

Content-based processing and analysis of endoscopic images and videos: A survey

Bernd Münzer¹  · Klaus Schoeffmann¹ ·
Laszlo Böszörményi¹

Received: 19 May 2016 / Revised: 14 October 2016 / Accepted: 25 November 2016 /
Published online: 11 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract In recent years, digital endoscopy has established as key technology for medical screenings and minimally invasive surgery. Since then, various research communities with manifold backgrounds have picked up on the idea of processing and automatically analyzing the inherently available video signal that is produced by the endoscopic camera. Proposed works mainly include image processing techniques, pattern recognition, machine learning methods and Computer Vision algorithms. While most contributions deal with real-time assistance at procedure time, the post-procedural processing of recorded videos is still in its infancy. Many post-processing problems are based on typical Multimedia methods like indexing, retrieval, summarization and video interaction, but have only been sparsely addressed so far for this domain. The goals of this survey are (1) to introduce this research field to a broader audience in the Multimedia community to stimulate further research, (2) to describe domain-specific characteristics of endoscopic videos that need to be addressed in a pre-processing step, and (3) to systematically bring together the very diverse research results for the first time to provide a broader overview of related research that is currently not perceived as belonging together.

Keywords Medical imaging · Endoscopic videos · Content-based video analysis · Medical multimedia

✉ Bernd Münzer
bernd@itec.aau.at
Klaus Schoeffmann
ks@itec.aau.at
Laszlo Böszörményi
laszlo@itec.aau.at

¹ Institute for Information Technology / Lakeside Labs, Klagenfurt University, Universitätsstraße 65-67, 9020 Klagenfurt, Austria

1 Introduction

In the last decades, medical endoscopy has emerged as key technology for minimally-invasive examinations in numerous body regions and for minimally-invasive surgery in the abdomen, joints and further body regions. The term “*endoscopy*” derives from Greek and refers to methods to “look inside” the human body in a minimally invasive way. This is accomplished by inserting a medical device called *endoscope* into the interior of a hollow organ or body cavity. Depending on the respective body region the insertion is performed through a natural body orifice (e.g., for examination of the esophagus or the bowel) or through a small incision that serves as artificial entrance. For surgical procedures (e.g., removal of the gallbladder), additional incisions are required to insert various surgical instruments. Compared to open surgery this still causes much less trauma, which is one of the main advantages of minimally invasive surgery (also known as buttonhole or keyhole surgery). Many medical procedures were revolutionized by the introduction of endoscopy, some were even enabled by this technology in the first place. For a more detailed insight into the history of endoscopy, please refer to [113] and [96].

Endoscopy is an umbrella term for a variety of very diverse medical methods. There are many sub-types of endoscopy, which have very different characteristics. They can be classified according to several criteria, e.g.,

- body region (e.g., abdomen, joints, gastrointestinal tract, lungs, chest, nose)
- medical speciality (e.g., general surgery, gastroenterology, orthopedic surgery)
- diagnostic vs. therapeutical focus
- flexible vs. rigid construction form

Unfortunately, the usage of the term “endoscopy” is neither consistent in common parlance nor in the literature. It is often used as synonym for gastro-intestinal endoscopy (of the digestive system), which mainly includes **colonoscopy** and **gastroscopy** (examination of the colon and stomach, respectively). A further special type is Wireless Capsule Endoscopy (WCE). The patient has to swallow a small capsule that includes a tiny camera and transmits a large number of images to an external receiver while it travels through the digestive tract for several hours. The images are then assessed by the physician after the end of this process. WCE is especially important for examinations of the small intestine because neither gastroscopy nor colonoscopy can access this part of the gastrointestinal tract.

The major therapeutic sub-types are **laparoscopy** (procedures in the abdominal cavity) and **arthroscopy** (orthopedic procedures on joints, mainly knee and shoulder). They are often subsumed under the term “minimally invasive surgery”. Laparoscopic operations span different medical specialities, particularly general surgery, pediatric surgery, gynecology and urology. Examples for common laparoscopic operations are cholecystectomy (removal of the gall bladder), nephrectomy (removal of the kidney), prostatectomy (removal of the prostate gland) and the diagnosis and treatment of endometriosis. Further important endoscopy types are thoracoscopy (thorax/chest), bronchoscopy (airways), cystoscopy (bladder), hysteroscopy (uterus) and further special procedures in the field of ENT (ear, nose, throat) and neurosurgery (brain).

In the course of an endoscopic procedure, a video signal is produced by the endoscopic camera and visualized to the surgical team to guide their actions. This inherently available video signal is predestinated for automatic content analysis in order to assist the physician. Hence, numerous research communities proposed methods to process and analyze it, either in real-time or for post-procedural usage. In both cases, image processing techniques

are often used to pre-process individual video frames, be it to improve the performance of subsequent processing steps or to simply improve their visual quality. Pattern recognition and machine learning methods are used to detect lesions, polyps, tumors etc. in order to aid physicians in the diagnostic analysis. The robotics community applies Computer Vision algorithms for 3D reconstruction of the inner anatomical structure in combination with detection and tracking of operation instruments to enable robot-assisted surgery. In the context of Augmented Reality, endoscopic images are registered to pre-operative CT or MRI scans to provide guidance and additional context-specific information to the surgeon during the operation. For readers who want to learn more about the workflow in this field, we recommend the following tutorial papers [151, 152].

In recent years, we can observe a growing trend to record and store videos of endoscopic procedures, mainly for medical documentation and research. This new paradigm of video documentation has many advantages: it enables to revisit the procedure anytime, it facilitates detailed discussions with colleagues as well as explanations to patients, it allows for better planning of follow-up operations, and it is a great source of information for research, training, education and quality assurance. The benefits of video documentation have been confirmed in numerous studies, e.g., [61, 100, 101, 137]. However, physicians can only benefit from endoscopic videos if they are easily accessible. This is where research in the Multimedia field comes in. Well researched methods like content-based video retrieval, video segmentation, summarization, efficient storage and archiving concepts as well as efficient video interaction and browsing interfaces can be used to organize an endoscopic video archive and make it accessible for physicians. Because of their post-processing nature, these techniques are not constrained by immediate OR requirements and therefore can be applied in real-world scenarios much easier than real-time assistance features. Nevertheless, they have to be adapted to the specific peculiarities of this very specific domain. Their practical relevance is steadily growing considering the fact that video documentation is on the rise in recent years. Once comprehensive video documentation is established as best practice and maybe even becomes mandatory, they will be essential cornerstones in Endoscopic Multimedia Information Systems.

As we can see, there are very diverse goals and perspectives on the domain of endoscopic video processing. This survey is intended to provide a broad overview of related research in this very heterogeneous and broad field that is currently not perceived as belonging together. It also tries to point up common problems that might be easier to solve when considering findings of other fields. In an extensive literature research, more than 600 publications were found. Based on titles and abstracts, we classified them into the following three main categories which are described in the subsequent sections:

1. pre-processing methods
2. real-time support at procedure time
3. post-procedural applications

Figure 1 illustrates the resulting categorization of research topics in the field of endoscopic image/video processing and analysis, representing the structure of the following sections as well. This classification should not be understood as the ultimate truth because many of the presented techniques and concepts have significant overlappings and cannot be distinctively delimited. For example, the traditionally post-procedural application of surgical quality assessment is currently being ported to real-time systems and in this context could as well be regarded as an application of Augmented Reality. Nevertheless, this categorization enables a structured and clear overview of the many topics that are covered in this review.

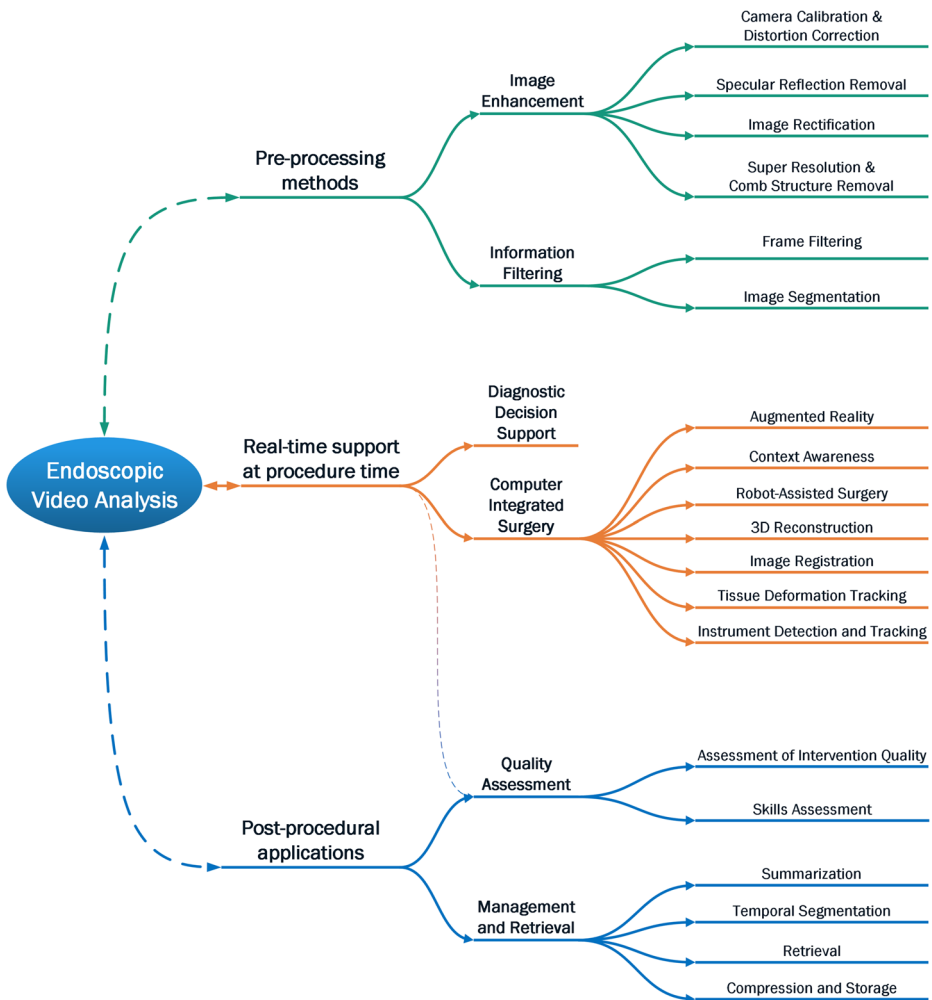


Fig. 1 Categorisation of publications in the field of endoscopic video analysis

2 Pre-processing methods

Endoscopic videos have various domain-specific characteristics that need to be addressed when dealing with this special kind of video. This section describes the most distinctive aspects and gives an overview of corresponding methods that are applied as a preparatory step prior to other analysis techniques and/or enhance the image quality for the surgeon.

2.1 Image enhancement

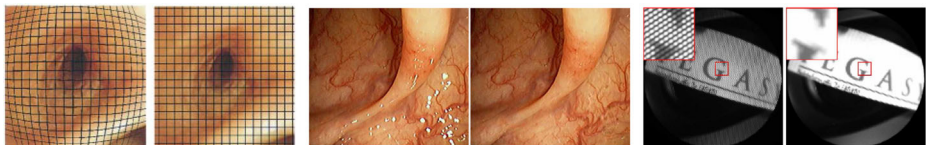
A number of publications deal with the enhancement of frames from endoscopic videos in order to improve the visual quality of the video. That means that the underlying data, i.e., the pixels of the individual frames are not only analyzed but also modified while other analysis approaches described in the upcoming sections only try to extract information without

changing the content. In this context a number of well-established general purpose image processing techniques can be applied, but this section will focus on techniques and research findings that specifically address the domain of endoscopy. Another aspect that is particularly important in this context is real-time capability because the optimized result should instantly be visible at the screen during a procedure. However, image enhancement and pre-processing is not only interesting for real-time applications but can also be of great importance as a preparation step for any kind of further automatic processing. Early work in this area includes:

- Automatic adjustment of contrast with the help of clustering and histogram modification [207].
- Removal of temporal noise, i.e., small flying particles or fast moving smoke only appearing for a short moment at one position, by using a temporal median filter of color values [251].
- Color normalization using an affine transformation in order to get rid of a reddish tinge caused by blood during therapeutic interventions and to obtain a more natural color [251].
- Correction of color misalignment: Most endoscopes do not use a color chipset camera but a monochrome chipset that only captures luminance information. To get a color video, red, green and blue color filters have to be applied sequentially. In case of rapid movements - which occur frequently in endoscopic procedures - the color channels become misaligned. This is not only annoying when watching the video but particularly hindering further automatic analysis. Dahyot et al. [47] propose to use color channels equalization, camera motion estimation and motion compensation to correct the misalignments.

2.1.1 Camera calibration and distortion correction

Typical endoscopes have a fish-eye lens to provide a wide-angle field of view. This characteristic is useful because the endoscopist can see a larger area. However, the drawback is a non-linear geometric distortion (barrel distortion). Objects located in the center of the image appear larger and lines get bended as illustrated in Fig. 2a. This distortion has to be corrected prior to advanced methods that rely on correct geometric information, e.g., 3D reconstruction or image registration. The basic problem is to find the distortion center and the parameters that describe the extent of the distortion, which is not constant but depends on the respective endoscope. This process is also known as camera calibration and includes the determination of intrinsic and extrinsic camera parameters. Vijayan et al. [247] proposed to use a calibration image showing a rectangular grid of dots. This image is captured by the



(a) Distortion correction of a bronchoscopic image [77] [222] © 2001 IEEE. **(b)** Specular reflection removal [57] © 2006 IEEE. **(c)** Comb structure removal [57] © 2006 IEEE.

Fig. 2 Illustration of image enhancement methods for endoscopy

endoscope, resulting in a distorted version of the calibration image. Then the transformation parameters from this distorted image to the original calibration image are calculated using polynomial mapping and least squares estimation. These parameters are used to build a model that can then be used to correct the actual frames from the endoscopic video. This approach was further improved in [277] and [77]. A further approach in [273] is not only applicable to forward viewing endoscopes but also to oblique viewing endoscopes. Their camera model is able to compensate the rotation but has a higher complexity and more parameters. For calibration, they use a chess pattern image instead of a grid of dots. Further publications using this calibration pattern are [11, 12, 223]. In [72], the authors investigate if distortion correction also affects the accuracy of CAD (Computer Aided Diagnosis). The surprising result was that for many feature extraction techniques the performance did not improve but was even worse than without distortion correction. Only for shape-based features that rely on geometrical properties a modest improvement was observed. Further research results in this field can be found in [90, 123, 264].

2.1.2 Specular reflection removal

Endoscopic images often contain specular light reflections, also called highlights, on the wet tissue surface. They are caused by the inherent frontal illumination and are very distracting for the observer. A study conducted in [252] shows that physicians prefer images where they are corrected. Even worse, reflections severely impair analysis algorithms because they introduce wrong pixel values and additional edges. This also impairs image feature extraction, which is an essential technique for reconstruction, tracking etc. Hence, a number of approaches for correction have been proposed as a supporting component for other analysis methods, e.g., detection of non-informative frames [169], segmentation and detection of surgical instruments [34, 201], tracking of natural landmarks for cardiac motion estimation [70], reconstruction of 3D structures [226] or correction of color channel misalignment [8].

Most approaches consist of two phases. First, the highlights are detected in each frame. This is rather straightforward and in most cases uses basic histogram analysis, thresholding and morphological operations. Pixels with an intensity above a threshold are regarded as highlights. Some authors additionally propose to check for low saturation as a further strong indication for specular highlights ([169, 274]). In this context, the usage of various color spaces has been proposed, e.g., RGB [8], YUV [222], HSV [169], HSI [274], CIE-xyY [143]. In a second phase, the pixels identified as reflections are “corrected”, i.e., modified in a way that the resulting image looks as realistic as possible. An example of a corrected image can be seen in Fig. 2b. An important aspect is that user should be informed about this image enhancement, because one cannot rule out the possibility that wrong information is introduced, e.g., a modified pit pattern on a polyp that can adversely affect the diagnosis. For this second phase, the following two different approaches can be distinguished:

- **Spatial interpolation:** Only the current frame is considered and the pixels that correspond to specular highlights are replaced by interpolated pixels from the surrounding. This technique is also called inpainting and has its origins in image restoration (for an overview of general inpainting techniques refer to [208]). For the interpolation, different methods have been proposed, e.g., spectral deconvolution [222], anisotropic confidence-based filling-in [70, 71] or a two-level inpainting where first the highlight pixels are replaced with the centroid of their neighborhood and finally a gaussian blur is applied to smooth the contour of the interpolated region [8].

- **Temporal interpolation:** With inpainting techniques, an actually correct reconstruction is not possible because the real color information for highlight pixels cannot be determined from the current frame. Hence, several approaches have been proposed that consider the temporal context [226, 253], i.e., try to find the corresponding position in preceding and subsequent frames and reuse the information from this position. This approach has a higher complexity than inpainting but it can be used to reconstruct the real information. However, this is not always possible, especially if there is too little or very abrupt motion or if the lighting conditions change too much. Moreover, it is not applicable to single frames or WCE (Wireless Capsule Endoscopy) frames.

