD. Studies of this stereo algorithms against noise [5] suggests that a window radius of 4 is most suitable. Since the disparity map is retrieved, a depth map can be generated using both the camera focal length obtained by the calibration technique, and the image pixel size.

2.2 Interactive Binocular Stereo

Here, three main steps are involved in this approach. First, the cameras are calibrated to attain the physical and optical properties of the acquisition system. Next, correspondence between a subset of the stereo-pair image pixels is achieved by finding similarities (usually by clicking on stereo corresponding pixels). The last step is to calculate the 3D coordinates of the corresponding points in the images by triangulation technique.

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. Extrinsic parameters, such as the rotation and translation coefficients, define the location and orientation of the camera axis with respect to a known world reference frame. Intrinsic parameters link the pixel coordinates of an image point with its corresponding points in the camera reference frame. In this project, Tsai's calibration algorithm is applied due to its simplicity and sufficient accuracy. Tsai's calibration is defined as a "two-step" calibration method [6] involving the direct computation of most of the calibration parameters while an iterative approach estimates the remaining parameters (namely the depth component of the translation vector, the focal length and the first order radial distortion parameter).

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 a tripod 20 centimetres apart. Two images of the calibration cube with 150 non-coplanar 3D reference points are taken simultaneously. Nine calibration parameters, namely six external (rotation angles and translation vectors) and three intrinsic (e.g. the focal length, the uncertainty scale factor and the radial distortion factor) coefficients, are then estimated [6].

In order to find the optimal distance between the cameras and the calibration object, tests on calibration accuracy at varying distance between the camera and the calibration object were performed. Experimental results indicate that given the current setup, calibration error is minimal at 115 cm.

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.

Experiment. After both cameras are calibrated, a stereo pair of images is taken for each test subject.

Next, corresponding points between the images are found manually in this experiment as small white dots are put on test subject's face as markers. Once 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.

3D coordinates of the feature points are calculated and mapped to a generic 3D face model (1808 vertices) inspired from CANDIDE3 [7] (see Fig 1 first image). Its encapsulated MPEG-4 standard defines vertices according to the MPEG4 Face Feature Points(FFP)[1]. Second image of Fig 1 shows an example of the reconstruction result.

2.3 Photometric Stereo Method (PSM)

The theory of Photometric Stereo for Lambertian surfaces was developed by Woodham [4]. 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. Three consecutive images are taken with light sources being switched on from three different directions in our experiments (see left and middle left Figure 1) while a fourth one with all lights on is acquired for texture mapping.

A depth map or a 2.5-D model is then reconstructed by the Photometric Stereo method (See Figure 1). 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 [8].

Experiment. In our experiment, Photometric Stereo has been developed by [8]. The experiment took place in a dark room where all external light sources were blocked 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 of PSM is to calibrate the light source direction. A sphere has been chosen as the calibration object due to its reflecting properties as well as its concave shape. Three images of the test subject are then acquired and processed to reconstruct the face depth map. The application also allows the mapping of the test subjects' texture on to the depth map, which is then presented in VRML format. Fig 1 shows some of our reconstruction results obtained via PSM.

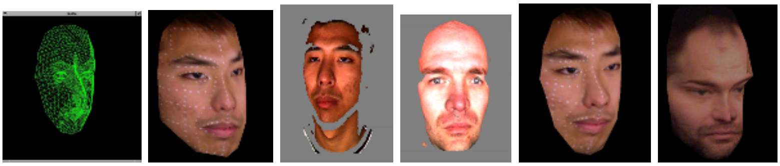

Fig. 1. From left to right: first 2 images: Interactive Binocular Stereo; Next 2 images: PSM results; last 2 images: Reconstruction results by Orthogonal Views

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.

Facial features such as the eyes, eyebrows, lips, nose and mouth can be extracted using image processing techniques [9]. These features are mapped to a 3D generic face model to reconstruct a 3D face. In our experiment, the frontal image is taken with test subject facing directly to the camera. Then, the camera is placed orthogonally (90 degree) and a side image of the face acquired. Fig 1 images show an example of orthogonal images for a test subject.

In this experiment, tiny white dots are placed on test subject's face as feature points. 29 facial feature points are extracted from the test subject's face manually. These points are then interpolated into the predefined face model.

