Survey of Computer Vision and Machine Learning in Gastrointestinal Endoscopy

Anant S. Vemuri

Index Terms—Computer assisted intervention, gastro-intestinal (GI) endoscopy, Barrett’s Oesophagus, biopsy relocalization, electromagnetic tracking, video synchronization.

This paper attempts to provide the reader a place to begin studying the application of computer vision and machine learning to gastrointestinal (GI) endoscopy. They have been classified into 18 categories. It should be noted by the reader that this is a review from pre-deep learning era. A lot of deep learning based applications have not been covered in this thesis.

I. ENDOSCOPIC APPLICATIONS

The clinical applications have been classified into the following 18 broad categories:

1) Polyp Detection and Classification (PD): All colorectal cancers (CRC) develop from dysplastic precursor lesions. This is true either in the presence of a predisposing factor such as in inflammatory bowel diseases (IBD) or lack thereof, with lesions occurring sporadically. Macroscopically the shape of lesions observed in the colon have been classified as described in [1], [2]. This class of application involves first detecting the polyps visible during colonoscopy and then presenting a classification based on their type. In [3] feature descriptor using the colour and pixel position in image is used for polyp detection using SVM and in [4], that feature is compared with Colour Wavelet covariance and LBP. In [5], [6], texture features using GLCM are compared with LBP using SVM for classification. [7] proposed using edges, followed by a hough transform of the image before using GLCM for texture features detection. They employed an adaboost classifier. In [8], firstly a model is defined for polyp appearance as enclosed by intensity valleys, including specular highlights and blood vessels to make the model robust. Using this a polyp localization energy map is generated, which is then used as an input for polyp segmentation. Also refer [9]–[11] for more details. [12], presented an algorithm, termed as the Classification of Regional Feature (CoRF), that is an extension of the sparse matrix and vector quantization for feature detection and segmentation. CoRF solves the intrinsic block selection problem of vector quantization by including training codebook about the shape of regional feature. They demonstrated that this approach works better for polyp detection and segmentation, as opposed to k-means or LGB clustering.

In [13], authors evaluated the discriminative power of image features extracted from sub-bands of the Gabor and the Dual-Tree Complex Wavelet Transform for the classification of zoom-endoscopy images. Further they also incorporated colour channel information and show, that this leads to superior classification results, compared to luminance-only based processing. Later, in [14], a colour wavelet cross co-occurrence matrix is proposed and use it to obtain statistical features for classification. This new wavelet-domain based colour texture feature extends the concept of classic co-occurrence matrices to capture information between detail sub-band pairs of different colour channels. The descriptor is then used for poly detection using a KNN classifier with Euclidean distance metric. Further work from the authors can be referred from [13]–[15]. In [16], an approach to poly classification is presented using vessel segmentation to extract, 22 features that describe the complex vessel topologies. Three feature selection strategies are compared with Simulated Annealing giving the best performance for polyp classification. [17], proposes a new descriptor by analysing shape of the connected components (blobs). The shape is described using convex hull, skeletonization, perimeter based features and contrast feature histograms for mucosal texture classification of polyps using Pentax iScan chromoendoscopy. The readers are referred to following notable works for further incite: [18]–[46] 

2) Ulcer Detection (UD): Oesophageal and gastric ulcers are caused as a result of GORD. In the colon ulcerative colitis (a type of IBD) occurs when the lining of the large intestine (colon) and the rectum become inflamed. This inflammation produces tiny sores called ulcers on the lining of the colon. It usually begins in the rectum and spreads upward.

The proposed method in [47] involves decomposition of images into components called as intrinsic mode functions, using bi-dimensional ensemble empirical mode decomposition. From the decomposition, two lacunarity based colour texture characteristics were obtained; the second and higher order correlation between intrinsic texture primitives and pixel intensity distribution. An SVM classifier is used on these features for UD.

In [48], an overview of the three image decomposition approaches is provided, a) Empirical mode decomposition or EMD ; b) Ensemble EMD and; c) Bidimensional EEMD , that provide the intrinsic mode functions (IMF). A differential Lacunarity (DLlac) metric is computed at each IMF and the responses matched with the characteristics of an ulcerated image. Those IMFs that are closely
related to the diseased condition are selected for reconstruction of the decomposed image. The DLac response vector computed earlier is used as the feature vector. Using this descriptor, classifier performance comparison between LDA, Quadratic discriminant analysis, NN using Mahalanobis distance and SVM was performed. In [49], on the other hand, authors use the lacunarity based colour texture features were used to investigate how the structural information of healthy and abnormal tissue is distributed on RGB, HSV and CIE Lab colour spaces. In [50], an HSV colour space feature histogram was used along with texture features extracted using the Contourlet transform and the Log Gabor filter, which were used to train an SVM classifier for UD. In [51], a curvelet based local binary pattern is proposed for texture feature extraction, to distinguish ulcer from normal regions, by training a multilayer perceptron neural network classifier.

Readers are referred to the following references: [28], [52], [52]–[61]

3) Celiac Disease Detection (CED): Celiac disease is an autoimmune disorder that can occur in genetically predisposed people where the ingestion of gluten leads to damage in the small intestine. During the course of Celiac disease, the mucosa loses its absorptive villi, leading to a strongly diminished ability to absorb nutrients. The gold standard for detection is based on extraction of biopsies from suspicious regions, which are identified during duodenoscopy using different imaging modalities. Computer aided detection methods to automatically mark suspicious regions during endoscopy have been widely explored in literature to decrease the miss-rates. [62], presents a CED approach by providing a comparison of classification between LBP, LTP, Multi-Fractal Spectrum, Dual-Tree Complex Wavelet Transform, Shape Curvature Histogram, Fisher vector and Vector of Locally Aggregated descriptors using a linear SVM classifier. They also provide a comparison under NBI, HD zoom endoscopy and standard white light endoscopy. Variants of DT-CWT are explored for automatic classification of endoscopic images using the Marsh classification, in [63]. The feature vector was composed of mean and standard deviations of the sub-bands from DT-CWT variant or Weibull parameter of the sub-bands. Enhanced scale invariance was obtained by applying DFT or DCT across the scale dimension of the feature vector. A k-NN classifier was used with leave-one-out cross-validation. In [64], [65], spatial domain (histogram) and transform domain (wavelet or Fourier) features are extracted from the images. A comparison between k-NN, SVM and bayes classifier is presented. The following references provide further details; [66]–[79].

4) Crohn’s Disease Detection (CRD): This is another kind of IBD, sometimes attributed to the aggressive immune response to harmless bacteria, by causing inflammation (normal immune system response), leading to chronic inflammation, ulceration, thickening of the intestinal wall, and eventually causing patient symptoms. CRD can occur anywhere from the mouth to the anus but most commonly observed in the ileum and beginning of the colon. [80], introduces a generic image matching methodology in presence of a complex scene by combining the output of multiple matchers using a single decision function. They provide a study on improving the SVM classifier performance under this framework. [81] provides the framework for lesion segmentation with application to Crohn’s disease.

