

# Eigenmotion-Based Detection of Intestinal Contractions

Laura Igual<sup>1</sup>, Santi Seguí<sup>1</sup>, Jordi Vitrià<sup>1</sup>, Fernando Azpiroz<sup>2</sup>,  
and Petia Radeva<sup>1</sup>

<sup>1</sup> Computer Vision Center and Universidad Autónoma de Barcelona,  
Bellaterra, Spain

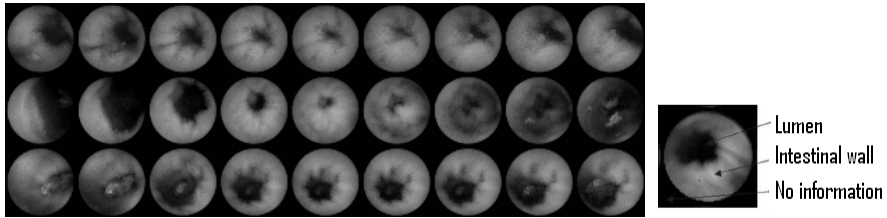
<sup>2</sup> Hospital de Vall d'Hebron, Barcelona, Spain

**Abstract.** Intestinal contractions are one of the main features for analyzing intestinal motility and detecting different gastrointestinal pathologies. In this paper we propose Eigenmotion-based Contraction Detection (ECD), a novel approach for automatic annotation of intestinal contractions of video capsule endoscopy. Our approach extracts the main motion information of a set of contraction sequences in form of eigenmotions using Principal Component Analysis. Then, it uses a selection of them to represent the high dimension motion data. Finally, this contraction characterization is used to classify the contraction sequences by means of machine learning techniques. The experimental results show that motion information is useful in the contraction detection. Moreover, the proposed automatic method is essential to speed up the costly examination of the video capsule endoscopy.

## 1 Introduction

The analysis of the small bowel contractions has been proved to be a meaningful method for diagnosing several intestinal dysfunctions [1], [2]. Currently, the most extended diagnosis test for motility disorders is intestinal manometry [3], which measures variations of pressure. This technique is highly invasive and disagreeable and requires hospitalization of the patient.

A recent and very promising alternative acquisition method is the Wireless Capsule Video Endoscopy (WCVE) [4] which consists of a capsule with a camera, a battery, and a set of led lamps for illumination. The capsule is swallowed by the patient and emits a radio frequency signal stored in an external device. The result is a video which records the "trip" of the capsule along the intestinal tract with a rate of two frames per second. This novel technique is non-invasive compared with the manometry and there is no need of hospitalization of the patient. The resulting images provide a view of the inner gut in which the gut wall and lumen are visualized (Fig. 1 (right)). The visual paradigm of intestine muscle contractions in WCVE is defined as a close-open movement of the lumen. Fig. 1 (left) shows three examples of intestinal contraction sequences. They are defined by 9 frames, since phasic contraction takes on average 4.5 seconds. The non-occlusive contractions (lumen is not completely closed) are hard to detect by classical manometry, since the intestinal walls are not accomplishing enough amount of pressure. However,



**Fig. 1.** Left: three different examples of contraction sequences. Right: example of a frame from capsule video endoscopy with a description of the parts in it.

they are clearly identified in WVCE (Fig. 1 (third row)). Unfortunately, the examination and interpretation of the capsule recordings by the specialists is time consuming - it takes approximately two hours for each study - and very tedious, since the prevalence of contractions in video is very low (1:50 frames). Thus, this procedure becomes not feasible as a clinical routine. Several works have been developed for the automatic detection in WCVE of intestinal affections, such as cancer, bowel Crohn’s disease and small bowel ulcers [5], [6], [7]. Nevertheless, to the author’s knowledge WCVE has not been used to analyze the intestinal motility disorders, except for the work presented in [8].

We propose Eigenmotion-based Contraction Detection (ECD), a new method for the automatic detection of intestinal contractions in WCVE based on the dynamic patterns present in the video. Unlike this method, the approach in [8] only uses static information of 9 frames and does not exploit the motion structure of the capsule recordings. For now on, we will refer to the method developed in [8] as Static-based Contraction Detection (SCD). In further sections we will show that including motion analysis improves drastically the results.

