Manuscript of:
Piotr M. Szczypinski, Ram D. Sriram,
Parupudi V.J. Sriram, D. Nageshwar Reddy,
A model of deformable rings for interpretation of wireless capsule endoscopic videos,

Published:
Elsevier, Medical Image Analysis, Volume 13, Issue 2,
April 2009, Pages 312-324, ISSN 1361-8415,
DOI: 10.1016/j.media.2008.12.002.

A Model of Deformable Rings for Interpretation of Wireless Capsule Endoscopic Videos

Piotr M. Szczypiński,
Ram D. Sriram,
Parupudi VJ Sriram,
D. Nageshwar Reddy

Piotr M. Szczypiński, PhD, is with the Institute of Electronics, Technical University of Lodz, Poland (phone: +4842 6312638, e-mail: pms@p.lodz.pl)
Ram D. Sriram, PhD, is with the National Institute of Standards and Technology, Gaithersburg, MD 20899 (phone: +1 3019753507, e-mail: sriram@nist.gov)
Parupudi VJ Sriram, MD, is with the Department of Internal Medicine (Gastroenterology), Texas Tech University Health Sciences Center, Lubbock, Texas
Nageshwar Reddy, MD, is the chief of Asian Institute of Gastroenterology, Hyderabad, India.

Abstract

Wireless Capsule Endoscopy (WCE) provides a means to obtain a detailed video of the small intestine. A single session with WCE may produce nearly 8 hours of video. Its interpretation is tedious task, which requires considerable expertise and is very stressful. The Model of Deformable Rings (MDR) was developed to preprocess WCE video and aid clinicians with its interpretation. The MDR uses a simplified model of a capsule's motion to flexibly match (register) consecutive video frames. Essentially, it computes motion-descriptive characteristics and produces a two-dimensional representation of the gastrointestinal (GI) tract's internal surface – a map. The motion-descriptive characteristics are used to indicate video fragments which exhibit segmentary contractions, peristalsis, refraction phases and areas of capsule retention. Within maps, certain characteristics that indicate areas of bleeding, ulceration and obscuring froth could be recognized. Therefore, the maps allow quick identification of such abnormal areas. The experimental results demonstrate that the number of discovered pathologies and gastrointestinal landmarks increases with the MDR technique.

Keywords

Model of deformable rings, deformable models, video analysis, wireless capsule endoscope, gastrointestinal system

1. Introduction

The human small intestine, which measures approximately 6 meters, is difficult to visualize completely using a traditional endoscopic approach. Wireless Capsule Endoscopy (WCE) [1-7], which is primarily marketed by Given Imaging®, is a relatively new technique that facilitates the imaging of the small intestine. The WCE system consists of a pill-shaped capsule (Fig. 1) with built-in video camera, light-emitting diodes, video signal transmitter and

battery, as well as a video signal receiver-recorder device. The capsule is ingested by a patient and passes through the gastrointestinal (GI) tract. It is propelled by peristaltic movements within the GI tract and by gravity. There is no mechanism to control the capsule's speed or direction as it traverses the GI tract [16, 17]. It is also not possible to measure the capsule's location or orientation during its traversal within the human body [18, 19].

The capsule transmits color images of the internal lumen of the GI tract at a rate of two frames per second for approximately 8 hours. The images are 256×256 pixels. The 140° field of view covers a circular area in the center of a frame which is about 240 pixels in diameter. The transmitted images are received and recorded by an external receiver-recorder device.

The investigation of WCE's video recordings is performed by a trained clinician. It is a tedious task that takes a considerable amount of time, usually more than an hour per recording. The video interpretation involves viewing the video and searching for bleedings, erosions, ulcers, polyps and narrow sections of the bowel due to disease or any other abnormal-looking entities. The interpretation always requires a high level of concentration for viewing the entire video recording, so as not to miss focal lesions that might be present in only a few frames and could be easily missed.

The WCE video frames may be classified into three categories: 1) cases of normal mucosal appearance, 2) abnormal findings (most likely diseased fragments of GI tract), and 3) frames showing an intestinal content obscuring a view of GI tract (e.g. bubbles, digested or partly digested food). Also useful would be the segmentation of video into fragments representing stomach, duodenum, small intestine and colon, or to detect topographic marks such as pylorum or ileo-cecal valve. An automatic method that would provide such a categorization for WCE video images or one that provides some additional information useful for such categorization (which is not immediately available from the video) would improve the effectiveness of the interpretation process. A trained gastroenterologist would decide which video fragment should be viewed with the highest concentration, and which fragments may be disregarded due to froth or partially digested luminal content that obscures the mucosal details and makes interpretation infeasible.

Several research groups have recently reported methods for GI tract segmentation into sections. In [12-15] image color and texture descriptors defined by MPEG-7 standard are used for topographic segmentation of GI tract into sections. A similar approach has been presented in [16-17], where the color and texture descriptors are used for image classification and pylorus detection.

Another classification procedure is presented in [18]. Image features, such as image intensity, contrast and selected geometrical parameters, are extracted from WCE video frames. Then, feature vectors computed for nearby frames are compared. The WCE video is divided into fragments containing contractions, when the vectors vary, and non-contractions fragments, when the vectors are similar. Authors demonstrate that fragments with contraction are of special interest to the physician [18].

Methods for continuous tracking of an endoscope camera motion were reported in [19-23]. The methods use epipolar geometry analysis supported by registration of real and virtual endoscopy images. Virtual endoscopy images can be created by 3D modeling on computed tomography (CT) data, whereas the real images are produced during endoscopic examination. These methods have proven their usefulness with image guided bronchoscopy. However, the problem of tracking motion within WCE video is different. A small intestine, unlike trachea or bronchi, is elastic and in motion due to peristalsis. Therefore, 3D visualization is difficult, and also epipolar equation cannot be directly applied. Furthermore, WCE examination is rarely combined with CT examination or any other 2D or 3D visualization technique.

Based on our earlier work in image processing [24-30] and experience in endoscopic examinations [31], we have developed the Model of Deformable Rings (MDR). The MDR is a unique technique that registers consecutive WCE video frames and follows the capsule's motion. The goal of the model is to process the video, which is of low spatial and temporal resolution, when no additional data on localization or orientation of the capsule is available, and no other visualization technique was used. The outcome should be medically useful to

make the video interpretation more efficient, to reduce number of missed abnormalities or reduce the time spent on the interpretation.

The MDR computes motion descriptors and a two-dimensional map of the internal surface of the digestive system. The map serves as a quick reference to the video sequences, supports identification of segments of the bowel and can be glanced through for quick identification of large-scale abnormalities. The capsule motion descriptors provide data sufficient for localization of video sequences that show narrow sections of the GI tract, where the capsule either stopped or slowed down considerably.

This paper is organized as follows. In the next section we describe the MDR technique, which includes an explanation of the motion tracking technique, the concept of elasticity and a method for local tension computation with regard to a variable size neighborhood. We follow this by a general explanation of the model's behavior, introduce the original concept of forming a digestive system map and present a method of motion descriptors computation. We also explain our approach to the selection of a motion estimation method and the model's parameters. Then, we demonstrate our technique through selected examples of GI tract maps and images, followed by comments on how the results of the WCE video preprocessing can support the interpretation of presented cases. Finally, we provide results that compare WCE video interpretations with and without use of the MDR, present our conclusions and enumerate areas for future work.

2. Model of deformable rings

2.1. Overview

The key concept behind WCE video processing with MDR involves tracking movements of digestive system walls by elastic matching (registration) of consecutive video frames. The model estimates intensity of image motion toward and outward from the center of the video frame.

If portions of the image shift outward from the image center it indicates the capsule passes through the GI tract and its camera points forward with respect to the motion's direction. When portions of the image shift toward its center, it means the camera points back. Thus, essentially the MDR is a tool for partial egomotion estimation of the capsule.

The MDR is a member of the deformable models family [24-30, 32-38]. As with many other deformable models it is composed of active interconnected nodes. The uniqueness of MDR lies in the arrangement of these nodes which form concentric rings with their centers located near the center of the image frame.

Essentially, the model works as follows: The nodes store portions of a current image frame, or local image properties, found at their locations. The image frame, then, changes to the subsequent one. Within the new frame, nodes search the image to find fragments resembling the fragments stored, or having similar properties to the properties stored. Then, nodes are shifted toward the locations of these image fragments that were found. To preserve the arrangement of MDR nodes, the model incorporates tension computation. Therefore, on one hand nodes push toward the locations they found to be similar to the stored ones and on the other hand the excessive movement is prevented by tension modeling. The final location of each node is found iteratively, after obtaining a balance between the two factors. The balance is found by minimization of the model's energy function, which we define as follows:

$$E_{MDR} = \sum_{p,q} \left(M_{p,q} + E_{p,q} \right) \tag{1}$$

The energy is a total of individual nodes' energies. A node's energy is composed of two components. The $M_{p,q}$ component models the image influence and should reach its minimum when the node is located at the image fragment which is most similar to the stored one. The $E_{p,q}$ component is responsible for tension modeling, and is minimal when there are no local deformations near the node. The E_{MDR} energy depends on the nodes' locations and is a function of their coordinates. Therefore, the minimization of energy function results in changing the model's shape, orientation and position.

