Towards Intelligent Capsules for Robust Wireless Endoscopic Imaging of the Gut

Dimitris K. Iakovidis$^1$, Roberto Sarmiento$^2$, Juan S. Silva$^3$, Aymeric Histace$^3$, Olivier Romain$^3$, Anastasios Koulaouzidis$^4$, Catherine Dehollain$^5$, Andrea Pinna$^6$, Bertrand Granado$^6$, Xavier Dray$^{3,7}$

$^1$Department of Gastroenterology and Hepatology, University of Paris 06, UMR 7606, LIP6, F-75005, Paris, France
$^2$Department of Gastroenterology and Hepatology, Paris 7 University & APHP Hôpital Lariboisière, Paris, France
$^3$University of Cergy-Pontoise, ETIS UMR CNRS 8051, Cergy-Pontoise, France
$^4$The Royal Infirmary of Edinburgh, Endoscopy Unit, Edinburgh, UK
$^5$Swiss Federal Institute of Technology (EPFL), SCI STI CD, Lausanne, Switzerland
$^6$UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
$^7$Department of Applied Microelectronics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

Abstract—Wireless capsule endoscopy (WCE) enables screening of the gastrointestinal tract by a swallowable imaging system. However, contemporary WCE systems have several limitations, which often result in low diagnostic yield. This paper introduces the concept of a next generation WCE system with embedded intelligence aiming to effectively minimize diagnostic errors. The proposed system is based on a novel wirelessly-powered hardware-software architecture integrating reconfigurable components that are optimized in terms of area-time complexity and power consumption. It integrates multispectral and 3D vision modules, and embedded intelligence for video quality control, accurate localization of the capsule and automatic detection of a broad spectrum of abnormalities. The feasibility of the proposed WCE system is qualitatively assessed with respect to the results obtained, and novel research directions are drawn.

Keywords—wireless capsule endoscopy; multispectral imaging; reconfigurable; intelligent system; medical decision support.

I. INTRODUCTION

For over a decade now since its introduction in clinical practice, wireless capsule endoscopy (WCE) [1] allows—by means of a swallowable, camera-enabled capsule—direct visualization of the gastrointestinal (GI) tract. During its 8-15 h voyage through the digestive tract, the capsule acquires two-dimensional (2-D), color images and wirelessly transmits them to an external receiver. Eventually, these data form lengthy video sequences of 50,000-120,000 frames each that require review by clinicians; this tedious task usually takes an average of 45-90 min (per video) of undivided concentration and attention [2]. Although capsule endoscopes (CEs) are propelled across the entire GI tract, WCE found its niche in the examination of the small-bowel, as this part of the gut is not easily accessible by conventional endoscopy. Other limitations of contemporary CEs include the low quality images, the low capsule/lesion localization accuracy, the stabilization and navigation inability, the inability to perform biopsies and to apply treatment, and more importantly, the overall low positive diagnostic yield, which is estimated to be less than 50% [3]).

In this paper, we introduce the concept of a next generation WCE system that aims to cope with major limitations and drawbacks of contemporary WCE technology, and excel in diagnostic performance and cost-effectiveness. It is based on a novel imaging architecture with embedded intelligence that enables the capsule to be aware of contents of the images it acquires. The capabilities of this system extend well-beyond the capabilities of the state of the art WCE systems by providing: a) significantly higher video quality, both in terms of spatial and temporal resolution, in video segments with higher suspicion of abnormality; b) more accurate capsule localization for precise surgical interventions and disease monitoring; c) diagnostic cues, and output that can be more efficiently visualized so that the diagnostic errors are minimized.

The rest of this paper consists of five sections. Section II provides an overview of the state of the art on WCE systems, section III reviews our algorithms for intelligent WCE video processing and analysis, and section IV discusses the feasibility of fully or partially embedding such algorithms in area-time optimized reconfigurable hardware. The proposed architecture and its integration with embedded intelligence and imaging technologies are presented in Section V. A discussion on feasibility issues and a summary of conclusions is provided in the last section.

II. STATE OF THE ART OVERVIEW

Various conceptual CE designs have been proposed to overcome the aforementioned limitations of the contemporary WCE systems. Such designs include CEs with articulated parts for controlled locomotion, expandable wheels for odometry-based localization, miniature robotic arms for biopsy and/or drug delivery [4, 5, 6]. However, several implementation issues are still unresolved such as the integration of multiple components into a single, energy-efficient capsule to be the most challenging one [7]. Wireless power transmission has been recently proposed as a technology that could reduce the energy limitations of current and future CEs [8, 9, 10].

