Soft Tissue Tracking for Minimally Invasive Surgery: Learning Local Deformation Online

Peter Mountney[1,2] and Guang-Zhong Yang[1,2]

[1] Department of Computing, [2] Institute of Biomedical Engineering
Imperial College, London SW7 2BZ, UK

Abstract. Accurate estimation and tracking of dynamic tissue deformation is important to motion compensation, intra-operative surgical guidance and navigation in minimally invasive surgery. Current approaches to tissue deformation tracking are generally based on machine vision techniques for natural scenes which are not well suited to MIS because tissue deformation cannot be easily modeled by using ad hoc representations. Such techniques do not deal well with inter-reflection changes and may be susceptible to instrument occlusion. The purpose of this paper is to present an online learning based feature tracking method suitable for *in vivo* applications. It makes no assumptions about the type of image transformations and visual characteristics, and is updated continuously as the tracking progresses. The performance of the algorithm is compared with existing tracking algorithms and validated on simulated, as well as *in vivo* cardiovascular and abdominal MIS data. The strength of the algorithm in dealing with drift and occlusion is validated and the practical value of the method is demonstrated by decoupling cardiac and respiratory motion in robotic assisted surgery.

Keywords: Feature, tracking, matching, tissue deformation.

1 Introduction

With the maturity of Minimally Invasive Surgery (MIS), the clinical uptake is steadily increasing because of its recognized benefits to patients and healthcare providers, particularly in terms of reduced patient trauma and hospital recovery times. The performance of MIS, however, is complicated by a number of visuomotor and ergonomic challenges including misaligned visuomotor axes, the fulcrum effect during instrument manipulation, limited field of view, and loss of 3D vision and tactile feedback. The introduction of robotic assisted MIS has provided surgeons with improved visualization and enhanced dexterity. It also offers the possibility of integrating patient-specific preoperative/intraoperative data to allow imaged guided surgical navigation and intervention. For these techniques to be successful, particularly for cardiovascular and gastrointestinal surgeries where large scale tissue deformation is common, an important prerequisite is the accurate estimation and tracking of dynamic tissue deformation.

D. Metaxas et al. (Eds.): MICCAI 2008, Part II, LNCS 5242, pp. 364–372, 2008.

Recent work has shown that it is possible to perform 3D tissue tracking by using both monocular and stereo depth cues [1-3]. Researchers have also relied on gaze vergence through binocular eye tracking to facilitate real-time 3D tissue deformation recovery [4]. Two major issues identified in current vision based techniques include tracked feature *density* and *persistency*. The former dictates the level of detail of the deforming surface that can be reconstructed, whereas the latter is affected by mutual and self-occlusion of the tissue and instrument during the surgical procedure. Feature persistency is also heavily influenced by changes in operating field-of-view and lighting conditions. Current research has made significant inroads into improving density of the tracked features by using multiple depth cues to cater for the complex tissue geometry *in vivo*. However, they generally do not explicitly model nonlinear tissue deformation and may be susceptible to drift and occlusion.

The purpose of this paper is to present an online learning based feature tracking method suitable for *in vivo* applications. Feature tracking is formalized as a classification problem where we propose solutions to training the classifier with unlabeled data and adaptive updates during the tracking process. The approach makes no assumptions about the type of image transformations or visual characteristics enabling it to deal with nonlinear tissue deformation. It is demonstrated in this paper that with the proposed technique for general MIS scenes, as little as just 0.5 seconds may be required to start building up a complete feature representation. The performance of the algorithm is compared with other conventional trackers, and validated on simulated as well as *in vivo* cardiovascular and abdominal MIS data. The strength of the algorithm in dealing with drift and occlusion as well as tissue deformation is demonstrated.

2 Methods

2.1 Learning Based Feature Tracking

The effectiveness of a feature tracking algorithm is largely determined by how the appearance of the feature is represented. This consists of two elements, firstly which information to encode (*e.g.* color, edges, intensity, texture, gradient) and secondly how to represent the encoded information (*e.g.* the use of probability density histogram, histogram of gradients, templates, points, contours, active appearance models). The choice of what information to encode and how to represent the encoded information is context specific. For example, in [5] mean-shift is used to track deformable objects by making the assumption that color is the most salient information to encode. Approaches such as SIFT [6] represent scaled information as gradient oriented histograms.