3 3D Face Comparison

The goal of this project is to find the optimal 3D face reconstruction solution for 3D face analysis and synthesis. Therefore, it is necessary to determine which technique has the most accurate reconstruction. To do so, *3D Surface Comparison* is investigated in the following section.

3.1 3D Surface Comparison

3D Surface comparison allows finding the surface differences from individual reconstruction results by different image processing techniques. In addition, surface comparison can show the variances on areas between the reconstructed face surfaces. The overall surface differences for the whole test subject's population are computed. In order to find the optimal solution for 3D face analysis and synthesis in term of reconstruction accuracy, a surface comparison with the same vertices in surfaces generated by three image-processing techniques is performed.

There are a few factors that make the comparison extremely difficult in this experiment. Firstly, each system obtains results with different orientation and scaling. Secondly, benchmarks of each test subject are unavailable. In order to solve this problem, surface normalization is required, which involves rotation, scaling and translation of data. In this comparison approaches, we intend to apply RBF data interpolation technique to scale the 3D surfaces. After the normalization process, surface distances between reconstruction results are computed. In this experiment, we assumed the results from PSM as benchmark as it generates a complete face dense depth map and contains a large amount of vertices.

3D Surface Normalization. Research into 3D face comparisons from different systems is at an exploratory stage and no methodology has been defined for this particular type of comparison. Therefore, the approach applied in this experiment is a new idea and may not be the optimal method. In this project, depth maps of 3D faces generated from different systems are used for this comparison approach. Distances between the 3D surfaces are computed and compared. However, normalization is required for the 3D data, so that all 3D face meshes have the same orientation and scale.

Surface normalization is made up of three stages: rotation, scaling and translation. Rotation is for adjusting all the surfaces to face the same direction. Scaling adjust all the 3D surfaces with all primary facial features are located approximately the same area. The last procedure of normalization is to translate all the face surfaces to the minimum distance apart.

Rotation. The aim of this step is to have all the face surfaces sitting in the same coordinate setting and facing the same direction. Depth maps of face surfaces are used and the face should point upward. Figure 2 shows three 3D face surfaces after the rotation process. Each face dense map has the same size (500 x 500) and sits on the same coordinate system.

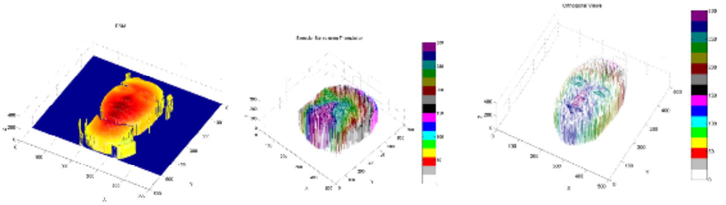

Fig. 2. Reconstruction Results after Rotation

Scaling. It is irrelevant to scale the whole face meshes by using just a few facial feature points. Ideally, all facial feature points should be used and these facial feature points should be distributed over the whole face surface. In this experiment, a new approach is investigated to scale 3D face surfaces. We intend to use Radial Basis Function (RBF), a data interpolation technique, for scaling 3D face surfaces. In this experiment, 18 points mostly located on the primary facial features are chosen in this normalization procedure. 3D data of these 18 points from the PSM result is extracted and interpolated into the Orthogonal Views' and Binocular Stereo with Triangulation's result. Since these 18 points are distributed over the whole 3D face, the whole face surfaces reconstructed by Orthogonal Views and Binocular Stereo with Triangulation technique is then deformed and scaled accordingly.

The Radial Basis Function (RBF) is a classical approximation function, defined as a weighted sum of translations of a radially symmetric basis function augmented by a polynomial term, and is widely used in surface reconstruction, image morphing, etc [10].

Translation. To simplify the comparison process, all these surfaces are translated as close as possible. In theory, the nose tip is the highest point among the whole face surface. In this normalization step, all the face surfaces are translated as the nose tips of all face surfaces are shifted to the centre of the depth map (250,250). Since all face surfaces are properly scaled, the location of facial features on each face surface such as the eyebrows, the eyes, nose and mouth should be located approximately in the same position. In addition, all the face surfaces are pulled to the same height. Again, the nose tip is used as the reference and all the face surfaces are translated until their nose tips are shifted to the same level.

