Artificial Intelligence in Medicine xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

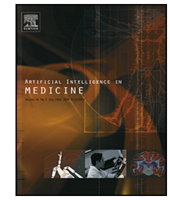


Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes

Q1 Charreau S. Bell[a,*], Keith L. Obstein[a,b], Pietro Valdastri[a,b]

[a] Department of Mechanical Engineering, Vanderbilt University, 2301 Vanderbilt Place, PMB 351826, Nashville, TN 37235-1826, USA
[b] Department of Gastroenterology, Hepatology and Nutrition, Vanderbilt University Medical Center, 1030C MRB IV, 2215 Garland Avenue, Nashville, TN 37232-0252, USA

ARTICLE INFO

Article history:
Received 1 November 2012
Received in revised form
20 September 2013
Accepted 23 September 2013

Keywords:
Artificial neural networks
Pose estimation
Localization
Optical flow
Visual odometry
Closed-loop control
Teleoperation
Narrow band imaging
Flexible endoscopes
Colonoscopes
Gastrointestinal endoscopy

ABSTRACT

Objective: Colorectal cancer is one of the leading causes of cancer-related deaths in the world, although it can be effectively treated if detected early. Teleoperated flexible endoscopes are an emerging technology to ease patient apprehension about the procedure, and subsequently increase compliance. Essential to teleoperation is robust feedback reflecting the change in pose (i.e., position and orientation) of the tip of the endoscope. The goal of this study is to first describe a novel image-based tracking system for teleoperated flexible endoscopes, and subsequently determine its viability in a clinical setting. The proposed approach leverages artificial neural networks (ANNs) to learn the mapping that links the optical flow between two sequential images to the change in the pose of the camera. Secondly, the study investigates for the first time how narrow band illumination (NBI) – today available in commercial gastrointestinal endoscopes – can be applied to enhance feature extraction, and quantify the effect of NBI and white light illumination (WLI), as well as their color information, on the strength of features extracted from the endoscopic camera stream.

Methods and materials: In order to provide the best features for the neural networks to learn the change in pose based on the image stream, we investigated two different imaging modalities – WLI and NBI – and we applied two different spatial partitions – lumen-centered and grid-based – to create descriptors used as input to the ANNs. An experiment was performed to compare the error of these four variations, measured in root mean square error (RMSE) from ground truth given by a robotic arm, to that of a commercial state-of-the-art magnetic tracker. The viability of this technique for a clinical setting was then tested using the four ANN variations, a magnetic tracker, and a commercial colonoscope. The trial was performed by an expert endoscopist (>2000 lifetime procedures) on a colonoscopy training model with porcine blood, and the RMSE of the ANN output was calculated with respect to the magnetic tracker readings. Using the image stream obtained from the commercial endoscope, the strength of features extracted was evaluated.

Results: In the first experiment, the best ANNs resulted from grid-based partitioning under WLI (2.42 mm RMSE) for position, and from lumen-centered partitioning under NBI (1.69° RMSE) for rotation. By comparison, the performance of the tracker was 2.49 mm RMSE in position and 0.89° RMSE in rotation. The trial with the commercial endoscope indicated that lumen-centered partitioning was the best overall, while NBI outperformed WLI in terms of illumination modality. The performance of lumen-centered partitioning with NBI was 1.03 ± 0.8 mm RMSE in positional degrees of freedom (DOF), and 1.26 ± 0.98° RMSE in rotational DOF, while with WLI, the performance was 1.56 ± 1.15 mm RMSE in positional DOF and 2.45 ± 1.90° RMSE in rotational DOF. Finally, the features extracted under NBI were found to be twice as strong as those extracted under WLI, but no significance in feature strengths was observed between a grayscale version of the image, and the red, blue, and green color channels.

Conclusions: This work demonstrates that both WLI and NBI, combined with feature partitioning based on the anatomy of the colon, provide valid mechanisms for endoscopic camera pose estimation via image stream. Illumination provided by WLI and NBI produce ANNs with similar performance which are comparable to that of a state-of-the-art magnetic tracker. However, NBI produces features that are stronger than WLI, which enables more robust feature tracking, and better performance of the ANN in terms of accuracy. Thus, NBI with lumen-centered partitioning resulted the best approach among the different

* Corresponding author. Tel.: +1 615 3220688.
E-mail addresses: charreau.s.bell@vanderbilt.edu (C.S. Bell), keith.l.obstein@vanderbilt.edu (K.L. Obstein), pietro.valdastri@vanderbilt.edu (P. Valdastri).

0933-3657/$ – see front matter
http://dx.doi.org/10.1016/j.artmed.2013.09.002

variations tested for vision-based pose estimation. The proposed approach takes advantage of components already available in commercial gastrointestinal endoscopes to provide accurate feedback about the motion of the tip of the endoscope. This solution may serve as an enabling technology for closed-loop control of teleoperated flexible endoscopes.

1. Introduction

Each year, colorectal cancer claims the lives of more than 600,000 people worldwide, and is the fourth leading cause of cancer-related death in the worldy [1]. Colorectal cancer commonly progresses to malignancy in approximately 5–10 years. However, this type of cancer has the unique quality that if the tumor is detected at an early enough stage, the prognosis for survival is 90%, whereas if detected too late, it decreases to 5% [2]. This emphasizes the relevance of timely screening for population at risk (i.e., people over 50 years of age or having family history of colorectal cancer), even in the case that no symptoms are observed.

The most commonly used method for diagnostic and therapeutic assessment of colorectal cancer is through colonoscopy, an endoscopic procedure which requires the insertion of a 1.5-m long flexible tube through the anus. The endoscope provides illumination for visualization of the colon lumen, which enables detection and removal of polyps. Standard colonoscopy is performed under white light illumination (WLI). However, this approach can fail to reveal important information [3]. Even experienced endoscopists can miss up to 30% of all potential cancer lesions when using standard WLI [4].

In the last decade, narrow band imaging (NBI) has been introduced to improve diagnosis. NBI uses filters to narrow projected light to blue (415 nm) and green (540 nm) wavelengths to generate a colored image. Blue-green light enhances superficial mucosal capillaries and mucosal surface patterns; greater absorption of illuminating bands by hemoglobin causes the blood vessels to look darker. Despite recent literature demonstrating that NBI does not increase the diagnostic yield when compared to WLI [5], this imaging modality is today increasingly common in commercial colonoscopes (e.g., H180AL/I, Olympus, Japan).

Although a colonoscopy usually takes less than 30 min and is performed in outpatient surgery under sedation, patient compliance with recommended screening is low (i.e., 1 in 3 adults are not being screened [6]) due to the preparation required, fear of pain during the procedure, and perceived embarrassment. The main technological improvements in the field of flexible endoscopy aim to help patients to overcome these hindrances.

An approach for accomplishing this goal is through the development of increasingly flexible endoscopes, wireless capsule endoscopy (WCE), and virtual colonoscopy [2]. Complementary to these advances is the emergence of computer-assisted technologies to aid the doctor, whose purpose is to increase detection of malignancies and control over the intended trajectory of the endoscope. Robotics is playing an increasingly important role in this field with the development of fully- or semi-automated endoscopic systems [2,7–11]. This technological breakthrough has the potential to widen the implementation of colorectal cancer screening and surveillance programs to rural areas, to mobile camps, or to in-field military bases, and the physical presence of an expert endoscopist may no longer be required.

Real-time pose (i.e., position and orientation) detection of the tip of an automated flexible endoscope is crucial to achieving reliable and effective teleoperation. These devices operate in an intricate, complex environment and by definition are compliant; many variables exist that cannot be accounted for in a model, which severely limits the efficacy of open-loop control. Furthermore, medical procedures require a high degree of precision and accuracy; implementing real-time pose detection allows for calculated, controlled movements which enhance system stability [12]. In particular, the real-time estimated pose of the endoscope head can be used as feedback signal for a closed-loop control strategy, as represented in Fig. 1. This allows us to minimize the error between the intended pose (i.e., where the user wants the endoscope to move and orient the camera), and the reached pose (i.e., the measured pose of the endoscope tip).

In order to achieve real-time pose detection, magnetic tracking has emerged as a reliable method and there are several commercial manufacturers of 5 or 6 degree of freedom (DOF) trackers [13,14]. Magnetic trackers placed along the entire length of the colonoscope, such as in the commercially available ScopeGuide® (Olympus, Japan), have been used to provide the endoscopist visual feedback of the instrument pose with respect to a global coordinate frame [15]. Within bronchoscopy, the endoscopic camera stream has been used in conjunction with image registration and fluoroscopy for tracking of the endoscope [16–19].

However, magnetic trackers require additional space in the endoscope; this results in an increase in the size of the device, and a corresponding reduction in the flexibility of the endoscope. For endoscopes with extremely limited operating space, such as encephaloscopes, rhinoscopes, and bronchoscopes, minimization of the size of the endoscope is fundamental. Furthermore, commercial players in the field of gastrointestinal endoscopy are proposing platforms that are based on magnetic manipulation of the endoscopic device [20,21]. This promising approach is also being pursued by several research labs worldwide [22–26]. Magnetic trackers interfere with magnetic manipulation due to the presence of metallic components or because the localization principle itself is based on triangulation of electromagnetic fields.

