Wide-Baseline Dense Feature Matching
for Endoscopic Images

Gustavo A. Puerto-Souza and Gian-Luca Mariottini

Dept. of Computer Science and Engineering,
Univ. of Texas at Arlington, 500 UTA Blvd., 76019 Texas, USA
gustavo.puerto@mavs.uta.edu
gianluca.mariottini@uta.edu

Abstract. Providing a feature-matching strategy to accurately recover tracked features after a fast and large endoscopic-camera motion or a strong organ deformation, is key in many endoscopic-imaging applications, such as augmented reality or soft-tissue shape recovery. Despite recent advances, existing feature-matching algorithms are characterized by limiting assumptions, and have not yet met the necessary levels of accuracy, especially when used to recover features in distorted or poorly-textured tissue areas. In this paper, we present a novel feature-matching algorithm that accurately recovers the position of image features over the entire organ's surface. Our method is fully automatic, it does not require any explicit assumption about the organ's 3-D surface, and leverages Gaussian Process Regression to incorporate noisy matches in a probabilistically sound way. We have conducted extensive tests with a large database of more than 100 endoscopic-image pairs, which show the improved accuracy and robustness of our approach when compared to current state-of-the-art methods.
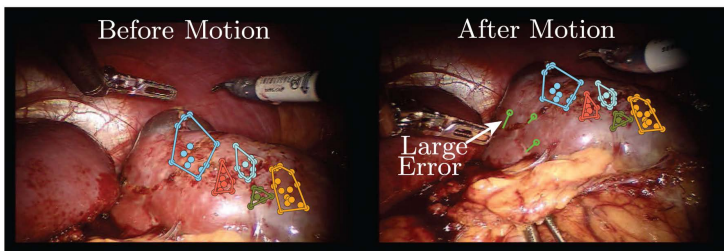
1   Introduction

Providing a reliable solution for the recovery of tracked image features [1] that were lost after a sudden endoscopic-camera motion, or strong organ deformation, is key in order to guarantee long-term performances in many endoscopic-imaging applications, such as augmented reality (AR) [2–4], structure and camera-motion estimation [5–7], and soft-tissue shape recovery [8–10].

Feature matching (or "tracking-by-detection") [11, 12] is a promising technique, as it deals with retrieving corresponding features between two distinct images, even when separated by a wide baseline. Differently from feature tracking [5, 13–16], feature matching does not make any temporal assumption about the two images, about the camera pose [17], or about the 3-D geometry of the observed scene (e.g., known organ motion due to periodic breathing).
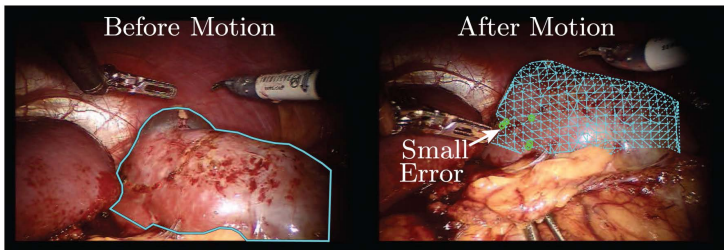
Despite recent advances [18–20], existing feature-matching algorithms are characterized by limiting assumptions and, most important, have not yet met the necessary levels of accuracy, especially when used to recover features in distorted or poorly-textured areas. Figure 1(a) illustrates an example obtained by

using the HMA feature-matching approach [21] (a toolbox is publicly available at [22]). Even if this method recovers a larger number of correct matches (inliers) when compared to existing methods, HMA exhibits a large error (green lines) when used to recover the position of features on the organ's boundaries (green circles). These large errors are caused by the absence of supporting geometric constraints (the colored polygons correspond to affine mappings [23]) in such highly-distorted or textureless areas.

Other regularization methods have been proposed to recover the non-rigid object's shape, such as Finite Element Methods (FEM) [24, 25], and point-set methods (such as Thin-Plate Splines - TPS) [26, 27]. However, FEM-based regularization approaches require an explicit 3-D mesh representation of the organ's surface, and cannot properly describe sharp surface boundaries. Point-set methods are computationally intensive and few of them can be applied to point sets extracted from real endoscopic images.

(a) State-of-the-art HMA Method

(b) Our Method

Fig. 1. (a) The recent feature-matching HMA method works accurately when used to predict the position of features near-by the supporting affine transformations (colored polygons), but exhibits large pixel errors (green line) when used to the feature position on organ's boundaries. (b) Our proposed strategy *accurately and densely* recovers the position of image features over the *entire* organ's surface.

