



2D reconstruction of small intestine's interior wall

Rahman Attar^{a,c}, Xiang Xie^b, Zhihua Wang^b, Shigang Yue^{c,*}

^a School of Computing, University of Leeds, Leeds, UK

^b Institute of Microelectronics, Tsinghua University, Beijing, China

^c School of Computer Science, University of Lincoln, Lincoln, UK

ARTICLE INFO

Keywords:

Wireless endoscopy
Image stitching
Image registration

ABSTRACT

Examining and interpreting of a large number of wireless endoscopic images from the gastrointestinal tract is a tiresome task for physicians. A practical solution is to automatically construct a two dimensional representation of the gastrointestinal tract for easy inspection. However, little has been done on wireless endoscopic image stitching, let alone systematic investigation. The proposed new wireless endoscopic image stitching method consists of two main steps to improve the accuracy and efficiency of image registration. First, the keypoints are extracted by Principle Component Analysis and Scale Invariant Feature Transform (PCA-SIFT) algorithm and refined with Maximum Likelihood Estimation Sample Consensus (MLESAC) outlier removal to find the most reliable keypoints. Second, the optimal transformation parameters obtained from first step are fed to the Normalised Mutual Information (NMI) algorithm as an initial solution. With modified Marquardt-Levenberg search strategy in a multiscale framework, the NMI can find the optimal transformation parameters in the shortest time. The proposed methodology has been tested on two different datasets - one with real wireless endoscopic images and another with images obtained from Micro-Ball (a new wireless cubic endoscopy system with six image sensors). The results have demonstrated the accuracy and robustness of the proposed methodology both visually and quantitatively - registration residual error of 0.93 ± 0.33 pixels on 2500 real endoscopic image pairs and residual error accumulation of 16.59 pixels and without affecting the visual registration quality on stitching 152 images of Micro-Ball.

1. Introduction

Development of wireless endoscopy has given gastroenterologists the opportunity to investigate the small intestine in a non-invasive way. Wireless capsule endoscopy is most commonly performed for obscure gastrointestinal (GI) bleeding, but also for other indications such as diagnosis of Crohns disease, suspicion of a small bowel tumour, diagnosis and surveillance of hereditary polyposis syndromes, non-steroidal anti-inflammatory drug-induced small intestinal lesions and celiac disease [1]. Wireless endoscopy offers several potential advantages over traditional esophagogastroduodenoscopy including the ability to perform 24 h a day, without sedation, and interpreted at the bedside by physicians. In addition, it is painless and much less invasive, so enables the patient to pursue daily routines after the procedure. Therefore, it has a high patient acceptability rate [2]. However, interpretation procedure is time consuming and tiresome. Once swallowed, the device which typically has one or two cameras moves through the digestive tract and takes two or three pictures (depending on the model) per second for about eight hours. Thus, a large number of images are

generated. Moreover, low illumination and random movements of the device and therefore non-perfect shooting angles make the images less understandable. Besides, there are some symptoms (such as ulcer or polyp) that are usually spread over large areas in GI tract, whereas the endoscopic image area is very small [3]. In order to detect these symptoms, it is crucial to build panoramic images of interior walls to extend the field of view. Consequently, if a 2D representation map of the whole GI tract is possible to generate, it will be much easier to interpret rather than investigating the images individually. However, it should be noted that the capsule has an average missing rate of , due to limited visual field, random movements and low capture rate of the cameras, that imposes a challenge [4]. To address the missing-rate issue, a smart wireless Micro-Ball endoscopy prototype system was proposed by Gu et al. [5]; where six cameras, constructing a cubic structure, are embedded in the Micro-Ball system. Determined by a smart control system, the device operates in monitor, normal and drastic modes with a capture rate of 1, 6 and 24 fps respectively. To reduce power consumption, a camera selection strategy is applied so that only one or two cameras work simultaneously without missing any

* Corresponding author.

E-mail address: syue@lincoln.ac.uk (S. Yue).

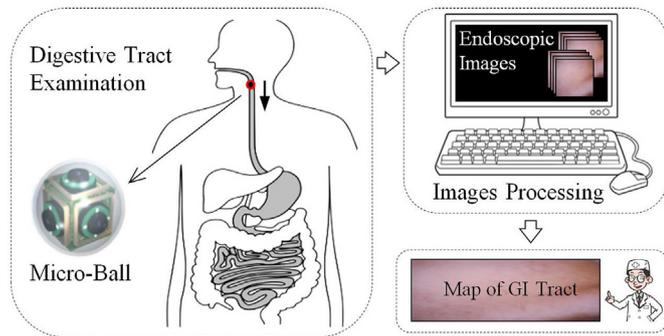


Fig. 1. 2D representation of GI tract images for clinical diagnosis.

areas.

As shown in Fig. 1, the GI tract map can then be constructed successfully provided the images are efficiently registered. In this context, image registration is defined as follows: given two images, I_A and I_B , to register the images is to find an optimal geometric transformation function T_μ , where $T_\mu(I_A)$ best matches I_B , where μ is a set of transformation parameters.

Considering that the GI tract is soft and flexible together with the fact that the endoscopy device is propelled by gravity and peristaltic movements within the GI tract, reconstruction of the interior wall seems to be a formidable task and current algorithms fail to achieve satisfactory results within a reasonable time frame. Therefore, a new approach is required to overcome this challenge.

To tackle the imposed challenges, attempts have been made by Wang et al. [6] to create a 2D representation of GI tract with the Micro-Ball. This work has proposed a method of a straightforward combination of two image registration algorithms, i.e. Phase Correlation Method (PCM) and Scale Invariant Feature Transform (SIFT) [7] to assess image registration algorithms on this particular application. However, in practice, further investigations are crucial to provide a better representation of the small intestinal tract for physicians.

The objective of the present work is to further improve the solution to the 2D reconstruction of the small intestines interior wall by proposing a new methodology to register images for easier inspection. Images are enhanced with a pre-processing method designed by Attar et al. [8]. The keypoints are extracted with Principle Component Analysis and SIFT (PCA-SIFT) algorithm [9], and then processed with Maximum Likelihood Estimation Sample Consensus (MLESC) [10] outlier removal to select the most correlated keypoints between paired images. The set of transformation parameters μ are considered as the candidate solutions of transformation function. To improve the accuracy and efficiency of image registration, the optimal set of transformation parameters μ are tuned to maximise the Normalised Mutual Information (NMI) value between paired images. As becomes evident in

the sequel, application of this technique on real wireless endoscopic images and also a unique database of images obtained from *in vitro* experiments with the Micro-Ball demo system, shows the capability of reconstructing a 2D representation for visual inspection. The new proposed method works well for images of the current endoscopy devices, also is a perfect fit for the capsules with multiple cameras and larger field of view like the Micro-Ball.

The paper is organised as follows. In Section 2, endoscopic images and related research are reviewed briefly. The whole problem of 2D representation and the contribution of this paper is given in Section 3. In Section 4, the image stitching algorithm is developed. The experimental results are presented in Section 5. Finally, Section 6 concludes the paper and proposes future work.

2. Endoscopic image analysis and related work

Bad imaging conditions such as low illumination, high compression ratio, limitations in power consumption, short-focal-length camera, communication bandwidth, and storage device all affect the image quality [11,12]. Observing a large number of histograms of endoscopic images, it can be seen that distribution of pixels concentrates in a certain range of intensity bins. Fig. 2 (left) shows distribution of pixels for ten samples. Furthermore, the most striking result to emerge from investigating all pixels of whole images in the dataset (including 89,800 images from two subjects) is that 81.86% of pixels fall in 28.12% of intensity bins. Moreover, the frequency domain analysis using Discrete Fourier Transform (DFT) for ten typical images of the GI tract is shown in Fig. 2 (right). In the frequency domain image, each point represents a particular frequency contained in the spatial domain image. The number of high frequency components are too small. It means that the features are not distinct and it brings challenges both for medics to interpret and engineers to process them for different applications.