5) Haemorrhoid and Bleeding Detection (HD and BD): Haemorrhoids are itching, painful or bleeding masses of swollen tissues and veins located in the anus and rectum. Bleeding on the other hand could be attributed to wide variety of reasons such as, Angiodysplasia (abnormalities in the blood vessels near the intestines), polyps, ulcers, Crohn’s disease, colon cancer, including haemorrhoids. Detection of bleeding thus is very important as it usually indicates a severe condition in the lumen. [82], use a descriptor comprising of the HSV histogram, dominant colour and texture features from the colour co-occurrence matrix. The dominant colour feature vector included, 8 representative colours, their variances and their percentages in the image. They propose a down-sampling strategy based on unsupervised clustering and probability driven sampling from each cluster to preserve the geometric structure while using fewer instances to train an ensemble of SVM classifiers. [83] present a study of all the MPEG-7 descriptors to determine the ones best suited for BD, UD and PD. Experiments indicated that the best results were obtained when using scalable colour and homogeneous texture descriptors, especially when only relevant coefficients are used using PCA. In [83], [84], an ANN classifier was trained using, texture features were extracted in RGB and HSV spaces. An alternate approach was proposed using CIE-LAB colour space with image covariance weighting. [85] introduced a clipped illumination invariant colour space, to compute an alternate binary feature vector, as opposed to the conventional colour histogram, by comparing similarity between local histograms instead of checking for the existence of a specified pattern. An SVM classifier is trained using this binary feature vector. In [86], pixels are grouped through a super-pixel segmentation and, for each super-pixel, the red ratio in RGB space is used as a feature descriptor, which is used to train an SVM classifier. [87] employed the statistical texture descriptors in the hue space to train a k-NN classifier. [88] defined an intrinsic colour model using YIO was proposed. This was used to extract statistical features to train an SVM classifier for BD. For further reading, please refer to the following references; HD - [89], [90] and BD - [56]–[58], [60], [91]–[107].

6) Oesophageal tissue Analysis (OA): There are two main types of oesophageal cancers; squamous cell cancer and oesophageal adenocarcinoma (OAC). Squamous cell cancer occurs most commonly in people who smoke cigarettes and drink alcohol excessively. Whereas, OAC occurs most commonly in people with gastro-oesophageal reflux disease (GORD). The latter condition has seen an
increase in frequency in the last two decades. GORD, a benign complication caused when the stomach acid escapes into the lower part of the oesophagus. As a chronic condition, it leads to changes in the oesophageal lining, causing the tissue to resemble the intestinal wall. This pathological condition is termed as Barrett’s oesophagus (BO). Several studies have indicated a direct link of BO with OAC. OAC appears to arise from the Barrett’s mucosa through progressive degrees of dysplasia [108], observed in the cells of the lower oesophagus.

The possibility of being able to perform staging of the precancerous tissue, provides room for early diagnosis and targeted treatments, avoiding emergency surgical interventions such as oesophagectomy. The literature reviews methods that include computer-aided detection of these conditions to aide diagnosis. [110] propose using heterogeneous descriptors computed from heterogeneous colour spaces. Instead of concatenating the descriptors to a super vector, a hierarchical heterogeneous descriptor SVM framework is proposed to simultaneously apply heterogeneous descriptors for GORD diagnosis and overcome the curse of dimensionality problem. [111] proposed a content-based image retrieval framework for detection of precancerous lesions in the oesophagus based on colour-texture analysis. The novelty of their approach lies in the interactive loop provided by a relevance feedback algorithm to improve detection accuracy. [112], presented a comparative evaluation of SVM, K-NN and boosting for detection of OA under NBI, WL and chromoendoscopy. [113] propose to train an SVM classifier using local colour and texture features, from the original and on the Gabor-filtered image. Based on the spectral characteristics of the cancerous tissue, specific filters were designed.

7) **Motility Detection (MD):** It is a term used to describe contraction of the muscles that mix and propel contents in the GI tract, with each of the four regions of the GI tract exhibiting specific characteristic movements and are separated by sphincter muscles and abnormal motility or sensitivity in any part of the tract can cause characteristic symptoms [114]. In [115], Laplacian of Gaussian filter is used to extract the lumen, then sum of the lumen area throughout the sequence of 9 frames which is compared with two certain thresholds empirically set with the help of the experts. Optical flow based, ego motion estimation is performed and a Relevance-Vector-Machine classifier is used on the ego-motion representation to extract images with motility. [116] tackles the problem learning a robust classification function from a very small sample set, when a related but unlabelled data set (for MD) is provided. In [117], at the first level of the system, each video was processed resulting in a number of possible contraction sequences. To encode the patterns of intestinal motility, a panel of textural and morphological features of the intestine lumen were extracted. In the second part, the final recognition of contractions sequences was carried out by means of a SVM classifier. [118], proposes a novel method based on anisotropic image filtering and efficient statistical classification of contraction features. In particular, the image gradient tensor was applied for mining informative skeletons from the original image and a sequence of descriptors for capturing the characteristic pattern of contractions. [119], [120] use linear radial patterns by means of the valleys and ridges detection. In this context, they propose descriptors of directional information using steerable filters. Self-organizing maps were used in general summarization for MD. Later, in [121], use textural, colour and blob features to train a classifier for MD. In [122], [124], propose two sets of features. First, motility based features in which, contractile activity characterization is performed using valley detection through use of Gabor-like filters. Then, the valley image is converted into a 1D signal representing the valley positions. Peak detection is performed that represent contractions in valley positions signal. Second, lumen perimeter estimation is performed, by applying mean-shift clustering to reduce noise in colour distribution. Then on grayscale image, thresholding is performed to segment the lumen. Morphological operators are then used for detection of smooth regions in the intestinal lumen. A combination of, histograms of SIFT Flow Directions to describe the flow course; SIFT descriptors to represent image intestine structure and; SIFT flow magnitude to quantify intestinal deformation, was proposed in [125].

8) **Endoscopic Abnormality Detection and Classification (ABD):** This is a broad category that encompasses, all the kinds of lesions or abnormalities that cannot be classified clinically, in any of the above mentioned classes. The methodologies presented here do not focus on any specific disease condition but aim to differentiate a normal tissue from abnormal one. [126] provides a review of various feature descriptors used in lesion detection in colonoscopic videos. [127] proposed, textural analysis of the different colour channels, using the wavelet transform to select the bands containing the most significant texture information. Later, in [128], the texture descriptors from co-occurrence matrix at two different scales was used in conjunction with second and higher order moments from the GLCM computed from the image recovered using specific selected scales of the wavelet decomposition of the original image as descriptors. In [129], statistical textural descriptors were computed taken from the Discrete Curvelet transform of the image in multiple directions and scales. The covariance of texture descriptors is used as the final feature vector. [130], performed a comparison between descriptors obtained from wavelet decomposition and discrete curvelet transform. In each case an ANN classifier was trained using the described feature vector. [131] proposed using image patches in the BoW model generated using a random forest based clustering which were used to train an SVM classifier. In [132], [133], colour histogram statistics were computed for images in R,G,B,H,S,V channels of the WCE images. Additionally, a local texture information was collected for each pixel.
by using a LTP and labelled as part of a texture unit. This complete information vector is used for classification using a neural network trained using the Bayesian ying-yang method to maximize entropy. 

used image level annotations to learn a set of online local features for adenoma detection in patches extracted in images. The BRISK based spatial structure is used for sampling pixels for learning visual descriptors. proposed an extended Gaussian filtered LBP descriptor, robust to illumination changes, noise. The algorithm is claimed to be able to capture more informative edge-like features. proposes a new method to choose a subset of cluster pairs based on the idea of Latent Semantic Analysis (LSA) and proposes a new inter-cluster statistics which captures richer information than the traditional co-occurrence information. In authors present two schemes. The first, working on the full-resolution image, the second on a multi-scale pyramid space. With this framework any feature descriptor could be employed; but a multi-resolution LBP was tested. In Root-SIFT and a multi-resolution local patterns descriptors were extracted from image patches, for each colour channel. For complete set of references, readers are referred to the following list: 9) Endoscopic Navigation (NAV) and 6-DOF Localization (LOC): Navigation refers to, using the current endoscopic image information, for determining where to go next. In some ways it charts the path ahead for endoscope. Whereas, localization uses the data from previous few seconds to estimate the current pose or anatomical location of endoscope in the GI tract. This information could be in two forms; as knowledge of the section of GI tract, such as oesophagus, stomach, duodenum, ileum etc., determined by classifying the tissue structure; or secondly, by estimating the complete endoscopic motion to obtain the 6-DOF pose of the endoscope.