In [28] the authors adopt a TPS warping coupled with bending constraints that gives robustness to self-occlusions. This deformable model is used as geometric

constraint to remove outliers based on their reprojection error. While some effort have been shown for applying this method in an AR surgical application [29], it is distant from a clinical use due to its computational complexity and the sensitivity to outliers; moreover, there is no reported quantitative analysis about its accuracy and robustness in real clinical cases. A different approach is followed in [25] where Gaussian Processes (GP) [30] are used to model the deformable mapping between images. The authors interleaved the GP estimation with an iterative noise-level thresholding to retain only the most certain matches. These filtered matches are used as training data for a weighted-GP regression that estimates the deformable mapping. However, this method tends to eliminate potential correct matches, especially those ones away from the majority of the inliers (e.g. on the object borders). Furthermore, this algorithm has not been evaluated on endoscopic scenarios and it usually fails in cases of large change of view and strong deformations, which are very usual in endoscopic images.

In this paper, we present a novel feature-matching algorithm that *accurately and densely* recovers the position of image features over the *entire* organ's surface. Our method is fully automatic, and it does not require any explicit assumption about the organ's 3-D surface. We adopt Gaussian Process Regression to incorporate noisy matches in a probabilistically sound way. A qualitative example of the performance of our method is illustrated in Fig. 1(b), which shows the dense mapping and the decreased reprojection error. Furthermore, this is the first work that uses GP for dense feature matching in real endoscopic images which are more challenging than the in-lab scenarios tested in [25]. We have conducted extensive tests using a large database with more than 100 endoscopic-image pairs showing the improved accuracy and robustness of our approach when compared to the current state-of-the-art HMA algorithm [21] and the other GP method in [25].

The paper is organized as follows: Sect. 2 presents our dense feature-matching algorithm. Sect. 3 illustrates the results of our extensive experimental evaluation and discusses the algorithms' performance. Finally, Sect. 4 draws the conclusions of this work.

2   Methods

In this section we present our dense feature-matching algorithm that uses as input a training (or template) image, $\mathcal{I}^t$, and a query (or output) image $\mathcal{I}^q$ (e.g., the one before and after a complete occlusion, respectively) and returns a dense mapping between them.

Our strategy consists of two phases: in the first *Sparse Feature-Matching* phase, a set of n sparse matches $\mathcal{M} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$ are obtained by processing a set of appearance-based (e.g., SIFT) initial matches extracted from $\mathcal{I}^t$ and $\mathcal{I}^q$. In this first phase, we adopt the state-of-the-art HMA algorithm [21], which returns a set of accurate matches $\mathcal{M}$, as well as a collection of *rigid* (affine) mappings between both images. As mentioned in Sect. 1, we observed that HMA can easily lead to large errors when these rigid mappings are used to predict the

position of some features located in an organ's area no associated to a specific affine mapping (e.g., on the organ's boundaries where only a few SIFT features are found due to high distortions).

In the second phase, the feature matches from HMA are used as input to a *Gaussian Process (GP) regression*. The goal of this GP phase is to find a latent dense mapping function $f(\mathbf{x})$ that models the *non-rigid motion* of any image point $\mathbf{x}_* \in \mathcal{I}^t$ to the corresponding point $\mathbf{y}_* = f(\mathbf{x}_*)$ in $\mathcal{I}^q$. This second phase also outputs a *dense* set of correspondences that adapt more appropriately to the shape of the organ. The GP phase is described in detail in the next section.

2.1 Gaussian Process Regression Phase

In this phase, the sparse image matches in $\mathcal{M} \triangleq \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$ are used to estimate the *non-rigid and dense* function $f(\mathbf{x})$ that maps any image point $\mathbf{x}_* \in \mathcal{I}^t$ to its corresponding image point $\mathbf{y}_* \in \mathcal{I}^q$. $f(\mathbf{x})$ can be written as, $\mathbf{y}_* = f(\mathbf{x}_*) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is an additive Gaussian noise with zero mean and standard deviation σ. Each knot in the mesh-grid in Fig. 1(b)-right, corresponds to each predicted point $\mathbf{y}_*$ on $\mathcal{I}^q$.