To minimize the E_{MDR} energy a gradient descent optimization is applied. The minimum of a function is estimated iteratively, by modifying a vector of the function arguments by a vector proportional to the negative of the functions gradient, given by equation (2).

$$\nabla E_{MDR} = \nabla\left(\sum_{p,q}\left(M_{p,q} + E_{p,q}\right)\right) = \sum_{p,q}\left(\nabla M_{p,q} + \nabla E_{p,q}\right) \tag{2}$$

After a specified number of iterations the image characteristics are again stored within the nodes' memories and the process repeats for the subsequent frame. It continues until the last frame of the video sequence is reached.

We assumed a simplified nature of a capsule's motion, that most of the motion is longitudinal with respect to the longitudinal axis of the capsule. The capsule is elongated and surrounded tightly by walls of the GI tract. This causes the capsule to align its longitudinal axis with the course of the tract and also to move along this direction most of the time. The model tracks the longitudinal motion and a roll rotation of the capsule. It turns around the image center in case of roll rotation. The model shrinks during backward motion, or grows during forward motion. The MDR size changes are used for longitudinal velocity estimation of the WCE within the GI tract. The MDR's ability to follow other motion components is restricted; however, type and strength of such motion components are characterized.

When processing video data, MDR rings are virtually cut and laid out flat as a two-dimensional map. The map is a key result produced by the model. The map can be correctly assembled when the capsule aligns its longitudinal axis with the course of the tract. Otherwise only one side of the surrounding walls is shown, which makes it impossible to do the complete tract's surface reconstruction. We believe that the ability of the model to track only the longitudinal motion and rotation satisfies requirements for the map assembly.

2.2. Representation

A model of deformable rings consists of nodes, which are connected to form a mesh. The mesh is positioned in the plane of a video frame. It forms concentric rings surrounding the center of the frame (Fig. 2). Every node of the mesh is referred to by a pair of indexes p = 1, 2,...P and q = 1, 2,...Q., where P is a number of rings and Q is a number of nodes per ring. The initial location of a node within the image coordinate system is given by the following formula:

$$\left[\hat{x}_{p,q} \quad \hat{y}_{p,q}\right]^{\mathrm{T}} = rw^{p-1}\left[\cos\left(\frac{2\pi q}{Q}\right) \quad \sin\left(\frac{2\pi q}{Q}\right)\right]^{\mathrm{T}} \tag{3}$$

The initial location is a location on the first frame of a video sequence, before the actual video processing starts. The r parameter is a radius of inner ring and the w is a ratio of radii of adjacent rings. The center of coordinate system here is placed in the center of a WCE video frame. Parameters r and w are adjusted to fit the model within a valid area of the video frame, leaving some margin of the valid image on the sides (Fig. 2.b). The shape given by (3) serves both as an initial shape of the model and as an undeformed reference shape used for tension computation, which will be explained in the following sections.

2.3. Image influence

Since the basic task of the MDR is to follow the motion of image content, the crucial problem is to select an appropriate form of the image influence energy ($M_{p,q}$). The energy should specify local image dissimilarity. The most common dissimilarity measures used in motion estimation are mean square error (*MSE*) and mean absolute difference (*MAD*). Similarity measures commonly used with video processing or image registration applications are normalized cross-correlation (*NCC*) [39, 40] and mutual information (*MI*) [41].

The *MAD* and *MSE* functions of $N\times N$ pixels block located at ($x+\Delta x$, $y+\Delta y$) inside the current frame and a $N\times N$ pixels block located at (x, y) in a previous frame are defined as follows:

$$MAD(\Delta x, \Delta y) = \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left|I_k(x+\Delta x+m,\ x+\Delta y+n) - I_{k-1}(x+m,\ y+n)\right| \quad (4)$$

$$MSE(\Delta x, \Delta y) = \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left(I_k(x+\Delta x+m,\ x+\Delta y+n) - I_{k-1}(x+m,\ y+n)\right)^2 \quad (5)$$

where I_k is a current frame and $I_{k\text{-}1}$ is a previous frame of a video sequence. In the MDR the required block of a k-1 frame is stored within a node memory. The minimum value of dissimilarity measure indicates that regions are alike. Since the digital image's space is discrete the x, y, Δx and Δy are integer numbers.

The *NCC* of two image blocks is defined as follows:

$$NCC(\Delta x, \Delta y) = \frac{\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left(I_k(x+\Delta x+m,\ x+\Delta y+n) - \mu_k\right)\left(I_{k-1}(x+m,\ y+n) - \mu_{k-1}\right)}{\sigma_k \sigma_{k-1}} \quad (6)$$

where μ is the mean value of block histogram and σ is the standard deviation.

The *MI* [41] of two discrete random variables A and B can be defined as follows:

$$MI(A,B) = \sum_a \sum_b p(a,b)\log\frac{p(a,b)}{p(a)\,p(b)} \quad (7)$$

where $p(a, b)$ is the joint probability distribution function of A and B, and $p(a)$ and $p(b)$ are the probability distribution functions of A and B, respectively. In our approach A is the distribution of grey values of $N\times N$ pixels image block located at (x, y) within the current frame and B is the distribution of grey values of such a block located at $(x+\Delta x, y+\Delta y)$ within a previous frame.

Dissimilarity measures or negative of similarity measures multiplied by a ξ parameter are used to model the image influence $M_{p,q}$ in equation (1). The ξ controls contribution of image influence component in the model's energy.

The gradient of dissimilarity function is computed by means of Sobel operator. For non-integer coordinates of nodes, it is bilinear interpolated.

Results of MDR performance and comparisons of various methods for image influence computation were reported in [30] and are presented in the section 3.1.

2.4. Tension computation

We use a novel technique for the tension computation. This technique compares current location of a node referenced to its location within the undeformed model against current location of neighboring nodes (n-neighborhood) referenced to their location within the undeformed model. We define the n-neighborhood (N) of node p, q as a set of all MDR nodes (k, l) connected with the node p, q by n or less number of connecting lines, and it includes the node p, q itself.

The *MDR* follows image motion, which may include translation, change in size and orientation. Therefore, the transformation of a neighborhood from its initial location given by (3) can be defined by a translation vector $\mathbf{T}_{p,q}$ and a scaling-rotation matrix $\mathbf{J}_{p,q}$ (Fig. 3).

$$\mathbf{J}_{p,q} = \begin{bmatrix} j_{11} & j_{12} \\ -j_{12} & j_{11} \end{bmatrix} \quad (8)$$

$$\mathbf{T}_{p,q} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \quad (9)$$

Vector $\mathbf{T}_{p,q}$ and matrix $\mathbf{J}_{p,q}$ are computed to minimize a mean square error $\varepsilon_{p,q}$ (10) between actual locations of neighborhood nodes and locations computed by means of the affine transformation.

$$\varepsilon_{p,q} = \sum_{k,l\in N}\left(\left|\begin{bmatrix} x_{k,l} \\ y_{k,l} \end{bmatrix} - \mathbf{J}_{p,q}\begin{bmatrix} \hat{x}_{k,l} \\ \hat{y}_{k,l} \end{bmatrix} - \mathbf{T}_{p,q}\right|^2\right) \quad (10)$$

where $[x_{k,l}, y_{k,l}]^{\mathrm{T}}$ is a current location of node k, l, which belongs to the neighborhood of the node p, q.

The $\varepsilon_{p,q}$ is a function of j_{11}, j_{12}, t_1 and t_2. The minimum of the function exists and hence, at the minimum, derivatives of $\varepsilon_{p,q}$ with respect to the j_{11}, j_{12}, t_1 and t_2 must be equal to zero. The p,q indexes are omitted for simplicity of notation.

$$\frac{\mathrm{d}\,\varepsilon}{\mathrm{d}\,j_{11}} = 2\sum\left(j_{11}\left(\hat{x}^2 + \hat{y}^2\right) - \hat{x}\left(x - t_1\right) - \hat{y}\left(y - t_2\right)\right) = 0 \tag{11a}$$

$$\frac{\mathrm{d}\,\varepsilon}{\mathrm{d}\,j_{12}} = 2\sum\left(j_{12}\left(\hat{x}^2 + \hat{y}^2\right) - \hat{y}\left(x - t_1\right) + \hat{x}\left(y - t_2\right)\right) = 0 \tag{11b}$$

$$\frac{\mathrm{d}\,\varepsilon}{\mathrm{d}\,t_1} = 2\sum\left(t_1 - x + j_{11}\hat{x} + j_{12}\hat{y}\right) = 0 \tag{11c}$$

$$\frac{\mathrm{d}\,\varepsilon}{\mathrm{d}\,t_2} = 2\sum\left(t_2 - y - j_{12}\hat{x} + j_{11}\hat{y}\right) = 0 \tag{11d}$$

The solution to the above system of equations is:

$$\begin{bmatrix} j_{11} \\ j_{12} \\ t_1 \\ t_2 \end{bmatrix} = \frac{\begin{bmatrix} \sum 1 & 0 & -\sum \hat{x} & -\sum \hat{y} \\ 0 & \sum 1 & -\sum \hat{y} & \sum \hat{x} \\ -\sum \hat{x} & -\sum \hat{y} & A & 0 \\ -\sum \hat{y} & \sum \hat{x} & 0 & A \end{bmatrix}}{A\sum 1 - \left(\sum \hat{x}\right)^2 - \left(\sum \hat{y}\right)^2} \begin{bmatrix} B \\ C \\ \sum x \\ \sum y \end{bmatrix} \tag{12}$$

where: $A = \sum \hat{x}^2 + \sum \hat{y}^2; B = \sum \hat{x}x + \sum \hat{y}y; C = \sum \hat{y}x - \sum \hat{x}y$