In order to study the motion present in the WCVE, we use a robust method for optical flow estimation [9]. Afterwards, we learn the main motion structures from a particular training set of optical flow fields using Principal Component Analysis (PCA) [10]. Then, we use a selection of the learned eigenmotions to characterize the contraction sequences. This new data representation is applied in the posterior classification of the video sequences as contraction or non-contraction sequences. In previous works, as [11] and [12], the authors propose frameworks for learning generic, expressive optical flow priors which later are used to improve the optical flow computation. We extract main optical flow patterns, but our goal is to use them for classifying.

The paper is organized as follows: in Section 2, we introduce the methodology employed to the automatic detection of intestinal contractions, in Section 3 we present the experimental results, and finally, in Section 4, we expose the conclusions and future work.

## 2 Methodology

The proposed methodology has five steps: 1) *Filtering*, 2) *Motion Estimation*, 3) *Eigenmotion Extraction*, 4) *Eigenmotion Selection* and 5) *Classification*.

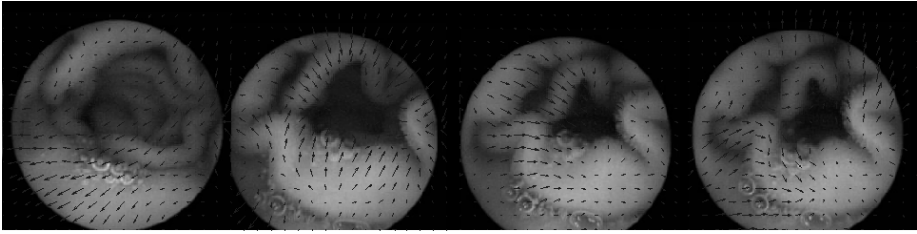
1) *Filtering*: The movement of intestinal contractions in WCVE can be hidden by technical and physiological effects. On one hand, the free movement of the camera into the gut leads to an unsteady orientation. On the other hand, intestinal juices mixed up with the remains of food can also prevent the correct visualization of contraction events. These visual artifacts present in the video make useless a subset of frames in it. The aim of this step is to reject this kind of frames, which are labelled as: *turbid*, *wall* and *tunnel* frames, and represent generally the 30% of the video frames.

Turbid frames are those where intestinal turbid liquid appears. As proposed in [8], we characterize them in terms of color and we classify them. Wall frames are due to the stable orientation of the camera towards the intestinal wall, keeping the intestinal lumen completely out of the field of view. Tunnel frames are the result of a continuous orientation of the camera focusing the intestinal lumen during an undefined period of time in which the gut does not move. Both wall and tunnel frames are described by means of the sum of the lumen area throughout the sequence of 9 frames which is compared with two certain thresholds empirically set with the help of the experts. In order to estimate the area of the lumen in each frame, a Laplacian of Gaussian (LoG) filter is applied [8].

2) *Motion Estimation*: Motion estimation is a key problem in computer vision which consists in computing the apparent motion of the elements present in an image sequence, called *optical flow field*. Many approaches exist to carry out this estimation, but we have chosen the method proposed in [9] for its proved robustness [9], [12]. It consists on a framework based on robust estimation that addresses the problem of improving accuracy of flow estimates in regions containing multiple motions by relaxing the single motion assumption. It estimates the dominant motion accurately, ignoring other existing motions by minimizing a functional with robust data and smoothness terms. In addition, the approach detects where the single motion assumption is violated (i.e., where the error measure is large), and these positions are examined to see if they correspond to a consistent motion.

Before computing the optical flow field of endoscopy video sequences, we try to correct any deviation of the lumen position due to the capsule movement. For that, we simply translate the image in such a way that the centroid of the lumen coincide with the center of the image. Fig. 2 shows the estimated optical flow fields of consecutive frames of a contraction sequence.

3) *Eigenmotion Extraction*: The purpose of this step is to find motion models from the class of motions of intestinal contraction sequences, denoted  $\mathcal{C}$ . This problem has many applications including dimensionality reduction, visualization, exploratory data analysis and pattern recognition. In some cases the models can be derived analytically imposing some hypothesis over data, e. g. the assumption that the addressed motions follow a parametric affine model. However, in many cases, as in the current problem, more complex motions are considered, and PCA can be used to find a basis of optical flow fields with an available training ensemble [10], [12], [13].



