Detection of wrinkle frames in endoluminal videos using betweenness centrality measures for images.

Santi Seguí*†, Michal Drozdzal*†, Ekaterina Zaytseva*†, Carolina Malagelada‡, Fernando Azpiroz‡, Petia Radeva*†, and Jordi Vitrià*†
*Computer Vision Center (CVC), Barcelona, Spain
†Dept. Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Barcelona, Spain
‡Digestive System Research Unit, Hospital Vall d'Hebron, Barcelona, Spain

***Abstract*—Intestinal contractions are one of the most important events to diagnose motility pathologies of the small intestine. When visualized by wireless capsule endoscopy (WCE), the sequence of frames that represents a contraction is characterized by a clear wrinkle structure in the central frames that corresponds to the folding of the intestinal wall. In this paper we present a new method to robustly detect wrinkle frames in full WCE videos by using a new mid-level image descriptor that is based on a centrality measure proposed for graphs. We present an extended validation, carried out in a very large database, that shows that the proposed method achieves state of the art performance for this task.**

***Index Terms*—Wireless Capsule Endoscopy, Small Bowel Motility Dysfunction, Contraction Detection, Structured Prediction, Betweenness Centrality.**

I. Introduction

The general function of the small intestine is the absorption of the food we eat. This absorption is performed with the help of muscle contractions that move the food content back and forth and mix it with the digestive enzymes that are secreted into the intestine. Contractions also move the contents of the intestine slowly towards the large intestine. The overall process of absorbing food and liquids and moving waste through the gut is called motility.

The motility process is the result of the integrated activity of nerves, muscles and hormones. Abnormalities in any of these components or in their integration can result in different motility dysfunctions [1], [2], [3].

The presence or absence of diverse physiological symptoms constitutes the first evidence for the diagnosis of a pathology of the small bowel. Nowadays, the main source of information, and the only one which leads to a conclusive diagnosis of intestinal motility disorders is the one obtained from the result of a motility test performed by using manometric devices [4]. Small bowel manometry uses a catheter that has pressure sensors to record the contractions of the gastrointestinal tract. The catheter is introduced via the mouth and positioned in such a way that different sensors are located in different regions of the stomach and the small intestine. Testing time is long because contractions in the small intestine do not occur frequently and measurements must be made over a long period in an effort to catch the contractions. The application of this technique presents several drawbacks: 1) It is restricted to few referral centers because it requires considerable technical expertise and knowledge for interpretation of the data, 2) it involves intestinal intubation with consequent discomfort for the patients and 3) it is limited to the analysis of pressure values, lacking of information about different content, structure, morphology and dynamics of the intestine.

Wireless Capsule Endoscopy (WCE) is a technology that allows to look at the intestine from inside with minimum discomfort for the patient [5]. One of the main drawbacks associated with the WCE videos is the long time needed by the physicians for the proper video analysis. In the literature this problem has been overcome by building Computer-Aided Decision Support Systems (CADSS) [6]. Generally, CADS have been used either for efficient video visualization, e. g. [7], [8], [9] or to detect different intestinal abnormalities such as bleeding [10], Crohn's disease [11], polyp [12], tumor [13] and ulcer detection [14].

Recently, this technology has been also proposed for evaluation of the small bowel motor function based on a fully computerized image analysis program [15], [16]. The proposed method is based on the detection of different visual events: contractile patterns (phasic luminal closure and radial wrinkles by wall texture analysis), non-contractile patterns (tunnel and wall appearance), intestinal content presence, and endoluminal motion. One of the main conclusions was that patients exhibited a significant deviation of the contractile activity level, higher or lower, with respect to the level of healthy subjects.

In WCE, intestinal contractions are visualized as a sequence representing, first, the closing of the lumen from the resting position, and then, the opening of the lumen to the resting position again. The main visual features to characterize these events are the changing lumen area (see Fig. 1) and the presence of characteristic wrinkles in central frames of the sequence.

Wrinkles are an omnipresent characteristic of contractions and have been mainly studied as a pattern to characterize a subset of intestinal contractions [17], [18]. This pattern is visually observed as a set of folds of intestinal wall, in star-like shape (see Fig. 2). Usually, the wrinkle pattern is observed in the central frames of intestinal contractions where strong pressure is produced by the nerve system. In [17], Vilariño et al. proposed a method for the detection of tonic contractions [17] based on wrinkle information. The proposed method categorizes wrinkle frames by using general linear radial patterns based on the valleys and ridges of the image. An alternative

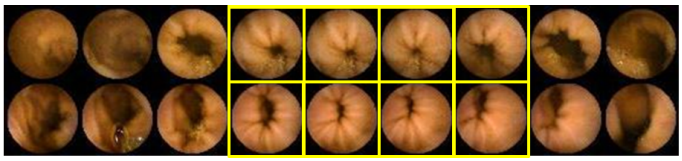


Fig. 1: Samples of intestinal contractions. Each row represents a set of frames depicting a different contraction. The presence of wrinkle patterns can be clearly observed at central frames (marked in yellow).

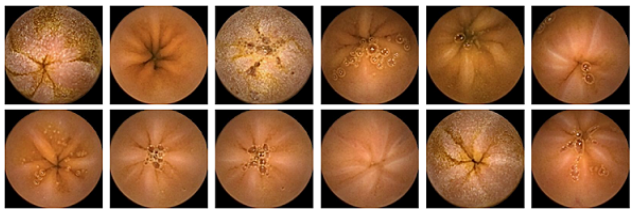
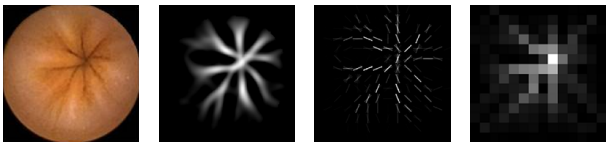


Fig. 2: Images corresponding to central frames of contractions clearly show a star-like pattern produced by the strong pressure of nerves when closing the intestinal wall.

method was proposed by Spyridonos et. al. in [18]. In this paper, a new image descriptor for categorization of wrinkle images was presented using directional information from the structure tensor matrix. In both works, the characterization of wrinkle frames was achieved by dividing the images into 4 different quadrants and computing the corresponding set of features for each quadrant. The weakest point of these approaches was the need to detect the lumen center in order to define the quadrants, since in wrinkle images the lumen is very small or even appears completely closed because of the wall closure. In [3] Vu et al. proposed a method to assess intestinal motility based on the characteristics of contractile patterns and information on their frequencies. Contractions are recognized by changes in the edge structure of the intestinal folds (wrinkles) and evaluating similarity features in consecutive frames. Additionally, the size of contraction windows is adjusted according to the passage of the capsule in order to minimize the number of false positives.