We define the $E_{p,q}$ energy as a sum of two components, weighted with ρ and ζ parameters. The first component is proportional to the square of distance between a current location of a node p,q and its location transformed with $\mathbf{T}_{p,q}$ and $\mathbf{J}_{p,q}$. It takes into account the local tension modeling. The second component keeps the MDR near the center of a video frame. It is proportional to the square of global MDR displacement from the image center.

$$E_{p,q} = \frac{\rho}{2}\left\| \begin{bmatrix} x_{p,q} \\ y_{p,q} \end{bmatrix} - \mathbf{J}_{p,q} \begin{bmatrix} \hat{x}_{p,q} \\ \hat{y}_{p,q} \end{bmatrix} - \mathbf{T}_{p,q} \right\|^2 + \frac{\zeta}{2PQ}\left(\sum_{k,l \in MDR} \begin{bmatrix} x_{k,l} \\ y_{k,l} \end{bmatrix} \right)^2 \tag{13}$$

The $E_{p,q}$ (13) is a function of $x_{p,q}$ and $y_{p,q}$. Therefore, the energy gradient is computed using the following formula:

$$\nabla E_{p,q} = \rho\left(\begin{bmatrix} x_{p,q} \\ y_{p,q} \end{bmatrix} - \mathbf{J}_{p,q} \begin{bmatrix} \hat{x}_{p,q} \\ \hat{y}_{p,q} \end{bmatrix} - \mathbf{T}_{p,q} \right) + \frac{\zeta}{PQ} \sum_{k,l \in MDR} \begin{bmatrix} x_{k,l} \\ y_{k,l} \end{bmatrix} \tag{14}$$

The proposed method for tension computation is computationally effective. Moreover, it is easy to compute tensions for nodes located at the border of a deformable model. This is not straightforward with methods, which utilize second order differential equations, commonly used with deformable models. The other advantage of the method is its potential to compute tensions within irregular meshes.

2.5. Elastic matching

The elastic matching is a process of successive displacements of nodes intended for finding a state, in which balance between the two effects, image motion and tension, is obtained. In successive iterations (i), nodes move toward locations, in which balance between the vectors is achieved. The displacement of a node is computed using the following formula:

$$\begin{bmatrix} x_{p,q}^{(i+1)} \\ y_{p,q}^{(i+1)} \end{bmatrix} = \begin{bmatrix} x_{p,q}^{(i)} \\ y_{p,q}^{(i)} \end{bmatrix} - \nabla M_{p,q}^{(i)} - \nabla E_{p,q}^{(i)} \tag{15}$$

The displacements are computed for all the nodes p,q and then process is repeated several iterations per frame. The process of displacing nodes is repeated until some state of equilibrium is reached, i.e., average displacement distance of the MDR nodes drops below some selected threshold value, or until an iteration index reaches some arbitrarily chosen maximum (I). When the process is completed, image properties found at node coordinates are stored, the next video frame is presented to the algorithm and a process of elastic matching is repeated. This continues until the last frame of the video sequence is reached.

It was experimentally found that the process of matching is more efficient during the initial stages, if a neighborhood n is large enough to comprise all the nodes. In this way, the $\mathbf{T}_{p,q}$ and the $\mathbf{J}_{p,q}$ are computed for the entire model and only once per iteration. Therefore, the model quickly adjusts its position, size and orientation. This phase of matching may be compared with rigid image registration. Later, the n parameter is decreased, and the MDR deforms locally, matching the image details. This phase may be compared with non-rigid image registration.

2.6. Rings swapping and map assembling

As already mentioned, if the capsule moves forward, the MDR expands while matching consecutive video frames. If the capsule moves backward the model shrinks. We determine the present size of MDR in relation to its initial state (3) by computing the $\mathbf{J}_g$ matrix. The $\mathbf{J}_g$ matrix is computed the same way as the $\mathbf{J}_{p,q}$ matrix, with a neighborhood including all the MDR nodes. The square root of determinant $\mathbf{J}_g$ is a measure of model's size in relation to its initial size.

To prevent the model from excessive expansion or from shrinking, limits on the model size are set. If the model is too big, i.e., $\det \mathbf{J}_g > w^2$, then the outer ring of a model is erased and a new inner ring is created. If it is too small, i.e., $\det \mathbf{J}_g < w^{-2}$, then the inner ring is erased and a new outer one is added. In either case, the image content is sampled along the outer ring to form a row of pixels. All such rows collected during the video processing are put together to make an image – the map of GI tract.

2.7. Velocity estimation and deformation coefficient

We use two approaches to estimate the capsule's velocity. In the first one, we assume that digestive system walls stick to the capsule's casing including its transparent dome (Fig. 4.a). The second approach assumes the digestive system is akin to a rigid pipe (Fig. 4.b) of a diameter close to the capsule's diameter.

In the first approach, since the dome of the capsule is spherical, the distance S the capsule moves about is proportional to the change in an angle φ, at which a selected fragment of the tract is viewed by the capsule's camera. When the surface of the image sensor is considered, the fragment appears at a distance $d = f\tan(\varphi)$ form the center of the sensor. Therefore, the velocity is estimated using the following equation:

$$V_D = \frac{r_D}{\Delta t}\left(\tan^{-1}\left(\frac{d_1}{f}\right) - \tan^{-1}\left(\frac{d_2}{f}\right)\right) \tag{16}$$

where d_1 and d_2 are sizes of the MDR observed at two different times, and Δt is a time between the two observations. The d parameters are derived from the $\mathbf{J}_g$ matrix determinant, using the following formula:

$$d = r\sqrt{w^{P-1}\left|\det \mathbf{J}_g\right|} \tag{17}$$

If the rings' swapping occurs, the d_1 parameter is computed afterward.

In the second approach the perspective projection scheme is applied. The distance to the object is proportional to the reciprocal of its image size produced on the surface of the image sensor. Thus, the capsule's velocity is estimated using the following equation:

$$V_P = \frac{r_D f}{\Delta t}\left(\frac{1}{d_2} - \frac{1}{d_1}\right) \tag{18}$$

The two estimated velocities are computed each time the rings are swapped and a new row of pixels is appended to the GI map. Therefore, every line of the digestive system map is assigned two velocity values: V_D and V_P.

Since the capsule's motion is not always smooth and parallel to its symmetry axis, the velocity estimation is usually inaccurate. It is also hard to determine which of the two approaches should be used to estimate the velocity in a particular case.

In our study of the MDR, we have also explored plots of the models motion descriptive coefficients, which are computed per frame after the matching process finalization. The models displacement from the image center determines amount of motion components caused by pitch and yaw rotations or sideways movement. It is defined by the following formula:

$$G_1 = \frac{1}{PQ}\left|\sum_{k,l\in MDR}\begin{bmatrix} x_{k,l} \\ y_{k,l}\end{bmatrix}\right| \tag{19}$$

The deformation coefficient measures the amount of the local model's tension. The coefficient may indicate video fragments which exhibit deformations of GI tract due to contractions, fragments in which the motion was jumpy or the image changed significantly from one frame to another. The deformation coefficient is defined as follows:

$$G_2 = \frac{1}{PQ}\sum_{p,q\in MDR}\left|\begin{bmatrix} x_{p,q} \\ y_{p,q}\end{bmatrix} - \mathbf{J}_{p,q}\begin{bmatrix} \hat{x}_{p,q} \\ \hat{y}_{p,q}\end{bmatrix} - \mathbf{T}_{p,q}\right|^2 \tag{20}$$

Both the G_1 and G_2 coefficients are useful in pointing these video fragments, which should be thoroughly investigated. We discuss this concept in section 3.2.

3. Results

The MDR was implemented in C++ as a video processing module for Microsoft® Direct Show technology. In this study the MDR with $P = 7$ rings with $Q = 128$ nodes per each ring, and parameters $r = 70$ pixels and $w = 1.04$ have been used.

Three experiments were performed to optimize the model's performance and to assess its usefulness in aiding the capsule endoscope video interpretation. The first one compares the MDR performance implementing various dissimilarity measures for image influence computation. The second experiment demonstrates the MDR ability to discriminate various motility patterns. The third one verifies whether the use of the MDR map improves the interpretation results. We also present example cases in which maps show informative details beneficial for interpretation.

3.1. Image dissimilarity measure

We considered four dissimilarity measures for image influence energy computation: - *MI*, - *NCC*, *MSE* and *MAD*. Therefore, we designed an experiment that verified these methods and let us select the most appropriate one for WCE video preprocessing. The selected method had to satisfy the following conditions: 1) the resulting map should present relevant details; 2) the model should at least follow the forward and angular roll motion; 3) the model should be immune to other types of motion that it cannot handle properly, and after such motion ends it must correctly proceed with video preprocessing; and 4) the time required for video preprocessing should be acceptable from a practical point of view. We expect the processing of 50000 frames would not exceed the typical time of interpretation (less then 144 ms/frame).

In our experiment we use an artificial video (Fig. 5), created with 3D graphics rendering software. The video demonstrates constant speed forward motion (0.19 mm/s) and the rotation in roll with a constant angular speed (7.2°/s).