The HMA matches in $\mathcal{M}$ are used as input to a Gaussian Process [30], which describes the posterior distribution of the latent function f as, $p(f|\mathcal{M}) \propto p(\mathcal{M}|f)p(f)$, where $f(\mathbf{p}) \sim \mathcal{GP}(\mu(\mathbf{p}), k(\mathbf{p}, \mathbf{q}))$, $\mathbf{p}$ and $\mathbf{q}$ being two generic vectors. Since a GP is completely determined by its mean and covariance matrix, we define, $\mu(\mathbf{p}) = \mathbf{p}$, as the deterministic mean of the non-rigid function f [25]. The GP covariance matrix is defined as, $k(\mathbf{p}, \mathbf{q}) = \lambda exp\left(-\frac{||\mathbf{p}-\mathbf{q}||^2}{2\rho^2}\right)$, where λ is the amplitude of the process variance and ρ is the length of the process variation.

Note that $p(f|\mathcal{M})$ captures the interaction between the *observed* correspondences and the latent function f, while $p(f)$ represents the *a-priori* knowledge about the non-rigid mapping in terms of the given mean and covariance of the GP.

We denote as $\mathbf{X} \in I\!\!R^{n \times 2}$ and $\mathbf{Y} \in I\!\!R^{n \times 2}$ the matrices containing all the n 2-D HMA matches in $\mathcal{I}^t$ and $\mathcal{I}^q$, respectively. The matrix $\mathbf{X}_* \in I\!\!R^{m \times 2}$ contains the m features to be predicted, and $\mathbf{Y}_* \in I\!\!R^{m \times 2}$ all the corresponding predicted points, so that $\mathbf{y}_{*i} = f(\mathbf{x}_{*i}) + \boldsymbol{\epsilon}$. Then, the joint distribution of both the observed values, $\mathbf{Y}$, and of the target values, $\mathbf{Y}_*$, under the prior is given by,

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{Y}_* \end{bmatrix} \propto \mathcal{N} \left(\begin{bmatrix} \mathbf{X} \\ \mathbf{X}_* \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I} & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right),$$

where each (i, j) entry for the matrices $\mathbf{K}(\mathbf{X}, \mathbf{X}) \in I\!\!R^{n \times n}$, $\mathbf{K}(\mathbf{X}_*, \mathbf{X}) \in I\!\!R^{m \times n}$, and $\mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \in I\!\!R^{m \times m}$ is defined as $k(\mathbf{x}_i, \mathbf{x}_j)$ for row vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ belonging, respectively, to $\mathbf{X}$ or $\mathbf{X}_*$.

Then, the points predicted by the non-rigid function [30] corresponding to $\mathbf{X}_*$ are given by:

$$\mathbf{Y}_* = \mathbf{X}_* + \mathbf{K}(\mathbf{X}_*, \mathbf{X})(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1}(\mathbf{Y} - \mathbf{X}). \tag{1}$$

Note that each entry $k(\cdot, \cdot)$ varies according to the values of its *hyper-parameters* λ and ρ. In order to improve the performance of the GP, we iteratively selected these parameters by maximizing the marginal log-likelihood of the model with respect to the hyper-parameters. The marginal log-likelihood of the model given the n training matches is given by: $\log\left(p(\mathbf{Y}|\mathbf{X}, \sigma, \boldsymbol{\theta})\right) = \sum_{i=1}^{2} -0.5(\mathbf{Y}_i - \mathbf{X}_i)^T \boldsymbol{\alpha}_i - 0.5\log|\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}| - 0.5\,n\log 2\pi$, where $\boldsymbol{\theta} \triangleq [\theta_1, \theta_2]^T = [\rho, \lambda]^T$, $\boldsymbol{\alpha}_i = (\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I})^{-1}(\mathbf{Y}_i - \mathbf{X}_i)$ and $\mathbf{Y}_i$ and $\mathbf{X}_i$ represent the i-th column of $\mathbf{Y}$ and $\mathbf{X}$, respectively. The gradients of this marginal log-likelihood over each component of the hyper-parameter vector are:

$$\frac{\partial \log\left(p\left(\mathbf{Y}|\mathbf{X}, \sigma, \boldsymbol{\theta}\right)\right)}{\partial \theta_j} = \frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T - \left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\right)\frac{\partial \mathbf{K}}{\partial \theta_j}\right).$$

This log-likelihood is maximized by using its gradients into a non-linear optimization method.

