

# Vision-based 3D Surface Motion Capture for the DIET Breast Cancer Screening System

Richard G. Brown

Department of Mathematics and Statistics  
University of Canterbury, Christchurch, New Zealand  
Email: R.Brown@math.canterbury.ac.nz

Chris E. Hann

and J. Geoffrey Chase  
Department of Mechanical Engineering  
University of Canterbury, Christchurch, New Zealand

**Abstract**—Breast cancer is one of the most prevalent forms of cancer in the world today. The search for effective treatment and screening methods is a highly active area of research. The Digital Image-based ElastoTomography (DIET) project is a new breast cancer screening system under development, where surface motion from the mechanically actuated breast is measured in 3D, and used as input to an inverse problem solving for breast elasticity. Cancerous lesions appear as high contrast features, being an order of magnitude stiffer than healthy tissue.

The 3D motion capture is measured by an array of digital cameras using computer vision techniques. This paper presents a computer vision imaging system for the capture of 3D breast surface motion for the DIET system, including the image acquisition system, camera calibration, and 3D surface and motion reconstruction.

Results are presented for experiments performed with silicone gel phantoms, with conditions designed to replicate the clinical procedure. Full 3D surface motion is successfully captured using an array of 5 cameras. Some successful results from the DIET inverse problem are also presented to demonstrate the viability of the system in practice.

## I. INTRODUCTION

Breast cancer is a significant health problem in both developed and developing countries. It is estimated that each year the disease is diagnosed in over one million women worldwide and is the cause of death of over 400,000 women [1]. In New Zealand, breast cancer accounts for the highest mortality rate of all cancers among women and New Zealand has the sixth highest breast cancer death rate of 173 developed countries.

There are many treatment options available, including surgery, chemotherapy, radiation therapy, and hormonal therapy. These treatments are significantly more effective in reducing the mortality of the disease if detected early through breast cancer screening programmes. The most common method for early detection of breast cancer is mammography. Mammography, however, can cause significant discomfort to the patient and requires radiation exposure, a further health concern.

Digital Image-based Elasto-Tomography (DIET) is an emerging technology for non-invasive breast cancer screening. The DIET system uses digital imaging of an actuated breast surface to determine tissue surface motion from a specified input. It then reconstructs the 3D internal tissue stiffness distribution from that motion. Regions of high stiffness suggest cancer since cancerous tissue is between 3 and 10 times stiffer than healthy tissue in the breast [2], [3], [4].

The DIET approach eliminates the need for X-Rays and excessive, potentially painful compression of the breast [5] required by a mammogram. Hence, screening could start much younger and might enjoy greater compliance [6].

Presently, there are other Elasto-Tomographic methods based on magnetic resonance [7] and ultrasound [8] modalities. Both methods are capable of measuring the tissue elasticity and are undergoing rapid development. However, they are also costly in terms of equipment and take significant time to use. They are therefore of limited utility in a practical wide-scale screening application.

The DIET system, in contrast, is computer-based and is thus potentially low cost and portable, so the technology could be used in any medical centre, particularly in remote areas. In addition, the use of silicon technology ensures that as it improves and scales upward in capability so will the DIET system performance. This scalability of performance is not true for X-Ray or ultrasound based approaches.

This paper is focussed on the imaging side of the DIET system and tests the ability to accurately capture actuated tissue motion on realistically shaped breast silicon phantoms.

## II. BACKGROUND

### A. DIET project

The DIET (Digital Image-based ElastoTomography) concept is a new methodology for breast cancer screening. The breast is mechanically actuated sinusoidally at a low frequency (around 50-100Hz). This motion is visible as a wave pattern on the breast surface. With a linear elasticity model, knowledge of the breast surface motion in 3D is theoretically sufficient information to reconstruct the stiffness distribution of the interior of the breast, and thus, because cancerous tumours are around an order of magnitude stiffer than healthy tissue, could be used to diagnose breast cancer. In practice the elasticity reconstruction will be performed by a parametrised inverse problem.

The surface motion is captured by using a combination of strobed lighting and a number of calibrated digital cameras. Motion in 3D is computed by matching moving points between different cameras, and using computer vision techniques to construct the corresponding points in 3D.