The parameters set during the experiment were constant except for N and ξ. The maximum number of iterations per frame was set to $I = 14$. Tension coefficients for the first 10 iterations were set to $\rho = 1$ and $\zeta = 0.05$, with neighborhood including all the nodes ($n = 64$); then these parameters were reduced to $\rho = 0.5$ and $\zeta = 0.01$, with $n = 2$. The motion estimation methods were tested with a range of ξ coefficients and several image block sizes ($N = 1$, 3, 5, 7 and 9). Parameter ξ during last four iterations was always reduced by 50% with

respect to its value during first 10 iterations. For each of the methods tested, we selected a range of ξ parameters, for which the method produced best results. Selected results produced by the MDR are shown in Fig. 6, Fig. 7, Fig. 8 and Table 1 (computation time per frame).

The results obtained with *MI* were produced for images for which the number of bits per pixel and per color component was reduced from 8 (256 levels of color component) to 3 (8 levels). The reduction was necessary to decrease computation time, which for 24 bits per pixel images reached 5 minutes per frame. We did not notice any significant deterioration of MDR performance caused by the reduction.

The plots of the estimated velocity V_P are presented in Fig. 7. The plots of the estimated angular speed ω of roll rotation are presented in Fig. 8. In all the tested methods there exists a pair of N and ξ parameters for which the MDR correctly computes both the linear and the angular speeds. In cases of *MAD*, *NCC* and *MSE* a range of N and ξ parameters, for which the estimated speeds are close to the expected ones, can be observed.

In case of *MAD* the results close to the expected ones (error less then 5%) are obtained for all the tested block sizes and for ξ parameter in range of 0.03 and 0.2 (6 times variability). A plateau of the *MAD* function can be observed within the range. In other cases (*NCC*, *MSE* and *MI*) ranges of acceptable N and ξ parameters are smaller. Moreover, the computation time required to process video is the smallest in case of *MAD* (Table 1).

For further experiments we selected the following method and parameters: *MAD*, $N = 3$, two phases matching process with $n = 64$ (all MDR nodes within the neighborhood), $\xi = 0.1$, $\rho = 1$ and $\zeta = 0.05$ for first 10 iterations, and $n = 2$, $\xi = 0.05$, $\rho = 0.5$ and $\zeta = 0.01$ for the following 4 iterations of the matching process.

The MDR performance was then tested on selected fragments of natural WCE video and on artificially generated videos demonstrating forward motion, the rotation in roll, as well as rotation in pitch and yaw. It was confirmed the MDR is not able to track rotation in pitch and yaw. However, the model was found immune to such types of motion, and after such motion ended it correctly proceeded with video preprocessing.

3.2. Motility patterns

The motility of the small intestine muscles includes two categories of contractions [42]. One category is peristalsis, which pushes the intestine content in one direction. The second category – segmentary contractions – is intended for mixing the content. The contraction phases are separated by the muscles repose – the refraction phases. Therefore, the motion of the capsule presented by the video data is irregular, characterized by quick changes of image content during the contractions and very slow changes during the refraction phases. From the clinician's perspective the fragments which present quick changes of image content are of special interest, whereas motionless fragments are usually fast-forwarded.

The MDR computed motion descriptors can be used for detection of various motility phases of a small intestine. To verify this thesis, three categories of video motion were defined: 1) motionless – capsule stopped due to a stricture or refraction, 2) smooth motion – capsule is slowly pushed by peristalsis and 3) rapid motion – capsule is turned and tossed by the strong contractions (Fig. 9). Then, for each the category, the representative video frames were selected – over 100 frames per category – chosen from 50 various video recordings.

The video recordings were processed by the MDR. The category labeled vectors of $|V_P|$, $|\omega|$, G_1, G_2 and *MAD* were stored for the selected frames. (The *MAD* is computed here at the final iteration of the matching process and averaged for all the MDR nodes.) Fig. 9 shows example maps produced from videos presenting three different capsule motion categories. Above the maps, plots of velocity estimation $|V_P|$ and the deformation coefficient G_1 are presented. Fig. 10 (a) presents distribution of vectors in $|V_P|$, G_1, *MAD* subspace. Each sample is represented by a digit symbol. The digits indicate the sample membership to an appropriate motion class. Samples, which belong to different categories, occupy different areas of the subspace. Therefore, the location of the vector within the subspace corresponds to the motion category.

The linear discriminant analysis [43, 44] was performed to determine ability of the stored vectors to separate the categories of motion. The distribution of vectors in space of two most discriminating features (MDF) is presented in Fig. 10 (b). The linear separability coefficient [44] is higher then 0.9 and the Fisher's separation coefficient equals 3.4. Therefore, the stored vectors are able to separate correctly the motion categories in over 90% of cases.

Next, a formula for controlling the video frame-rate was devised. It was assumed that motionless video frames should be presented for $^1/_{75}$ of a second (which is a typical refreshing time of computer monitors). For smooth motion video, the frame presentation time was accepted to be $^1/_{10}$ of a second. The presentation time for frames exhibiting rapid changes in image content was chosen to be $^1/_2$ of a second which corresponds with original frame rate of the capsule. The linear regression analysis was performed to find the best linear combination of the parameters to control frame presentation time. The combinantion found is:

$$T = 0.0135\ MAD + 0.1332\ |V_P| + 0.0037\ |\omega| - 0.3871\ G_1 + 0.0007\ G_2 - 0.0315$$

The frame presentation time in our video player is proportional to the T factor. The ratio between the presentation time and the T factor may be adjusted by the clinician during interpretation. In addition, the frame presentation time is limited from the top by half a second, which is long enough to pause video if required, and from the bottom by the actual computer monitor refreshing time.

3.3. Map assessment

The MDR was tested extensively with over 60 WCE videos. The time required to preprocess a 50,000 frame video is roughly 30 minutes with an Intel Pentium IV 3 GHz processor. The maps used for evaluation were 256 pixels wide (sampled at coordinates of 128 nodes of the outer ring and in between the adjacent nodes) and 15000-25000 pixels high.

Certain characteristics that indicate areas of bleeding, ulceration and obscuring froth can be recognized within maps (Fig. 11). Therefore, the map can be useful for rapid identification of such abnormal areas designated for further detailed investigation. Moreover, low values of velocity plot indicate areas of capsule retention (e.g., due to structural abnormalities like strictures or functional abnormalities like achalasia, gastroparesis, dysmotility or ileus). Also, the map shows video fragments useless for investigation due to froth obscuring the view. Therefore, use of the map should increase the number of abnormalities discovered in time.

To verify the above we designed the following experiment. Two gastroenterology experts were recruited to interpret ten WCE videos. The experts did not see the videos before the experiment. The ten videos were preprocessed with the MDR to produce the MDR map and velocity plot. Further, for comparison, the videos were preprocessed to produce a color bar image with number of rows equal to the number of video frames, each row of constant color which is an average of the respective frame content. Similar color bars are available in commercially available software for WCE video interpretation.

The 10 videos were split randomly into two groups, 5 videos each. Expert no. 1 (Parupudi Sriram, MD) investigated the first group of videos with the aid of MDR maps and the other group using color bar only. Expert no. 2 (Rajesh Gupta, MD) investigated the first group with the color bar and the second with the MDR map. The time of investigation was measured in every case. The so-called findings list, with a location and short description of abnormality discovered, was stored. All the findings were reviewed. It was acknowledged, all the findings (lesions or landmarks) were relevant. Table 2 depicts the results obtained during the experiment.

The overall time spent by two experts on interpretation of the videos was 360 minutes in case the videos were supplemented with MDR maps and 367 minutes without maps. The number of findings (gastro-intestinal pathologies and landmarks) discovered by the two experts was 181 with an aid of MDR map and 147 without the map.

The results show there is a 2% reduction in time of interpretation and the number of discovered pathologies and gastrointestinal landmarks is over 20% higher in case the MDR map was applied. We believe that once the doctors get used to the system the interpretation time would also reduce considerably.

4. Conclusions

The model of deformable rings (MDR) is a new technique for aiding in the interpretation of capsule endoscope video. The model processes the video to characterize capsule's motion and to produce a map – a rough representation of gastrointestinal system's endo-luminal surface. The novelty of the model lies in a shape of its mesh, designed particularly to follow specific type of WCE motion, in a novel way of tension computation, and in the concept of transforming a video into the map.

The maps enable a clinician to quickly get an overview of the entire recording in terms of completeness of the procedure (whether the capsule endoscope traversed the entire GI tract) and adequacy of the procedure in terms of the recording quality (absence of obscuring froth or luminal content including gross bleeding). It facilitates the identification of abnormal areas and focuses the efforts on interpreting the mucosal details of the selected areas before examination of the entire recording.

The pace at which rows of the map are assembled depends on the amount of motion detected. The map fragments are suppressed at motionless video and expanded if rapid motion is detected. Therefore, the map's area corresponds with importance of represented information.

The MDR motion tracking ability is limited. It tracks motion when the capsule moves smoothly forward or backward. When it moves sideways, is jumpy or flips around, the produced map may not be an accurate depiction of the GI tract. Such movements cause discontinuities within the map. However, the map discontinuities can be spotted, to indicate video fragments for thorough examination.

