Pathological Site Retargeting under Tissue Deformation Using Geometrical Association and Tracking

Menglong Ye[1], Stamatia Giannarou[1], Nisha Patel[1,2], Julian Teare[2], and Guang-Zhong Yang[1]

The Hamlyn Centre for Robotic Surgery
Department of Surgery and Cancer
Imperial College London, UK
{menglong.ye11}@imperial.ac.uk

Abstract. Recent advances in microscopic detection techniques include fluorescence spectroscopy, fibred confocal microscopy and optical coherence tomography. These methods can be integrated with miniaturised probes to assist endoscopy, thus enabling diseases to be detected at an early and pre-invasive stage, forgoing the need for histopathological samples and off-line analysis. Since optical-based biopsy does not leave visible marks after sampling, it is important to track the biopsy sites to enable accurate retargeting and subsequent serial examination. In this paper, a novel approach is proposed for pathological site retargeting in gastroscopic examinations. The proposed method is based on affine deformation modelling with geometrical association combined with cascaded online learning and tracking. It provides online *in vivo* retargeting, and is able to track pathological sites in the presence of tissue deformation. It is also robust to partial occlusions and can be applied to a range of imaging probes including confocal laser endomicroscopy.

1 Introduction

Gastroscopy is the gold-standard method of visualising the upper gastrointestinal tract. In addition to direct visualisation of the mucosa, the procedure enables the procurement of histological samples important for the diagnosis and assessment of gastrointestinal pathologies. Recently, there has been an increasing demand for *in vivo*, *in situ* real-time tissue characterisation and assessment. Techniques that enable microscopic detection and characterisation of tissue include fluorescence spectroscopy, fibred confocal microscopy, and optical coherence tomography. These methods can be packaged into miniaturised probes that can be deployed through endoscopic instrument channels. They allow changes that occur in diseases such as cancer to be detected at an early and pre-invasive stage (without taking histopathological tissue samples). One of the major problems with the use of probe-based systems is its small field-of-view, hence systematically screening larger surface areas with surface deformation, such as the oesophagus becomes difficult. Retargeting previously probed mucosal sites is also

K. Mori et al. (Eds.): MICCAI 2013, Part II, LNCS 8150, pp. 67–74, 2013.

challenging due to the lack of visible marks left on the mucosal surface. Recent adoption of Narrow Band Imaging (NBI) enhances the visibility of mucosal features, allowing the possibility of using computer vision approaches to assist pathological site retargeting in examinations of the upper gastrointestinal tract.

For addressing the issue of accurate optical biopsy site retargeting, a number of approaches have been proposed. These include the use of Markov Random Fields (MRF) [1], Simultaneous Localisation and Mapping (SLAM) [2], and a multi-view based approach using epipolar geometry [3]. To permit simultaneous retargeting and localisation during an entire examination procedure, endoscopic image classification in the manifold space has been used for surveillance endoscopy [4]. Key issues highlighted in the above studies include multiple biopsy site tracking under tissue deformation, tool-tissue interaction and partial occlusion. The purpose of this paper is to propose a new approach for pathological site retargeting during endoscopic examinations. In contrast to the previous methods, the proposed method enables online *in vivo* retargeting based on a learning-based tracker. A local geometrical association scheme is proposed to mitigate the problem of partial occlusion and feature drop-out during tracking, which also facilitates simultaneous tracking of multiple biopsy sites. The method provides reliable retargeting of pathological sites in the presence of tissue deformation where existing methods such as SLAM fail. Detailed experimental validation with both phantom and *in vivo* endoscopic data has been performed and the results derived demonstrate its potential clinical value.

2 Methods

2.1 Local Affine Tissue Deformation Modelling

Tissue deformation causes general tracking methods in computer vision to fail. In this paper, we assume regional tissue deformation can be modelled by local affine transformations, as similarly adopted in [3]. With this assumption, the appearance of a local surface patch in two different views can be linked with an affine transformation.

We denote the camera's intrinsic matrix as K and extrinsic matrix as $[R|T]$ between two different views. For an *in vivo* environment with tissue deformation, the changes in the 3D position of the surface patch can also be incorporated into $[R|T]$. As shown in Fig. 1 a), an arbitrary 3D point in the surface patch is denoted as P. The goal is to show that there is a homography H that satisfies $p_2 = Hp_1$, where p_1 and p_2 are the 2D reprojection points of P in View 1 and View 2, respectively. The homography can be obtained by projecting p_1 and p_2 to the same 3D point P, such that

$$H = \left[\frac{f_1}{f_2}KRK^{-1}|\frac{1}{f_2}KT\right], \tag{1}$$

where f_1 and f_2 are the normalising factors for the two views. Since we consider small regions, H can be approximated using a local affine transformation.