As noted earlier, thousands of images with the aforementioned features are recorded for physicians to inspect. Depending on the experience of the examiner, the average time required for the visual inspection ranges from 45 to 120 min. This will result in an unnecessary waste of time and efforts. Bergen and Wittenberg [13] have done a comprehensive study of the endoscopic image stitching and surface reconstruction methods. In literature, to the best of our knowledge, apart from those methods that directly detect abnormal frames, there are two main groups of algorithms that have been proposed to reduce the visual inspection time, yet improve the quality of examination.

First group reduce the number of images by removing redundancy from dataset. For example: In Ref. [14], a registration methodology has been proposed to reduce video length by removing same or similar frames. It utilises a segmentation scheme and a graph method to register similar regions in two consecutive frames. In a similar work [3], it is shown that for image registration, the sparse graph cut method yields a computation speed up by an order of magnitude in comparison to a

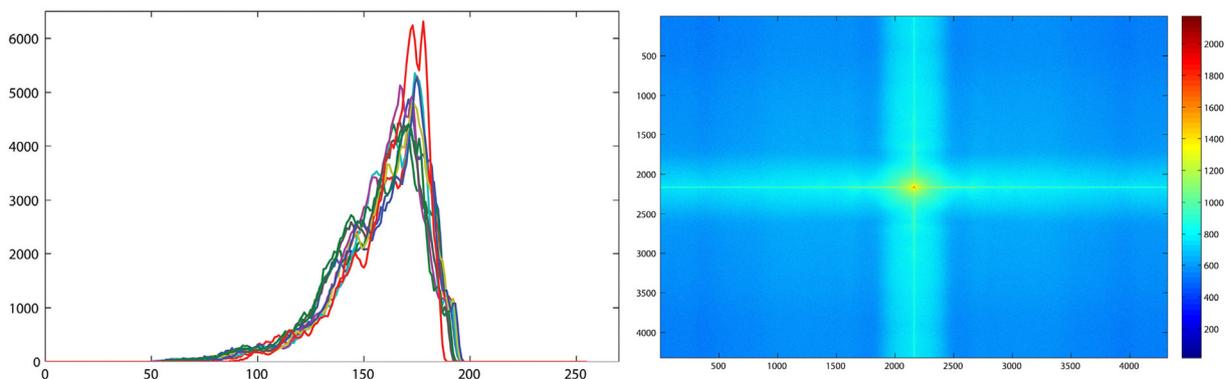


Fig. 2. Histogram (left) and Frequency analysis (right) of 10 sample endoscopic images.

standard dense graph cut. In a different study [15], a combination of an intensity correction method with a method based on optical flow and features is proposed to detect and reduce near-duplicate images. With the same objective in Ref. [16], an approach is proposed based on a data reduction algorithm to extract representative video frames from the full length endoscopic images.

Second group use image processing and computer vision algorithms to produce clearer, more detailed, and wider view of region of interest. That is to say, information from a set of images introduce a 2D/3D informative representative. For example: In Ref. [17], a method has been proposed to facilitate clinical diagnosis by building a panoramic image of the bladder using images acquired from different viewpoints. Consecutive images are pairwise registered and then all are placed in a common and global coordinate system. More recently, a method for automatic construction of panoramic visual summaries was proposed by Spyrou et al. [18]. It involves extraction of visual features from each frame, and frame registration by Speeded Up Robust Features (SURF) algorithm. According to an investigation by Karargyris and Bourbakis [19]; methodologies are proposed that can create new interpolated frames for smoothing video representation and having a more realistic 3D view of a few regions of the digestive tract surface, but not a 3D representation of the whole tract. In Ref. [20], a method is put forward based on the model of deformable rings. The model processes the video to characterise capsule's motion and to produce a map - a rough representation of the GI system's endoluminal surface. The novelty of the model lies in the shape of its mesh, designed particularly to follow the specific type of motions.

However, as Bergen and Wittenberg [13] reported in their study, GI tract is one of the most challenging organs for stitching and surface reconstruction and only a few contributions on the topic can be found in the literature. Also, most of the research has focused on surface reconstruction of only a few patches of the region of interest. Therefore, in practice, further investigations are needed to provide a better solution. Despite the other works that have been done so far, this work focuses on developing a fully automatic method to simultaneously address concerns of both above groups on a large sequence of images, i.e. by reducing the number of images in the dataset as well as providing a wider representation of region of interest for easier inspection.

3. Methodology and contribution

Generally, image registration algorithms can be classified into feature-based and intensity-based methods.

Feature-based methods establish a correspondence between a number of especially distinct points (using keypoints detector algorithms) in images, then a geometrical transformation is determined to map the target image to the reference images. One of the main advantages of these approaches is that they are fast and robust to noises, complex geometric distortions, and significant radiometric differences. Supremacy of SIFT in comparison with other keypoint detectors/descriptors is reported in Refs. [21,22]. Although SIFT performs best for most of the cases it is unable to provide satisfactory result when it's directly applied to the endoscopic images. Thus, using SIFT alone cannot produce optimal results because there are only a small number of correct matches, keypoints are not evenly spread over the image, and many outliers exist in feature matches.

Intensity-based methods, which are more popular in medical image registration, employ the intensity of pixels. Mutual Information (MI) is the most popular measure because of its robustness to intensity variation and noise, and high accuracy. It indicates any linear/non-linear correlation [23], and produces consistently sharper peaks at the correct registration values than other intensity-based methods, which is very important for accurate registration. Despite the general promising results, using MI can result in misregistration when the images have low resolution or when the images contain little information [24,25]. However, the registration algorithm performs much better when it

starts maximisation of the Normalised MI (NMI) from an accurate initial solution which is close to ground truth.

Therefore, in this work, a hybrid algorithm is put forward based on the combination of an intensity-based and a feature-based method that has shown the best performance in images registration. The advantages of both methods are employed to complement each other. Based on the experiments that have been done on endoscopic images in this work, Normalised Cross Correlation (NCC) which is an intensity-based method produces very similar results to NMI. When NMI is replaced with NCC in the algorithm the overall error is slightly more (1.27%). SURF (a feature-based method) also was tested instead of SIFT. Although SURF is faster than SIFT (37.88%) the results obtained by using SIFT is more accurate (14.71%). All these results declare that above combination of the methods works the best on endoscopic images but someone may get better results in future works by changing/modifying the method in the initialisation step and/or fine-tuning step. The structure of the algorithm presented here is in a way that initialisation step can be done with a feature-based method and then an intensity-based method is used to tune the transformation function.

The problem of registration is being solved for any two adjacent images of input image dataset: $I_{GI} = \{I_1, I_2, \dots, I_n\}$, it gradually develops along the dataset to form the GI tract map (I_{map}), as illustrated in Fig. 3.

