# Hybrid Retargeting for High-Speed Targeted Optical Biopsies

André Mouton[1], Menglong Ye[1], François Lacombe[2], and Guang-Zhong Yang[1]

[1] The Hamlyn Centre for Robotic Surgery, Imperial College London, UK
[2] Mauna Kea Technologies, Paris, France
{a.mouton,menglong.ye11}@imperial.ac.uk

**Abstract.** With the increasing maturity of optical biopsy techniques, routine clinical use has become more widespread. This wider adoption of the technique demands effective tracking and retargeting of the biopsy sites, as no visible markers are left following examination. This study presents a high-speed framework for intra-procedural retargeting of probe-based optical biopsies in gastrointestinal endoscopy. A probe tip localisation method using active shape models and geometric heuristics, which eliminates the traditional dependency on shaft visibility, is proposed for automated initialisation. Partial occlusion and tissue deformation are addressed by exploiting the benefits of indirect and direct tracking through a novel combination of geometric association and online learning. Robustness to rapid endoscope motion and improvements in computational efficiency are achieved by restricting processing to the automatically detected video content area and through a feature-based rejection of non-informative frames. Performance evaluation in phantom and *in-vivo* environments demonstrates accurate biopsy site initialisation, robust retargeting and significant improvements over the state-of-the-art in processing time and memory usage.

## 1 Introduction

Gastrointestinal (GI) endoscopy together with histopathological tissue examination is the gold-standard for the diagnoses of pathologies in the digestive tract. Endoscopy-guided probe-based optical biopsies, however, allow for *in-vivo* visualisation of tissues at a cellular level, forgoing the need for tissue excision. Macroscopic retargeting of previously examined sites in this context is challenging due to the absence of physical scars at the biopsy sites [1]. Traditional retargeting, performed by tattooing the examined tissue with ink or Argon Plasma Coagulation (APC), is limited as the ink tends to diffuse and APC causes tissue damage. This study proposes a non-invasive alternative for automated intra-procedural retargeting using purely vision-based techniques, seeking high-speed performance and robustness to rapid endoscope movements, natural tissue deformation and partial site occlusion.

Mountney *et al.* [6] perform intra-procedural retargeting using a visual SLAM-based approach (assuming large-scale rigidity) to generate a 3D model of the
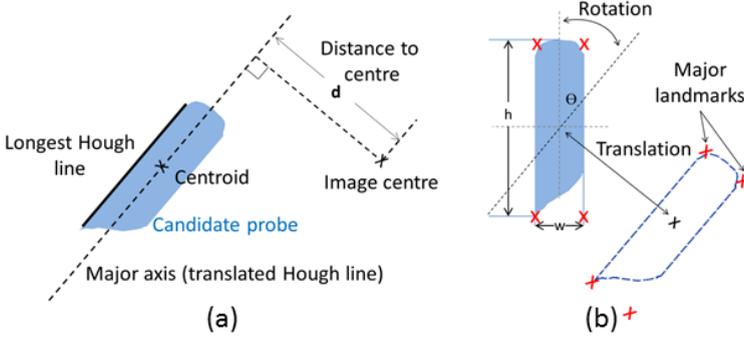
**Fig. 1.** Tip localisation: (a) Orientation constraint. (b) Major landmarks computation.

tissue surface. Initialisation is automated by a probe detection technique which exploits the achromatic properties of the probe shaft. Atasoy *et al.* [2] address tissue deformation using a geometrically-constrained MRF model, while Allain *et al.* [1] propose retargeting based on epipolar lines, derived from multiple views of a given biopsy site. Ye *et al.* [9] model tissue deformation as locally affine and perform indirect retargeting by geometric association and online-learning. Partial occlusion is addressed by reinitialising the associated sites when the optical probe exits the field of view (determined by simple blob detection).

This work expands upon these prior studies, each of which addresses only part of the retargeting problem, through the development of a novel, end-to-end framework for high-speed, automated and robust retargeting. The principal contributions include: real-time detection of non-informative frames; geometry-independent content area detection; shape-based probe tip localisation for automated initialisation and hybrid retargeting for improved performance under partial occlusion and tissue deformation.

## 2   Methods

An automated, high-speed solution for targeted optical biopsies in GI endoscopy, comprised of the following five components, is proposed: **1)** non-informative frame rejection; **2)** video content area detection; **3)** probe tip localisation; **4)** retargeting under partial occlusion and **5)** retargeting without occlusion.