In this work, we address the specific problem of detecting wrinkle frames using a new mid-level image descriptor that measures the continuity of certain image features. First, the image is locally described by a second-order differential quantity of the image, the Hessian, a matrix derived from the second derivatives of the image [19]. This image descriptor is used because it summarizes the predominant directions of the image curvature in the neighborhood of a pixel. Second, we compute a histogram of oriented features from the Hessian field building a data structure similar to the well known HoG descriptor [20]. Then, the resulting image descriptor is transformed into a graph structure that is analyzed to produce mid-level information related to the continuity of the image curvature fields. More specifically, we show that the application of a centrality measure [21] on graphs that represent local curvature orientation distributions can produce a mid-level image representation that is very useful to represent



(a) (b) (c) (d)

Fig. 3: An example representing the important steps of the proposed methodology: a) Original image; b) $\max(0, \lambda_1)$; c) Histogram of oriented features computed from $(\lambda_1, \mathbf{e_2})$; d) Centrality descriptor calculated by transforming the histogram of oriented features into a graph (darker cells indicate low centrality values and lighter cells indicate high centrality values).

wrinkle frames and radial image structures in general. Finally, we train an image window classifier using a structured training paradigm [22]. We show that by posing our problem as a structured output prediction problem, we can significantly increase the detection performance of our method.

This paper is a substantial revision of the work presented in [23]. Improvements over [23] include a critical review of different centrality measures, extended validation experiments, and a better classification strategy based on Structured Support Vector Machines.

The organization of the paper is as follows: Section 2 presents the proposed image descriptor. Section 3 presents a qualitative and quantitative discussion of our results. Finally, we discuss our contribution and draw some conclusions in Section 4.

II. Method

Given a WCE frame, the proposed method to detect wrinkle structures is divided into the following four steps:

1) In the first step we compute, for each image pixel, a matrix that represents the predominant curvature directions in the neighborhood of the pixel by using the Hessian matrix.
2) In the second step, the eigen-decomposition is applied to the Hessian matrix to compute its eigenvalues (λ_1, λ_2) and the corresponding eigenvectors, $(\mathbf{e}_1, \mathbf{e}_2)$. Let λ_1 be the eigenvalue with the highest absolute value (see Fig. 3.b). We construct a set of local histograms describing the orientation distribution of the $\mathbf{e_2}$ eigenvector in a similar way as it is done with well known Histogram of Gradients (HoG) [20] (see Fig. 3.c).
3) Then, a mid-level image descriptor is obtained by transforming the set of local histograms into a graph and computing the centrality measure of each node (see Fig. 3.d).
4) Finally, a Structured Output Support Vector Machine classifier trained with mid-level features is applied, by following a sliding window approach, to detect the presence of wrinkle structures in the image.

In the following subsections, we give details of each step of our approach.

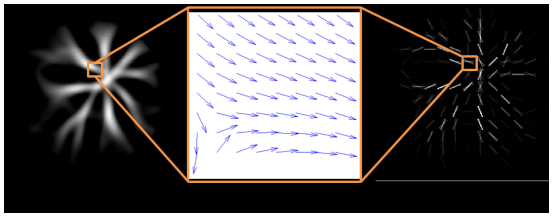


Fig. 4: From image to histogram: a) $\max(0, \lambda_1)$; b) Orientations of the second eigenvector of one selected cell; c) Histogram of oriented features computed from $(\lambda_1, \mathbf{e_2})$;.

A. Feature Extraction: Hessian Matrix

In order to detect the intestinal wrinkles, we need a descriptor that is able to discriminate the shape of these specific image features.

The Hessian Matrix (HM) is a matrix derived from the second derivatives of the image that summarizes the predominant directions of the local curvatures and their magnitudes in a neighborhood of a point. The Hessian Matrix, HM_σ of an image I is a symmetric 2×2 matrix of functions. Each entry is given by:

$$HM_\sigma(p) = \begin{bmatrix} G(\sigma) * I_{xx}(p) & G(\sigma) * I_{xy}(p) \\ G(\sigma) * I_{xy}(p) & G(\sigma) * I_{yy}(p) \end{bmatrix} \quad (1)$$

where I_{xx}, I_{xy}, I_{yy} are the second order partial derivatives of the image I with respect to x and y coordinates, $p = (x, y)$ is an image point, $*$ is the convolution operator and $G(\sigma)$ is the Gaussian function with scale parameter σ. Let λ_1 be the largest eigenvalue by absolute value, $|\lambda_1| > |\lambda_2|$. $|\lambda_1|$ shows the strength of the local image curvature and its corresponding eigenvector, $\mathbf{e}_1$, is aligned with the dominant curvature direction of the image within a window defined by σ. The second eigenvector is orthogonal to the dominant curvature direction, generally pointing towards the direction of the least curvature.

Wrinkle structures can be associated to image valleys [24]. For this reason, λ_1 represents at every image pixel a *wrinkleness* measure that can be used to detect foldings of the intestinal wall. In order to select these points we consider for every pixel the map represented by $\max(0, \lambda_1)$. An example that illustrates this procedure is presented in Fig. (4). In Fig. (4.a), it is shown that the considered map perfectly defines tubular image structures at scale σ and can be used as an indicator of wrinkle presence. In fact, we have observed that pixels corresponding to low curvature regions do not carry any interesting information for wrinkle detection. For this reason, we apply an adaptive threshold, fixed by cross-validation, that selects the 30% of image pixels with the highest values of $\max(0, \lambda_1)$. In Fig. (4.b) it is shown that the second eigenvector $\mathbf{e}_2$ is aligned with wrinkle direction, and consequently it points towards the closed lumen.