## B. Camera Model

Digital CCD cameras can be accurately modeled as perspective projection pinhole cameras. The 3D world 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  $P \in \mathbb{R}^{3 \times 4}$ , 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

$$\lambda \mathbf{u} = P\mathbf{X} \quad (1)$$

where  $\lambda$  is a nonzero scalar. The calibration matrix  $P$ , if computed by real world and image point correspondences, can be factored into  $P = K[R \ \mathbf{t}]$  where  $K$  is an upper triangular matrix representing the intrinsic camera parameters,  $R \in \mathbb{R}^{3 \times 3}$  is a rotation matrix describing the relative orientations of the camera and world frames, and  $\mathbf{t} \in \mathbb{R}^3$  is the origin of the world coordinate system in the camera frame. A camera is said to be fully calibrated if  $K, R, \mathbf{t}$  and hence  $P$  are known.

## III. METHOD

### A. Fiducial system

Human skin lacks high-contrast features which are easily extractable from digital images. In order to reconstruct individual points in 3D, points on the surface must be able to be extracted from images taken from different cameras, i.e. they must be viewpoint invariant. The easiest way to define such features is to introduce artificial fiducials to the surface. For reasons that will be described in the following sections, the fiducials used in this paper are randomly applied identical points in three colours, red, green, and blue.

### B. Feature tracking

The features are tracked using the novel Euclidean-invariant algorithm described in [9] which is based on a system of coloured fiducial points as mentioned in the previous section. This method of feature tracking uses geometrically invariant properties of local configurations of the coloured point locations to match points between frames, rather than using image-based correlation techniques. The advantage of this approach is that points can be matched over a wider range of transformations, notably those involving a large translational component. Additionally, a significant advantage in computational speed is observed. If the motion is small, then an even simpler approach can be used, nearest neighbour matching.

The output of this process is the path of each fully tracked point over one phase of motion in each set of images. Note that

because the breast can be well-modelled by a linear stiffness model, motion of each point on the breast surface resulting from sinusoidal actuation will be an ellipse, and hence the paths in images from each camera will also be elliptical. This property can potentially be used to decrease the number of frames necessary to identify the motion path of each point, as in principle an ellipse can be computed from only five points.

### C. 3D Reconstruction

The algorithm outlined in this section is a method of reconstructing the coordinates of a dense set of points on the surface in 3D based on their image locations. Conventional approaches to this problem involve finding means of identifying corresponding points between images, and then triangulating these to find the resulting 3D position of the point. In this research an alternative approach is utilised. The epipolar constraint is used to find a set of candidate matches for each point. All potential matching pairs are then reconstructed in 3D, giving a large point cloud, contained within which is the desired surface. With suitable point density and surface smoothness, the desired 3D surface can be extracted from this set of 3D points.

1) *Epipolar constraint*: Consider two images  $I_1$  and  $I_2$  from two fully calibrated cameras. A point  $p$  in  $I_1$  can be represented in homogeneous coordinates as  $\mathbf{p} = (p_1, p_2, p_3)$ . The geometry of the two-camera configuration constrains any potential match for  $p$  in  $I_2$  to fall upon an *epipolar line*  $l$  [10]. This constraint can be represented algebraically by the following matrix equation

$$\mathbf{p}^T \mathbf{F} \mathbf{q} = 0 \quad (2)$$

where  $\mathbf{p}, \mathbf{q}$  are the homogeneous coordinates of  $p$  and  $q$ ,  $F$  is a rank 2 matrix known as the *Fundamental Matrix* which encapsulates the epipolar geometry and can be computed from the two camera projection matrices. The homogeneous coordinates of the epipolar line  $l$  are given by  $\mathbf{l} = \mathbf{p}^T F$ , i.e. Eq. 2 becomes  $\mathbf{l} \mathbf{q} = 0$ . In practice, points  $\mathbf{p}$  and  $\mathbf{q}$  are said to satisfy the epipolar constraint if

$$\mathbf{p}^T \mathbf{F} \mathbf{q} < \epsilon_{ep} \quad (3)$$

for some threshold  $\epsilon_{ep}$  and appropriate scaling of  $\mathbf{l}$  such that Eq. 3 represents the perpendicular distance between  $q$  and  $l$ . The threshold  $\epsilon_{ep}$  is determined by the error in calibration and image point measurement.

2) *3D Reconstruction procedure*: It is assumed that the points on the surface are sufficiently dense that surface points within a radius  $r$  of each individual surface point lie on a plane, to good approximation. Given an approximate density of surface points,  $\rho$ , the radius  $r$  can be chosen so that on average  $n$  surface points fall within a ball of radius  $r$  of each point  $\mathbf{x}$ , denoted  $B_r(\mathbf{x})$ .