2.1.3 Image rectification

In surgical practice, a commonly used type of endoscopes are oblique-viewing endoscopes (e.g., 30°). The advantage of this design is the possibility to easily change the viewing direction by rotating the endoscope around its axis. This enables a larger field of view. The problem is that also the image rotates, resulting in a non-intuitive orientation of the body anatomy. The surgeon has to unrotate the image in their mind in order to not lose their orientation. The missing information about the image orientation is especially a problem in Natural Orifice Translumenal Endoscopic Surgery (NOTES), where a flexible endoscope is used (as opposed to rigid endoscopes like in laparoscopy). Some approaches have been proposed that use modified equipment to tackle this problem, e.g., an inertial sensor mounted on the endoscope tip [80], but hardware modifications always limit the practical applicability. Koppel et al. [102] propose an early vision-based solution. They track 2D image features to estimate the camera motion. Based on this estimation, the image is rectified, i.e., rotated such that the natural “up” direction is restored. Moll et al. [149] improve this approach by using the SURF descriptor (Speeded Up Robust Features), RANSAC (Random Sample Consensus) and a bag-of-visual-words approach based on Integrated Region Matching (IRM). A different approach [60] exploits the fact that endoscopic images often feature a “wedge mark”, a small spike outside the image circle that visually indicates the rotation. By detecting the position of this mark, the rotation angle can easily be computed.

2.1.4 Super resolution and comb structure removal

In diagnostic endoscopic procedures like colonoscopy it is important to visualize very fine details - e.g., patterns on the colonic mucosa surface - to make the right diagnosis. Super-resolution [237] has been proposed as a means to increase the level of detail of HD videos and enable a better diagnosis - both manual and automatic [75, 76]. The idea of super resolution is to combine the high frequency information of several successive low resolution images to an image with higher resolution and more details. However, the authors come to the conclusion that their approach neither has a significant impact on the visual quality nor on the classification accuracy. Duda et al. propose to apply super-resolution [55] for WCE images. Their approach is very fast because it simply computes a weighted average of the upsampled and registered frames and is shown to perform better than bilinear interpolation.

Rupp et al. [200] use super-resolution for a different task, namely to improve the calibration accuracy of flexible endoscopes (also called fiberscopes). This type of endoscope uses a bundle of coated glass fibers for light transmission. This produces image artifacts in the form of a comb-like pattern (see Fig. 2c). This comb structure hampers an exact calibration, but also many other analysis tasks like feature detection. Several methods for comb structure removal have been proposed, e.g., low pass filtering [50], adaptive reduction

via spectral masks [266], or spatial barycentric or nearest neighbor interpolation between pixels containing fiberoptic content [57]. These methods typically contain some kind of low pass filtering, meaning that edges and contours are blurred. These lost high frequency components can be restored by applying super-resolution algorithms.

2.2 Information filtering

Endoscopic videos typically contain a considerable amount of frames that do not carry any relevant information and therefore are useless for content-based analysis. Hence, it is desirable to automatically detect such frames and sort them out, i.e., perform a *temporal filtering*. This can be regarded as a different kind of pre-processing, with the difference that not the pixels of individual frames are modified but the video as such is modified to the effect that frames are removed. This idea is closely related to video summarization (see Section 4.2.4), which can be seen as an intensification of frame filtering. In video summarization, the goal is to select especially informative frames or sequences and reduce the video to an even higher extent. Moreover, it is often the case that only parts of an image are non-informative, but other regions are indeed relevant for the analysis. To concentrate analysis on such selected regions, several image segmentation techniques have been proposed to perform a *spatial filtering*.

2.2.1 Frame filtering

In the literature, different criteria can be found for a frame to be considered as informative or non-informative. The most important criterion is blurriness. According to [9], about 25 % of the frames of a typical colonoscopy video are blurry. Oh et al. [169] propose to use edge detection and compute the ratio of isolated pixels to connected pixels in the edge image to determine the blurriness. As this method depends very much on the selection of thresholds and further parameters, they propose a second approach using discrete Fourier transformation (DFT). Seven texture features are extracted from the gray-level co-occurrence matrix (GLCM) of the resulting frequency spectrum image and used for k-means clustering to differentiate between blurry and clear images. A similar approach by [9] uses the 2D discrete wavelet transform with a Haar wavelet Kernel to obtain a set of approximation and detail coefficients. The L^2 norm of the detail coefficients of the wavelet decomposition is used as feature vector for a Bayesian classifier. This method is nearly 10-times faster than the DFT-based method and also has a higher accuracy. Rangseekajee and Phongsuphap [189] and Rangseekajee et al. [199] on the other side took up the edge-based approach for the domain of thoracoscopy and added adaptive thresholding as pre-processing step to reduce the effect of lighting conditions. Besides, they claim that the Sobel edge detector is more appropriate for this task than the Canny edge detector because it detects less edges due to irrelevant details caused by noise. Another approach [10] uses inter-frame similarities and the concept of manifold learning for dimensionality reduction to cluster indistinct frames. Grega et al. [68] compared the different approaches for the domain of bronchoscopy and reported results for F-measure, sensitivity, specificity and accuracy of at least 87 % or higher. According to them, the best-scoring alternative is a transformation-based approach using discrete cosine transformation (DCT).

Especially in the context of WCE (Wireless Capsule Endoscopy), the presence of intestinal juices is another criterion for non-informative images. Such images are characterized by bubbles that occlude the visualization field. Vilarino et al. [248] use Gabor filters to detect

them. According to their studies, 23 % of all images can be discarded, meaning that the visualization time for the manual diagnostic assessment as well as the processing time for automatic diagnostic support can be considerably reduced. In [13], a similar approach is proposed that uses a Gauss Laguerre transform (GLT)-based multiresolution texture feature and introduces a second step that uses spatial segmentation of the bubble region to classify ambiguous frames.

A further type of non-informative frames are out-of-patient frames, i.e., frames from scenes that are recorded outside the patients body. They often occur at the beginning or end of a procedure because it is not always possible to start and stop the recording exactly at the right time. The need for manual recording triggering in general deters many endoscopists from recording videos at all. To address this issue, [218] propose a system that automatically detects when a colonoscopic examination begins and ends. Every time a new procedure is detected, the system starts recording and writes a video file to the disk until the end of the procedure is detected. The proposed approach uses simple color features that work well for the domain of colonoscopy. In [217], the authors extend their approach by various temporal features that take into account the amount of motion to avoid false positives.

2.2.2 Image segmentation

Instead of discarding complete frames, some authors try to identify non-informative regions in endoscopic images. In further processing steps, only the informative regions have to be considered, which speeds up processing and improves accuracy. Such a spatial filtering can also be used as basis for temporal filtering by defining a threshold ratio between the size of informative and non-informative regions. A typical irrelevant region is the border area outside the characteristic circular content area of endoscopic images. It contains no useful information but only noise that impairs analysis as well as compression. In [159], an efficient domain-specific algorithm is proposed to detect the exact circle parameters. Bernal et al. [20] propose a model of appearance of non-informative lumen regions that can be discarded in a subsequent CAD (Computer Aided Diagnosis) component. Prasath et al. [181] also use image segmentation to differentiate between lumen and mucosa, but they use the result as a basis for 3D reconstruction. For WCE images, [135] apply morphological operations, fuzzy k-means, sigmoid function, statistic features, Gabor Filters, Fisher test, neural network, and discriminators in the HSV color space to differentiate between informative and non-informative regions. In the context of CAD, image segmentation is also used as basis for shape-based features. Here, the goal is to determine the boundaries of polyps, tumors and lesions [21, 83] or other anomalies like bleeding or ulceration in WCE images [232].

In the case of surgical procedures, the most frequently addressed target of image segmentation are surgical instruments. They can be tracked in order to understand the surgical workflow or assess the skills of the surgeon. For more details on instrument detection and tracking please refer to Section 3.2.6. Few approaches have been proposed for segmentation of anatomical structures. Chhatkuli et al. [38] show how segmentation of the uterus in gynecological laparoscopy using color and texture features improves the performance of Shape-from-Shading 3D reconstruction and feature matching. Bilodeau et al. [24] combine graph-based segmentation and multistage region merging to determine the boundary of the operated disc cavity in thoracic discectomy, which is a useful depth cue for 3D reconstruction and very important in this surgery type to correctly estimate the distance to the spinal cord.

3 Real-time support at procedure time

The use case of endoscopic video analysis that has been studied most extensively in the literature is to directly support the physician during the procedure in various ways. The application scenarios can be categorized into (1) Diagnostic Decision Support and (2) Computer Integrated Surgery, which includes Augmented Reality as well as Robot-Assisted Surgery.

3.1 Diagnostic decision support

In case of diagnostic procedures like colonoscopies or gastroscopies, the main goal is to assist physicians in their diagnosis by deciding whether the anatomy is normal or abnormal. This is done by detecting and classifying suspicious patterns in images that correspond to abnormalities like polyps, lesions, inflammations and tumors. Figure 3 illustrates examples for normal (first row) and abnormal (second row) images. Such decision support systems are often called CAD (Computer Aided Diagnosis) systems and are already used to some extent in clinical practice. In general, CAD systems strive to be real-time capable to provide immediate feedback during the examination, e.g., [261]. If the physician misses a suspicious structure, the system can highlight the corresponding region to indicate that it should be investigated in detail and maybe a biopsy should be taken. If CAD is applied as post-processing, a reaction of the physician is not possible anymore. However, some state-of-the-art approaches are still too computationally expensive and can currently only be applied offline after the examination. In the special case of WCE, the diagnostic support does not have to be real-time, because the physician anyway looks at the images after the actual procedure is finished and all images have been acquired. For WCE, aside from the detection of structures like polyps or tumors, the detection of images showing bleedings is of particular interest [66, 119].

CAD systems typically use pattern recognition and machine learning algorithms to identify abnormal structures. After various pre-processing steps, visual features are extracted and fed into a classifier, which then delivers the diagnosis in form of a classification result. The classifier has to be trained in advance with a possibly great number of labeled examples. The most frequently used classifiers are Support Vector Machines (SVM) and Neural

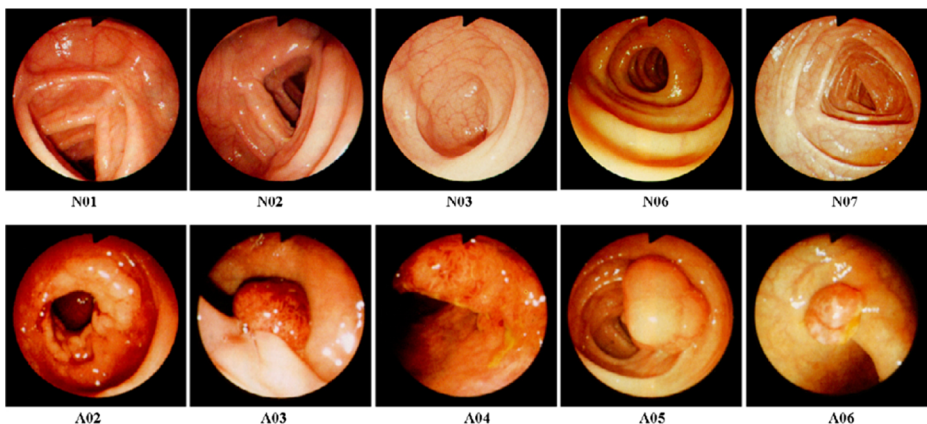


Fig. 3 Typical normal (first row) and abnormal (second row) images [238]

Networks. Often, dimensionality reduction techniques like Principal Component Analysis (PCA) are used, e.g., in [238]. Numerous alternatives for the selection of features and classifiers have been proposed. Some approaches exploit the characteristic shape of polyps, e.g., [21]. Considering the shape also enables unsupervised methods, which require no training, e.g., the extraction of geometric information from segmented images [83] or simple counting of regions of a segmented image [49]. Many approaches use texture features, e.g., based on wavelets or local binary patterns (LBP), e.g., [6, 86, 118, 257]. In many cases, color information adds a lot of additional information, so color-texture combinations are also common [93]. Also simple color and position approaches have been shown to perform reasonable despite their low complexity compared to more sophisticated approaches [4].

Most publications concentrate on one specific feature, but there are also attempts to use a mix of features. As example, Zheng et al. propose a Bayesian fusion algorithm [279] to combine various feature extraction methods, which provide different cues for abnormality. A very recent contribution by Riegler et al. [193] proposes to employ Multimedia methods for disease detection and shows promising preliminary results. A detailed survey of gastrointestinal CAD-systems can be found in [122].

3.2 Computer integrated surgery

In the case of surgical endoscopy, we can differentiate between “passive” support in the form of Augmented Reality and “active” support in the form of robotic assistance.

In the former case, supplemental information from other image modalities (MRT, CT, PET etc.) is displayed to improve navigation, enhance the viewing conditions or provide context-aware assistance. In the latter case, surgical robots are used to improve surgical precision for complex and delicate tasks. While early systems acted as direct extender of the surgeons movements, recent research activity strives for more and more actions carried out by the robot autonomously. Both cases pose a number of typical Computer Vision problems (object detection and tracking, reconstruction, registration etc. in order to “understand” the surgical scene), hence video analysis is an essential component.

All these ideas and techniques can be subsumed under the concept of Computer Integrated Surgery (CIS), or sometimes also referred to as surgical CAD (due to the popularity of the term CAD) [235]. The underlying idea is to integrate all phases of treatment with the support of computer systems, and in particular medical imaging. This includes intra-operative endoscopic imaging as well as pre-operative diagnostic imaging modalities like CT (X-ray Computed Tomography), MRI (Magnetic Resonance Imaging), PET (Positron Emission Tomography) or sonography (“ultrasound”). These modalities are used for diagnosis and planning of the actual procedure and often are essential for the precise navigation to the surgical site [14]. This is especially the case for surgeries that require a very high accuracy due to the risk of damaging healthy tissue (e.g., endonasal skull base surgery [145]). Navigation support is also important for diagnostic procedures like bronchoscopy (examination of the lung airways) where the flexible endoscope has to traverse a complex tree structure with many branches to find the biopsy site that has been identified prior to the examination [48, 79]. These pre-operative images or volumetric models are aligned with general information about human anatomy (anatomy atlases) in order to create a patient-specific model that enables a comprehensive and detailed procedure planning. This pre-operative model is then registered to the intra-operative video images in real-time to guide the surgeon by overlaying additional information, performing certain tasks autonomously or increasing surgical safety by imposing constraints on surgical actions that could harm the patient. For such an assistance, the system has to monitor the progress of

the procedure and in case of complications automatically adapts/updates the surgical plan [234].

3.2.1 Augmented reality

An essential concept for CIS is Augmented Reality (AR) - a technique to augment the perception of the real world with additional virtual components that are generated by the computer and have to be aligned to the real-world environment in real-time. In the case of endoscopic surgery, the “virtual” components are usually obtained from the pre-operative patient model and surgical plan. They can be visualized in several ways, e.g., on the ordinary monitor, through a head mounted display (HMD), which can either be an optical or a video see-through HMD, or as projection directly on the patient body [209]. Without AR, the surgeon has to mentally merge this isolated information with the live endoscopic view, which causes additional mental load. By augmenting the endoscopic video images with this kind of information, target structures can be highlighted for easier localization and hidden anatomical structures (e.g., vessels or tumors below the organ surface) can be visualized as overlay in order to improve safety and avoid surgical errors and complications. The key challenge for AR is to align the endoscopic video with the pre-operative data, i.e., to fuse them to a common coordinate system - a technique called image registration. The problem is massively exacerbated by the fact that the soft tissue is not rigid but shifting and deforming. Hence, another research challenge is to track the tissue deformation to derive deformation models in order to update the registration. In addition to Computer Vision algorithms (e.g., calibration, registration, 3D reconstruction), AR systems often rely on external optical tracking systems, which are used to determine the position and motion of the endoscope or an instrument [23, 29]. This implies that instruments have to be modified by attaching markers, which are tracked by an array of infrared cameras rigidly mounted on the ceiling of the operating room. The drawback of such methods is the limited applicability in a practical scenario due to the necessary hardware modification. An example for the application of AR is depicted in Fig. 4.

Surgical navigation AR is especially helpful in the field of surgical oncology [167], i.e., the surgical management of cancer. The exact position and size of the tumor is often not directly visible in the endoscopic images, hence an accurate visualization helps to choose an optimal dissection plane that minimizes damage of healthy tissue. Such systems are often called “Surgical Navigation Systems” because they support the navigation to the surgical site. They have been proposed for various procedures, e.g., prostatectomy (removal of the prostate gland) [82, 210]. Mirota et al. [146] provides a comprehensive overview of vision-based navigation in image-guided interventions for various procedure types.