It should be noted that these methods make *ad hoc* assumptions about which information will be most discriminative and how to encode it. They work well if the underlying assumptions hold. In MIS, changes in lighting and specular highlights can significantly alter the 2D appearance of the tissue. These environmental factors are exacerbated by 3D nonlinear tissue deformation. This makes *ad hoc* modeling of tissue appearance for consistent tracking difficult. Alternatively, it is possible to learn which information is most discriminative and how best to encode it. This concept has been adopted in hand writing recognition [8], object detection [9] and corner detection

[10] by learning offline the most discriminative representation of the data. Offline learning requires prior knowledge of the data which is not available during tissue tracking. In this paper, an online learning scheme is developed to extract discriminative information adaptively as the tracking process progresses such that it can cater for tissue deformation and environment changes while remaining robust to drift and occlusion. The main steps of the algorithm are outlined in Fig. 1 and detailed in the following sections.

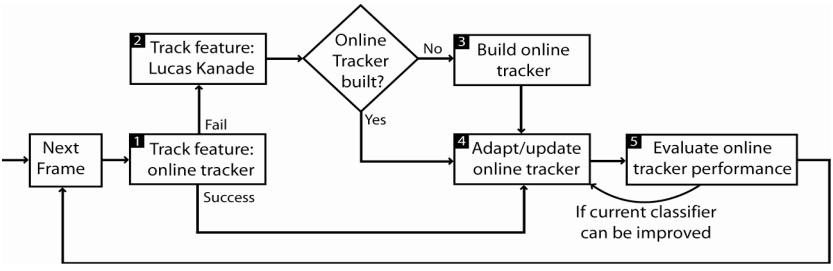

Fig. 1. A diagrammatic overview of the proposed learning based online tracking system

The algorithm is comprised of five main steps; 1-2) feature tracking is initially performed using a Lucas Kanade (LK) [7] algorithm and then with our online approach, 3) building the online tracker and learn a representation for a feature, 4) adapted and updated the feature representation and 5) tracker evaluation.

Building the Online Tracker

The feature tracking problem can be formalized as a classification problem where the goal is to classify the feature in a new image as a true match and classify all other features as false matches. In order to train a classifier we require a set of training data with true and false labels. Such training data can be obtained synthetically if the appearance of the feature can be well modeled or through manual labeling for offline learning. Neither of these approaches is suitable for tracking tissue therefore the classifier will need to be trained from unlabeled data.

To solve this problem, we propose to extract the training data online while the features are tracked. In this paper, features are extracted using Difference Of Gaussian [6] and Shi and Tomasi [11] detectors and initially tracked using a LK tracker. The key to our proposed method is to learn what information will be most appropriate for tracking, therefore the training data will consist of a number of image patches extracted directly from the image. The LK tracker enables the generation of a labeled set of true matches for the classifier. The set of false matches are then taken from the local area around the tracked feature. The labeled data provides the information which enables a set of patches S to be partitioned into two sets S_t and S_f representing 'true' and 'false' matches. An ID3 [12] decision tree is then used to iteratively partition S. For each patch in set S, a test compares two pixel values to identify if the first pixel is greater, similar or less in value than the second pixel. The entropy of

each subset is measured to identify the test that provides the maximum information and therefore the best partition, *i.e.*,

$$H(S) = |S|\log_2|S| - |S_t|\log_2|S_t| - |S_f|\log_2|S_f| \tag{1}$$

Exhaustive search using Eq. (1) is computationally prohibitive, this is solved instead by computing the log likelihood ratios [13] between distributions of S_t and S_f and applying a variance ratio to find the optimum solution. At each pixel location, we create histograms $t(x,y)$ and $f(x,y)$ and calculate the log likelihood

$$L(x,y) = \log\frac{\max(t(x,y),\delta)}{\max(f(x,y),\delta)} \tag{2}$$

where δ is set to be 0.001 to avoid dividing by zero. The variance ratio of the log likelihood is used to quantify the distance of the two classes, *i.e.*,

$$V(L;x,y) = \frac{\mathrm{var}(L;(t+f)/2)}{[\mathrm{var}(L;t) + \mathrm{var}(L;f)]} \tag{3}$$

where

$$\mathrm{var}\left(L;(a)\right) = \sum_i a(i)L^2(i) - \left(\sum_i a(i)L(i)\right)^2 \tag{4}$$

given the discrete probability density function a_i. This provides a measure of intra- and inter-class variance, and is capable of handling multimodal distributions unlike linear discriminate analysis.

Tracking Features in New Frames

A search area is defined based on the position of the feature in the last frame and at each point $p(x,y)$ in the search region, the patch around that point is classified using the decision tree. This classification can be performed quickly as the tests are simple and the false matches can be readily identified with only a few tests. This classification step results in a number of candidate points in the search region which represent the potential location of the feature. The feature is localized by examining the probability distribution $P(N_j) = |S_{tj}||S_t|^{-1}$ at the tree node N_j to determine if it is a correct match, where $|S_{tj}|$ is the number of true matches classified by node j and $|S_t|$ is the number of true matches classified by the entire tree. The best candidate point $p_{x,y}$ in the search area is then selected using the node distribution and a Gaussian kernel centered on the last known position.