Every image registration algorithm requires three basic ingredients: a spatial transformation model which determines the set of possible solutions, an objective similarity measure which estimates the quality of each potential solution, and an optimisation algorithm which looks for the best solution [26]. Therefore, the whole process is to solve the image registration problem formulated by the following equations for any two adjacent images.

$$\hat{T}_\mu = arg\ opt_{T_\mu} sim((\Omega_A), T_\mu(\Omega_B)) \tag{1}$$

$$\Omega_A = f_A(I_A^{ov}(d_{GI}, d_{CA})) \tag{2}$$

$$\Omega_B = f_B(I_B^{ov}(d'_{GI}, d'_{CA})) \tag{3}$$

where μ is a set of transformation parameters which can lead to rotation, scale, shear, shift, and projection between the two images. Due to the small surface deformation, highly overlapped frames, and slow movement of the camera there are only small displacements between consecutive images of an endoscopic video sequence. Therefore it is possible to successfully register image pairs using 2D perspective transformations [3,17]. I_A^{ov} and I_B^{ov} are the overlapping parts of two successive images I_A and I_B , respectively. Usually, in image registration, homologous information Ω_A and Ω_B are extracted from images using a particular algorithm f . However, due to lack of enough distinct features in endoscopic images, it's difficult to select Ω and f . Meanwhile, image deformation d_{GI} (caused by soft tissue deformation of GI tract) and d_{CA} (caused by random motion of the endoscopy device) have negatively affected the image registration process. Generally, the homologous structures are superimposed by an optimisation procedure opt_{T_μ} , which

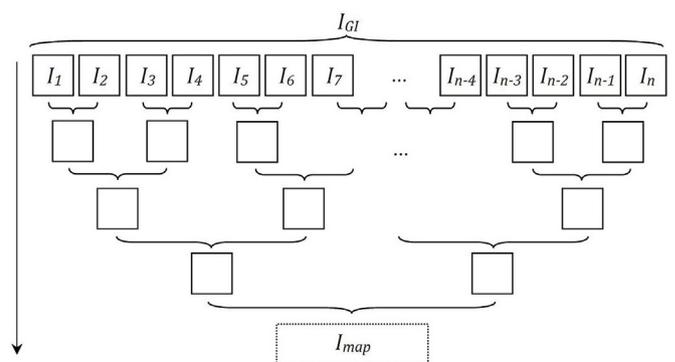


Fig. 3. Scheme of endoscopic image registration of each two adjacent images.

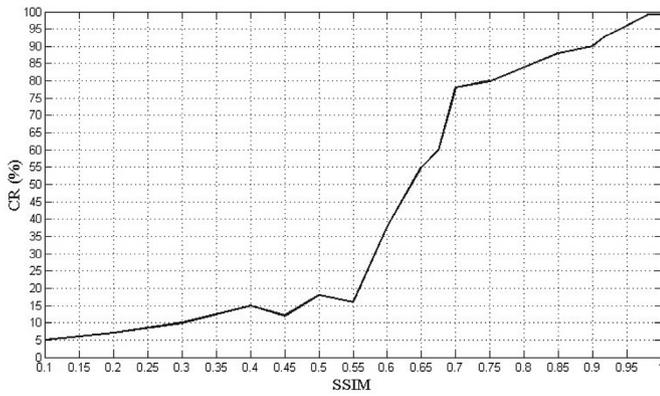


Fig. 4. Behaviour of SSIM vs CR; higher SSIM shows higher correct rate.

maximises the similarity function *sim*, or minimise dissimilarity.

Due to the unpredictable motions (i.e. bouncing) of the capsule camera, the probability that two successive frames are very dissimilar is relatively high. In such cases it is crucial that the methodology should not perform registration. In order to handle these cases, to create a decision-making model, a simple check between two successive images by Structural SIMilarity (*SSIM*) index is used. *SSIM* basically shows the degree of similarity of two images, scaled on the range $0 \leq SSIM \leq 1$, the higher the value of *SSIM*, the more similarity there is [8,27]. Therefore, the registration method is applied to those image pairs whose *SSIM* is greater than a threshold (*th*). According to Fig. 4 which shows Correct Rate (CR; explained in Section 5) versus *SSIM*, the threshold is set to $th = 0.70$. Clearly, for the values lower than *th*, the CR is too low and for the values greater than *th*, CR increases. Thus, based on this predictable behaviour a judgement is possible.

Consequently, the hybrid registration scheme proposed in this work, after preprocessing, is composed of the following main stages, visualised in Fig. 5. To begin with, an initial solution for the transform function is obtained through a preregistration procedure based on scale invariant keypoints which have been extracted and refined with a reliable outlier removal. Then, fine tuning of the registration parameters

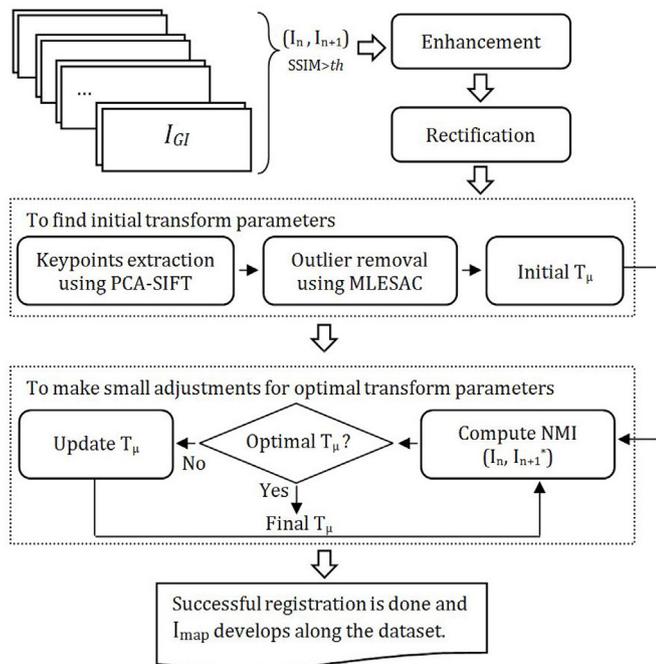


Fig. 5. Flowchart of the proposed method. I_n and I_{n+1} = two consecutive images; I_{n+1}^* = transformed image; T_μ = transform function.

is accomplished in order to maximise NMI over each paired images. The main innovation and contribution of this study are summarised as follows:

- 1) This work proposes an efficient automatic algorithm that employs the robustness of feature-based methods and the accuracy of intensity-based methods to construct a hybrid scheme for endoscopic image registration.
- 2) By examining the performance of the best feature-based keypoints detectors/descriptors algorithm [7], i.e. SIFT algorithm and its two variants (PCA-SIFT and SIFT + GC) with the most widely used outlier removal algorithms (i.e. HT, RANSAC, MSAC and MLESAC) on wireless endoscopic images dataset, the best couple is selected to produce the best primary results to guarantee that the preregistration results are close enough to the ground truth.
- 3) For maximisation of NMI, an excellent initial solution selection strategy is put forward by using PCA-SIFT with MLESAC outlier removal at the coarsest level of search stage. Registration accuracy has improved significantly compared with standard NMI which uses a random initial solution.
- 4) A modification of Marquardt-Levenberg search strategy in a multi-scale framework is exploited to refine the coarser level registration results and reach the most precise registration parameters for maximising NMI.
- 5) The proposed registration strategy, besides common wireless endoscopic images, is tested on a unique database which has been obtained from *in vitro* experiments with the Micro-Ball demo system.

In the following sections, different steps are separately described and discussed in detail.

4. Proposed method

The proposed algorithm, as illustrated in Fig. 5, consists of the following steps:

4.1. Image enhancement

As an efficient preprocessing step to enhance the quality of endoscopic images to make their feature (Ω) more distinct and easier to examine, a simple approach proposed by Ref. [28] is used. This method by histogram information correction and edge preserving mask accentuate features of the interest in endoscopic images.

4.2. Image rectification

Rectification is the process of transforming an acquired image into a rectified image, for which the resulting overall transformation from physical object to rectified image has the ideal form. The distortion correction problem can be solved by an efficient approach as a separate preprocessing step to improve the quality of registration.

To overcome the issue of d_{CA} which is caused by random motion of the Micro-Ball, perspective distortion of wireless endoscopic images needs to be rectified. There are several rectification methods using the advantage of reference corners in images, but it is not easy to automatically find such keypoints in wireless endoscopic images. Fortunately, the location information given by the attitude-sensing system embedded in the Micro-Ball can be used to solve the image rectification problem [29]. In addition to that sensor which includes a magnetometer and an accelerometer, there is another attitude-sensing sensor which is worn on the patient body and it consists of magnetometer, accelerometer, and gyroscope. Using these two attitude sensing systems, the intestine's orientation can be driven, more details are available in Refs. [5,29].

