Selecting a Motion Estimation Method for a Model of Deformable Rings

Piotr M. Szczypiński
Institute of Electronic,
Technical University of Łódź,
Wólczańska 223, Łódź 90-924, Poland
e-mail: pms@p.lodz.pl

Abstract — **The Model of Deformable Rings (MDR) was developed to preprocess a Wireless Capsule Endoscopy (WCE) video and aid clinicians with its interpretation. WCE provides a means to obtain a detailed video of a small intestine, not feasible with other endoscopic techniques. The role of MDR is to analyze a motion and extract significant information from the video. One of the important issues of the MDR design is the selection of efficient technique for motion estimation. The earliest MDR implementation involved simple, time effective, yet inaccurate procedure for motion estimation. The goal of the study, presented in here, was to implement other selected methods for motion estimation within the MDR, then test and compare results produced by these methods, to point out the most reliable and efficient one.**

I. Introduction

The WCE system [7] 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. The capsule transmits video 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. There is no mechanism to control the capsule's speed or orientation as it traverses the GI tract. However, since the shape of the capsule is elongated and the GI tract is akin to a collapsed tube, most of the time the capsule aligns in a direction parallel to the GI tract, heading the camera lenses forward or backward.

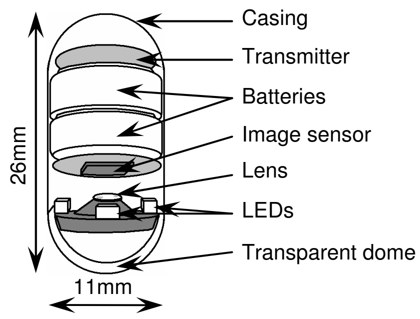

Fig. 1. Wireless capsule endoscope

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. The video frames are color images, 256 × 256 pixels. The field of view, which is about 240 pixels* in diameter (measured in horizontal or vertical direction), covers a circular area in the center of a frame (fig. 2).

The investigation of WCE video 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 or any other abnormal-looking entities.

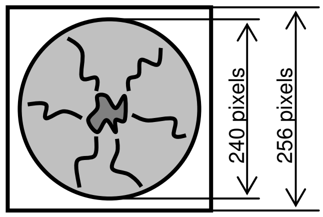

Fig. 2. Video frame outline

II. The Concept of the MDR

The aim of MDR [10] is to track movement of digestive system walls, relative to the capsule's camera, by elastic matching a content of consecutive video frames. Additionally, as a consequence of the movement tracking, the MDR estimates capsules velocity within a GI tract, and also scans the texture of digestive system walls. As a result, it produces a plot of velocity estimate, and an image of inner surface of the GI tract, simply called a GI map. Both may be used as supplementary information during the interpretation process, as a reference to the video data or for rapid identification of significant abnormal areas.

The MDR approach presumes the majority of the capsule motion is either straight forward or straight backward. In this case some distinctive image fragments shift in direction

* Pixel here is a measure of a distance between centers of two horizontally or vertically adjacent image elements

toward or outward the image center. The MDR follows this motion in general and in addition it follows some local fluctuations caused by elastic deformations of GI walls. In other cases our model follows some component of image motion toward or outward the image center and rotation rather then an actual motion.

The MDR as most of other deformable models [1-3, 5, 6] is composed of interconnected nodes, located within an image space. The uniqueness of MDR lies in the arrangement of these nodes which form concentric rings with their centers located in a center of the image frame. Each node of MDR is equipped with a memory for storage of portion of image content.

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 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. Since nodes search the image independently, individual nodes would "go their own way", and as the result the MDR arrangement would adversely change. To preserve the arrangement of MDR the tension within the model structure is modeled. Therefore, on one hand nodes push toward the locations they found to be similar to the stored ones, on the other hand the excessive movement is prevented by tension modeling. The final location of each node is found iteratively, after balance between the two factors is obtained. After that, again image characteristics are stored within nodes' memories and the process repeats. It continues until the last frame of the video sequence.