It was demonstrated that motion descriptors produced by the model can be used to categorize the WCE capsule's motion patterns. Also, the motion descriptors facilitate automatic control of video frame-rate during the video playback. Therefore the video playback is automatically slowed-down at fragments which present quick changes of image content and are of special interest, whereas motionless fragments are fast-forwarded. Moreover, the motion descriptors can be useful in detection of capsule retention areas.

The experiment, which compared results of video interpretation with and without use of the map, demonstrated that the use of the MDR map can reduce errors in the number of overlooked abnormalities. The experimental results show the number of pathologies and gastrointestinal landmarks found is over 20% higher in case the MDR map was applied.

Finding more lesions is very important. When one finds lesions like angiomas or angioectasia, the number of lesions detected has impact on the treatment options. When only a few lesions are found one would contemplate another endoscopic procedure such as push enteroscopy or double-balloon enteroscopy to treat the lesions rather than operate. If a number of such lesions are localized in a segment of the small intestine, one may prefer to treat by surgical resection of that segment of the intestine. In the context of another condition such as Crohn's disease, finding more lesions in different segments of the intestine connotes widespread involvement. Thus finding more lesions has a bearing on the overall interpretation of the extent of involvement of the gastrointestinal tract.

Additionally, there was a slight reduction in time required for the interpretation. This is a promising result, which demonstrates, the MDR map focuses investigator's attention on essential video fragments that could be mistakenly ignored during the standard interpretation procedure.

Acknowledgments

Authors would like to thank Dr. Rajesh Gupta from the Asian Institute of Gastroenterology, Hyderabad, India for his help in assessment of the MDR technique.

This work was supported by the U.S. Department of Commerce's National Institute of Standards and Technology. Any mention of organizations, agencies, vendors or commercial products in this document is for illustration only. It does not imply sponsorship, contract, recommendation or endorsement by NIST.

References

[1] G. Iddan, G. Meron, A. Glukhowsky, P. Swain, Wireless capsule endoscopy, Nature 405 (6785) (2000), 417-418..

[2] D. G. Adler, and C. J. Gostout, Wireless capsule endoscopy, Hospital Physician, 5 (2003) 16-22

[3] P. Swain, Wireless capsule cndoscopy, GUT, 52 (2003) 48-50.

[4] P. Swain, A. Fritscher-Ravens, Role of video endoscopy in managing small bowel disease, GUT, 53 (2004) 1866-1875.

[5] L.B. Gerson, J. Van Dam, Wireless capsule endoscopy and double-balloon enteroscopy for the diagnosis of obscure gastrointestinal bleeding, Techniques in Vascular and Interventional Radiology, 7 (3) (2004) 130-135.

[6] H. Yamamoto, H. Kita, Enteroscopy – Review, J. of Gastroenterology, Springer-Verlag, 40 (2005) 555-562.

[7] E. Scapa, H. Jacob, S. Lewkowicz, M. Migdal, D. Gat, A. Glukhovski, N. Gutmann, Z. Fireman, Initial experience of wireless-capsule endoscopy for evaluating occult gastrointestinal bleeding and suspected small bowel pathology, The American Journal of Gastroenterology, 97 (2002) 2776-2779.

[8] J. Kwon, S. Park, B. Kim, J.-O. Park, Bio-material property measurement system for locomotive mechanism in gastrointestinal tract, Proc. IEEE International Conference on robotics and Automation, Barcelona, Spain, April 2005, 1303-1308.

[9] X. Wang, M. Q.-H. Meng, Y. Chan, Physiological factors of the small intestine in design of active capsule endoscopy, Proc. IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, September 2005, 2942-2945

[10] T. Nagoka, A. Uchiyama, Development of small wireless position and bleeding detection sensor, Proc. 3rd Annual International IEEE EMBS Special Topic, Conference on Microtechnologies in Medicine and Bilology, Kahuku, Oahu, Hawaii, May, 2005, 130-131

[11] C. Hu, M. Q. -H. Meng, M. Mandal, Efficient magnetic localization and orientation technique for capsule endoscopy, International Journal of Information Acquisition, 2 (1) (2005) 23-36.

[12] M. Coimbra, P. Campos, J.P.S. Cunha, Extracting clinical infromation from endoscopic capsule exams Using MPEG-7 Visual Descriptors, EWIMT 2005, 105-110.

[13] M. Coimbra, P. Campos, J. P. Silva Cunha, Topographic segmentation and transit times estimation for endoscopic capsule exams, Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France, May 2006, (2) 1164-1167.

[14] M. Coimbra, J. Kustra, P. Campos, J.P. Silva Cunha, Combining color with spatial and temporal position of the endoscopic capsule for improved topographic classification and segmentation, Proc. of SAMT 2006, Athens, Greece, 17-18.

[15] M. Coimbra, J.P. Silva Cunha, MPEG-7 visual descriptors - contributions for automated feature extraction in capsule endoscopy, IEEE Trans. Circuits and Systems for Video Technology, 16 (5) (2006) 628-637.

[16] M. Mackiewicz, J. Berens, M. Fisher, G. D. Bell, Colour and texture based gastrointestinal tissue discrimination, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP, Toulouse, France, May 2006, (2) 597-600.

[17] M. Mackiewicz, J. Berens, M. Fisher, Wireless capsule endoscopy video segmentation using support vector classifiers and hidden Markov models, Proc. of Medical Image Understanding and Analyses Conference, Manchester, UK, July 2006.

[18] F. Vilarinao, L. I. Kuncheva, P. Radeva, ROC curves and video analysis optimization in intestinal capsule endoscopy, Pattern Recognition Letters, Special Issue on ROC Analysis (2005) 875-881.

[19] K. Mori, D. Deguchi, J. Hasegawa, Y. Suenaga, J. Toriwaki, H. Takabatake, H. Natori, A method for tracking the camera motion of real endoscope by epipolar geometry analysis and virtual endoscopy system, MICCAI, Proc. 4th International Conference, Utrecht, The Netherlands, October 2001, 14-17.

[20] K. Mori, D. Deguchi, J. Sugiyama, Y. Suenaga, J. Toriwaki, C.R. Maurer, H. Takabatake H. Natori, Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images, Med. Image Anal. 6 (2002) 321-336.

[21] J. Nagao, K. Mori, T. Enjouji, D. Deguchi, T. Kitasaka, Y. Suenaga, J. Hasegawa, J. Toriwaki, H. Takabatake, H. Natori, Fast and accurate bronchoscope tracking using image registration and motion prediction MICCAI, Proc. 7th International Conference Saint-Malo, France, September 2004, 2 551-558

[22] K. Mori, D. Deguchi, T. Kitasaka, Y. Suenaga, H. Takabatake, M. Mori, H. Natori, C.R. Maurer, Bronchoscope tracking based on image registration using multiple initial starting points estimated by motion prediction, MICCAI, Proc. 9th International Conference, Copenhagen, Denmark, October 2006, 2 645-652.

[23] J.P. Helferty, W.E. Higgins, Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance, Proc. International Conference on Image Processing, 2002, 2 961-964.

[24] P. Szczypinski, P. Strumillo, Application of an active contour model for extraction of fuzzy and broken image edges, Machine Graphics and Vision, 5 (4) (1996) 579-594.
[25] P. Szczypinski, A. Materka, Variable-flexibility elastic model for digital image analysis, Bulletin of the Polish Academy of Sciences, Technical Sciences, 47 (3) (1999) 263-269.
[26] P. Szczypinski, A. Materka, Object tracking and recognition using deformable grid with geometrical templates, International Conference On Signals And Electronic Systems 2000, 169-174.
[27] P.M. Szczypinski, P.V.J. Sriram, R.D. Sriram, D. Reddy, Computerized image analysis of wireless capsule endoscopy videos using a dedicated web-like model of deformable rings – A feasibility study, UEGW Sept. 2004, Endoscopy 36 (Suppl. I), A76
[28] P.M. Szczypinski, P.V.J. Sriram, R.D. Sriram, D.N. Reddy, Model of deformable rings for aiding the wireless capsule endoscopy Video Interpretation and Reporting, Computer Vision and Graphics, International Conference, ICCVG 2004, Warsaw, Poland, September 2004, 167-172.
[29] P.M. Szczypinski, P.V.J. Sriram, R.D. Sriram, D.N. Reddy, Preprocessing and aiding in interpretation of wireless capsule endoscope data with model of deformable rings, (in Polish), Zeszyty Naukowe Elektronika, Instytut Elektroniki PŁ, 10 (2005) 129-147.
[30] P.M. Szczypinski, Selecting a motion estimation method for a model of deformable rings, Proc. International Conference on Signals and Electronic Systems, September 2006, Lodz, Poland, 297-300.
[31] P.V. Sriram, G.V. Rao, D.N. Reddy, Wireless capsule endoscopy: experience in a tropical country, J Gastroenterol Hepatol. 19 (1) (2004) 63-67.
[32] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, Int. J. of Computer Vision, 1 (4) (1988) 321-331.
[33] H. Delingette, Adaptive and deformable models based on simplex meshes, IEEE Workshop on Motion of Non-Rigid and Articulated Objects, (1994) 152-157.
[34] W. Neuenschwander, P. Fua, O. Kuebler, From ziplock snakes to velcro surfaces, automatic extraction of man made objects from aerial and space images, Monte Verita, 1995 Birkhaeuser Verlag Basel, 105-114.
[35] T. McInerey, D. Terzopoulos, Deformable models in medical image analysis: A survey, Medical Image Analysis, 1 (2) (1996) 91-108
[36] L. D. Cohen, I. Cohen, Finite-element methods for active contour models and balloons for 2-D and 3-D images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15 (11) (1993) 1131-1147.
[37] Yao Wang, O. Lee, A. Vetro, Use of two-dimensional deformable mesh structures for video coding, Part I and II, IEEE Transactions on Circuits and Systems for Video Technology, 6 (12) (1996) 636-659.
[38] P. van Beek, A. M. Tekalp, N. Zhuang, I. Celasun, M. Xia, Hierarchical 2-D mesh representation, tracking, and compression for object-based video, IEEE Transactions on Circuits and Systems for Video Technology, 9 (3) (1999) 353-369.
[39] R. O. Duda, P. E. Hart, Pattern Classification and Scene Analysis, New York: Wiley, 1973.
[40] R. C. Gonzalez, R. E. Woods, Digital ImageProcessing (third edition), Reading, Massachusetts:Addison-Wesley, 1992.
[41] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, Mutual information based registration of medical images: a survey, IEEE Trans. On Medical Imaging, 22 (8) (2003) 986-1004.
[42] J. Christensen, D.L. Wingate, A Guide to gastrointestinal motility, Bristol John Wright Publisher, 1983.
[43] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1991.
[44] J. Mao, A. Jain, Artificial Neural Networks for Feature Extraction and Multivariate Data Projection, IEEE Trans. Neural Networks, 6 (1995) 296-316.