The low video quality of the current CEs is strongly related to the energy limitations, since higher resolution and/or frame
rates require more energy. Therefore, a main approach to improve video quality relies on image/video compression techniques [11, 12]. In addition, the variation of the frame rate based on the velocity of the CE, has proved to be particularly effective [13]. The latter has already been incorporated in PillCam Colon and lately in the PillCamSB3, offering a frame rate ranging from 4-35 frames/second (fps) and 4-6 fps, respectively, in order to enhance mucosal visualization at times of accelerated peristalsis.

Contemporary CEs address localization issues using a wearable radio-frequency (RF) sensor array that enable tracking through triangulation. Recent approaches indicate that more accurate localization can be achieved with the use of magnetic sensor arrays; however, this approach is still in an experimental phase and it has not been yet adopted in camera-enabled CEs [14]. Previous work addressing the enhancement of the CEs with diagnostic capabilities has mainly focused on bleeding detection. For this purpose, light-tissue interaction has been exploited using transmission spectroscopy and low energy wireless Raman spectrometry [15, 16, 7].

Low cost alternatives to video quality enhancement, capsule localization, and diagnosis support include algorithms that can be implemented in software [17, 18]. Our position is that the tradeoff between software and hardware should be optimized with respect to cost and effectiveness in diagnostic performance. To this end, a review of our state of the art algorithms is provided, and hardware implementation constraints are discussed.

III. INTELLIGENT VIDEO PROCESSING AND ANALYSIS

A. Visual saliency, abnormality detection & recognition

Visual saliency is generally associated with observables that immediately attract our attention within an image. Saliency detection can be implemented in various ways; however, being inspired by the way domain experts perceive saliency has proved to be quite effective. In [19] color is indicated as an important saliency carrier for lesion detection in endoscopy. Based on this fact we have recently proposed a computationally simple algorithm for color-based saliency detection [20, 21]. This algorithm has two steps: a) color transformation of the input image from RGB (red, green, blue) to CIE-Lab (or \( L^*a^*b^* \)) color space [22]; b) automatic selection of salient points in the image by application of the Hessian matrix-based approach of the Speeded-Up Feature Extraction (SURF) algorithm [23] on the chromatic component \( a \) of CIE-Lab color space.

The probability of a salient point to correspond to an abnormality can be significantly increased by incorporating machine intelligence in the process. The detected salient points can be classified based on their color, the color of their neighborhood area [21] and/or texture features [24] with use of a learning machine, such as a support vector machine (SVM) or boosting-based algorithms, trained with a set of previously annotated samples. Considering color features, experiments performed on a dataset of 137 images and 10 different kinds of abnormalities indicate that this general approach for abnormality detection can result in >90% accuracy.

The saliency of a video frame can be characterized by the saliency of its contents, and this process can be easily embedded into high-density reconfigurable hardware, as it is demonstrated in [24]. The semantic interpretation of the salient frames, including the recognition of their abnormal content, can be performed by the domain experts; however, machine learning algorithms can effectively support this task [18]. State of the art approaches of this kind usually address the recognition of only a single abnormality type e.g. polyps, and rarely, up to three types, e.g. ulcers and polyps [25]. For example, using texture features (computed from co-occurrence matrices) in a chained-adaboost configuration focusing on the particular task of polypl recognition, a sensitivity of 95% with a false-positive rate of only 5% was obtained on a dataset of 300 images presenting various types of polyps. The development of a general approach enabling the recognition of a larger, increasing number of abnormality types is still a challenge.

B. Visual capsule localization algorithms

Accurate CE localization is important for the monitoring and assessment of diseases, such as Crohn’s disease and accurate interventions for surgical treatment, e.g., polyp dissection. We have demonstrated the feasibility of CE localization based solely on visual cues in [26]. The method proposed in that study was inspired by visual odometry [27], which is a popular approach for the measurement of distances in robotics and in vehicular technologies research. It is also based on the saliency detection algorithm of the SURF extraction process. The salient points identified in consecutive video frames are tracked, thus enabling the estimation of capsule’s motion. Such a visual odometry approach is not affected by wheel slip in uneven terrain or other adverse conditions, and contrary to wheel odometry [6] it can provide more accurate trajectory estimates, with relative position error ranging from 0.1 to 2% [28].

Challenges in visual capsule localization are mainly associated with real-world applicability issues. These include coping with the motility of the GI tract, and the presence of intestinal content, both of which can interfere with the vision-based measurements. These issues can be addressed by considering methods for recognition of intestinal content [29], and techniques for intestinal motility assessment, either computationally [30] or by exploiting the output of a magnetic localization method [31]. The validation of the accuracy of the visual odometry approach constitutes also another challenge to be coped with realistic phantom models of the intestine that include visible landmarks facilitating gold standard measurement.