Comparison Result. After all reconstructed 3D face surfaces are normalized, comparison can be made. All the face surfaces should have a uniform scaling, orientation and unit. The surface comparison is performed where the distances between the 3D surfaces are computed. Table 1 shows the depth map comparison result of the test population using the percentage of vertices having less than 5, between 5 and 10, between 10 and 15, and between 15 and 20 pixels variation between two surfaces. It indicates that 3D surface generated from Binocular stereo using Triangulation is closer to the 3D surface generated from PSM (benchmark) than any others. It has higher proportion of vertices (51.76% and 26.79%) with 5 and 10 pixels difference against PSM than Orthogonal Views.

Table 1. Overall Comparison Result on different 3D surfaces

	$\leq$ 5	**$\leq$ 10**	**$\leq$ 15**	**$\leq$ 20**	**$\geq$ 20**	**Max.**	**Mean**	**Variance**	**Std Dev.**
PSM vs OV	49.3	26.1	13	4.7	6.8	80.9	9.2	111.5	26.3
PSM vs Tri	51.7	26.7	10.4	5.4	5.5	80.2	8.4	136.2	10.5
OV vs TRI	74	18	4.5	1.6	1.9	36.0	4.1	26.2	4.6

Table 1 also shows that the 3D faces generated by Orthogonal Views and Binocular Stereo using Triangulation are very similar. 74 % of the vertices are less than 5 pixels between these two face surfaces. This result was expected since both techniques interpolate the extracted 3D data from the test subjects into the same predefined face model.

In order to investigate which areas on the face surfaces has the biggest and smallest difference to the benchmark, we tend to display the pixel difference between two surfaces graphically. Result shows that there is much less vertex differences between Binocular Stereo using Triangulation and Orthogonal Views' results than others. However, further work is required to work out the vertices difference for particular areas on the 3D face surface for all test subjects. Areas to investigate would be mainly around primary facial features such as the eyebrows, eyes, nose and mouth.

4 Conclusion

In this paper, stereo vision, photometric stereo, and orthogonal views are compared for the purpose of 3D face analysis and synthesis. For sake of comparison, we assumed 3D faces generated by PSM as benchmarks since PSM generates denser depth map. 3D surface comparison indicates that results generated from Binocular Stereo using Triangulation are closest to PSM.

We are currently investigating a proper method to perform a face model comparison of accuracy using laser scan of a test subject as a benchmark. We are also investigating Binocular Stereo using Stereo Correspondence Algorithm with USB driven digital cameras. Currently we use PSM and Binocular stereo to generate animatable 3D faces for realistic expressions generation.

References

1. Wang, Q., Zhang, H., Riegeland, T., Hundt, E., Xu, G., Zhu, Z.: Creating animatable MPEG4 face. In: International Conference on Augmented Virtual Environments and Three Dimensional Imaging, Mykonos, Greece (2001)
2. Zhang, Z., Faugeras, O.: 3D Dynamic Scene Analysis: a stereo based approach. Springer Verlag (1992)
3. Kurihara, T., Arai, K.: A transformation method for modeling and animation of the human face from photographs. In: Proceedings of Computer Animation, Tokyo, Japan (1991) 45–58
4. Woodham, R.: Photometic method for determining surface orientation from multiple images. In: Optimal Engineering. Volume 19. (1980) 139–144
5. Leclercq, P., Morris, J.: Robustness to noise of stereo matching. In: International Conference on Image Analysis and Processing, Mantova, Italy (2003) 606–611
6. Tsai, R.: 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. (1987) 323–344
7. Ahlberg, J.: Candide3 – an updated parameterized face. In: Report No.LiTH-ISY-R-2326, Department of Electrical Engineering, Linkoping University, Sweden (2001)
8. Ng, A., Schlöns, K.: Towards 3D model reconstruction from photometric stereo. In: Image and Vision Computing New Zealand, Auckland, New Zealand (1998)
9. Goto, T., Lee, W., Magnenat-Thalmann, N.: Facial feature extraction for quick 3D face modeling. In: Signal Processing: Image Communication. Volume 17. (2002) 243–259
10. Carr, J., Fright, W., Beatson, R.: Surface interpolation with radial basis functions for medical imaging. In: IEEE Transactions on Medical Imaging. Volume 16. (1997) 96–107