For each point  $\mathbf{x}$  in the point cloud, the RANSAC[11] algorithm is used to robustly fit a plane to the  $r$ -neighbourhoods  $B_r(\mathbf{x})$ , using some inlier distance threshold  $\epsilon$ . The fitted planes are constrained to pass through  $\mathbf{x}$ . If image point measurements are sufficiently small, then the density of points in the

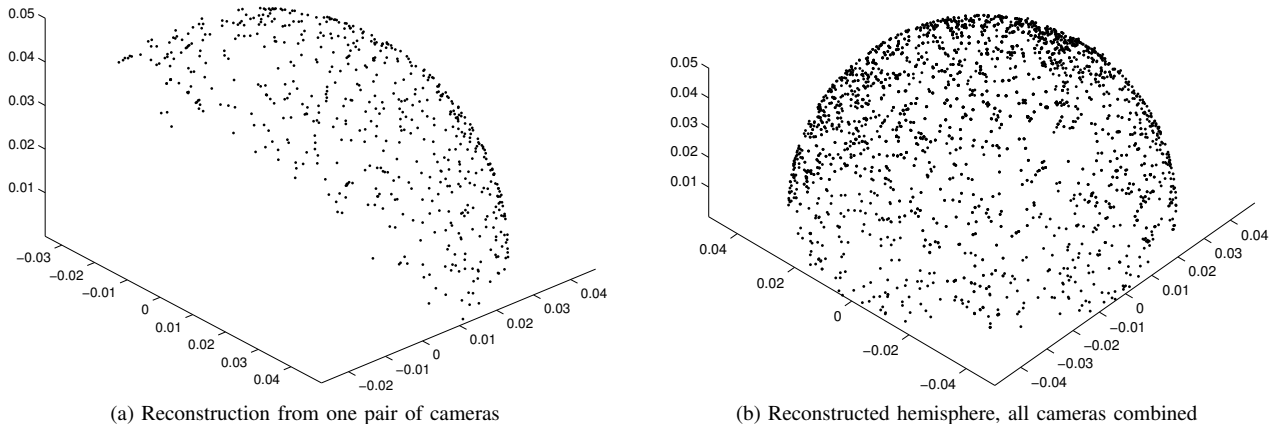


Fig. 1. 3D Reconstruction from computer simulation

vicinity of the surface will be much higher than elsewhere in the point cloud. Therefore, if the point  $\mathbf{x}$  is a surface point, then the inliers to the best robust plane fit are also likely to be surface points, and the fitted plane can be viewed as an estimate of the tangent plane to the surface at  $\mathbf{x}$ . If  $\mathbf{x}$  is not a surface point, then the fitted plane is essentially meaningless, as it is the best fit to a random cloud of points.

A means of reducing the density of random points in the point cloud, while preserving all of the surface points, is to use points of different colours. The use of colours adds a further constraint to the correspondence problem, and therefore reduces the number of matches generated by the epipolar constraint.

To find the surface, an *adjacency graph* of the points is constructed. Two points  $\mathbf{x}$  and  $\mathbf{y}$  are said to be adjacent if each is an inlier to the RANSAC tangent plane fit of the other, and if the normals of the fitted planes only differ by a small angle, dependent on the distance between  $\mathbf{x}$  and  $\mathbf{y}$ . The surface points are then extracted from the cloud by choosing the largest connected component from the resulting adjacency graph.

The surface density parameter  $\rho$  is required to be estimated for any experiment using this procedure. Experiments have shown that accuracy in  $\rho$  is not critical, for an experiment with points of density of around 10 pts/cm<sup>2</sup>, the reconstruction procedure works equally well for  $\rho$  anywhere between around 5 and 20. The parameter  $\epsilon$  is dependent on the camera geometry and on the accuracy of image point measurements. The parameter  $n$  (which determines  $r$ ) is chosen so that for each point, there is a high probability that sufficient neighbouring surface points are contained in the neighbourhood  $B_r(\mathbf{x})$  to construct the plane, while being small enough so that the region of the surface contained in  $B_r(\mathbf{x})$  is still approximately planar. Experiments have also shown that the parameters  $n$  and  $\epsilon$  are also relatively insensitive to variations, a good value of

$n$  for approximately uniformly randomly distributed points is 15.

