

Linear Radial Patterns Characterization for Automatic Detection of Tonic Intestinal Contractions

Fernando Vilariño¹, Panagiota Spyridonos¹, Jordi Vitrià¹,
Carolina Malagelada², and Petia Radeva¹

¹ Computer Vision Center. Universitat Autònoma de Barcelona,
Bellaterra 08193, Spain

² Hospital Vall d'Hebron, Barcelona 08035, Spain

fernando@cvc.uab.es

<http://www.cvc.uab.es/~fernando>

Abstract. This work tackles the categorization of general linear radial patterns by means of the valleys and ridges detection and the use of descriptors of directional information, which are provided by steerable filters in different regions of the image. We successfully apply our proposal in the specific case of automatic detection of tonic contractions in video capsule endoscopy, which represent a paradigmatic example of linear radial patterns.

1 Introduction

In this work, we propose a general methodology for the achievement of suitable features for the categorization of linear radial patterns. The interest about categorizing radial patterns is underpinned by the fact that a considerable number of points of interest in computer vision are to be points with an intrinsic radial structure -an interesting and deep review about this issue can be found at [7]-. Recent works presented the characterization of radial patterns to tackle the facial feature extraction, finding blobs in images, processing astronomical images, and starfish tracking, among others. Most of these approaches involve the analysis of the gradient image in a constrained neighborhood, and the application of a symmetry transform which codifies the degree of radial symmetry for each pixel of the image.

Our approach is specifically focused towards the characterization of linear radial patterns by means of the analysis of the overall directional information surrounding the centroid of a given linear pattern. We achieve this by means of the division of the image into several quadrants, and by calculating a set of features which carry the directional information for each quadrant. We also propose an alternative method based on the previous polar transform of the linear pattern. On the one hand, our approach is directly addressed to the specific issue of linear radial patterns ranging the whole image, which lays out of the scope of previous methods, which are focused to the localization of points of interest in a constrained neighborhood by using gradient analysis. On the other

hand, this orientation appears specially suitable for the specific issue of the automatic detection of tonic intestinal contractions in capsule video endoscopy for intestinal motility assessment [8].

Video capsule endoscopy [4] is a novel clinical technique based on the analysis of the video images provided by a camera attached to a capsule, which is ingested by the patient, in a frame rate of 2 frames per second. Tonic intestinal contractions are rendered in video capsule endoscopy images as a sequence of frames showing a radial pattern of wrinkles, corresponding to the intestinal folds. The manual annotation of these kind of sequences provides a useful source of information for intestinal motility assessment, but this is tedious and time consuming task, which may span for several hours. The use of the linear radial pattern characterization approach which we present in this work appears to deal in a successful way with the automatic detection of this type of events.

The organization of this paper is as follows: Section 2 introduces the visual paradigm of tonic contractions. Section 3 explains the procedure for the obtention of the linear pattern based on the valleys and ridges detection. Section 4 explains the categorization procedure of the linear radial patterns. Section 5 shows the results of our approach for the automatic detection of tonic contractions. We finish our contribution with our conclusions in Section 6.

2 *Tonic Intestinal Contractions: Visual Paradigms and Video Annotation*

Tonic contractions are produced by muscular tone [5], and can be visualized in capsule endoscopy as a continuous closing of the intestinal lumen with a high variability in length. Thus, the visual pattern of tonic contractions corresponds to a sequence of a closed intestinal lumen in a undefined number of frames. This pattern is highly recognizable for the presence of the characteristic wrinkles which the continuous muscular tone produces when the intestinal walls are folded. Figure 1 shows a characteristic example of a tonic contraction -frames are sequentially deployed from left to right and from top to bottom-