The distortion correction process is briefly shown in Fig. 6. We do image correction to find actual relationship $V(\theta, \varphi, \alpha, focal)$ between

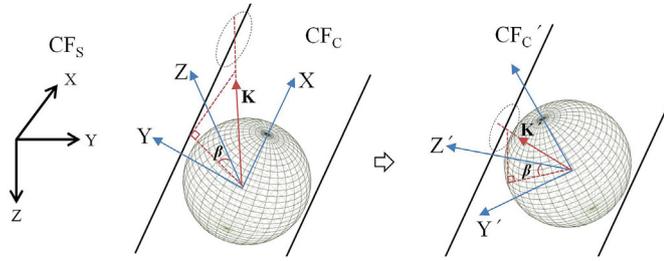


Fig. 6. The diagrammatic sketch of perspective distortion correction.

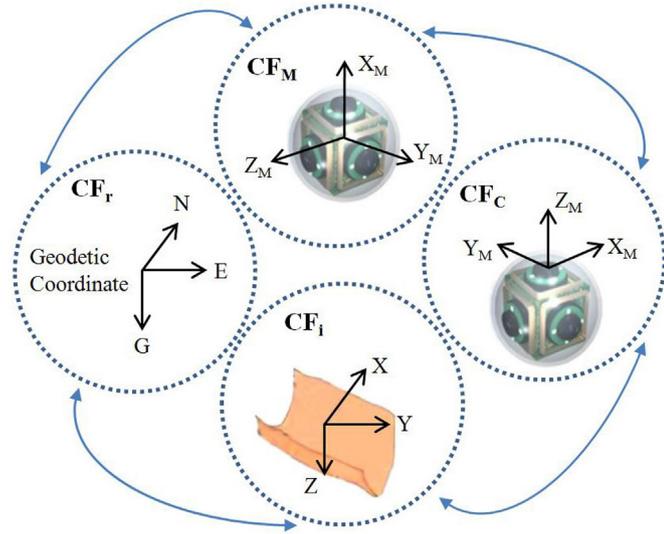


Fig. 7. Coordinate systems in image correction for Micro-Ball images.

pre-corrected image (CF_c) and post-corrected image (CF'_c). θ , φ , α , and $focal$ are, respectively, roll, pitch, yaw, and the focal length of the camera, available from Micro-Ball device. CF_s is the coordinate system relative to the practical plane. Because of the invariability of any physical point coordinates in the real world, then $V(\theta, \varphi, \alpha, focal)$ can be obtained.

In the Micro-Ball endoscopy system, as the captured image plane is changeable and each camera has its own coordinate system, the problem of image geometry correction is more complex. As shown in Fig. 7, the coordinate systems are illustrated as follows:

- CF_r : reference frame of attitude-sensing system;
- CF_M : carrier coordinate system of the Micro-Ball;
- CF_c : frame of the camera to be corrected;
- CF_i : frame of the intestine;
- CF_s : frame corresponding to the image plane captured.

As mentioned earlier, the relationship between CF_c and CF_s must be acquired to obtain image correction, and it is possible through the following four steps:

- Step 1 $CF_M = R \cdot CF_r$; R is the rotation transform matrix given by the endoscopy device.
- Step 2 $CF_c = T_1 \cdot CF_M$; This step is essential, because CF_M is fixed but CF_c is changing with the camera to be corrected.
- Step 3 $CF_r = T_2 \cdot CF_i$; This relationship is acquired based on the intestinal lumen detection combined with coordinate geometry location information from the Micro-Ball.
- Step 4 $CF_i = T_3 \cdot CF_s$. As shown in Fig. 6, the shooting direction of the camera, vector K in CF_i , is denoted as k_{nt}

$$k_{nt} = T_2^{-1}R^{-1} \cdot k_{n0} \quad n = 1,2, \dots,6 \quad (4)$$

where k_{n0} is K 's vector representation in CF_M are as follow

$$k_{10} = (0,0, -1)^T \quad k_{20} = (0,0,1)^T \quad k_{30} = (1,0,0)^T$$

$$k_{40} = (0, -1,0)^T \quad k_{50} = (-1,0,0)^T \quad k_{60} = (0,1,0)^T$$

With k_{nt} , the angle β in Fig. 6 can be calculated and then T_3 can be obtained by

$$T_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix}^T \quad (5)$$

According to above-mentioned steps, the relationship between CF_c and CF_s can be obtained by

$$CF_c = T_1 \cdot R \cdot T_2 \cdot T_3 \cdot CF_s \quad (6)$$

Moreover, $T_1 \cdot R \cdot T_2 \cdot T_3$ can be used to obtain $V(\theta, \varphi, \alpha, focal)$ and realise image correction. Although the image perspective distortion correction is made before the registration process the following perspective transformation function is used to have more accurate results.

$$T_\mu(x, y) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

Fig. 8 shows the result of registration of an image pair with and without image perspective distortion correction. Thus, perspective distortion correction for registration of images obtained through Micro-Ball is a crucial step.

4.3. Scale-invariant feature transform

This section provides a brief overview of SIFT algorithm and its two variants namely PCA-SIFT [9] and SIFT + GC [30]. Following are different steps of SIFT algorithm for generating keypoints and their descriptors: 1) Scale Space creation 2) Laplacian of Gaussian (LoG) Approximation 3) Keypoints detection 4) Keypoints elimination 5) Keypoints orientation 6) Descriptors formation.

After the keypoint descriptor has been calculated, keypoints are matched by using the minimum distance method. However, not every pair of matched keypoints will be useful.

In this work, two different editions of SIFT also are examined to find the best result for the application, in terms of time efficiency and accuracy; SIFT + GC and PCA-SIFT; SIFT + GC integrates global information into SIFT and PCA-SIFT applies PCA to the normalised gradient patch. In Section 5, the performance of three above scale invariant keypoints detectors are investigated and PCA-SIFT is selected for the application as it has better performance than standard SIFT or SIFT + GC.

4.4. Mismatch detection

Although PCA-SIFT has a very good performance, when it is used alone for image registration, many false matches arise. These false

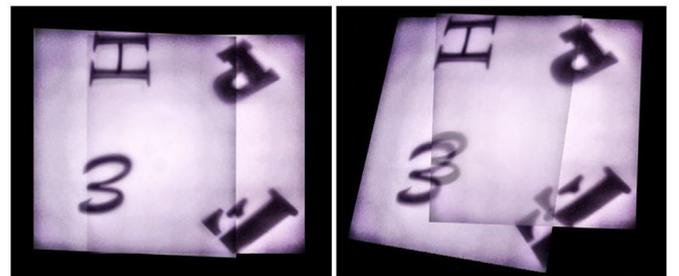


Fig. 8. Result of registration with and without rectification method. Images are captured through simulation experiment of Micro-Ball which is explained in section 5.

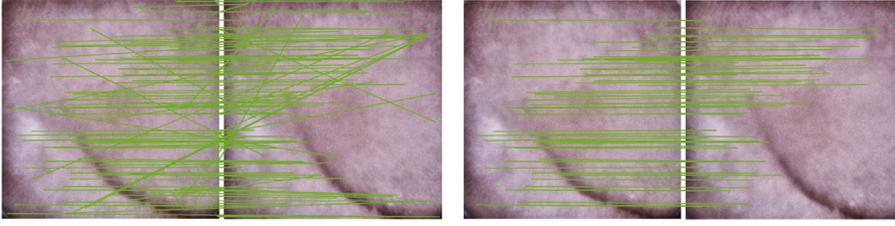


Fig. 9. Matches found without and with applying external outlier removal.

matches lead to further incorrect geometric correction. Therefore, a reliable outlier removal procedure is needed to ensure correct registration. As an example, Fig. 9 shows corresponding keypoints between an image pair which are extracted and matched by PCA-SIFT algorithm and then mismatches are removed by an outlier removal.