Tables

Table 1. Average computation time per frame in milliseconds

N	MAD	MSE	MI	NCC
1	18	18	N/A	N/A
3	37	38	132	84
5	75	75	184	214
7	127	132	238	397
9	200	209	327	654

Table 2. Selected results of the WCE video interpretation

	Expert no. 1		Expert no. 2	
Video	**Findings**	**Time [min]**	**Findings**	**Time [min]**
	Color Bar		**MDR Map**	
1	9	54	11	38
2	11	32	12	11
3	18	38	6	18
4	22	55	12	24
5	10	35	11	11
	MDR Map		**Color Bar**	
6	24	44	14	25
7	55	49	25	45
8	14	40	4	19
9	8	65	16	34
10	28	60	18	30

Figures

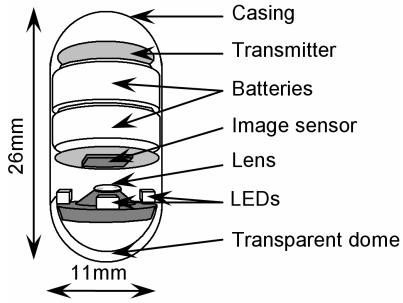

Fig. 1. Wireless capsule endoscope

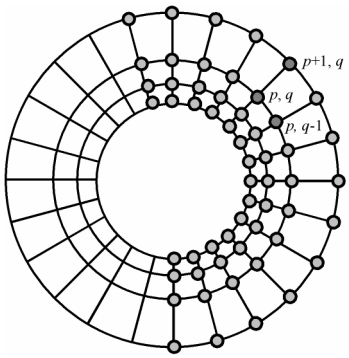

Fig. 2.a

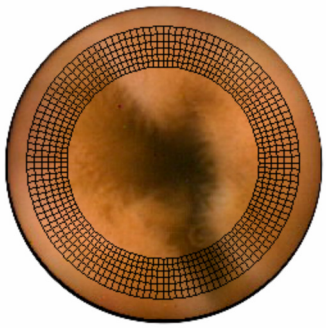

Fig. 2.b

Fig. 2. MDR structure: (a) initial form of MDR structure with node indexing style, and (b) the model superimposed on a WCE video frame.

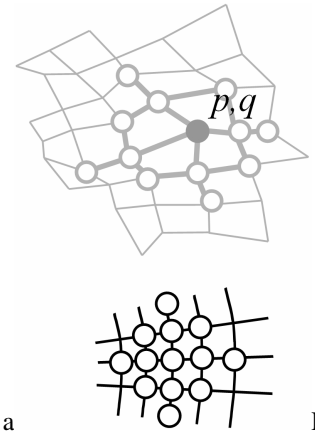

Fig. 3. a

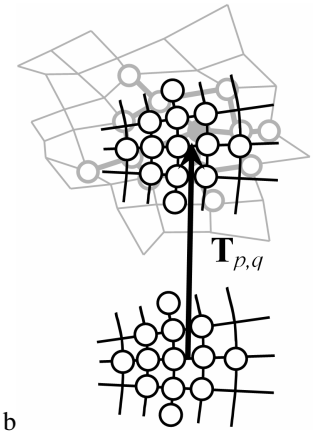

Fig. 3. b

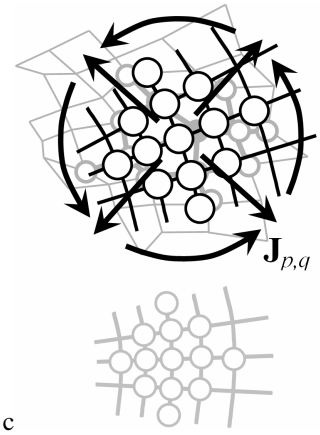

Fig. 3. c

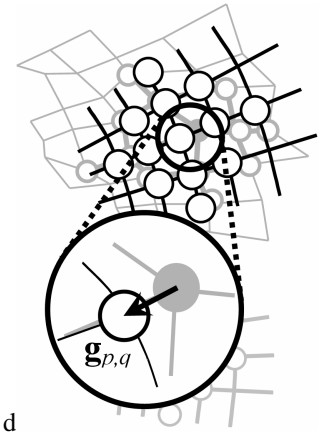

Fig. 3. d

Fig. 3. Tension computation procedure illustration: (a) neighborhoods of a node *p, q* in the MDR and in the reference model, (b, c) transformation of reference model neighborhood, and (c) tension vector.

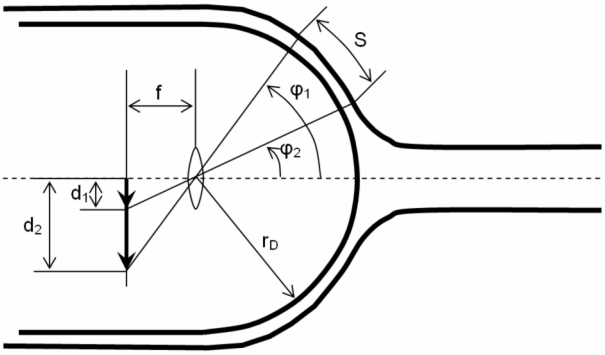

Fig. 4. a

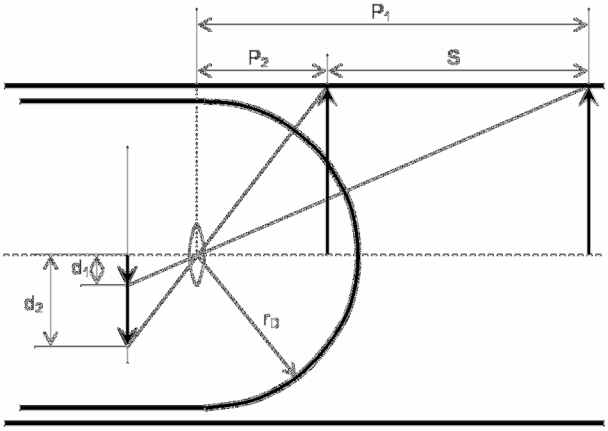

Fig. 4. b

Fig. 4. Image projection: (a) capsule within elastic, body-tight hosepipe and (b) capsule within a rigid cylinder

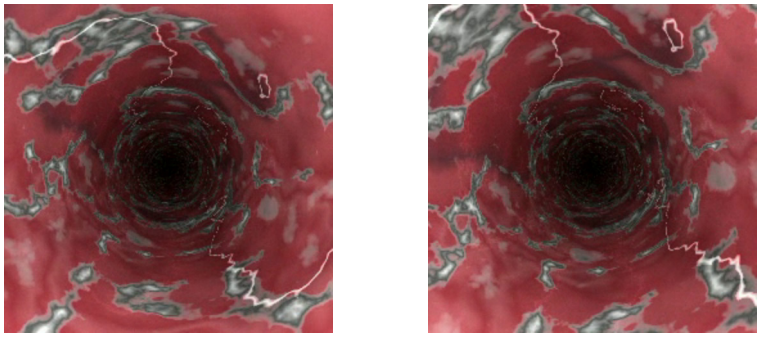

Fig. 5. Frame number 15 and 20 of artificial video showing constant forward motion and rotation

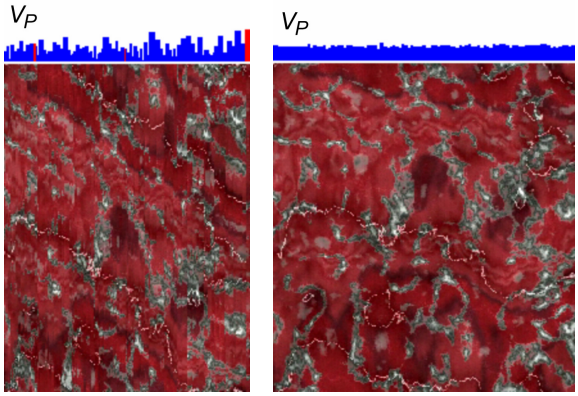

Fig. 6. a

Fig. 6. b

Fig. 6. Resulting maps and estimated velocity plots produced from the same artificial video with: (a) MI N=5 ξ=1.00, and (b) MAD N=5 ξ=0.05. The blue color of the velocity plot depicts forward motion and the red – backward motion.