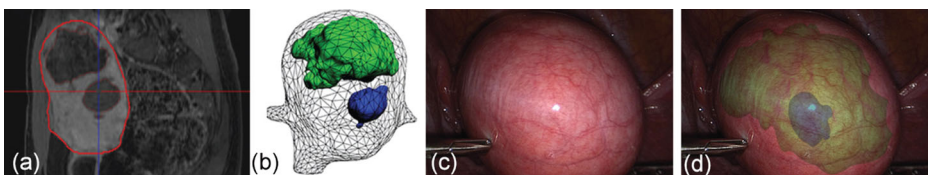


Fig. 4 MRI image showing a uterus with two myomas, the corresponding pre-operative model and the visualization of the overlay on the endoscopic image [45] © 2014 IEEE

Viewport enhancement A further possible application of AR is to improve the viewing conditions of the surgeon. This can be done by expanding the restricted viewport and visualizing the surrounding area using image stitching methods [19, 153, 241], potentially even in 3D [265]. Similar techniques have also been investigated for the purpose of video summarization, e.g., to obtain a condensed representation of an examination for a medical record (see Section 4.2.4). Other approaches even provide an alternative point of view with improved visualization. In [23], a “Virtual Mirror” is proposed that enables the surgeon to inspect the virtual components (e.g., a volumetric model of the liver from a pre-operative CT scan) from different perspectives in order to understand complex structures (e.g., blood vessel trees) and improve depth perception as well as navigational tasks. Fuchs et al. [59] propose a prototypical system that restores the physicians’ natural viewpoint and visualizes it via a See-Through HMD. The underlying idea is to free the surgeon from the inherent technical limitations of the imaging system and enable a more “direct” view on the patient, similar to that in traditional open surgery. The surgeon can change the viewing perspective by moving his head instead of moving the laparoscope.

3.2.2 Context awareness

Several medical studies prove the clinical applicability of Augmented Reality in various endoscopic operation types, e.g., nephrectomy [236], prostatectomy [246], laparoscopic gastrointestinal procedures [229] or splenectomy [88]. However - despite all the potential benefits - too much additional information may distract surgeons from the actual task [51]. The goal should be to automatically select the appropriate assistance for the current state of the procedure in a context-aware manner, providing hints or situation-specific additional information. Such assistance can also go beyond visualizations of the pre-operative model, e.g., it can support decision making by finding situations similar to the current one and showing how other surgeons handled a similar exceptional situation [95]. Another use case is to simulate the effect of a surgical step without actually executing it [115]. Speidel et al. [214] and [228] propose an AR system for warning in case of risk situations. The system alerts the surgeon if an instrument comes too close to a risk structure (ductus cysticus or arteria cystica in the case of cholecystectomy).

The basis for such context-aware assistance is the semantic understanding of the current situation. The field of Surgical Process Modeling (SPM) is concerned with the definition of appropriate workflow models [110, 166]. The main challenge is to formalize the existing expert knowledge, be it formal knowledge from textbooks or experience-based knowledge that also considers variations and deviations from theory. First, typical surgical situations or tasks have to be defined. Some approaches focus on fine-granular gestures like for example “insert a needle”, “grab a needle”, “position a needle” [18], or more generic actions like tool-tissue interaction in general [255]. The detection of such “low-level” tasks can also be used as basis for the assessment of surgical skills (see Section 4.1.1). On a higher abstraction level, a procedure is subdivided into pre-defined operation phases that describe the typical sequence of a surgery. Existing approaches focus on well standardized procedures, in most cases cholecystectomy (removal of the gallbladder) [25, 95, 99, 109, 174], which can be broken down to distinct phases very well. Surgical workflow understanding is also of particular interest for post-procedural applications, especially for temporal video segmentation and summarization (see Section 4.2.3).

A very discriminative feature to distinguish between phases is the presence of operation instruments, which can be detected by video analysis (see Section 3.2.6 for more information). However, metadata obtained from video analysis is only one of many possible

inputs for surgical situation understanding systems proposed in the literature. Often, various additional sensor data are used, e.g., weight of the irrigation and suction bags, the intra-abdominal CO_2 pressure and the inclination of the surgical table [221] or a coagulation audio signal [262]. The focus in this research area is not on how to obtain the required information from the video, but how to map the available signals (e.g., binary information about instrument presence) to the corresponding surgical phase. Therefore, instrument detection is often achieved by hardware modifications like RFID tags or color markers or even by manual annotations from an observer [166].

Several authors propose to use statistical methods and machine learning to recognize the current situation. The most popular method are Hidden Markov Models (HMM) [25, 111, 174, 196]. An HMM is a directed graph that defines possible states and transition probabilities and is built from training data. Another frequently used method is Dynamic Time Warping (DTW) [58], which can also be used without explicit pre-defined models by temporally aligning surgeries of the same type [3]. Besides HMM and DTW also alternative methods like Random Forests have been proposed [221]. A fundamentally different approach is to use formal knowledge representation methods like ontologies that use rules and logical reasoning to derive the current state. For example, Katic et al. use Description Logic in the OWL-standard (Web Ontology Language) [94, 95, 214].

3.2.3 Robot-assisted surgery

The majority of publications dealing with endoscopic video analysis have their roots in the robotics community and aim at integrating robotic components into the surgical workflow. Medical robots are not intended to replace human surgeons, but to extend their capabilities and improve efficiency and effectiveness by overcoming certain limitations of traditional laparoscopic surgery. They considerably enhance the dexterity, precision and repeatability by using mechanical wrists that are controlled by a microprocessor. Motion can be stabilized by filtering hand tremor and scaled for micro-scale tasks which are not possible manually [115]. Robotic surgery systems are practically used for several years, especially for very precarious procedures like prostatectomy [250]. However, current systems like the daVinci system [74] are pure “surgeon extenders”, i.e., the surgeon directly controls the slave robot via a master console. In this telemanipulation scenario, the robot has no autonomy and only “enhances” the surgeons movements, e.g., by hand tremor filtering and motion scaling. An overview of popular surgical robot systems can be found in [230] and [250]. State-of-the-art research tries to extend the robots autonomy, which requires numerous image/video analysis and Computer Vision techniques. Robotic systems inherently provide additional data that can be used to facilitate video analysis, e.g., kinematic motion data, information about instrument usage and stereoscopic images that allow for an easier 3D reconstruction of the scene.

An important application with mediocre degree of autonomy is the automation of endoscope holding [36, 254, 256, 278]. This task is usually carried out by an assistant, but during lengthy procedures, humans suffer from fatigue, hand tremor etc., therefore automation of this task is very appreciated by surgeons. The endoscope should always point at the current area of interest, which is typically characterized by the presence of the instrument tips. Hence, instrument positions have to be detected and the robot arm has to be moved such that the endoscope is adequately positioned without colliding with tissue, and the right zoom level is chosen [211]. Another application are automatic safety checks, e.g., in the form of active constraints respectively virtual fixtures. A virtual fixture [140, 177] is a constraint that reduces the precision requirements. It can be used to define forbidden regions

or a safety margin around critical anatomical structures, which must not be damaged, in order to prevent erroneous moves, or to simplify the execution of a task by “guiding” the instrument motion along a safe corridor [30]. The long-term vision is to enable commonly occurring tasks like suturing to be executed autonomously by high-level command of the surgeon (e.g., by pointing at the target position with a laser pointer) [105, 175, 220]. The main challenge is to safely move the instrument to the desired 3D position without harming the patient, a process referred to as visual servoing. Such an assistance requires a very detailed understanding of the surgical scene, including a precise 3D model of the anatomy, registered to the pre-operational model and also considering tissue deformations, as well as the exact location of relevant anatomical objects and instruments. Also surgical task models as discussed above are of great importance for this scenario. A survey of recent advances in the field of autonomous and semi-autonomous actions carried out by robotic systems is given in [156].

3.2.4 3D reconstruction

The reconstruction of the local geometry of the surgical site is an essential requirement for Robot-Assisted Surgery. It produces a three-dimensional model of the anatomy in which the instruments are positioned. Also for many Augmented Reality applications a 3D model is required for registration with volumetric pre-operative models (e.g., from CT scans). For diagnostic procedures, the analysis of the 3D shape of suspicious objects can be more expressive than the 2D shape. The fundamental challenge of 3D reconstruction is to map the available 2D image coordinates to 3D world coordinates.

In the context of Robot-Assisted Surgery, usually stereoscopic endoscopes are used to improve the depth perception of the surgeon. The stereo images also facilitate correspondence-based 3D reconstruction [22, 194, 225]. The challenge is to identify matching image primitives (e.g., feature points) between the left and right image, which can then be used to calculate the depth information by triangulation. However, this task is still far from being trivial because of various aggravating factors like homogenous surfaces with few distinct visual features, occlusions, specular reflections, image perturbations (smoke, blood etc.) and tissue deformations. In traditional laparoscopy, which is not supported by a robot, the used endoscopes are usually monoscopic. In this case, Structure-From-Motion (SfM) methods can be applied [44, 81, 138]. For SfM, the different views are obtained when the camera is moved to a different position. Camera motion estimation is required to estimate the displacement of the camera position, which is necessary for the triangulation, while in the stereoscopic case, the displacement is inherently known. A related method that is often used in the robotics domain is SLAM (Simultaneous Localization And Mapping), which iteratively constructs a map of the unknown environment and at the same time keeps track of the camera location. Traditional SLAM assumes a rigid environment, which does not hold for the endoscopic case. Therefore, attempts have been made to extend SLAM with respect to tissue deformations [67, 154, 242]. A further common problem for both SfM and SLAM is the often scarce camera motion. An alternative approach that deals with single images and therefore does not depend on the problem of finding correspondences is Shape-from-Shading (SfS), where the depth information is derived from the shading of the surface of anatomic objects. However, again some basic assumptions of generic SfS do not hold for endoscopic images, hence adaptations are necessary to obtain acceptable results [43, 171, 269]. The generalization of SfS to multiple light sources is referred to as photometric stereo and has also been proposed for 3D reconstruction of endoscopic videos [42, 178]. This technique requires hardware modifications to obtain different illumination conditions. Other

active methods requiring hardware modifications that have been proposed are Structured Light [1] and Time-of-Flight [179]. The former projects a known light pattern on the surface and reconstructs depth information from the deformation of the pattern. The latter uses non-visible near-infrared light and measures the time until it is reflected back.

All these approaches have their drawbacks, hence several attempts have been made to improve the performance by fusing multiple depth cues, e.g., stereoscopic video and SfS [127, 249], SfM and SfS [139, 239], or by incorporating patient specific shape priors extracted from pre-operative images [7]. However, 3D reconstruction still remains a very challenging task in the endoscopic domain. Several recent surveys about this topic are available [63, 69, 136, 176].

3.2.5 Image registration and tissue deformation tracking

The process of bringing two images of the same scene together to one common coordinate system is referred to as image registration. This includes the computation of a transformation model that describes how one image has to be modified to obtain the second image. This is a typical optimization problem that can be solved with optimization algorithms like Gauss-Newton etc. The classical application of medical image registration is to align images from different modalities, e.g., pre-operative 3D CT images and intra-operative 2D X-ray projection images [141]. We can distinguish between different types of registration, depending on the dimensionality of the underlying images (2D slice, 3D volumetric, video), as reviewed in [121].

For Augmented Reality scenarios where the endoscopic video stream is used as intra-operative modality, various use cases have been addressed in the literature, e.g., registering 3D CT models of the lungs with bronchoscopic videos [79, 131]. Similar examples relying on registration are sinus surgery (nose) [31] or skull base surgery [147]. In terms of laparoscopic surgery, interesting contributions are 3D-to-3D registration of the liver during laparoscopic surgery [227] and coating of the pre-operative 3D model with texture information from the live view [258]. An important prerequisite for registration is an accurate camera calibration (see Section 2.1.1) to obtain correct geometric correspondences.

A topic closely related to image registration is object tracking, i.e., following the motion of a region of interest over time, either in the two- or three-dimensional space. It can be seen as an intra-modality (as opposed to inter-modality) registration, i.e., successive frames of a video are registered. In case of camera motion between frames, the transformation can be described by translation, rotation and scaling. Estimating the camera motion is an important technique for 3D reconstruction (especially SfM and SLAM) and generating panoramic images. However, in the endoscopic video domain, the transformation is usually much more complex. The main reason is the fact that the soft tissue is not rigid but is deforming non-linearly, requiring adaptations of many established methods which assume a rigid environment (e.g., SLAM). Tissue deformation occurs for three main reasons, (1) organ shift due to insufflation (mainly relevant for inter-modality registration), (2) periodic motion caused by respiratory and cardiac cycles as well as muscular contraction and (3) tool-tissue interaction. Tracking the tissue deformation is a challenging research topic that has strongly gained attention in the last years. It is particularly important for updating the reconstructed 3D model that has to be registered with the static pre-operative model for Augmented Reality applications and Robot-Assisted Surgery, but also to track anatomical modifications for surgical workflow understanding. Periodic motion can be well described by a model, e.g., Fourier series [192]. Estimating the periodic motion of the heart is of particular interest for robot-assisted motion compensation and virtual stabilization [202].

Both registration and tissue tracking share the basic problem of finding a set of correspondences between two images in order to compute the transformation model. This is usually based on some kind of “landmarks” that can be identified in both images. In a matching step, an algorithm decides which landmarks represent the same position. Landmarks can either be artificial, e.g., represented by color markers attached to the tracking target like in [202], or natural. For tissue tracking, artificial markers (also called fiducials or fiducial markers) can hardly be used, therefore tracking algorithms have to rely on natural landmarks, i.e., salient image features that can clearly be distinguished from their surrounding and are unique, e.g., vessel junctions and surface textures. One possibility is to work in the image space and use region-based representations of regions of interest in the form of pixel patches. However, these representations are often not sufficiently expressive and not very robust against illumination changes, specular highlights and occlusions. Hence, feature-based representations have established as preferred method. They allow to detect natural landmarks and extract specific information that is represented by a feature descriptor. The most common descriptors are SIFT (Scale Invariant Feature Transform) [130] and SURF (Speeded-Up Robust Features) [15]. They are popular because of their beneficial characteristics like scale and rotation invariance and robustness against illumination changes and noise. Further feature descriptors that have been used for tissue tracking are MSER (Maximally Stable Extremal Regions) [142, 224], STAR, which is a modified version of the Center Surrounded Extremas for Real-time Feature Detection (CenSuRE) [2], and BRIEF (Binary Robust Independent Elementary Features) [32, 275]. Moutney et al. [150] provide a comprehensive evaluation and comparison of numerous feature descriptors and present a framework for descriptor selection and fusion. Figure 5 illustrates correspondences between several pairs of images.

The matching strategy typically used by feature-based approaches is “tracking-by-detection”, as opposed to recursive tracking methods, e.g., Lucas Kanade, which is based on the optical flow. The latter search locally for a best match for image patches, while the former extract features for each frame and then compare them to find the best matches. While recursive methods work well on small deformations, they have problems with illumination changes and reflections and suffer from error propagation. In contrast, tracking-by-detection in combination with feature-based region representation is fairly robust against large deformations and occlusions due to the abstracted feature space. A comparative evaluation of state-of-the-art feature-matching algorithms for endoscopic images has been carried out in [186].

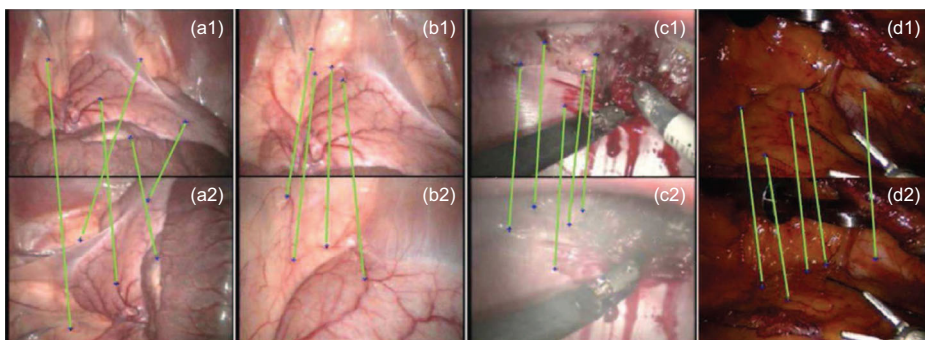


Fig. 5 Illustration of finding corresponding natural landmarks between pairs of images with (a) significant rotation, (b) scale change, (c) image blur (d) tissue deformation, combined with illumination changes [64]
© 2009 IEEE

However, although promising advances have been made recently [155, 187, 188, 205, 231, 268], deforming tissue tracking is a very hard research challenge that still requires a lot of further work. Endoscopic videos feature many domain-induced problems like scarcity of distinctive landmarks because of homogenous surfaces and indistinctive texture that makes it hard to find good points to track. Moreover, occlusions, specular reflections, cauterization smoke, blood and fluids lead to tracking points being lost. Hence, one of the main problems is a robust long-term tracking. Last but not least, the real-time requirement poses a demanding challenge. A survey about three-dimensional tissue deformation recovery and tracking is available in [152].

3.2.6 Instrument detection and tracking

Besides anatomical objects, surgical instruments are the most important objects of interest in the surgical site. Therefore, a key requirement for scene understanding, Robotic-Assisted Surgery and many other use cases is to detect their presence and track their position as well as their motion trajectories. The precision requirements differ with the application scenario. For surgical phase recognition, it is often sufficient to know which instruments are present at all. In this case, it is already sufficient to equip the instruments with a cheap RFID tag to detect the insertion and withdrawal [103]. In terms of visual analysis, a classification of the full frame can be carried out to detect the presence of instruments [183]. The next level is to determine the position of the instrument in the two-dimensional image, or more specifically the position of the instrument tip, which is the main differentiation characteristic between different types of instruments [213]. Also for instrument tracking, the position of the tip is usually considered as reference point.