Tracking of the endoscope head is even more crucial in soft body cavities (e.g., colon, small intestine), since image registration is not effective. Thus, a localization system which is independent of the technology platform to which it is applied, provides accurate pose estimation for real-time feedback, and neither creates unwanted disturbance to the system nor adds additional size to the endoscope will be beneficial for enabling closed-loop control of teleoperated flexible endoscopes.

1.1. Related work

The problem of real-time localization and steering of flexible endoscopes has a number of challenges [27,28]. Concerning the use of the image stream to steer the endoscope through the lumen, possible approaches include finding the darkest region of the image for lumen detection [29,30], identifying features such as the ring-like contours surrounding the lumen [31–33], and using highlights resulting from illumination [34]. Several works have used lumen center detection schemes to correct the current heading of the camera towards the lumen center in each control loop [8,35], providing a mechanism for automation. However, these solutions do not measure the change in pose of the endoscope, and thus cannot be used to implement closed-loop control (i.e., although the motors can be actuated towards the center of the lumen, there is no feedback as to whether the actuation was successful).

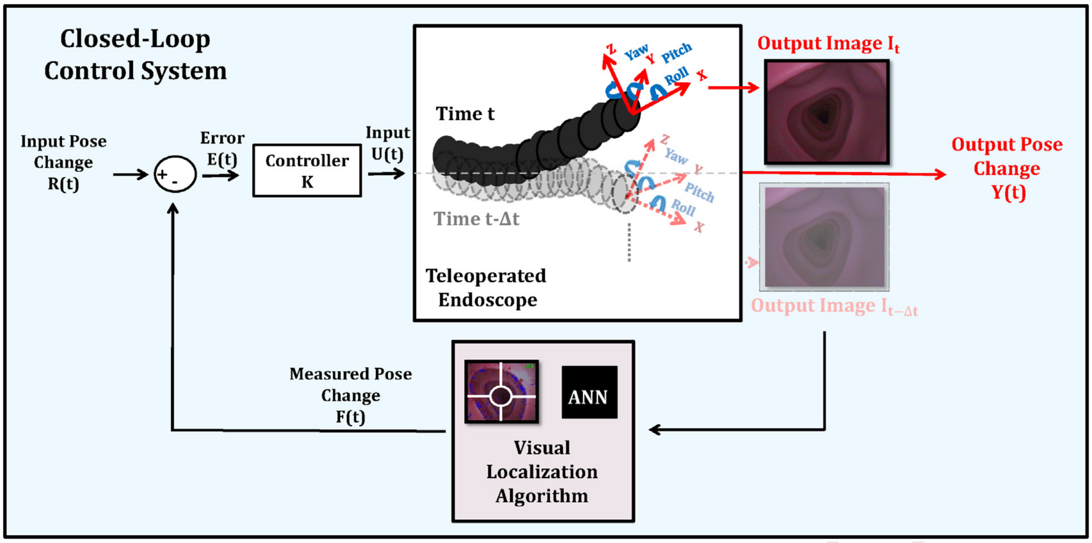


Fig. 1. Closed-loop control system taking advantage of the proposed pose detection approach to guide a teleoperated endoscope.

Localization and tracking based image motion analysis has been quite successful in other fields, including mobile robots, unmanned vehicle navigation [36,37], and egomotion estimation [38]. The most popular techniques include optical flow [39], visual simultaneous localization and mapping [40], and structure from motion (SFM)/stereopsis [41]. These approaches have also been explored in gastrointestinal endoscopy. A 3-dimensional reconstruction of the colon was achieved using SFM reconstruction with the image sequence from a monocular camera [42]. However, this implementation assumes zero rotation for simplification, and is thus able to calculate just the 3 DOF related to camera translation between two images. Furthermore, the SFM algorithm is able to calculate 6 DOF motion relative to the translation along the optical axis; as a consequence, the metric translation along the optical axis (i.e., the longitudinal axis of the lumen) cannot be accurately calculated using this model. To remedy this depth estimation problem, the spherical camera model has been used, although this again requires simplifying assumptions about the rotation of the camera [8]. Focus of expansion has also been used to avoid the numerical instability of optical flow and SFM calculations, and was successfully employed on a virtual colonoscopy and other image sets [43,10]. However, algorithm performance on computer-generated datasets can differ significantly from a colon simulator or human colon [8].

As for artificial intelligence and machine learning, techniques within endoscopy have been mostly limited to signal filtering and facilitation of computer-aided diagnosis (e.g., segmentation, object recognition, etc.) [44,45]. Localization of an endoscopic capsule within general anatomical regions of the gastrointestinal tract was achieved by moving picture expert group (MPEG)-7 features (commonly used in video and audio compression) with pattern recognition classifiers [46]. Rule-based systems using fuzzy logic have also been used for extraction of the lumen [31]. However, the efficacy of these algorithms for teleoperated systems is again limited since they cannot produce an accurate quantitative measurement of pose.

The approach presented in this paper aims to build upon the previous body of work by using machine learning techniques to estimate the pose variation of the endoscope from optical flow-based features. This algorithm estimates the change in pose $F(t)$ caused by the actuation of the device. Fig. 1 illustrates how this quantity is used within a control loop. Using two sequential images from the endoscope at the current time I_t and a previous time $I_{t-\Delta t}$, the algorithm calculates the feedback $F(t)$, which reflects if the actuation of the endoscope has produced the intended pose change $R(t)$. The controller K then works to minimize the potential error $E(t)$. Since the changes in pose are utilized directly within the control loop, no integration is necessary, thus avoiding inaccuracies related to numerical drift. Our approach describes the optical flow of features between sequential images, and then relates this description of feature movements to the achieved pose using artificial neural networks (ANNs). The applicability of the proposed method to clinical use was also tested by using a commercial endoscope operated by an expert endoscopist (>2000 lifetime procedures).

The second contribution of this paper is to quantitatively compare the effect of illumination on the strength of image features extracted. This is done by calculating the eigenvalues of individual pixels, which provides a numerical description of the strength of a feature. By comparing the average maximum eigenvalues obtained from the NBI and WLI image streams obtained using a commercial endoscope, the illumination mechanism most effective in providing better features can be determined.

1.2. Outline

The outline of this paper is as follows: Section 2 describes the extraction and construction of the feature set based on the estimated optical flow between sequential images, and the training and testing of the ANNs used for learning the pose from these extracted features. It furthermore explains the process of validating the algorithm using a commercial endoscope and the method for determining the strength of features extracted using NBI and WLI. Section 3 discusses the impact of the illumination modalities on the performance of the ANNs trained on feature vectors created by the different imaging modalities and spatial partitioning, as well as the results of the validation performed with a commercial endoscope. Section 3 also discusses the extracted feature strengths using NBI and WLI. Section 4 summarizes the relevance of these findings and discuss the future of the work.

2. Methodology

The proposed technique calculates the change in the endoscope camera pose (i.e., 6 DOF transformation matrix, a common representation in robotics [47], with three DOF for position and three Euler angles describing orientation) between sequential frames