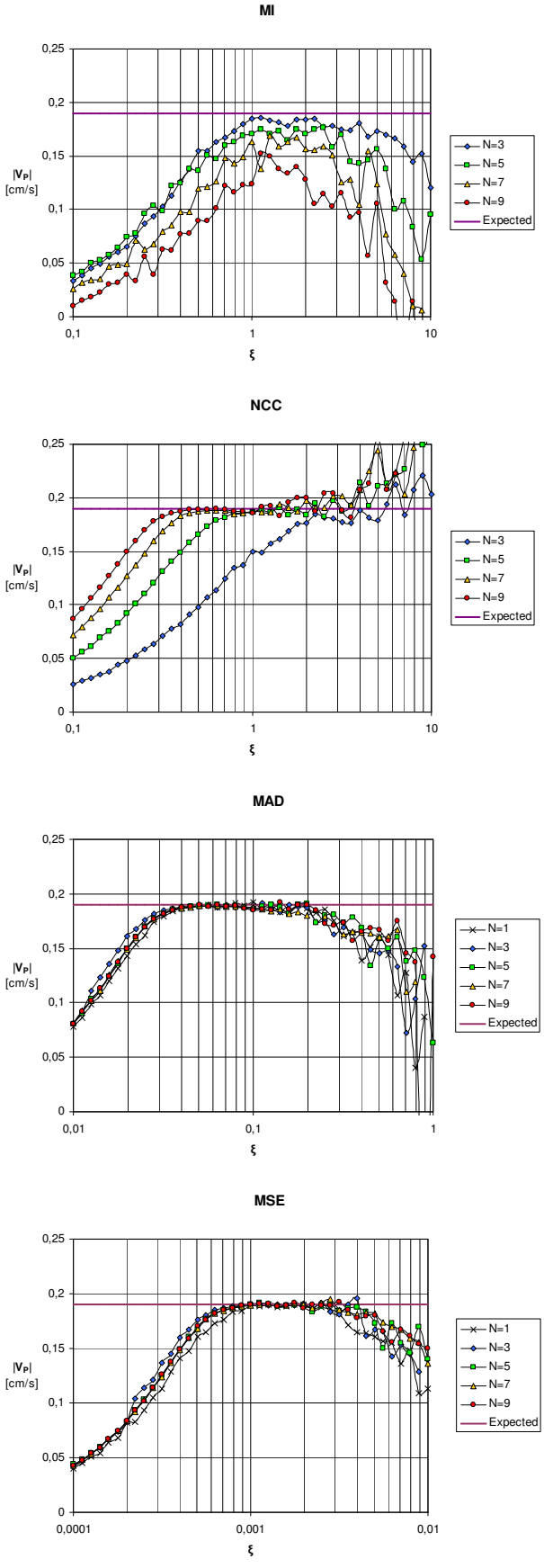

Fig. 7. V_P velocity plots as function of N and ξ parameters.