B. Feature Representation: Histogram of Features

The computation of the Hessian-based descriptor on an image produces a two-dimensional vector for every pixel (see Fig. 4.b). In order to reduce the dimensionality of this representation and also to increase its invariance to scale and position variations, we decompose the image into a set of M small squared cells and compute a histogram over orientation bins. The angle of $\mathbf{e_2}$ is used to vote on the corresponding orientation bin with the vote value represented by λ_1. Votes are accumulated over all the pixels within each cell. Following the classical HoG, an image descriptor is then built by concatenating the values of the bins of all histograms, getting a high-dimensional vector $H = (\mathbf{h}_1, \ldots, \mathbf{h}_M)$ that represents the image, where $\mathbf{h}$ is a cell histogram. This image representation is shown in Fig. (4.c).

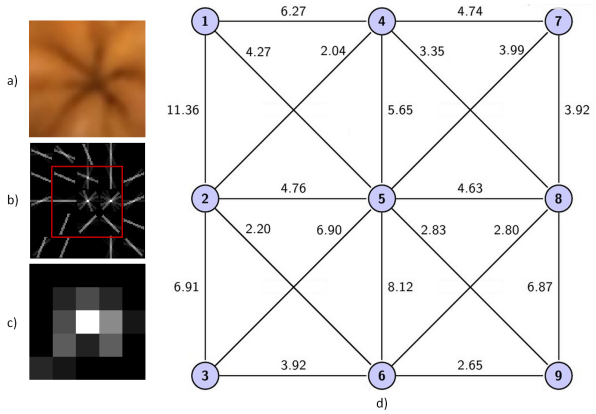


Fig. 5: An example of centrality measure calculation for an image (a). A histogram of oriented features (b) can be transformed into a graph (d) by defining a node for each cell and a set of edges connecting all neighbor cells. Each edge can be labeled with a weight value representing the connectivity degree of the pair of nodes (i, j) it links, which can be computed from the histograms of their corresponding cells, $\mathbf{h_i}$ and $\mathbf{h_j}$. Then, a centrality measure (c) can be computed on this weighted graph structure.

C. Mid-level descriptor

If we consider $(\lambda_1, \mathbf{e_2})$ to be a low-level image descriptor, a mid-level descriptor would be a continuous or discrete numeric measurement obtained after a global analysis of the interactions between the values of $(\lambda_1, \mathbf{e_2})$ in the whole image. In our case, we are interested in a discriminant descriptor to characterize the prototypical star-like pattern that represents wrinkle frames. To this end, we define a new mid-level image descriptor, that we call *image centrality*, which is based on the betweenness centrality measure that was originally proposed to analyze social networks [25]. If we consider the graph where each image cell corresponds to a node and links are only defined for pairs of neighboring cells, this descriptor will define for each image cell a robust measure of its centrality in terms of its probability to occur on a randomly chosen shortest path between two randomly chosen cells. The following subsections describe how the histogram is transformed into a graph and how centrality of each graph node can be computed.

1) From histograms to graph: The graph can be formally defined in the following way. Let M be the number of image cells resulting from dividing the image I into cells of $n \times n$

pixels. Let $H = (\mathbf{h}_1, \ldots, \mathbf{h}_M)$ be the histogram of oriented features computed from $(\lambda_1, \mathbf{e_2})$. Let $\mathbf{h}_i(\alpha)$ be the value of the orientation bin of cell i corresponding to the votes of $\mathbf{e}_2$ oriented at α degrees. Then, we can built a weighted graph $\mathcal{G} = (V, E, \mathbf{A})$ by considering a set of M nodes V, where each node v_i corresponds to cell i of I, a set of edges E, where edge $e_{i,k}$ connects two neighboring cells i, k with corresponding histograms $\mathbf{h}_i$ and $\mathbf{h}_k$, and a matrix $\mathbf{A}$ of weights. We have considered the set of 8-connected cells of the image.

Let i and k be two neighboring image cells. Let β be the orientation of the vector extending from the center position of i to the center position of k. Then, the cost of the edge $e_{i,k}$ is assigned by taking into account the relationship between the value of β and the values of $\mathbf{h}_i$ and $\mathbf{h}_k$. When considering 8-connected cells, the values of β (in degrees) can be $(0, 45, 90, 135, 180, 225, 270, 315)$. Taking into account that we have considered 8 orientation bins for each orientation histogram, corresponding to $0, 22.5, 45, 67.5, 90, 112.5, 135$ and 157.5 degrees, the cost of $e_{i,k}$ can be defined as:

$$\begin{aligned} e_{i,k} &= \frac{1}{4}(\mathbf{h}_i(\beta) + \mathbf{h}_k(\beta)) \\ &+\frac{1}{8}(\mathbf{h}_i(\beta - 22.5) + \mathbf{h}_k(\beta + 22.5)) \\ &+\frac{1}{8}(\mathbf{h}_i(\beta + 22.5) + \mathbf{h}_k(\beta - 22.5)) \end{aligned} \quad (2)$$

All angle operations are defined $(\mathrm{mod}\ 180)$.

This is a symmetric measure ($e_{i,k} = e_{k,i}$) that assigns high values to neighboring cells which present curvature fields (represented by their histograms) that are similar to β and low values (or even zero) to neighboring cells which present curvature fields with orientations different from β.

We show an example of the graph corresponding to the 3×3 cells of the central part of a wrinkle image in Fig. 5.