Even small estimation errors can reduce the efficiency of the selection method drastically. In practice, the optimal set of features is usually unknown, and it is common to have irrelevant or redundant features at the beginning of process.

Sum of squared differences between corresponding keypoints is used to measure the strength of candidate matches. Because the matching process is only based on proximity and similarity, mismatches often occur. The threshold for match acceptance is deliberately conservative at this stage to minimise incorrect matches. Consequently robust methods must also be adopted, which can provide a good estimation of the solution even if some of the data are mismatches (outliers). Correct matches will obey the epipolar geometry. The aim then is to obtain a set of inliers consistent with the epipolar geometry using a robust technique. In this case outliers are putative matches inconsistent with the epipolar geometry.

The first criterion is $T_{ih} = d_1/d_2$, an indicator of the ambiguity of each matching; where d_1 is the Euclidean distance in feature space between a keypoint in the input image and its nearest matched keypoint in the other image and d_2 is the Euclidean distance to the second nearest keypoint. If T_{ih} is close to one, it means d_1 is close to d_2 . That is to say, for a certain keypoint in the input image, algorithm has detected two possible matching keypoints in the reference image. This is an ambiguous situation, and the matched pair will be deleted if $T_{ih} > 0.80$. The second criterion is a spatial constraint based on an external algorithm. Following algorithms are examined one by one as second step to eliminate falsely matched keypoints using the spatial relationship: HT [31], RANSAC [32], MSAC [33], and MLESAC [10].

Consequently, performance of three different scale invariant keypoints detectors (SIFT, PCA-SIFT, and SIFT + GC) accompanied with four different outlier removal algorithms (HT, RANSAC, MSAC, and MLESAC) are examined one by one, i.e. totally 12 different cases. The results show that a combination of PCA-SIFT and MLESAC has the best performance among others for the application. In Section 5, the experimental results are shown.

Based on the experimental results, the most reliable keypoints that have been extracted by PCA-SIFT-MLESAC are used to form the initial transform function T_μ so as to be used in the next step of the proposed method for further improvement.

4.5. Mutual information

There are two main characteristics that distinguish MI from other dependency measures: *First*, the capacity of measuring any kind of relationship between variables; *Second*, its invariance under space transformations. MI between two images takes its maximum value when images are absolutely dependent. Registration is assumed to correspond to maximising MI: the images have to be aligned in such a manner that the amount of information they contain about each other is maximal.

4.5.1. MI formulation

Giving two images to register, $f_A(x, y)$ and $f_B(x, y)$, defined on a

continuous domain V^c , the coordinates (x_i, y_i) are samples of V^c , and the discrete set of these samples is called V . Let L_A and L_B be discrete sets of intensities of two images, and $T(\mu)$ be the geometric transformation with associated parameters μ . MI is

$$MI(\mu) = \sum_{a \in L_A} \sum_{b \in L_B} p(a, b; \mu) \cdot \log_2 \left(\frac{p(a, b; \mu)}{p_A(a; \mu) p_B(b; \mu)} \right) \quad (8)$$

where $p_A(a; \mu)$ and $p_B(b; \mu)$ are the marginal probability distribution, $p(a, b; \mu)$ is the joint probability distribution, $a \in L_A$, and $b \in L_B$.

While MI measures the degree of dependency of $f_A(x, y)$ and $f_B(x, y)$ by measuring the distance between the joint distribution $p(a, b; \mu)$ and the distribution associated to the case of complete independence $p_A(a; \mu) \cdot p_B(b; \mu)$ through KullbackLeibler measure Vajda [34]. These distributions can be obtained by

$$p(a, b; \mu) = \frac{h(a, b; \mu)}{\sum_{a \in L_A} \sum_{b \in L_B} h(a, b; \mu)} \quad (9)$$

$$p(a; \mu) = \sum_{a \in L_A} p(a, b; \mu) \text{ and } p(b; \mu) = \sum_{b \in L_B} p(a, b; \mu) \quad (10)$$

where h is the joint discrete Parzen histogram given by (11), the Parzen histogram is built based on the partial overlapping domain of $f_A(x, y)$ and $f_B(T_\mu(x, y))$, namely, the grey values (a, b) of those pairs of pixels which lie in the same position.

$$h(a, b; \mu) = \frac{1}{\varepsilon_A \varepsilon_B} \sum_{(x_i, y_i) \in V} \omega \left(\frac{a/\varepsilon_A - f_A(x_i, y_i)/\varepsilon_A}{\varepsilon_A} \right) \cdot \omega \left(\frac{b/\varepsilon_B - f_B(T_\mu(x_i, y_i))/\varepsilon_B}{\varepsilon_B} \right) \quad (11)$$

where ε_A and ε_B are two positive scaling factors related to $\text{card}(L_A)$ and $\text{card}(L_B)$, respectively, in order to control the width of the Parzen window ω . In this work, B-spline functions $\beta^n(x)$ are used as the Parzen window. A B-spline of order n can be generated by convolving the B-spline of order 0 with itself $(n + 1)$ times. This is a piece-wise polynomial of integer degree $n \geq 0$ that can be recursively defined by the following convolution Thévenaz and Unser [35].

$$\beta^n(x) = (\beta^{n-1} * \beta^0)(x) = \int_{-\infty}^{\infty} \beta^{n-1}(x) \beta^0(x - t) dt, \quad n > 0 \quad (12)$$

where β^0 is a unit square pulse

$$\beta^0(x) = \begin{cases} 1 & ; \quad x \in [-0.5, 0.5] \\ 0 & ; \quad x \notin [-0.5, 0.5] \end{cases} \quad (13)$$

The cubic B-spline function β^3 is selected as model kernel since it could provide a high-quality model.

Moreover, as MI is sensitive to the amount of overlap between the images, Normalised Mutual Information NMI (I') in (14) is used to overcome this problem [24]. Here, any change in uncertainty of the image values and, therefore, the marginal entropies, will not result in a change in the alignment measure. Maximisation of I' seeks a transformation where joint entropy is minimised with respect to the marginal entropies.

$$I'(\mu) = 1 + \frac{MI(\mu)}{E(a, b)} \quad (14)$$

where $E(a, b)$ is the joint entropy given in

$$E(a, b) = - \sum_{a \in L_A} \sum_{b \in L_B} p(a, b; \mu) \log p(a, b; \mu). \quad (15)$$

4.5.2. Multiscale framework

A multiscale framework is a solution to solve the problem of high computation cost. The task is first solved at a coarse scale. Then, the results are propagated to the next finer level and used as a starting guess for solving the task at that level. This procedure is iterated until the finest level is reached. An image pyramid is made, a set of gradually reduced versions of the original image. Based on an image of size $M \times N$, a set of sequence of images are created with different sizes $\{(M_1 \times N_1), \dots, (M_i \times N_i), \dots, (M \times N)\}$ where M_i and N_i are obtained from $M_i = M_{i+1}/2$ and $N_i = N_{i+1}/2$. It is very important that the initial condition for last level is the best possible in order to reduce the amount of refinement necessary to reach convergence. To get optimal starting conditions, it is crucial that the coarse levels of the pyramid most represent the finest level. For all cases, NMI between the whole overlap of sub-band images of two images is computed at each level and maximised successively, and the search is performed on an interval around the optimal transformation parameters found at the previous level and is refined at the next level. Thus, the accuracy of the search increases from coarse resolution to fine resolution. At each level, the iteration stops when the maximum number of 110 iterations is achieved, which are sufficient for the search to converge to a set of registration parameters with a sub-pixel accuracy based on our experiments.

4.5.3. Search strategy

For the multiscale framework, a suitable optimiser is desired that converges in a few criterion evaluations when initialised with good starting conditions. The Levenberg-Marquardt (LM) algorithm [36] is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions. It has become a standard technique for non-linear least-squares problems, widely adopted in a broad spectrum of disciplines. LM can be thought of as a combination of steepest descent and the Gauss-Newton method. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a Gauss-Newton method.