**Fig. 2.** Optical flow pairwise computed between the 1th to 5th frames of a contraction sequence. The optical flow fields are superimposed to the 2th to 5th frames.

Let us define the training set of representative samples of the class  $\mathcal{C}$  as a matrix  $X = \{v_i\}_{i \in \{1, \dots, N\}}$ , where  $v_i$  denotes a particular optical flow field, and  $N$  is the number of optical flow fields in the set. Note that, since there are 8 motions in each contraction sequence,  $N = 8c$ , where  $c$  is the number of contraction sequences in the set. Moreover, we only consider the optical flow within a defined region of interest, which contains the lumen of the video endoscopy images. Thus, each vector  $v_i$  has dimension  $2D$ , where  $D$  is the area of the region of interest (15373 pixels). Limiting the optical flow to this region avoids data to be disturbed by possible boundary effects.

We apply PCA to the  $2D \times N$  matrix  $X$  to obtain a decomposition  $X = WH$ , where  $W = \{w_i\}_{i \in \{1, \dots, R\}}$ , is a  $2D \times R$  matrix containing an orthonormal basis of optical flow fields in its columns, and  $H = \{h_i\}_{i \in \{1, \dots, R\}}$ , is a  $R \times R$  matrix. The elements of this basis are the eigenvectors of the covariance matrix of the set of optical flow fields, and we call them *eigenmotions*.

An approximation  $\tilde{v}_j$  of optical flow fields in the class  $\mathcal{C}$  can be obtained as a linear combination of eigenmotions:

$$\tilde{v}_j = \sum_{k=1}^r w_k h_k, j = 1, \dots, R, \quad (1)$$

where  $r \in \{1, \dots, R\}$ . The question now arises is how "better" the new basis of eigenmotions are as data representation and how much we can reduce the data dimensionality. The more extended way to select the best dimension  $r$  for the approximations is evaluating the cumulative energy content of the eigenmotions. Nevertheless, we propose a more intelligent way to select the basis elements which helps us to better understand the data indicating which ones are the important eigenmotions. This subject is argued in the following step.

4) *Eigenmotion Selection:* In this step we deal with the problem of selecting the "best" eigenmotions. Eigenmotions are actually features of the data, thus, this procedure is called feature selection or eigenmotion selection. By removing the most irrelevant and redundant eigenmotions the effect of the curse of dimensionality is alleviated, facilitating the subsequent classification step.

In our case, we apply the method presented in [14] for feature selection. It requires the optimization of a global loss function, which contains one term associated with the empirical loss and a regularization constraint on the parameters  $\mathbf{w}$ , but adding another constraint to also promote sparsity on the distribution of the features. Given  $\mathbf{v} \in \mathbb{R}^d$  a new parameter vector, the global loss is defined as follows:

$$G(\mathbf{w}, \mathbf{v}) = L(X, \mathbf{w}, \mathbf{v}) + \lambda(\|\mathbf{w}\|_1 + \|\sigma_k^d(\mathbf{v})\|_1). \quad (2)$$

$L(X, \mathbf{w}, \mathbf{v})$  is the negated log-likelihood estimator for the parameter set,  $\lambda$  is a positive real value, and the two terms  $\|\mathbf{w}\|_1 + \|\sigma_k^d(\mathbf{v})\|_1$  form the imposed  $L1$  restriction on the parameters  $(\mathbf{w}, \mathbf{v})$ , where  $\sigma_k^d : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is defined as  $\sigma_k^d(a) = (\sigma_k(a_1), \dots, \sigma_k(a_d)), \forall a = (a_1, \dots, a_d) \in \mathbb{R}^d$ ,  $k$  any positive real value, and  $\sigma_k : \mathbb{R} \rightarrow \mathbb{R}$  is the sigmoid function defined as  $\sigma_k(a) = \frac{1}{1 + \exp(-ka)}, \forall a \in \mathbb{R}$ .