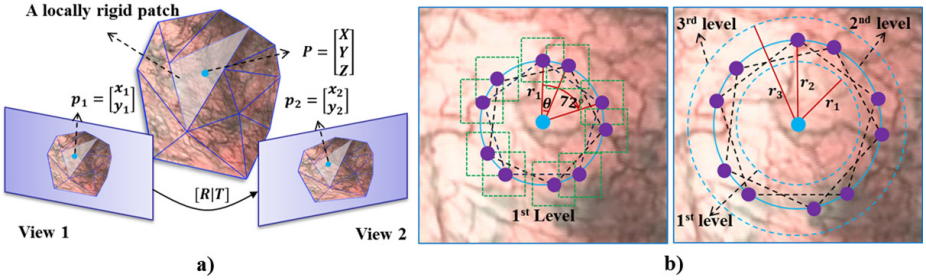

Fig. 1. a) 3D-2D reprojection. An arbitrary point P in a locally rigid patch is reprojected back onto View 1 and View 2, represented by p_1 and p_2, respectively; b) Initialisation of geometrically associated sites. Start by creating associated sites with size r_1, three levels of association are initialised; the green dash boxes represent the patches independently tracked by Tracking-Learning-Detection (TLD).

In our work, an optical biopsy site is retargeted using the affine transformation obtained from its tracked adjacent regions. This transformation is used to estimate the position of the pathological site in the current image. This is important as the target site does not always have identifiable surface features and the use of multiple surrounding landmarks would ensure both reliability and consistency under camera motion, tissue deformation and partial occlusion.

2.2 Tracking with Cascaded Online Learning

Tracking objects with online learning is a popular topic in computer vision. Methods [5,6] have been proposed to combine learning and tracking in an online framework. The tracker adopted in our work is the Tracking-Learning-Detection (TLD) method [7]. TLD consists of a median-flow tracker, a cascaded classifier and scanning-windows detection. When the tracker is able to provide the object location, the cascaded classifier, which contains a variance filter, a randomised fern filter and Nearest Neighbour (NN) classifier, is trained with the samples generated from detection and tracking. When the tracker fails in providing the object location, the patches generated from detection are regarded as potential candidates. These candidates are then filtered by the cascaded classifier that has been trained. The final remaining patch is identified as the object, and this reboots the tracking.

In this work, due to the requirement of tracking multiple surrounding regions, TLD is extended to independently track multiple surface patches. Therefore, we define the model map that contains the models of N surrounding regions as $\{M_1, M_2, \ldots, M_N\}$, where $M_i = \{M_i^+, M_i^-\}$. $M_i^+ = \{s_1^+, s_2^+, \ldots, s_{l_i}^+\}$ and $M_i^- = \{s_1^-, s_2^-, \ldots, s_{k_i}^-\}$ represent the collections of l_i positive and k_i negative samples of region i, respectively. Similar to [7], the initialisation of M_i is achieved by generating affine-transformed positive samples, which enables rotation invariance, and high-variance negative samples.

2.3 Geometrical Association for Retargeting

Using the local affine transformation and the multiple-object TLD mentioned earlier, a retargeting framework based on local geometrical association has been proposed. A pathological site is retargeted using the affine transformation obtained from its surrounding regions, which are called associated sites.

Least-squares [8] is applied to calculate the transformation using at least four pairs of corresponding associated sites. It is essential for least-squares that the positions of these associated site should be non-colinearly distributed during initialisation. To ensure enough non-colinear associated sites can be tracked for affine transformation estimation, the following steps are used for initialisation. Firstly, a regular pentagon with size r_1 (Fig. 1 b)) is created and centred at the 2D position of the pathological site, p_{init}, which can be obtained from the operator's input or probe detection during a probe-based confocal laser endomicroscopy (pCLE) procedure. The corner points of the pentagon are defined as the centres of the associated sites independently tracked by TLD. The pentagon is then rotated with regard to the centre by angles $\{\theta, 2\theta, \ldots, (n-1)\theta\}$ until $n\theta = 72^\circ$ (note the pentagon is rotationally symmetric), where n is the number of pentagons. For each angle setting, five associated sites are created for tracking. We define these as the 1^{st} level for the associated sites. To take into account changes in scale, pentagons of multiple sizes are used to create associated sites. In a similar procedure as above, pentagons of size r_2 and r_3 ($r_3 > r_2 > r_1$) are used to create the 2^{nd} and 3^{rd} level of the associated sites, respectively.