It may be observed that if the capsule moves forward, the MDR expands while processing consecutive video frames. If the capsule moves backward the model shrinks. To prevent the model from excessive expanding or from shrinking, limits on the model size are set. If the model is too big, then the outer ring of a model is erased and a new inner ring is created. If it is too small, 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 raw of pixels. All such rows collected during the video processing are put together to make an image – the map of GI tract.

III. The Capsule's Velocity Estimation

There are two approaches for estimating the capsule's velocity with use of the MDR. In the firs one, it is assumed that digestive system walls stick to the capsule casing including its transparent dome (fig. 3). The second approach assumes the digestive system is akin to a rigid pipe (fig. 4) 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 image sensor is considered, the fragment appears at a distance $d = f\, tg(\varphi)$ form the center of the sensor. Therefore, the velocity is estimated with the following equation:

$$V_D = \frac{r_D}{\Delta t}\left(atg\left(\frac{d_1}{f}\right) - atg\left(\frac{d_2}{f}\right)\right) \quad (1)$$

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.

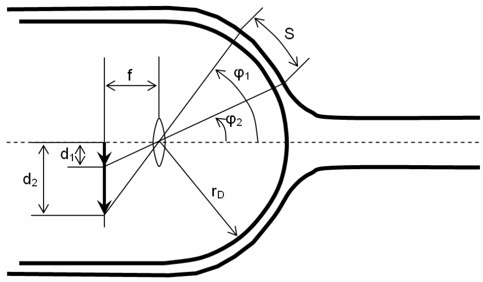

Fig. 3. Image projection case: capsule within an elastic, constricted pipe

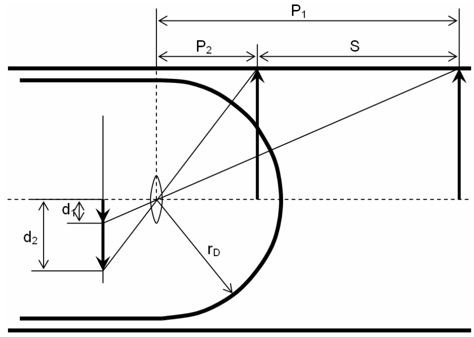

Fig. 4. Image projection case: capsule within a rigid pipe

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 with the following equation:

$$V_P = \frac{r_D f}{\Delta t}\left(\frac{1}{d_2} - \frac{1}{d_1}\right) \quad (2)$$

IV. Motion Estimation Methods

Since the basic task of the MDR is to follow motion of image content, the crucial issue is to select appropriate method for motion estimation. There are a number of methods for motion vector computation [4, 8, 9] that are used for video compression or in motion detection applications. These methods usually compare image content searching for similar regions within two or more consecutive video frames. Therefore, since they search for similar regions, the measure of similarity, or dissimilarity, of regions has to be defined. The most common dissimilarity measures are mean square error (MSE) and mean absolute difference (MAD). The MAD or MSE functions of image region of size $N \times N$ pixels located at (x, y) inside the current frame and a $N \times N$ region located at $(x+\Delta x, y+\Delta y)$ in a previous frame are defined as follows:

$$MAD(\Delta x, \Delta y) = \\ = \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left|I_k(x+\Delta x+m, x+\Delta y+n) - I_{k-1}(x+m, y+n)\right| \quad (3)$$

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

where I_k is a current and $I_{k\text{-}1}$ a previous frame of a video sequence (In the MDR the required $I_{k\text{-}1}$ block is stored within memories the nodes are equipped with.) The minimum value of dissimilarity measure indicates that regions are alike, and the $(-\Delta x, -\Delta y)$ vector is usually considered to be a motion vector at location (x, y).