Many approaches proposed in the literature use modified equipment for localizing instruments. The most common modification are color markers on the shaft that can easily be detected and segmented [240, 263]. To distinguish between multiple instruments, different colors can be used [26]. Nageotte et al. [165] use a pattern of 12 black spots on a white surface. This modification also enables pose estimation to some extent. Krupa et al. [105] use a laser pointing device on the tip of the endoscope to derive the relative orientation of the instrument with respect to the organ. This approach even works if the instrument is outside the current field of view. Besides the fact that these methods have a limited applicability to arbitrary videos from an archive, another disadvantage of such modifications is that the biocompatibility and sterilizability has to be ensured, as they have direct contact to human tissue. Also internal kinematic data from a robot can be used to estimate the position of instruments, but is generally not accurate enough, especially when force is applied to a surface [30]. However, kinematics can be useful as supplementary source of information to get a coarse estimation that is refined by visual analysis [219].

Also purely vision-based approaches without any hardware modification have been proposed. Doignon et al. [53] perform a color segmentation mainly using the saturation attribute to differentiate the achromatic instrument from the background. Voros et al. [256] define a cylindrical shape model and use geometric information to detect the edges of the tool shaft and the symmetry axis. A similar approach using the Hough transform to detect the straight lines of the instrument shaft is presented in [41]. These approaches face a number of challenges like homogenous color distribution, indistinct edges, occlusions, blurriness, specular reflections and artifacts like smoke, blood or liquids. Moreover, methods have to deal with multiple instruments, as surgical procedures rarely rely on one single instrument.

The final stage of instrument identification and one of the current research challenges is to determine the exact position of the tip in the three-dimensional space and track its

motion trajectories [5, 33]. In this context, also the estimation of the instrument pose is of particular importance [54]. This knowledge is necessary for use cases like visual servoing, context-aware alerting and skills assessment. Given geometric constraints can be exploited to facilitate tracking, particularly the motion constraint, which is imposed by the stationary incision point. Knowledge about this point restricts the search area for region seeds for color-based segmentation [52] and enables modeling of possible instrument motion [267]. Recently, a number of advanced and very sophisticated approaches for 3D instrument tracking and pose estimation have been proposed, e.g., training of appearance models of individual parts of an instrument (shaft, wrist and finger) using color and texture features [180], learning of fine-scaled natural features in the form of particular 3D landmarks on instrument tips using Randomized Trees [191], learning the shape of instruments using HOG descriptors and Latent SVM to probabilistically track them [107], and approaches to determine semantic attributes like “*open/closed, stained with blood or not, state of cauterizing tools*” etc. [108].

4 Post-procedural applications

In recent years, it became more and more common to capture videos of endoscopic procedures. We experience a trend towards a comprehensive documentation where entire procedures are recorded, stored and archived for documentation, retrospective analysis, quality assurance, education, training and other purposes. The question arises how to handle this huge corpus of data? The emerging huge video archives pose a challenge to management and retrieval systems. The Multimedia community has proposed several methods to enable summarization, retrieval and management of such data, but this research topic is clearly understudied as compared to the real-time assistance scenario. This section gives an overview of these methods, which can be regarded as kind of post-processing. They do not have the requirement to work in real-time because they operate on captured video data and can be executed offline. Nevertheless, performance is important in order to keep up with the constantly growing data volume.

4.1 Quality assessment

An important application for post-procedural usage of endoscopic videos that has been studied intensively is quality assessment of individual surgeon skills and of entire procedures. Quality control of endoscopic procedures is a very important, but also very difficult issue. In current practice, an experienced surgeon has to review videos of surgeries to subjectively assess the skills of the surgeon and the quality of the procedure. This is a very time-consuming and cumbersome task that cannot be carried out extensively due to the high effort. Hence, it is desirable to provide automatic methods for an objective quality assessment that can be applied to each and every procedure and has the potential to strongly improve surgical quality by pointing out weak points and suggesting potentials for improvement. Recent works even assess quality in real-time during the procedure to provide immediate feedback to the physician and thus improve their performance [216].

4.1.1 Surgical skills assessment

Several attempts have been made to assess the psychomotor skills of surgeons by analyzing how they perform individual surgical tasks like cutting, grasping, clipping, drilling

or suturing. This requires a decomposition of a procedure into individual atomic surgical gestures (often called *surges*), which can also be seen as a kind of temporal segmentation. Similar to surgical workflow understanding (Section 3.2.2), Hidden Markov Models (HMM) are typically used to model and detect the different tasks. A reference model is trained by an expert and serves as basis for the skills assessment. The main parameter for the assessment is the motion trajectory of instruments. Various metrics like path length, motion smoothness, average acceleration etc. have been proposed as quality indicators.

Most classic approaches use non-standard equipment to obtain this motion data in a straightforward manner, e.g., inherently available kinematic data from a simulator or surgical robot [124], trajectory data from an external optical tracker [116] and/or haptic data from an additional three-axis force/torque sensor [197]. However, such a simple but expensive data acquisition is not suitable for a comprehensive quality assessment on a daily basis, but rather interesting for special applications like training, simulation and Robot-Assisted Surgery.

The more practical alternative for an extensive evaluation of surgeon skills is to extract the motion data directly from the video data with content-based analysis methods [173]. The advantage of this approach is that it can be applied to any video without any hardware modification. Furthermore, videos can provide contextual information with regard to the anatomical structures and instruments involved that can act as additional hints. To obtain the required motion data, instruments have to be detected and tracked (see Section 3.2.6), preferably in the three-dimensional space. Some contributions concerning this matter have recently been published, e.g., [89, 172, 276].

4.1.2 Assessment of screening quality

The skills assessment methods discussed above are rather applied for surgeries and focus mainly on the instrument handling. In diagnostic procedures, usually no instruments are used, therefore other quality criteria have to be defined. Hwang et al. [84, 168] define objective metrics that characterize the quality of a colonoscopy screening, e.g., the duration of the withdrawal phase (based on temporal video segmentation), the ratio of non-informative frames, the number of camera motion changes and the ratio of frames showing close inspections of the colon wall to frames showing a global lumen view. This framework is extended in several further publications. In [126], a “quadrant coverage histogram” is introduced that determines to what extent all sides of the mucosa have been inspected. A similar approach is proposed in [91] for the domain of cystoscopy. Here the idea is to determine to what extent the inner surface of the bladder has been inspected and if parts have been missed. A further quality criterion for colonoscopies is the presence of stool that occludes the mucosa and thus may lead to missed polyps. Color features are used to measure the ratio of “stool images” and consequently the quality of the preceding bowel cleansing [85, 164]. On the other side, the presence of frames showing the appendiceal orifice indicates a high quality examination, because it means that the examination has been performed thoroughly [259]. Also the occurrence of retroflexion, which is a special endoscope maneuver to discover polyps that are hidden behind deep folds in the peri-anal mucosa, is a quality indicator and can be detected with a method proposed in [260]. In [168], therapeutic actions are detected and also considered for quality assessment metrics. The assessment of intervention quality is mainly applied post-operatively on recorded videos, but also first attempts have been made to include these techniques in the clinical workflow and directly notify the physician about quality deficiencies [163, 170].

4.2 Management and retrieval

The goal of video management and retrieval systems is to enable users to efficiently find exactly the information they are looking for, either within one specific video or within an archive. They have to provide means to articulate some kind of query describing the information need, or special interaction mechanisms for efficient content browsing and exploration. Especially the latter aspect has rarely been addressed for this specific domain yet and therefore provides a large potential for future work.

4.2.1 Compression and storage

If an endoscopic video management system is to be deployed in a realistic scenario, domain-specific concepts for compression, storage organization and dissemination of videos are required. As for compression of endoscopic videos, literature research only showed very few contributions. For the field of bronchoscopic examinations, we found a statement that “*it is possible to use lossy compressed images and video sequences for diagnostic purposes*” [56, 185]. In terms of storage organization, [27] propose a system with a distributed architecture that uses a NoSQL database to provide access to videos within a hospital and across different health care institutions [28]. They also present a device for video acquisition and an annotation system that enables content querying [112]. Münzer et al. [160] show that the circular content area of endoscopic videos can be exploited to considerably improve encoding efficiency and the discarding of irrelevant segments (blurry, out-of-patient or dark and noisy) can save up to 20 % of the storage space [161]. In [162], a subjective quality assessment study is conducted that shows that it is not necessary to archive videos in the original HD resolution, but lower quality representations still provide sufficient semantic quality. This study also contains the first set of encoding recommendations for the domain of endoscopy. A follow-up study in [148] evaluates the effective savings in storage space by using domain-specific video compression on an authentic real-world data set. It comes to the conclusion that by using these encoding presets together with circle detection, relevance segmentation and a long-term archiving strategy, the overall storage capacity can be reduced by more than 90 % in the long term without losing relevant information.

4.2.2 Retrieval

One possibility to retrieve specific information in endoscopic videos is to annotate them manually. Lux et al. present a mobile tool for efficient manual annotation of surgery videos [73, 133]. It provides intuitive direct interaction mechanisms on a tablet device that make the annotation task less tedious. For colonoscopy videos, an annotation tool called Arthemis [125] has been proposed that supports the Minimal Standard Terminology (MST) of the European Gastrointestinal Society for Endoscopy (ESGE), which is a standardized terminology for diagnostic findings in GI endoscopy. However, manual annotation and tagging cannot be carried out for each and every video of a video archive due to time restrictions. It is rather interesting to use manual annotation to obtain expert knowledge from a limited number of representative examples and extract this knowledge for automatic retrieval techniques like content-based image retrieval (CBIR).

The typical use case of CBIR is to find images that are visually similar to a query image. Up to now, research in image retrieval in the medical domain rather focuses on other image modalities (CT, MRI etc.) [106, 157], but some publications also deal with endoscopic images. An early approach [272] uses simple color histograms in HSV color space

to determine the similarity between colonoscopy images. A more recent approach [270] for gastroscopic images uses image segmentation in the CIE $L^*a^*b^*$ color space to extract a “color change feature” and dominant color information and compares images using the Kullback-Leibler distance. Tai et al. [233] also incorporate texture information by using a color-texture correlogram and the Generalized Tersky Index as similarity measure. Furthermore, they employ an interactive relevance feedback to refine the search result. Xia et al. [271] propose to use multi-feature fusion to combine color, texture and shape information. A very recent and more sophisticated approach [39] that obtains very promising results uses Multiscale Geometric Analysis (MGA) of Nonsubsampled Contourlet Transform (NSCT) and the statistical framework based on Generalized Gaussian Density (GGD) model and Kullback-Leibler Distance (KLD). Another task relying on similarity search is to link a single still image captured by the surgeon during the procedure to the according video segment in an archive [195]. The latest approach for this task uses Feature Signatures and the Signature Matching Distance and achieves reasonable results [16]. For laparoscopic surgery videos, a technique has been proposed to find scenes showing a specific instrument that is depicted in the query image [37]. In terms of video retrieval, [245] use the HOG (Histogram of Oriented Gradients) descriptor and a Fisher kernel based similarity to find similar video sequences in other videos based on a video snippet query. This technique can on the one hand be used to compare similar situations, but on the other hand also to automatically assign a semantic tag annotation based on the existing annotation of similar sequences in an existing (annotated) video archive. The same authors also propose a method for surgery type classification that automatically differentiates between 8 types of laparoscopic surgery [244]. The method uses RGB and HSV color histograms, SIFT (Scale Invariant Feature Transform) and HOG features together with an SVM classifier and obtains promising results.

4.2.3 Temporal segmentation

While CBIR seems to work quite well for diagnostic endoscopy, it is not well studied for surgical endoscopy types like laparoscopy. In this case, the “query by example” paradigm is not very expedient. It is based on the naive assumption that visual similarity correlates with semantic similarity. However, this assumption does not necessarily hold because the semantics of a laparoscopic image or video sequence depend on a very complex context that cannot be thoroughly represented with simple low-level features like color, texture and shape. This discrepancy between low-level representation and high-level semantics is referred to as semantic gap. The key to close this gap is to additionally take into account the dynamic aspects of videos that images do not have. One of the key techniques for general video retrieval is temporal segmentation, i.e., the subdivision of a video into shots and semantic scenes. This abstraction can help a lot to better understand a video and, e.g., find the position of a certain surgical step. Unfortunately, established generic techniques cannot be applied because they typically assume that a video is composed of shots. Endoscopic videos usually have exactly one shot and hence only one scene according to commonly accepted definitions. Therefore, conventional shot detection cannot be used. This means that a new definition of shots and scenes is necessary for endoscopic videos. Some authors tried to introduce a new domain-specific notion of shots. Primus et al. [182] define “shot-like” segments based on significant motion changes. Their idea is to differentiate between typical motion types, namely no motion, camera motion and instrument motion. This approach produces a very fine-grained segmentation that can be used as basis for a more coarse-grained semantic segmentation. Cao et al. [34] define operation shots in colonoscopy videos. They

are detected by the presence of instruments, which are only used in special situations in this diagnostic endoscopy type.

A more promising approach for endoscopic video segmentation is to define a model of the progress of the procedure and associate video segments with the individual phases of this model. This idea is closely related to surgical workflow understanding as discussed above (see Section 3.2.2). However, here the purpose is not to provide context-aware assistance, but to structure a video file to facilitate efficient review of procedures and retrieval of specific scenes. The fact that complete information about the whole procedure is available at processing time makes this task easier as compared to the intra-operative real-time scenario. For diagnostic procedures like colonoscopy, where the endoscope has to follow a predetermined path through several anatomic regions, it is straightforward to consider these regions as phases. Cao et al. [35] observed that transitions from one section of the colon to the next feature a certain pattern of sharp and blurry frames. This is because the physician has to steer the endoscope around anatomic “corners”. This pattern can be exploited to segment a video into semantic phases. Similarly, WCE videos can also be segmented into subvideos showing specific topographic areas like esophagus, stomach, small intestine and large intestine [46, 114, 134, 206].

In the case of laparoscopy, the presence of certain instruments can be used to distinguish between surgical phases [184]. This approach works very well for standardized procedures like Cholecystectomy. For more individualized procedures, additional cues need to be incorporated. Some alternative approaches, which are not based on surgical process modeling, have been proposed in the literature, e.g., probabilistic tissue tracking to generate motion patterns that are used to identify surgical episodes [65], smoke detection as indication of electrocautery based on *ad hoc* kinematic features from optical flow analysis of a grid of particles [129] and an unsupervised approach that extracts and analyzes multiple time-series data sets based on instrument occurrence and motion [97]. Another interesting contribution [98] does not incorporate temporal information, but classifies individual frames according to their surgical phase. The used features are automatically generated by genetic programming in order to overcome the challenge of choosing the right features a priori. The drawback of this approach is the long processing time for feature evolution.

4.2.4 Summarization

Endoscopic videos often have a duration of several hours, but surgeons usually do not have the time to review the whole footage. Therefore, summarization of endoscopic videos is a crucial feature of a video management system. Dynamic summaries can be used to reduce the duration by determining the most relevant parts of the video. Static summaries are useful to visualize the essence of an endoscopic procedure at a glance and can easily be archived in the patient’s medical record or even be passed down to the patient. Many general approaches for summarization of videos have been proposed in the literature, but only a few that consider the specific domain characteristics of endoscopic videos.

Lux et al. [132] present a static summarization algorithm for the domain of arthroscopy (an endoscopy type that is hardly ever addressed in the literature). It generates a single result image composed of a predefined number of most representative frames. The representativity is determined based on k-medoid clustering of color and texture features. In this context, such frames are often referred to as *keyframes*. Another method for keyframe extraction in endoscopic videos, which can also be used as basis for temporal segmentation, is presented in [204]. It uses the ORB (Oriented FAST and Rotated BRIEF) keypoint descriptor [198] and an adaptive threshold to detect significant differences between two frames.

Lokoč et al. [128] present an interactive tool for dynamic browsing of such keyframes. The keyframes are clustered and presented in a hierarchical manner to get a quick overview of a procedure. A similar interactive presentation of keyframes from hysteroscopy videos in the form of a video segment tree is proposed in [203] and further refined in [62], together with a summarization technique that estimates the clinical relevance of segments based on the attention attracted during video acquisition. These interactive techniques build the bridge between video summarization and temporal segmentation. However, video summaries do not have to be static, but can also consist of a shortened version of the original video, as proposed for the domain of bronchoscopy in [117]. This is achieved by discarding a large number of completely non-informative frames and keeping frames that are representative or clinically especially relevant (e.g., showing the branching of airways or pathological lesions).

Many WCE-specific techniques can also be considered as kind of summarization [40, 87, 243]. The task is to reduce a very large collection of images (with some temporal relationship, but less than in a typical video because the frame rate is much lower) to the ones that are diagnostically relevant. This could as well be seen as a pre-processing or filtering step. The WCE scenario differs from the usual post-procedural review scenario because it is not an additional optional task where the physician can watch the video material again, but it is the mandatory core step of the screening. Thus, the optimization of time efficiency is especially important here. A recent survey of various image analysis techniques for WCE can be found in [92].