In this paper, three levels of pentagons $\{O_1, O_2, \ldots, O_{3n}\}$ are initialised. Once there are k ($k \geq 4$) associated sites of one pentagon O_i successfully identified in the current image, an affine transformation can be obtained by minimising:

$$e\left(H_i\right) = \frac{1}{k} \sum_{j=1}^{k} \left(q_j - H_i p_j\right)^2, \tag{2}$$

with p_j and q_j being one pair of the corresponding associated sites between two images. It should be noted that the 2D positions of the sites are defined as the centres of the patches tracked by TLD. The transformation H_i estimated by Eq. 2 is then applied to calculate a candidate position of the site in the current image with $p'_i = H_i p_{init}$. A collection of candidate sites $\{p'_1, p'_2, \ldots, p'_m\}$, where $m \leq 3n$, is obtained by searching for all pentagons that provide affine transformations. The final coordinates of the site in the current image are chosen as the median values of all the candidate sites, which provides robustness against noise.

3 Results and Validation

3.1 Parameter Configuration

Prior to the use of the proposed method based on multi-object TLD, the parameters mentioned above related to geometrical association need to be determined. To this end, a detailed sensitivity and specificity analysis has been performed

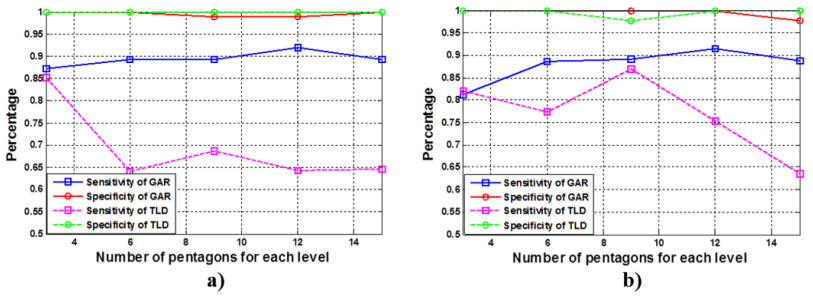

Fig. 2. Sensitivity and specificity test of the proposed Geometrically Associated Retargeting (GAR) and TLD. a) Results for site 1 with five different pentagon settings; b) Results for site 2 with five different pentagon settings. True Positives: correctly identified sites that are pathological; True Negatives: correctly rejected sites that are not pathological; False Positives: incorrectly identified sites that are not pathological; False Negatives: incorrectly rejected sites that are pathological.

with known ground truth. The patch size ($width \times height$) of the associated sites tracked by TLD is typically 60×60. The sizes of the pentagons in different levels are then defined as $r_1 = \lambda width$, $r_2 = (\lambda + 0.1)\, width$ and $r_3 = (\lambda + 0.2)\, width$, where $\lambda = 0.2$ is the control element that is determined empirically to constrain the total area of the associated sites to be a locally rigid patch (for the approximation of deformation). We perform our method on phantom data with $n =$ 3, 6, 9, 12, and 15 pentagons for each level to retarget two different sites. The sensitivity and specificity measures are summarised in Fig. 2 a) and b). TLD has also been performed on the same two sites. It should be noted that our proposed method can achieve an improved performance than the original TLD. It is evident that the setting of 12 pentagons gives the best performance, and therefore is used for all the experiments conducted in this paper.

3.2 Validation

Deforming Phantom. The proposed method is tested on an oesophageal phantom that simulates the textural 'mucosal' features enhanced with NBI as well as irregular deformation (see supplementary videos[1]). Fig. 3 a) shows the retargeting of multiple pathological sites. Three video sequences of a deforming phantom have been processed and three sites have been selected for validation. As the phantom surface is deforming, ground truth (reprojections of the sites in images) can only be generated with an expert's observations. By comparing the results with the ground truth, visual angle errors adopted in [2] have been estimated using the camera intrinsic parameters. The average visual angle errors are presented in Table 1 (Site No. $1-3$) and the sensitivity (recall) and specificity values range from 0.9231 to 0.9722 and 0.9310 to 1, respectively. The precision values range from 0.9459 to 1.