#### IV. RECONSTRUCTION - COMPUTER SIMULATION

The test case for the reconstruction algorithm was a hemisphere with randomly distributed points generated by a homogeneous Poisson process on the surface. The hemisphere had radius 5cm, and synthetic images were taken from five cameras evenly spaced around the hemisphere. The cameras were evenly positioned on a circle of radius 20cm, elevated 20cm above the base of the hemisphere. Five cameras were used, as this was the minimum required to have each region of the hemisphere viewed by two overlapping cameras. The camera matrices were defined with parameters such that the image of the hemisphere approximately filled a 1600 × 1200 pixel region. The surface was reconstructed by performing the reconstruction pairwise for pairs of adjacent cameras, and then combining the resulting surfaces for the final result.

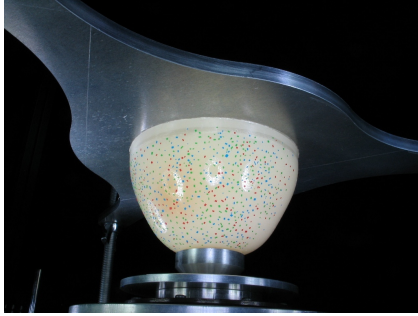
Gaussian noise of standard deviation  $\sigma = 0.5$  pixels was added to each coordinate of the image measurements. Parameters used were  $\rho = 10$  pts/cm<sup>2</sup>,  $n = 15$ , and  $\epsilon = 0.5$ mm. The results are shown in Figs 1a, 1b

#### V. RECONSTRUCTION - LABORATORY EXPERIMENT

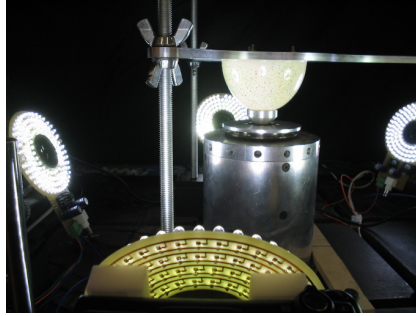
##### A. Experimental setup

1) *Gel phantom*: A gel phantom was created from molding silicon gel with similar elasticity characteristics to human breast tissue. The silicon was pigmented to yield a more natural flesh colour. Small red, green, and blue points were randomly applied to the surface. See Fig. 2a for a picture.

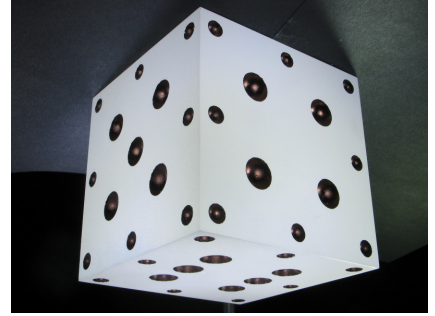
2) *Actuator*: The actuator is a voice-coil style electromechanical actuator, and provides sinusoidal actuation at frequencies between 20 and 200 Hz at amplitudes of up to 1mm. Amplitude is controlled via feedback from an LVDT position sensor from a dSpace real time control environment.



(a) Silicon phantom under actuation with random coloured dots applied



(b) Actuator, gel phantom, and four of the five cameras fitted with LED ring flashes



(c) Calibration: Image of the calibration die

Fig. 2. Experimental Setup

3) *Cameras and lighting*: The cameras are standard consumer Canon Powershot G5 cameras operating at 2 megapixels. Each camera is fitted with an LED ring flash which can be synchronised with an external trigger signal (see Fig. 2b). These LED flashes are strobed at the same frequency as the actuation frequency via a control signal from the dSpace control system. The phase of the strobe signal can be varied with respect to the actuator signal, allowing images to be taken at arbitrary points in the actuation cycle.

4) *Software*: The dSpace control system is designed in Simulink, and has a real-time interface in Control Desk. Custom software was written to automatically coordinate the taking of the images and the adjusting of the strobe phase.

5) *Experiment parameters*: For the experiment presented in this paper, the phantom was actuated at 80Hz, with an amplitude of 0.75mm. 20 images were taken at even phase increments over the cycle from each camera.

6) *Calibration*: In order to reconstruct the surface, it is required that the camera is fully calibrated. The projection matrix  $P$  can be computed from correspondences between known world locations of a number of points on a calibration object and their corresponding image locations.  $P$  is estimated using nonlinear least squares minimising the reprojection error. See [10] for details.