3   Experimental Results and Discussion

We present here our experimental validation conducted on a large and publicly-available benchmark containing: an *In-lab* dataset, with 18 image-pairs of a textured non-planar object, subject to controlled rotations and deformations, and a *Surgical-Image* dataset with more than 100 cases extracted from real partial-nephrectomy surgeries [18]. Figure 2 illustrates some representative examples of the adopted benchmark.

This benchmark is particularly appealing for our method as it contains manually-annotated (ground-truth) correspondences, thus allowing us to quantitatively assess the performances of our proposed algorithm and compare them with other state-of-the-art methods by measuring the pixel reprojection error. We compared the accuracy of our algorithm with respect to Zhu's method [25] and HMA [18]. In the case of Zhu and HMA, the training data consists of a set of appearance-based matches (SIFT). The parameters in Zhu and HMA were selected based on the reported values on their respective papers with slight changes in order to maximize their accuracy.

In our method, SIFT features were extracted and the set of sparse correspondences are obtained by means of the HMA Toolbox (publicly available at [22]). In our experimental validation, we propose to increase the number of correspondences, $\mathcal{M}$, by using the affine transformation mappings provided by HMA and by also including those initial SIFT matches with a pixel reprojection error below to fifteen pixels. This choice is based on the ROC analysis reported in [21] which shows that increasing this threshold introduces other correct matches that were not captured by the affine model and few slightly incorrect matches. These extra matches are useful to allow other image features on the entire organ's surface (in particular in the organ's boundaries) to become a support for the GP estimation phase. Also note that the noise introduced by those erroneous matches is already modeled by σ in the GP. The correspondences $\mathcal{M}$ were then used as the training data $\mathbf{X}$ and $\mathbf{Y}$ for the GP regression phase. Based on the precision of HMA's

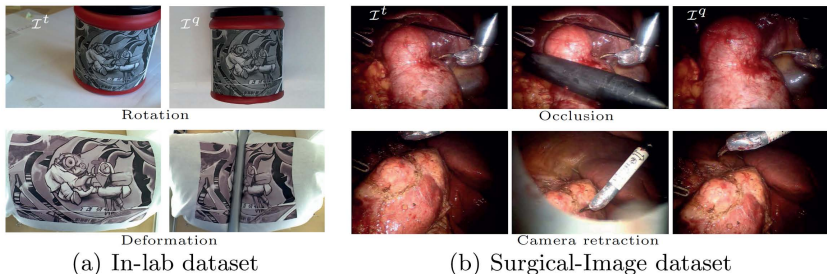(a) In-lab dataset   (b) Surgical-Image dataset

Fig. 2. *Datasets used in our experiments. (a)* Some of the images of the rotation and deformation cases for the In-lab dataset. The left image was fixed as training image, while the object in the query images was rotated or deformed. *(b)* Some of the images in the Surgical image dataset. The training images (left) were acquired before sudden occlusions, camera retraction or strong changes of illumination (center), while the query images (right) were obtained after such events.

affine transformations (set at 5 pixels), we set the noise as $\sigma = \frac{5}{3}$ pixels; the hyper-parameters λ and ρ are estimated as described in Sect. 2. For a better numerical stability, we center each column of $\mathbf{X}$ and $\mathbf{Y}$ to have mean zero and standard deviation one, as well as appropriately scaling σ by the inverse of the larger standard deviation of the columns of $\mathbf{Y}$.

3.1 In-lab Dataset

This dataset includes 7 image-pairs of a rotated textured object whose rotation angles range from -30 to $+30$ degrees [1], as well as 11 cases of a controlled object deformation divided as small, medium and strong deformations. Fig. 2(a) shows some representative examples of this dataset. The image resolution of these images is 640×480, and the number of ground-truth correspondences is in average 55 for the rotation, and 30 for the deformation cases.

Our results are reported in Table. 1: the columns show the mean and standard deviation of the reprojection errors for different object's rotation angles, and object's deformations for our proposed approach, HMA and Zhu's algorithm.

