MODEL OF DEFORMABLE RINGS FOR AIDING THE WIRELESS CAPSULE ENDOSCOPY VIDEO INTERPRETATION AND REPORTING

Piotr M. Szczypinski[1], Parupudi V. J. Sriram[2],
Ram D. Sriram[1], D. Nageshwar Reddy[2]
[1]*National Institute of Standards and Technology, Gaithersburg, USA*
[2]*Asian Institute of Gastroenterology, Hyderabad, India*

Abstract: The wireless capsule endoscopy (WCE) imaging technique provides detailed images of the gastrointestinal (GI) tract, in particular the small intestine, not feasible with earlier techniques. We present a model of deformable rings (MDR) for aiding in the interpretation of WCE videos. The model flexibly matches consecutive video frames with regard to displacement of distinctive portions of the GI tract's tube-like surface. It creates a map, 2D representation of the surface, and estimates relative velocity of a capsule endoscope as it traverses the GI tract. The map can be glanced through for rapid identification of the possible abnormal areas for further detailed endoscopic video investigation. This method significantly reduces the total time spent in interpretation of the WCE videos.

Key words: deformable models, wireless capsule endoscopy, WCE, motion analysis

1. INTRODUCTION

The human small intestine, which measures approximately 6 meters, cannot be visualized using a traditional endoscopic approach. The WCE [6], [9] is a relatively new technique that facilitates the imaging of the small intestine. The WCE system consists of a pill-shaped capsule 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 and passes through the GI tract. Currently, the capsule transmits video images at a rate of two frames per second for approximately 8 hours. The transmitted images are received and recorded by the external receiver-recorder device.

The investigation of video recordings is performed by a trained clinician. It is a tedious task that takes 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 MDR aims at aiding the WCE video interpretation. It preprocesses the WCE video recording to produce a map of the internal surface of the digestive system. The MDR also computes a rough estimate of capsule velocity as it passes through the GI tract. 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 estimate of the capsule velocity provides data sufficient for localization of video sequences that show narrow sections of GI tract, where the capsule stopped or considerably slowed down.

The paper is organized as follows. Section 2 depicts a class of the video data that is processed by the MDR. Section 3 presents the novel structure of the MDR, which includes a unique technique for variable neighborhood tension computation, provides a general explanation of the model implementation and behavior, and introduces the original conception of forming a digestive system map. Finally, Section 4 demonstrates selected examples of GI tract maps and concludes the paper.

2. WCE VIDEO PROPERTIES

The wireless capsule endoscope used in this study produces color images of the internal lumen of the GI tract, covering a circular 140° field of view [9]. Since the shape of the capsule is elongated and the GI tract is akin to a collapsed tube, most of the time the wireless capsule endoscope aligns in a direction parallel to the GI tract, heading the camera lenses forward or backward. Folded walls of digestive tract or intense peristaltic movements may cause the capsule to change its pitch or yaw in relation to the axis of the GI tract. However, such changes do not last long and the capsule eventually repositions to stay parallel with the axis of the GI tract. Therefore, it was assumed that most of the video frames contain images of the GI tract walls, which converge in perspective at the point located near the center of an image (figure 1 a). As the capsule passes through, portions of the tract shift outward or toward the center of an image.

In the current approach to WCE video processing, we follow movements [2]-[4], [8] of the digestive system walls by elastic matching of consecutive video frames. We also estimate the average speed of these movements toward and outward the center of the video frame and collect data on the texture of internal lumen of the digestive system.

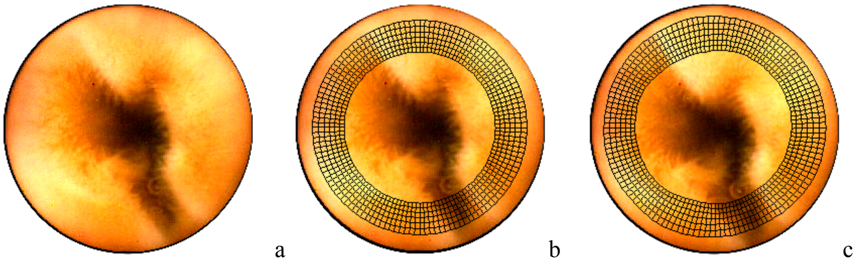

Figure 1. Example of WCE video frames: (a) video frame showing a fragment of small intestine interior, (b) initial form of MDR superimposed on the video frame and (c) frame with MDR after completion of the matching process

3. MDR MATCHING AND MAP ASSEMBLING

The MDR comprises nodes, which are connected to form a mesh. The mesh is positioned in the plane of the video frame. It forms concentric rings surrounding the center of the frame (figure 1.b). Every node of the mesh is referred to by a pair of indexes $p = 1, 2,...P$ and $q = 1, 2,...Q$. The initial location of a node within the image coordinate system is given by the following formula:

$$\begin{bmatrix} \hat{x}_{p,q} & \hat{y}_{p,q} \end{bmatrix}^{\mathrm{T}} = rw^{p-1} \begin{bmatrix} \cos\left(\frac{2\pi q}{Q}\right) & \sin\left(\frac{2\pi q}{Q}\right) \end{bmatrix}^{\mathrm{T}} \quad (1)$$

where r is a radius of inner ring and w is a ratio of radii of adjacent rings.