[1] `http://www.imperial.ac.uk/hamlyn/surgicalvision`

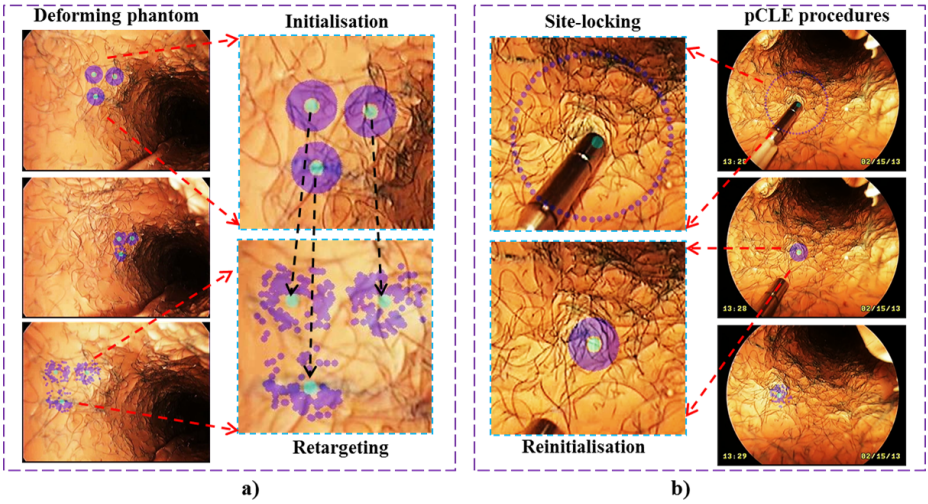

Fig. 3. a) Retargeting on deforming phantom with multiple site tracking; b) Retargeting in simulated pCLE procedures with site-locking and reinitialisation. Purple dots represent the locations of associated sites; green dots represent pathological sites.

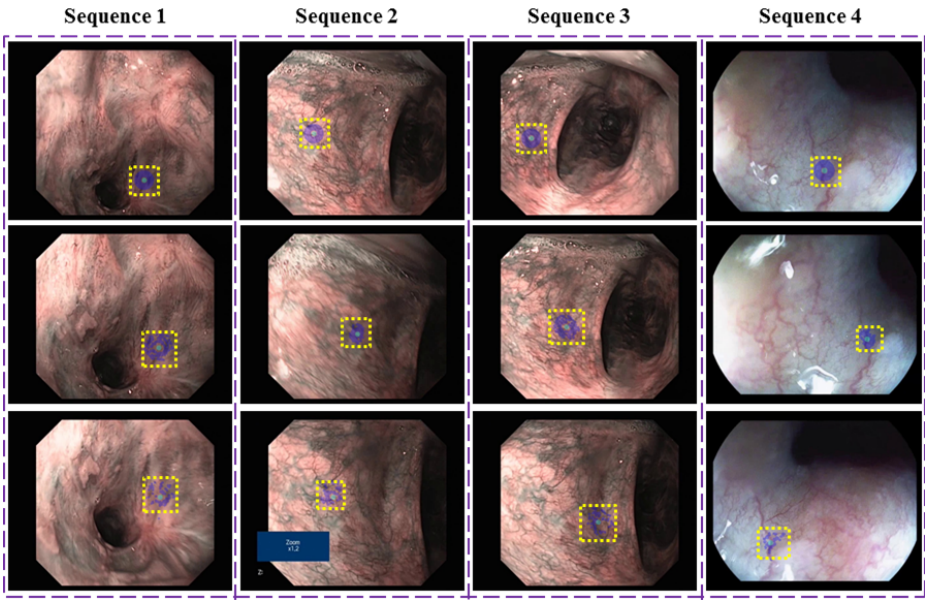

Fig. 4. The proposed retargeting method performed on *in vivo* data, showing four *in vivo* video sequences acquired using Olympus NBI and PENTAX i-Scan endoscopies.

Table 1. Quantitative evaluation results of the proposed retargeting approach on ten pathological sites, being represented as Average Visual Angle Error (AVAE) [2], AVAE in X-dimension (AVAEX), and AVAE in Y-dimension (AVAEY). The numbers in the brackets represent the percentage of field-of-view.

	Deforming phantom			**PCLE procedure**			***In vivo* data**			
Site No.	1	2	3	4	5	6	7	8	9	10
AVAE	0.51° (0.54%)	0.64° (0.67%)	0.35° (0.37%)	2.49° (2.61%)	2.79° (2.93%)	2.62° (2.75%)	2.71° (1.51%)	0.79° (0.44%)	1.68° (0.94%)	0.87° (0.48%)
AVAEX	0.36° (0.48%)	0.27° (0.37%)	0.24° (0.32%)	1.73° (2.32%)	1.69° (2.27%)	1.68° (2.25%)	1.32° (0.94%)	0.39° (0.28%)	0.95° (0.67%)	0.63° (0.45%)
AVAEY	0.27° (0.46%)	0.56° (0.94%)	0.20° (0.33%)	1.29° (2.19%)	1.98° (3.35%)	1.74° (2.94%)	2.24° (2.00%)	0.64° (0.57%)	1.28° (1.14%)	0.51° (0.44%)

