AUTOMATIC LESION DETECTION IN WIRELESS CAPSULE ENDOSCOPY - A SIMPLE SOLUTION FOR A COMPLEX PROBLEM

Dimitris K. Iakovidis[1], Anastasios Koulaouzidis[2]

[1]Dept. of Computer Engineering, Technological Educational Institute of Central Greece, Lamia, Greece
[2]The Royal Infirmary of Edinburgh, Endoscopy Unit, Edinburgh, UK

ABSTRACT

Wireless capsule endoscopy (WCE) is performed with a swallowable miniature optical endoscope which transmits color images wirelessly during its journey in the gastrointestinal tract. In this paper we present a computationally efficient and effective approach to cope with automatic detection of possible abnormalities in the WCE videos and consequently with the reduction of the time required for the WCE inspection. It involves automatic detection of salient points based on color information and supervised classification of simple color vectors extracted from the neighborhood of each point. The experiments performed aim to determine the optimal color space components for feature extraction, and identification of abnormalities. Main advantages of this approach are its computational efficiency, its sensitivity to detect small lesions, and its generality. The results obtained from experimentation with a dataset with various types of abnormalities and non-ideal normal frames, approximate 0.9 in terms of the area under receiver operating characteristic (ROC).

Index Terms— Wireless capsule endoscopy, lesion detection, color features, classification.

1. INTRODUCTION

Wireless capsule endoscopy (WCE) is a relatively new imaging technique that poses a variety research challenges to the scientific community of image processing and pattern recognition. Once a swallowable capsule endoscope (CE) enters the human body it begins to wirelessly transmit color images for an average duration of 8 hours. This can eventually result to a large volume of data (currently in the order of 50,000-120,000 depending on the CE capture frame rate) [1]. Considering that the average time required – even by experienced reviewers – for WCE video review and report is approx. 45-90 minutes of intensely focused and undistracted attention [2], it becomes evident that one of the prime challenges is to develop software tools for automated WCE video analysis that will enable faster visual inspection of the WCE videos and, eventually, more accurate diagnosis.

Endoscopic video analysis has attracted the interest of researchers since the early 2000 [3]. In the ensuing decade the technological revolution of WCE [4] led to a consequent increase of scientific contributions focused solely on WCE, which currently has reached its peak [5]. State-of-the-art approaches to the analysis of WCE videos include methods for automatic reduction of video frames to be inspected, e.g. by selection of representative video frames [6], and methods for the recognition of specific types of abnormalities. These are mainly focusing to the detection of bleeding [7, 8], mucosal ulcers [9, 10], Crohn's disease (CD) lesions [11], and polyp(oid) lesions [12], whereas fewer studies have addressed detection of more than a single type of abnormalities e.g. ulcers and polyps [13], or bleeding and ulcers [14].

Most of these approaches are based on color and/or textural features. Bleeding detection has been based on chromaticity moments and uniform local binary patterns (ULBP) in the *HSI* color space [7], as well as on hue histograms [8]; ulcer detection has been based on color rotation and ULBP features extracted from the *RGB* color space [9], as well as on scale invariant feature transform (SIFT) features [10]; the detection of CD lesions has been based on the MPEG-7 dominant color and homogeneous texture descriptors, and edge histograms [11]; and polyp(oid) lesion detection has been based on ULBP features extracted from discrete wavelet transform (DWT) subbands in *RGB* and *HSI* spaces [12], and speeded up robust features (SURF) [15]. Combined ulcer-polyp detection has been addressed using log Gabor filtering on the hue component of the *HSV* color space, co-occurrence matrix features and geometric curvature features; and combined bleeding-ulcer detection has been investigated in the context of selecting subsets of a total of 5859 features extracted from a variety of color spaces including *HSV*, CIE-*Lab*, *YUV*, and *YIQ* [14]. In the majority of the afore-mentioned studies support vector machines (SVMs) and neural networks (NN) have been considered as the classification methods of choice.