IV. EMBEDDABLE INTELLIGENCE

In order to optimize the tradeoff between cost and effectiveness to diagnostic performance, we consider embedding a critical mass of intelligent video processing and analysis algorithms into reconfigurable hardware within the CE.

A. Reconfigurable hardware technologies for WCE

Reconfigurable computing (RC) is a paradigm of using hardware with functionality and structure that can change on demand. The strength of reconfigurability is based on its potential to combine software-like flexibility with the high-performance ability of hardware [32]. The interest of the worldwide research community on RC has grown rapidly during the last 20 years [33]. Nowadays, reconfigurable
systems are highly preferred for the implementation of multimedia applications, which involve variable parameters [34]. The experience from this domain is transferable to WCE for high-throughput video processing and analysis.

Field programmable gate arrays (FPGAs) make it feasible to include reconfigurable hardware on a CE. Different algorithms for WCE video processing and analysis can be mapped on different reconfigurable units inside an FPGA. The algorithms implemented by these units and their interconnections determine the functionality of the FPGA-based CE. The FPGAs are reconfigurable based on a defined configuration process described by a configuration bitstream [35]. This enables the construction of CEs with functionalities that can change during an examination, e.g., by changing the parameters used in an algorithm implementation, or even the algorithm itself, so as to adapt in particular conditions. This adaptability to predefined scenarios is the most important advantage of reconfigurable computing applied to WCE, while requiring lower power than general-purpose microprocessors (GPPs) and graphical processing units (GPUs).

FPGAs are also more flexible and easier to design than application-specific integrated circuits (ASICs), involving a lower non-recurrent engineering cost. The FPGA reconfiguration process could be either static (off-line) or dynamic (on-line). In the former case, the system is stopped in order to load a new configuration, whereas in the latter case parts of the FPGA logic can be substituted on demand, without interrupting operation of the rest of the device. Current research focuses on dynamic and partial reconfiguration (DPR), where a small part of the system can be changed on the fly. The time-overhead introduced for such a change should not affect the diagnostic process in WCE, e.g., it is unacceptable to leave parts of the intestine unexplored. This is feasible by optimizing the size of the configuration bitstream, the power consumption and the memory requirements.

B. Implementation aspects, constraints & expected benefits

In the past and recent literature, the presented approaches for detection and recognition of abnormalities are definitely of primary interest but they may not fully compel to the hardware constraints imposed for the construction of an intelligent CE. These methods were designed mainly for an off-line use by the clinicians and can fully benefit from the high computing capabilities of the last-generation processors. The related processing schemes usually include demanding image analysis algorithms such as feature extraction, active contour segmentation, blob detection or local curvature estimation, that have not been proved yet to be easily embeddable on a low-resource hardware such as an FPGA. In that context, an optimized FPGA implementation in terms of area-time complexity was proposed for implementation of co-occurrence-based image feature extraction [36], and in [37] and [24], two feasibility studies were presented for embedding particular detection tasks. In [37] an SVM classifier using 3D features of the scene computed from an original imaging process (see section V.B) was developed on an FPGA (Virtex 2) platform and showed very satisfying performance. In [24], it is shown that using a simple detection scheme based on texture features in a boosting scenario, a complete VHDL embedding of the process can be achieved. In both cases, the massive parallelization leads to algorithms fully compatible with real-time and hardware constraints.

Currently, a major issue to address is in the expected benefits from such embedding of the abnormality detection process and above all in terms of energy consumption when compared with a software implementation. Having in mind an ASIC-implementation, the fact that an FPGA consumes 12 times more dynamic power than an equivalent ASIC on average [38] and the current power consumption of Virtex 5, we can estimate that the power consumption due to the processing proposed in [24] for instance, will be approximately under the hundred of μW. This feature is less than the power consumption of the usual eight white LED used for illumination and the RF transceiver.

V. PROPOSED ARCHITECTURE

Taking into account the different aspects presented and discussed in previous sections, the proposed WCE system relies on the concept that a WCE system should be able to automatically recognize the contents of the images it acquires. This way, the system should be able to intelligently control its output video quality so that images that are more likely to be abnormal are presented at a higher spatiotemporal resolution. This process is supported by additional imaging components providing essential cues for a robust diagnostic performance.