2) Centrality descriptor: Given an image I and its corresponding weighted undirected graph $\mathcal{G}$, the main assumption of our method is that the image cell that corresponds to the closed lumen is an *important* node of $\mathcal{G}$ when considering the shortest paths between all pairs of nodes in $\mathcal{G}$. Recall that, from our graph definition, the shortest paths will lie on regions of the image with a high curvature and will be aligned with the direction of least curvature. Then, due to the wrinkle structure, *the node corresponding to the closed lumen position will have a high probability to occur on the shortest path between two randomly chosen nodes of the image*.

This assumption advises for an image descriptor based on the concept of shortest paths in a graph, and more specifically for a measure of the importance of a node based on the number of shortest paths it belongs to. In the literature, we can find several proposals regarding the measurement of the importance of a node based on this idea, all representing different approaches to the concept of *node centrality*.

Closeness centrality [26] is the inverse of the average shortest-path distance from the vertex to any other vertex in the graph. It can be viewed as the efficiency of each individual vertex in spreading information to all other vertices. Graph centrality was introduced implicitly in [27] to identify the *center* of a network by using only the maximum value of the shortest-path distances. Stress centrality was introduced in [28] to measure how much *work* is done by each vertex in a communication network. It assumes that the set of paths used for communication as the set of shortest paths. Finally, betweenness centrality [25] is the most important one and it constitutes a fundamental measurement concept that was originally proposed for the analysis of social networks.

More formally the four centrality measures can be defined as follows:

1) Closeness centrality: $C_1(v) = \frac{1}{\sum_{t\in V} d_{\mathcal{G}}(v,t)}$.
2) Graph centrality: $C_2(v) = \frac{1}{\max_{t\in V} d_{\mathcal{G}}(v,t)}$.
3) Stress centrality: $C_3(v) = \sum_{s\neq v\neq t\in V} \sigma_{st}(v)$.
4) Betweenness centrality: $C_4(v) = \sum_{s\neq v\neq t\in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$.

The parameter, σ_{st} is the number of shortest paths from node s to node t, $d_{\mathcal{G}}(s,t)$ is the distance between nodes s and t (i.e. the length of the shortest path from node s to node t) and $\sigma_{st}(v)$ denote the number of shortest paths from s to t that some $v \in V$ lies on.

The resulting n^2-dimensional vector, C, stores the centrality measure of all graph vertices. This vector can be seen as a mid-level descriptor that represents the importance of a region of the image with regard to the shortest paths that run by following the field defined by $\mathbf{e}_2$. From this point of view, it contains global information that cannot be captured by any means by using local operators. In our application, this vector represents fairly well the wrinkle structures and its maximum value component is located in the position of the closed lumen. Fig. 6 shows different centrality measures. It can be seen that the betweenness centrality is the measure that best aligns with the wrinkle structures and for this reason, it is our choice for representation of mid-level visual information.

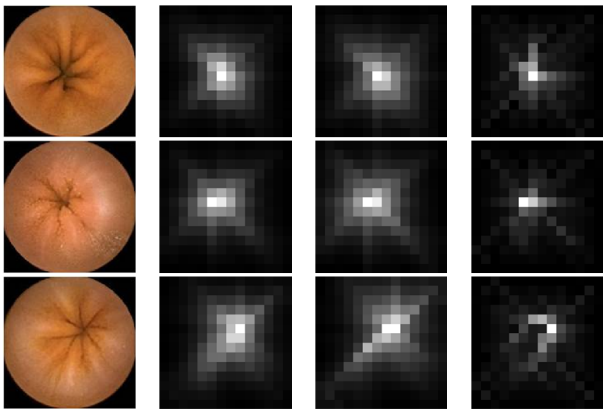
A naive implementation of betweenness centrality would result in an algorithm complexity of $\Theta(|V|^3)$, where $|V|$ is the number of nodes of $\mathcal{G}$, which would make the computation of this measurement for large graphs prohibitive. An algorithm for the calculation of the betweenness centrality that runs in $\mathcal{O}(|V|\, m)$, where m is the number of edges, was proposed in [29]. This algorithm allows a very fast computation of the betweenness centrality measure of all image cells.

D. Automatic detection of wrinkle frames

The last step of the method deals with the detection of a wrinkle pattern in a WCE video. To this end, we propose to learn a linear classifier from a set of positive and negative examples and, then, to apply this classifier to image frames by using a sliding window that scans the image cells looking for the presence of a wrinkle pattern.

Structured Output Support Vector Machines. Support Vector Machines (SVM) [30] are widely used to solve linear classifier problems in binary data, but in their classical formulation they are not easily applicable to multiclass problems. Structured Output Support Vector Machines (SO-SVM) ([22], [31]) is a recently proposed extension to SVM that is able to deal even with problems with infinite number of classes.

Let $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{X}, \mathcal{Y}, i = 1, \ldots, N$, be a set of N training instances from a sample space $\mathcal{X}$ and a label space $\mathcal{Y}$. Let $f(\mathbf{x}) = \arg\max_{\mathbf{y}\in\mathcal{Y}} \langle \mathbf{w}, \phi(\mathbf{x}, \mathbf{y}) \rangle$ be a decision rule that



(a) (b) (c) (d)

Fig. 6: Centrality measures for different images: (a) Original image, (b) Closeness centrality, (c) Graph centrality and (d) Betweenness centrality. Stress centrality is not shown because in most of the cases, given our specific graph structure, it is equivalent to (d).

assigns a label to a data sample. Then, the SO-SVM problem is defined as:

$$\min_{\mathbf{w},\{\xi_i\}} \frac{1}{2}||\mathbf{w}||^2 + \frac{C}{N}\sum_{i=1}^{N}\xi_i$$

s.t. for $i = 1, \ldots, N$:

$$\Delta(\mathbf{y}_i, \mathbf{y}) + \langle \mathbf{w}, \phi(\mathbf{x}_i, \mathbf{y})\rangle - \langle \mathbf{w}, \phi(\mathbf{x}_i, \mathbf{y}_i)\rangle \leq \xi_i,$$

for all $\mathbf{y} \in \mathcal{Y}\backslash\{\mathbf{y}^n\}$.