Considering that the types of abnormalities that can be found in the gastrointestinal (GI) tract are several and

indeed diverse [16], the interest of gastroenterologists in methods that aim to the detection of specific types of abnormalities is rather limited. Furthermore, most of the current methods are based on statistical features extracted from image regions. A drawback of this approach is that regions should be adequately large for the extracted features to represent statistically significant information; therefore, small lesions, occupying only a few pixels area of the whole frame are likely to be missed.

A method capable of detecting any possible abnormality, regardless of its size and type, would be more useful since it would enable the WCE reviewer to focus on specific video segments, and perform the overall WCE video inspection much faster. From a practical viewpoint, it is more crucial to detect an abnormality than recognize and/or classify it, as the latter task can be performed by the WCE reviewer. Therefore, the method we proposed in [6] is closer to this concept; however, it has a high computational complexity [17], and it returns several video frames that are representative of the normal GI tract. In the context of abnormality detection this is translated into a high false positive (FP) rate.

In this paper we propose a pattern recognition methodology that can efficiently cope with the afore-mentioned issues. This methodology is based on very simple but effective color features that enable the detection even of very small lesions of various types in WCE images. In accordance with previous studies the feature classification task is handled by SVMs.

The rest of this paper consists of four sections. In section 2 we describe the proposed methodology. The experiments performed and the results obtained are presented in section 3, and the conclusions of this study are summarized in the last section.

2. METHODOLOGY

Color is extremely important in the endoscopic diagnosis of digestive tract diseases [18]. During the inspection of a WCE video, differences in the coloring of the mucosa attract the attention of the reviewer to assess more thoroughly the respective regions of interest. During thorough examination of these regions, color is important for the discrimination of the majority of the pathologies from normal tissues or from normal luminal content [16]. The texture of a tissue can be important in the case of lesions with similar colors; however, this mostly applies in magnification or high definition (HD) endoscopy, where image resolution is higher than the resolution of WCE images and allows for the assessment of details [19, 20].

The proposed methodology is conceptually inspired by this process and aims to classify image pixels into two classes; those belonging to normal and those belonging to posssibly abnormal tissue. It involves four steps:

1) Color transformation of the image from *RGB* (red, green, blue) to a color space where chromatic components are approximately decorrelated from luminance components. Such spaces include *HSV* (hue, saturation, value), CIE-*Lab* (or $L^*a^*b^*$), *YCbCr* and several others proposed in the context of different applications [21].

2) Automatic selection of points of interest based solely on color information. This can be achieved by application of the SURF algorithm on an informative chromatic component of the image. SURF points are detected by using a fast approximation of the Hessian Matrix and convolutions of the input image with box filters at multiple scales [22]. This process results in a significant reduction of the pixels to be classified; however, an adequate number of points should be selected so as to minimize the possibility of an abnormality to be missed.

3) Feature extraction from each selected point. Each point is represented by a low-dimensional feature vector formed by:

(i) The values of the color space components of the selected pixels e.g. the values of L, a and b in the CIE-*Lab* space. These are important for pixel-level assessment of the image content.

(ii) The maximum, M_X , and minimum, m_X, values of the color space components X in a $N\times M$ pixel neighborhood around the selected point. Histograms are rather inefficient color descriptors for small neighborhoods because of their high dimensionality (color quantization is not recommended since slight color differences could signify an abnormality). By extensive experimentation using various first-order statistical measures we have found out that simple, min-max features can capture more robustly the significant differences between the normal and the abnormal tissues.

4) The feature vectors are classified by a supervised learning machine such as SVMs or NNs.

3. EXPERIMENTS AND RESULTS

A total of 252 WCE procedures with MiroCam® (IntroMedic® Co., Seoul, South Korea) were performed at the Royal Infirmary of Edinburgh (University Hospital and referral centre for WCE for the southeast of Scotland, UK). The MiroCam® CE is benefited by a white light illumination system of 6 Light Emitting Diodes (LEDs) and has a field and depth of view of 170 degrees and 7-20 mm, respectively. It captures images at a rate of 3 frames per second (fps) with a resolution of 320×320 pixels. The aforementioned WCE video sequence volume is translated to a total of 1,370 images (either for land-marking or capture of pathology). The images were classified by experts of our research group following the Capsule Endoscopy Standardized Terminology (CEST) [23] into the following categories: a) vascular lesions, including angiectasias and/or intraluminal bleeding; b) inflammatory lesions, including mucosal aphthae and ulcers, mucosal erythema, mucosal cobblestone, and luminal stenosis; c)

lymphangiectasias, including chylous cysts, nodular lymphangiectasias, punctuate lymphangiectasias, and d) polypoid lesions.

