Surface motion reconstruction for a DIET system

Richard G. Brown[1], Christopher E. Hann[1], J. Geoffrey Chase[1], Larry E. Ray[2] and Xiaoqi Chen[1]
[1]Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand
[2]Eastman Kodak Company, Kodak Research Laboratory, Rochester, New York, USA
rgb57@student.canterbury.ac.nz, chris.hann@canterbury.ac.nz

Abstract

A Digital Image-based Elasto-Tomography (DIET) system for breast cancer screening has been proposed in which the elastic properties of breast tissue are recovered by solving an inverse problem on the surface motion of a breast under low frequency (50-100 Hz) mechanical actuation. The proposed means for capturing the surface motion of the breast in 3D is to use a stroboscope to capture images from multiple digital cameras at preselected phase angles. Photogrammetric techniques are then used to reconstruct matched point features in 3D. Since human skin lacks high contrast visual features, it is necessary to introduce artificial fiducials which can be easily extracted from digital images. The chosen fiducials are points in three different colours in differing proportions randomly applied to the skin surface. A three-dimensional signature which is invariant to locally Euclidean transformations between images is defined on the points of the lowest proportion colour. The approximate local Euclidean invariance between adjacent frames enables these points to be matched using this signature. The remaining points are matched by interpolating the transformation of the matched points. The points between the two cameras are matched by matching ellipses in image space. Successful results are presented for simulated image sequences and for images of a mechanically actuated viscoelastic gel phantom showing that the overall procedure is suitable for a DIET system.

Keywords: Breast cancer detection, digital imaging, surface motion reconstruction,invariant signatures

1 Introduction

Breast Cancer is a serious health problem amongst women. It is the most common form of female cancer and is the second most fatal cancer among women worldwide. One in ten women will suffer from breast cancer during their lives and 25% of those who develop it are predicted to die from the disease.

Mammography represents the principal and most effective technology currently available for breast cancer screening. However patients often report a great level of pain and discomfort during the process of compressing the breast. The patients are also constantly subjected to radiation if regular screening occurs and this provides an additional health risk.

Digital Image-Based Elasto-Tomography (DIET) is an emerging new technology that uses digital imaging of a sinusoidally actuated breast surface to determine the surface motion of the tissue. The surface motion is then used to reconstruct the three dimensional internal elasticity distribution. The DIET system could potentially increase screening compliance rates, as the portable and low cost nature of the device means that screening is more accessible, as well as eliminating the disincentive for screening that the often uncomfortable mammogram process initiates.

1.1 Problem Description

This paper is concerned with the problem of acquiring the breast surface motion using multiple digital cameras, or more precisely, reconstructing in 3D the motion of a dense set of feature points on the breast surface.

3D motion reconstruction involves a number of steps. These are: synchronising images between cameras, tracking the motion of feature points in the individual camera frames, calibrating the cameras, and finding correspondences between the tracked sequences of points between cameras. Once the correspondence problem has been solved, the 3D location of each point at each time step can be found by triangulation.

In this paper, large numbers of coloured fiducial points are randomly applied to the surface. These points are tracked by a novel method which matches local Euclidean-invariant signatures based on closest point properties. The correspondence problem is then solved by taking advantage of the fact that points on the surface of a sinusoidally actuated elastic object will follow elliptical paths in space, and hence elliptical paths in each camera frame. Unlike point and line features, there exists a correspondence condition for matching ellipses between cameras [1] which is used in conjunction with the epipolar constraint to match the ellipses.

Results are presented from simulations using points on the surface of a finite element model of a harmonically actuated elastic cylinder.

2 Preliminaries

2.1 Camera Model

Digital CCD cameras can be accurately modeled as perspective projection pinhole cameras. The 3D world

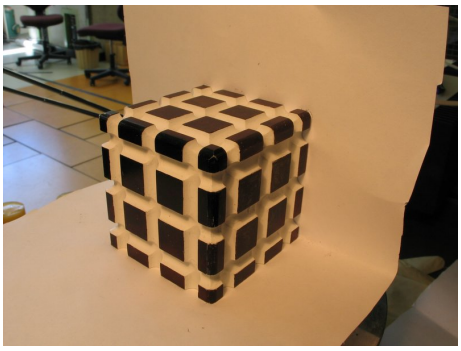

Figure 1: Cube used to calibrate cameras

space $\mathbb{R}^3$ can be embedded in projective space $\mathbb{P}^3$ with homogeneous coordinates (X, Y, Z, W) and the image coordinate space $\mathbb{R}^2$ can be embedded in the projective plane $\mathbb{P}^2$ with homogeneous coordinates (u, v, w). The corresponding coordinates in $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively, are given by $(\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W})$ and $(\frac{u}{w}, \frac{v}{w})$. Similarly, measured 3D coordinates and 2D image coordinates can be embedded in $\mathbb{P}^3$ and $\mathbb{P}^2$ respectively by the maps $(X, Y, Z) \mapsto (X, Y, Z, 1)$ and $(u, v) \mapsto (u, v, 1)$.