The calibration object used for this experiment is an precisely machined  $100 \times 100 \times 100$ mm anodised aluminium die. The point locations used to compute the projection matrix are the centroids of the die "dots". Note that because the dots are circular, images of these will be ellipses, and the centroid can be computed in a viewpoint-invariant manner by mapping each face to a square by a homography. These dots can also be used to uniquely identify which faces are visible, and hence calibration is fully automatic from a single image of the cube. See Fig. 2c for an image of the calibration cube used.

Error in calibration was computed by reconstructing in 3D the measured image point locations of feature points on the calibration object in 3D, and comparing with their known coordinates. For these experiments, the mean reconstruction error thus obtained was of the order 0.1mm.

7) *Feature tracking*: The coloured points are extracted by applying simple colour thresholds to the individual RGB

channels of the images. The point locations are taken to be the centroids of the detected coloured blobs. As the motion was relatively small with respect to the point spacing, feature tracking was performed using nearest neighbour matching. The motion of the feature points from one camera is depicted in Fig. 3a.

### B. 3D Reconstruction

3D reconstruction was performed on the data points from the first frame. Parameters used were  $\epsilon_{ep} = 2$  pixels,  $r = 10$ mm, and  $e_{max} = 0.5$ mm. Points were reconstructed in 3D by performing 3D reconstructions for each of the 5 adjacent camera pairs individually and collating the results.

An example of the point cloud created by reconstructing all potential matches from the epipolar constraint stage is depicted in Fig. 3b. Because we have points of three different colours, we can eliminate a number of these mismatches, by only allowing potential matches to have the same colour. The same resulting point cloud is shown in Fig. 3c. The effect of this is to significantly decrease the density of points which are not on the surface.

The outlier rejection algorithm was then applied to the point cloud. For the cases considered, the surface was successfully reconstructed both with and without the colour information. The case with the colour taken into account performed significantly faster, however. The resulting reconstructed surface is shown in Figs. 3d and 3e.

## VI. CONCLUSION

This paper implemented an algorithm for reconstructing surface motion from digital images of an actuated gel phantom. All the key issues of digital image acquisition, camera calibration, surface reconstruction and point tracking were addressed. The end result is highly accurate tissue surface motion tracking which 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.