modelled the colon as a cylinder. By estimating the camera motion parameters between each consecutive frame, circumferential bands from the cylinder of the colon surface were extracted. Registering these extracted band images from adjacent video frames provided a visibility map, that could reveal unexplored areas by clinicians from colonoscopy videos. proposed, learning the pose from the optical flow fields in WCE images. Feature descriptors were generated using lumen centred and grid based methodology. ANN was used to evaluate the strength of descriptors extracted from WL and NBI images. In authors propose extraction of SURF features and use RANSAC based matching to estimate homography between consecutive frames to provide navigational help. proposes to use lumen detection for image-guided visual servoing in endoscopy. For NAV application, the readers are invited to review the following additional references propose to use, MPEG-7 features along with vector quantization and PCA for descriptor compression. A neural network was trained to classify different section of the GI tract using the computed features. 

propose multi-scale elastic registration of consecutive frames of the WCE and extraction of projective geometry to determine the pose of the capsule endoscope. employed Gabor filter based texture descriptors to detect duodenum in WCE video stream. For the paper proposes a roll angle estimation for complete 6-DOF pose recovery. proposes a hybrid tracking method of WCE motion, integrating magnetic sensing and image-based localization. presents an approach to use the intestinal motility to localize the endoscope. These additional references complete the list for NAV in literature:

10) Intra and Inter-Operative Re-localization (IAO and IRO): The IAO based approaches focus on detecting, tracking and localizing biopsy sites during a single procedure. Primarily, these approaches focus on BSR. One of the first methods in IAO re-localization, was published by Allain et al. In their approach, the authors proposed to compute feature points in scale-space around the biopsy location and then extracted descriptors for these points using scale invariant feature transform (SIFT) for the two endoscopic views to be matched. Then employing the epipolar constraint, a fundamental matrix was computed between the two views, that mapped the biopsy site to facilitate re-targeting. In a framework for characterizing and propagation of the uncertainty in the localization of the biopsy points was presented. Mountney et al. performed a review of various feature descriptors applied to deformable tissue tracking and in proposed an Extended Kalman filter (EKF) framework for simultaneous localization and mapping (SLAM) based method for feature tracking in deformable scene, such as in laparoscopic surgery. This EKF framework was then extended for maintaining a global map of biopsy sites for endoluminal procedures, intra-operatively. The authors presented an evaluation of the EKF-SLAM on phantom models of stomach and oesophagus. Giannarou et al. presented an affine-invariant anisotropic region detector robust to soft tissue deformations. This was used along with SIFT descriptors. The feature matching problem was then modelled as a global optimization of an Markov Random Field (MRF) labelling. Recently, Ye et al. accomplished the biopsy site re-targeting in three stages. First using the Tracking-Learning-Detection (TLD) method proposed by Kalal et al. TLD was used for tracking multiple regions around the selected biopsy site. Under the assumption that the regional tissue deformations can be approximated using local affine transformations, a local homography between matched region centres was estimated. In this way multiple regions around the biopsy sites are tracked, which were then used for homography estimation and mapping the biopsy sites. Wang et al. proposed to learn a graph (atlas) from a sequence of images from several gastroscopic interventions. Considering that the stomach’s deformation as not being large between similar frames the nodes of the learnt graph atlas were connected by an estimated rigid transformation. Thus, the mapping
of the biopsy sites from a single (reference) frame to subsequent frames for any given intervention was reduced to a graph search problem. Firstly, for the reference frame and the moving frame their corresponding matching nodes in the graph were computed. Using Dijkstra’s algorithm, the shortest path between these matched nodes was obtained. Hence, the transformation between the reference frame and moving frame was obtained as the associated combination of rigid transforms along the shortest path between the corresponding matched nodes of the graph.

In contrast, the IRO methods attempt to provide localization between interventions. [242] proposed the use of electromagnetic tracking system (EMTS) for localizing the biopsy sites in the stomach. They construct a 3D model of the stomach using SLAM and map the biopsy points tracked using the EMTS on to the 3D model. The inter-operative registration was performed by selecting five reference points manually, during each intervention. In [243], [244], Atasoy et al. proposed to formulate the relocalization as a image-manifold learning process. The method involved firstly, building an adjacency graph between the images of a surveillance intervention. Normalized cross-correlation was used as the similarity measure between image frames to compute the adjacency graph. Then using laplacian eigenmaps decomposition that was proposed in [245], a linear projection matrix was computed. This approximation for projection on to the manifold was used to compute the low-dimensional representation for all the images in the intervention. Then, two separate methodologies for performing inter-operative re-localization was proposed using scene association. In [244], the scene association is performed by computing the nearest neighbour directly over the low-dimensional representation from an earlier surveillance endoscopy. However, in [243] a two-run surveillance endoscopy was suggested, in which a dummy surveillance is performed before, that was used for scene association with the actual surveillance. The authors claimed that the modified approach in [243] allowed for scene association in presence of significant structural changes in the tissue. For colonoscopic procedures, the need to provide navigational assistance is substantial. One of the earliest approaches involved combination of 3D reconstruction from pre-operative CT with endoscopic video known as virtual colonoscopy. The chief aspect of it involved computation of optical flow to estimate the ego-motion of the colonoscope. Ego-motion or visual odometry involves firstly, extracting features from the image and computing optical flow fields. Then, using the flow fields, the camera motion would be estimated. In [246] the authors presented a comparison of two ego-motion estimation schemes, supervised and unsupervised. Supervised methods, as shown in [192] require training data to be available in the form of optical-flow measurements and corresponding camera motion data. Unsupervised approaches, however, used image correspondences between video frames and multiple-view geometry to estimate endoscope motion, as was shown in [247]. Theoretically, these methods can be applied to oesophageal procedures as well. The first endoscopy can be used to obtain a 3D reconstruction of the oesophagus and can be used in the follow-up surveillance procedures. But, the video based 3D reconstruction in GI procedures is still an open area for research. However, an additional pre-operative imaging such as CT can be used for the reconstruction of the oesophagus. Due to which, such methods were not cost-effective and aren’t used as part of routine procedures.

11) **Lumen Detection (LD):** By itself LD can be employed for NAV, LOC, MD etc. [248] proposes, global thresholding, followed by a differential region growing using dynamic hill clustering optimization to extract the lumen. In [249], Haar like feature combined with adaboost were used to select the most discriminative features. Then, a boosted cascade of classifiers was employed for lumen detection. Otsu thresholding was employed for segmenting darker regions of the image in [250]. A pyramidal structure of binarized images was constructed and from the smallest image, the region seed is grown back to the original image resolution to detect the lumen. In [251], the proposed method is based on the appearance and geometry of the lumen, which we defined as the darkest image region whose centre is a hub of image gradients. In [252], the proposed technique applied the Otsu’s procedure recursively to obtain a coarse ROI, which is then subjected to an Iris filter operation so that a smaller enhanced region can be identified. The enhanced region was then subjected to the Otsu’s procedure recursively and the process of performing Iris filter operation repeated. [253] developed a deformable region model approach to extract lumen from the endoscopic image by giving an approximate boundary plan of the lumen using minimum cross-entropy algorithm, that was then deformed to the compute the real boundary automatically.