The steepest-gradient descent is a minimisation algorithm that can be succinctly described in (16). Its local convergence is guaranteed, although it may be very slow.

$$\mu^{(k+1)} = \mu^{(k)} - \Gamma \nabla I'(\mu^{(k)}) \quad (16)$$

where $\nabla I'$ is the gradient of I' . A key problem is determination of the appropriate scaling diagonal matrix Γ . The Newton method is described by

$$\mu^{(k+1)} = \mu^{(k)} - (\nabla^2 I'(\mu^{(k)}))^{-1} \nabla I'(\mu^{(k)}) \quad (17)$$

where $\nabla^2 I'$ is the Hessian of I' , defined as the matrix of the second derivative of I' . As its convergence to an optimum is not guaranteed, it may converge to a saddle point. Even worse, it diverges from the desired solution when the problem is not convex.

A modification of Marquardt-Levenberg proposed by Thevenaz and Unser [25] exhibits super-linear convergence when close enough to the optimum. This specifically designed optimiser introduce a modified Hessian HI' (18) in which we retain the off-diagonal entries of $\nabla^2 I'$ and multiply its diagonal entries by some factor.

$$[HI'(\mu)]_{i,j} = [\nabla^2 I'(\mu)]_{i,j} \cdot (1 + \delta_{i,j} \lambda) \quad (18)$$

where $\delta_{i,j}$ is the Kronecker symbol, and where λ is a tuning factor that represents the compromise between the gradient method and the Newton method. Therefore, the new update is

$$\mu^{(k+1)} = \mu^{(k)} - (HI'(\mu^{(k)}))^{-1} \nabla I'(\mu^{(k)}) \quad (19)$$

In fact, search for a maximum of a function is nothing else than the search for the zeroes of the gradient of that function. Thus, above-mentioned steps are used to find the optimal parameters μ to maximise I' between an image and its transformed image (obtained by the transform function T_μ).

4.6. Map construction

Once the transformation is applied to one of the images to take it to the co-ordinate system of the other image, then two images are aligned and stitched together to gradually construct the map. However, visible seams (due to exposure differences, misregistration, or moving objects) can occur during stitching process. In order to reduce such artefacts and achieve a smooth transition across any image pair, there should be a mechanism to decide which pixels to use and how to weight or blend them. In this work, in order to make the transition across images smooth, a simple but effective algorithm proposed in Farbman et al. [37] is used. In order to blend source image to the target image, let c_1, c_2, \dots, c_n be the n points on the seam, D_1, D_2, \dots, D_n be the intensity differences at those points between the images, and b be a pixel of the source image. Then the intensity differences at pixel b is computed as

$$D(b) = \sum_{i=1}^n w_i(b) D(c_i) \quad (20)$$

where the weights $w_i(b)$ are

$$w_i(b) = \frac{1/\|c_i - b\|}{\sum_{j=1}^n 1/\|c_j - b\|}. \quad (21)$$

5. Results and discussion

In this section, results from experimental analysis are given to evaluate the overall performance of the proposed algorithm. Implementations have been carried out using MATLAB 8.3.0.532 on an Intel(R) Xeron(R) CPU E5-1620 0 @ 3.60 GHz with 32 GB of RAM. Two datasets are considered.

The first dataset concerns normal endoscopic images of the internal wall of the GI tract for two different subjects obtained by PillCam Capsule Endoscopy. The dataset of two different subjects totally contains almost 89,800 colour images of size 256×256 in JPEG format.

The second dataset is based on the scenario which considers the simulation of the Micro-Ball performance, since the device is currently available as a prototype with a diameter of 4 cm. The dataset in this scenario consists of 152 images of size 480×480 (in modified JPEG format with PSNR 40.7 dB and compression rate of 86%) obtained from an experiment of moving the Micro-Ball through a cylinder. Although simulation of the surface deformation of GI tract can be neglected in this experiment (due to highly overlapped images and slow movement of the camera Miranda-Luna et al. [17] surface deformation of GI tract and its impact on image stitching is considered in this work. The Micro-Ball is placed into a transparent, soft and flexible plastic pipe (see Fig. 10) and images are captured while the Micro-Ball is moving and rotating through the slow-moving pipe. The outer wall of the pipe is covered with a reference picture printed on a white leather-like material.

Similarly to the approach by Froehlich et al. [38]; observation of image sequences of the two subjects, are used to estimate the actual movement of the Micro-Ball inside the small bowel. The rotation occurrence frequency is approximately deduced from the perspective changes of the image sequences. The simulated movement is obtained by altering the translation frequency, average translation distance, and rotation occurrence frequency to adjust the prototype (40 mm) with the actual size (15 mm). Translation frequency, average translation

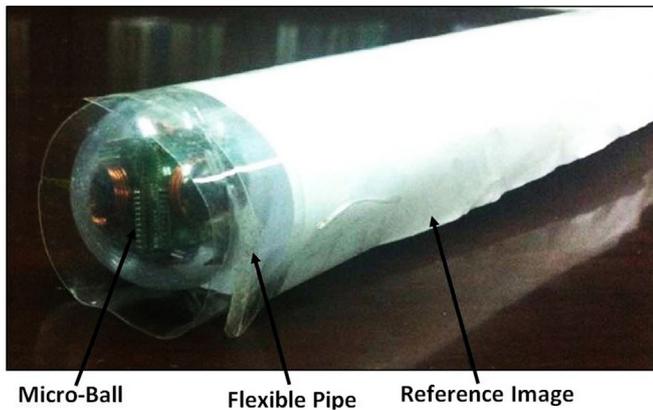


Fig. 10. Experiment of moving the Micro-Ball through the cylinder.

distance, and rotation occurrence frequency for real situation and simulation are 10 times/min, 3 mm, 1 time/min and 10 times/min, 8 mm, 1 time/min, respectively.

In the following subsections, performance of the first two steps of the algorithm, namely keypoints detection and outlier removal are examined, then, the accuracy of the proposed algorithm is verified quantitatively and visually to show how the algorithm would help to reach the goal of constructing a 2D reconstruction of the small intestinal tract. All results are based on the first dataset except Fig. 13. To quantitatively evaluate the registration performance, a ground truth is required to compare with the obtained registration parameters. Therefore, 2500 different pairs of images are generated by cropping and manipulating (rotating, scaling, shifting, shearing and projecting) of the real endoscopic images (from the first dataset) with different overlap ratios. The reported results in Tables 1 and 2 are based on this dataset.

5.1. Comparison of scale invariant keypoints detectors

Outlier removal methods are set to eliminate false matches while preserving the correct ones. To select the optimum method, Correct Rate (CR) which is defined as follow is used, similar to Mikolajczyk and Schmid [22].

$$CR (\%) = ((c - f)/a) \times 100 \tag{22}$$

where c , f , and a are the number of remaining correct matches, remaining false matches, and all correct matches, respectively.

From the experimental results shown in Table 1, it follows: 1) MLESAC is more reliable than other methods (HT, RANSAC, and MSAC) for different keypoint detectors being capable of removing most of the false matches on the tested datasets. As an illustrative example, correct and false matches are shown in Fig. 9. 2) PCA-SIFT performs better than

Table 1

Comparison of scale-invariant keypoints detectors in terms of Correct Rate (CR) using outlier removal algorithms, before and after enhancement.

Keypoints Detection		CR(%) (before)	CR(%) (after)
SIFT	HT	75.12	76.98
	RANSAC	79.88	80.17
	MSAC	84.67	85.97
	MLESAC	88.01	88.99
PCA-SIFT	HT	79.80	81.76
	RANSAC	83.87	88.45
	MSAC	89.97	92.25
	MLESAC	93.74	95.02
SIFT + GC	HT	76.12	78.31
	RANSAC	79.75	81.09
	MSAC	83.54	86.77
	MLESAC	86.92	89.91

Table 2

Mean Euclidean Distance (MED) and computation time of registration by running PCA-SIFT, NMI, and the proposed method on 2500 different pairs of images.