The camera represents a projection between $\mathbb{P}^3$ and $\mathbb{P}^2$, which can be represented by a homogeneous matrix $P \in \mathbb{R}^{3\times4}$, whose kernel is the projection centre of the camera. The projection from world coordinates $\mathbf{X} = (X, Y, Z, W)$ to image coordinates $\mathbf{u} = (u, v, w)$ is then described by the linear equation

$$\lambda\mathbf{u} = P\mathbf{X} \tag{1}$$

where λ is a nonzero scalar. The calibration matrix $\mathbf{P}$ can be factored into $P = K[RT]$ where K is an upper triangular matrix representing the intrinsic camera parameters, $R \in \mathbb{R}^{3\times3}$ is a rotation matrix describing the relative orientations of the camera and world frames, and $T \in \mathbb{R}^3$ is the origin of the world coordinate system in the camera frame.

3 Method

3.1 Calibration

The cameras are calibrated from images of a precisely machined calibration cube as depicted in Figure 1. The vertices on three faces of the cube are automatically detected and identified with the corresponding known locations of squares. See [2] for more details.

3.2 Fiducials

Because of the large density of features required, designing a system of fiducials, or marker points, which would be viable in a clinical context is a somewhat difficult problem. The markers need to be very easy to apply and remove, and yet still be able to be uniquely identified in images. The solution devised to this problem is to randomly apply small dots in different colours to the breast surface, and to design an invariant signature on the dots which can be used to identify the points. For the purposes of this article, the dots are red, green, and blue. The location of each of these points is easily extracted from colour images by converting the images into the HSV colour space, threshholding the hue (H) channel, and choosing the centroid of each resulting blob as the point location. Note that the centroid of a circle/ellipse is invariant under perspective projection.

3.3 Feature tracking

3.3.1 Tracking using invariant signatures

The idea behind feature tracking using signatures is to construct a mapping from the image space to some abstract *signature* space which is invariant under the image point motion. The points can hence be matched by matching their signatures in this signature space. In this paper, a simple signature based on invariants of the Euclidean group is presented, which is suitable for tracking applications in which the non-Euclidean component of the transformation in image space is small. The concept of a signature based on geometric invariants for registration was presented in [3], where a Euclidean-invariant signature curve was constructed for smooth curves in $\mathbb{R}^2$ and used for applications in object recognition.

The signature presented here is based on a 3-point joint invariant of the Euclidean group which is composed of the three 2-point joint invariants between each pair of points, i.e. the interpoint distances. The Euclidean group was chosen of a less restrictive transformation group such as the affine or projective group because the invariants, because they involve only distances, and not ratios or other nonlinearities are significantly more resilient to noise. The point triples are chosen to be nearby each other, which means the transformation must be only locally Euclidean, as only local Euclidean distances are used. Because the transformations are not exactly Euclidean, the signature space does distort between frames, but the signature space is sufficiently spread that the signature points are still able to be uniquely identified, i.e. the motion to point spacing ratio is significantly lower in the signature space.

3.3.2 Signature definition

Consider image points in two-dimensional Euclidean space, $\mathbf{u} \in \mathbb{E}^2$, and denote the Euclidean distance metric by $d(\dot{;}\dot{)}$. A signature function $f : (\mathbb{E}^2)^{(\times3)} \mapsto \mathbb{R}^3$ can be defined on ordered triples of these points as follows:

$$f : (\mathbf{u}^1, \mathbf{u}^2, \mathbf{u}^3) \mapsto \\ (d(\mathbf{u}^2, \mathbf{u}^1), d(\mathbf{u}^3, \mathbf{u}^1), d(\mathbf{u}^3, \mathbf{u}^2)) \tag{2}$$

Note that Eq. 2 defines a complete set of functionally independent joint invariants for a three point configuration, as the 2D configuration of points has six

degrees of freedom, and a Euclidean transformation three, hence there are three functionally independent joint invariants.

To find the ordered triples, the signature is defined on the points of one colour only (red). The ordered triple for each red point is defined to be itself, along with its nearest blue, and nearest green neighbour by Euclidean distance. To facilitate matching high point densities, the red points are applied at a significantly lower density than the green and blue points, which means that a smaller quantity of points is matched using the signature method, allowing the rest to be matched by interpolating the motion of the matched red points. The matching procedure is summarised in the next section.