Table 1. *In-lab dataset:* Rotation and Deformation Sets

Method	$-30°$	$-20°$	$-10°$	$0°$	$10°$	$20°$	$30°$	Small Def.	Med. Def.	Str. Def.
Our meth.	5.2±6.6	2.8±3.2	2.2±2.4	1.6±1.9	1.7±1.4	2.4±3.3	2.3±3.6	2.7±4.3	2.2±2.4	2.5±2.7
HMA	9.3±10.3	6.7±6.4	3.7±4.3	3.2±3.8	3.3±3.7	3.6±4.2	4.0±5.5	4±4.5	5.9±10.6	6.5±3.8
Zhu	19.6±14.3	17.6±12.7	13.4±11.15	13.7±9.7	12.6±8.7	13.8±11.6	14.2±12.8	10.9±8.01	21.1±14.0	23.9±14.3

From the results of Table 1 we can observe the enormous improvement of our method over Zhu and HMA. This improvement of our algorithm with respect to Zhu's method is also evident in the qualitative examples shown in Fig. 3. An example of the rotation case is shown in Fig. 3(a) which shows that our method

[1] These angles were chosen to guarantee a good percentage of reliable SIFT matches [31].
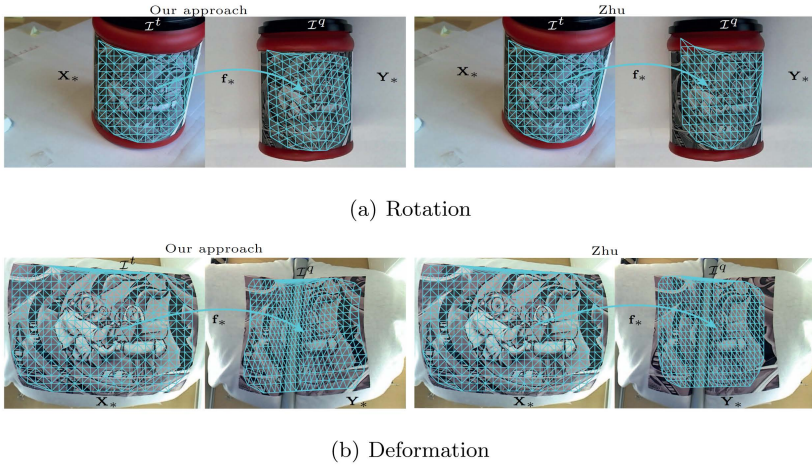
(a) Rotation

(b) Deformation

Fig. 3. *Qualitative comparison*: *(a)* Estimated correspondences $\mathbf{X}_*$ and $\mathbf{Y}_*$ using our approach and Zhu's method, for image pair 4 of the rotation set of the In-lab dataset. *(b)* Estimated correspondences for image pair 6 of the deformation set of the In-lab dataset.

can successfully match the features between images (blue meshes). Conversely, Zhu has problems to match the boundaries due to the lack of matches in that area and the reduced robustness to outliers of their adaptive thresholding approach. As a result, the estimated mapping for these features is very similar to their original position in the training image. Moreover, Fig. 3(b) shows the resulting example for a strong deformation. Note that our approach obtains a very precise mapping as noticeable in the borders of the (blue) warped mesh. On the other hand, Zhu shows problems in the borders. We observed that Zhu is in fact very conservative when filtering its training data; as a result, those matches on the borders are usually removed or strongly penalized when there is a large viewpoint change or a strong deformation between images.

3.2   Surgical-Images Dataset

This dataset contains 100 images extracted before and after camera occlusions, organ deformation, camera retraction/re-insertion, and large changes of illumination, as shown in the examples of Fig. 2(b).

