The three dimensional reconstruction and monitoring of facial surfaces

A. Gregory and R.T. Lipczynski

School of Electrical Engineering, University of Bath, Claverton Down, Bath BA2 7AY, UK

ABSTRACT

The study and monitoring of facial appearance is particularly important in the field of orthodontics and reconstructive maxillo-facial surgery. The introduction of new medical imaging techniques, such as X-ray computerized tomography (CT), has enabled surgery to be planned in three dimensions, compared with conventional methods using lateral and antero-posterior cephalographic images. Whilst it is possible to visualize bone and soft tissue structure the repeated use of CT scanning is undesirable because of the high doses of radiation involved. The research involves the development of a simple, low cost and non-invasive three-dimensional scanning method for facial surfaces. This is achieved using a technique known as structured light, using a standard commercial slide projector, CCD cameras and a framestore linked to a PC. The eventual aim is to provide clinicians a software assisted surgical procedure, by merging X-ray and skin data, allowing the manipulation of bone and soft tissue. This will give the facility to predict and monitor post-surgical appearance, detailing the effects of surgery benefitting both the clinician and patient. The research focuses on how the images are obtained using the scanning system and the subsequent image processing of data to give a realistic 3D image using a standard PC.

Keywords: Structured light scanning, reconstruction of facial surfaces

INTRODUCTION

There were a number of criteria identified that are essential or desirable in a scanning method for three dimensional body surfaces such as the face:

1. Non-invasive, and using a non-coherent light source.
2. Instant or near instant recording, to eliminate errors due to subject movement.
3. Relatively simple optics and operation.
4. Capable of being linked to a standard PC for the manipulation and plotting of data.
5. Scanning system should map the patient’s face at a resolution of 1 mm or better in x, y and z.

EXISTING METHODS FOR FACIAL MEASUREMENT

There are many well established techniques for the measurement of three dimensional body shape including laser range finding, moiré topography, stereo photogrammetry and contour photography, reviewed by Gallup. Dunn et al. applied structured light to the quantification of surface burns. It was evaluated for medical applications in the 1980s by Lievesley. To a large extent it was preferred to contemporary methods such as traditional stereo photogrammetry, since despite offering significantly better accuracy, <0.2 mm, this required complicated and expensive optical equip-
HARDWARE

The hardware consists of two monochrome CCD cameras (Hitachi KP-111) connected to a real-time frame grabber capable of storing four video frames. The cameras have a scanning area of 6.5 mm by 4.85 mm, with 585(h) and 470(v) pixel resolution. The frame grabber has 512 x 512 pixel resolution with 255 quantization levels (8 bits).

In addition, there is a monitor facility using a monitor directly connected so that either the 'live' image, during patient positioning, or the 'frozen' grabbed image can be displayed. The field of view of the system, as currently arranged, is approximately 600 x 500 mm giving a spatial resolution at the object surface of about 1 mm per pixel. An important point to note is the need for synchronization between the cameras and the image acquisition hardware, to prevent any mismatch or pixel jitter.

The data is downloaded from the framestore to a 386 IBM PC compatible using the standard Centronics port.

PROBLEMS AND CONSIDERATIONS

The image processing on the captured structured light pattern is complicated by the requirement to produce distinct and good contrast stripes and automating the whole process of 3D reconstruction is a non-trivial task. Essentially the system is a mono-photogrammetric one, if we treat each camera/projector pair in isolation, with one active camera and one passive camera, i.e. the projector. As a consequence, many of the traditional problems found in stereo systems also have to be solved. In common with other optical techniques such as moiré topography there are difficulties when dealing with concave areas. In practice the shading experienced in concave areas can be minimized with careful orientation of the camera/projector pair and the subject, and to a large extent is away from the areas of interest.

As outlined, since this is a photogrammetric technique using the underlying principle of triangulation to recover the three-dimensional shape, it is reasonable to suggest that the accuracy obtainable is comparable to classical stereo photogrammetry, if the optical components are of similar quality and resolution. Whilst CCD cameras are significantly cheaper and more readily available than conventional metric cameras, they offer poorer resolution than standard metric types. However, it has been found that with smoothing algorithms and sub-pixel interpolation (El-Hakim3) this level of camera resolution is adequate.

IMAGE PROCESSING

The success of the low level image processing provides the key to achieving an eventual and accurate 3D reconstruction. Standard textbook solutions to segmenting and extracting the contour patterns from the captured images proved to be far from optimal. The most evident of the problems that exist is the poor contrast level in the images, and this can be seen in a digitized image shown in Figure 2. In order to utilize all the possible grey levels (255 in total, 0 representing black and 255 white) some form of contrast stretching or enhancement needs to be applied. Simple techniques find the maximum and
minimum grey levels used and apply a global factor accordingly so that all the quantization levels are occupied. One method suggested the use of a harmless skin lotion applied to the subject to compensate for the poor contrast level but this immediately nullifies the advantage of the method being non-invasive.