In order to efficiently investigate the performance of the proposed methodology, our experimental study involved a random subset of 77 from the available images with abnormalities. These images were graphically annotated in pixel-level detail by the experts using the Ratsnake annotation tool [24]. More specifically this subset consisted 5 aphthae, 5 frames with intraluminal bleeding, 9 nodular lymphangiectasias, 8 chylous cysts (a subgroup of lymphangiectatic small-bowel lesions), 27 angiectasias with different bleeding potentials, 6 polyps, 6 stenoses, 9 ulcers, and 2 villous oedemas. In order to account for more realistic conditions we included a set of 60 images without visible abnormalities in the dataset, half of which including bubbles and/or luminal debris or opaque luminal fluid.

The medical problem posed is the automatic detection of even the smallest visible abnormalities within the 137 (i.e. 77+60) images of our dataset. To solve this problem we need to consider the classification of each of the $1.4 \cdot 10^7$ (i.e. $137 \cdot 320 \cdot 320$) pixels into two classes; normal and abnormal, depending on their belongingness to the respective categories of tissues. In order to reduce the computational complexity of this problem we investigated the application of the SURF point detection method on various color space components so as to automatically select a significantly smaller subset of salient pixels that are more likely to belong to abnormalities. Subsequently, the proposed computationally simple feature extraction approach is evaluated and compared with state-of-the-art feature extraction approaches in terms of classification accuracy, sensitivity (SN) and specificity (SP). The classification accuracy is determined by the area under the receiver operating characteristic (ROC) curve (AUC), as it represents an intuitive performance measure even if the datasets have imbalanced class distributions [25].

3.1. Automatic point selection

We applied the SURF algorithm for salient point selection on the components of various color spaces that have been previously considered in WCE video analysis; these include *RGB*, *HSV*, CIE-*Lab* (using standard illuminant D65), and *YCbCr* spaces [21]. We have focused on the chromatic components due to the importance of color in the detection of abnormalities by the endoscopists. The lightness *L* component of CIE-*Lab* was considered as a representative of the greylevel image intensity since it is highly correlated with both *V* and *Y* components of the *HSV* and *YCbCr* color spaces respectively.

The results are illustrated in Fig. 1. Figure 1(a) shows the number of selected points with respect to the SURF threshold, and Fig. 1(b) shows the percentage of detected abnormalities by these points. An abnormality is considered