A qualitative comparison between our approach and Zhu's method is presented in the challenging examples of Figs. 4 and 5, which illustrate two cases of very poor textured and highly-cluttered images. Figures 4(a) show the mapping (blue meshes) between the images for both algorithms. Note that our approach results in a higher accurate mapping (e.g., see the organ boundaries indicated by the yellow arrows). In particular, our approach achieves better results than Zhu because it uses more informative training data (this can be observed in Figs. 4(b) which show the matches used as training data). Note that, even if both methods used the *same set* of input appearance-based matches, Zhu's adaptive thresholding
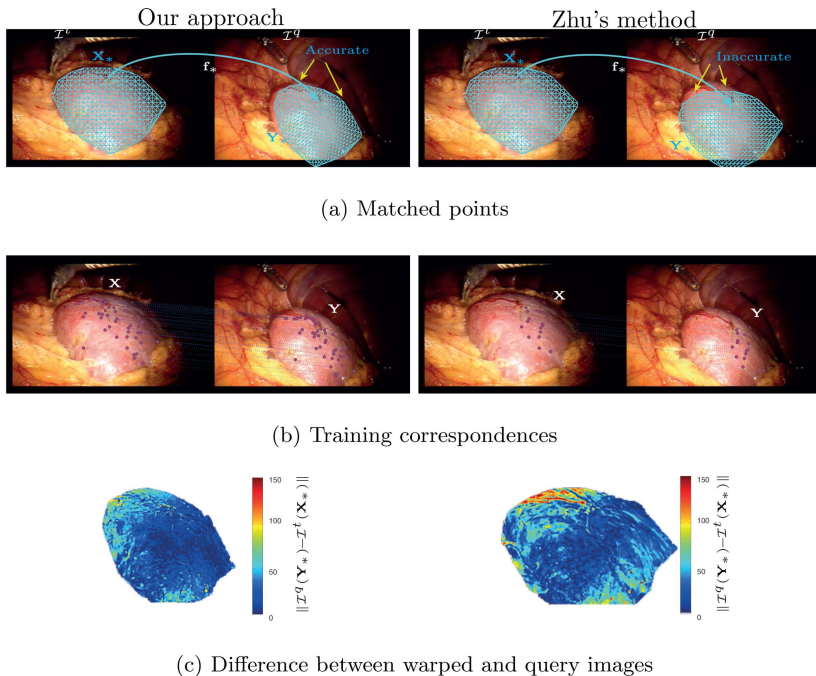
Our approach   Zhu's method

(a) Matched points

(b) Training correspondences

(c) Difference between warped and query images

Fig. 4. *Qualitative comparison*: *(a)* Estimated correspondences $\mathbf{X}_*$ and $\mathbf{Y}_*$ using our approach and Zhu's method, for image pair 10 of the Surgical-Image database. *(b)* Set of sparse matches used as training data for both methods (dashed lines). *(c)* Absolute pixel difference between the warped and the reference regions.

removes a large number of matches, in particular those on the organ's boundaries. As a result, the estimated mapping has large errors. This is a very common issue for GP where the uncertainty in the prediction quickly grows on the borders of the training data. We also evaluate, in Figs. 4(c), the quality of the mapping by computing the difference of intensities between the warped ($\mathcal{I}^t(\mathbf{X}_*)$) and the corresponding query ($\mathcal{I}^q(\mathbf{Y}_*)$) images. Despite illumination changes, observe that the areas with large similarity (blue areas) between warped and query images are more evident with our approach than with Zhu.

Moreover, Figs. 5(a) show a case where Zhu fails to prune the noisy appearance-based matches due to high clutter of the images and because Zhu's tight constraints remove the matches on the top and left-parts of the organ. Conversely, our algorithm is able to estimate a very precise mapping because of its capacity to deal with noisy matches. This is evident from Figs. 5(b) which show the matches used as training sets by each algorithm. As a result, the intensity similarities between warped and query images are significantly lower for our approach than Zhu. Note that for our method, the similarity errors, in Figs. 5(c), are mostly caused by illumination changes between views.
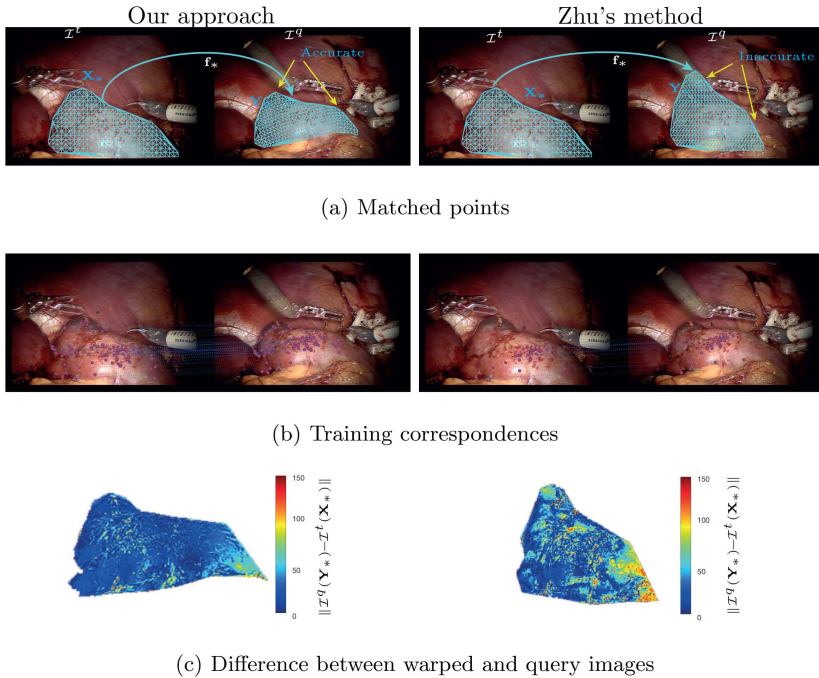
Our approach   Zhu's method