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002

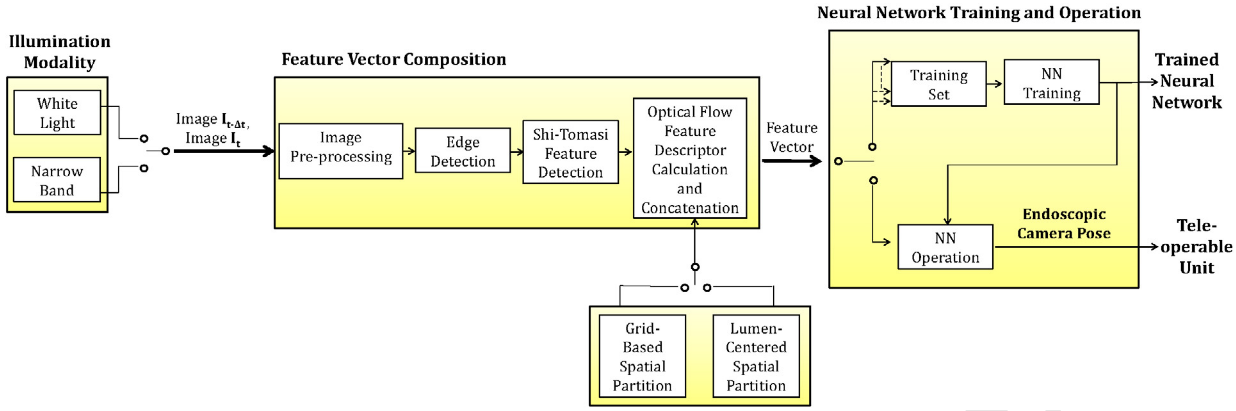


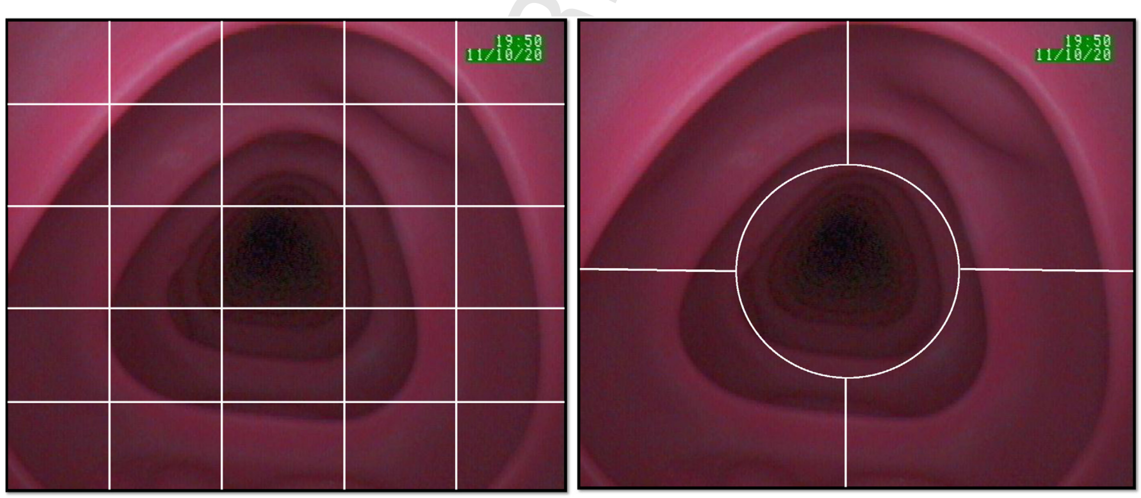
Fig. 2. Flow diagram for the proposed method for calculating the change in the position and orientation of the endoscopic module, including the investigated variations in illumination (WLI or NBI) modality and spatial partitions (grid-based or lumen-centered).

using the image stream. Essential to the robust calculation of the camera motion parameters is the extraction of stable features from the endoscopic image stream. The gastrointestinal lumen is well-known for its lack of texture and brightness constancy complications due to changes in illumination from the movement of the endoscope [8,43,46]. The proposed algorithm involves finding strong correspondences in two sequential images using optical flow, applying a spatial grid to form a feature vector which expresses the visual representation of the change in pose, and then using this information to train the ANNs, as represented in Fig. 2. The performance of the ANN is then tested on a separate diverse test set.

The only assumption made is that the scene is static; thus, the movement perceived in the image can be assumed to be due to the change in the pose of the camera only. This is a valid assumption since there are only three major contributors to the movement of the colon: respiration, wall deformation due to the endoscope, and haustral contractions. It is assumed that the effects of respiration will be minimal, since the colon is insufflated during colonoscopy. Additionally, on average, a displacement of only approximately 7.85 mm occurs in the anterior/posterior plane during deep respiration [48]. Furthermore, the colon wall does deform due to the movement of the endoscope; however, this only affects the colon in regions behind the camera on the endoscope. Thus, this does not contribute significantly to a change in the scene captured by the camera. Haustral contractions, which move the content of the colon forward, are the only movements which significantly violate the inertness of the scene. Since these only occur every 25–30 min [49], a specialized control loop within the teleoperation software can be used to handle this exception.

In order to perform more accurate pose estimation for endoscope localization, we investigate two different imaging modalities – WLI vs. NBI - and we apply two different spatial partitions – lumen-centered vs. grid-based. We then compare the performance of the ANNs, trained on ground truth provided by a robotic arm moving the camera during the trials, within these four variations as concerns pose detection. The four variations are then compared to a commercially available state-of-the-art magnetic tracker. Performance is measured by calculating the root mean square error (RMSE), where this error is the deviation of the estimated pose from ground truth. We additionally measure the time to complete the algorithm using a standard laptop (Lenovo Thinkpad T520, Intel Core i5-2520M CPU at 2.50 GHz, Windows 7 Professional; Lenovo; USA).

Given these results, we then assess the validity of this approach in a clinical setting. This is achieved by implementing the technique on a commercial endoscope with both WLI and NBI capabilities. These experiments are performed in a human colon simulator using fresh blood, and the endoscope is driven by an expert endoscopist. Within this experiment, we again compare the performance of the



(a) Grid-based spatial partition. (b) Lumen-centered spatial partition.

Fig. 3. (a and b) Spatial partitioning rules for feature vector composition.

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002

four variations, but in this case, the ANNs are trained on noisy data provided by a magnetic tracker. The metric used to evaluate plausibility of this algorithm for clinical use is the RMSE between the pose reported by the ANNs and the magnetic tracker readings. Using the image stream acquired from the commercial endoscope, we also calculate the power of features obtained by WLI and NBI.

2.1. Feature vector composition

Fig. 2 shows the flow diagram of the algorithm used to acquire each image. Frames are captured from the video processor at times $t-\Delta t$ and t, and first cropped down to their effective pixels and converted to grayscale. Using the Shi-Tomasi (S-T) features [50] found in image $I_{t-\Delta t}$, the locations of the corresponding features in image I_t are found using the Lucas-Kanade optical flow algorithm [39], and thus, the optical flow from the previous to the current image is encoded.

Feature descriptors, which summarize the nature of these correspondences in specific regions of the image, are constructed based on the partitioning method adopted. The region boundaries defined by the spatial partitioning divisions, which are shown in Fig. 3.

The first – grid-based spatial partitioning (Fig. 3a) – is a basic partition of the image in 25 equal rectangular regions (i.e., 5 × 5 grid). This represents a simplistic static grid system, a common partitioning method in computer vision applications [43,51–53], which could easily be achieved by a simple image overlay in an endoscopic module. For each grid location $g \in G$, two feature descriptors are calculated as

$$\overline{dx}_g = \frac{\sum_{i=1}^{n_g} dx_i}{n_g}$$

and

$$\overline{dy}_g = \frac{\sum_{i=1}^{n_g} dy_i}{n_g}$$

where n_g is the number of feature correspondences present in image I_t at grid location g, and dx and dy are the change in coordinates in the X and Y directions between corresponding features in image $I_{t-\Delta t}$ and I_t. These features are then concatenated into a feature vector of size 50 (25 grid regions with 2 feature descriptors each) as input to the ANN.

Lumen-centered spatial partitioning, shown in Fig. 3b, is conceived considering that the colon is a tubular structure with a dark region which usually corresponds to the center of the lumen. This partitioning method is based on consistently aligning the center of the partition with the lumen center. This approach requires first segmenting the image into the lumen center and surrounding area. A lumen segmentation approach similar to [8] was taken, by first histogram equalizing the image to increase contrast and then applying a threshold. In the resultant image, the lumen appears white, while the rest of the image appears black.

The centroid of the lumen (x_c, y_c) is then calculated. The circumference of the lumen is calculated as a summation of the pixels on the edge of the lumen in this thresholded image. The radius r is calculated using this circumference estimate by dividing by 2π. The centroid of lumen in conjunction with this calculated radius defines the first region of the lumen-centered approach. The other four quadrants are defined by dividing the image horizontally at y_c, and vertically at x_c not including the area labeled as the lumen center. For each of these 5 regions, the two feature descriptors are calculated as

$$\overline{dr}_g = \frac{\sum_{i=1}^{n_g} \sqrt{dx_i^2 + dy_i^2}}{n_g}$$

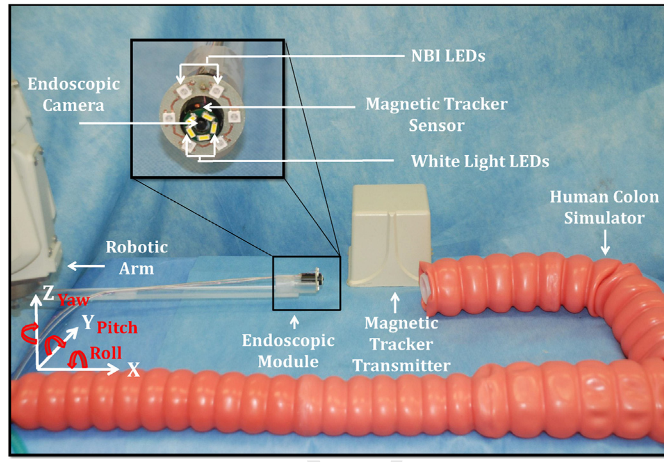


Fig. 4. Experimental setup for training and testing of the proposed pose detection approach.

and

$$\overline{\theta}_g = \frac{\sum_{i=1}^{n_g} tan^{-1}(dy_i/dx_i)}{n_g}$$

where $\overline{dr}_g$ is the average distance of optical flow between corresponding features in region g, and $\overline{\theta}_g$ is the average inclination of the flow of the features in region g. These features are then concatenated into a feature vector of size 10 (5 regions described by 2 feature descriptors each) as input to the ANN.

2.2. ANN training and operation

The set of feature vectors generated from either the grid-based or the lumen-centered partitions are then used as input into a multi-layer feedforward ANN. ANNs are computational networks which are useful in function approximation and pattern recognition due to their rejection of noise in the training set, high accuracy, and speed of computation during operation [54,55]. ANNs are powerful for learning complex mappings, given the correct number and size of hidden layers and certain characteristics of the function to be mapped from a set of exemplars and their target outputs [56].