MDR nodes store information on local image properties, such as RGB color components, that were found at their locations within a preceding video frame. The nodes search within the vicinity of the current frame for locations having similar properties. Each node is pushed toward such a location. On the other hand, the arrangement of nodes within the MDR mesh has to be retained. This requirement is satisfied by modeling tensions within the model structure. The elastic matching is a process of successive displacements of nodes intended for finding a state, in which balance between the two effects is obtained [1].

For a computation of the image influence vector, several locations within the image, in the vicinity of a node (circular area with radius v), are randomly chosen. Image properties at these locations are compared with properties stored within the node by means of an Euclidean distance d in the properties' space. The image influence vector **f** is directed from the node toward the location with the smallest d value and its length is equal to a value of some parameter ξ.

For computation of tension, let us define the n-neighborhood of node p, q. The n-neighborhood is a set of all MDR nodes connected with node p, q by n or less number of lines and it includes the node p, q itself. The averaged transformation of a neighborhood from its initial position given by (1) can be defined by translation vector $\mathbf{T}_{p,\, q}$ and matrix $\mathbf{J}_{p,\, q}$ of scaling, rotation and shear. The $\mathbf{J}_{p,\, q}$ is a square 2x2 matrix, which can be viewed as an average local Jacobian of MDR mesh transformation. Vector $\mathbf{T}_{p,\, q}$ and matrix $\mathbf{J}_{p,\, q}$ are computed in the MDR for minimum mean square error of node positions.

The tension vector for node p, q is computed by means of the following formula:

$$\mathbf{g}_{p,q} = \rho\left(\mathbf{J}_{p,q}\left[\hat{x}_{p,q} \quad \hat{y}_{p,q}\right]^{\mathrm{T}} + \mathbf{T}_{p,q} - \left[x_{p,q} \quad y_{p,q}\right]^{\mathrm{T}}\right) \tag{2}$$

where $[x_{p,q}, y_{p,q}]^{\mathrm{T}}$ is a vector of node p, q actual location and ρ is a tension parameter. Displacement of a node is computed using the following formula:

$$\left[x_{p,q}^{(i+1)} \quad y_{p,q}^{(i+1)}\right]^{\mathrm{T}} = \left[x_{p,q}^{(i)} \quad y_{p,q}^{(i)}\right]^{\mathrm{T}} + \mathbf{g}_{p,q}^{(i)} + \mathbf{f}_{p,q}^{(i)} \tag{3}$$

where an index i refers to discrete time – iteration number.

The process of displacing nodes is repeated until some state of equilibrium is reached, i.e., average displacement distance of MDR nodes drops below some selected threshold value, and until the iteration index (i) is lower than some arbitrary chosen maximum.

It was experimentally found that the process of matching is more efficient if at the start, neighborhood n and vicinity v parameters are relatively large, and then are reduced gradually throughout the process. In this manner, at the beginning of the matching process, the model quickly adjusts its position with regard to some global image changes then, after decreasing n and v parameters, it deforms locally [5], [7] matching local GI tract deformations.

The process of elastic matching is repeated for every frame (m) of the WCE video recording. It must be noted that as the capsule moves forward the MDR expands while matching consecutive video frames. If the capsule moves backward the model shrinks. The average size of MDR in relation to its initial state (1) can be evaluated by determinant of matrix **J** computed for a neighborhood of all the model nodes.

To prevent the model from excessive expanding or from collapsing, a det **J** is computed after completion of every matching. When det $\mathbf{J} > w^2$ the outer ring of a model is erased and a new inner ring is created. If det $\mathbf{J} < w^{-2}$ then the inner ring is erased and a new outer one is added. In either case of rings swapping, RGB vectors sampled at locations of nodes forming the outer ring are arranged in a row of pixels. All such rows are collected during the video processing to form an image or a map of the GI system surface (figure 2). At every swapping, capsule pace is evaluated with the formula:

$$v = \gamma \frac{\sqrt{\det \mathbf{J}_k} - \omega\sqrt{\det \mathbf{J}_{k-1}}}{(m_k - m_{k-1})\omega\sqrt{\det(\mathbf{J}_k \mathbf{J}_{k-1})}} ;\omega = \begin{cases} w & if \quad \det \mathbf{J}_{k-1} < w^{-2} \\ w^{-1} & if \quad \det \mathbf{J}_{k-1} > w^{2} \end{cases} \tag{4}$$

where k is a swapping event index, m is a frame index and γ a parameter.