(a) Matched points

(b) Training correspondences

(c) Difference between warped and query images

Fig. 5. *Qualitative comparison*: *(a)* Estimated correspondences $\mathbf{X}_*$ and $\mathbf{Y}_*$ using our approach and Zhu's method, for image pair 15 of the Surgical-Image database. *(b)* Set of sparse matches used as training data for both methods (dashed lines). *(c)* Absolute pixel difference between the warped and the reference regions.

To provide a quantitative comparison, we decided to partition the dataset in three categories according to their percentage of correct matches[2] as: less than 15% (very hard), between $15 - 40\%$ (hard) and more than 40% (regular) correct matches. Table 2 summarizes the results of the reprojection errors (mean and standard deviation) through the three categories.

Table 2. *Surgical-Image dataset:* Average errors (pixels) over the 100 image pairs

Algorithm	Avg.Err. [pixels] $0 - 15\%$ inliers	Avg.Err. [pixels] $15 - 40\%$ inliers	Avg.Err. [pixels] $40 - 100\%$ inliers	Failures
Our method	**4.49±4.68**	**3.4±3.23**	**1.84±1.66**	**5**
HMA	5.89±6.27	4.59±4.79	2.26±2.29	5
Zhu	16.07±9.39	9.43±7.11	7.07±4.63	9

The results in Table 2 show a 70% error reduction of our approach with respect to Zhu, thus supporting the observations from the qualitative examples. Moreover, Zhu tends to fail in almost twice number of cases due to convergence problems. In addition to providing a dense mapping, our proposed approach has an accuracy improvement of approximately 23% when compared to HMA.

---

[2] The appearance-based matches and the percentage of final correct matches (inliers) were provided from the dataset.

4   Conclusions

We presented a feature-matching algorithm that accurately and densely recovers the position of image features over the entire organ's surface. Our method is fully automatic, it does not require any explicit assumption about the organ's 3-D surface, and it combines sparse feature-matching with the probabilistic modeling of Gaussian Process Regression. We demonstrated the effectiveness of our algorithm under a highly-controlled in-lab dataset, and a large endoscopic dataset with more than 100 images of challenging scenarios. We measured the reprojection errors with respect to manually annotated ground-truth and observed that one of the main advantages of our proposed algorithm is its higher accuracy when compared to state-of-the-art methods, in particular for features closer to the tissue's boundaries. This happens because our algorithm can leverage the extra information provided from (initially-discarded and noisy) SIFT matches, thanks to the probabilistic noise modeling of Gaussian Processes and the hyper-parameters' estimation.

References

1. Shi, J., Tomasi, C.: Good features to track. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 593–600 (June 1994)
2. Cohen, D., Mayer, E., Chen, D., Anstee, A., Vale, J., Yang, G.Z., Darzi, A., Edwards, P.: Augmented reality image guidance in minimally invasive prostatectomy. In: Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention, pp. 101–110 (2010)
3. Su, L.M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D.: Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3D-CT to stereoscopic video registration. Urology 73(4), 896–900 (2009)
4. Higgins, E.W., Helferty, P.J., Lu, K., Merritt, A.S., Lav, R., Kun-Chang, Y.: 3d ct-video fusion for image-guided bronchoscopy. Computerized Medical Imaging and Graphics 32(3), 159–173 (2008)
5. Mountney, P., Yang, G.-Z.: Motion compensated SLAM for image guided surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 496–504. Springer, Heidelberg (2010)
6. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking. IEEE Signal Processing Magazine 27(4), 14–24 (2010)
7. Hu, M., Penney, G.P., Rueckert, D., Edwards, P.J., Bello, F., Casula, R., Figl, M., Hawkes, D.J.: Non-rigid reconstruction of the beating heart surface for minimally invasive cardiac surgery. In: Proc. of the 12th Int. Conf. on Med. Image Comp. and Comp.-Ass. Int., pp. 34–42 (2009)
8. Lo, B.P.L., Visentini-Scarzanella, M., Stoyanov, D., Yang, G.Z.: Belief propagation for depth cue fusion in minimally invasive surgery. In: Proc. of the 11th Int. Conf. on Med. Image Comp. and Comp.-Ass. Int, pp. 104–112 (2008)
9. Visentini-Scarzanella, M., Mylonas, G.P., Stoyanov, D., Yang, G.Z.: *i*-brush: A gaze-contingent virtual paintbrush for dense 3d reconstruction in robotic assisted surgery. In: Proc. of the 12th Int. Conf. on Med. Image Comp. and Comp.-Ass. Int., pp. 353–360 (2009)