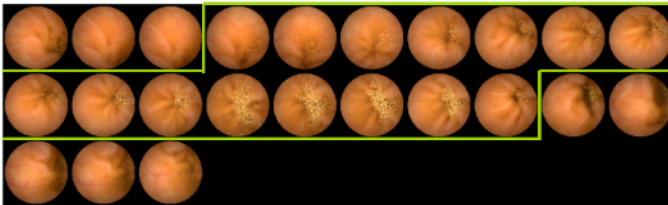


Fig. 1. Sustained contraction (bounded by the green line) spanning 15 frames

The procedure of annotation of tonic contractions in capsule endoscopy consists of the visualization of the whole video by the specialist, and the labelling of the contraction frames. As long as the duration of tonic contractions is not

constrained, and varies in a wide range, the specialist labels the first and last frames of the sequence which contains the contraction. The threshold for the consideration of a motility event as a tonic contraction is set by the specialist in at least 10 frames. This procedure becomes a tedious and time consuming task. For a typical study spanning 2 hours of video, and containing about 40 tonic contractions, a trained specialist may last about 3 hours for an accurate annotation.

3 *Intestinal Wrinkles* Detection Through Valleys and Ridges Analysis

Our hypothesis is that the characteristic pattern of radial wrinkles which tonic contractions show is associated with the valleys and ridges analysis of its intensity image. But before the application of a method for the valleys and ridges detection, endoscopy images must be pre-processed in order to smooth them, which we successfully performed by means of a median filter, with a fixed rectangular window. The size of the median filter window is set to the mean width of wrinkles in sustained contractions, which is set to 6.5 pixels. The application of this filter is justified by the sharpness of the images in areas where a homogeneous view of the intestinal walls is rendered. This is mainly due to the physiological structure of the intestinal walls tissue, and some amount of image noise. Figure 2 (b) renders one example of the results of the median filter smoothing. We tried other sophisticated smoothing methods, such as anisotropic diffusion, which respects the structural information of the image in order to apply the smoothing by keeping sharpness in edges, obtaining similar results but with a dramatic increase in the computation time.

The valley and ridge detection procedure is performed in the following way: we create a filter mask by calculating the second derivative of an anisotropic gaussian kernel [6]. The implemented anisotropic kernel uses $\sigma_1 = 1$ and $\sigma_2 = 2$ for each direction. We obtain 4 different filter responses $F^i(n)$ for an input image I_n as:

$$F^i(n) = I_n * kern_{\alpha_i}, \quad \alpha_i = \frac{i\pi}{4} \quad (1)$$

where, α_i represents 4 different orientations $0^0, 45^0, 90^0$ and 135^0 , $kern_{\alpha_i}$ represents the anisotropic kernel rotated α_i radians, and $*$ represents the convolution operator. The valleys and ridges images F^{val} and F^{rid} are calculated as:

$$\begin{aligned} F^{val}(n) &= \max_{(x,y)} \{F^i(n)\} \\ F^{rid}(n) &= \max_{(x,y)} \{(-1) * F^i(n)\} \end{aligned} \quad (2)$$

where $\max_{(x,y)}$ represents the maximum value of the F^i functions for the (x, y) pixel. Figure 2 (c) shows the valleys image F^{val} for the example image of Figure 2 (a). In the next step of our procedure, we create a binary image by keeping the 75% percentile of $F^{val}(n)$ and $F^{rid}(n)$. Figure 2 (d) shows the binary image created by this procedure for F^{val} . Finally, we apply a morphological skeletonisation [2] in order to obtain the lines with one pixel connectivity which describe

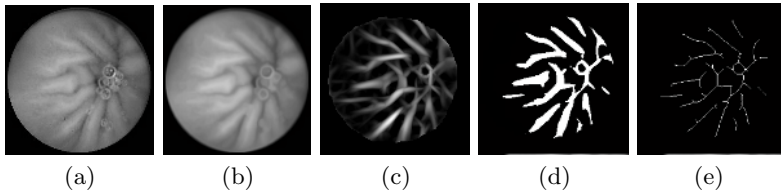


Fig. 2. (a) Original image. (b) Smoothed image. (c) Valley detection. (d) Binary image. (e) Skeletonisation results.