**Frame Rejection:** Redundant processing is mitigated by discarding clinically non-informative frames using a novel quantification of image saliency based on the number of FAST keypoints detected [8] (Fig. 2). In particular, frames are rejected by placing a lower bound, $\tau_\sigma$, on the number of detected keypoints. High frame rates in endoscopic sequences ($\geq$ 25fps) result in significant information overlap between a certain number of consecutive frames, making the selection of $\tau_\sigma$ non-critical as information lost in incorrectly discarded frames is likely to be recaptured in subsequent frames.
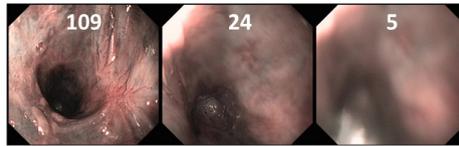
**Fig. 2.** Number of FAST keypoints detected.

**Content Detection:** Subsequent processing is restricted to the automatically-detected central content area (Fig. 3). While Munzer *et al.* [7] proposed a method to automatically detect circular content regions, here a technique for annotating the content area of all endoscopic sequences, regardless of the FoV geometry, is presented. Binary thresholding is applied to a greyscale input image to eliminate low-intensity noise in the background regions. The largest connected component in the binary image is then found and refined via morphological opening, which reduces unwanted spurious details.

**Probe Tip Localisation:** The initial biopsy site is assumed to be located at the tip of the optical probe and is found via probe detection and tip localisation. A limitation of prior work in vision-based endoscopic instrument detection is the reliance on the visibility of the achromatic instrument shaft [6]. This limitation is addressed via a markerless probe detection approach that combines colour-based segmentation with active shape modelling and geometric heuristics. The technique focusses directly on the detection of the probe tip, ignoring the instrument shaft and thereby compensating for scenarios where the shaft is not visible. Candidate segmentations are obtained by thresholding and connected component analysis in a *texture-reduced* image, obtained by greyscale morphological closing. Closing attenuates darker features in the image, reducing the characteristic mucosal and vascular patterns (i.e. reducing the texture) without impacting the appearance of the comparatively homogeneous probe tip and thereby simplifies the segmentation task. The candidate set is refined by applying two geometric constraints based on *a priori* knowledge of probe size and orientation. All components falling outside of a predefined size range are eliminated, while the orientation constraint is formulated based on the knowledge that the optical probe is inserted through the operating channel of the endoscope and is thus radially orientated. The constraint is enforced by placing a threshold on the distance between the extended major axis of the component and the image centre. The major axis is found by computing the longest straight line segment in the component using the Hough transform. This line (which is expected to represent one of the two longitudinal edges of the probe) is translated to pass through the centroid of the component, yielding the major axis (Fig. 1). The final coarse segmentation is that component with the longest straight line segment satisfying these constraints.

The texture-reduced image is searched using an Active Shape Model (ASM) [3] of the probe tip, the accuracy of which is optimised by initialising the search with the coarse segmentation. This is achieved by defining four *major* landmarks

in the ASM located at the end-points of the longitudinal edges of the probe. The corresponding points in the coarse segmentation are estimated by centring the probe at the origin (by translation) and rotating it according to the orientation of its longest line segment (Fig. 1). The height and width of the probe are estimated as the maximum differences in the $(x, y)$ components of any two points in the probe. The landmarks are then approximated as the corresponding offsets from the probe centroid, translated and rotated back to the initial pose. A linear transformation that minimises the Procrustes distance between the major landmarks in the mean and initial shapes is used to align the ASM with the coarse segmentation. This approach significantly improves segmentation accuracy, compared to the traditional approach of initialising the search with the mean shape and a fixed pose.

**Pathological Site Retargeting (PSR):** A hybrid retargeting approach is proposed, whereby the underlying methodology is based on the degree of partial site occlusion. A site is considered occluded if the overlap $O$ between its bounding box (defined by the set of pixels $A$) and the optical probe (defined by the set $B$) exceeds a predefined threshold where $O = \frac{|A \cap B|}{|A|}$.

**PSR with Partial Occlusion** is performed by geometric association (Fig. 3). Similarly to [9], the associated sites are defined as the vertices of a set of $n$ concentric, fixed radius pentagons. The set is constructed by incremental rotation of an initial pentagon by $\{0, \theta, 2\theta, \ldots, (n-1)\theta\}$ such that $n\theta = 72$. Based on the assumption that tissue deformation is locally affine [9], the biopsy site location is computed using the perspective transformation that minimises the reprojection error between the associated sites in the previous and current frames. The critical difference to [9], where each associated site is retargeted independently using TLD [5], is that the sites are tracked directly using Lucas-Kanade optical flow, which significantly reduces computational overhead. This approach is motivated by the fact that initial tracking of the site, while the probe is visible, does not generally involve endoscope motion and is thus unlikely to require site re-detection. This makes a short-term tracking solution more suitable in terms of computational overhead.