For the MDR application two different methods employing dissimilarity measures were considered, a full search (FS) and a gradient based (GB) methods. In the FS approach the dissimilarity measure is computed individually for each node, for all the variations of $\Delta x = -M \ldots -1, 0, 1 \ldots M;$ and $\Delta y = -M \ldots -1, 0, 1 \ldots M$. The motion vector pushes the node from its actual location toward the location where the dissimilarity function reaches its minimum value. The magnitude of the vector is a difference between the dissimilarity measure at the two locations.

In the GB approach, a gradient of dissimilarity measure is computed at current locations of MDR nodes. The gradient horizontal component at (x, y) location is a difference of the dissimilarity function at right $(x+1, y)$ and left $(x-1, y)$ adjacent pixels. Vertical component is a difference of the dissimilarity function at top $(x, y+1)$ and bottom $(x, y-1)$ adjacent pixels. The motion vector in this method is equal to a reversed gradient vector of the dissimilarity function.

V. Results and Conclusions

In this study the MDR with 8 rings and 128 nodes per ring has been used. The six variations of motion estimation methods were implemented and examined with the MDR: the FS method employing MAD (with $N = 5$ and $N = 1$), the FS with MSE function ($N = 5$), as well as GB method with MAD function ($N = 5$ and $N = 1$) and MSE function ($N = 5$).

Since there are two contradictory factors affecting the MDR progression, the image motion and internal tension, it is possible to control the model behavior by setting proportions between these factors. Hence, five different settings of motion-tension ratio (ξ) were used per every variation of motion estimation method. The choice of ξ was made to assure different levels of MDR flexibility. Altogether, there were investigated 30 different setups of MDR algorithm.

The MDR was tested on WCE video and on artificially generated videos. The artificial videos demonstrate motion of a strictly specified type, within a rigid pipe, e.g. they represent a constant speed forward motion, the rotation in roll with a constant angular speed, variation in pitch and yaw, and selected combinations of the aforesaid. For all the videos used in this study, the MDR produced acceptable results.

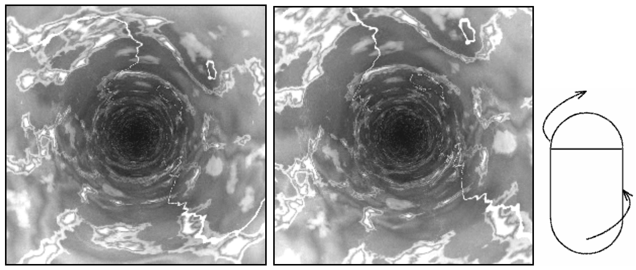

Fig. 5. Frame no. 15 and 20 of artificial video showing constant forward motion and rotation

In table 1, the selected results produced by the MDR are gathered. The results were obtained on an artificial video showing combination of constant forward motion and rotation in roll by 360° (fig. 5). Expected is: a high and constant value of estimated velocity, minimal deformation of the MDR, rotation by 360° and possibly small computation time.

The highest values of velocity where obtained basically for gradient-based methods. This velocity parameter indicates that maps produced are large and most likely the finest. Along with velocity, the ratio of standard deviation of velocity to velocity has been computed. The smaller values of the ratio indicate methods for which the computed velocity is stable.

The rotation angle detected by MDR varies between the methods from 101° (which is a poor result) up to almost 360°. The values close to 360° were achieved with gradient-based methods using both MSE and MAD function, with high values of motion-tension ratio (flexible model). Unfortunately, for these methods, values of the model deformation are pretty high.

If the computation time is a consideration, the fastest are methods with parameter N set to one. In this case only single pixels are compared instead of comparing blocks 5×5 pixels, which results in computation times about 10 times smaller. Despite, the methods are fast, they still produce acceptable results of both the angle and the forward motion tracking.

Although full search methods are regarded as optimal for motion estimation in video coding, the study shows that gradient-based methods fit the MDR better. Also increasing the N parameter, which grossly increases the computation time, does not contribute much to the accuracy being achieved.