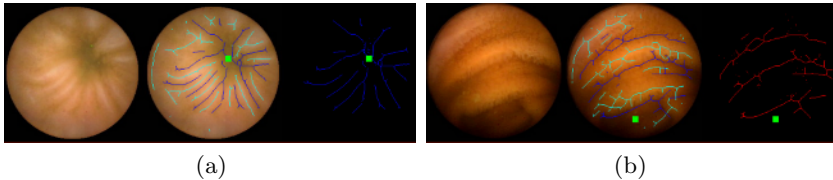


Fig. 3. Original image, wrinkles detected as valleys (blue) and ridges (cyan), and wrinkles pattern for (a) a frame from a tonic contraction and (b) a randomly selected frame

the valleys and the ridges. Figure 2 (e) shows the skeletons created by this procedure. Figure 3 shows the super-imposition of the valleys and wrinkles for two test frames: (a) a frame from a sustained contraction, and (b) a random frame. The green square corresponds to the center of the intestinal lumen. This point can be straightforwardly estimated as the centroid of the binary blob obtained by applying a greater-than-zero threshold on the response image of laplacian of gaussian detector with $\sigma = 6$, as [8] demonstrated. Notice that the centroid of the lumen appears in the middle of the radial wrinkle pattern for the contraction frame. On the contrary, for the random frame the position of the centroid of the lumen does not follow a fixed pattern.

4 Descriptors for Linear Radial Patterns

We propose two different strategies for the linear radial patterns characterization. The first approach is based on 4 descriptors which code the multi-directional information of the wrinkles. The second approach simplifies the number of descriptors to 2, by means of a polar transform of the original data. Although the wrinkle pattern is defined both by valleys and ridges, we used only the valleys pattern as a source of wrinkle information because valleys usually show a higher degree of contrast, a more regular radial organization, and keep a more constant orientation for each wrinkle. The inclusion of ridges is shown not to provide a better performance in the specific case of intestinal contractions.

4.1 Multidirectional Approach

For a given wrinkle pattern, we define 4 different quadrants -see Figure 4 (a)-, which we denote by quadrant 1, 2, 3 and 4, using the centroid of the lumen as the

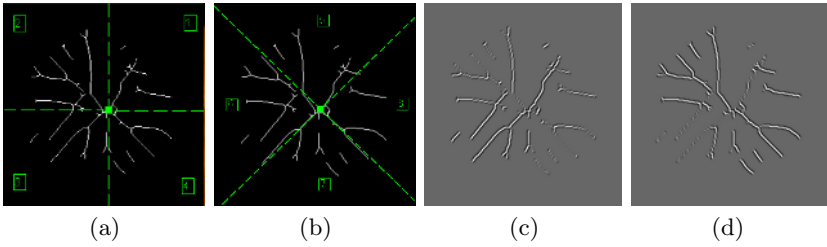


Fig. 4. Quadrant divisions of a frame -(a) and (b)-, and two filter responses for two different orientations: (c) 45°, and (d) 135°

quadrant division middle point. In our preliminary tests, we checked other approaches for the center estimation, such as using as a centroid the center of mass of the wrinkle pattern. However, the results provided by these alternatives were not as satisfactory as those provided by the proposed method. The directional information is obtained by means of 4 second derivative of gaussian steerable filters [1], oriented to 45° and 135°. For both filters, $\sigma = 1$, so they basically operate as line detectors in the direction towards they are oriented. The former output is used to define two descriptors f^1 and f^2 as:

$$\begin{aligned}
 f^1(n) &= G_{1,3}^{45^\circ}(n) - G_{1,3}^{135^\circ}(n) & \left| \right. & f^3(n) = G_{6,8}^{0^\circ}(I_n) - G_{6,8}^{90^\circ}(I_n) \\
 f^2(n) &= G_{2,4}^{135^\circ}(n) - G_{2,4}^{45^\circ}(n) & \left| \right. & f^4(n) = G_{5,7}^{90^\circ}(I_n) - G_{5,7}^{0^\circ}(I_n)
 \end{aligned} \tag{3}$$