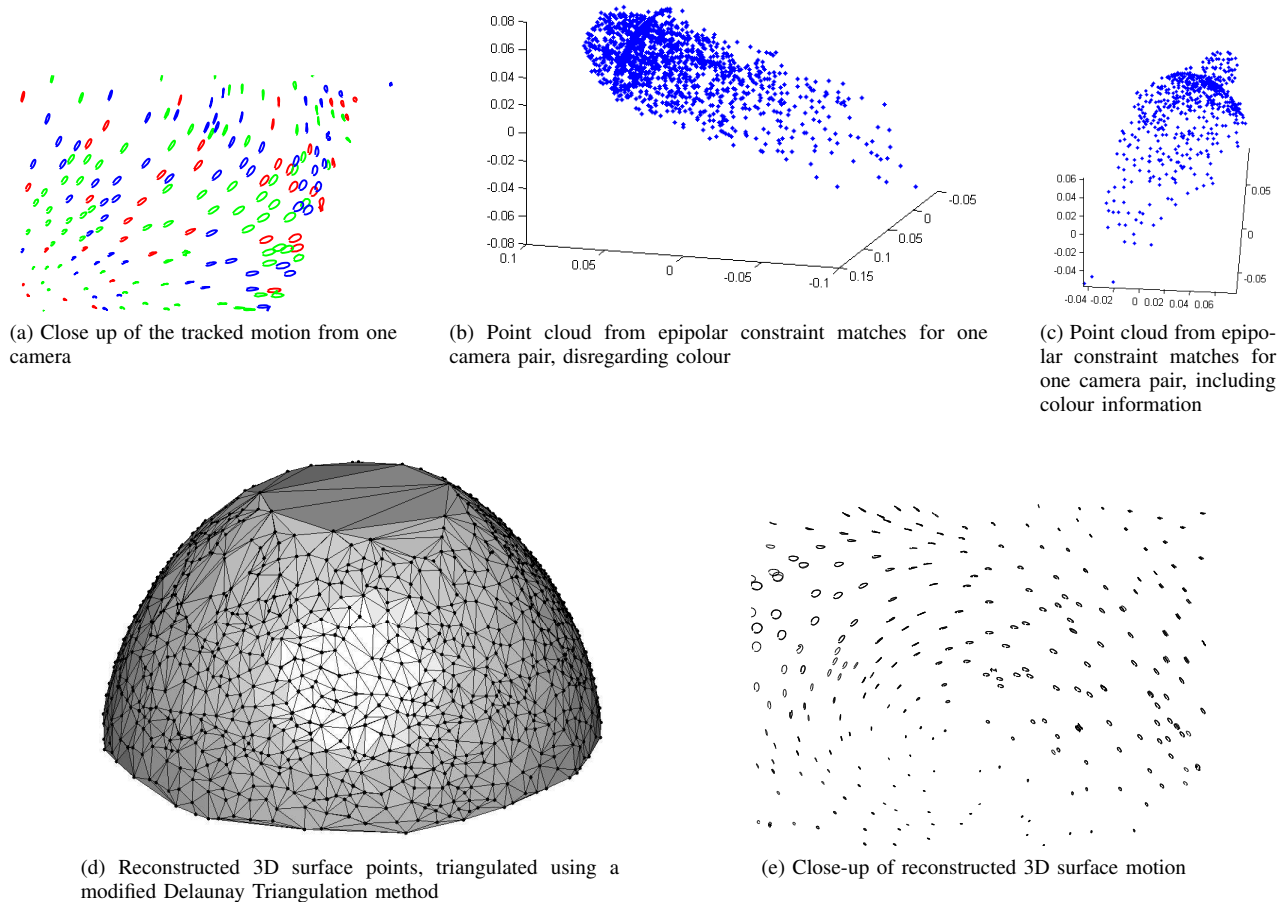


Fig. 3. Experimental Results

High frequencies of 100 Hz were captured using a combination of standard digital cameras and a stroboscope providing a simple cost-effective approach. Overall the results show good potential for practical implementation in a DIET system with the potential for low cost and portable breast cancer screening.

#### ACKNOWLEDGMENT

The authors would like to thank the Tertiary Education Commission (TEC) for a Top Achiever Doctoral Scholarship which supported (RGB) in this research.

#### REFERENCES

- [1] I. S. Fentiman, "Breast cancer in older women," *Breast Cancer Online*, vol. 5, 2002.
- [2] A. Samani, J. Bishop, C. Luginbuhl, and D. B. Plewes, "Measuring the elastic modulus of ex-vivo small tissue samples," *Phys. Med. Biol.*, vol. 48, pp. 2183–2198, 2003.
- [3] T. A. Krouskop, T. M. Wheeler, F. Kallel, B. S. Garra, and T. Hall, "Elastic moduli of breast and prostate tissues under compression," *Ultrasonic Imaging*, vol. 20, pp. 2183–2198, 1998.
- [4] P. W. Wellman and R. D. Howe, "Breast tissue stiffness in compression is correlated to histological diagnosis," *Harvard BioRobotics Laboratory Technical Report*, 2000.
- [5] A. Peters, A. Milsant, J. Rouze, L. Ray, J. G. Chase, and E. E. W. Houten, "Digital image-based elasto-tomography: Proof of concept studies for surface based mechanical property reconstruction," *Japanese Society of Mechanical Engineers (JSME)*, vol. 47, pp. 1117–1123, 2005.
- [6] M. Robertson, "Commercialisation and ip strategies for digital image-based elasto-tomography (diet)," Master's thesis, University of Canterbury, 2005.
- [7] T. Oida, A. Amano, and T. Matsuda, "Magnetic resonance elastography: in vivo measurements of elasticity for human tissue," *International Conference in Informatics Research for Development of Knowledge Society Infrastructure*, 2004.
- [8] R. L. Maurice, J. Ohayon, Y. Fretigny, M. Bertrand, G. Soulez, and G. Cloutier, "Non-invasive vascular elastography: theoretical framework," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 164–180, 2004.
- [9] R. G. Brown, C. E. Hann, J. G. Chase, and L. A. Ray, "Discrete color-based Euclidean-invariant signatures for feature tracking in a DIET breast cancer screening system," in *Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*. Edited by Manduca, Armando; Hu, Xiaoping P. *Proceedings of the SPIE, Volume 6511*, pp. 65110D (2007)., vol. 6511, Mar. 2007.
- [10] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [11] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.