12) **Uninformative Frame or Region Detection (UI):** Section sec:challenges had earlier presented a description of what constitutes an UI frame. It is important to note in this context that any endoscopic frame need not be completely informative or entirely UI. Thus, some methods proposed in literature also try to identify the UI regions. [254] propose UI region detection using a multi-stage approach with Chan-Vese segmentation, color range ratio, adaptive gamma correction (AGCM), and finally using canny colour edge detection operator with morphological processing. [255], propose using texture analysis of image DFT and use k-means clustering to classify UI frames. [256] propose to use L2 norm of DWT decomposition as features given to a Bayesian classifier. In [257], the local colour moments in Ohta space, along with HSV colour histogram were used as features to train an SVM classifier in the first stage of UI frame removal. In the second phase, the Gauss laguerre transform based multi-resolution decomposition was performed and the responses were thresholded. The authors also present a comparison with Gabor and wavelet based descriptors. In the methods proposed by [258], two values are computed
over a grid on the image; a) Dark Region Identification (DRI) using convolution with gaussian kernel. b) Directed Gradient Accumulation (DGA) . A UI region is then defined by low(DRI) and high(DGA). [259] proposed to perform, watershed segmentation followed by morphological closing and Frontier based region merging. After the first merging, region-based merging is performed using mean grey value to threshold over a sliding window. Five empirically chosen region profiles were used for thresholding. In [260], proposed approach involves, lumen detection based on mean shift and evaluation of coherent motility for selecting informative frames. [261], use texture feature extracted from bank of Gabor filters with a feed-forward neural network for UI classification. [262] propose three methods for UI region detection in WCE frames, using feature extracted from morphological operations, statistical features and Gabor filter based features in HSV colour space. Fuzzy k-means, Fisher test and neural network based discriminators were used. The following references give additional methods from this category proposed in literature: [263]–[270].

13) Specular Highlight Detection and Removal (SHD): Although, specular highlights in the image constitute UI regions, this particular category of methods attempt to not only identify such regions, but also correct them. In [271], specular highlights is addressed using:a segmentation method based on non-linear filtering and colour image thresholding followed by a fast in painting method. The proposed method in [272], aims to decouple the specular and diffuse components of endoscopic imagery in order to suppress specular reflectance. A stochastic Bayesian estimation approach is introduced to estimate the specular component of endoscopic imagery. A Monte-Carlo sampling of image regions is performed for computing posterior probability. [273], describe a specularity removal framework using a Dichromatic Reflection Model (DRM) and multi-resolution inpainting technique to obtain the corrected region.

14) Endoscopic Reconstruction (REC): 3D Reconstruction in flexible endoscopic procedures is quite a challenging task. Apart from the already discussed, UI frames, presence of repeatable features in a deformable environment poses significant difficulties, if overcome, can aide in assisted diagnosis, pre-operative planning and post-operative review. The feature detectors that were discussed in ?? are used frequently used to recover the 3D from images. In [274], the tracked feature points are used for estimating camera parameters and providing an estimate of the polyp size. [275] proposed to use SIFT features using normalized SSD based monoSLAM for 3D reconstruction of the oesophagus. [276] employed Shi-Tomasi features and used them in shape from shading framework for reconstruction. [277], [278] proposed to use, affine invariant version of SIFT detector and descriptor to estimate the epipolar geometry and recover the 3D. In [279], [280], edges of colon fold contours were first detected and processed to generate the wire frame of the reconstructed virtual colon. A colon fold contour estimation algorithm using a single colonoscopy image was proposed and the depth and shape estimation of colon folds using brightness intensity of pixels was introduced. In [281], shape-from-shading was used to reconstruct polyps for better recognition. [282] describes an approach to perform a gastric panorama by visual tracking. [283] proposes a structure from motion based method that takes advantage of a 6-DOF tracking device that is used to record the endoscope’s position during a procedure. After feature tracking, a space constraint strategy is applied to remove the outliers and recover the missing data. In [45], a method to reconstruct the 3D texture surface of the GI tract using single WCE image using Shape from Shading technique is presented. [284] used, a circular generalized cylinder as a basis for 3D reconstruction of the GI tract. The model was decomposed as a series of 3D circles and a MRF framework was proposed to maximize the a posteriori estimation. In [216], a 3D model and panoramic view are incorporated into the navigation system with three improvements: selection of reference and tracking of features; perspective projection for constructing local and global panoramic view. 3D surface modelling is performed using structure from motion. The system was evaluated for three clinic applications: broadening the endoscopic view, performing non-invasive re-targeting, and determining the overall lesion locations. [285] proposes to use lumen detection in WCE to create a 3D map using inertial information from the WCE trackers.

15) Endoscopic Image Enhancement (IE): This category refers to a class of approaches directed towards pre-processing steps to improve the quality of visible image and feature response. In [286], authors evaluate different reconstruction-based super-resolution algorithms in order to enhance the spatial resolution of endoscopic images acquired with an HD endoscope and to determine the its feasibility to study fine mucosal structures in HD endoscopy. To overcome the rather dark WCE images a adaptive contrast diffusion filtering is proposed in [288]. [287] a colour enhancement of WCE frames is proposed to obtain robust texture based features. [288], proposes use of Homomorphic filtering and [29] describe an adaptive anisotropic diffusion pre-processing for image enhancement before feature extraction.

16) Endoscopic Video Summarization (ES): This is primarily a category ascribed to wireless capsule endoscopy. Due to the large volume of frames to analyse methods have been developed to minimize this time using different methods. [289] proposes a new fast spatio-temporal technique that detects an operation scene a video segment corresponding to a single purpose diagnosis action or a single purpose therapeutic action. In [270], an approach is presented to segment WCE video. To accomplish this, firstly, colour and wavelet texture features are used to denote UI regions. Then boundaries between adjacent organs of WCE video are estimated in two levels. At course level, colour feature is utilized to draw a dissimilarity curve between
frames and the aim is to find the peak of the curve, which represents the approximate boundary. At the fine level, Hue-Saturation histogram colour feature in HSV colour space and uniform LBP texture feature from grayscale images are extracted. These features are used to train an SVM classifier for video segmentation. In [290], a two step approach to summarization is proposed. The first step consists of a semi-supervised clustering and Local Scale Learning (SS-LSL) algorithm. This algorithm is used to group video frames into prototypical clusters that summarise the CE video with constraints that are deduced from the training frames. The second step consists of a novel relational motion histogram descriptor that is designed to represent the local motion distribution between two contiguous frames. [291] proposes the use of textons for classifying video segments corresponding to different regions in the GI tract. [292] reviews various colour and texture descriptor for WCE image analysis. The segments of constant intestinal activity are detected with a robust statistical test that is based on Hoeffding’s inequality in [293]. [294] propose using HSV histograms compressed using a combination of DCT and PCA for for identifying different regions in the WCE video. [295] proposes a hierarchical key frame extraction algorithm based on a saliency map to automatically select a small number of key informative frames. [296] propose a method that is based on clustering using symmetric non-negative matrix factorization, initialized by the fuzzy c-means algorithm and supported by non-negative Lagrangian relaxation, to extract a subset of video scenes containing the most representative frames from an entire examination. [297], [298] propose using SURF feature points from consecutive frames, and RANSAC based matching to estimate a homography between consecutive frames for fast video browsing of WCE. An unsupervised k-window clustering is presented in [158] to cluster video frames. Each cluster is trained on a different neural network for summarization. [299] describes a novel colour-texture feature to describe the contents of the frame in a WCE video. Spectral clustering is applied to segment a WCE video into meaningful parts via shot boundary detection using the extracted features. The following references have not been detailed here: [300]–[315].

17) Segmentation of Specific Tissues (IAS): [316] propose a three step approach to segmentation of WCE image frame. a) Local polynomial approximation algorithm which finds locally-adapted neighbourhood of each pixel; b) Colour texture analysis which describes each pixel by a vector of numerical attributes that reflect this pixel local neighbourhood characteristics; c) Performing k-means clustering based on the colour feature vector. For chroendoendoscopy and NBI imaging, [317] describes the usage of various visual features individually and in combinations (edgemaps, creaseness, and color), in normalized cuts image segmentation framework. [318] describes an approach to segmenting bubbles in colonscopic images.