Measure	Method	Mean	SD
MED (pixel)	PCA-SIFT	3.12	0.98
	NMI	2.68	0.71
	Proposed	0.93	0.33
	PCA-SIFT	6.87	1.68
TIME (second)	NMI	10.54	2.10
	Proposed	8.98	0.98

SIFT and SIFT + GC for the intended application. Thus, the combination of MLESAC and PCA-SIFT is a promising approach to detect the set of most reliable scale-invariant keypoints as the input that maximises the NMI. 3) Observing the CR results in Table 1, the correct rate has expectedly increased (3%–5%) for all cases after applying the image enhancement method.

5.2. Image registration accuracy

Registration accuracy criterion is computed as the Mean Euclidean Distance (MED) between homologous pixels of the target images and the registered images. The computation time (seconds) and registration accuracy in terms of the Mean and Standard Deviation (SD) for 2500 image pairs with different methods are shown in Table 2.

It can be seen from the results in Table 2 that the proposed algorithm shows a better performance compared with the results of either PCA-SIFT or NMI. Although the computation time has increased about two second on average, the accuracy has improved significantly compared to that of PCA-SIFT. Analogously, comparing with NMI, the proposed method shows a very good improvement in terms of time and accuracy.

5.3. 2D map construction

The proposed algorithm was applied on sets of consecutive images obtained by the conventional wireless capsule endoscopy. In Fig. 11, five segments of 2D representation of the stitched consecutive images are shown. Many cases are examined to generate a wider map; however, many regions are not covered by the cameras, only segments of the GI tract are able to reconstruct a map, instead of one image of whole intestinal tract. This is because the wireless capsule endoscopy devices are equipped with 1 or 2 cameras, they can never provide enough information to enable reconstruction of a full view of the intestinal tract. The 2D reconstruction method relies on the overlapping areas of paired images and coverage rate of intestine's interior wall, the traditional wireless capsule devices could not provide enough information for the reconstruction of the whole intestinal tract due to their limited visual field and the negative effect of the random movement of the capsule. However, the Micro-Ball can achieve nearly 100% coverage with its multiple cameras equipped with a smart imaging system.

The abnormal region was found in Fig. 11 which reveals an arteriovenous malformation with active bleeding, confirmed by physicians, indicates the proposed method performed well to show surroundings of an abnormal region for more efficient diagnosis.

The proposed algorithm is not able to create one continues map representing the whole area of GI tract as explained above, however, registration of any successive images that contain similar information is led to 64.89% reduction in total number of images in the dataset. Since there is no control on movement of wireless capsule endoscopy, its random movement causes producing a lot of redundant data when it stops moving, and still captures images from the same region repeatedly. The proposed algorithm by registering any similar adjacent images considerably reduces the number of images. For instance, Fig. 12, shows two sets of successive images of almost the same region



Fig. 11. Five different regions of GI tract are reconstructed by successful registration of 102, 78, 66, 86, and 98 successive images (top to bottom).

of the small intestinal wall. Running the proposed algorithm on these two sets, the output will be one image that has all the information from the 40 images, as shown in Fig. 12.

Running the proposed method with the same implementation on the second dataset which is taken from moving the Micro-Ball through the cylinder, as shown in Fig. 10, proves that the proposed method together with the Micro-Ball have been successful to make the entire field of view possible in 30.75 min with MED error accumulation of 16.59 pixels and without affecting the visual registration quality, compared with the reference image, as shown in Fig. 13. Using the Micro-Ball technology together with the proposed method, a map was produced to represent the whole information of the dataset. It provides a unique wide area visualisation on which the observation and examination can be performed in an efficient way. Besides, from the clinical point of view, it is valuable to observe at once both the region of interest and its surrounding area, without passing from one image to another.

The importance of this work included but not limited to the

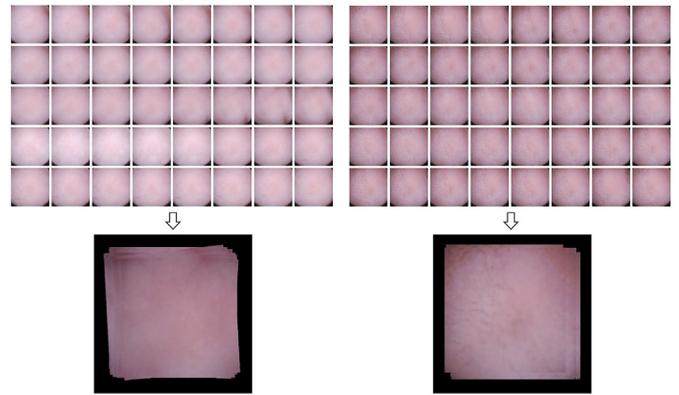


Fig. 12. Two sample sets, each set with 40 consecutive images captured two different regions of intestinal wall. Images are registered by the proposed method, each image in second row represents the whole information of each image set in first row.

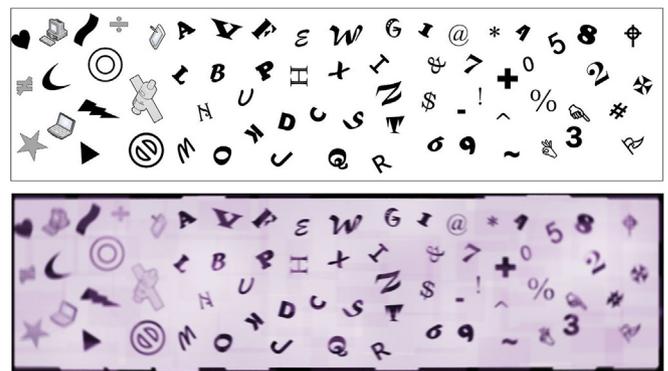


Fig. 13. The reference picture printed on a white leather-like material which covers the whole outer wall of the cylinder (top) and reconstructed map from 152 smaller images captured by Micro-Ball (bottom).

following items: First, this work proposes an efficient automatic algorithm that employs the robustness of feature-based methods and the accuracy of intensity-based methods to construct a hybrid scheme for endoscopic image registration. The results show that this arrangement of two different groups of image registration algorithm has been successful. Current version of the method works the best but someone may get better results in the future by changing the method in initialisation step and/or fine-tuning step. The structure of the algorithm presented here is in a way that initialisation step can be done with any feature-based method and then an intensity-based method is used to tune the transformation function. Second, the proposed registration strategy, besides common wireless endoscopic images, was tested on a unique database which has been obtained from *in vitro* experiments with the Micro-Ball demo system. The algorithm has been successful to construct the whole inner wall of the *in vitro* experiment and also many regions of the GI tract using real endoscopic images. The construction of the whole tract was not possible with real endoscopic images not because of the algorithm but due to limitations of the endoscopy devices i.e. high data missing rate, no location information, no smart imaging strategy, etc. These results are of significance for the challenging task of endoscopic image registration for different applications. Third, it should be noted that the aim of this work is not only to produce some improvement in the registration algorithm but the whole workflow is of high importance in the wireless endoscopy field. In fact, using the capability of the new developed smart Micro-Ball to construct the whole inner wall is also a promising result for the future of wireless endoscopy devices with multiple cameras.

6. Conclusion and future work

A new fully automatic image registration approach was proposed in this paper for reconstruction of a 2D representation of the small intestine's interior wall. The proposed method was based on maximisation of normalised mutual information between any two consecutive images. Aiming at this, the initial solution for search strategy was obtained by the PCA-SIFT algorithm which was equipped with MLESAC outlier removal. Then, in a multiscale framework, the algorithm attempted to find the best transformation parameters for registration and gradually formed the whole map. The method was examined on two different datasets and the registration accuracy and computation time were very satisfying.