The function $\Delta : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ is a loss function that represents the price we are willing to pay by predicting an estimated value instead of the true value for an instance of data, and $\phi(\mathbf{x}, \mathbf{y}) : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^D$ is a problem-dependent feature function that measures the correspondence between a data sample and a label. This is an optimization problem with $N|\mathcal{Y}|$ linear constraints and a convex, differentiable objective function that can be solved with off-the-shelf optimization software by following the iterative algorithm proposed in [31].

Learning problem formulation. We formulate our problem as to learn a localization function that predicts the bounding box of a wrinkle structure, centered on the lumen position, in a WCE frame. That is, $f(\mathbf{x})$: {all images $\rightarrow$ all image squared bounding boxes}. We are given a set of training pairs $(\mathbf{x}_1, \mathbf{y}_1) \ldots (\mathbf{x}_N, \mathbf{y}_N)$, where $\mathbf{x}_i$ are images and $\mathbf{y}_i$ consists of a label o indicating whether an object is present, and a four dimensional vector $(x_{tl}, y_{tl}, x_{br}, y_{br})$ indicating the top-left and bottom-right coordinates of the bounding box within the image: $\mathbf{y}_i \in \{(o, x_{tl}, y_{tl}, x_{br}, y_{br}) | o \in \{+1, -1\}, (x_{tl}, y_{tl}, x_{br}, y_{br}) \in \Re^4\}$. The objective is to learn a mapping f that generalizes from given examples.

To define the feature function $\phi(\mathbf{x}, \mathbf{y})$, we note that we have two different kinds of labels which combined define a class: a binary label indicating whether the bounding box contains a wrinkle structure or not, and four numerical labels representing the bounding box coordinates. Bounding box coordinates are clearly irrelevant for learning a good mapping f, because wrinkle patterns can be found at any image position, but the binary label defines a partition of the input space that should be represented in the feature function. Following the model proposed in [32], we define $\phi(\mathbf{x}, \mathbf{y})$ as a $2M$-dimensional vector $\phi(\mathbf{x}, \mathbf{y}) = (x_1, \ldots, x_M, 0, \ldots, 0)$ when $\mathbf{x}$ is a positive example (o = +1) and $\phi(\mathbf{x}, \mathbf{y}) = (0, \ldots, 0, x_1, \ldots, x_M)$ when $\mathbf{x}$ is a negative example (o = -1). Consequently, $\mathbf{w}$ must be also a $2M$-dimensional vector.

This feature representation induces the simultaneous learning of two weight vectors: a weight vector $\mathbf{w}_+ = (w_1, \ldots, w_M)$ for positive examples and a weight vector $\mathbf{w}_- = (w_{M+1}, \ldots, w_{2M})$ for negative examples. This scheme, in spite of the fact that increases the dimensionality of $\mathbf{w}$, has been shown useful for training Structured Output Support Vector Machines [32].

In SO-SVM, the loss function $\Delta(\mathbf{y}, \mathbf{y}')$ plays similar role as the margin in classical SVM. It measures how far a prediction $\mathbf{y}'$ is from a true label $\mathbf{y}$. Let $\mathbf{y}|_{bb}$ be the set of image pixels included in the bounding box represented in $\mathbf{y}$. Then, given a prediction $\mathbf{y}'$ and a true label $\mathbf{y}$, we defined the following loss function:

$$\begin{cases} \Delta(\mathbf{y}, \mathbf{y}') = 1 - A \text{ iff } \mathbf{y}|_{bb} \cap \mathbf{y}'|_{bb} \neq \emptyset. \\ \Delta(\mathbf{y}, \mathbf{y}') = 0 \text{ iff } \mathbf{y} = \mathbf{y}'. \\ \Delta(\mathbf{y}, \mathbf{y}') = 1 \text{ otherwise.} \end{cases}$$

where A is the Jaccard coefficient [33]:

$$A = \frac{\text{Area of } (\mathbf{y}|_{bb} \cap \mathbf{y}'|_{bb})}{\text{Area of } (\mathbf{y}|_{bb} \cup \mathbf{y}'|_{bb})}$$

.

Algorithm 1 Wrinkle frame detection

Input: A video frame $\mathbf{x}_i$ and a wrinkle frame model $\mathbf{w}^*$.

1: Compute for each pixel the Hessian matrix HM.
2: Compute for each pixel $\lambda_1, \lambda_2, \mathbf{e}_1$ and $\mathbf{e}_2$.
3: Compute an adaptive threshold value T and apply it to the image by selecting those pixels where $\lambda_1 > T$.
4: Divide the image in $n \times n$ non-overlapping cells.
5: Compute the histogram of features H corresponding to the $\mathbf{e}_2$ direction for each cell.
6: Build the graph $\mathcal{G}$ from H.
7: Compute the $n \times n$ vector C corresponding to the betweenness centrality of each image cell.
8: Define a $m \times m$ window W such that $m \leq n$.
9: **for** every possible position of W on the image **do**
10: Build a vector $\phi(\mathbf{x}_i, \mathbf{y}_i)$ by considering the centrality values of all the image cells inside W.
11: Compute $\mathbf{y}^*_{jk} = \arg\max_{\mathbf{y}\in\mathcal{Y}} \langle \mathbf{w}^*, \phi(\mathbf{x}_i, \mathbf{y}_i)\rangle$
12: **end for**
13: **if** there is a $\mathbf{y}^*_{jk} \neq (0, 0, 0, 0)$ **then**
14: **return** $\mathbf{y}_i = \mathbf{y}^*_{jk}$.
15: **else**
16: **return** $\mathbf{y}_i = (0, 0, 0, 0)$.
17: **end if**

Wrinkle frame detection. Finally, given a video frame $\mathbf{x_i}$ and $\mathbf{w}^*$ (computed by the SO-SVM), the detection process follows the steps presented in Algorithm 1. The dimension of the wrinkle model is directly related to the size (or number of cells) of the samples we have used to train the model. In our case, positive and negative samples are half of the size of a WCE frame in order to have a more localized response in the image.