18) Clinical Decision Support (CDS): The methods described in this category discuss approaches to build a generic tool for clinicians to provide decision support. The references described here do not target a specific disease category, but rather a system for detecting lesions and categorizing them to determine the disease type. This category can also be classified as a content based image retrieval platform, which essentially uses computer vision concepts for retrieving similar images from database to aide clinical diagnosis.

In the method proposed by [319], images are transformed to CIE-LAB space. Non-subsampled contourlet transform is used to decompose the chromaticity and intensity components, representing colour and texture features. The decomposed sub-bands are modelled using Generalized Gaussian Density using a ML estimator. The resulting feature vector is then compressed using PCA. Using Least Square-SVM to perform pre-classification, which is followed by computing the kullback-leibler divergence between the features of the query image and the database. [320], proposes using GLCM, colour histogram, GIST and Gabor, wavelet, Maximum response (MR8), Leung-Malik (LM) filter bank, and the Schmid Filter banks responses as feature descriptors for image retrieval using a naive Bayes Nearest neighbour classifier. [321] presents a review of various colour and texture descriptors in to retrieve the types of frames in the endoscopic scene. A comparison is drawn using these feature descriptors by training using ans SVM and an ANN classifier. [322], proposes computing a 10-bin normalized hue and saturation histograms and training a SVM classifier for retrieving the class of the tissue in the ROI. [323] proposes representing the localized features in HD endoscopy images in semantic space to generate a CBIR system for clinicians to review online selected regions. In [324], using texture features extracted from DT-CWT, the authors propose a generative model based strategy closely related to CBIR for online tissue classification. [325] proposes a novel approach to the design of a semantic, low-dimensional, encoding for endoscopic imagery. [326] discusses the development of a platform for image annotation and retrieval in GI endoscopy. A CBIR system is presented in [111], for identifying precancerous lesions in the oesophagus based on color-texture analysis. [327], explores local features (extracted by using sampling schemes such as Difference-of-Gaussians and grid sampling), BoW, and provides extensive experiments on a variety of technical aspects for feature description. For the CBIR system, an SVM classifier is investigated and its performance under different kernel types, sampling strategies for the local features, the number of classes to be considered etc. is studied. [328] reviews various feature descriptors and classifications methodologies for WCE images. In [43], authors propose combining information from multiple images, to design a supervised classification approach using an hidden markov model (HMM) framework. This framework, is prototyped with weak (k-NN) classifier to evaluate its performance for regions of the GI tract containing polyps. [329] proposes the use of
colour features in the form of Hue-Saturation Histograms and texture as SVD of local regions to develop a CBIR system for detecting Pylorus valve between stomach and intestine in WCE.

In the thesis, authors from the decomposition of an image using the Hilbert Huang Transform, selected modes were compressed using PCA to generate a representative feature vector. In LBP and its variants were used firstly, for uninformative frame removal and then in the detection of lesions developed due to Celiac disease, Crohn’s disease, intestinal polyps and tumours. proposes a software system that uses various colour and texture features, combined into a single feature vector. Then a feature selection model is presented, that uses, Deep Sparse SVM (DSSVM) which assigns a suitable weight to the feature dimensions like the other traditional feature selection models and directly excludes useless features from the feature pool. presents an intelligent system for online endoscope image analysis. The method discusses extraction of texture features in chromatic and achromatic domains from histograms of each colour component to train an ANN. proposes the use of Hue-Saturation histogram in combination with LBP to classify precancerous and cancerous lesions in multi-spectral imaging. A comparison between Logistical model trees, Naive Bayes, NN and SVM classifiers was made. In authors study an adaptive texture classification strategy to achieve robustness to varying degrees of degradation in training images. The papers also discuss various similarity measures in this context. presents a comparison of various texture based feature descriptors; LBP and variants, multi-fractal spectrum, edge co-occurrence matrix and local phase quantization to train an SVM classifier. Approach to achieve blur invariance through blur-equalization has also been studied in the context of Celiac disease classification.

In scale invariant features are extracted from different variants of the DT-CWT of image, in order to classify high-magnification colon endoscopy imagery with respect to the pit pattern scheme. To enhance the scale invariance, the DCT is applied to the feature vectors. The final descriptor contains either consist of the means and standard deviations of the subbands from a DTC-WT variant or of the Weibull parameter of these sub-bands. Readers are referred to review further publications by Hafner et al to study the various use of spatial frequency domain descriptors presents a modified version of LBP descriptor over individual colour channels to train a NN classifier using Bhattacharyya distance presents a comparison of four cross-validation approaches leave-one-image-out, leave-one-parent-image-out, leave-one-lesion-out and leave-one-patient-out for colon polyp classification. discusses the application areas for decision support in GI endoscopy and presents a review of various features for detection of adenomas in video endoscopy. presents a CDS that uses geometrical and colour features from the endoscopic image. The thesis , provides an analysis on the detection of various generic scene categories in the colonoscopy videos. discusses usage of edge features and the extraction of most discriminative subsets using a greedy feed forward selection. The descriptors are used with a NN classifier to detect various scenes in GI endoscopy. The reader is invited to the study following references for further incite:

This completes the review of scene understanding and classification in GI endoscopy. Although the target application in this thesis has been the oesophagus, a review of the complete GI anatomy was performed since, no such comprehensive review was encountered in literature and a clear understanding of the application domains was felt necessary. Firstly, a description of the various state of the art algorithms was performed, to highlight the key approaches. Then, a classification based on endoscopic application was performed. 18 categories were identified ranging from disease-specific cases, such as CED and CRD, to generalized CDS systems. Application domain of reconstruction, navigation and localization form important parts of an intelligent support systems and hence have also been reviewed. The secondary aim of this categorization was to elucidate the strategic thinking observed in biomedical community, on transfer of technology to GI clinical domain.

References

[1] L. Laine, T. Kaltenbach, A. Barkun, K. R. McQuaid, V. Subramanian, and R. Soetikno, “Scenic international consensus statement on surveillance and management of dysplasia in inflammatory bowel disease,” Gastroenterology, vol. 148, no. 3, pp. 639–651, 2015.

[2] H. Inoue, H. Kashida, S. Kudo, M. Sasaki, T. Shimoda, H. Watanabe, S. Yoshida, M. Guelrud, C. Lightdale, K. Wang et al., “The paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to december 1, 2002,” Gastrointestinal Endoscopy, vol. 58, no. 6 Suppl, pp. S3–S43, 2003.

[3] L. a. Alexandre, J. Casteloire, and N. Nobre, “Polyp detection in endoscopic video using SVMs,” PKDD 2007 Proc. 11th European conference on Principles and Practice of Knowledge Discovery in Databases, vol. 4702, pp. 358–365, 2007.

[4] L. a. Alexandre, N. Nobre, and J. Casteloire, “Color and position versus texture features for endoscopic polyp detection,” Biomedical Engineering and Informatics: New Developments and the Future - Proc. 1st International Conference on Biomedical Engineering and Informatics, BMEI 2008, vol. 2, pp. 38–42, 2008.

[5] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, “Texture-based polyp detection in colonoscopy,” in Bildverarbeitung für die Medizin 2009. Springer, 2009, pp. 346–350.

[6] S. Ameling, S. Wirth, D. Paulus, and Others, Methods for Polyp Detection in Colonoscopy Videos: A Review. Inst. für Computervisualistik, 2009.

[7] Q. Angermann, A. Histace, O. Romain, X. Dray, A. Pinna, and B. Granado, “Smart Videocapsule for Early Diagnosis of Colorectal Cancer: Toward Embedded Image Analysis,” in Computational Intelligence in Digital and Network Designs and Applications, ser. Part 2: Digital, Network Designs and Applications, Chapter 12. Springer, May 2015, p. 25.