where $G_{i,j}^\theta(n)$ represents the sum of the response of the filter with orientation θ over all the pixels of the image I_n in the quadrants i and j . Thus, f^1 and f^2 codify the global amount of directional information in the diagonal radial direction for each quadrant. This same analysis was repeated for a 45° rotated version of the quadrant distribution as shown Figure 4 (b), defining the new quadrants labelled by 5, 6, 7 and 8. This new quadrant distribution provides two more descriptors f^3 and f^4 defined in equation (3), which codify the global amount of directional information in the vertical and horizontal directions for each quadrant. The paradigmatic response of two filters oriented 45° and 135° is rendered in Figure 4 (c) and (d). In order to illustrate the behavior of this set of descriptors, we edited a pool of synthetic images and calculated f^1 , f^2 , f^3 and f^4 . Figure 5 shows the feature values for several test images: the four features tend to express high positive values for the linear radial patterns and high negative values for concentric distributions.

4.2 Polar Transform Approach

Polar transform [3] consists of a mapping from the original cartesian image, in which each pixel is referred to by the pair $(row, column)$, into a transformed image in which each pixel is referred to by a pair $(angle, dist)$. In order to perform a polar transform, we need to fix a center. For each pixel with cartesian coordinates $(row, column)$, the $dist$ value is its Euclidean distance to the center, while the angle value is the angle which the vector connecting the center and the pixel

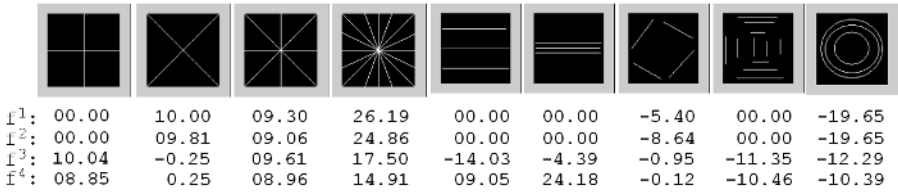


Fig. 5. Directional features for test images

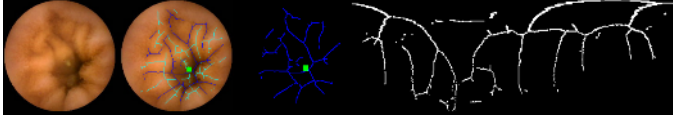


Fig. 6. Wrinkles detected as valleys in polar coordinates

forms with the horizontal axis. In the polar image, the horizontal axis represents the angle parameter, ranging in the interval $[0^0, 360^0]$ from left to right. The vertical axis represents the $dist$ parameter, ranging in the interval $[0, max_{dist}]$ from top to bottom, where max_{dist} corresponds to the maximal distance between two pixels within the camera field of view -which in capsule endoscopy frames corresponds to 240 pixels-. We set the center of the polar transform to be the centroid of the lumen, relating the origin of the transform to the center of the intestinal lumen. Figure 6 shows the result of the polar transform on the wrinkles associated to valleys (blue) for a given frame. Finally, we calculate two descriptors $f^{1'}$ and $f^{2'}$ as follows:

$$\begin{aligned} f^{1'}(n) &= G^{0^0}(I_n^{polar}) \\ f^{2'}(n) &= G^{90^0}(I_n^{polar}) \end{aligned} \quad (4)$$

where $f^{1'}(n)$ and $f^{2'}(n)$ codify the global amount of directional information in the horizontal and vertical direction of the polar image I_n^{polar} of the n frame.