**PSR without Partial Occlusion:** The retargeting methodology switches to TLD when the overlap between the optical probe and the local neighbourhood of the biopsy site falls below a threshold (Fig. 3). TLD is an established long-term tracking method that integrates a median flow tracker, a cascaded classifier and online learning. In [9] tissue deformation is addressed by modelling the local neighbourhood with 180 separate bounding boxes. Here it is proposed that the same local information is captured by initialising TLD with a single, suitably-sized bounding box centred on the biopsy site. Since TLD is robust to affine transformations, the tracking of this bounding box is not affected by affine tissue deformation. This direct retargeting approach offers a significant reduction in computational overhead (1 site vs. 180 sites). In order to avoid corrupting the positive training samples of the cascaded classifier (generated online), the

**Table 1.** Probe tip localisation results.

| Method | P | R | FPR | ALE (pixels) | FPS |
|---|---|---|---|---|---|
| TipLoc | 0.91 | 0.89 | 0.05 | $3.84 \pm 1.80$ | 10 |
| Mount [6] | 0.09 | 0.05 | 0.41 | $43.26 \pm 49.30$ | 25 |

retargeting methodology reverts back to geometric association at any stage that the probe re-enters the local neighbourhood of the site.

## 3   Results

Performance was evaluated using both phantom and *in-vivo* gastrointestinal data, captured with an Olympus Narrow Band Imaging (NBI) endoscope. Two phantoms (modelling the textural mucosal and vascular characteristics of the human oesophagus) were used. Ground-truth annotations were performed by an experienced observer. Since probe-based optical biopsies are not currently practised in GI endoscopy in the UK, the procedure was simulated in the phantom environment using a Cellvizio optical probe manufactured by Mauna Kea Technologies. Tissue deformation was simulated by manually applying an external force to the phantom. All components of the proposed methodology related to probe tip localisation were evaluated in this environment.

**Probe Tip Localisation:** The ASM model was built using 50 training images and the number of search iterations was restricted to 10, such that overall processing times were not significantly compromised. Accuracy was measured as the Average Localisation Error (ALE) in pixels between the computed and ground-truth probe tip positions. Performance was quantified further according to the recall rate (R), False Positive Rate (FPR), and precision (P). Localisation errors of less than 10 pixels were considered true positives. Performance of the proposed method (denoted *TipLoc*) was compared to [6] (denoted *Mount*) - chosen as its methodology is representative of the majority of the related vision-based methods that rely on the visibility of the achromatic instrument shaft. With reference to Tab. 1, the proposed technique significantly outperformed [6] in terms of both tip localisation and detection accuracy. The poor performance of [6] is attributed to the failed initial segmentations and not to inaccurate tip localisations, as the current dataset consists predominantly of images where the shaft is not visible. Despite the lower average frame rate, Tab. 1 highlights the advantages of the proposed technique over the popular shaft-based segmentation approaches.

**Retargeting:** Performance was evaluated using three phantom and three *in-vivo* sequences and quantified using precision, recall and the ALE between the centres of the ground-truth and predicted bounding boxes. In particular, an object-based evaluation (c.f. pixel-based) was considered, whereby a true positive instance is defined as an overlap $\sigma_i = \frac{B_i^G \cap B_i^T}{B_i^G \cup B_i^T}$ of greater than a threshold $\tau_\sigma$ between the ground-truth and predicted bounding boxes ($B^G, B^T$ respectively).
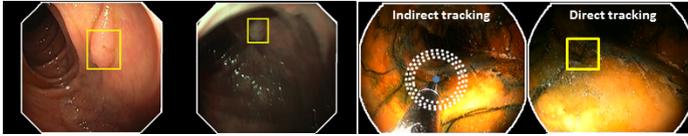
**Fig. 3.** Sample *in-vivo* (left) and phantom (right) PSR results. Content areas in white.