Also **panoramic images** of the surgical site can be seen as a special kind of summary that gives a visual overview, especially of examinations. Techniques for panorama generation are sometimes also calling mosaicing or stitching algorithms. If a panorama is generated during a procedure, the term dynamic view expansion is often used (see Section 3.2.1). The basic idea is to combine different frames of the video, which were recorded from a different perspective, to one image that extends the restricted field of view. Image stitching is closely related to image registration (see Section 3.2.5). The underlying challenge is to find corresponding points and compute the transformation between the two frames, in order to convert them to a common coordinate system. Finally, the registered images have to be blended together to create the panorama.

Several authors addressed panoramas in the context of cystoscopy, i.e., the examination of the interior of the bladder [78, 144, 212]. The bladder can be modeled as a sphere, so geometric constraints can be imposed. Behrens et al. [17] present a method using graphs to identify missed regions during a cystoscopy, which would lead to gaps in a panoramic image, in order to assess the completeness of the examination. Spyrou et al. [215] propose a stitching technique for WCE frames. In the context of fetoscopy (examination of the interior of the uterus during pregnancy), [190] propose a method to create a panorama of the placenta to support intrauterine fetal surgery. Liao et al. [120] extend this idea by a method to map the panorama to a three-dimensional ultrasound image. A review about recent advances in the field of endoscopic image mosaicing and panorama generation can be found in [19].

5 Conclusion

The main goal of this extensive literature review was to give a broad overview of research that deals with the processing and analysis of endoscopic videos. A further goal is to draw attention to this research field in the Multimedia community. We hope to stimulate further research, especially in terms of post-processing, which is probably the most relevant topic

with regard to common Multimedia methods and offers a broad range of open research questions. Moreover, we give insights into domain-specific characteristics of the endoscopy domain and how to deal with them in a pre-processing phase (e.g., lense distortion, specular reflections, circular content area, etc.).

In the literature research, numerous contributions were found and classified into three categories: (1) pre-processing methods, (2) real-time support at procedure time and (3) post-procedural applications. However, many methods and approaches have been found to be relevant for multiple use-cases, e.g., instrument detection and tracking methods developed for robotic assistance that can also be helpful for post-procedural video indexing. Currently, the respective research communities are often not aware that complementary contributions exist in seemingly unrelated research fields.

The domain-specific peculiarities of the endoscopic video domain require specific pre-processing methods like distortion correction or specular reflection detection. Moreover, pre-processing is often used to enhance the image quality or filter relevant content, both temporally and spatially. These enhancements are both important to improve the viewing conditions for physicians and as pre-processing step for advanced analysis methods. In this context, it is also important to distinguish between different types of endoscopy, mainly between diagnostic (examinations) and therapeutic (surgeries) types, but also between the subtypes that often have very heterogeneous domain characteristics. Most approaches focus on one specific endoscopy type and cannot be transferred to other types without significant modifications, i.e., they are strongly domain-specific.

Furthermore, we have to distinguish between methods that are applied during the procedure and methods that operate on recorded videos. The former have the requirement to work in real-time and can only use the information up to the current moment, i.e., they cannot “read into the future”. The latter have this possibility because the entire video is available at analysis time. In terms of diagnostic endoscopy types, the focus is on pattern recognition for Diagnostic Decision Support, mainly in the form of polyp/lesion/tumor detection in colonoscopic screenings. In the special domain of WCE (Wireless Capsule Endoscopy), numerous approaches have been proposed to differentiate between diagnostically relevant and irrelevant content in order to increase time efficiency without impairing the diagnostic accuracy.

The largest part of all found publications originates in the robotics community and has the goal to enable real-time assistance during surgeries by (1) Augmented Reality and (2) (semi-)autonomous actions carried out by surgical robots. These visionary goals require a number of classical Computer Vision techniques like 3D-reconstruction, object detection and tracking, registration etc. Existing methods fail in most cases because of the special characteristics of endoscopic images and aggravating factors like the restricted and distorted viewport, scarce camera motion, specular reflections, oclusions, image quality perturbations, textureless surfaces, tissue deformation etc., and therefore have to be adapted. Many of these techniques could also be useful for post-procedural analysis of videos for more efficient management, retrieval and interactive retrospective review.

The post-procedural use case turned out to be extremely understudied in comparison to the real-time scenario. However, it involves a number of very interesting and challenging research questions, including indexing, retrieval, segmentation, summarization and interaction.

One of the reasons might be that recording of endoscopic procedures is just becoming a general practice and is not yet commonly widespread. This can also be explained by the lack of efficient compression and archiving concepts for comprehensive recording. As a consequence, the acquisition of appropriate data sets is the first challenge for new

researchers in this field. Data availability in general is a critical issue since sensitive medical data is involved and hospital policies regarding video recording and transfer are often very strict. Only very few public data set are available and usually target a very specific application (like tumor recognition). In order to draw expressive comparisons between alternative approaches it is very important for the future to create and publish more and bigger public data sets. The extent of a data set is especially important for machine learning methods that require large amounts of training samples. The shortage of a sufficiently large and representative training corpus often hampers the usage of popular techniques like deep learning with Convolutional Neural Networks (CNN) [104].

An additional challenge is to find medical experts who are willing to take the time to cooperate and share their medical expertise, which is an absolutely essential ingredient to successful research in this extremely specific domain. As an example, ground truth labels need to be annotated to training examples. This task cannot be carried out by medical laymen. Moreover, at a first glance, it seems that the demand by physicians is limited because they might see post-procedural usage of endoscopic videos as additional workload, although in fact it has a huge potential for quality improvement. Nevertheless, it is a tough job to familiarize physicians with the benefits of post-procedural usage of videos and the need for research in this area. A survey conducted in [158] showed that physicians often do not have a clear notion of potential benefits until they actually see it in action. After watching a prototype demonstration of a content-based endoscopic video management system, they stated a significantly higher interest in such a system than before.

However, it should not be expected that all problems in the post-operative handling of endoscopic videos can be solved by automatic analysis. An extremely important aspect is to combine these methods with easily understandable visualization concepts and intuitive interaction mechanisms for efficient content browsing and exploration. Especially the latter aspect has rarely been addressed in the literature yet and therefore provides a huge potential for future work.

In the foreseeable future, we assume that video documentation of endoscopic procedures will become required by law. This will lead to huge archives of visually very similar video content and techniques for video storage, retrieval and interaction will become essential. When it comes to that point, research should already have appropriate solutions for these problems.

Acknowledgments Open access funding provided by University of Klagenfurt. This work was supported by Universität Klagenfurt and Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF-20214 U. 3520/26336/38165.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Ackerman JD, Keller K, Fuchs H (2002) Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality. In: *Electronic Imaging 2002*, pp 39–46. International Society for Optics and Photonics
2. Agrawal M, Konolige K, Blas MR (2008) CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. In: *Computer Vision – ECCV 2008*, no. 5305 in LNCS, pp 102–115. Springer

3. Ahmadi SA, Sielhorst T, Stauder R, Horn M, Feussner H, Navab N (2006) Recovery of Surgical Workflow Without Explicit Models. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, no. 4190 in LNCS, pp 420–428. Springer
4. Alexandre L, Nobre N, Casteleiro J (2008) Color and Position versus Texture Features for Endoscopic Polyp Detection. In: Int'l Conf on BioMedical Engineering and Informatics. BMEI 2008, vol 2, pp 38–42
5. Allan M, Ourselin S, Thompson S, Hawkes D, Kelly J, Stoyanov D (2013) Toward Detection and Localization of Instruments in Minimally Invasive Surgery. *IEEE Trans Biomed Eng* 60(4):1050–1058
6. Ameling S, Wirth S, Paulus D, Lacey G, Vilarino F (2009) Texture-based polyp detection in colonoscopy. In: Bildverarbeitung für die Medizin 2009, pp 346–350. Springer
7. Amir-Khalili A, Peyrat JM, Hamarneh G, Abugharbieh R (2013) 3d Surface Reconstruction of Organs Using Patient-Specific Shape Priors in Robot-Assisted Laparoscopic Surgery. In: Abdominal Imaging. Computation and Clinical Appl., no. 8198 in LNCS, pp 184–193. Springer
8. Arnold M, Ghosh A, Ameling S, Lacey G (2010) Automatic Segmentation and inpainting of Specular Highlights for Endoscopic Imaging. *EURASIP Journal on Image and Video Processing*:1–12
9. Arnold M, Ghosh A, Lacey G, Patchett S, Mulcahy H (2009) Indistinct Frame Detection in Colonoscopy Videos. In: Machine Vision and Image Processing Conf, 2009. *IMVIP '09*. 13th International, pp 47–52
10. Atasoy S, Mateus D, Lallemand J, Meining A, Yang GZ, Navab N (2010) Endoscopic video manifolds. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*:437–445
11. Barreto J, Roquette J, Sturm P, Fonseca F (2009) Automatic Camera Calibration Applied to Medical Endoscopy. In: 20th British Machine Vision Conference (BMVC '09)
12. Barreto J, Swaminathan R, Roquette J (2007) Non Parametric Distortion Correction in Endoscopic Medical Images. In: 3DTV Conference, 2007, pp 1–4
13. Bashar M, Kitasaka T, Suenaga Y, Mekada Y, Mori K (2010) Automatic detection of informative frames from wireless capsule endoscopy images. *Med Image Anal* 14(3):449–470
14. Baumhauer M, Feuerstein M, Meinzer HP, Rassweiler J (2008) Navigation in Endoscopic Soft Tissue Surgery: Perspectives and Limitations. *J Endourol* 22(4):751–766
15. Bay H, Tuytelaars T, Gool LV (2006) SURF: Speeded Up Robust Features. In: Computer Vision – ECCV 2006, no. 3951 in LNCS, pp 404–417. Springer
16. Beecks C, Schoeffmann K, Lux M, Uysal MS, Seidl T (2015) Endoscopic Video Retrieval: A Signature-based Approach for Linking Endoscopic Images with Video Segments. In: IEEE Int'l Symposium on Multimedia. *ISM*, Miami
17. Behrens A, Takami M, Gross S, Aach T (2011) Gap detection in endoscopic video sequences using graphs. In: 2011 Annual Int'l Conf of the IEEE Engineering in Medicine and Biology Soc, pp 6635–6638
18. Béjar Haro B, Zappella L, Vidal R (2012) Surgical Gesture Classification from Video Data. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*:34–41
19. Bergen T, Wittenberg T (2014) Stitching and Surface Reconstruction from Endoscopic Image Sequences: A Review of Applications and Methods. *IEEE Journal of Biomedical and Health Informatics* PP(99):1–1
20. Bernal J, Gil D, Sánchez C, Sánchez FJ (2014) Discarding Non Informative Regions for Efficient Colonoscopy Image Analysis. In: Computer-Assisted and Robotic Endoscopy, LNCS, pp 1–10. Springer
21. Bernal J, Sánchez J, Vilarino F (2012) Towards automatic polyp detection with a polyp appearance model. *Pattern Recogn* 45(9):3166–3182
22. Bernhardt S, Abi-Nahed J, Abugharbieh R (2013) Robust dense endoscopic stereo reconstruction for minimally invasive surgery. In: Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. Springer, pp 254–262
23. Bichlmeier C, Heining SM, Feuerstein M, Navab N (2009) The Virtual Mirror: A New Interaction Paradigm for Augmented Reality Environments. *IEEE Trans Med Imaging* 28(9):1498–1510
24. Bilodeau GA, Shu Y, Cheriet F (2006) Multistage graph-based segmentation of thoracoscopic images. *Comput Med Imaging Graph* 30(8):437–446
25. Blum T, Fußner H, Navab N. (2010) Modeling and segmentation of surgical workflow from laparoscopic video. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*:400–407
26. Bouarfa L, Akman O, Schneider A, Jonker PP, Dankelman J (2012) In-vivo real-time tracking of surgical instruments in endoscopic video. *Minim Invasive Ther Allied Technol* 21(3):129–134
27. Braga J, Laranjo I, Assunção D, Rolanda C, Lopes L, Correia-Pinto J, Alves V (2013) Endoscopic Imaging Results: Web based Solution with Video Diffusion. *Procedia Technol* 9:1123–1131

28. Braga J, Laranjo I, Rolanda C, Lopes L, Correia-Pinto J, Alves V (2014) A Novel Approach to Endoscopic Exams Archiving. In: *New Perspectives in Information Systems and Technologies*, Volume 1, no. 275 in *Advances in Intelligent Systems and Computing*. Springer, pp 239–248
29. Buck SD, Cleynenbreugel JV, Geys I, Koninckx T, Koninck PR, Suetens P (2001) A System to Support Laparoscopic Surgery by Augmented Reality Visualization. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001*, no. 2208 in *LNCS*, pp 691–698. Springer
30. Burschka D, Corso JJ, Dewan M, Lau W, Li M, Lin H, Marayong P, Ramey N, Hager GD, Hoffman B et al (2005) Navigating inner space: 3-D assistance for minimally invasive surgery. *Robot Auton Syst* 52(1):5–26
31. Burschka D, Li M, Ishii M, Taylor RH, Hager GD (2005) Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. *Med Image Anal* 9(5):413–426
32. Calonder M, Lepetit V, Strecha C, Fua P (2010) BRIEF: Binary Robust Independent Elementary Features. Springer
33. Cano A, Gayá F, Lamata P, Sánchez-González P, Gómez E (2008) Laparoscopic tool tracking method for augmented reality surgical applications. *Biomedical Simulation*:191–196
34. Cao Y, Liu D, Tavanapong W, Wong J, Oh J, de Groen P (2007) Computer-Aided Detection of Diagnostic and Therapeutic Operations in Colonoscopy Videos. *IEEE Trans Biomed Eng* 54(7):1268–1279
35. Cao Y, Tavanapong W, Li D, Oh J, Groen PCD, Wong J (2004) A Visual Model Approach for Parsing Colonoscopy Videos. In: *Image and Video Retrieval*, no. 3115 in *LNCS*, pp 160–169. Springer
36. Casals A, Amat J, Laporte E (1996) Automatic guidance of an assistant robot in laparoscopic surgery. In: *IEEE International Conference on Robotics and Automation*, pp 895–900
37. Chattopadhyay T, Chaki A, Bhowmick B, Pal A (2008) An application for retrieval of frames from a laparoscopic surgical video based on image of query instrument. In: *TENCON 2008-2008 IEEE Region 10 Conference*, pp 1–5
38. Chhatkuli A, Bartoli A, Malti A, Collins T (2014) Live image parsing in uterine laparoscopy. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*
39. Chowdhury M, Kundu MK (2015) Endoscopic Image Retrieval System Using Multi-scale Image Features. In: *Proceedings of the 2Nd International Conference on Perception and Machine Intelligence, PerMin '15*. ACM, New York, NY, USA, pp 64–70
40. Chu X, Poh C, Li L, Chan K, Yan S, Shen W, Htwe T, Liu J, Lim J, Ong E, Ho K (2010) Epitomized Summarization of Wireless Capsule Endoscopic Videos for Efficient Visualization. In: *MICCAI 2010*, no. 6362 in *LNCS*, pp 522–529. Springer
41. Climent J, Hessel RA (2012) Particle filtering in the Hough space for instrument tracking. *Comput Biol Med* 42(5):614–623
42. Collins T, Bartoli A (2012) 3d reconstruction in laparoscopy with close-range photometric stereo. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012*. Springer, pp 634–642
43. Collins T, Bartoli A (2012) Towards Live Monocular 3d Laparoscopy Using Shading and Specularity Information. In: *Inf. Proc. in Computer-Assisted Interventions*, no. 7330 in *LNCS*, pp 11–21. Springer
44. Collins T, Compté B, Bartoli A (2011) Deformable shape-from-motion in laparoscopy using a rigid sliding window. In: *Medical Image Understanding and Analysis Conference*
45. Collins T, Pizarro D, Bartoli A, Canis M, Bourdel N (2014) Computer-Assisted Laparoscopic myomectomy by augmenting the uterus with pre-operative MRI data. In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp 243–248
46. Cunha J, Coimbra M, Campos P, Soares J (2008) Automated Topographic Segmentation and Transit Time Estimation in Endoscopic Capsule Exams. *IEEE Trans. Med. Imaging* 27(1):19–27
47. Dahyot R, Vilarinho F, Lacey G. (2008) Improving the Quality of Color Colonoscopy Videos. *EURASIP Journal on Image and Video Processing* 2008:1–7
48. Deguchi D, Mori K, Feuerstein M, Kitasaka T, Maurer Jr. CR, Suenaga Y, Takabatake H, Mori M, Natori H (2009) Selective image similarity measure for bronchoscope tracking based on image registration. *Med Image Anal* 13(4):621–633
49. Dhandra BV, Hegadi R, Hangarge M, Malemath VS (2006) Analysis of abnormality in endoscopic images using combined hsi color space and watershed segmentation. In: *18th International Conference on Pattern Recognition, 2006*, vol 4, *ICPR 2006*. pp 695–698
50. Dickens M, Bornhop DJ, Mitra S (1998) Removal of optical fiber interference in color micro-endoscopic images. In: *11th IEEE Symposium on Computer-Based Medical Systems*, pp 246–251
51. Dixon B, Daly M, Chan H, Vescan A, Witterick I, Irish J (2013) Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surg Endosc* 27(2):454–461
52. Doignon C, Nageotte F, de Mathelin M (2006) The role of insertion points in the detection and positioning of instruments in laparoscopy for robotic tasks. *MICCAI*:527–534