[8] J. Bernal del Nozal, “Polyp Localization and Segmentation in Colonoscopy Images by Means of a Model of Appearance for Polyps,” Ph.D. dissertation, Universitat Autònoma de Barcelona, 2014.

[9] J. Bernal, J. Sánchez, and F. Vilarino, “Towards automatic polyp detection with a polyp appearance model,” Pattern Recognition, vol. 45, no. 9, pp. 3166–3182, 2012.

[10] J. Bernal, J. M. Núñez, F. J. Sánchez, and F. Vilarino, “Polyp Segmentation Method in Colonoscopy Videos by Means of MSA-DVOA Energy Maps Calculation,” in Clinical Image-Based Procedures. Translational Research in Medical Imaging. Springer, 2014, pp. 41–49.

[11] J. Bernal, F. J. Sánchez, J. Fernández-Esparza, D. Gil, C. Rodríguez, and F. Vilarino, “WM-DVOA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians,” Computerized Medical Imaging and Graphics, vol. 43, pp. 99–111, 2015.
[12] H. C. Wang, W. M. Chen, Y. P. Lin, and W. C. Shen, “Tumor detecting in colonoscopic narrow-band imaging data,” in Proc. International Symposium on Intelligent Signal Processing and Communications Systems. IEEE, 2012, pp. 564–568.

[13] R. Kwitt and A. Uhl, “Multi-directional multi-resolution transforms for zoom-endoscopy image classification,” in Advances in Soft Computing. Springer, 2007, vol. 45, pp. 35–43.

[14] R. Kwitt and A. Uhl, “Color wavelet cross co-occurrence matrices for endoscopy image classification,” in Proc. 3rd International Symposium on Communications, Control, and Signal Processing (ISCCSP). IEEE, 2008, pp. 715–718.

[15] R. Kwitt and A. Uhl, “Color eigen-subband features for endoscopy image classification,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2008, pp. 589–592.

[16] S. Gross, S. Palm, J. J. W. Tischendorf, A. Behrens, C. Trautwein, and T. Aach, “Automated classification of colon polyps in endoscopic image data,” in Proc. of SPIE Medical Imaging, vol. 8315. International Society for Optics and Photonics, 2012, pp. 83 150W–83 150W–8.

[17] M. Hafner, A. Uhl, and G. Wimmer, “A novel shape feature descriptor for the classification of polyps in HD colonoscopy,” in Medical Computer Vision. Large Data in Medical Imaging. Springer, 2014, pp. 205–213.

[18] J. Ayoub, B. Granado, Y. Mhanna, and O. Romain, “SVM based colon polyps classifier in a wireless active stereo endoscope,” in Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2010, pp. 5585–5588.

[19] S. H. Bae and K.-J. Yoon, “Polyp Detection via Imbalance d Learning with unlabeled regions for NBI image recognition,” in International Conference on Pattern Recognition (ICPR). IEEE, 2012, pp. 25–28.

[20] L. R. P. Viana, Y. Iwahori, K. Funahashi, and K. Kasugai, “Automated Polyp Detection from Endoscopy Images,” in Proc. 6th International Conference on Soft Computing and Intelligent Systems and 13th International Symposium on Advanced Intelligent Systems, 2012, pp. 2272–2275.

[21] Y. Wang, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen, “Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy,” IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4, pp. 1379–1389, 2014.

[22] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, “Polyp-Alert: Near real-time feedback during ‘colonoscopy.’” Computer methods and programs in biomedicine, vol. 120, no. 3, pp. 164–179, 2015.

[23] Y. Yuan and M. Q.-H. Meng, “Improved Bag of Feature for Automatic Polyp Detection in Wireless Capsule Endoscopy Images,” Automation Science and Engineering, IEEE Transactions on, vol. PP, no. 99, pp. 1–7, 2015.

[24] Y. Yuan and M. Q.-H. Meng, “A novel feature for polyp detection in wireless capsule endoscopy images,” in Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2014, pp. 5010–5015.

[25] Q. Zhao, T. Dassopoulos, G. Mullin, G. Hager, M. Q.-H. Meng, and R. Kumar, “Towards integrating temporal information in capsule endoscopy image analysis,” in Proc. Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS). IEEE, Aug 2011, pp. 6627–30.

[26] M. Q. Meng, “Polyp Detection in Wireless Capsule Endoscopy Images Using Novel Color Texture Features,” in Intelligent Control and Automation (WCICA), 2011 9th World Congress on. IEEE, 2011, pp. 948–952.

[27] Q. Zhao and M. Q.-H. Meng, “3D reconstruction of GI tract texture surface using Capsule Endoscopy Images,” in Proc. of IEEE International Conference on Automation and Logistics. 2012, pp. 277–282.

[28] Z. Zhou, G. Bao, Y. Geng, B. Alkandari, and X. Li, “Polyp detection and radius measurement in small intestine using video capsule endoscopy,” in Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on. IEEE, 2014, pp. 237–241.

[29] V. Charisio, L. Hadjiileontiadis, and G. Sergiadiis, “Lacunarity-Based Inherent Texture Correlation Approach for Wireless Capsule Endoscopy Image Analysis,” in IFMBE Proc., vol. 41. Springer, 2014, pp. 297–300.

[30] V. S. Charisi, L. J. Hadjiileontiadis, C. N. Liatos, C. C. Mavrogianis, and G. D. Sergiadiis, “Capsule endoscopy image analysis using texture information from various colour models,” Computer Methods and Programs in Biomedicine, vol. 107, no. 1, pp. 61–74, 2012.

[31] V. Charisi, A. Tsiligiri, L. J. Hadjiileontiadis, C. N. Liatos, C. C. Mavrogianis, and G. D. Sergiadiis, “Ulcer Detection in Wireless Capsule Endoscopy Images Using Bidimensional Nonlinear Analysis,” in Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2010, pp. 236–239.
D. Mitrea, P. Mitrea, R. Badea, and M. Socaciuc, “Computerized methods for the assessment and characterization of the inflammatory bowel diseases and colon cancer from ultrasound and endoscopic images,” in *Proceedings of 10th WSEAS International Conference*. World Scientific and Engineering Academy and Society (WEAS), 2011, pp. 336–343.

R. Miyaki, S. Yoshida, S. Tanaka, Y. Kominami, Y. Sanomura, T. Matsuo, S. Oka, B. Raytchev, T. Tanaki, T. Koide, K. Kaneda, M. Yoshihara, and K. Chayama, “Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement,” *Journal of Gastroenterology and Hepatology*, vol. 25, no. 5, pp. 841–847, 2013.

R. D. Nawarathna, “Detection of temporal events and abnormal images for quality analysis in endoscopy videos,” Ph.D. dissertation, University of North Texas, 2013.

M. S. Neofytou, M. S. Pattichis, C. S. Pattichis, V. Tanos, E. C. Kyriacou, and D. D. Koutsouris, “Texture-based classification of hysteroscopy images of the endometrium,” in *Proc. Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS)*. IEEE, 2006, pp. 3005–3008.

V. B. S. Prasath, I. N. Figueiredo, P. N. Figueiredo, and K. Palaniappan, “Mucosal region detection and 3D reconstruction in wireless capsule endoscopy videos using active contours,” in *Proc. Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS)*, vol. 2012. IEEE, 2012, pp. 4014–7.

V. B. S. Prasath and R. Delhibabu, “Automatic Image Segmentation for Video Capsule Endoscopy,” in *Computational Intelligence in Medical Informatics*. Springer, 2015, pp. 73–80.

G. A. Puerto-Souza, S. Manivannan, M. Trujillo, J. Hoyos, E. Trucco, and G. Mariotti, “Enhancing normal-abnormal classification accuracy in colonoscopy videos via temporal consistency,” in *Proc. Computer-Assisted and Robotic Endoscopy Workshop, MICCAI*, 2015.