An outline of the architecture of the proposed WCE system is illustrated in Fig. 1. It is composed of four blocks:

- An instrumentation block with an imager, a laser for 3D active reconstruction and light emission diodes (LEDs) for multispectral processing.
- A processing block that implements the intelligent video processing and analysis algorithms including 3D reconstruction, feature extraction, saliency and abnormality detection, and localization algorithms, described in section III.
- An RF block that transmits data outside the body.
- A magnetic function block with coils for wireless power transmission and localization.

The processing block is implemented in reconfigurable hardware. In this part, processing is made in a co-design fashion, i.e., time consuming processing is parallelized and implemented in hardware, while the rest of the processing is implemented in software. For example, innovative algorithms for fast, automatic saliency and abnormality detection can be implemented in the processing block of the CE so as to obtain a high frame rate (e.g. 25 fps) and resolution (e.g. full high definition). The intelligent processing logic implemented within the capsule should be able to automatically control which frames are more salient (thus likely to contain abnormalities), and transmit them to the external receiver at a higher quality than the less salient ones.
Figure 2 shows the block diagram of a DPR scheme implementing the reconfigurable hardware processing block of the proposed CE architecture. It is composed of a Dynamic Resource Manager (DRM) and a Dynamic Reconfigurable Area (DRA) [39]. DRM architecture works on two levels: a) the scheduler, in charge of the evaluation of the most efficient way to perform a partial reconfiguration requested by the application, and b) the reconfiguration engine (RE), is the responsible for the dynamic deployment of the reconfigurable components on the reconfigurable area. Most of these tasks are device and technology dependent and, for that reason, the main purpose of the RE is to provide a common abstraction, so that the dynamic reconfiguration of the components can be encapsulated as a transparent service to the upper layers.

### A. Multispectral imaging

Multispectral imaging (MSI) is a technique originating from the field of remote sensing, and recently its application has been extended to a variety of fields, including medicine. A multispectral sensor (MS) provides a number (usually in the order of tens) of spectral bands for the same image. If the number of bands is higher than a hundred and the resolution is less than 10 nm, the sensor is called hyperspectral sensor (HS). This spectral information, represented as a data cube, provides the opportunity of differentiating the elements of a scene because the absorption, reflection and scattering of light of each element are related to its molecular structure. Such elements, in the case of WCE, can be different tissues of a mucosa of the GI tract. Considering that the characteristics of tissues change during the progression of a disease [40], the spectral information of MS images carries quantitative diagnostic information about tissue pathology.

Various research groups have investigated the diagnostic value of near-infrared technology in animals and in humans. An excellent review is presented in [41]. An intestinal segment from a pig was clamped and several HS images were captured by two cameras in an in-vivo experiment [42]. The results showed that the normal and ischemic intestine spectra are clearly different in 765-830 nm wavelength regions, without interrupting the surgical procedure. Another hyperspectral imaging (HSI) system was proposed for cancer detection in [43]. This system is based on state-of-the-art liquid crystal tunable filter technology coupled to an endoscope. The diagnostic capabilities of this system are demonstrated by observing mouse carcinomas in-vivo, where the instrument successfully distinguished between benign and malignant mouse skin. In another study HSI was used to distinguish between tumor from normal breast during surgery in induced rat breast cancer [44]. The results demonstrated that the system performance was at least equal to that of histopathological examination of the tumor. It demonstrated a sensitivity of 89% and a specificity of 94% for detection of residual tumors.

Studies on human beings indicate that hyperspectral video endoscopy provides additional diagnostic information about tissues. In [45] it was possible to obtain measurements from nine different tissues. The results proved that hyperspectral video endoscopy is useful for the in-vivo discrimination between healthy and abnormal tissues, although it was not applied to malignant tissues identification. In [46], HSI is used to classify images of colon tissue cells into benign and malignant, with an average accuracy of 84%. In [47] an HSI technique is proposed for studying cancer cells in liquid-based Pap test slides. Cell types were characterized (cervical cells, squamous cell carcinoma, fibroblast, etc.) with varying degrees of dysplasia, which were grown on glass slides. The system was capable of identifying normal cervical cells with a specificity of 95.8%. The sensitivity of low-grade and high-grade precancerous cells was 66.7% and 93.5%, respectively. The sensitivity for squamous cell carcinoma was 98.6%.

The main challenge in the application of MSI into an intelligent miniature CE device is the integration of a spectral camera in such a tiny space. Traditional techniques cannot be applied and it is only recently that IMEC® has designed a prototype HIS device [48] using monolithically integrated Fabry-Perot filters. These filters are directly post-processed at wafer level on top of a standard CMOS sensor. It can acquire multispectral image cubes of 256x256 pixels over 32 bands in the spectral range of 600-1000 nm at up to 340 cubes per second. The number of filters and resolution are adaptable.