53. Doignon C, Nageotte F, de Mathelin M (2007) Segmentation and Guidance of Multiple Rigid Objects for Intra-operative Endoscopic Vision. In: Proceedings of the 2005/2006 International Conference on Dynamical Vision, WDV'05/WDV'06/ICCV'05/ECCV'06, pp 314–327. Springer-Verlag, Berlin, Heidelberg
54. Doignon C, Nageotte F, Maurin B, Krupa A (2008) Pose Estimation and Feature Tracking for Robot Assisted Surgery with Medical Imaging. In: Unifying Perspectives in Computational and Robot Vision, no. 8 in Lecture Notes in Electrical Engineering, pp 79–101. Springer, USA
55. Duda K, Zielinski T, Duplaga M (2008) Computationally simple Super-Resolution algorithm for video from endoscopic capsule. In: Int'l Conf. on Signals and Electronic Systems, 2008. ICSES '08, pp 197–200
56. Duplaga M, Leszczuk M, Papir Z, Przelaskowski A (2008) Evaluation of Quality Retaining Diagnostic Credibility for Surgery Video Recordings. In: Visual Information Systems. Web-Based Visual Information Search and Management, no. 5188 in LNCS, pp 227–230. Springer
57. Elter M, Rupp S, Winter C (2006) Physically Motivated Reconstruction of Fiberscopic Images. In: 18th International Conference on Pattern Recognition, 2006. ICPR 2006, vol 3, pp 599–602
58. Forestier G, Lalys F, Riffaud L, Trelhu B, Jannin P (2012) Classification of surgical processes using dynamic time warping. *J Biomed Inform* 45(2):255–264
59. Fuchs H, Livingston MA, Raskar R, State A, Crawford JR, Rademacher P, Drake SH, Meyer AA (1998) Augmented reality visualization for laparoscopic surgery. In: Proceedings of the 1st Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention. Springer, pp 934–943
60. Fukuda N, Chen YW, Nakamoto M, Okada T, Sato Y (2010) A scope cylinder rotation tracking method for oblique-viewing endoscopes without attached sensing device. In: 2010 2nd International Conference on Software Engineering and Data Mining (SEDM), pp 684–687
61. Gambadauro P, Magos A (2012) Surgical Videos for Accident Analysis, Performance Improvement, and Complication Prevention: Time for a Surgical Black Box? *Surg Innov* 19(1):76–80
62. Gavião W, Scharcanski J, Frahm JM, Pollefeys M (2012) Hysteroscopy video summarization and browsing by estimating the physician's attention on video segments. *Med Image Anal* 16(1):160–176
63. Geng J, Xie J (2014) Review of 3-D Endoscopic Surface Imaging Techniques. *IEEE Sensors J* 14(4):945–960
64. Giannarou S, Visentini-Scarzarella M, Yang GZ (2009) Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery. In: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on Biomedical Imaging, pp 1059–1062
65. Giannarou S, Yang GZ (2010) Content-Based Surgical Workflow Representation Using Probabilistic Motion Modeling. In: *Med. Imaging and Aug. Reality*, no. 6326 in LNCS, pp 314–323. Springer
66. Giritharan B, Yuan X, Liu J, Buckles B, Oh J, Tang S (2008) Bleeding detection from capsule endoscopy videos. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008. EMBS 2008, pp 4780–4783
67. Grasa O, Civera J, Montiel JMM (2011) EKF monocular SLAM with relocalization for laparoscopic sequences. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp 4816–4821
68. Grega M, Leszczuk M, Duplaga M, Fraczek R (2010) Algorithms for Automatic Recognition of Non-informative Frames in Video Recordings of Bronchoscopic Procedures. In: *Information Technologies in Biomedicine, Advances in Intelligent and Soft Computing*, vol 69. Springer, pp 535–545
69. Groch A, Seitel A, Hempel S, Speidel S, Engelbrecht R, Penne J, Höller K, Röhl S, Yung K, Bodenstedt S, Pflaum F, dos Santos TR, Mersmann S, Meinzer HP, Hornegger J, Maier-Hein L. (2011) 3d surface reconstruction for laparoscopic computer-assisted interventions: comparison of state-of-the-art methods. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conf Series*, vol 7964, p 796415
70. Gröger M, Ortmaier T, Hirzinger G (2005) Structure Tensor Based Substitution of Specular Reflections for Improved Heart Surface Tracking. In: *Bildverarbeitung f. d. Medizin 2005, Informatik aktuell*. Springer, pp 242–246
71. Gröger M, Sepp W, Ortmaier T, Hirzinger G (2001) Reconstruction of Image Structure in Presence of Specular Reflections. In: *Proceedings of the 23rd DAGM-Symp. on Pattern Recognition*. Springer, pp 53–60
72. Gschwandtner M, Liedlgruber M, Uhl A, Vécsei A (2010) Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In: 2010 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB), pp 1–6
73. Guggenberger M, Riegler M, Lux M, Halvorsen P (2014) Event Understanding in Endoscopic Surgery Videos. In: *Proceedings of the 1st ACM International Workshop on Human Centered Event Understanding from Multimedia, HuEvent '14*, pp 17–22. ACM, New York, NY, USA

74. Guthart G, Salisbury Jr JK (2000) The IntuitiveTM Telesurgery System: Overview and Application. In: ICRA, pp 618–621
75. Hafner M, Liedlgruber M, Uhl A (2013) POCS-based super-resolution for HD endoscopy video frames. In: 2013 IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS), pp 185–190
76. Hafner M, Liedlgruber M, Uhl A, Wimmer G (2014) Evaluation of super-resolution methods in the context of colonic polyp classification. In: 12th Int'l Workshop on Content-Based Multimedia Indexing
77. Helferty JP, Zhang C, McLennan G, Higgins WE (2001) Videoendoscopic distortion correction and its application to virtual guidance of endoscopy. *IEEE Trans Med Imaging* 20(7):605–617
78. Hernández-Mier Y, Blondel W, Daul C, Wolf D, Guillemin F (2010) Fast construction of panoramic images for cystoscopic exploration. *Comput Med Imaging Graph* 34(7):579–592
79. Higgins WE, Helferty JP, Lu K, Merritt SA, Rai L, Yu KC (2008) 3d CT-Video Fusion for Image-Guided Bronchoscopy. *Comput Med Imaging Graph* 32(3):159–173
80. Holler K, Penne J, Hornegger J, Schneider A, Gillen S, Feussner H, Jahn J, Gutierrez J, Wittenberg T (2009) Clinical evaluation of Endorientation: Gravity related rectification for endoscopic images. In: Proceedings of 6th Int'l Symp. on Image and Signal Processing and Analysis, ISPA 2009, pp 713–717
81. Hu M, Penney G, Figl M, Edwards P, Bello F, Casula R, Rueckert D, Hawkes D (2012) Reconstruction of a 3d surface from video that is robust to missing data and outliers: Application to minimally invasive surgery using stereo and mono endoscopes. *Med Image Anal* 16(3):597–611
82. Hughes-Hallett A, Mayer EK, Marcus HJ, Cundy TP, Pratt PJ, Darzi AW, Vale JA (2014) Augmented Reality Partial Nephrectomy: Examining the Current Status and Future Perspectives. *Urology* 83(2):266–273
83. Hwang S, Celebi M (2010) Polyp detection in Wireless Capsule Endoscopy videos based on image segmentation and geometric feature. In: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp 678–681
84. Hwang S, Oh JH, Lee JK, Cao Y, Tavanapong W, Liu D, Wong J, de Groen PC (2005) Automatic measurement of quality metrics for colonoscopy videos. In: Proceedings of the 13th annual ACM international conference on Multimedia, pp 912–921
85. Hwang S, Oh JH, Tavanapong W, Wong J, de Groen PC (2008) Stool detection in colonoscopy videos. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008. EMBS 2008. pp 3004–3007
86. Iakovidis DK, Maroulis DE, Karkanis SA, Brokos A (2005) A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. In: Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems, pp 575–580
87. Iakovidis DK, Tsevas S, Polydorou A (2010) Reduction of capsule endoscopy reading times by unsupervised image mining. *Comput Med Imaging Graph* 34(6):471–478
88. Ieiri S, Uemura M, Konishi K, Souzaki R, Nagao Y, Tsutsumi N, Akahoshi T, Ohuchida K, Ohdaira T, Tomikawa M, Tanoue K, Hashizume M, Taguchi T (2012) Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative CT image using optical tracking device. *Pediatr Surg Int* 28(4):341–346
89. Jun SK, Narayanan M, Agarwal P, Eddib A, Singhal P, Garimella S, Krovi V (2012) Robotic Minimally Invasive Surgical skill assessment based on automated video-analysis motion studies. In: 2012 4th IEEE RAS EMBS Int'l Con. on Biomedical Robotics and Biomechatronics (BioRob), pp 25–31
90. Kallemeyn NA, Grosland NM, Magnotta VA, Martin JA, Pedersen DR (2007) Arthroscopic Lens Distortion Correction Applied to Dynamic Cartilage Loading. *Iowa Orthop J* 27:52–57
91. Kanaya J, Koh E, Namiki M, Yokawa H, Kimura H, Abe K (2008) A system for detecting locations of oversight in cystoscopy. In: Proceedings of the 9th Asia Pacific Industrial Engineering & Management Systems Conf., pp 2388–2392
92. Karargyris A, Bourbakis N (2010) Wireless Capsule Endoscopy and Endoscopic Imaging: A Survey on Various Methodologies Presented. *IEEE Eng Med Biol Mag* 29(1):72–83
93. Karkanis S, Iakovidis D, Maroulis D, Karras D, Tzivras M (2003) Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans Inf Technol Biomed* 7(3):141–152
94. Katić D, Wekerle AL, Gärtner F, Kenngott HG, Müller-Stich BP, Dillmann R., Speidel S. (2014) Knowledge-Driven Formalization of Laparoscopic Surgeries for Rule-Based Intraoperative Context-Aware Assistance. In: Information Processing in Computer-Assisted Interventions, no. 8498 in LNCS. Springer, pp 158–167
95. Katić D, Wekerle AL, Görtler J, Spengler P, Bodenstedt S, Röhl S, Suwelack S, Kenngott HG, Wagner M, Müller-Stich BP, Dillmann R, Speidel S (2013) Context-aware Augmented Reality in laparoscopic surgery. *Comput Med Imaging Graph* 37(2):174–182

96. Kelley WE (2008) The Evolution of Laparoscopy and the Revolution in Surgery in the Decade of the 1990s. *JLS : Journal of the Society of Laparoendoscopic Surgeons* 12(4):351–357
97. Khatibi T, Sepehri MM, Shadpour P (2014) A Novel Unsupervised Approach for Minimally-invasive Video Segmentation. *Journal of Medical Signals and Sensors* 4(1):53–71
98. Klank U, Padoy N, Feussner H, Navab N (2008) Automatic feature generation in endoscopic images. *Int J Comput Assist Radiol Surg* 3(3–4):331–339
99. Ko SY, Lee WJ, Kwon DS (2010) Intelligent interaction based on a surgery task model for a surgical assistant robot: Awareness of current surgical stages based on a surgical procedure model. *Int J Control Autom Syst* 8(4):782–792
100. Kondo W, Zomer MT (2014) Video recording the Laparoscopic Surgery for the Treatment of Endometriosis should be Systematic! *Gynecol Obstet* 04(04)
101. Koninckx PR (2008) Videoregistration of Surgery Should be Used as a Quality Control. *J Minim Invasive Gynecol* 15(2):248–253
102. Koppel D, Wang YF, Lee H (2004) Image-based rendering and modeling in video-endoscopy. In: *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*, vol 1, pp 269–272
103. Kranzfelder M, Schneider A, Fiolka A, Schwan E, Gillen S, Wilhelm D, Schirren R, Reiser S, Jensen B, Feussner H (2013) Real-time instrument detection in minimally invasive surgery using radiofrequency identification technology. *J Surg Res* 185(2):704–710
104. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
105. Krupa A, Gangloff J, Doignon C, de Mathelin M, Morel G, Leroy J, Soler L, Marescaux J (2003) Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *IEEE Trans Robot Autom* 19(5):842–853
106. Kumar A, Kim J, Cai W, Fulham M, Feng D (2013) Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data. *J Digit Imaging* 26(6):1025–1039
107. Kumar S, Narayanan M, Singhal P, Corso J, Krovi V (2013) Product of tracking experts for visual tracking of surgical tools. In: *2013 IEEE International Conference on Automation Science and Engineering (CASE)*, pp 480–485
108. Kumar S, Narayanan M, Singhal P, Corso J, Krovi V (2014) Surgical tool attributes from monocular video. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp 4887–4892
109. Lahane A, Yesha Y, Grasso MA, Joshi A, Martineau J, Mokashi R, Chapman D, Grasso MA, Brady M, Yesha Y et al (2011) Detection of Unsafe Action from Laparoscopic Cholecystectomy Video. In: *Proceedings of ACM SIGHT International Health Informatics Symposium*, vol 47, pp 114–122
110. Lalys F, Jannin P (2014) Surgical process modelling: a review. *Int J Comput Assist Radiol Surg* 9(3):495–511
111. Lalys F, Riffaud L, Morandi X, Jannin P (2011) Surgical phases detection from microscope videos by combining SVM and HMM. In: *Proceedings of the 2010 Int'l MICCAI conference on Medical computer vision: recognition techniques and applications in medical imaging, MCV'10*. Springer, pp 54–62
112. Laranjo I, Braga J, Assunção D, Silva A, Rolanda C, Lopes L, Correia-Pinto J, Alves V (2013) Web-Based Solution for Acquisition, Processing, Archiving and Diffusion of Endoscopy Studies. In: *Distributed Computing and Artificial Intelligence*, no. 217 in *AISC*, pp 317–324. Springer
113. Lau WY, Leow CK, Li AKC (1997) History of Endoscopic and Laparoscopic Surgery. *World J Surg* 21(4):444–453
114. Lee J, Oh J, Shah SK, Yuan X, Tang SJ (2007) Automatic classification of digestive organs in wireless capsule endoscopy videos. In: *Proce. of the 2007 ACM Symp. on Applied Computing, SAC '07*. ACM, pp 1041–1045
115. Lee SL, Lerotic M, Vitiello V, Giannarou S, Kwok KW, Visentini-Scarzanella M, Yang GZ (2010) From medical images to minimally invasive intervention: Computer assistance for robotic surgery. *Comput Med Imaging Graph* 34(1):33–45
116. Leong JJ, Nicolaou M, Atallah L, Mylonas GP, Darzi AW, Yang GZ (2006) HMM assessment of quality of movement trajectory in laparoscopic surgery. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, pp 752–759. Springer
117. Leszczuk MI, Duplaga M (2011) Algorithm for Video Summarization of Bronchoscopy Procedures. *BioMedical Engineering OnLine* 10(1):110
118. Li B, Meng MH (2012) Tumor Recognition in Wireless Capsule Endoscopy Images Using Textural Features and SVM-Based Feature Selection. *IEEE Trans Inf Technol Biomed* 16(3):323–329
119. Li B, Meng MH (2009) Computer-Aided Detection of Bleeding Regions for Capsule Endoscopy Images. *IEEE Trans Biomed Eng* 56(4):1032–1039