Evaluating and Improving the Online Tracker Performance

Building the decision tree can be computationally intensive if the data set is large. However, testing the performance of our classifier is relatively fast. This is exploited in the proposed algorithm to adaptively build classification trees that best fit the observed data. The tree is built initially with a small set of data, this is then followed by evaluating its classification performance and further improving the classifier. The

online tracker's performance is evaluated by measuring the classification accuracy on the current data set. The metric used here is the false negative rate of the classifier, for which a high value indicates the classification tree is not suited to the data and its inherent information needs to be further exploited. False negatives indicate mismatches by the tracker where the test and distribution $P(N_j)$ at node N_j are not ideal representations of the data. Instead of rebuilding the entire tree, we can simply reclassify all the patches at node N_j adding the incorrectly classified patch. This has the effect of shifting the distribution to better represent all the observed data and may lead to new nodes being added to the tree. The final adaptive step in the update is to select the most discriminative color space for tracking. This follows the criterion set out in [13] where 49 color spaces are searched to identify the most discriminative. This uses the variance ration equations outlined in Eq. 3.

2.2 Extracting Intrinsic Tissue Motion

To demonstrate the practical application of this technique, the cardiac and respiratory motion of the tissue are extracted by performing Independent Component Analysis (ICA) of the tracked features. ICA is a statistical technique for separating signals into additive subcomponents assuming mutual statistical independence. ICA can be formulated to consider the recovered 3D motion (computed using stereo geometry) of the surface of the tissue to be the latent variables $m = (x, y, z)$ and the components of intrinsic motion as $s = (h, r)$. It attempts to find the transformation W such that $s = Wm + n$ where n is zero mean Gaussian noise. The components of m can be written as the weighted sum of the independent components, *i.e.*, $m = \sum a_i s_i$, where a_k is a vector of mixing weights which make up the mixing matrix $A = (a_1 \ldots a_n)$ where $W = A^{-1}$. The source s and the mixing matrix A are estimated adaptively with cost function $s_k = w^T m$ to maximize nongaussianality.

3 Experiments and Results

To evaluate the performance of the proposed online learnt tracker, results from simulated, porcine and *in vivo* data are compared to those from four conventional tracking techniques (Lucas Kanade[7] with template update, SIFT[6], and two mean-shift algorithms [13]).

3.1 Simulated Experiment with Known Ground Truth

For synthetic data, an image from a MIS procedure was taken and textured onto a 3D mesh, which was then warped with a mixture of Gaussian model to simulate the cardiac and respiratory induced tissue deformation. The mesh was projected onto a virtual camera for subsequent feature tracking. To better represent the real-life data, Guassian noise was added to the images. Fig. 2 (b) illustrates the tracking result for the synthetic data. It is evident that the LK tracker performed relatively well at the beginning of the experiment, but the performance rapidly declines due to error propagation resulting from tissue deformation. The detect/match approach of SIFT in this case also performed poorly. The number of points does not decline over time, it

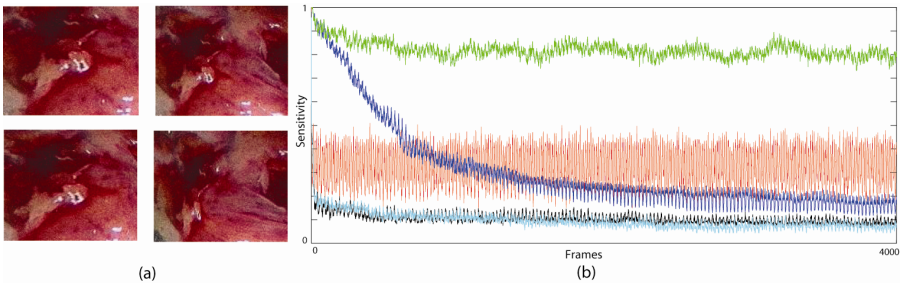

Fig. 2. Relative tracking performance for a synthetic data for the different tracking algorithms considered. **(a)** The simulated data by warping an image taken from a MIS procedure with known ground truth deformation characteristics. **(b)** Relative performance values for the five different tracking techniques compared; green – our online learnt tracker, red – SIFT, dark blue – Lucas Kanade, black – mean-shift 1 and light blue – mean-shift 2.