Lines (1-7) of the algorithm show the computation of the values of the betweenness matrix, C. The algorithm uses a sliding window approach in order to evaluate the response of the model $\mathbf{w}^*$ at every location of the frame $\mathbf{x_i}$ (lines (8-14)). If at least one evaluation reports a positive detection, the image is considered a wrinkle frame.

Note that up to now we have been considering features (low- and mid-level) that were derived from frame intensity. We can get an improvement on the performance of this algorithm by considering the addition of color features, since color is an important visual cue that can improve wrinkle detection. Color information can be easily added by concatenating a few values representing the color inside the window hypothesis to the betweenness matrix C.

III. Results

In this section, we perform an evaluation of the proposed method. We compare the mid-level features based on betweenness centrality to several low-level images features such as the Histogram of Gradients (HoG) and the Histogram of Features (HoF) based on the Hessian using a standard linear SVM [34]. We show that the best results are obtained for the betweenness centrality descriptor. Moreover, we show that the color information is an important cue for wrinkle frame detection. Finally, we compare a standard linear SVM to the linear SO-SVM. Parameters in both classifiers were tuned using a cross-validation.

Database. In order to validate the proposed system, a training and a testing set were created using different videos obtained with a PillCam SB2 capsule provided by Given Imaging Ltd. Both, the training and the testing set were collected and labeled by experts at the original 256×256 image resolution. The training set consists of 1000 wrinkle frames and 1000 non wrinkles frames from 4 videos. The lumen center was manually labeled in all training wrinkle images. For each positive sample in the training set, 4 partial and 1 full wrinkles windows were considered, where partial windows means a window with partial overlapping with ground truth. Both full and partial windows consist of 128x128 image pixels with the corresponding label defined as the bounding box coordinates. Negative samples consist of 128x128 pixel image windows located at random locations of negative samples. The testing set consists of 1500 wrinkles frames and 2500 non wrinkles frames from 5 videos (not considered in the training set). All negative frames, from both training and testing set, were obtained by a random subset of non-wrinkle frames.

Measurements. In order to compare different wrinkle descriptors and classifiers the area under precision/recall curve (AUC), accuracy, Precision and recall are used. The wrinkle frames are considered the positive samples.

Validation. Table I presents the obtained results when using four different image descriptors and the sliding window approach: (i) the standard HoG descriptor computed from image gradients; (ii) a histogram descriptor built from the $\mathbf{e_2}$ values of the Hessian matrix, HoF; (iii) the proposed centrality C descriptor; (iv) the concatenation of the betweenness centrality and the color information, C_c. Color information has been defined as a 3-dimensional vector representing the mean RGB color of each cell.

Since wrinkles are tubular structures, intuitively they should be better represented by the Hessian matrix than by the distribution of Gradient vectors on the image. This hypothesis is confirmed by analyzing the obtained results presented in Table I and Fig. 7. By comparing AUC value of HoG and of HoF, an improvement of more than 10% can be seen. Also, it can be seen that the method presented in [18] is relatively worse than the HoF SVM method. The main problem of the method in [18] is that it relies on a good lumen detection instead of an sliding window approach. By further analysis of the obtained results, it can be seen that the proposed mid-level centrality descriptor outperforms the low-level information coded in the histogram by increasing AUC from 85.41% to 87.12% and the Accuracy from 83.53% to 87.76%. This result confirms that the relation between different image cells that is coded in the mid-level centrality vector C is useful for wrinkle detection. Finally, results confirm that color information is an important cue. The inclusion of color information provides further improvement in AUC from 87.12% to 92.35%. The precision/recall curves presented in Fig. 7 show that the centrality descriptor obtains a better compromise between precision and recall. The inclusion of the color information increases the performance of the detector by allowing the discrimination of frames with food content that sometimes resemble wrinkle structures.

Finally, we consider to use a linear SO-SVM instead of the linear SVM (see Table I and Fig. 7). The main difference between these two approaches is the consideration of *partial windows* during the learning process. These hypotheses, that correspond to bounding boxes that intersect the true bounding box in a wrinkle frame, have a clear regularization effect on the learned decision function that allows a better generalization to unseen samples. As it can be seen in the Table I, SO-SVM outperforms the standard SVM. In the case of using only the mid-level centrality descriptor the AUC increases from 87.12 to 92.04, and in the case of using the centrality descriptor plus color information the AUC increases from 92.35 to 96.07.

An statistical t-student test with $p-value = 0.05$ have been performed in order to validate the statistical significance of the results. The experiment has been done with 20 random runs using 50% of training data. The test has shown that C_c $SO-SVM$ method is statistically better than C_c SVM method.

A qualitative evaluation is presented in Fig. 8, 9a and 9b showing, respectively, True Positive (TP) detections, False Negative (FN) detections and False Positive (FP) detections. Figures present both, the original frame and a visualization of its corresponding centrality descriptor. As it can be seen in Fig. 8, the centrality descriptor for most TP samples shows the star-like shape and the cell related to the closed lumen is

TABLE I: Classification performance using different descriptors 1) Standard HoG, 2) Histogram of Features HoF, 3) Betweenness Centrality C and 4) Betweenness Centrality plus color information C_c. In the validation two types of linear classifiers were used: Support Vector Machines (SVM) and Structured Output Support Vector Machines (SO-SVM).