120. Liao H, Tsuzuki M, Mochizuki T, Kobayashi E, Chiba T, Sakuma I (2009) Fast image mapping of endoscopic image mosaics with three-dimensional ultrasound image for intrauterine fetal surgery. Minimally invasive therapy & allied technologies: MITAT: official journal of the Society for Minimally Invasive Therapy 18(6):332–340
121. Liao R, Zhang L, Sun Y, Miao S, Chefed'hotel C (2013) A Review of Recent Advances in Registration Techniques Applied to Minimally Invasive Therapy. IEEE Trans Multimedia 15(5):983–1000
122. Liedlgruber M, Uhl A (2011) Computer-Aided Decision Support Systems for Endoscopy in the Gastrointestinal Tract: A Review. IEEE Rev Biomed Eng 4:73–88
123. Liedlgruber M, Uhl A, Vecsei A (2011) Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In: 2011 17th Int'l Conf. on Digital Signal Processing, pp 1–6
124. Lin HC, Shafran I, Yuh D, Hager GD (2006) Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. Comput Aided Surg 11(5):220–230
125. Liu D, Cao Y, Kim KH, Stanek S, Doungtranaex-Chai B, Lin K, Tavanapong W, Wong J, Oh J, de Groen PC (2007) Arthemis: Annotation software in an integrated capturing and analysis system for colonoscopy. Comput Methods Prog Biomed 88(2):152–163
126. Liu D, Cao Y, Tavanapong W, Wong J, Oh JH, de Groen PC (2007) Quadrant coverage histogram: a new method for measuring quality of colonoscopic procedures. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007, pp 3470–3473
127. Lo B, Scarzanella MV, Stoyanov D, Yang GZ (2008) Belief Propagation for Depth Cue Fusion in Minimally Invasive Surgery. In: MICCAI 2008, no. 5242 in LNCS, pp 104–112. Springer
128. Lokoč J, Schoeffmann K, del Fabro M (2015) Dynamic Hierarchical Visualization of Keyframes in Endoscopic Video. In: MultiMedia Modeling, no. 8936 in LNCS, pp 291–294. Springer
129. Loukas C, Georgiou E (2015) Smoke detection in endoscopic surgery videos: a first step towards retrieval of semantic events. Int J Med Rob Comput Assisted Surg 11(1):80–94
130. Lowe DG (2004) Distinctive Image Features from Scale-Invariant Keypoints. Int J Comput Vis 60(2):91–110
131. Luó X, Feuerstein M, Deguchi D, Kitasaka T, Takabatake H, Mori K (2012) Development and comparison of new hybrid motion tracking for bronchoscopic navigation. Med Image Anal 16(3):577–596
132. Lux M, Marques O, Schöffmann K, Böszörményi L, Lajtai G (2009) A novel tool for summarization of arthroscopic videos. Multimedia Tools and Applications 46(2–3):521–544
133. Lux M, Riegler M (2013) Annotation of endoscopic videos on mobile devices: a bottom-up approach. In: Proceedings of the 4th ACM Multimedia Systems Conference. ACM, pp 141–145
134. Mackiewicz M, Berens J, Fisher M (2008) Wireless Capsule Endoscopy Color Video Segmentation. IEEE Trans Med Imaging 27(12):1769–1781
135. Maghsoudi O, Talebpour A, Sotanian-Zadeh H, Alizadeh M, Soleimani H (2014) Informative and Uninformative Regions Detection in WCE Frames Journal of Advanced Computing
136. Maier-Hein L, Mountney P, Bartoli A, Elhawary H, Elson D, Groch A, Kolb A, Rodrigues M, Sorger J, Speidel S, Stoyanov D (2013) Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. Med Image Anal 17(8):974–996
137. Makary MA (2013) The Power of Video Recording: Taking Quality to the Next Level. JAMA 309(15):1591
138. Malti A (2014) Variational Formulation of the Template-Based Quasi-Conformal Shape-from-Motion from Laparoscopic Images Int'l Journal of Advanced Computer Science and Applications (IJACSA)
139. Malti A, Bartoli A (2014) Combining Conformal Deformation and Cook-Torrance Shading for 3-D Reconstruction in Laparoscopy. IEEE Trans Biomed Eng 61(6):1684–1692
140. Marayong P, Li M, Okamura A, Hager G (2003) Spatial motion constraints: theory and demonstrations for robot guidance using virtual fixtures. In: IEEE International Conference on Robotics and Automation, 2003. Proceedings. ICRA '03, vol 2, pp 1954–1959
141. Markelj P, Tomazevic D, Likar B, Pernuš F (2012) A review of 3d/2d registration methods for image-guided interventions. Med Image Anal 16(3):642–661
142. Matas J, Chum O, Urban M, Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. Image Vis Comput 22(10):761–767
143. Meslouhi O, Kardouchi M, Allali H, Gadi T, Benkaddour YA (2011) Automatic detection and inpainting of specular reflections for colposcopic images. Central European Journal of Computer Science 1(3):341–354
144. Miranda-Luna R, Daul C, Blondel W, Hernandez-Mier Y, Wolf D, Guillemin F (2008) Mosaicing of Bladder Endoscopic Image Sequences: Distortion Calibration and Registration Algorithm. IEEE Transactions on Biomedical Engineering 55(2):541–553

145. Mirotta D, Wang H, Taylor R, Ishii M, Hager G (2009) Toward video-based navigation for endoscopic endonasal skull base surgery. *MICCAI* 2009:91–99
146. Mirotta DJ, Ishii M, Hager GD (2011) Vision-Based Navigation in Image-Guided Interventions. *Annu Rev Biomed Eng* 13(1):297–319
147. Mirotta DJ, Uneri A, Schafer S, Nithiananthan S, Reh DD, Gallia GL, Taylor RH, Hager GD, Siewerdsen JH (2011) High-accuracy 3d image-based registration of endoscopic video to C-arm cone-beam CT for image-guided skull base surgery. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol 7964, p 79640J
148. Münzer B, Schoeffmann K, Böszörmenyi L (2016) Domain-Specific Video Compression for Long-Term Archiving of Endoscopic Surgery Videos. In: *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, pp 312–317. doi:10.1109/CBMS.2016.28.00000
149. Moll M, Koninckx T, Van Gool LJ, Koninckx PR (2009) Unrotating images in laparoscopy with an application for 30° laparoscopes. In: *4th European conference of the international federation for medical and biological engineering*. Springer, pp 966–969
150. Mountney P, Lo B, Thiemjarus S, Stoyanov D, Yang GZ (2007) A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery. *MICCAI* 2007:34–41
151. Mountney P, Stoyanov D, Yang GZ Recovering Tissue Deformation and Laparoscope Motion for minimally invasive Surgery. Tutorial paper. <http://www.mountney.org.uk/publications/SPM%202010.pdf>
152. Mountney P, Stoyanov D, Yang GZ (2010) Three-Dimensional Tissue Deformation Recovery and Tracking. *IEEE Signal Process Mag* 27(4):14–24
153. Mountney P, Yang GZ (2009) Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009*, pp 1184–1187
154. Mountney P, Yang GZ (2010) Motion compensated SLAM for image guided surgery. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*:496–504
155. Mountney P, Yang GZ (2012) Context specific descriptors for tracking deforming tissue. *Med Image Anal* 16(3):550–561
156. Moustris GP, Hiridis SC, Deliparaschos KM, Konstantinidis KM (2011) Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *Int J Med Rob Comput Assisted Surg* 7(4):375–392
157. Müller H, Michoux N, Bandon D, Geissbuhler A (2004) A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. *Int J Med Inform* 73(1):1–23
158. Münzer B (2011) Requirements and Prototypes for a Content-Based Endoscopic Video Management System. Master's Thesis. Alpen-Adria Universität Klagenfurt, Klagenfurt
159. Münzer B, Schoeffmann K, Böszörmenyi L (2013) Detection of circular content area in endoscopic videos. In: *2013 IEEE 26th Int'l Symp. on Computer-Based Medical Systems (CBMS)*, pp 534–536
160. Münzer B, Schoeffmann K, Böszörmenyi L (2013) Improving encoding efficiency of endoscopic videos by using circle detection based border overlays. In: *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp 1–4
161. Münzer B, Schoeffmann K, Böszörmenyi L (2013) Relevance Segmentation of Laparoscopic Videos. In: *2013 IEEE International Symposium on Multimedia (ISM)*, pp 84–91
162. Münzer B, Schoeffmann K, Böszörmenyi L, Smulders J, Jakimowicz J (2014) Investigation of the Impact of Compression on the Perceptual Quality of Laparoscopic Videos. In: *2014 IEEE 27th International Symposium on Computer-Based Medical einfarbig Systems (CBMS)*, pp 153–158
163. Muthukudage J, Oh J, Nawarathna R, Tavanapong W, Wong J, de Groen PC (2014) Fast Object Detection Using Color Features for Colonoscopy Quality Measurements. In: *Abdomen and Thoracic Imaging*, pp 365–388. Springer
164. Muthukudage J, Oh J, Tavanapong W, Wong J, de Groen PC (2012) Color Based Stool Region Detection in Colonoscopy Videos for Quality Measurements. In: *Advances in Image and Video Technology*, no. 7087 in LNCS, pp 61–72. Springer
165. Nageotte F, Zanne P, Doignon C, de Mathelin M (2006) Visual servoing-based endoscopic path following for robot-assisted laparoscopic surgery. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp 2364–2369
166. Neumuth T, Jannin P, Schlomberg J, Meixensberger J, Wiedemann P, Burgert O (2011) Analysis of surgical intervention populations using generic surgical process models. *Int J Comput Assist Radiol Surg* 6(1):59–71
167. Nicolau S, Soler L, Mutter D, Marescaux J (2011) Augmented reality in laparoscopic surgical oncology. *Surg Oncol* 20(3):189–201

168. Oh J, Hwang S, Cao Y, Tavanapong W, Liu D, Wong J, de Groen P (2009) Measuring Objective Quality of Colonoscopy. *IEEE Trans Biomed Eng* 56(9):2190–2196
169. Oh J, Hwang S, Lee J, Tavanapong W, Wong J, de Groen PC (2007) Informative frame classification for endoscopy video. *Med Image Anal* 11(2):110–127
170. Oh JH, Rajbal MA, Muthukudage JK, Tavanapong W, Wong J, de Groen PC (2009) Real-time phase boundary detection in colonoscopy videos. In: *ISPA 2009. Proceedings of 6th International Symposium on Image and Signal Processing and Analysis*, 2009, pp 724–729
171. Okatani T, Deguchi K (1997) Shape Reconstruction from an Endoscope Image by Shape from Shading Technique for a Point Light Source at the Projection Center. *Comput Vis Image Underst* 66(2):119–131
172. Oropesa I, Sánchez-González P, Chmarra MK, Lamata P, Fernández A, Sánchez-Margallo JA, Jansen FW, Dankelman J, Sánchez-Margallo FM, Gómez EJ (2013) EVA: Laparoscopic Instrument Tracking Based on Endoscopic Video Analysis for Psychomotor Skills Assessment. *Surg Endosc* 27(3):1029–1039
173. Oropesa I, Sánchez-González P, Lamata P, Chmarra MK, Pagador JB, Sánchez-Margallo JA, Sánchez-Margallo FM, Gómez EJ (2011) Methods and Tools for Objective Assessment of Psychomotor Skills in Laparoscopic Surgery. *J Surg Res* 171(1):e81–e95
174. Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO, Navab N (2012) Statistical modeling and recognition of surgical workflow. *Med Image Anal* 16(3):632–641
175. Padoy N, Hager G (2011) Human-Machine Collaborative surgery using learned models. In: *2011 IEEE International Conference on Robotics and Automation (ICRA)*, pp 5285–5292
176. Parchami M, Mariottini GL (2014) A comparative study on 3-D stereo reconstruction from endoscopic images. In: *Proceedings 7th Int'l Conf. on Pervasive Technologies Related to Assistive Environments*, p 25. ACM
177. Park S, Howe RD, Torchiana DF (2001) Virtual Fixtures for Robotic Cardiac Surgery. In: *MICCAI 2001*, no. 2208 in LNCS, pp 1419–1420. Springer
178. Parot V, Lim D, González G, Traverso G, Nishioka NS, Vakoc BJ, Durr NJ (2013) Photometric stereo endoscopy. *J Biomed Opt* 18(7):076,017–076,017
179. Penne J, Höller K, Stürmer M, Schrauder T, Schneider A, Engelbrecht R, Feußner H, Schmauss B, Hornegger J (2009) Time-of-Flight 3-D Endoscopy. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, no. 5761 in LNCS, pp 467–474. Springer
180. Pezzementi Z, Voros S, Hager G (2009) Articulated object tracking by rendering consistent appearance parts. In: *IEEE International Conference on Robotics and Automation*, 2009. *ICRA '09*, pp 3940–3947
181. Prasath V, Figueiredo I, Figueiredo P, Palaniappan K (2012) Mucosal region detection and 3d reconstruction in wireless capsule endoscopy videos using active contours. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp 4014–4017
182. Primus MJ, Schoeffmann K, Böszörmenyi L (2013) Segmentation of recorded endoscopic videos by detecting significant motion changes. In: *2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp 223–228
183. Primus MJ, Schoeffmann K, Böszörmenyi L (2015) Instrument Classification in Laparoscopic Videos. In: *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*. Prague
184. Primus MJ, Schoeffmann K, Böszörmenyi L (2016) Temporal segmentation of laparoscopic videos into surgical phases. In: *2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp 1–6. doi:[10.1109/CBMI.2016.7500249](https://doi.org/10.1109/CBMI.2016.7500249)
185. Przelaskowski A, Jozwiak R (2008) Compression of Bronchoscopy Video: Coding Usefulness and Efficiency Assessment. In: *Information Technologies in Biomedicine*, no. 47 in *Advances in Soft Computing*, pp 208–216. Springer
186. Puerto GA, Mariottini GL (2012) A Comparative Study of Correspondence-Search Algorithms in MIS Images. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, no. 7511 in LNCS, pp 625–633. Springer
187. Puerto-Souza G, Mariottini GL (2013) A Fast and Accurate Feature-Matching Algorithm for Minimally-Invasive Endoscopic Images. *IEEE Trans Med Imaging* 32(7):1201–1214
188. Puerto-Souza GA, Mariottini GL (2014) Wide-Baseline Dense Feature Matching for Endoscopic Images. In: *Image and Video Technology*, no. 8333 in LNCS, pp 48–59. Springer
189. Rangseekajee N, Phongsuphap S (2011) Endoscopy video frame classification using edge-based information analysis. In: *Computing in Cardiology*, 2011, pp 549–552
190. Reeff M, Gerhard F, Cattin PC, Székely G (2006) Mosaicing of endoscopic placenta images. *GI Jahrestagung* 2006(1):467–474

191. Reiter A, Allen PK, Zhao T (2014) Appearance learning for 3d tracking of robotic surgical tools. *Int J Robot Res* 33(2):342–356
192. Richa R, Bó APL, Poignet P (2011) Towards robust 3d visual tracking for motion compensation in beating heart surgery. *Med Image Anal* 15(3):302–315
193. Riegler M, Lux M, Gridwodz C, Spampinato C, de Lange T, Eskeland SL, Pogorelov K, Tavanapong W, Schmidt PT, Gurrin C, Johansen D, Johansen HA, Halvorsen PA (2016) Multimedia and Medicine: Teammates for Better Disease Detection and Survival. In: Proceedings of the 2016 ACM on Multimedia Conference, MM '16, pp 968–977. ACM, New York, NY, USA. doi:10.1145/2964284.2976760. 00000
194. Röhl S, Bodenstedt S, Suwelack S, Kenngott H, Müller-Stich BP, Dillmann R, Speidel S (2012) Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. *Med Phys* 39(3):1632–1645
195. Roldan Carlos J, Lux M, Giro-i Nieto X, Munoz P, Anagnostopoulos N (2015) Visual information retrieval in endoscopic video archives. In: 2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI), pp 1–6
196. Rosen J, Brown J, Chang L, Sinanan M, Hannaford B (2006) Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. *IEEE Trans Biomed Eng* 53(3):399–413
197. Rosen J, Hannaford B, Richards C, Sinanan M (2001) Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Trans Biomed Eng* 48(5):579–591
198. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB :An efficient alternative to SIFT or SURF. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp 2564–2571
199. Rungseekajee N, Lohvithee M, Nilkhamhang I (2009) Informative frame classification method for real-time analysis of colonoscopy video. In: 6th Int'l Conf on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2009, ECTI-CON 2009, vol 2, pp 1076–1079
200. Rupp S, Elter M, Winter C (2007) Improving the Accuracy of Feature Extraction for Flexible Endoscope Calibration by Spatial Super Resolution. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007, pp 6565–6571
201. Saint-Pierre CA, Boisvert J, Grimard G, Cherié F (2011) Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Mach Vis Appl* 22(1):171–180
202. Sauvée M, Noce A, Poignet P, Triboulet J, Dombre E (2007) Three-dimensional heart motion estimation using endoscopic monocular vision system: From artificial landmarks to texture analysis. *Biomed Signal Process Control* 2(3):199–207
203. Scharcanski J, Gaviao W (2006) Hierarchical summarization of diagnostic hysteroscopy videos. In: 2006 IEEE International Conference on Image Processing, pp 129–132
204. Schoeffmann K, del Fabro M, Szkaliczki T, Böszörményi L, Keckstein J (2014) Keyframe extraction in endoscopic video. *Multimedia Tools and Applications*:1–20
205. Selka F, Nicolau S, Agnus V, Bessaid A, Marescaux J, Soler L (2015) Context-specific selection of algorithms for recursive feature tracking in endoscopic image using a new methodology. *Comput Med Imaging Graph* 40:49–61
206. Shen Y, Guturu P, Buckles B (2012) Wireless Capsule Endoscopy Video Segmentation Using an Unsupervised Learning Approach Based on Probabilistic Latent Semantic Analysis With Scale Invariant Features. *IEEE Trans Inf Technol Biomed* 16(1):98–105
207. Sheraizin S, Sheraizin V (2006) Endoscopy Imaging Intelligent Contrast Improvement. In: IEEE-EMBS 2005. 27th Annual Int'l Conf of the Engineering in Medicine and Biology Society, 2005, pp 6551–6554
208. Shih T, Chang RC (2005) Digital inpainting - survey and multilayer image inpainting algorithms. In: Third Int'l Conf. on Information Technology and Applications, 2005. ICITA 2005, vol 1, pp 15–24
209. Sielhorst T, Feuerstein M, Navab N (2008) Advanced Medical Displays: A Literature Review of Augmented Reality. *J Disp Technol* 4(4):451–467
210. Simpfendorfer T, Baumhauer M, Müller M, Gutt CN, Meinzer HP, Rassweiler JJ, Guven S, Teber D (2011) Augmented Reality Visualization During Laparoscopic Radical Prostatectomy. *J Endourol* 25(12):1841–1845
211. Song KT, Chen CJ (2012) Autonomous and stable tracking of endoscope instrument tools with monocular camera. In: 2012 IEEE/ASME Int'l Conf. on Advanced Intelligent Mechatronics (AIM), pp 39–44
212. Soper TD, Porter MP, Seibel EJ (2012) Surface Mosaics of the Bladder Reconstructed From Endoscopic Video for Automated Surveillance. *IEEE Trans Biomed Eng* 59(6):1670–1680