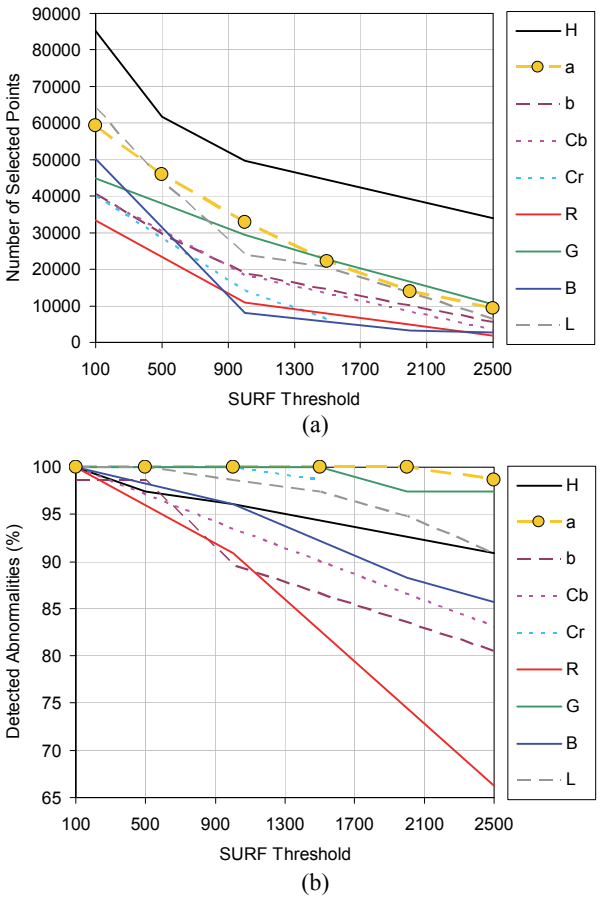

Figure 1. Point selection results. (a) Number of selected points (pixels) of interest over all dataset using SURF algorithm on various color space components. (b) Percentage of detected abnormalities by the SURF points.

as detected if the respective image region includes at least one of the points selected by the SURF algorithm. It can be observed that the use of the chromatic component *a* of CIE-*Lab* color space results in 100% detection of the abnormalities while it results the smaller number of points (only 13835 pixels for a threshold value of 2000, which is $9.8 \cdot 10^{-2}$% of the pixels of the whole dataset).

Representative examples of automatically selected points in images with various types of abnormalities are illustrated in Fig. 2. The selected points are indicated by cross marks and their subset belonging to abnormal regions (according to the ground truth annotations provided by the experts) are indicated by circular marks. It can be noticed that using the *a* component the SURF algorithm manages to automatically select points on all abnormalities, even if these are very small lesions e.g. Fig. 2(a), (b), (d).

3.2. Feature extraction and classification

Motivated by the results described in the previous paragraph we investigated the performance of various feature extraction

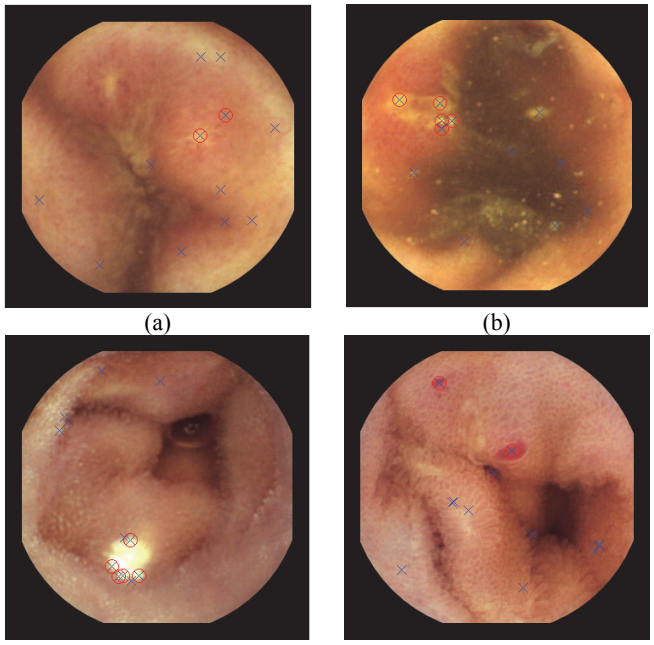

Figure 2. Representative abnormalities detected by the application of the SURF algorithm on the 'a' component of the CIE-Lab color space. Cross marks indicate the detected points and circular marks indicate those belonging to abnormal regions. (a) Aphthae. (b) Ulcer. (c) Nodular lymphangiectasia. (d) Angiectasia.

schemes in the CIE-*Lab* color space. The feature extraction schemes considered are based on the pixel values of the selected points, represented by lightness L and chromaticities a and b.