TABLE I

SELECTED RESULTS PRODUCED BY THE MDR

	ξ	V_P [mm/s]	$\frac{S(V_P)}{V_P}$	V_D [mm/s]	$\frac{S(V_D)}{V_D}$	Angle [deg.]	Defor-mation	Time[†] [ms]
FS-MAD N=5	0.80	0.161	0.064	0.125	0.064	283.4	0.173	291
	0.40	0.163	0.065	0.127	0.065	284.9	0.140	291
	0.20	0.167	0.056	0.130	0.056	286.7	0.109	304
	0.10	0.164	0.053	0.128	0.053	263.9	0.072	299
	0.05	0.152	0.057	0.119	0.057	216.3	0.036	292
FS-MAD N=1	0.80	0.101	0.116	0.079	0.117	111.5	0.181	25
	0.40	0.102	0.124	0.079	0.124	111.9	0.163	27
	0.20	0.103	0.115	0.080	0.114	115.1	0.134	28
	0.10	0.104	0.106	0.080	0.106	113.2	0.098	28
	0.05	0.102	0.114	0.079	0.114	101.9	0.060	26
FS-MSE N=5	0.80	0.168	0.053	0.131	0.053	283.2	0.114	232
	0.40	0.166	0.056	0.130	0.056	269.2	0.095	232
	0.20	0.160	0.054	0.125	0.054	239.1	0.071	231
	0.10	0.147	0.066	0.114	0.066	193.9	0.044	232
	0.05	0.124	0.087	0.096	0.088	139.3	0.022	231
GB-MAD N=5	0.40	0.177	0.100	0.138	0.101	359.7	0.541	153
	0.20	0.181	0.115	0.141	0.116	354.2	0.153	155
	0.10	0.181	0.102	0.141	0.102	353.9	0.044	155
	0.05	0.184	0.043	0.143	0.043	343.0	0.014	156
	0.02	0.157	0.094	0.123	0.094	178.0	0.003	153
GB-MAD N=1	0.40	0.180	0.178	0.140	0.178	354.6	1.858	22
	0.20	0.181	0.120	0.141	0.120	355.1	0.481	22
	0.10	0.185	0.073	0.144	0.073	355.9	0.124	23
	0.05	0.184	0.060	0.143	0.060	336.5	0.034	22
	0.02	0.155	0.113	0.121	0.114	168.6	0.011	22
GB-MSE N=5	0.80	0.173	0.405	0.135	0.400	359.1	0.761	123
	0.40	0.186	0.131	0.145	0.132	354.1	0.241	123
	0.20	0.188	0.059	0.146	0.059	356.3	0.072	123
	0.10	0.184	0.054	0.144	0.054	336.6	0.019	124
	0.05	0.159	0.119	0.123	0.119	184.5	0.006	123

REFERENCES

[1] M. Kass, A. Witkin, D. Terauzopoulos, "Snakes: Active Contour Models", Int. J. of Computer Vision 1988; 1 (4): 321-331.

[2] H. Delingette. "Adaptive and deformable models based on simplex meshes", IEEE Workshop on Motion of Non-Rigid and Articulated Objects 1994: 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 1996; 6: 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] P. Szczypiński, A. Materka. "Object Tracking and Recognition Using Deformable Grid with Geometrical Templates" ICSES 2000:169-174.

[7] G. Iddan, G. Meron, A. Glukhowsky, P. Swain, "Wireless capsule Endoscopy", Nature 2000; 405-417.

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

[9] A. M. Tourapis, O. C. Au, and M. L. Liou, "Predictive Motion Vector Field Adaptive Search Technique (PMVFAST) - Enhancing Block Based Motion Estimation", Proceedings of Visual Communications and Image Processing 2001 (VCIP-2001), San Jose, CA, January 2001

[10] 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, 2006 Springer: pp.167-172

[†] Average time of processing a single frame on Intel Centrino 1.6 GHz machine