Adaptive contrast enhancement methods operate on a small neighbourhood at any one time and apply a local scaling factor dependent on the mean and standard deviation of the available grey levels. To prevent indiscriminate noise enhancement the gain is limited between two limits. Prior to this a median filter is applied to remove any random noise present.

To extract the contour information from the digitized images some form edge enhancement or edge detection operator needs to be applied. After careful investigation a technique called line thinning, by Zhang and Suen, was adopted and developed for this specific application. It has inherent edge detection and produces lines that are one pixel wide essential for the later tracking stage. It was modified to give improved noise suppression and smoothing. The result of the above image preprocessing steps can be seen in Figure 3.

**CALIBRATION**

The subject of non-metric camera calibration has been a topic of recent research, the emphasis being on CCD cameras and the particular problems associated with their use. A comparison of several solid state cameras and their accuracy performance was made by Beyer et al. The inherent distortions, such as radial lens distortion, need to be accurately modelled or compensated for. The approach being researched is the one developed by Tsai. It is a linear technique that involves determining unknown camera parameters. It is also important to consider the exact relationship between the number of samples on one line of the CCD array and the corresponding number of samples in the framestore memory. This calibration procedure has been the one adopted by several close range photogrammetric applications.

3D RECONSTRUCTION

The tracking algorithm uses certain a priori knowledge of the structured light pattern. All spurs less than a certain number of pixels and not part of a continuous contour are assumed to be noise and deleted. It is also possible to state that the stripes are vertically preponderant and parallel in nature (contours can never cross). The algorithm also takes into account the direction and length of the currently tracked contour. Labelling of each individual contour is also carried out at this stage. The correspondence problem or grid ambiguity is solved using special grid patterns, an example is shown on the subject in Figure 2.

The final stage of the 3D reconstruction is the triangulation process common to all stereo systems that recovers the actual co-ordinate information so
that it can be plotted. It relates the 3D position of a point on the object surface to the detected image point in the camera plane and to the corresponding projected point in the slide. This is done by a similar technique to the solution of the camera and projector calibration matrices, and uses the computed coefficients. Using standard CAD and graphics packages the data can then be plotted as a wireframe model or rendered to give the appearance of soft tissue.

**DISCUSSION**

The technique described compares well and has distinct advantages over other existing facial measurement methodologies. It is in a sense an extension of the classical stereo photogrammetric case. The use of CCD cameras has been made possible due to research into calibration schemes and the mathematical modelling of geometric errors and distortions. This indicates that results are possible which are accurate, reproducible and reliable. The use of smoothing and sub-pixel interpolation schemes coupled with initial findings from experimentation into calibration methods also confirms that accuracy to the desired level of 1 mm can be obtained, a figure of merit given by clinicians and orthodontists. If required the use of higher resolution CCD arrays would further improve the accuracy of the method and will become more viable as the cost decreases.

The next stage is to compare the results and give an independent verification of accuracy with reference data sets from anthropometric data from a human face or polystyrene model, with similar properties and characteristics. The comparative analysis, involving two sets of collected data, as in the case of postsurgical monitoring, or the composite analysis using two different modes of collection i.e. merging with X-ray data, requires the registration problem to be adequately solved. The use of visible or standard anatomical landmarks is the current solution, but their recognition can be a subjective process. This issue must also be addressed in the future.

The technique requires no skilled operators to drive the system or the subsequent image processing, and patient preparation is also negligible, i.e. no makeup is needed due to the contrast enhancement algorithms. Incorporation of the surface along with X-ray data into CAD style programs will give clinicians an extremely versatile tool for analysis of soft tissue and bone movement prior to surgery.

**CONCLUSIONS**

The results obtained thus far concentrate on the essential low level processing of the digitized data tackling the problem of poor contrast images and the task of reliably extracting the 3D depth information using dedicated image processing algorithms. This has enabled much of the data analysis to be automatic, rendering the technique far more viable in a clinical environment. A non-invasive and inexpensive facial scanning system has been developed to enable pre-surgical analysis, in conjunction with X-ray data, and especially long term postsurgical monitoring. The data capture time is insignificant and the proposed level of accuracy adequate. As such, the technique is extremely versatile and can also be applied to other body surfaces and hence a variety of other medical applications. All the algorithms discussed have been implemented on a standard PC with the view to eventual widespread use in hospitals and clinics.

**ACKNOWLEDGEMENTS**

Thanks are due to the Wessex Health Authority who have collaborated with the work and the Science and Engineering Research Council who have provided the funding.

**REFERENCES**