It is worth noting that ALE is not scale-invariant [ **??** ] and that it is thus not generally possible to perform ROC analysis using object-based metrics (as the false-positive rate cannot be computed) [4]. Therefore, in accordance with [4,5, **??** ], the scalar-valued $F1$-measure ($F1 = 2\left(\frac{P \cdot R}{P+R}\right)$) was computed in addition to P, R and ALE, to provide an objective overall performance measure. The *in-vivo* performance of the proposed methodology (denoted *PSR*) was compared to the baseline intra-procedural approach of Ye *et al.* [9] (denoted *PSR_Ye*). A direct comparison was not possible in the phantom environment as PSR_Ye does not consider probe detection, thus requiring manual initialisation. Optimal parameter values were determined empirically, based on performance in a single sequence (not included in the results). The chosen parameter set was used for all tests (*in-vivo* and phantom).

PSR performed well in the phantom environment (Tab. 2), with perfect precision and a worst-case F1 of 0.91. Retargeting localisation accuracy was high for all three sequences (9.75-13.72 pixels). Performance is comparable to the state-of-the-art [9] (achieved in a similar phantom environment). PSR maintained high precision (0.91-0.96), recall (0.80-0.94) and F1 (0.85-0.94) scores *in-vivo* (Tab. 2). In contrast, PSR_Ye performed poorly, particularly with regard to recall (0.12-0.44), resulting in considerably lower F1 scores (0.23-0.61). In all three cases, PSR_Ye lost track of the biopsy site before it had exited the FoV and was not able to re-detect it in subsequent frames. As noted in [9], a decrease in textural saliency and an increase in the number of out-of-focus frames relative to the phantom data (due to rapid endoscope motion) negatively affects the feature-based classification in TLD. PSR_Ye, which relies on multi-site TLD and does not discard non-informative frames, is thus affected more significantly by these factors than PSR. The relative decline in performance of PSR *in-vivo* may be further attributed to the absence of partial probe occlusion, which biases overall phantom performance due to the relative ease of short-term tracking in partially occluded frames. The *in-vivo* and phantom results suggest that direct tracking is favourable to geometric association in non-occluded scenarios.

Content detection used in conjunction with frame rejection was found to yield optimal performance, when considering F1 scores in conjunction with processing times (Tab. 3). Although a negligible drop in recall was caused by the inevitable rejection of a small number of informative frames, frame-rejection crucially improved precision by reducing the number of false-positive containing frames and significantly decreased processing times by eliminating redundant sliding window searches.

**Computational Performance:** The improvements in computational efficiency of PSR relative to PSR_Ye were evaluated according to CPU usage, memory

**Table 2.** Phantom and *in-vivo* performance.

| Name | Method | P | R | F1 | ALE (pixels) |
|---|---|---|---|---|---|
| **Phantom 1** | PSR | 1.00 | 0.98 | 0.99 | 13.72 |
| **Phantom 2** | PSR | 1.00 | 0.96 | 0.98 | 13.06 |
| **Phantom 3** | PSR | 1.00 | 0.91 | 0.95 | 9.75 |
| *In-vivo* **1** | PSR | 0.96 | 0.81 | 0.88 | 10.94 |
| | PSR_Ye | 1.00 | 0.27 | 0.43 | 4.60 |
| *In-vivo* **2** | PSR | 0.94 | 0.94 | 0.94 | 15.89 |
| | PSR_Ye | 1.00 | 0.44 | 0.61 | 3.40 |
| *In-vivo* **3** | PSR | 0.91 | 0.80 | 0.85 | 12.24 |
| | PSR_Ye | 1.00 | 0.13 | 0.23 | 14.19 |

**Table 3.** Impact of content detection (CD) and frame rejection (FR). Optimal results in bold.

| Name | CD | FR (discarded) | P | R | F1 | Time (s) |
|---|---|---|---|---|---|---|
| | no | no | 0.79 | 0.37 | 0.50 | 131.87 |
| *In-vivo* **1** | yes | no | 0.94 | 0.83 | 0.88 | 120.73 |
| | no | yes (275) | 0.77 | 0.87 | 0.82 | 101.00 |
| | **yes** | **yes (325)** | **0.96** | **0.81** | **0.88** | **91.12** |
| | no | no | 0.94 | 0.67 | 0.78 | 46.73 |
| *In-vivo* **3** | yes | no | 0.87 | 0.80 | 0.83 | 41.18 |
| | no | yes (6) | 0.63 | 0.69 | 0.66 | 41.47 |
| | **yes** | **yes (15)** | **0.91** | **0.80** | **0.85** | **39.00** |