F. Riaz, M.-D. Ribeiro, P. Pimentel-Nunes, and M. Tavares Coimbra, “Integral scale histogram local binary patterns for classification of narrow-band gastroenterology images,” in *Proc. Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS)*, vol. July 2013, pp. 3714–3717.

F. Riaz, M. Areia, F. B. Silva, M. Dinis-Ribeiro, P. P. Nunes, and M. Coimbra, “Gabor textons for classification of gastroenterology images,” in *Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*. IEEE, Mar 2011, pp. 117–120.

F. Riaz, F. B. Silva, M. D. Ribeiro, and M. T. Coimbra, “Invariant Gabor Texture Descriptors for Classification of Gastroenterology Images,” *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 10, pp. 2893–2904, 2012.

X. Shen, K. Sun, S. Zhang, and S. Cheng, “Lesion detection of electronic gastroscopy images based on multiscale texture feature,” in *IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)* 2012). IEEE, 2012, pp. 756–759.

R. Sousa, D. C. Moura, M. Dinis-Ribeiro, and M. T. Coimbra, “Local Self Similar Descriptors : Comparison and Application to Gastroenterology Images,” in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 4635–4638.

K. Sun, Y. Wu, X. Lin, S. Cheng, Y. M. Zhu, and S. Zhang, “Mean shift-based lesion detection of gastroscopic images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, 2012, vol. 7202 LNCS, pp. 167–174.

Kai Sun, S. Zhang, R. Yao, W. Yang, S. Cheng, and S. Zhang, “Lesion detection of gastroscopic images based on cost-sensitive boosting,” in *2011 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE, Sep 2011, pp. 1–6.

Z. Sun, B. Li, R. Zhou, H. Zheng, and M. Q. H. Meng, “Removal of Non-Informative Frames for Wireless Capsule Endoscopy Video Segmentation,” in *IEEE International Conference on Automation and Logistics, ICAL Aug 2012*, pp. 294–299.

D. Surangsrirat, M. Tapia, and W. Z. W. Zhao, “Classification of endoscopic images using support vector machines,” in *IEEE SoutheastCon 2010 (SoutheastCon)*. *Proc. of the*. IEEE, 2010, pp. 436–439.

P. M. Szczypiński, R. D. Srima, P. V. Srima, and D. N. Reddy, “A model of deformable rings for interpretation of wireless capsule endoscopic videos,” *Medical Image Analysis*, vol. 13, no. 2, pp. 312–324, 2009.

P. Szczypiński and A. Klepaczko, “Automated recognition of abnormal structures in WCE images based on texture most discriminative descriptors,” in *Advances in Intelligent and Soft Computing*. Springer, 2010, vol. 84, pp. 263–270.

P. Szczypiński, A. Klepaczko, and M. Strzelecki, “An intelligent automated recognition system of abnormal structures in wce images,” in *Proc. 6th International Conference on Hybrid Artificial Intelligent Systems - Volume Part 1*, ser. HAIS’11. Springer, 2011, pp. 140–147.

M. P. Tjoa and S. M. Krishnan, “Feature extraction for the analysis of colon status from the endoscopic images,” *BioMedical Engineering Online*, vol. 2, no. 1, p. 9, 2003.

A. Uhl, G. Wimmer, and M. Hafner, “Shape and size adapted local fractal dimension for the classification of polyps in HD colonoscopy,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*. IEEE, Oct 2014, pp. 2299–2303.

H. Vu, T. Echigo, Y. Imura, Y. Yanagawa, and Y. Yagi, “Segmenting Reddish Lesions in Capsule Endoscopy Images Using a Gastrointestinal Color Space,” in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 3263–3268.

H. Wang, D. Chen, M.-H. Meng, C. Hu, and Z. Liu, “Robust abnormal wireless capsule endoscopy frames detection based on least squared density ratio algorithm,” in *2011 IEEE International Conference on Information and Automation*. ICIA 2011. IEEE, 2011, pp. 324–328.

Y. Yanagawa, T. Echigo, H. Vu, H. Okazaki, Y. Fujiwara, T. Arakawa, and Y. Yagi, “Tracking abnormalities in video capsule endoscopy using surrounding features with a triangular constraint,” in *Proc. 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2012, pp. 578–581.

R. Yao, S. Zhang, W. Yang, S. Cheng, and Y. Chen, “Abnormality detection on gastroscopic images using patches assembled by local weights,” in *2010 International Conference of Medical Image Analysis and Clinical Application (MIACA)*. IEEE, 2010, pp. 38–41.

X. Yuan, B. Giriharan, M. Abouelnien, J. Liu, and X. Yuan, “Geometric Incremental Support Vector Machine for Object Detection from Capsule Endoscopy Videos,” in *Computer-Aided Cancer Detection and Diagnosis: Recent Advances*. Society of Photo-Optical Instrumentation Engineers, 2013.

M. A. Armin, H. De Visser, G. Chetty, C. Dumas, D. Conlan, F. Grimpen, and O. Salvador, “Visibility Map: A New Method in Evaluation Quality of Optical Colonoscopy,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, pp. 396–404.

C. S. Bell, K. L. Obstein, and P. Valaistri, “Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes,” *Artificial intelligence in medicine*, vol. 59, no. 3, pp. 185–196, 2013.

D. K. Iakovidis, E. Spyrou, and D. Diamantis, “Efficient homography-based video visualization for wireless capsule endoscopy,” in *13th IEEE International Conference on BioInformatics and BioEngineering*. IEEE, 2013, pp. 1–4.

D. K. Iakovidis, E. Spyrou, D. Diamantis, and I. Tsiampionis, “Capsule endoscopy localization based on visual features,” in *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1–4.

M. Stafiotakis, X. Zabulis, and D. P. Tsakiris, “Endoscopic capsule line-of-sight alignment by visual servoing,” in *7th Intl. Conf. on Wearable Micro and Nano Technologies for Personalized Health* (pHealth 2010), 2010.

G. Bao, L. Mi, Y. Geng, M. Zhou, and K. Pahlavan, “A video-based speed estimation technique for localizing the wireless capsule endoscopy inside gastrointestinal tract,” in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 5615–5618.

G. Bao, Y. Ye, U. Khan, X. Zheng, and K. Pahlavan, “Modeling of the Movement of the Endoscopy Capsule inside GI Tract based on the Captured Endoscopic Images,” in *International Conference on Modeling, Simulation and Visualization Methods*, Las Vegas, 2012.

G. Bao and K. Pahlavan, “Motion estimation of the endoscopy capsule using region-based kernel svm classifier,” in *ElectroInformation Technology (EIT), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–5.

G. Bao, K. Pahlavan, and L. Mi, “Hybrid Localization of Microscopic Endoscopic Capsule Inside Small Intestine by Data Fusion of Vision and RF Sensors,” *Sensors Journal, IEEE*, vol. 15, no. 5, pp. 2669–2678, 2015.

G. Bao, L. Mi, and K. Pahlavan, “A video aided RF localization technique for the wireless capsule endoscope (WCE) inside small intestine,” in *Proc. 8th International Conference on Body Area Ne-
A. Klepaczko and P. Szczypiński, “Automated segmentation of endoscopic images based on local shape-adaptive filtering and color descriptors,” in *Advances in Conceptual Intelligence Systems*. Springer, 2010, vol. 6474 LNCS, pp. 245–254.

F. Riaz, F. B. Silva, M. D. Ribeiro, and M. T. Coimbra, “Impact of Visual Features on the Segmentation of Gastroenterology Images Using Normalized Cuts,” *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 5, pp. 1191–1201, 2013.