In order to train the ANN (i.e., tune the weights to learn the mapping between optical flow features and pose change), the full training set is first divided further into a slightly smaller training set, a validation set, and a test set. During training, each input vector in the new training set is presented to the ANN, and forward propagated through the network. After the outputs are generated, the error between the network output and the ground truth obtained from the true motion of the endoscopic module is calculated. The weights of the network are then adjusted based on this error using Levenberg–Marquardt error backpropagation [57,58]. Training is stopped when the error in the validation set begins increasing over a specified number of epochs. This technique is referred to as early stopping, and allows the ANN to maintain its ability to generalize by preventing overtraining (i.e., memorization of the training set). In addition, the test set garnered from the training set is used as an independent gauge to assess the learning of the network by testing before training and after training. At this point, the network is considered trained, and testing on an independently generated test set (i.e., not the small test set created by segmenting the training set) is performed.

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002

2.3. Benchtop validation of proposed method

The purpose of this benchtop experiment is to assess the performance of the four variations of ANNs along the straight sections of the colon to compare the RMSE in pose detection to that of a state-of-the-art magnetic tracker. The experimental setup – illustrated in Fig. 4 – consists of a tethered endoscopic module (22 mm in length × 27 mm in diameter) rigidly connected to a 6 DOF industrial robotic arm (RV-6SDL; Mitsubishi Corporation, Japan). Thanks to the rigid connection with the robotic arm, the actual position of the endoscopic module can be derived from the robot encoders at any given point in time. This data is used as ground truth for pose detection assessment. The endoscopic module contains a 500 × 582 resolution endoscopic camera (291,000 effective pixels, cross-section 3 mm × 3 mm, and 140° field of view; Introspicio 110, Medigus, Ltd., Israel), 5 white light emitting diodes (LEDs) (NESW007BT; Nichia Corporation, Japan), and 6 blue light (450 nm) LEDs (Kingbright Electronic Company, Ltd., Taiwan) for NBI. The two illumination systems were designed and driven so as to have approximately the same light intensity (6.5 candela). The unit also contains a 6 DOF magnetic tracker sensor (1.4 mm positional nominal RMSE, 0.5° rotational nominal RMSE, 240–420 Hz update rate; 3D Guidance trakStar, Mid-range; Ascension Technology Corporation, USA) to compare the accuracy of the algorithm with a commercially available tracker. The validation software provided with the device was then used to appropriately position the magnetic tracker transmitter. This was done in order to ensure the highest fidelity readings from the sensor by minimizing interference from other metallic objects.

During training and testing, the endoscopic module is moved along the straight sections of a plastic human colon simulator (Kyoto Kagaku, Japan). This phantom model is commonly used for training medical doctors in performing colonoscopy and possesses the gross anatomy of a human colon. In order to recreate features that are enhanced by NBI – such as the blood vessels and capillaries in the colon – fresh porcine blood was applied to the interior of the colon simulator. To accurately model the lighting environment of the colon, the simulator was covered by an opaque black cloth (not shown in figure).

Control software written in C++was used to send positional commands via TCP ethernet connection to the robot controller to move the arm in real-time and capture the resulting movement of the camera and tracker in the endoscopic module. Frame couples were compared at each iteration of the procedure outlined in Algorithm 1, using a frame grabber connected to a camera video processor (Introspicio; Medigus, Ltd., Israel), and were then read into the control software and processed using OpenCV [59] library functions. The magnetic tracker pose was read at the same time of the camera and robot encoders using functions from Ascension's 3D Guidance Application Programming Interface (API).

Table 1
Magnitude, direction, and number of training repetitions for generating ANN training set.

Degree of freedom tested	Magnitude of training repetitions	Total number of training repetitions
X only	±0.5 mm to ±5 mm	180
Y only	±0.5 mm to ±3 mm	120
Z only	±0.5 mm to ±3 mm	120
Roll only	±0.5 mm to ±2°	80
Pitch only	±0.5 mm to ±2°	80
Yaw only	±0.5 mm to ±2°	80
Translation only	Variable	120
Rotation only	Variable	80
All degrees	Variable	320

Algorithm 1. Algorithm for training set generation and training of ANN.

Result: Trained ANN
Select desired illumination modality;
Read in image $I_{t-\Delta t}$;
Record robot encoders and magnetic tracker sensor position at time t;
while *Not finished with training trajectory* **do**
  Move robot/endoscopic camera to next training pose;
  Read in image I_t;
  Generate and record optical flow-based feature vector;
  Record robot and magnetic tracker sensor position;
  Set image $I_{t-\Delta t} = I_t$
end
Calculate change in pose (ΔP_{target}) from ground truth to be used as target vectors for ANN;
while *Validation training error has not increased for six epochs* **do**
  Forward propagate input vector through ANN to get estimated pose$\Delta P_{predicted}$;
  Backpropagate error $\frac{1}{2}||\Delta P_{target} - \Delta P_{predicted}||^2$ to train network;
end

The procedure for generating the training and testing sets is shown in Algorithm 1. Each time the robot/endoscopic module assembly is moved, the resultant optical flow-based feature vectors are calculated, and the robot and magnetic tracker positions are recorded. This training trajectory is shown in Table 1, where the coordinates refer to the Cartesian axes of Fig. 4. The 1180 steps of the training trajectory are representative of endoscopic module when it moves in each DOF independently, and then in combinations of these DOFs. All these types of movements are common during colonoscopy [60]; however, in this trajectory, the corners of the colon are not traversed. The training set allows 10 repetitions of varying magnitudes for each DOF independently tested, and 5 training repetitions for any combinational movement. When combinational movement is tested, the magnitudes of the movements vary between 0 mm or 0° and the absolute maximum of the range shown in Table 1. The conclusion of this training trajectory execution marks the end of the training set generation and the beginning of ANN training.

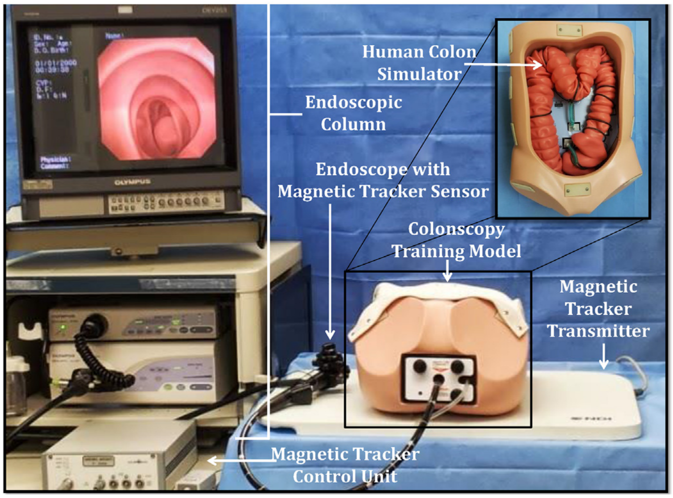


Fig. 5. Experimental setup for training and testing of the proposed pose detection approach using a commercial colonoscope equipped with a magnetic tracker via tool channel.

The offline training of the ANN proceeds as follows: as shown in Fig. 2, the inputs to the ANN are optical flow feature vectors, which are a compact representation of the evolution of the scene at each time step. The corresponding ANN training targets are calculated using the previous pose as the reference frame. Then, the change in pose is calculated using the current pose, and these differences are used as targets for the ANN. Using these data as the training set, the ANN learns a numerical estimation of the change in the 6 DOF pose of the endoscopic tip as a function of the optical flow between two images.

Testing starts by moving the endoscopic module along an arbitrary trajectory and recording the positions and orientations of the robot and magnetic tracker, as well as the optical flow-based feature vectors between successive images. The testing set was randomly generated to fall within 0 ± 5 mm in the Z direction; 0 ± 3 mm in the X and Y directions; and rotations of $0 \pm 2^\circ$ in roll, pitch, and yaw as defined in Fig. 4. The main difference between training and testing is that during testing, the inputs are simply forward propagated to output the approximated pose (that is, there is no calculation of or backpropagation of error). The RMSE is then calculated for the ANNs and commercial tracker with respect to ground truth.

Matlab's Neural Network Toolbox was used for training and computing outputs for network training and testing. 85% of the training set was allocated for pure training, whereas 10% and 5% of the training set were used for validation and performance testing, respectively. To terminate training, early stopping was invoked if the error in the validation test set increased for six successive epochs. The ANNs are constructed to have $2n+1$ hidden layer architecture [61], where n is the number of nodes in the input layer (i.e., number of features in the input vector); the ANNs, then, have either $50 \times 101 \times 6$ or $10 \times 21 \times 6$ architecture for grid-based and lumen-centered spatial partitioning, respectively.

2.4. Application of proposed method to commercial endoscope

The training method proposed in the previous subsection provides the most reliable data on which to train the ANN. However, ground truth may not always be available. To specifically address this case, we performed an experiment which more accurately reflects the conditions to be expected in a clinical setting.