4.3 Definition of the Area of Analysis

As we previously showed, intestinal folds and wrinkles in tonic contractions appear as radial lines in the cartesian image, and as nearly vertical lines in the polar transformed image. However, it must be noticed that this pattern undergoes deformations which are more severe around the the lumen center, and also in distant points from the center -distant parts of the wrinkles usually occur to be curved and no longer respect the radial orientation-. These both regions correspond to the top and bottom areas in the polar plot. To minimize the influence of this phenomenon, we tested the exclusion of the area defined by the lumen detector from the wrinkle analysis. In addition to this, we also excluded all the distant pixels. This exclusion was performed by a simple morphological procedure of dilation and subtraction as defined in Figure 7. The region of analysis

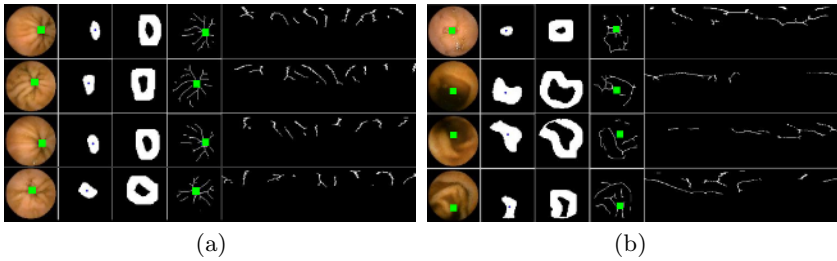


Fig. 7. (a) Tonic contractions. (b) Randomly selected frames.

is defined by a ring-wise or donut mask, which is to be applied to the valleys wrinkle pattern previously to the feature extraction procedure. Figure 7 shows several examples of (a) frames from tonic contractions and (b) random frames, with their corresponding lumen blobs, masks, wrinkles and polar transform of the valleys wrinkles pattern, restricted to the donut-mask.

5 Results

We tested the features explained in last section, defining 4 different sets: 1) *Quadrant*: features f^1 , f^2 , f^3 and f^4 , 2) *Quadrant-donut*: features f^1 , f^2 , f^3 and f^4 , restricted to the mask defined in Figure 7, 3) *Polar*: features $f^{1'}$ and $f^{2'}$, and 4) *Polar-donut*: the same like quadrant-donut, but using the polar features $f^{1'}$ and $f^{2'}$. For the dilatation step of the donut generation, we used a 40 pixels squared structural element. For each feature set, we run 3 different experimental tests in order to assess the performance of our approach in the detection of frames belonging to tonic contractions, and the final detection rate of tonic contractions: Our first experiment consisted of testing the performance of our system with patterns of frames belonging to sustained contraction without any further restriction. The second experiment tried to quantify the number of *frames* belonging to sustained contractions which were labelled as contraction frames -sensitivity over frames belonging to tonic contractions sequences-. Finally, the third experiment provides the global performance, over one whole video, in terms of sensitivity and precision.

For the performance assessment on general patterns of frames belonging to tonic contractions -from now on, we refer this set to as *wrinkle frames*-, the specialists selected a pool of 2,414 wrinkle frames and 2,414 randomly selected frames which did not belong to any tonic contraction -from now on, we refer this set to as *non-wrinkle frames*-. For all our experiments we trained a SVM classifiers with a radial basis function kernel and $\gamma = 0.05$. We used 80% of the samples for training and 20% for testing, performing 10 runs. Figure 8 show the ROC and PR-curves of the classification experiments for the different features sets. Both *quadrant* and *quadrant-donut* feature sets showed similar behaviors in the ROC curve, although the PR-curve analysis showed that quadrant presented optimal performance in precision in the middle region of recall (sensitivity) of wrinkle frames.