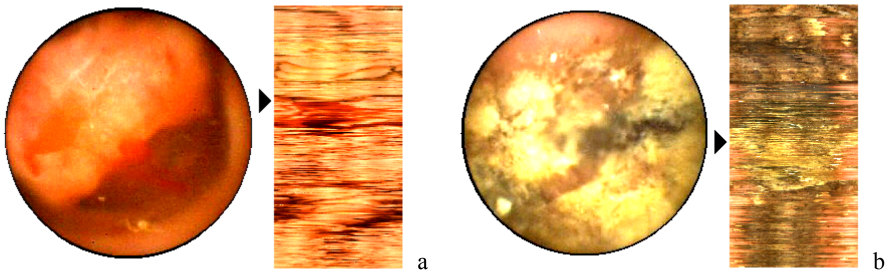

Figure 2. Examples of map fragments produced by MDR with corresponding video frames: (a) area of bleeding and (b) froth content

4. RESULTS AND CONLUSIONS

The MDR was implemented in C++ in DirectShow technology [10] as a video-processing module (filter). It was experimentally established that acceptable results could be obtained with models having $P = 7$ rings, with $Q = 128$ nodes per ring. Coefficients of image influence ξ and tension ρ were set to 1. The matching process is performed in two phases. During the first phase (10 iterations) the transformation matrix $\mathbf{J}$ is computed for all the nodes ($n > 64$); it is assumed that vector $\mathbf{T} = [0, 0]^{\mathrm{T}}$ and search vicinity is of radius $v = 20$. During the second phase (5 iterations), the model deforms locally matching image details with parameters $n = 2$ and $v = 5$. The average time of processing the 8-hour video recording is less then 30 minutes using a PC with an Intel Pentium 4, 1.8 GHz processor.

The database used for MDR testing consisted of over 30 video recordings. Twenty of them were used for preliminary tests and tuning of model parameters. Ten other known recordings consisting of 5 normal and 5 abnormal cases were used for the assessment of MDR utility. It was found that certain characteristics that indicate areas of bleeding, ulceration and obscuring froth could be recognized within maps. Therefore, the maps can be glanced through for quick identification of such abnormal areas, which noticeably reduces examination time. Also, by means of capsule velocity estimation graphs, areas of capsule retention can be detected. These indicate narrow sections of the gastrointestinal tract.

The model deals with images, which are obtained by a smooth forward/backward movement of a capsule. When the capsule moves sideways or jumps, the produced map may be ambiguous. Therefore, our future work will be focused on the development of models capable of tracking various kinds of capsule movements.

REFERENCES

[1] M. Kass, A. Witkin, D. Terzopoulos, *Snakes: Active Contour Models*, Int. J. of Computer Vision, vol. 1, no. 4, 1988, pp. 321-331

[2] H. Delingette, *Adaptive and deformable models based on simplex meshes*, IEEE Workshop on Motion of Non-Rigid and Articulated Objects, 1994, pp.152-157

[3] Yao Wang, O. Lee, A. Vetro, *Use of two-dimensional deformable mesh structures for video coding*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, Dec. 1996, pp. 636-659

[4] P. van Beek, A. M. Tekalp, N. Zhuang, I. Celasun, Minghui Xia, *Hierarchical 2-D mesh representation, tracking, and compression for object-based video*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, March 1999, pp. 353-369

[5] P. Szczypiński, A. Materka, *Variable-Flexibility Elastic Model for Digital Image Analysis*, Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol.47, No.3, 1999, pp. 263-269

[6] G. Iddan, G. Meron, A. Glukhowsky, P. Swain, *Wireless Capsule Endoscopy*, Nature 2000, pp. 405-417

[7] P. Szczypiński, A. Materka, *Object Tracking and Recognition Using Deformable Grid with Geometrical Templates*, ICSES 2000, pp.169-174

[8] W. Badaway, *A structured versus unstructured 2D hierarchical mesh for video object motion tracking*, Canadian Conference on Electrical and Computer Engineering, Vol. 2, May 2001, pp. 953-956

[9] D. G. Adler, C. J. Gostout, *Wireless Capsule Endoscopy*, Hospital Physician, May 2003, pp.16-22

[10] Microsoft Corporation, *DirectShow*, http://msdn.microsoft.com/library