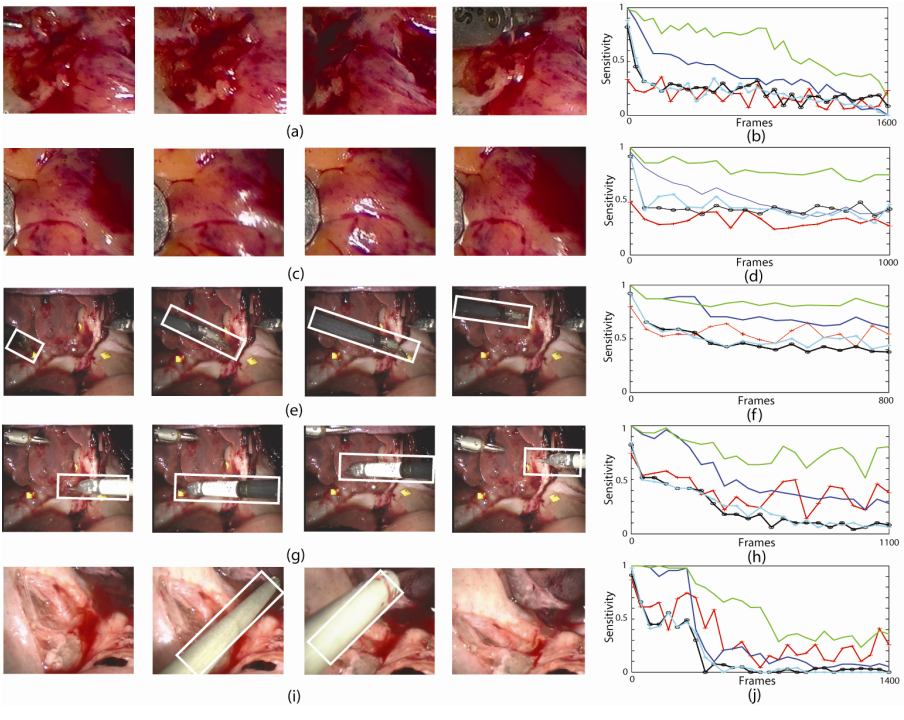

Fig. 3. Relative tracking performance for *in vivo* sequences. **(a,c,e,g,i)** Example frames taken from *in vivo* data available at [14], **(b,d,f,h,j)** the associated quantitative analysis results of the tracking algorithms. Five trackers are compared; green – our online learnt tracker, red – SIFT, dark blue – Lucas Kanade, black – mean-shift 1 and light blue – mean-shift 2.

oscillates as the tissue deforms, making it less attractive for continuous *in vivo* tracking. The performance of the two mean-shift algorithm is similar. This is not surprising as mean-shift only works well on self contained blobs of distinct color, which is

difficult to hold for *in vivo* applications. Large movements can also result in the feature falling outside the trackers basin of attraction which further contributes to the relatively poor performance achieved. For the proposed tracker with online learning, the overall performance is maintained, and the derived sensitivity outperforms all of the alternative techniques compared.

3.2 *In Vivo* Experiments

The performance of the proposed learning based online tracker was quantitatively evaluated on five *in vivo* sequences. The ground truth for the tracked features was obtained manually at 50 frame intervals. Fig. 3 demonstrates the five sequences used and the corresponding tracking results as compared to the four conventional tracking algorithms. Figs. 3 (a-d) show two beating heart sequences where artifacts due to bleeding, specular reflections and instrument occlusion have introduced significant problems to the LK, SIFT and mean-shift trackers. Similar to the synthetic experiment, the LK tracker exhibits drift and its performance degrades as the tracking process progresses. The SIFT and mean-shift trackers perform worse in the sequence shown in Fig. 3 (a) than Fig. 3 (b) as deformation in this sequence is more pronounced. The graphs in Fig. 3 (b) and (d) show more features can be tracked using the learning based method.

The effect of introducing instrument occlusion into the surgical field-of-view is shown in Figs. 3 (e-j). In Figs. 3 (e,g,i), instrument occlusion was introduced to the surgical field-of-view. Deformation in these sequences is mainly from respiration. In Fig 3 (e), only a small number of features are occluded, whereas in Fig 3 (g) the number is increased. In Fig 3 (i), almost all features are occluded at some point. Tracking in the last sequence is made more difficult as a suction device is used to remove blood, thus significantly changing the appearance of features.