10. Totz, J., Mountney, P., Stoyanov, D., Yang, G.Z.: Dense surface reconstruction for enhanced navigation in MIS. In: Proc. of the 14th Int. Conf. on Med. Image Comp. and Comp.-Ass. Int., pp. 89–96 (2011)
11. Lepetit, V., Fua, P.: Monocular model-based 3-d tracking of rigid objects: A survey. Foundations and Trends in Computer Graphics and Vision 1, 1–89 (2005)
12. Puerto-Souza, G.A., Mariottini, G.L.: A comparative study of correspondence-search algorithms in mis images. In: Proc. of the 15th Int. Conf. on Med. Image Comp. and Comp.-Ass. Int., pp. 625–633 (2012)
13. Stoyanov, D., Mylonas, G., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In: Proc. of the 8th Int. Conf. on Med. Image Comp. and Comp.-Ass. Int., pp. 139–146 (2005)
14. Richa, R., Bo, A.P., Poignet, P.: Towards robust 3d visual tracking for motion compensation in beating heart surgery. Medical Image Analysis 15(3), 3012–3315 (2010)
15. Giannarou, S., Visentini-Scarzanella, M., Yang, G.Z.: Probabilistic tracking of affine-invariant anisotropic regions. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2012)
16. Yip, M., Lowe, D., Salcudean, S., Rohling, R., Nguan, C.: Real-time methods for long-term tissue feature tracking in endoscopic scenes. In: Information Processing in Computer-Assisted Interventions, pp. 33–43 (2012)
17. Tola, E., Lepetit, V., Fua, P.: Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(5), 815–830 (2010)
18. Puerto-Souza, G.A., Mariottini, G.L.: Hierarchical multi-affine (HMA) algorithm for fast and accurate feature matching in minimally-invasive surgical images. In: Proc. IEEE/RSJ Int. Conf. Intel. Robots Syst., pp. 2007–2012 (October 2012)
19. Del Bimbo, A., Franco, F., Pernici, F.: Local shape estimation from a single key-point. In: Proc. Comp. Vis. Patt. Rec. Workshops, pp. 23–28 (2010)
20. Cho, M., Lee, J., Lee, K.M.: Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: Proc. 9th Int. Conf. Comp. Vis., pp. 1280–1287 (2009)
21. Puerto-Souza, G.A., Mariottini, G.L.: A Fast and Accurate Feature-Matching Algorithm for Minimally-Invasive Endoscopic Images. IEEE Transactions on Medical Imaging (in Press, 2013)
22. HMA feature-matching toolbox (Web),
    `http://ranger.uta.edu/~gianluca/feature_matching/`
23. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge Univ. Press (2000)
24. Pilet, J., Lepetit, V., Fua, P.: Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation. International Journal of Computer Vision 76(2) (2008)
25. Zhu, J., Hoi, S., Lyu, L.: Nonrigid shape recovery by gaussian process regression. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1319–1326 (2009)
26. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 509–522 (2001)
27. Myronenko, A., Song, X.: Point Set Registration: Coherent Point Drift. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(12), 2262–2275 (2010)

28. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion reasoning. International Journal of Computer Vision 97(1), 54–70 (2012)
29. Kim, J.-H., Bartoli, A., Collins, T., Hartley, R.: Tracking by detection for inter-active image augmentation in laparoscopy. In: Dawant, B.M., Christensen, G.E., Fitzpatrick, J.M., Rueckert, D. (eds.) WBIR 2012. LNCS, vol. 7359, pp. 246–255. Springer, Heidelberg (2012)
30. Rasmussen, C., Williams, C.: Gaussian processes for machine learning, vol. 1. MIT press, Cambridge (2006)
31. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)