We considered four different feature sets: $S_1=\{L, a, b, M_L, m_L, M_a, m_a, M_b, m_b\}$, $S_2=\{L, a, M_L, m_L, M_a, m_a\}$, $S_3=\{M_L, m_L, M_a, m_a, M_b, m_b\}$, and $S_4=\{a, M_a, m_a\}$, where M_X and m_X, $X = L, a, b$, represent the maximum and minimum pixel values over the neighborhood of each point. Preliminary classification experiments determined the size of this neighborhood at 36×36 pixels. For comparison purposes we also provide baseline results obtained using plain color component vectors (L, a, b) from the selected points, and results from the application of well-known texture descriptors of the neighborhoods of these points. The texture descriptors include 64-dimensional SURF features estimated over the lightness L (SURF_L) and the image chromatic components a (SURF_a), and the ULBP descriptors [9].

Initially we performed comparisons using an SVM classifier with radial basis function (RBF) kernel, which is widely used and well-known for its learning capacity [26]. Repetitive classification experiments were carried out using the 10-fold cross validation strategy for the selection of the training and the testing sets. The results obtained are summarized in Table 1. It can be observed that best performance is obtained with the use of feature set S_1. The performance of this feature set was tested also with a linear SVM which is less complex than RBF SVM; however, the results obtained indicate that the classification problem is highly non-linear.

It can be noticed that the standard SURF descriptors over the lightness component of CIE-*Lab* (SURF_L) results in very low classification performance (AUC 0.50±0.01), although they have been proved adequate for WCE image registration tasks [27, 28]. Combinations of SURF_a and SURF_L with color component values of the selected points resulted also in such a low performance (not included in Table 1 due to space limitations). The results obtained with the ULBP approach are higher than those obtained with the SURF features but still lower than those obtained with the proposed approach.

The time required for the extraction of the proposed features per image, with not optimized Matlab code, on an Intel i5 laptop, was of the order of 10 ms.

Table 1. Classification results.

Features	Classifier	AUC	SN	SP
S_1	**RBF-SVM**	**0.89±0.01**	**0.95±0.01**	**0.83±0.02**
	Linear-SVM	0.56±0.02	0.40±0.11	0.72±0.11
S_2	RBF-SVM	0.81±0.01	0.87±0.01	0.92±0.02
S_3	RBF-SVM	0.73±0.01	0.75±0.01	0.71±0.03
S_4	RBF-SVM	0.69±0.01	0.78±0.04	0.60±0.05
L, a, b	RBF-SVM	0.63±0.02	0.59±0.08	0.66±0.11
SURF_a	RBF-SVM	0.69±0.02	0.64±0.02	0.75±0.04
SURF_L	RBF-SVM	0.50±0.01	0.67±0.15	0.30±0.18
ULBP	RBF-SVM	0.68±0.02	0.71±0.01	0.66±0.02

4. CONCLUSIONS

We presented a computationally simple but effective approach to feature extraction for automatic detection of various abnormalities in WCE videos. The proposed approach was experimentally evaluated on a realistic dataset with 10 types of abnormalities and non-ideal normal frames. The results of the experiments validate that it is:

- accurate in the detection of lesions, outperforming state of the art approaches;
- robust in the presence of luminal content;
- capable of detecting small lesions.

Future work includes experimentation with even more types of abnormalities and full length WCE videos.

5. ACKNOWLEDGMENTS

This work was partially funded by the Technological Educational Institute of Central Greece, Lamia, Greece.

6. REFERENCES

[1] A. Koulaouzidis, E. Rondonotti, and A. Karargyris, "Small-bowel capsule endoscopy: a ten-point contemporary review," *World journal of gastroenterology: WJG*, vol. 19, no. 24, p. 3726, 2013.

[2] S. K. Lo, "How should we do capsule reading?" *Techniques in Gastrointestinal Endoscopy*, vol. 8, no. 4, pp. 146–148, 2006.

[3] S. A. Karkanis, D. K. Iakovidis, D. Karras, and D. Maroulis, "Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2. IEEE, 2001, pp. 833–836.

[4] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, p. 417, 2000.

[5] M. Liedlgruber and A. Uhl, "Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review," *Biomedical Engineering, IEEE Reviews in*, vol. 4, pp. 73–88, 2011.