### B. 3D imaging

Since primary intestinal lesions, such as polyps, could be considered as protruberating structures, the extraction of 3D features from the scene is of first importance for a reliable in situ detection of that kind of structures. Nevertheless, none of the available WCE on the market provides 3D-imaging capabilities. The proposed intelligent CE will embed such an imaging technique. The main principle of the 3D imaging system, based on the active stereovision principle was fully described in [49]. Briefly, a structured pattern is projected on the scene using a monochromatic light going through a diffractive element. By using simple thresholding scheme on the gray-scale luminance of the images, it is possible to get back the deformed pattern and compare it to the original one to compute quantitative 3D features from the illustrated scene. A large-scale demonstrator has already proved the feasibility of implementing the approach on an FPGA, and the 3D features extracted using the 3D imaging technique described has proved to be efficient in the context of a SVM classifier for a polyp detection task: on a dataset of 181 synthetic polyps, more than 90% were well detected using only 3D features.

### C. Hybrid visual-magnetic capsule localization

The proposed WCE system architecture considers visual CE localization: a) as a low cost alternative to the current localization approaches, and b) as a technique to complement magnetic localization for enhanced robustness. The visual localization algorithm can be implemented as a hardware-software component for online operation (e.g., feature extraction, such as SURF, ORB [50] etc, which is also useful for saliency detection, can be implemented on hardware), and as a software tool to support offline measurements.
Several feasibility studies indicate that magnetic localization can be very accurate for WCE [14]. However, there are still several challenges to be overcome before accurate magnetic localization becomes possible in real WCE examinations [51]. These are mainly associated with the unpredictable, abrupt changes in capsule’s motion during an examination, and with issues related to the magnetic field distribution over a patient’s body.

Hybrid visual-magnetic capsule localization has been proposed in the context of 3D recovery of pathological tissues [52]. In that study, visual localization is considered only for rotation estimation. We have shown that our recent approach to visual CE localization produces a significantly smaller error even for large rotation angles, and that it can also be used accurately for travel distance estimation. The combination of magnetic with visual localization can be based on measurement aggregation, knowledge-based temporal fusion, considering the performance of each methodology as a function of location. Uncertainty-aware approaches to fusion, including fuzzy fusion models can be considered [53, 54].

D. Wireless power transmission by magnetic coupling

Wireless power transmission opens new perspectives in the development of wireless autonomous or semi-autonomous in-vivo medical devices, such as implants and endoscopes [55, 56]. To this end, a major challenge is to achieve high power efficiency between the external base station and the capsule by magnetic coupling as the capsule is at a certain distance from the surface of the body. Moreover, the dimensions of the coil included in the capsule are limited by the sizes of the capsule. Therefore, the attenuation between the external base station and the capsule is high. In [56], for which the capsule is wirelessly powered by the intelligent jacket worn by the patient, it has been shown that the co-design of the coil of the external base station included in the intelligent jacket and of the coil of the capsule (choice of the geometries of coils, quality factor, position, etc) is very important to maximize the power efficiency. Indeed, with respect to the optimization of the remote powering chain, it is highly important to maximize the power efficiency between the DC power delivered at the output of the rectifier (AC to DC converter) included in the capsule and the DC power consumed by the power amplifier of the external base station. Therefore, all this chain has to be optimized to obtain good power efficiency.

VI. DISCUSSION AND CONCLUSIONS

Current limitations of WCE technology require that experience clinicians go through endless hours of (often normal) video sequences – with intense concentration – for lesion detection. The image video quality is far from perfect and there is no accurate localization system [57]. For clinicians who offer WCE on a regular basis, it comes as no surprise that the lack of 3D space appreciation, from a non-controllable device such as the CE, is an additional diagnostic limitation [58]. Therefore, in order for a CE to excel in the field of digestive endoscopy, accurate lesion localization (to allow function of future actuation systems and drug delivery but more importantly planning of medical/surgical treatment in the present state), 3-D ability and improved image quality together with potential for automatic lesion detection will move the technology of wireless endoscopy to an entirely different level.

Our consortium aims to effectively combine various technological bricks from different scientific areas in order to propose a next generation of WCE that will overcome limitations of the state of the art CE devices, with respect to imaging. In this article, with the benefit of the expertise of each member of our consortium, we showed the feasibility of implementing an original architecture for a WCE with embedded intelligence in accordance with clinical and technological constraints. This is the starting point of the next step of our project.

REFERENCES