3.3.3 Tracking procedure

The tracking procedure consists of a few simple steps.

(1) extract all of the red, green, and blue point locations from the images

(2) Find the nearest blue and green neighbours to each red point to form the point triples

(3) Compute the signature (Eq. 2) for each red point

(4) Match triples by matching their signatures in signature space, discarding matches if they violate some upper bound on the transformation magnitude

(5) Match the remaining unmatched points by interpolating the transformation of the matched points

3.4 Feature correspondence between two views

The human breast can be well-modelled with a linear model. This means that if it is sinusoidally actuated, points on the breast surface will move in an ellipse in space. Moreover, the perspective projection of these ellipses onto camera images will also be ellipses. The tracked contours from the previous section over one period of the actuation should thus form an ellipse. Ellipses can be fitted by nonlinear least squares, minimising geometric error, to these points, see [4]

3.5 Ellipse correspondence

The fitted ellipses are represented as 3x3 symmetric matrices in homogeneous coordinates. Following the derivation in [1], let C, C' be two corresponding ellipses in two different images, i.e. $\mathbf{u}^T C \mathbf{u} = 0$ and $\mathbf{u'}^T C' \mathbf{u'} = 0$ are the equations for the ellipses. Then, in homogeneous world coordinates $\mathbf{X}^T A \mathbf{X} = 0$ and $\mathbf{X}^T B \mathbf{X} = 0$ where $A = P^T C P$, $B = P'^T C' P'$ and P, P' are the projection matrices of the two cameras. A, B are rank 3 quadric surfaces, i.e. cones with vertices at the centres of projection of the two cameras. Enforcing the constraint that the intersection of these two cones be a conic section yields the following constraint on the entries of A and B

$$\Delta = I_3^2 - 4 I_2 I_4 = 0 \tag{3}$$

where $p(\lambda) = I_1 \lambda^4 + I_2 \lambda^3 + I_3 \lambda^2 + I_4 \lambda + I_5$ is the characteristic polynomial of the matrix pencil $A + \lambda B$. See [1, 5] for more details.

3.6 Correspondence Procedure

Firstly, the epipolar constraint is used to identify potential corresponding ellipses within a threshhold between images. This saves having to evaluate Eq. 3 for all ellipse pairs between the images. This narrows down the potential ellipse matches to a much more manageable number per ellipse. Because the expression 3 is based on 4th order polynomials, the values can be fairly large. It is thus more practical to deal with $\log \Delta$ instead.

Δ is evaluated for each pair of potentially corresponding ellipses. The resulting value is considered to be the *cost* of the matching. The best matching of ellipses is determined by solving the resulting weighted matching problem by the standard Hungarian Method, see e.g. [6] for details.

4 Results

4.1 Simulation

A finite element of an elastic cylinder with similar elastic properties to a human breast with a high stiffness inclusion was used to generate realistic data. The model was developed by Ashton Peters, from the University of Canterbury. The 17000 node finite element mesh was produced in GAMBIT, and the model simulated in Fortran 90, with the matrix inversion being done with the direct sparse matrix inversion package, MUMPS. The model was simulated with a 50Hz actuation frequency, with a 0.5mm peak to peak amplitude. See Figure 2 for a rendering of the cylinder.

400 Points, 80 red, 160 green, and 160 blue were projected onto one quarter of the surface of the cylinder. The motion of the finite element nodes was interpolated onto these points. Two camera models were set up 90° apart at a distance of 30cm from the cylinder. The resulting image point locations were computed over one phase of the actuation, with 20 frames. Gaussian noise was added to the measured image point locations, with standard deviation in each orthogonal direction of 0.2 pixels. This value is reasonable, as the image point locations in practice will be determined by taking the centroid of a blob of $O(100)$ pixels. The expected measurement error should therefore be well under one pixel.