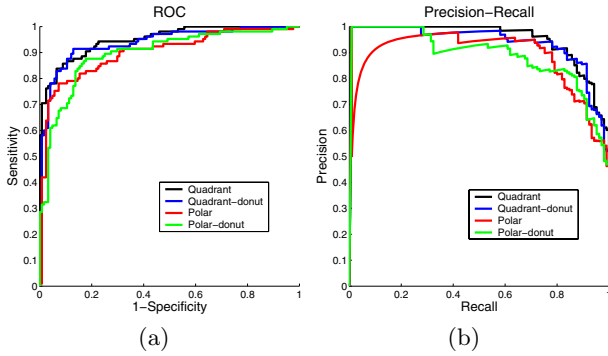


Fig. 8. (a) ROC, and (b) PR curves for the wrinkle frames detection

We used the classifier trained in the last experiment with the *quadrant* feature set in order to label all the frames of the sequences of tonic contractions. The aim of this experiment is to assess the sensitivity of our system for frames of tonic contractions. Table 1 summarizes the results provided by our system. We analyzed 42 sustained contractions with 737 frames. The mean length of the contractions was $14.2(\pm 5.3)$ frames. Our system detected 454 frames, which represents 61.5% of all the frames belonging to tonic contractions. In average, for each contraction sequence, our system labelled 60.9% of frames as frames showing the radial wrinkle pattern.

Table 1. Detection of frames belonging to tonic contractions

Number of tonic contractions:	42	(737 frames)
Number of annotated frames:	454	(61.5%)
Avg. frame annotation by sequence:	60.9%	(± 0.3)

The final step for the validation of our system consists of the automatic labelling of tonic contractions for one video. In order to define the criteria for the detection of a sequence of tonic contraction, we followed the subsequent highlights proposed by the physicians: we consider that we detect a sustained contraction if we detect 5 or more radial wrinkle frames within a window of ± 5 frames. Thus, we define a contraction detection by providing at least one frame holding the previous requirement. Our criteria for the calculus of the system performance consisted on the following lines: 1) We automatically annotate all the wrinkle frames in the video. 2) We create all the sequences of sustained contractions following the criterion described above: all the frames belonging to the same sequence have, at least, 5 wrinkle frames within a ± 5 frames neighborhood. 3) We consider that a sequence labelled by the experts is detected (a true positive) if there exists a sequence provided by the system which has, at least, one frame in common. 4) We consider that a sequence provided by the system is a false positive if none of its frames belong to a tonic contraction labelled by the experts. Following this criteria we obtain the results showed in Table 2:

Table 2. Detection rate of sequences of sustained contractions

Sustained contractions:	42
System sequences:	106
True positives:	30 (71.4% sensitivity)
False positives:	76 (28.3% precision)

Our system successfully detected 71.4% of the sustained contractions provided by the experts. In the final output 1 out of 3 suggested sequences are real tonic contractions. In this sense, the visual validation on the false positives sequences shows the difficulty of the labelling of this kind of contractions by the specialists. The visual patterns of the sequences obtained as false positives match, in many cases, with the paradigm of more than 10 frames of sustained contraction, although in some cases it is really difficult to separate the threshold between a phasic and a sustained for such a sort span of time. For the case of the longest contractions, this is not happening, and the specialist and the system provided sequences which intersect for all the cases without exception. Figure 9 (a) shows a set of representative sequences of sustained contractions detected by the system (true positives). The frames detected as wrinkle frames are surrounded by a green square. Figure 9 (b) shows a set of representative false negatives (missed findings). Finally, Figure 9 (c) shows a mosaic of the output of the system for the test video. The sequences detected by the system are surrounded by a green square, the experts' sequences are surrounded by a blue square, and the coincidences are in cyan.