Simulation with pCLE Retargeting. The proposed method has also been validated in simulated pCLE procedures in a phantom environment. Partial instrument occlusions exist in the procedures. To deal with this, we perform an additional step that locks the site where the probe has touched. This site-locking step is achieved using the proposed geometrical association scheme (see Section 2.3) but only with one level of pentagons of size $2width$. After the probe moves away from the site, the locked site is reinitialised with the aforementioned multi-scale pentagons (Fig. 3 b)). Blob detection is performed to identify the movement of the probe. For validation, electromagnetic sensors (Aurora, NDI) are attached to the camera and the probe (Mauna Kea Technologies) so that the 3D position of the site can be obtained. Given that the phantom is static, the site position in the 3D space is reprojected onto the image using the world-to-sensor transformation from Aurora and sensor-to-camera transformation from hand-eye calibration [9]. The reprojections of the site in the images are defined as the ground truth, which are then compared with the results from the proposed approach. Quantitative results are presented in Table 1 (Site No. $4-6$).

***In Vivo* Data Validation.** In addition to phantom validation, *in vivo* gastroscopic data is also processed. The validation is performed by comparing the ground truth from expert observations and the estimated site locations from our approach. Four *in vivo* sequences (three sequences with Olympus NBI, one sequence with PENTAX i-Scan endoscopy) are presented in Fig. 4. Our method is able to perform retargeting in the presence of tissue deformation. The average visual angle errors provided in Table 1 (Site No. $7-10$) range from 0.87° to 2.71°. The precision values range from 0.5833 to 1. The sensitivity (recall) and specificity values range from 0.3529 to 0.8824 and 0.6000 to 1, respectively. The results show lower sensitivity and specificity measures in *in vivo* data than that of the phantom experiment. This is because fast *in vivo* endoscopic motion results in blurred images, causing poor textural information to be gleaned for

retargeting. In addition, dark illuminating conditions on the pathological site also leads to less reliable textural information. Nevertheless, our proposed method provides consistently low visual angle errors for both phantom and *in vivo* data.

4 Conclusion

In this paper, we have proposed an online deformable approach for pathological site retargeting in endoscopic examinations. The proposed approach performs online learning-based tracking to re-identify optical biopsy sites. By modelling tissue deformation as local affine transformations, our approach is able to handle realistic motion as encountered in typical *in vivo* endoscopic examinations. The proposed geometrical association scheme provides the flexibility of being able to be applied to general probe-based microscopic detection techniques such as pCLE. The method has been validated using deforming phantom data, simulated pCLE procedures, as well as *in vivo* gastroscopic data. It has been demonstrated that the method can accurately retarget pathological sites online in deformable environments, thus enabling its practical use in endoscopy.

Acknowledgements. The authors would like to thank Dr. Selen Atasoy, Dr. Jonathan Hoare and Prof. Alexander Meining for data sharing and collection.

References

1. Atasoy, S., Glocker, B., Giannarou, S., Mateus, D., Meining, A., Yang, G.-Z., Navab, N.: Probabilistic region matching in narrow-band endoscopy for targeted optical biopsy. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 499–506. Springer, Heidelberg (2009)
2. Mountney, P., Giannarou, S., Elson, D., Yang, G.-Z.: Optical biopsy mapping for minimally invasive cancer screening. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 483–490. Springer, Heidelberg (2009)
3. Allain, B., Hu, M., Lovat, L.B., Cook, R.J., Vercauteren, T., Ourselin, S., Hawkes, D.J.: Re-localisation of a biopsy site in endoscopic images and characterisation of its uncertainty. Med. Image Anal. 16(2), 482–496 (2012)
4. Atasoy, S., Mateus, D., Meining, A., Yang, G., Navab, N.: Endoscopic video manifolds for targeted optical biopsy. IEEE Trans. Med. Imag. 31(3), 637–653 (2012)
5. Mountney, P., Yang, G.-Z.: Soft tissue tracking for minimally invasive surgery: Learning local deformation online. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 364–372. Springer, Heidelberg (2008)
6. Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell. 33(8), 1619–1632 (2011)
7. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. 34(7), 1409–1422 (2012)
8. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. 13(4), 376–380 (1991)
9. Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Trans. Robot. Autom. 5(3), 345–358 (1989)