[6] D. K. Iakovidis, S. Tsevas, and A. Polydorou, "Reduction of capsule endoscopy reading times by unsupervised image mining," *Computerized Medical Imaging and Graphics*, vol. 34, no. 6, pp. 471–478, 2010.

[7] B. Li and M.-H. Meng, "Computer-aided detection of bleeding regions for capsule endoscopy images," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 4, pp. 1032–1039, 2009.

[8] G. Lv, G. Yan, and Z. Wang, "Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 6643–6646.

[9] V. S. Charisis, C. Katsimerou, L. J. Hadjileontiadis, C. N. Liatsos, and G. D. Sergiadis, "Computer-aided capsule endoscopy images evaluation based on color rotation and texture features: An educational tool to physicians," in *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*. IEEE, 2013, pp. 203–208.

[10] L. Yu, P. C. Yuen, and J. Lai, "Ulcer detection in wireless capsule endoscopy images," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 45–48.

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

[12] B. Li and M. Q.-H. Meng, "Automatic polyp detection for wireless capsule endoscopy images," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10952–10958, 2012.

[13] A. Karargyris and N. Bourbakis, "Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 10, pp. 2777–2786, 2011.

[14] P. Szczypinski, A. Klepaczko, M. Pazurek, and P. Daniel, "Texture and color based image segmentation and pathology detection in capsule endoscopy videos," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 396–411, 2014.

[15] S. Hwang, "Bag-of-visual-words approach to abnormal image detection in wireless capsule endoscopy videos," in *Advances in Visual Computing*. Springer, 2011, pp. 320–327.

[16] M. B. Wallace and A. Buchner, "Atlas of gastrointestinal endomicroscopy," *Gastroenterology*, vol. 144, no. 4, pp. 854–854, 2013.

[17] I. Kanaris, S. Tsevas, I. Maglogiannis, and D. K. Iakovidis, "Enabling distributed summarization of wireless capsule endoscopy video," in *Imaging Systems and Techniques (IST), 2010 IEEE International Conference on*. IEEE, 2010, pp. 17–21.

[18] M. Tanaka, Y. Kidoh, M. Kamei, T. Terasaki, A. Watanabe, T. Sakamoto, and M. Fujimaki, "A new instrument for measurement of gastrointestinal mucosal color," *Digestive Endoscopy*, vol. 8, no. 2, pp. 139–146, 1996.

[19] J. Tischendorf, R. Schirin-Sokhan, K. Streetz, N. Gassler, H. Hecker, M. Meyer, F. Tacke, H. Wasmuth, C. Trautwein, and R. Winograd, "Value of magnifying endoscopy in classifying colorectal polyps based on vascular pattern," *Endoscopy*, vol. 42, no. 01, pp. 22–27, 2010.

[20] J. Mannath and K. Ragunath, "Endoscopy: High-definition imaging and nbi—improving colonic imaging?" *Nature Reviews Gastroenterology and Hepatology*, vol. 8, no. 11, pp. 604–605, 2011.

[21] G. Wyszecki and W. S. Stiles, *Color science*. Wiley New York, 1982, vol. 8.

[22] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[23] L. Korman, M. Delvaux, G. Gay, F. Hagenmuller, M. Keuchel, S. Friedman, M. Weinstein, M. Shetzline, D. Cave, and R. de Franchis, "Capsule endoscopy structured terminology (cest): proposal of a standardized and structured terminology for reporting capsule endoscopy procedures," *Endoscopy*, vol. 37, no. 10, pp. 951–959, 2005.

[24] D. Iakovidis, T. Goudas, C. Smailis, and I. Maglogiannis, "Ratsnake: A versatile image annotation tool with application to computer-aided diagnosis," *The Scientific World Journal*, vol. 2014, 2014.

[25] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[26] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[27] E. Spyrou and D. K. Iakovidis, "Video-based measurements for wireless capsule endoscope tracking," *Measurement Science and Technology*, vol. 25, no. 1, p. 015002, 2014.

[28] D. K. Iakovidis, E. Spyrou, and D. Diamantis, "Efficient homography-based video visualization for wireless capsule endoscopy," in *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1–4.