In true positives the wrinkle frames show the sought radial pattern. Regarding the false negatives, the origin of the misclassifications is twofold: some bubbles in the lumen center hinder the detection of the lumen. For other cases the wrinkle

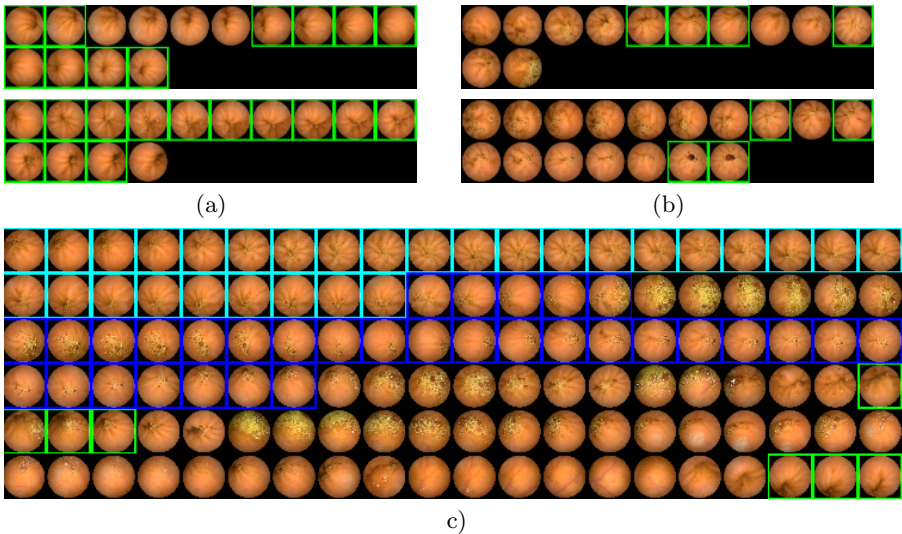


Fig. 9. (a) True positive, (b) false negative sequences of tonic contractions, and (c) a mosaic: system output (green), experts' findings (blue) and coincidences (cyan)

pattern is so weak that no blob was in the binary laplacian image. Finally, it must be noticed that, even with a precision about 30%, the output of our system supposes a valuable tool for the specialists, because they are directly driven to the suggested contractions, and they must not visualize the whole video. For the proposed example of our experiments, assuming the experts to take 10 seconds in the discrimination between a false positive and a true positive of tonic contractions, the total amount of analysis time is reduced to $106 \times 10 = 1060$ seconds, less than 20 minutes. The time consumed by the expert in the labelling of 42 tonic contractions in the analyzed video was reported to be more than 3 hours.

6 Conclusions

In this work, we presented our own general approach for the categorization of linear radial patterns. The obtention of the patterns is performed through a valley and ridge detection algorithm, followed by a morphological processing stage. For the pattern characterization, we propose two different alternatives based on a directional analysis by steerable filters: 1) taking into account the the information of different quadrant divisions of the image, and 2) by means of a polar transform. This general strategy appears to deal well with the specific problem of automatic detection of tonic contractions in capsule endoscopy, providing the experts with a useful tool which achieves a relevant reduction in inspection time.

Acknowledgements

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, H. U. Vall d'Hebron, Barcelona, Spain, as well as the projects FIS-G03/1085, FIS-PI031488, TIC2003-00654 and MI-1509/2005. The technology and methods embraced by this disclosure has been filed for patent protection.

References

1. W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.
2. R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision*, volume I. Addison-Wesley, 1992.
3. A. Hernandez, D. Gil, et al. Anisotropic processing of image structures for adventitia detection in IVUS images. In *Proceedings of Computers in Cardiology*, volume 31, pages 229–232, September 19–22, 2004.
4. G. Iddan, G. Meron, et al. Wireless capsule endoscopy. *Nature*, 405:417, 2000.
5. J.E. Kellow, M. Delvaux, F. Aspiroz, et al. Principles of applied neurogastroenterology: physiology motility-sensation. *Gut*, 45(2):1117–1124, 1999.
6. A. M. Lopez, F. Lumbreras, J. Serrat, and J. J. Villanueva. Evaluation of methods for ridge and valley detection. *IEEE Trans. on PAMI*, 21(4):327–335, 1999.
7. G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):959–973, 2003.
8. F. Vilarino, L. Kuncheva, et al. Roc curves and video analysis optimization in intestinal capsule endoscopy. *Pat. Recog. Let.*, 27(8):875–881, 2006.