usage and processing time for single-site retargeting in a 419-frame *in-vivo* sequence. Additionally, the feasibility of multi-site PSR was evaluated by measuring computational performance for one, two, three and five sites. Multi-site PSR was implemented by brute-force, whereby each site was retargeted independently (resource sharing and/or parallelisation were not considered). Processing was performed on an Intel® Core™ i7, 3.4GHz quad-core CPU with 16GB of RAM. PSR outperformed PSR_Ye in every aspect of computational performance for single-site retargeting with reductions of approximately 50% in mean CPU usage and 90% in mean memory usage (Fig. 4a-c). Most significantly (with regard to clinical feasibility), PSR gave an improvement of two orders of magnitude in frame rate over PSR_Ye (10-30fps vs. $\approx$ 0.5fps). These are significant improvements in computational efficiency and highlight the benefits of the proposed hybrid retargeting approach. Fig. 4d shows the results of regression analyses for each metric (averaged over the entire sequence) for multi-site PSR. Memory usage and CPU usage exhibited quadratic and linear relationships with the number of retargeted sites respectively, while frame rate decreased exponentially. While this is not encouraging for clinical feasibility (where approximately 10-20 sites would need to be retargeted), it is proposed that parallelisation and resource sharing would improve performance significantly. This claim is substantiated by the results for single-site PSR_Ye (Fig. 4a-c). PSR_Ye, which has been parallelised, simultaneously retargets 180 associated sites for a single biopsy site, yet operates with manageable computational overhead and maintains a frame rate of approximately 0.5fps.
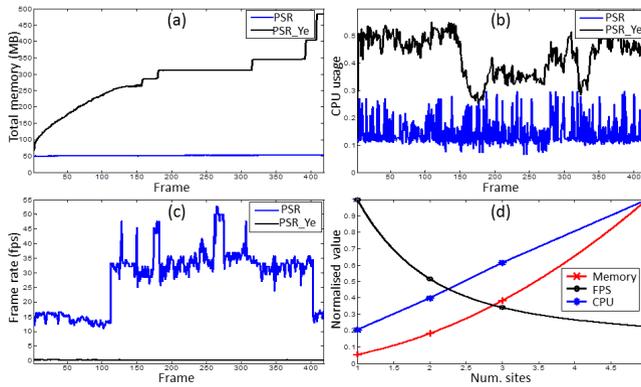
**Fig. 4.** Computational performance: (a) Single-site memory usage. (b) Single-site CPU usage. (c) Single-site frame rates. (d) Multi-site regression analyses.

## 4    Conclusion

This study has presented a high-speed, automated solution for targeted optical biopsies in GI endoscopy. Initialisation is performed via a novel probe tip local-isation method using active shape modelling and geometric heuristics. Partial occlusion and tissue deformation have been addressed by adopting a hybrid re-targeting approach which combines indirect tracking by geometric association and direct tracking by online learning. Additionally, a method for the detection of non-informative frames using the FAST feature detector has been presented, improving robustness to rapid endoscope motion and facilitating high-speed im-plementation. Computational efficiency has been optimised further by restrict-ing processing to the automatically-detected video content area. State-of-the-art probe tip localisation accuracy has been presented with an average error of 3.84 pixels and consistently high recall (0.89) and precision (0.91) rates. Crucially, the traditional dependency on the visibility of the probe shaft has been elimi-nated. Retargeting is achieved with consistently high F1 scores (0.85-0.99) for both phantom and *in-vivo* data. Together with an improvement of two orders of magnitude in processing time, this corresponds to a significant improvement over the state-of-the-art [9]. Future work will seek to further improve computational efficiency to facilitate the implementation of multi-site retargeting.

## References

1. Allain, B., Hu, M., Lovat, L.B., Cook, R.J., Vercauteren, T., Ourselin, S., Hawkes, D.J.: A system for biopsy site re-targeting with uncertainty in gastroenterology and oropharyngeal examinations. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 514–521. Springer, Heidelberg (2010)

2. 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)
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. CVIU 61(1), 38–59 (1995)
4. Hua, Y., Alahari, K., Schmid, C.: Occlusion and motion reasoning for long-term tracking. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VI. LNCS, vol. 8694, pp. 172–187. Springer, Heidelberg (2014)
5. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE TPAMI 34(7), 1409–1422 (2012)
6. 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)
7. Munzer, B., Schoeffmann, K., Boszormenyi, L.: Detection of circular content area in endoscopic videos. In: IEEE Inter. Symp. on Computer-Based Medical Systems, pp. 534–536 (2013)
8. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
9. Ye, M., Giannarou, S., Patel, N., Teare, J., Yang, G.-Z.: Pathological site retargeting under tissue deformation using geometrical association and tracking. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 67–74. Springer, Heidelberg (2013)