Descriptor Classifier	HoG SVM	HoF SVM	C SVM	C SO-SVM	C_c SVM	C_c SO-SVM
AUC	73.25	85.41	87.12	92.04	92.35	**96.07**
Accuracy	71.92	82.79	83.53	87.76	89.11	**92.38**
Precision	65.04	79.70	79.66	82.55	83.45	**91.85**
Recall	64.93	67.90	69.90	78.80	81.70	**84.48**

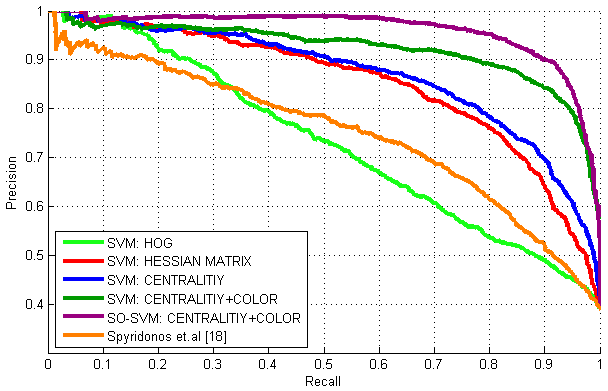


Fig. 7: Precision/Recall curves.

the one with highest centrality value. On the other hand, we can see in Fig. 9a that most of FN present very smooth folds of intestinal wall and a completely closed lumen. These issues make it difficult to properly characterize the frame with the centrality descriptor, since in most of these cases the star-like shape is not observed. Finally, we can see in Fig. 9b that some of the FP are difficult to be labeled, for instance, images from the third row (second and third image from left) present several folds. Moreover, the majority of FP contains some intestinal content which makes difficult to see the lumen, and so, the proper image classification.

For the physician, it is important to see how the wrinkle frames are distributed along the small intestine. To this end, we display in Fig. 10 the regions where our system detects a high percentage of wrinkle frames together with a motility bar video representation obtained with the tool proposed in [35]. The corresponding video segment is 20 minutes long, which

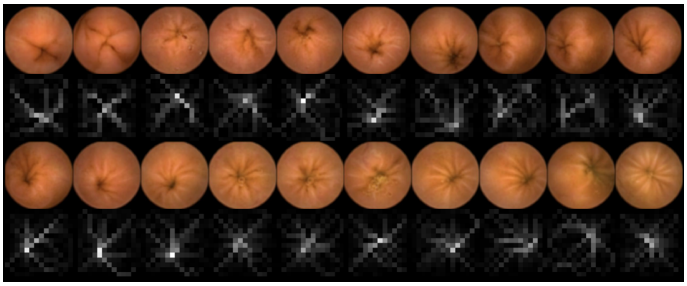
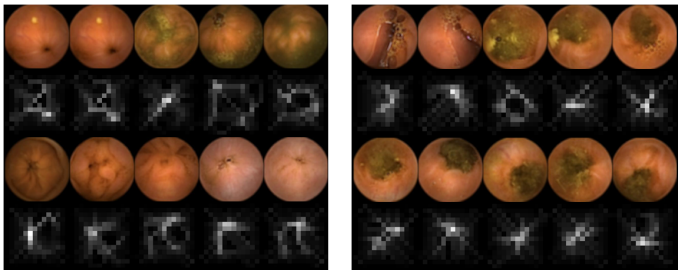


Fig. 8: True Positive detections.



(a) (b)

Fig. 9: Visual evaluation of the results: (a) False Negative detections, (b) False Positive detections.

means 2.400 frames. We implemented the method in Matlab and run it on Intel I5-2520 CPU machine. The time needed to obtain wrinkle score for this video segment was approx. 30 minutes.

IV. Conclusions

In this paper, we presented a new image descriptor for the classification of WCE wrinkle frames. The proposed image descriptor is based on a new image centrality descriptor which is based on the histogram of oriented features extracted from the Hessian matrix of an image. This mid-level descriptor integrates global image information that is useful to detect star-like shape patterns.

The detection process is based on a model learned by using a Structural Output Support Vector Machine approach. This approach not only uses positive and negative samples, but also samples that correspond to partial hypotheses. This inclusion produces better detection models.

The detection process is performed by a sliding window procedure that scans the image looking for a positive label. This allows to train more accurate models that can be applied in a multi-scale architecture in order to get better localization. A second advantage of this approach is that there is no need to detect the lumen center previously to the classifier application. The low complexity of all involved algorithms allows for near real time processing of WCE frames.

The validation, carried out on a large database, shows that the proposed descriptor successfully detects this particular event of WCE videos, outperforming previous methods and defining a new state of the art for this problem.

Acknowledgements

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, as well as by Spanish MINECO Grants TIN2009-14404-C02 and TIN2012-38187-C03.

References

[1] H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa, and Y. Yagi, "Contraction detection in small bowel from an image sequence of wireless capsule endoscopy," in *Proc. of the 10th international Conf. on Med. Image Computing and Computer-Assisted Intervention - Volume Part I*, ser. MICCAI'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 775–783.

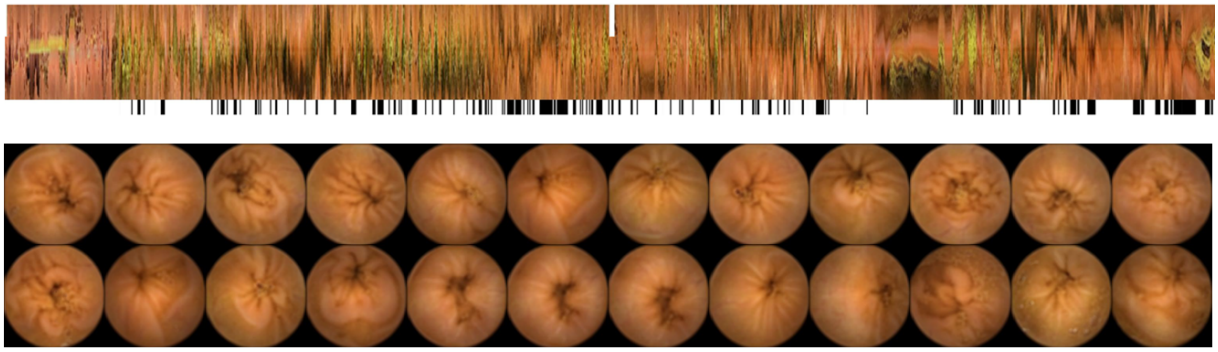


Fig. 10: The image on the top shows the results of the proposed wrinkle detector on a video mosaic. Black vertical bars correspond to the video segments with high density of frames detected as wrinkles. The bottom image shows some random frames detected as wrinkles by the system.