The setup used in this experiment is shown in Fig. 5. In this experiment, an expert gastroenterologist performed a set of four colonoscopies on a colonoscopy training model (Kyoto Kagaku, Japan), in which a plastic human colon simulator was arranged in a basic anatomical configuration (Fig. 5, upper-right corner). In order to ensure the presence of randomized features, the colon was filled with porcine blood, and then adequately drained. A 5 DOF magnetic tracking system (1.20 mm positional nominal RMSE, 0.5° rotational nominal RMSE, 40Hz update rate; Aurora, Tabletop Transmitter; Northern Digital Inc. (NDI), USA) was inserted into the tool channel of a state-of-the-art flexible endoscope (H180AL/I Colonovideoscope; Olympus, Japan), which was then used to perform 4 colonoscopies – 2 under WLI, and 2 under NBI. Following each colonoscopy, the endoscope was completely removed from the simulator, and the interior of the colon agitated so as to prevent bias due to the ANNs learning the specific blood patterns.

Control software written in C++utilizing the NDI API and OpenCV was used to capture image frames and read the pose from the sensor for the duration of the trial. A trial is defined as a full traversal from the sigmoid colon to the cecum and a subsequent return to the sigmoid colon. As in the previous experiment, feature vectors were composed as outlined in Section 2.1 for grid-based and lumen-centered partitioning. Two amendments were made to the procedure for the image processing of NBI images. Since histogram equalization is ineffective in increasing image contrast for NBI, this step was ignored. Furthermore, an image mask was applied to the image such that only the endoscopic image was processed; this excludes the black border surrounding the image. The rest of the steps were performed identically to the previous experiment. Thus, the resultant thresholded image was comparable to that which would be obtained with WLI.

The same procedure employed in the previous experiment was used in order to train the ANNs; using the feature vectors as inputs and the change in pose calculated from the magnetic tracking system, each of the four ANNs was trained. One trial under both WLI and NBI was used for training. The remaining trial was used in order to test the performance of the ANNs by evaluating the RMSE between the ANN outputs and the pose reported by the tracker.

2.5. Assessment of feature strength based on illumination and color

In order to quantitatively compare the strength of the features extracted from WLI and NBI, the criteria for trackable corners or edges used in the S-T algorithm for good features was employed [50]. The S-T algorithm, a well-established and common image processing method, extracts strong features from an image by calculating the eigenvalues of a pixel of interest in a local neighborhood. This algorithm identifies two types of good features – corners and edges. A corner is indicated when these eigenvalues are both large (i.e., there is a large variation in both directions), and an edge is indicated by one large eigenvalue.

In order to evaluate these features, a single image was first divided into its respective red, green, and blue channels, and additionally converted to grayscale. For each of these 4 derived images, the S-T algorithm was first applied in order to find the locations of the good features, and at these points, the maximum eigenvalue was recorded. Then, the total number of features, as well as the mean and standard deviation of these maximum eigenvalues was found for each image. This was repeated for 200 images collected from the trial described in Section 2.4 under both WLI and NBI. In this way, the strength of the features based on illumination were quantified, and an assessment of the role of each of the color channels was performed.

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002

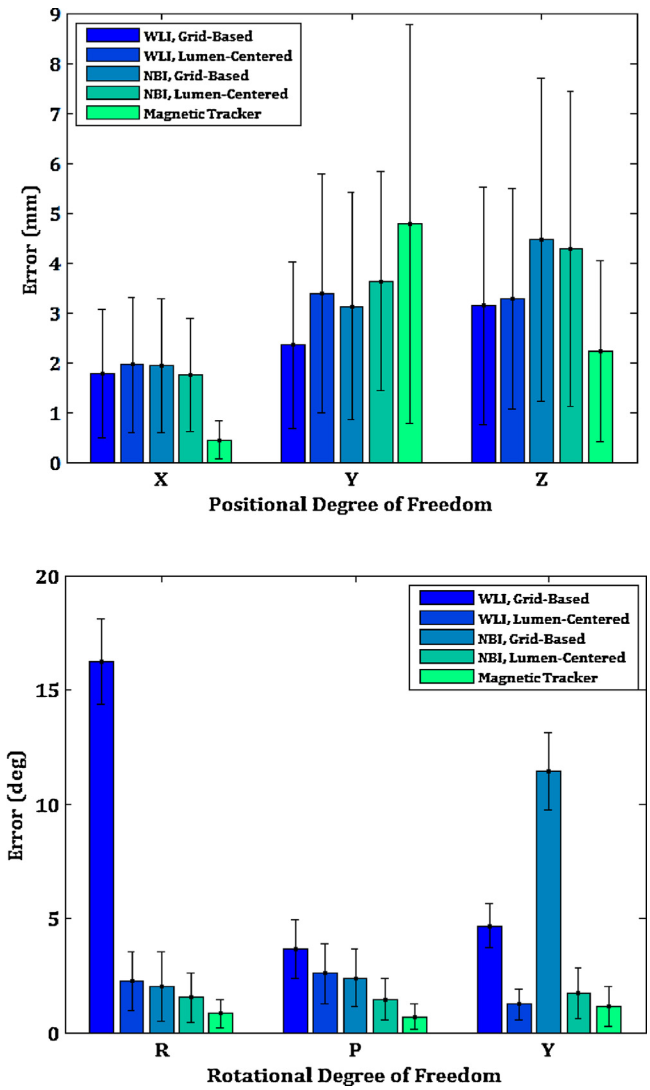


Fig. 6. RMSE and standard deviation of the ANN variations and state-of-the-art magnetic tracker with respect to ground truth based on robot encoder readings.

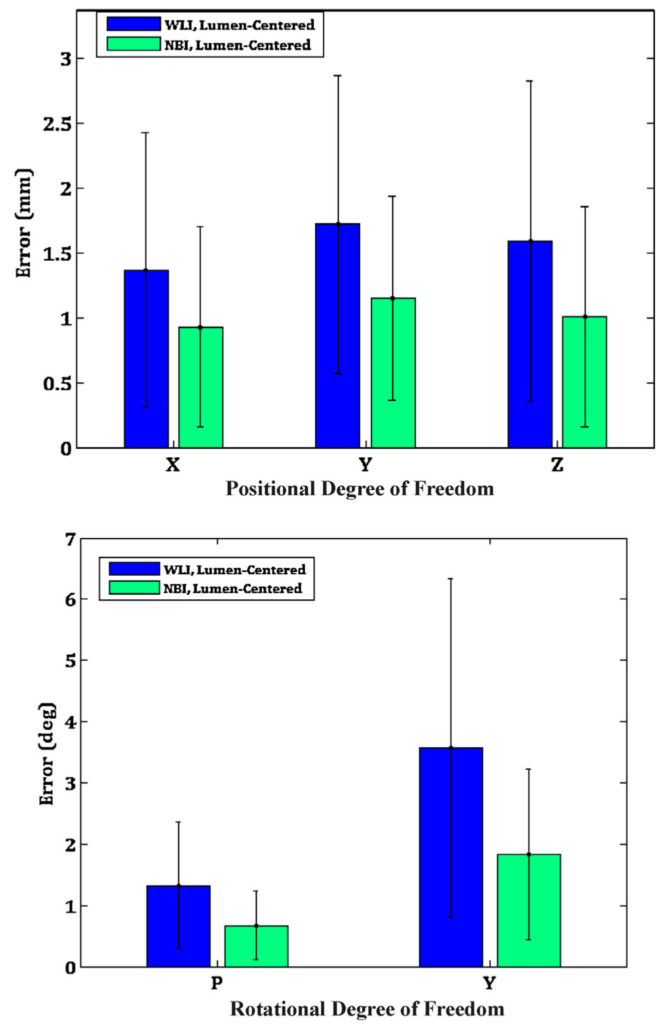


Fig. 7. RMSE and standard deviation of ANN pose estimations against magnetic tracker readings during clinical evaluation of the algorithm.

3. Results and discussion

3.1. Benchtop validation results

A comparison of the RMSE of the performance of the ANNs and magnetic tracker over the entire testing set is shown in Fig. 6. As shown, for all variations of the ANNs in all positional DOF, the RMSE is less than 5 mm. The ANNs best estimate the *X* DOF, which is the direction along the optical axis, and arguably the most important. Although there is similar performance in all of these ANNs, the grid-based partitioning method under WLI is the best performing ANN. This is again confirmed in the *Y* and *Z* DOF. Noteworthy is the *Y* DOF, in which all of the ANNs are able to achieve better performance than the tracker.