For the future work, we plan to further reduce the computation time, by searching the most salient initial solution at the first stage without outlier removal algorithms and by adding high performance computing hardware such as GPU. Moreover, as the GI tract wall is a soft live object and its texture changes slightly during the examination, it is also important to consider the image deformation issue in endoscopic image registration. Another objective is to extend the methodology to process not every two images at a time but several images which may lead to higher quality reconstructed map.

Acknowledgement

This work was supported by: EU FP7 projects EYE2E (Grant No. 269118) and LIVCODE (Grant No. 295151); EU Horizon 2020 projects ENRICHME (Grant No. 643691) and STEP2DYNA (Grant No. 691154).

References

- [1] C. Van de Bruaene, D. De Looze, P. Hindryckx, Small bowel capsule endoscopy: where are we after almost 15 years of use? *World J. Gastrointest. Endosc.* 7 (1) (2015) 13–36.
- [2] A.C. Meltzer, S. Bashir, Promise and potential of video capsule endoscopy in the emergency department, *Tech. Gastrointest. Endosc.* 17 (1) (2015) 2–4.
- [3] T. Weibel, C. Daul, D. Wolf, R. Rösch, A. Ben-Hamadou, Endoscopic bladder image registration using sparse graph cuts, 2010 IEEE International Conference on Image Processing, IEEE, 2010, pp. 157–160.
- [4] M.-H. Meng, T. Mei, J. Pu, C. Hu, X. Wang, Y. Chan, Wireless robotic capsule endoscopy: state-of-the-art and challenges, *Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on*, vol. 6, IEEE, 2004, pp. 5561–5555a.
- [5] Y. Gu, X. Xie, G. Li, T. Sun, D. Wang, Z. Yin, P. Zhang, Z. Wang, Design of endoscopic capsule with multiple cameras, *IEEE transactions on biomedical circuits and systems* 9 (4) (2015) 590–602.
- [6] D. Wang, X. Xie, G. Li, Y. Gu, Z. Yin, Z. Wang, Research on 2d representation method of wireless micro-ball endoscopic images, 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 1145–1148.
- [7] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.
- [8] A. Attar, A. Shahbahrani, R.M. Rad, Image Quality Assessment Using Edge Based Features. *Multimedia Tools and Applications*, (2015), pp. 1–16.
- [9] Y. Ke, R. Sukthankar, Pca-sift: a more distinctive representation for local image descriptors, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Volume 2, Pages II-506, IEEE, 2004.
- [10] P. Torr, A. Zisserman, Robust computation and parametrization of multiple view relations, *Computer Vision, 1998. Sixth International Conference on*, IEEE, 1998, pp. 727–732.
- [11] B. Li, M.Q.-H. Meng, Wireless capsule endoscopy images enhancement using contrast driven forward and backward anisotropic diffusion, 2007 IEEE International Conference on Image Processing, Volume 2, Pages II-437, IEEE, 2007.
- [12] B. Li, M.Q.-H. Meng, Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments, *Comput. Biol. Med.* 39 (2) (2009) 141–147.
- [13] T. Bergen, T. Wittenberg, Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods, *IEEE J. Biomedical and Health Informatics* 20 (1) (2016) 304–321.
- [14] A. Karargyris, N. Bourbakis, A video-frame based registration using segmentation and graph connectivity for wireless capsule endoscopy, *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH, IEEE, 2009*, pp. 74–79.
- [15] H.-G. Lee, M.-K. Choi, B.-S. Shin, S.-C. Lee, Reducing redundancy in wireless capsule endoscopy videos, *Comput. Biol. Med.* 43 (6) (2013) 670–682.
- [16] D.K. Iakovidis, S. Tsevas, A. Polydorou, Reduction of capsule endoscopy reading times by unsupervised image mining, *Comput. Med. Imag. Graph.* 34 (6) (2010) 471–478.
- [17] R. Miranda-Luna, C. Daul, W.C. Blondel, Y. Hernandez-Mier, D. Wolf, F. Guillemin, Mosaicing of bladder endoscopic image sequences: distortion calibration and registration algorithm, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 55 (2) (2008) 541–553.
- [18] E. Spyrou, D. Diamantis, D.K. Iakovidis, Panoramic visual summaries for efficient reading of capsule endoscopy videos, *Semantic and Social Media Adaptation and Personalization (SMAP), 2013 8th International Workshop on*, IEEE, 2013, pp. 41–46.
- [19] A. Karargyris, N. Bourbakis, Three-dimensional reconstruction of the digestive wall in capsule endoscopy videos using elastic video interpolation, *IEEE Trans. Med. Imag.* 30 (4) (2011) 957–971.
- [20] P.M. Szczypiński, R.D. Sriram, P.V. Sriram, D.N. Reddy, A model of deformable rings for interpretation of wireless capsule endoscopic videos, *Med. Image Anal.* 13 (2) (2009) 312–324.
- [21] J. Bauer, N. Sünderhauf, P. Protzel, Comparing several implementations of two recently published feature detectors, *IFAC Proceedings* 40 (15) (2007) 143–148.
- [22] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (10) (2005) 1615–1630.
- [23] R. Shams, P. Sadeghi, R.A. Kennedy, R.I. Hartley, A survey of medical image registration on multicore and the gpu, *IEEE Signal Process. Mag.* 27 (2) (2010) 50–60.
- [24] J.P. Plum, J.A. Maintz, M.A. Viergever, Image registration by maximization of combined mutual information and gradient information, *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, 2000, pp. 452–461.
- [25] P. Thevenaz, M.A. Unser, Spline pyramids for intermodal image registration using mutual information, *Optical Science, Engineering and Instrumentation'97, International Society for Optics and Photonics, 1997*, pp. 236–247.
- [26] J.A. Maintz, M.A. Viergever, A survey of medical image registration, *Med. Image Anal.* 2 (1) (1998) 1–36.
- [27] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [28] A. Attar, X. Xie, C. Zhang, Z. Wang, S. Yue, Wireless micro-ball endoscopic image enhancement using histogram information, 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 3337–3340.
- [29] J. Hu, X. Xie, G. Li, Y. Gu, Y. Zeng, Z. Wang, Attitude sensing system design for wireless micro-ball endoscopy, 2011 IEEE International Symposium of Circuits and Systems (ISCAS), IEEE, 2011, pp. 2357–2360.
- [30] E.N. Mortensen, H. Deng, L. Shapiro, A sift descriptor with global context, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, IEEE, 2005, pp. 184–190.
- [31] V.C. H. P. (1962). Method and means for recognizing complex patterns. *US Patent 3,069,654*.
- [32] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* 24 (6) (1981) 381–395.
- [33] P.H. Torr, A. Zisserman, Mlesac: a new robust estimator with application to estimating image geometry, *Comput. Vis. Image Understand.* 78 (1) (2000) 138–156.
- [34] I. Vajda, *Theory of Statistical Inference and Information* vol. 11, Kluwer Academic Pub, 1989.
- [35] P. Thévenaz, M. Unser, Optimization of mutual information for multiresolution image registration, *IEEE Trans. Image Process.* 9 (12) (2000) 2083–2099.
- [36] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.* 11 (2) (1963) 431–441.
- [37] Z. Farbman, G. Hoffer, Y. Lipman, D. Cohen-Or, D. Lischinski, Coordinates for instant image cloning, *ACM Transactions on Graphics (TOG)*, vol. 28, ACM, 2009, p. 67.
- [38] J.M. Froehlich, M.A. Patak, C. von Weymarn, C.F. Juli, C.L. Zollkofer, K.-U. Wentz, Small bowel motility assessment with magnetic resonance imaging, *J. Magn. Reson. Imag.* 21 (4) (2005) 370–375.