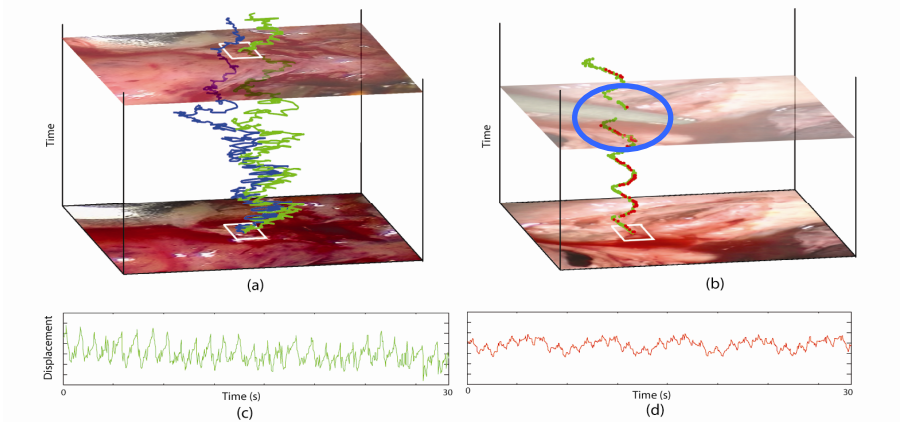

Fig. 4. (a) A single feature tracked over time showing drift with LK tracking in blue and the robustness of our approach in green. **(b)** Illustrates the problem of occlusion by a tool. Green – our online learnt tracker, red – SIFT. SIFT tracking is not continuous. **(c)** The first and **(d)** second components recovered using ICA from online tracking for motion compensation.

It is evident that once the feature is lost in the LK and mean-shift trackers, it can no longer be recovered. Detect and match tracking approaches such as SIFT are naturally suited to dealing with occlusion, but are not well suited to continuous tracking of deforming tissue. In contrast, the proposed learning based online tracker holds well for the experiment performed.

Figs. 4 (a) and (b) illustrate the problems of drift and occlusion in 3D spatio-temporal plots for the different techniques considered in this study. It demonstrates how feature representation with online learning can successfully overcome these problems. In this example, the online tracker was used to track deformation of the epicardial surface and the resulting ICA motion extraction shown in Fig 4 (c) and (d) clearly depicts cardiac and respiratory induced deformation.

4 Conclusion

In this paper, we have proposed a novel approach for feature tracking with online learning. The approach has been validated on simulated, porcine and *in vivo* data and compared to four conventional tracking techniques. We have demonstrated that the technique is robust to drift and capable of recovering from occlusion. The proposed technique is well suited to dealing with deforming tissue and unknown image transformations. Robust feature tracking is important for a range of applications in robotic assisted MIS including real-time depth recovery, pre- and intra-operative image registration, as well as prescribing dynamic active constraints.

References

1. Wengert, C., Bossard, L., Häberling, A., Baur, C., Székely, G., Cattin, P.C.: Endoscopic Navigation for Minimally Invasive Suturing. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 620–627. Springer, Heidelberg (2007)
2. Ortmaier, T., Groger, M., Boehm, D.H., Falk, V., Hirzinger, G.: Motion Estimation in Beating Heart Surgery. IEEE Trans. on Biomedical Engineering (52), 1729–1740 (2005)
3. Mountney, P., Lo, B.P.L., Thiemjarus, S., Stoyanov, D., Yang, G.-Z.: A Probabilistic Framework for Tracking Deformable Soft Tissue in Minimally Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 34–41. Springer, Heidelberg (2007)
4. Mylonas, G., Stoyanov, D., Deligianni, F., Darzi, A., Yang, G.-Z.: Gaze-contingent soft tissue deformation tracking for minimally invasive robotic surgery. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 843–850. Springer, Heidelberg (2005)
5. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-Based Object Tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence (25), 564–577 (2003)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (60), 91–110 (2004)
7. Lucas, B.D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. IJCAI, pp. 674–679 (1981)
8. Amit, Y., Geman, D.: Shape Quantization and Recognition with Randomized Trees. Neural Computation 9(7), 1545–1588 (1997)

9. Lepetit, V., Lagger, P., Fua, P.: Randomized trees for real-time keypoint recognition. In: Proc CVPR, vol. (2), pp. 775–781 (2005)
10. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
11. Shi, J., Tomasi, C.: Good Features to Track. In: Proc of CVPR, pp. 593–600 (1994)
12. Quinlan, J.R.: Induction of decision trees. Machine Learning 1 (1986)
13. Collins, R., Liu, Y., Leordeanu, M.: On-Line Selection of Discriminative Tracking Features. IEEE Trans Pattern Analysis and Machine Intelligence 10(27), 1631–1643 (2005)
14. Imperial College Visual Information Processing In: Vivo database, http://vip.doc.ic.ac.uk/