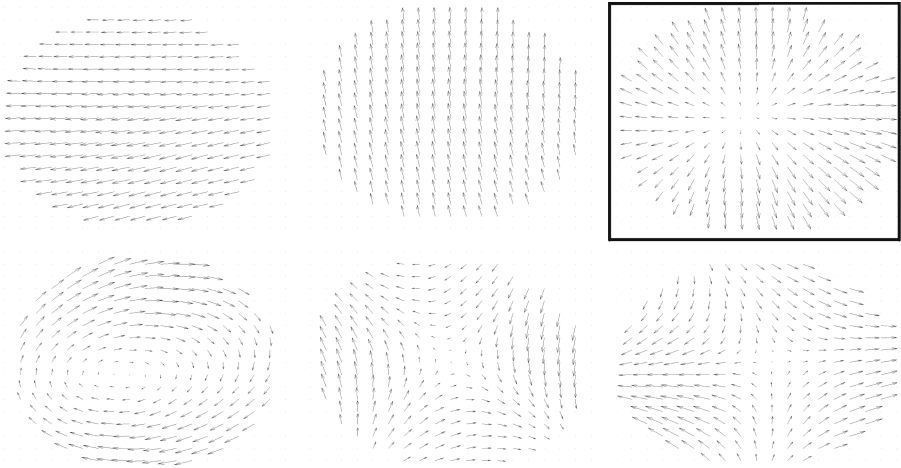
The loss function (2) represents a preference for solutions that use a small set of components from the parameters. The optimization of this functional is carried out using the Boosted Lasso algorithm as explained in [14]. We need for the optimization an appropriate and carefully selected training set where samples are very representative of contractions. As a result of this task we obtain a subset of  $p$  eigenmotions of the basis  $W$ , represented as a  $2D \times p$  matrix,  $\bar{W}$ . The selected subset of eigenmotions of the basis is used to represent optical flow data. A set of  $M$  contraction sequences provide us  $8M$  optical flow fields stored in a  $2D \times 8M$  matrix  $U$ . Then, we transform  $U$  into a  $p \times 8M$  matrix  $Y$  as follows:  $Y = \bar{W}^T U$ . The matrix  $Y$  contains  $M$  data of  $8p$  dimensions. Note that we reduce the dimension of the data from  $16D$  to  $8p$ .

5) *Classification*: Once the optical flow fields are transformed to the eigenmotion-based representation, this new codification is the input of a classifier. This classifier decide whether new video sequences are intestinal contraction sequences or not. In particular, we use the Relevance-Vector-Machine (RVM) classifier [16], since it has desired properties, namely, soft decisions and less complexity (number of kernels). Unlike the previous step, the present training process requires a rather large data set of video sequences, since our goal now is to achieve best generalization error.

### 3 Experimental Results

Our tests were carried out with data provided by the Digestive Diseases Research Group of Hospital "Vall d'Hebron" in Barcelona, Spain, who are currently working with the endoscopy capsule developed by Given Imaging, Ltd., Israel [17]. Videos from healthy subjects and subjects with intestinal dysfunctions (here called patients) were considered.

We first generated a set of intestinal contraction sequences extracted from a healthy subject video endoscopy, with the supervision of the specialists for assuring the representativeness of the selected group. This set was used to extract



**Fig. 3.** The six first eigenmotions. The selected one is pointed by a black box.

the eigenmotions, applying step (2) and (3). In Fig. 3 we show the first six extracted eigenmotions (ordered by decreasing eigenvalues). The first and second ones correspond to translation motions, they are probably due to the global camera motion. The next ones are dilatation, rotation and two shear deformations. The eigenmotion selection performed in step (4), with the parameters set to  $\lambda = 0.5$  and  $k = 10$ , returned only one selected eigenmotion. This one is pointed by a black box in Fig. 3. As can be seen, it represents contraction-dilatation movements. This motion is the one we would expect to find in the paradigm of intestinal contraction motion. Once the selected eigenmotion was available, new motion data were transformed to the eigenmotion-based representation, resulting in data of dimension 8, since  $p = 1$ .

We evaluated the performance results of step (5) using a ten fold cross validation strategy [10] on a larger data set. We considered a pool of 10 different videos from healthy subjects and patients. We first filtered these videos using step (1). The remaining frames were revised by a specialist who indicated the center of the contraction sequences in it. This generated a set of 11030 sequences. These findings were used as gold standard for testing our system. The test set were composed by these contraction sequences and an under-sampled set of non-contraction sequences randomly chosen from the same pool. The results were validated using several measures, described in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), as follows:  $Error = FP + FN$ ,  $Sensitivity = TP / (TP + FN)$ ,  $Specificity = TN / (TN + FP)$ ,  $Precision = TP / (TP + FP)$  and  $False\ Alarm\ Ratio, FAR = FP / (TP + FN)$ .