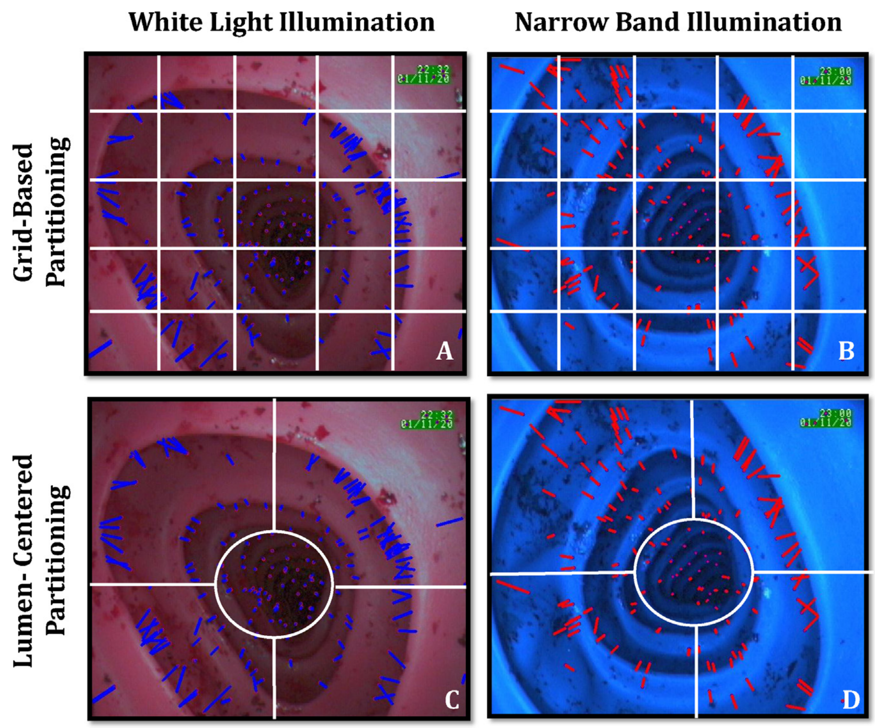
With regards to the rotational DOF, the lumen-centered partitioning using NBI is the best performing ANN with RMSE of less than 1.7° in each case. In the yaw DOF, the WLI lumen-centered partition performs marginally better than the ANN resulting from NBI and lumen-centered partitioning. In the rotational DOF, the tracker consistently outperforms the ANNs, but the average difference is a trivial 0.7°.

3.2. Results of validation with commercial endoscope

Fig. 7 shows the RMSE produced by the ANNs for the positional and rotational DOF. In this case, only 5 DOF are shown since the adopted magnetic tracker is only able to report 5 DOF. Approximately 1% of the data was removed to account for outliers. Additionally, the results for grid-based partitioning are not shown in Fig. 7 since their error is up to 10 times greater than that of the lumen-centered partitioning approach. In comparing the illumination modalities for lumen-centered partitioning, the ANNs trained with NBI are able to constantly achieve slightly better performance in terms of accuracy and precision than the WLI ANNs. Thus, lumen-centered partitioning using NBI is a superior mechanism to WLI for vision-based motion estimation in this application, although both have RMSE less than 2 mm in position and 3° in orientation.

An important result of this trial is that the ANNs can be trained on noisy data, and still produce valid results, especially in the *X*, *Y*, and yaw directions. Indeed, one of the applications of ANNs is to filter noisy data. Furthermore, this experiment with a commercial endoscope verifies that this approach is robust; during these trials, the endoscope water channel was used to clean the lens, the endoscope was moved with sharp and sudden motions, and blood frequently obscured the image – all of which produce significant noise and disturbances in the image. Even further, the effect of roll

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002



Q2 **Fig. 8.** (A–D) Typical optical flow patterns for a 5-mm translation along the *Z* axis with combinations of illumination modalities and spatial partitions. Tests were performed in a human colon simulator with porcine blood staining.

was essentially filtered out from the data set by the ANNs since it was unmeasured, but still present in the image stream.

A major contributor to the error in the pose estimates is likely due to this kind of noise, which obscures the image. Smooth, controlled movements, and dedicated algorithms for compensating for lens cleaning will aid in minimizing the error throughout the entire trajectory of the colon. One source of error that cannot be overlooked is the impact of the corners in this experiment, which particularly affects the performance of the grid-based method. This likely explains the poor estimation ability of the grid-based partitioning when tested on the entirety of the colon simulator rather than just the straight trajectories as described in Section 2.3.

Lastly, given the frame $I_{t-\Delta t}$, the time required to acquire the current frame I_t and estimate the pose variation is approximately 280 ms for lumen-centered partitioning – the most demanding in terms of computational time – during the highest magnitude of

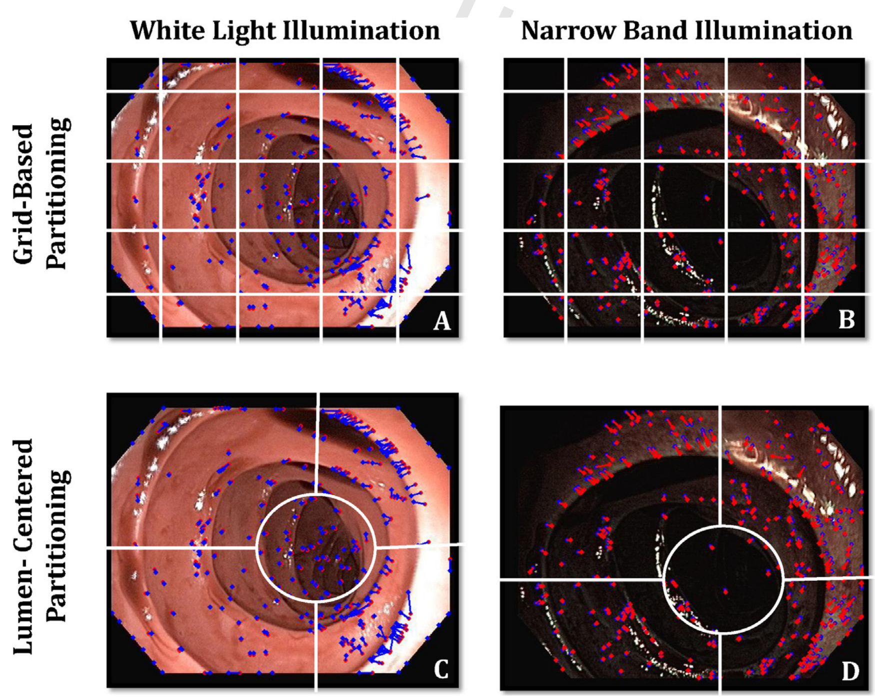


Fig. 9. (A–D) A comparison of the image of the lumen of the colon under WLI and NBI using a commercial endoscope.

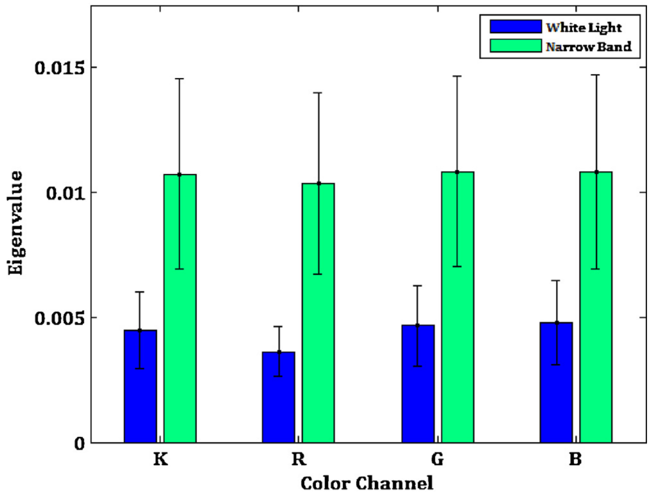


Fig. 10. A comparison of the strength of features between WLI and NBI per color channel. K represents a grayscale version of the image, R is the red channel, B is the blue channel, and G is the green channel.

movement tested. However, as mentioned in Section 2, this rate is highly dependent on the number of S-T feature correspondences found in each iteration of the testing procedure, which is approximately 300–10,000 in this implementation. Also, this algorithm was tested on a standard laptop with unoptimized code. Therefore, we expect that the computational time will be significantly reduced by parallelizing the computation of S-T feature correspondences, optimizing the code, and using a faster computer.

3.3. Feature analysis

The difference in appearance of the colon during the first trial comparing illumination modalities and spatial partitions is shown in Fig. 8. This corresponds to a translation of 5 mm along the *Z* axis. Visual inspection of the images reveals that there are distinct patterns in the optical flow due to the movements of the camera relative to the static image.

Fig. 9 shows an analogous set of images, but using WLI and NBI provided by the commercial endoscope. As is evident from a comparison of Fig. 8 and this figure, there is a visual difference in utilizing the artificial NBI LEDs versus NBI from the commercial endoscope. However, the effect is still the same; under qualitative inspection of the image, the blood features using NBI are much stronger and obvious as compared to those of WLI. In the figures shown, the partitioning approaches are overlaid on the image to visualize the impact of the segmentation on the feature set; however, in practice such a division is not created on the image. A quantitative comparison of feature strength for grayscale representation of the image versus the red, green, and blue (RGB) color channels for WLI and NBI based on the images from the commercial endoscope is shown in Fig. 10. Corners and edges represent areas of an image that have high variation; these variations correspond to high eigenvalues. As shown in this figure, NBI features have more than twice the strength of WLI features. Although it has a larger standard deviation in terms of these eigenvalues, even the lowest value of the mean – given this standard deviation – still far exceeds that of WLI. The average number of WLI features is 12,600 for all color channels; on average, NBI images have 11,700 features. This suggests that although WLI has more features than NBI, they have half the quality of NBI features. In addition, our adoption of an averaged grayscale image and discarding the other color components is justified since the grayscale value has nearly the same mean feature strengths as the other color channels.