[2] F. Vilariño, P. Spyridonos, F. D. Iorio, J. Vitrià, F. Azpiroz, and P. Radeva, "Intestinal motility assessment with video capsule endoscopy: Automatic annotation of phasic intestinal contractions," *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 246–259, 2010.

[3] H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa, and Y. Yagi, "Detection of contractions in adaptive transit time of the small bowel from wireless capsule endoscopy videos." *Comp. in Bio. and Med.*, vol. 39, no. 1, pp. 16–26, 2009.

[4] E. M. Quigley, "Gastric and small intestinal motility in health and disease." *Gastroenterol Clin North Am*, vol. 25, no. 1, pp. 113–45, 1996.

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

[6] 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.

[7] H. Liu, N. Pan, H. Lu, E. Song, Q. Wang, and C.-C. Hung, "Wireless capsule endoscopy video reduction based on camera motion estimation," *J. of Digital Imaging*, vol. 26, no. 2, pp. 287–301, 2013.

[8] S. Seguí, M. Drozdzal, F. Vilariño, C. Malagelada, F. Azpiroz, P. Radeva, and J. Vitrià, "Categorization and segmentation of intestinal content frames for wireless capsule endoscopy," *Information Technology in Biomedicine, IEEE Trans. on*, vol. 16, no. 6, pp. 1341–1352, 2012.

[9] J. Cunha, M. Coimbra, P. Campos, and J. M. Soares, "Automated topographic segmentation and transit time estimation in endoscopic capsule exams," *Med. Imaging, IEEE Trans. on*, vol. 27, no. 1, pp. 19–27, 2008.

[10] Y. Fu, W. Zhang, M. Mandal, and M. Meng, "Computer-aided bleeding detection in wce video," *Biomedical and Health Informatics, IEEE J. of*, vol. PP, no. 99, pp. 1–1, 2013.

[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 Trans. on*, vol. 59, no. 2, pp. 355–362, 2012.

[12] Y. Wang, W. Tavanapong, J. Wong, J. Oh, and P. de Groen, "Part-based multi-derivative edge cross-section profiles for polyp detection in colonoscopy," *Biomedical and Health Informatics, IEEE J. of*, vol. PP, no. 99, pp. 1–1, 2013.

[13] B. Li and M.-H. Meng, "Tumor recognition in wireless capsule endoscopy images using textural features and svm-based feature selection," *Information Technology in Biomedicine, IEEE Trans. on*, vol. 16, no. 3, pp. 323–329, 2012.

[14] "Texture analysis for ulcer detection in capsule endoscopy images," *Image and Vision Computing*, vol. 27, no. 9, pp. 1336 – 1342, 2009.

[15] C. Malagelada, F. De Iorio, F. Azpiroz, A. Accarino, S. Seguí, P. Radeva, and J. Malagelada, "New insight into intestinal motor function via noninvasive endoluminal image analysis," *Gastroenterology*, vol. 135, no. 4, pp. 1155–1162, 2008.

[16] C. Malagelada, S. Seguí, S. Mendez, M. Drozdzal, J. Vitrià, P. Radeva, J. Santos, A. Accarino, J. Malagelada, F. Azpiroz *et al.*, "Functional gut disorders or disordered gut function? small bowel dysmotility evidenced by an original technique," *Neurogastroenterology & Motility*, 2012.

[17] F. Vilariño, P. Spyridonos, J. Vitrià, C. Malagelada, and P. Radeva, "Linear radial patterns characterization for automatic detection of tonic intestinal contractions," in *CIARP*, 2006, pp. 178–187.

[18] P. Spyridonos, F. Vilariño, J. Vitrià, F. Azpiroz, and P. Radeva, "Anisotropic feature extraction from endoluminal images for detection of intestinal contractions," in *MICCAI (2)*, 2006, pp. 161–168.

[19] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1994.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Comp. Soc. Conf. on*, vol. 1. IEEE, 2005, pp. 886–893.

[21] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[22] T. Joachims, T. Hofmann, Y. Yue, and C.-N. Yu, "Predicting structured objects with support vector machines," *Communications of the ACM*, vol. 52, no. 11, pp. 97–104, 2009.

[23] S. Seguí, M. Drozdzal, E. Zaysteva, C. Malagelada, F. Azpiroz, P. Radeva, and J. Vitrià, "A new image centrality descriptor for wrinkle frame detection in WCE videos," in *13th IAPR Conf. on Machine Vision Applications*. IAPR, 2013.

[24] R. M. Haralick, "Ridges and valleys on digital images," *Computer Vision, Graphics, and Image Processing*, vol. 22, no. 1, pp. 28–38, 1983.

[25] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[26] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[27] P. Hage and F. Harary, "Eccentricity and centrality in networks," *Social networks*, vol. 17, no. 1, pp. 57–63, 1995.

[28] A. Shimbel, "Structural parameters of communication networks," *The bulletin of mathematical biophysics*, vol. 15, no. 4, pp. 501–507, 1953.

[29] U. Brandes, "A faster algorithm for betweenness centrality*," *J. of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[30] V. Vapnik, "The nature of statistical learning theory," *Data mining and knowledge discovery*, pp. 1–47, 1995.

[31] I. Tsochantaridis, T. Joachims, T. Hofmann, Y. Altun, and Y. Singer, "Large margin methods for structured and interdependent output variables," *J. of Machine Learning Research*, vol. 6, no. 2, p. 1453, 2006.

[32] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *J. of Machine Learning Research*, vol. 2, pp. 265–292, 2002.

[33] P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.

[34] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[35] M. Drozdzal, S. Seguí, J. Vitrià, C. Malagelada, F. Azpiroz, and P. Radeva, "Adaptable image cuts for motility inspection using WCE," *Computerized Med. Imaging and Graphics*, vol. 37, no. 1, pp. 72–80, 2013.