213. Speidel S, Benzko J, Krappe S, Sudra G, Azad P, Müller-Stich BP, Gutt C, Dillmann R (2009) Automatic classification of minimally invasive instruments based on endoscopic image sequences. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol 7261, p 72610A
214. Speidel S, Sudra G, Senemaud J, Drementschew M, Müller-Stich BP, Gutt C, Dillmann R (2008) Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conf. Series, vol 6918, p 69180X
215. Spyrou E, Diamantis D, Iakovidis D (2013) Panoramic Visual Summaries for Efficient Reading of Capsule Endoscopy Videos. In: 2013 8th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), pp 41–46
216. Srinivasan N, Stanek S, Szewczynski MJ, Enders F, Tavanapong W, Oh J, Wong J, Groen PCD (2013) Automated Real-Time Feedback During Colonoscopy Improves Endoscopist Technique As Intended: a Controlled Clinical Trial. *Gastrointest Endosc* 77(5):AB500. 00000
217. Stanek SR, Tavanapong W, Wong J, Oh JH, De Groen PC (2012) Automatic real-time detection of endoscopic procedures using temporal features. *Comput Methods Prog Biomed* 108(2):524–535
218. Stanek SR, Tavanapong W, Wong JS, Oh J, de Groen PC (2008) Automatic real-time capture and segmentation of endoscopy video. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol 6919, p 69190X
219. Staub C, Lenz C, Panin G, Knoll A, Bauernschmitt R (2010) Contour-based surgical instrument tracking supported by kinematic prediction. In: 2010 3rd IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp 746–752
220. Staub C, Osa T, Knoll A, Bauernschmitt R (2010) Automation of tissue piercing using circular needles and vision guidance for computer aided laparoscopic surgery. In: IEEE International Conference on Robotics and Automation (ICRA), 2010, pp 4585–4590
221. Stauder R, Okur A, Peter L, Schneider A, Kranzfelder M, Feussner H, Navab N (2014) Random Forests for Phase Detection in Surgical Workflow Analysis. In: Information Processing in Computer-Assisted Interventions, no. 8498 in LNCS, pp 148–157. Springer
222. Stehle T (2006) Removal of specular reflections in endoscopic images. *Acta Polytechnica: Journal of Advanced Engineering* 46(4):32–36
223. Stehle T, Hennes M, Gross S, Behrens A, Wulff J, Aach T (2009) Dynamic Distortion Correction for Endoscopy Systems with Exchangeable Optics. In: Bildverarbeitung für die Medizin 2009, Informatik aktuell, pp 142–146. Springer
224. Stoyanov D, Mylonas GP, Deligianni F, Darzi A, Yang GZ (2005) Soft-Tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005, no. 3750 in LNCS, pp 139–146. Springer
225. Stoyanov D, Scarzanella MV, Pratt P, Yang GZ (2010) Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010, no. 6361 in LNCS, pp 275–282. Springer
226. Stoyanov D, Yang GZ (2005) Removing specular reflection components for robotic assisted laparoscopic surgery. In: IEEE International Conference on Image Processing, 2005. ICIP 2005, vol 3, pp III–632–5
227. Su LM, Vagvolgyi BP, Agarwal R, Reiley CE, Taylor R, Hager GD (2009) Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3d-CT to Stereoscopic Video Registration. *Urology* 73(4):896–900
228. Sudra G, Speidel S, Fritz D, Müller-Stich BP, Gutt C, Dillmann R (2007) MEDIASSIST: medical assistance for intraoperative skill transfer in minimally invasive surgery using augmented reality. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol 6509, p 65091O
229. Sugimoto M, Yasuda H, Koda K, Suzuki M, Yamazaki M, Tezuka T, Kosugi C, Higuchi R, Watayo Y, Yagawa Y, Uemura S, Tsuchiya H, Azuma T (2010) Image overlay navigation by markerless surface registration in gastrointestinal, hepatobiliary and pancreatic surgery. *J Hepatobiliary Pancreat Sci* 17(5):629–636
230. Sung G, Gill IS (2001) Robotic laparoscopic surgery: a comparison of the da Vinci and Zeus systems. *Urology* 58(6):893–898
231. Suwelack S, Röhl S, Dillmann R, Wekerle AL, Kenngott H, Müller-Stich B, Alt C, Speidel S (2012) Quadratic Corotated Finite Elements for Real-Time Soft Tissue Registration. In: Computational Biomechanics for Medicine, pp 39–50. Springer, New York
232. Szczypiński P, Klepaczko A, Pazurek M, Daniel P (2014) Texture and color based image segmentation and pathology detection in capsule endoscopy videos. *Comput Methods Prog Biomed* 113(1):396–411

233. Tai XY, Wang LD, Chen Q, Fuji R, Kenji K (2009) A New Method Of Medical Image Retrieval Based On Color–Texture Correlogram And Gti Model. *Int J Inf Technol Decis Mak* 08(02):239–248
234. Taylor R (2006) A Perspective on Medical Robotics. *Proc IEEE* 94(9):1652–1664
235. Taylor R, Stoianovici D (2003) Medical robotics in computer-integrated surgery. *IEEE Trans Robot Autom* 19(5):765–781
236. Teber D, Guven S, Simpfendörfer T, Baumhauer M, Güven EO, Yencilek F, Gözen AS, Rassweiler J (2009) Augmented Reality: A New Tool To Improve Surgical Accuracy during Laparoscopic Partial Nephrectomy? Preliminary In Vitro and In Vivo Results. *Eur Urol* 56(2):332–338
237. Tian J, Ma KK (2011) A survey on super-resolution imaging. *SIViP* 5(3):329–342
238. Tjoa MP, Krishnan SM et al (2003) Feature extraction for the analysis of colon status from the endoscopic images. *BioMedical Engineering OnLine* 2(9):1–17
239. Tokgozoglu H, Meisner E, Kazhdan M, Hager G (2012) Color-based hybrid reconstruction for endoscopy. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 8–15
240. Tonet O, Thoranaghatte RU, Megali G, Dario P (2007) Tracking endoscopic instruments without a localizer: a shape-analysis-based approach. *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery* 12(1):35–42
241. Totz J, Fujii K, Mountney P, Yang GZ (2012) Enhanced visualisation for minimally invasive surgery. *Int J Comput Assist Radiol Surg* 7(3):423–432
242. Totz J, Mountney P, Stoyanov D, Yang GZ (2011) Dense Surface Reconstruction for Enhanced Navigation in MIS. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, no. 6891 in LNCS, pp 89–96. Springer
243. Tsevas S, Iakovidis DK, Maroulis D, Pavlakis E (2008) Automatic frame reduction of Wireless Capsule Endoscopy video. In: 8th IEEE International Conference on BioInformatics and BioEngineering, 2008. BIBE 2008, pp 1–6
244. Twinanda AP, Marescaux J, de Mathelin M, Padoy N (2015) Classification approach for automatic laparoscopic video database organization. *Int J Comput Assist Radiol Surg*:1–12
245. Twinanda AP, de Mathelin M, Padoy N (2014) Fisher Kernel Based Task Boundary Retrieval in Laparoscopic Database with Single Video Query. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, no. 8675 in LNCS, pp 409–416. Springer
246. Ukimura O, Gill IS (2008) Imaging-assisted endoscopic surgery: Cleveland Clinic experience. *Journal of Endourology / Endourological Society* 22(4):803–810
247. Vijayan Asari K, Kumar S, Radhakrishnan D (1999) A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation. *IEEE Trans Med Imaging* 18(4):345–354
248. Vilarino F, Spyridonos P, Pujol O, Vitria J, Radeva P, de Iorio F (2006) Automatic Detection of Intestinal Juices in Wireless Capsule Video Endoscopy. In: 18th International Conference on Pattern Recognition, 2006. ICPR 2006, vol 4, pp 719–722
249. Visentini-Scarzanella M, Mylonas GP, Stoyanov D, Yang GZ (2009) i-brush: A gaze-contingent virtual paintbrush for dense 3d reconstruction in robotic assisted surgery. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp 353–360. Springer
250. Vitiello V, Lee SL, Cundy T, Yang GZ (2013) Emerging Robotic Platforms for Minimally Invasive Surgery. *IEEE Rev Biomed Eng* 6:111–126
251. Vogt F, Krüger S, Niemann H, Schick C (2003) A system for real-time endoscopic image enhancement. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2003*:356–363
252. Vogt F, Paulus D, Heigl B, Vogelgsang C, Niemann H, Greiner G, Schick C (2002) Making the Invisible Visible: Highlight Substitution by Color Light Fields. *Conference on Colour in Graphics Imaging, and Vision* 2002(1):352–357
253. Vogt F, Paulus D, Niemann H (2002) Highlight substitution in light fields. In: 2002 International Conference on Image Processing. 2002. Proceedings, vol 1, pp I–637–I–640
254. Voros S, Haber GP, Menuet JF, Long JA, Cinquin P (2010) ViKY Robotic Scope Holder: Initial Clinical Experience and Preliminary Results Using Instrument Tracking. *IEEE/ASME Trans Mechatron* 15(6):879–886
255. Voros S, Hager GD (2008) Towards real-time tool-tissue interaction detection in robotically assisted laparoscopy. In: *BioRob 2008. 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, 2008*, pp 562–567
256. Voros S, Long JA, Cinquin P (2007) Automatic Detection of Instruments in Laparoscopic Images: A First Step Towards High-level Command of Robotic Endoscopic Holders. *Int J Robot Res* 26(11–12):1173–1190

257. Wang P, Krishnan SM, Kugean C, Tjoa MP (2001) Classification of endoscopic images based on texture and neural network. In: Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2001, vol 4, pp 3691–3695
258. Wang X, Zhang O, Han Q, Yang R, Carswell M, Seales B, Sutton E (2010) Endoscopic Video Texture Mapping on Pre-Built 3-D Anatomical Objects Without Camera Tracking. *IEEE Trans Med Imaging* 29(6):1213–1223
259. Wang Y, Tavanapong W, Wong J, Oh J, de Groen P (2010) Detection of Quality Visualization of Appendiceal Orifices Using Local Edge Cross-Section Profile Features and Near Pause Detection. *IEEE Trans Biomed Eng* 57(3):685–695
260. Wang Y, Tavanapong W, Wong J, Oh J, de Groen P (2013) Near Real-Time Retroflexion Detection in Colonoscopy. *IEEE Journal of Biomedical and Health Informatics* 17(1):143–152
261. Wang Y, Tavanapong W, Wong J, Oh JH, de Groen PC (2015) Polyp-Alert: Near real-time feedback during colonoscopy. *Comput Methods Prog Biomed* 120(3):164–179
262. Weede O, Dittrich F, Worn H, Jensen B, Knoll A, Wilhelm D, Kranzfelder M, Schneider A, Feussner H (2012) Workflow analysis and surgical phase recognition in minimally invasive surgery. In: 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp 1080–1074
263. Wei GQ, Arbter K, Hirzinger G (1997) Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation. *IEEE Eng Med Biol Mag*, IEEE 16(1):40–45
264. Wengert C, Reef M, Cattin P, Székely G (2006) Fully Automatic Endoscope Calibration for Intraoperative Use. In: Bildverarbeitung für die Medizin 2006, Informatik aktuell, pp 419–423. Springer
265. Wieringa FP, Bouma H, Eendebak PT, van Basten JPA, Beerlage HP, Smits GA, Bos JE (2014) Improved depth perception with three-dimensional auxiliary display and computer generated three-dimensional panoramic overviews in robot-assisted laparoscopy. *J Med Image* 15001:1
266. Winter C, Rupp S, Elter M, Munzenmayer C, Gerhauser H, Wittenberg T (2006) Automatic Adaptive Enhancement for Images Obtained With Fiberscopic Endoscopes. *IEEE Trans Biomed Eng* 53(10):2035–2046
267. Wolf R, Duchateau J, Cinquin P, Voros S (2011) 3d Tracking of Laparoscopic Instruments Using Statistical and Geometric Modeling. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011, no. 6891 in LNCS, pp 203–210. Springer
268. Wong WK, Yang B, Liu C, Poignet P (2013) A Quasi-Spherical Triangle-Based Approach for Efficient 3-D Soft-Tissue Motion Tracking. *IEEE/ASME Trans Mechatron* 18(5):1472–1484
269. Wu C, Narasimhan SG, Jaramaz B (2010) A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. *Int J Comput Vis* 86(2–3):211–228
270. Wu C, Tai X (2009) Application of Color Change Feature in Gastroscopic Image Retrieval. In: Sixth Int'l Conf on Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09, vol 3, pp 388–392
271. Xia S, Ge D, Mo W, Zhang Z (2005) A Content-Based Retrieval System for Endoscopic Images. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005, pp 1720–1723
272. Xia S, Mo W, Yan Y, Chen X (2003) An endoscopic image retrieval system based on color clustering method. In: Third International Symposium on Multispectral Image Processing and Pattern Recognition, vol 5286, pp 410–413
273. Yamaguchi T, Nakamoto M, Sato Y, Konishi K, Hashizume M, Sugano N, Yoshikawa H, Tamura S (2004) Development of a camera model and calibration procedure for oblique-viewing endoscopes. *Computer aided surgery: official journal of the Int'l Society for Computer Aided Surgery* 9(5):203–214
274. Yao R, Wu Y, Yang W, Lin X, Chen S, Zhang S (2010) Specular Reflection Detection on Gastroscopic Images. In: 2010 4th Int'l Conf. on Bioinformatics and Biomedical Engineering (iCBBE), pp 1–4
275. Yip M, Lowe D, Salcudean S, Rohling R, Nguan C (2012) Tissue Tracking and Registration for Image-Guided Surgery. *IEEE Trans Med Imaging* 31(11):2169–2182
276. Zappella L, Béjar B, Hager G, Vidal R (2013) Surgical gesture classification from video and kinematic data. *Med Image Anal* 17(7):732–745
277. Zhang C, Helferty JP, McLennan G, Higgins WE (2000) Nonlinear distortion correction in endoscopic video images. In: Proceedings. 2000 International Conference on Image Processing, 2000, vol 2, pp 439–442
278. Zhang X, Payandeh S (2002) Application of visual tracking for robot-assisted laparoscopic surgery. *J Robot Syst* 19(7):315–328
279. Zheng MM, Krishnan SM, Tjoa MP (2005) A fusion-based clinical decision support for disease diagnosis from endoscopic images. *Comput Biol Med* 35(3):259–274



Bernd Münzer is currently a postdoc researcher in the applied research project “EndoViP2”. He received his Master’s degrees in 2011 (Information Management, Mag.) and 2012 (Informatics, Dipl.-Ing.), and his Ph.D. degree (Dr. techn.) in September 2015 from Klagenfurt University, all with distinction. His research focus is on content-based analysis of videos from endoscopic surgeries in order to enable a comprehensive video documentation and efficient post-procedural analysis of procedures. In his diploma theses he first surveyed the current situation of video recording and usage in the medical domain of endoscopy and gave a comprehensive market overview of currently available systems. Based on that, he further investigated in detail how content-based video analysis could be incorporated in such systems in order to improve management efficiency and generate an added value for physicians. In his Ph.D. thesis he investigated how the enormous data volume of endoscopic video recordings can be substantially reduced by domain-specific video compression that exploits domain-specific characteristics and also considers perceptual aspects of video quality. In EndoViP 2, he continues his research in this challenging field.



Klaus Schoeffmann is an Associate Professor in the distributed multimedia systems research group at the Institute of Information Technology (ITEC) at Klagenfurt University, Austria. He received his Ph.D. in 2009 and his Habilitation (*venia docendi*) in 2015, both in Computer Science and from Klagenfurt University. His research focuses on human-computer-interaction with multimedia data, multimedia content analysis, and multimedia systems, particularly in the domain of endoscopic video systems. He has co-authored more than 70 publications on various topics in multimedia and he has co-organized several international conferences, special sessions and workshops. He is co-founder of the Video Browser Showdown (VBS), an editorial board member of the Springer International Journal on Multimedia Tools and Applications (MTAP), Springer International Journal on Multimedia Systems, and a temporary steering committee member of the International Conference on MultiMedia Modeling (MMM). Additionally, he is member of the IEEE and the ACM and a regular reviewer for international conferences and journals in the field of multimedia.



Laszlo Böszörményi is a full professor and the head of the Institute of Information Technology at Klagenfurt University. He received his M.Sc. in 1972 and his Ph.D. in 1980 from the Technical University Budapest, Hungary. His main research areas are distributed and parallel systems with emphasis on adaptive distributed multimedia systems, self-organizing multimedia delivery, interactive multimedia, resource scheduling in distributed multimedia systems, and programming languages and compilers. Laszlo Böszörményi is author of several books and publishes regularly in refereed international journals and conference proceedings. He is organizer and program committee member of several international workshops and conferences. He serves as a reviewer at several scientific organizations. He is a senior member of the ACM, member of IEEE, and the OCG.