This aligns with the results found in Section 3.2, as well as visual inspection of the images in Fig. 9, which show much higher contrast between the colon and blood features. Furthermore, since the features found in NBI are twice as strong as those found for WLI, this enables more robust tracking of features from one frame to another. This results in more consistency and coherence between the training set for the ANN and the testing data that are encountered in practice.

4. Conclusions

Teleoperated and automated flexible endoscopes have the potential to impact the lives of people worldwide by reducing the perceived indignity and discomfort of colonoscopy and other clinical procedures. Pose detection algorithms provide positional and rotational feedback about the movement of the tip of the endoscope as a result of actuating the device. This is essential for controlling devices in dynamic environments which cannot be accurately modeled, especially in the presence of disturbances caused by the environment and noise inherent in the system. The research presented in this paper has the potential to become an enabling technology for teleoperated and automated colonoscopy.

This work first investigated how pose feedback can best be estimated by using components that are already available in many modern endoscopes, including the endoscopic camera, WLI, and NBI. In order to do this, an image-based approach using optical flow between two successive frames was used to train ANNs to estimate a change in pose of the endoscope tip. The inputs to the ANN were feature vectors created using two different spatial partitioning approaches (grid-based or lumen-centered). To assess the proposed approach and compare it to commercially available magnetic trackers, a benchtop experiment was performed using a human colon simulator with blood. All the ANNs achieved positional RMSE of less than 5 mm, and in one case, the error in all the ANNs was lower than that of the commercial magnetic tracker. The best combination of illumination and partitioning was WLI with grid-based partitioning (2.42 mm RMSE). However, in terms of rotational RMSE, the most accurate ANN was the one using NBI and lumen-centered partitioning (1.69° RMSE). During this trial, the tracker obtained an accuracy of 2.49 mm in positional DOF and 0.89° in rotational DOF. With these results, we can conclude that the optical flow-based ANN has performance comparable to that of a state-of-the-art commercial tracker.

To confirm these results in a clinical setting, 4 colonoscopies were performed with a commercial endoscope operated by an expert endoscopist on a colonoscopy training model with fresh porcine blood. Four sets of images were produced – 2 under WLI, and 2 under NBI. During both these trials, the position and orientation of a magnetic tracker placed in the tip of the endoscope via the tool channel was also recorded. In each case of illumination, one image set was used for ANN training using the magnetic tracker readings as the target values. The other image set was used for testing, in which the performance of the ANNs was measured with respect to the magnetic tracker readings. The performance of lumen-centered partitioning with NBI was superior, with 1.03 ± 0.8 mm RMSE in positional DOF, and $1.26 \pm 0.98^\circ$ RMSE in rotational DOF, while with WLI, the performance was 1.56 ± 1.15 mm RMSE in positional DOF and $2.45 \pm 1.90^\circ$ RMSE in rotational DOF.

A secondary purpose of this study was to assess the impact of illumination and color channel on feature strength. This was achieved by analyzing a series of images collected from the experiment using the commercial endoscope. The features were compared based on their eigenvalues, a common image processing measure of feature strength. A comparison of these eigenvalues showed that features obtained from NBI were on average twice as

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002

strong as the features extracted under WLI. No significant difference between the features strengths obtained from the RGB color channels or grayscale for any illumination was observed.

This work demonstrated that an image-based approach using ANNs to learn the relationship between optical flow and change in pose of teleoperated flexible endoscopes is comparable to that of a commercially available magnetic tracker. The performance obtained by the ANNs was enhanced by the NBI modality, which corresponds to stronger features and better pose estimation. These findings indicate that NBI combined with a dynamic feature partitioning based on the anatomical structure of the colon – given the feature descriptors used for quantification – provides reliable and accurate feedback about the change in pose resulting from actuation of the endoscope. Regardless, the pose estimation algorithm presented can also be used with commercial endoscopes without NBI, although the accuracy will be slightly lower.

It is worth mentioning that the ANN trained using the benchtop experimental setup was not used for the assessment performed with the commercial endoscope. This is because the benchtop experiment did not include training data with corner folds of the colon or irregularly spaced or oriented haustral folds; this was left for validation with the commercial endoscope. However, the training portion of this algorithm is meant to be performed once for the lifetime of the endoscope, assuming that the camera optics/illumination do not change significantly. Initial training of the algorithm would require a calibration endoscopy to be performed once by the endoscopist; however, numerical training and usage of the neural network would proceed in a software automated fashion. As a future work, we will quantify the effect of the variance of the appearance of the colon among different patients and we will verify to what extent a new calibration/training of the ANN is required.

Although this work cannot be used to detect looping or colon perforation, it is a novel method which uses components native to commercial endoscopes for pose feedback to teleoperated endoscopes. Future work includes a further exploration of feature descriptors used for input to the ANNs, particularly those that leverage the strength of features provided by the illumination modality employed, RGB color features [62], and aggregated features. Additionally, this approach must still be confirmed by *in vivo* trials, repeating the experiment inside a living colon. Therefore, we plan porcine model experiments using a commercial NBI endoscope as next step of this work. These experiments will allow a more accurate description of features produced by NBI due to the presence of blood vessels. This will also enable us to find optimal features and feature descriptors for each control loop in order to decrease the computational time for real-time pose estimation, while maintaining or improving the current accuracy. Furthermore, these trials will provide the opportunity to assess the robustness of the proposed method with respect to haustral contractions. To cope with this issue, we plan to freeze the endoscope motion during the haustral contraction and resume pose detection once the contraction is over. Finally, the methodology presented will be integrated as real-time closed-loop feedback into the control system of a teleoperated platform to achieve reliable remote manipulation of a teleoperated flexible endoscope.

Acknowledgment

The authors would like to thank Trevor Bruns, Marco Beccani, Christian Di Natali, and Jack Noble for their assistance in this work.

References

[1] Fact sheet #297. Cancer. World Health Organization (WHO) [online]. Available at: www.who.int/mediacentre/factsheets/fs297/en [last accessed: 01.02.2012].
[2] Valdastri P, Simi M, Webster III RJ. Advanced technologies for gastrointestinal endoscopy. Annual Review of Biomedical Engineering 2012;14: 397–429.
[3] Hasan M, Wallace M. Image-enhanced endoscopy. American Society for Gastrointestinal Endoscopy 2009;16(4):1–4.
[4] Postic G, Lewin D, Bickerstaff C, Wallace M. Colonoscopic miss rates determined by direct comparison of colonoscopy with colon resection specimens. American Journal of Gastroenterology 2002;97(12):3182–5.
[5] Pasha S, Leighton J, Das A, Harrison M, Gurudu S, Ramirez F, et al. Comparison of the yield and miss rate of narrow band imaging and white light endoscopy in patients undergoing screening or surveillance colonoscopy: A meta-analysis. American Journal of Gastroenterology 2011;107(3):363–70.
[6] Vital Signs Cancer screening, colorectal cancer. Centers for Disease Control and Prevention [online]. Available at: www.cdc.gov/vitalsigns/CancerScreening/indexCC.html [last accessed: 01.02.2012].
[7] Obstein KL, Valdastri P. Advanced endoscopic technologies for colorectal cancer screening.". World Journal of Gastroenterology 2013;19(4):431–9.
[8] Reilink R, Stramigioli S, Misra S. Three-dimensional pose reconstruction of flexible instruments from endoscopic images. In: Amato N, editor. Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2011. p. 2076–82.
[9] Reilink R, Stramigioli S, Misra S. Image-based flexible endoscope steering. In: Luo R, editor. Proceedings of the 2010 IEEE/RSJ international conference on intelligent robots and systems. Institute of Electrical and Electronics Engineers (IEEE), Piscataway, NJ, USA; 2010. p. 2339–2344.
[10] van der Stap N, Reilink R, Misra S, Broeders I, van der Heijden R. The use of the focus of expansion for automated steering of flexible endoscopes. In: Desai J, Phee Soo Jay L, Zollo L, editors. 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2012. p. 13–8.
[11] Ruiter J, Rozeboom E, van der Voort M, Bonnema M, Broeders I. Design and evaluation of robotic steering of a flexible endoscope. In: Desai J, Phee Soo Jay L, Zollo L, editors. 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2012. p. 761–7.
[12] Troccaz J, editor. Medical robotics. London, UK/Hoboken, NJ, USA: ISTE Ltd./John Wiley & Sons, Inc.; 2012.
[13] Northern Digital Inc. [online]. Available at: www.ndigital.com/medical/aurora.php [last accessed: 01.05.2013].
[14] Ascension Technology Corporation [online]. Available at: www.ascension-tech.com/medical [last accessed: 01.05.2013].
[15] Szura M, Bucki K, Matyja A, Kulig J. Evaluation of magnetic scope navigation in screening endoscopic examination of colorectal cancer. Surgical Endoscopy 2012;26:632–8.
[16] Deguchi D, Mori K, Suenaga Y, Hasegawa J, Toriwaki J, Natori H, et al. New calculation method of image similarity for endoscope tracking based on image registration in endoscope navigation. International Congress Series 2003;1256(0):460–6.
[17] Deguchi D, Mori K, Feuerstein M, Kitasaka T, Maurer JC, Suenaga Y, et al. Selective image similarity measure for bronchoscope tracking based on image registration. Medical Image Analysis 2009;13(4):621–33.
[18] Asano F. Virtual bronchoscopic navigation. Clinics in Chest Medicine 2010;31(1):75–85.
[19] Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki J, C.M. Jr, et al. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Medical Image Analysis 2002;6(3):321–36.
[20] Rey J, Ogata H, Hosoe N, Ohtsuka K, Ogata N, Ikeda K, et al. Blinded nonrandomized comparative study of gastric examination with a magnetically guided capsule endoscope and standard videoendoscope. Gastrointestinal Endoscopy 2012;75(2):373–81.
[21] Keller J, Fibbe C, Volke F, Gerber J, Mosse AC, Reimann-Zawadzki M, et al. Inspection of the human stomach using remote-controlled capsule endoscopy: a feasibility study in healthy volunteers (with videos). Gastrointestinal Endoscopy 2011;73(1):22–8.
[22] Valdastri P, Ciuti G, Verbeni A, Menciassi A, Dario P, Arezzo A, et al. Magnetic air capsule robotic system: proof of concept of a novel approach for painless colonoscopy. Surgical Endoscopy 2011;26(5):1238–46.
[23] Keller H, Juloski A, Kawano H, Bechtold M, Kimura A, Takizawa H, et al. Method for navigation and control of a magnetically guided capsule endoscope in the human stomach. In: Desai J, Phee Soo Jay L, Zollo L, editors. Proceedings of the 4th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2012. p. 859–65.
[24] Mahoney A, Abbott J. Control of untethered magnetically actuated tools with localization uncertainty using a rotating permanent magnet. In: Desai J, Phee Soo Jay L, Zollo L, editors. Proceedings of the 4th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2012. p. 1632–7.
[25] Gao M, Hu C, Chen Z, Zhang H, Liu S. Design and fabrication of a magnetic propulsion system for self-propelled capsule endoscope. IEEE Transactions on Biomedical Engineering 2010;57(12):2891–902.
[26] Wang X, Meng M, Chen X. A locomotion mechanism with external magnetic guidance for active capsule endoscope. In: Desai J, Phee Soo Jay L, Zollo L, editors. 32nd Annual International Conference of the IEEE Engineering in Medicine