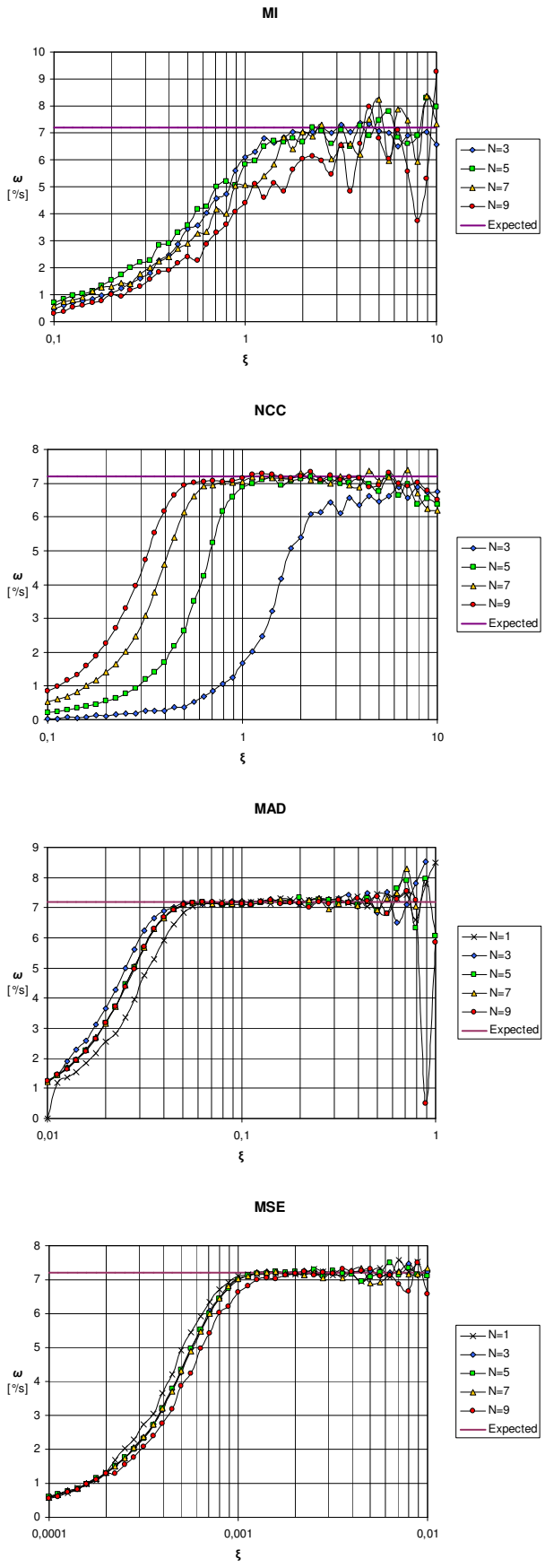

Fig. 8. Angular speed in roll as function of N and ξ

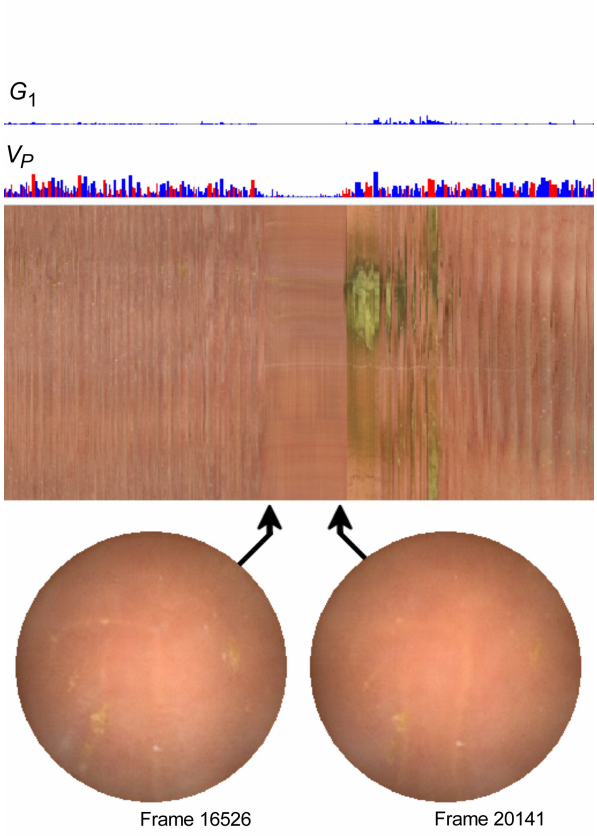

Fig. 9. a

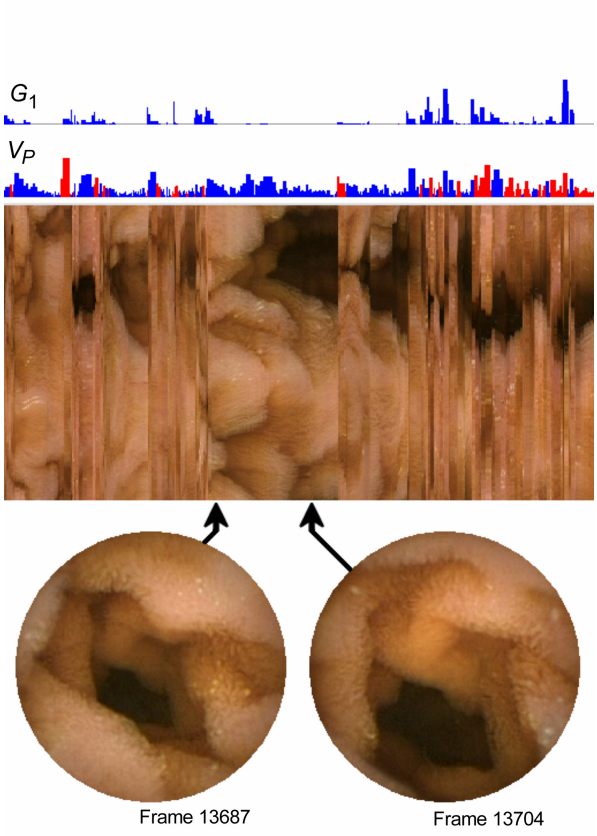

Fig. 9. b

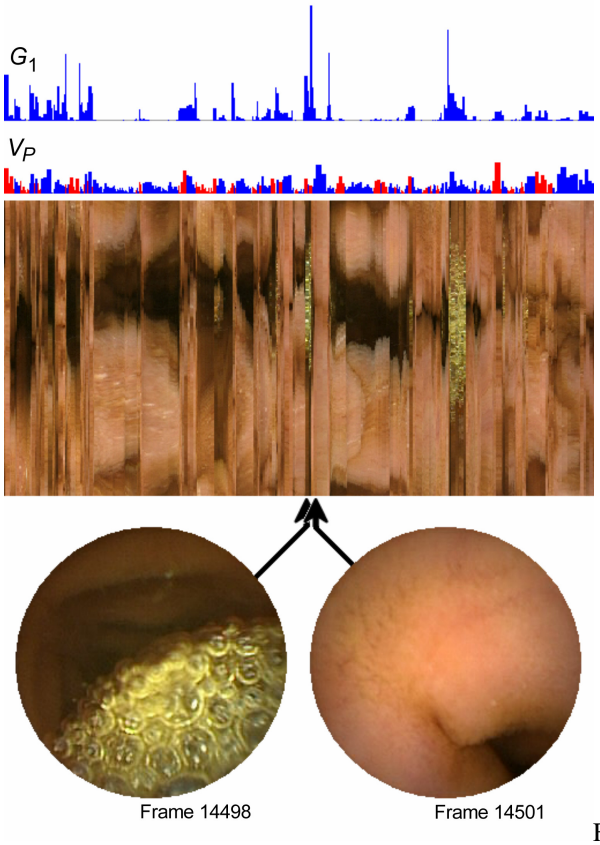

Fig. 9. c

Fig. 9. Examples of map fragments produced by the MDR together with corresponding video frames, velocity plots and deformation plots. Examples indicate: (a) motionless video fragment of refraction phase, (b) fragment demonstrating peristalsis and (c) segmentary contractions. The blue color of the velocity plot depicts forward motion and the red – backward motion.

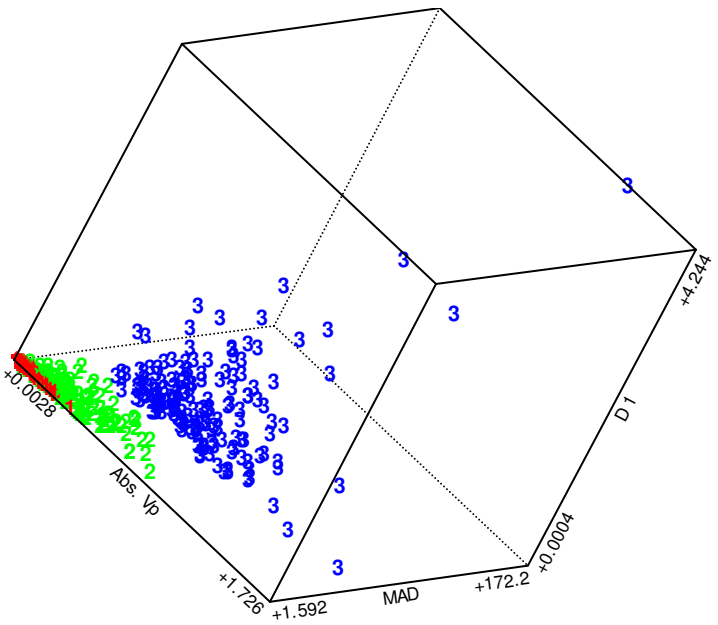

Fig. 10. a

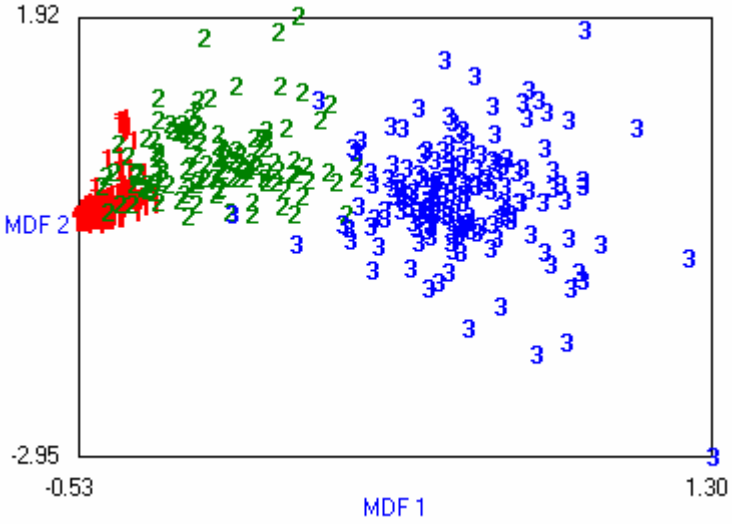

Fig. 10. b

Fig. 10. Distributions of MDR motion descriptor vectors: (a) in $|V_P|$, G_1 and *MAD* subspace, and (b) in transformed LDA space of two MDFs.

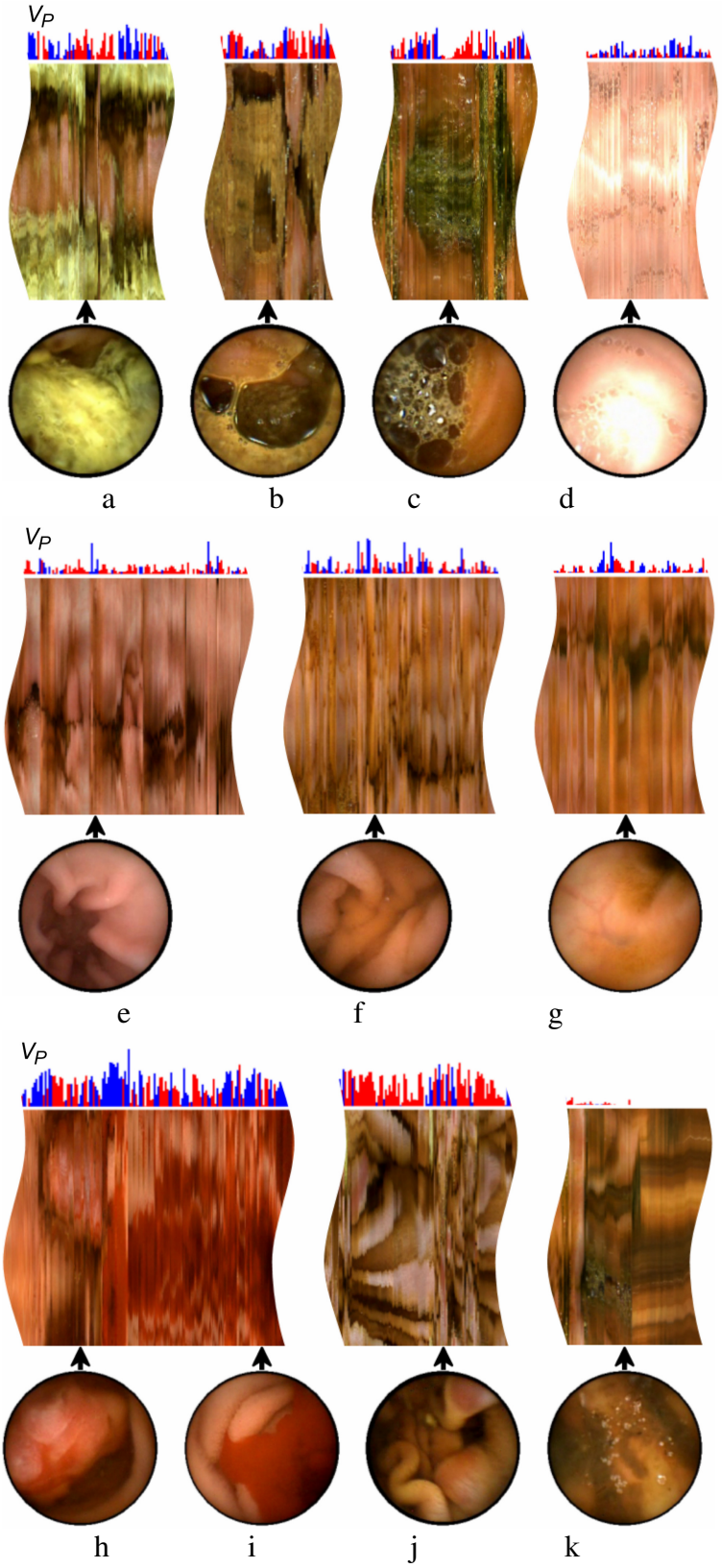

Fig. 11. Selected examples of map fragments (here rotated by 90º) produced by MDR together with corresponding video frames (at the bottom of maps), and velocity plots (above maps, blue represents forward and red backward motion). Examples show: (a) partially digested intestinal content, (b, c) luminal froth and bubbles, (d) overexposed image, (e) distal body antral region of stomach, (f) jejunum, (g) ileum, (h) diffuse ulcer due to Crohn's disease, (i) area of active bleeding, (j) focal hemorrhagic erosion and (k) area of capsule retention with no obvious mucosal pathology.