The points were then tracked using the algorithm

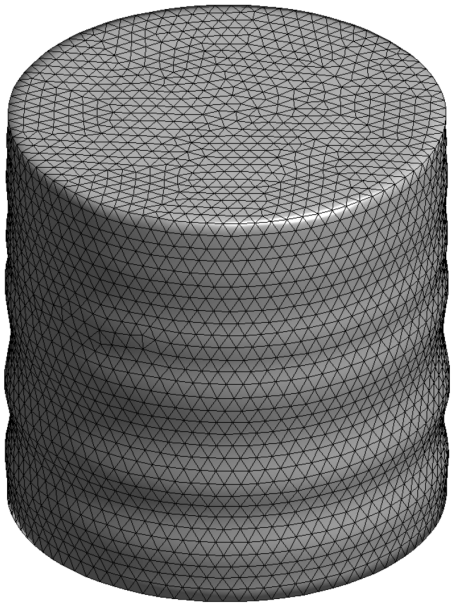

Figure 2: Finite element cylinder model used for simulation

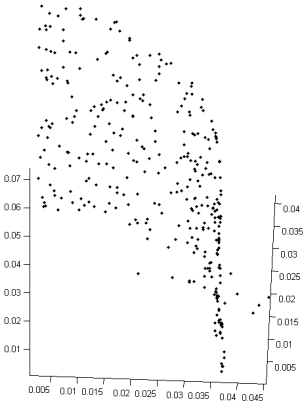

Figure 3: Reconstructed surface

Figure 4: Reconstructed surface

described in Section 3.3.3 over 20 frames, including the last frame to the first frame, giving elliptical contours. Those points that were not successfully tracked through all 20 frames (i.e. the 20th point in the sequence doesn't track back to 1) were discarded.

The reconstructed points are shown in Figure 3.

4.2 Gel Phantom Simulation

4.2.1 Experimental setup

The computer experiment from Section 4.1 was duplicated on real laboratory equipment. A cylindrical gel phantom made from an A-431 and LSR-05 silicone gel composite was placed on the platform of a mechanical actuator. A number of red, green, and blue coloured speckles made from finely cut coloured paper were applied to the surface of the cylinder, see Figure 4. The actuator was controlled by a simple control system written in Simulink and implemented in real time in dSpace to give a 50Hz, 1mm peak to peak sinusoid at the actuator plate. Still images at arbitrary phase were generated by strobing the cylinder in a dark enclosure at the actuation frequency at a user-specified phase separation from the actuator input signal, and the images were captured using a 6 megapixel Canon Powershot G5 digital CCD camera. A set of 20 images was captured in this manner to cover the entire sinusoid at 18 degree phase differences.

4.2.2 Results

Red, green, and blue regions were determined by directly threshholding the RGB image and comparing colour channels, and the point locations were taken to be the centroids of the coloured blobs,. This process takes a little under a second per image for a 6 megapixel image in Matlab. An example of the extracted points is depicted in Figure 5. Note that with this simple approach not all the image points are recognised, however this is not a concern as only a certain density of points is needed to be matched, rather than every individual point. The matching procedure was tested on various image pairs, and it was found that it successfully matched over 90% of the points with fewer than 1% mismatches (from visual inspection). The points that didnt get matched werent matched for one of three simple reasons. Firstly, some points were not picked up by the simple feature detection process. Therefore some points simply did not have corresponding points. Secondly, for some points only part of the coloured region comprising the point was detected. Finally, in some regions there

Figure 5: Reconstructed surface

were not sufficient red points, and hence insufficient point triples, to allow accurate interpolation of the motion of the points matched by the signature method in order to match the remaining points. As was done with the simulation example, the remaining number of mismatched points are easily ruled out by applying the constraint that the 20th point matches the first point.

5 Conclusion

This paper implemented an algorithm for reconstructing surface motion from digital images of an actuated gel phantom. The results in both the real and simulated cases validate the procedure and show that sufficiently accurate motions can be obtained that will would then go into a finite element based inverse problem that identifies the tissue distribution of the phantom. In the case of a breast, regions of high stiffness would suggest a tumour.

To implement the tracking, a Euclidean-invariant signature has been presented and successfully tested as a proof of concept for using local signatures for feature tracking and matching. The points were identified between the different cameras by matching ellipses in image space. Overall the results show good potential for practical implementation in a DIET system with the potential for low cost and portable breast cancer screening.

6 References

[1] L. Quan, “Conic Reconstruction and Correspondence from Two Views”, *IEEE Proceedings on Pattern Matching and Machine Intelligence*, 18(2), pp 151–160 (1996).

[2] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition (2004).

[3] E. Calabi, P. J. Olver, C. Shakiban, A. Tannenbaum and S. Haker, “Differential and numerically invariant signature curves applied to object recognition”, *International Journal of Computer Vision*, 26, pp 107–135 (1998).

[4] W. Gander, G. H. Golub and R. Strebel, “Least-Squares Fitting of Circles and Ellipses”, *BIT Numerical Mathematics*, 34(4), pp 558–578 (1994).

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, ISBN: 080185413X, third edition (1996).

[6] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization*, Dover Publications, Inc. (1998).