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002

and Biology Society. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2010. p. 4375–8.
[27] Nageotte MNF, Zanne P, Mathelin MD. In vivo comparison of real-time tracking algorithms for interventional flexible endoscopy. In: Karl W, editor. Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2009. p. 1350–3.
[28] van der Stap N, Heijden F, Broeders I. Towards automated visual flexible endoscope navigation. Surgical Endoscopy 2013:1–9.
[29] Khan G, Gillies D. Vision based navigation system for an endoscope. Image and Vision Computing 1996;14(10):763–72.
[30] Bricault I, Ferretti G, Cinquin P. Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy. IEEE Transactions on Medical Imaging 1998;17(5):703–14.
[31] Krishnan S, Tan C, Chan K. Closed-boundary extraction of large intestinal lumen. In: Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 1994. p. 610–1.
[32] Tian H, Srikanthan T, Asari K. Automatic segmentation algorithms for the extraction of lumen region and boundary from endoscopic images. Medical and Biological Engineering and Computing 2001;39(1):8–14.
[33] Xia S, Krishnan S, Tjoa M, Goh P. A novel methodology for extraction colon's lumen from colonoscopic images," Journal of Systemics. Cybernetics and Informatics 2003;1(2):7–12.
[34] Zabulis X, Argyros A, Tsakiris D. Lumen detection for capsule endoscopy. In: Chatila R, Merlet J, editors. Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2008. p. 3921–6.
[35] Zhen Z, Jinwu Q, Yanan Z, Linyong S. An intelligent endoscopic navigation system. In: Guo S, editor. Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2006. p. 1653–7.
[36] Stein GP, Mano O, Shashua A. A robust method for computing vehicle egomotion. In: Rillings J, editor. IEEE Intelligent Vehicles Symposium (IV2000). Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2000.
[37] Suzuki T, Kanade T. Measurement of vehicle motion and orientation using optical flow. In: Proceedings of the 1999 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems. Piscataway, NJ, USA: IEEE Service Center; 1999.
[38] Erdemir E, Wilkes D, Kawamura K, Erdemir A. Learning structural affordances through self-exploration. In: Blazevic P, editor. Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2012.
[39] Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. In: Hayes P, editor. Proceedings of the 7th International Joint Conference on Artificial Intelligence – vol. 2, ser. IJCAI'81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1981. p. 674–9.
[40] Davison AJ. Real-time simultaneous localisation and mapping with a single camera. In: Werner B, editor. Proceedings of the Ninth IEEE International Conference on Computer Vision – vol. 2, ser. ICCV'03. Washington, DC, USA: Institute of Electrical and Electronics Engineers (IEEE) Computer Society Press; 2003. p. 1403.
[41] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. 2nd ed. New York, NY, USA: Cambridge University Press; 2003.
[42] Thormählen T, Broszio H, Meier P. Three-dimensional endoscopy," Falk Symposium No. 124. Medical Imaging in Gastroenterology and Hepatology 2002.
[43] Liu J, Yoo T, Sabramanian K, Uitert RV. A stable optic-flow based method for tracking colonoscopy images. In: Huang T, editor. Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2008. p. 1–8.
[44] Suzuki K. Pixel-based machine learning in medical imaging. International Journal of Biomedical Imaging 2012;2012:1–18.
[45] Shi Z, He L. Application of neural networks in medical image processing. In: Yu F, Yue G, Leng M, Peng X, editors. Proceedings of the Second International Symposium on Networking and Network Security, April. Oulu, Finland: Academy Publisher; 2010. p. 23–6.
[46] Bulat J, Duda K, Duplaga M, Fraczek R, Skalski A, Socha M, et al. Data processing tasks in wireless GI endoscopy: image-based capsule localization & navigation and video compression. In: Dittmar A, editor. Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2007. p. 2815–8.
[47] Siciliano B, Sciavicco L, Villani L, Oriolo G. Robotics: Modelling, Planning and Control. 1st ed. London, UK: Springer-Verlag Limited; 2008.
[48] Korin H, Ehman R, Reiderer S, Felmlee J, Grimm R. Respiratory kinematics of the upper abdominal organs: A quantitative study. Magnetic Resonance in Medicine 1992;23.
[49] Sherwood L. Human Physiology: From Cells to Systems. 6th ed. Pacific Grove, CA, USA: Thomson Brooks/Cole; 2007.
[50] Shi J, Tomasi C. Good features to track. In: Shapiro L, editor. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: Institute of Electrical and Electronics Engineering (IEEE) Computer Society Press; 1994. p. 593–600.
[51] Saxena A, Schulte J, Ng AY. Depth estimation using monocular and stereo cues. In: Veloso MM, editor. IJCAI'07 Proceedings of the 20th International Joint Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press; 2007.
[52] Wilson S, Lovell B, Chang S, Masters B. Visual odometry for quantitative bronchoscopy using optical flow. In: Lovell B, Maeder A, editors. WDIC2005 APRS workshop on digital image computing proceedings, vol. 1, no. 1. The University of Queensland, St. Lucia, Queensland, Australia; 2005.
[53] Guizilini V, Ramos F. Semi-parametric learning for visual odometry. International Journal of Robotics Research 2013 April;32(5):526–46.
[54] Schalkoff R. Artificial Neural Networks. New York, NY, USA: McGraw-Hill; 1997.
[55] Schalkoff R. Pattern recognition: statistical, structural, and neural approaches. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1992.
[56] Cybenko G. Approximation by superpositions of a sigmoidal function,". Mathematics of Control Signals, and Systems (MCSS) 1989;2(4):303–14.
[57] Levenberg K. A method for the solution of certain nonlinear problems in least squares. Quarterly of Applied Mathematics 1944;2:164–8.
[58] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. SIAM Journal on Applied Mathematics 1963;11(2):431–41.
[59] Bradski G. The OpenCV Library. Dr Dobb's Journal of Software Tools 2000.
[60] Obstein K, Patil V, Jayender J, Estépar R, Spofford I, Lengyel B, et al. Evaluation of colonoscopy technical skill levels by use of an objective kinematic-based system. Gastrointestinal Endoscopy 2011;73(2):315–21.
[61] Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem. In: Grossberg S, editor. Proceedings of the IEEE First Annual International Conference on Neural Networks. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 1987. p. 11–4.
[62] Collins R, Liu Y, Leordeanu M. Online selection of discriminative tracking features. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005;27(10):1631–43.

Please cite this article in press as: Bell CS, et al. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. Artif Intell Med (2013), http://dx.doi.org/10.1016/j.artmed.2013.09.002