M. Suenaga, Y. Fujita, S. Hashimoto, T. Shuji, I. Sakaida, and Y. Hamamoto, “A Method of Bubble Removal for Computer-Assisted Diagnosis of Capsule Endoscopic Images,” in *Modern Advances in Image Intelligence*. Springer, 2014, pp. 228–233.

M. Chowdhury and M. K. Kundu, “Endoscopic Image Retrieval System Using Multi-scale Image Features,” in *Proc. 2nd International Conference on Perception and Machine Intelligence*. ACM, 2015, pp. 64–70.

J. Kalpathy-Cramer, “Classification and retrieval of endoscopic images from the clinical outcomes research initiative (CORI) collection,” Master’s thesis, Oregon Health and Science University, 2009.

P. C. Khun, Z. Zhuo, L. Z. Yang, L. Liyuan, and L. Jiang, “Feature selection and classification for Wireless Capsule Endoscopic frames,” in *Proc. International Conference on Biomedical and Pharmaceutical Engineering*. IEEE, 2009, pp. 1–6.

R. Kumar, P. Rjan, S. Bejakovic, S. Seshamani, and G. Mullin, “Learning Disease Severity for Capsule Endoscopy,” in *Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*. IEEE, 2009, pp. 1314–1317.

M. Hafner, N. V. Vasconcelos, A. Uhl, M. Hafner, and F. Wrba, “Learning pit pattern concepts for gastrointestinal training,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2011, vol. 6893, pp. 280–287.

R. Kwitt, A. Uhl, M. Hafner, A. Gangl, F. Wrba, and A. Vecsei, “Predicting the histology of colorectal lesions in a probabilistic framework,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2010, pp. 103–110.

R. Kwitt, N. Vasconcelos, N. Rasiaswa, A. Uhl, B. Davis, M. Hafner, and F. Wrba, “Endoscopic image analysis in semantic space,” *Medical Image Analysis*, vol. 16, no. 7, pp. 1415–1422, 2012.

D. Liu, Y. Cao, K.-H. Kim, S. Stanek, B. Dongratanaex-Chai, K. Lin, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen, “Arthemis: annotation software in an integrated capturing and analysis system for colonoscopy,” *Computer Methods and Programs in Biomedicine*, vol. 88, no. 2, pp. 152–63, 2007.

T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, “Computer-aided colorectal tumor classification in NBI endoscopy: Using local features,” *Medical Image Analysis*, vol. 17, no. 1, pp. 78–100, 2013.

G. Yang, Y. Yin, and H. Man, “Biomedical Image Analysis on Wireless Capsule Endoscopy Images and Videos,” in *Selected Topics in MicroNano-robotics for Biomedical Applications*. Springer, 2013, pp. 23–43.

M. Mackiewicz, J. Berens, M. Fisher, and D. Bell, “Colour and texture based gastrointestinal tissue discrimination,” in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proc.*. 2006 IEEE International Conference on, vol. 2. IEEE, 2006, pp. II–II.

M. Biswas, “Hilbert Huang Transform Based Video Analysis for Detecting Various Colon Diseases using Composite Similarity Measure,” Master’s thesis, Jadavpur University, 2014.

A. F. Constantinescu, M. Ionescu, I. Rogoveanu, M. E. Ciurea, C. T. Streba, V. F. Iovanescu, S. A. Ariene, and C. C. Vere, “Analysis of Wireless Capsule Endoscopy Images using Local Binary Patterns,” *Applied Medical Informatics*, vol. 36, no. 2, pp. 31–42, 2015.

Y. Cong, S. Wang, J. Liu, J. Cao, Y. Yang, and J. Luo, “Deep sparse feature selection for computer aided endoscopy diagnosis,” *Pattern Recognition*, vol. 48, no. 3, pp. 907–917, 2015.

M. Boulougoura, E. Wadge, V. S. Kodogiannis, and H. S. Chowdrey, “Intelligent systems for computer-assisted clinical endoscopic image analysis,” in *Proc. 2nd International Conference on Biomedical Engineering*. ACTA Press, 2004, pp. 405–408.

A. M. C. de Sousa, “Analysis of colour and texture features of vital magnification-endoscopy images for computer diagnosis of precancerous and cancer lesions,” Master’s thesis, Universidade do Porto, 2008.
cation.” in 2014 22nd International Conference on Pattern Recognition. IEEE, 2014, pp. 2739–2744.

[355] M. Häfner, M. Liedlgruber, A. Uhl, and G. Wimmer, “Evaluation of super-resolution methods in the context of colonic polyp classification,” in Content-Based Multimedia Indexing (CBMI), 2014 12th International Workshop on. IEEE, 2014, pp. 1–6.

[356] M. Häfner, T. Tamaki, S. Tanaka, A. Uhl, G. Wimmer, and S. Yoshida, “Local fractal dimension based approaches for colonic polyp classification,” Medical Image Analysis, vol. 26, no. 1, pp. 92–107, Dec 2015.

[357] A. Karargyris and N. Bourbakis, “Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented,” Engineering in Medicine and Biology Magazine, IEEE, vol. 29, no. 1, pp. 72–83, 2010.

[358] M. Keuchel, N. Kurniawan, P. Baltes, D. Bandorski, and A. Koulaouzidis, “Quantitative measurements in capsule endoscopy,” Computers in biology and medicine, 2015.

[359] R. Kumar, Q. Zhao, S. Sheshamani, G. Mullin, G. Hager, and T. Dassopoulos, “Assessment of crohn’s disease lesions in wireless capsule endoscopy images,” Biomedical Engineering, IEEE Transactions on, vol. 59, no. 2, pp. 355–362, 2012.

[360] I. Laranjo, J. Braga, D. Assunção, C. Rolanda, L. Lopes, J. Correia-pinto, and V. Alves, “Video Processing Architecture: A Solution for Endoscopic Procedures Results,” in Ambient Intelligence-Software and Applications. Springer, 2014, vol. 291, pp. 117–125.

[361] M. Liedlgruber and A. Uhl, “Computer-Aided Decision Support Systems for Endoscopy in the Gastrointestinal Tract: A Review,” IEEE Reviews in Biomedical Engineering, vol. 4, pp. 73–88, 2011.

[362] M. Liedlgruber and A. Uhl, Predicting Pathology in Medical Decision Support Systems in Endoscopy of the Gastrointestinal Tract. INTECH Open Access Publisher, 2011.

[363] M. Mackiewicz, Capsule Endoscopy - State of the Technology and Computer Vision Tools After the First Decade. INTECH Open Access Publisher, 2011.

[364] D. E. Maroulis, D. K. Iakovidis, S. a. Karkanis, and D. a. Kararas, “CoLD: A versatile detection system for colorectal lesions in endoscopy video-frames,” Computer Methods and Programs in Biomedicine, vol. 70, no. 2, pp. 151–166, 2003.

[365] D. Mikhailov, A. Starikovsky, V. Konev, A. Grigorenko, and S. Larisa, “Review of Software for Automated Analysis of Digestive Tract Images,” Biosciences Biotechnology Research Asia, vol. 11, no. 3, pp. 1109–1114, 2014.

[366] B. Wang and D. Yang, “Computer-Assisted Diagnosis of Digestive Endoscopic Images Based on Bayesian Theory,” in 2009 International Conference on Information Engineering and Computer Science. IEEE, 2009, pp. 1–4.

[367] Y. M. Yacob, H. Amylia, M. Sakim, N. Baharudin, L. Y. Yeh, N. Ashidi, and M. Isa, “A Survey on Medical Digital Imaging of Endoscopic Gastritis,” in IEEE TENCON 2009 : IEEE Region 10 Conference. IEEE, 2009, pp. 1–6.

[368] M. M. Zheng, S. Krishnan, and M. P. Tjoa, “A fusion-based clinical decision support for disease diagnosis from endoscopic images,” Computers in biology and medicine, vol. 35, no. 3, pp. 259–74, 2005.