Table 1 summarizes the results of the Eigenmotion-based Contraction Detection (ECD) using two different data representation. The first one used the 51 first eigenmotions, which returned the 99% of the data variance, and led to data of dimension 408. The second one used the selected eigenmotion in step (4), and the data had dimension 8. As can be seen, the more simple representation gave

**Table 1.** Classification results of our ECD method for two different data representation

	Error	Sensitivity	Specificity	Precision	FAR
<i>ECD</i> (dim= 408)	31.23%	69.29%	68.30%	68.96%	32.04%
<i>ECD</i> (dim= 8)	29.51%	70.30 %	70.68%	70.57%	29.32%

**Table 2.** Classification results of the methods SCD and CCD

	Error	Sensitivity	Specificity	Precision	FAR
<i>SCD</i> (dim= 54)	22.90%	73.55%	80.62%	79.14.%	28.46%
<i>CCD</i> (dim= 62)	18.73%	77.58 %	84.95%	83.75%	24.20%

better results (second row), than the more complex one (first row). This means that the selected eigenmotion represents an excellent discriminant feature.

Finally, in Table 2 (first row), we display the results obtained using the Static-based Contraction Detection (SCD) [8]. This method characterized intestinal contractions with 54 features, instead of 8 as in our case, and the evaluation measures were, in general, better than the ones obtained with ECD. Anyway, we concluded that motion information is important when we studied the results of the *Combined Contraction Detection* (CCD) which used both, the static and dynamic features. In this case, sequences were defined in a 62-dimensional space. We can see that this method led to the best results (second row of Table 2). Moreover, these results represent high specificity and sensitivity rates compared to the rates obtained by the specialists.

## 4 Conclusions and Future Work

Our contribution represents a significant step towards the introduction of the WCVE for intestinal motility diseases diagnosis. Our proposal combine classical and new procedures to achieve a robust and efficient method for automatic detection of intestinal contractions. Our method obtains high specificity and sensitivity rates and can be used to drastically reduce the time of analyzing the video capsule endoscopy.

The presented work can be extended in several ways. There are other methods for dimensionality reduction that could be used here instead of PCA, as for instance Independent Component Analysis (ICA) or Non-negative Matrix Factorization (NMF). These alternative approaches will be contemplated in future works. Moreover, as well as in general contraction detection, the characterization of the motility based on eigenmotions can be useful in other applications, as for instance in the definition, classification and detection of different types of contractions. In this direction, future work imply multiclassification tasks and other sophisticated techniques.

## Acknowledgements

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, and Hospital Universitari "Vall d'Hebron" Barcelona, Spain, as well as the project TIN2006-15308-C02.

## References

1. Kellow, J., Delvaux, M., Aspiroz, F., et al.: Principles of applied neurogastroenterology: physiology motility-sensation. *Gut* 45(2), 1117–1124 (1999)
2. Quigley, E.M.: Gastric and small intestinal motility in health and disease. *Gastroenterology Clinics of North America* 25, 113–145 (1996)
3. Hansen, M.B.: Small intestinal manometry. *Physiological Research* 51, 541–556 (2002)
4. Iddan, G., Meron, G., et al.: Wireless capsule endoscopy. *Nature* 405, 417 (2000)
5. Tjoa, M.P., Krishnan, S.M.: Feature extraction for the analysis of colon status from the endoscopic images. *Biomedical Engineering OnLine* 2, 3–17 (2003)
6. Karkanis, S.A., Iakovidis, D.K., et al.: Computer aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans. on Information Technology in Biomedicine* 7, 141–152 (2003)
7. Zheng, M.M., Krishnan, S.M., et al.: A fusion-based clinical support for disease diagnosis from endoscopic images. *Comp. Biology and Medicine* 35, 259–274 (2005)
8. Vilariño, F., Spyridonos, P., Vitrià, J., Azpiroz, F., Radeva, P.: Cascade analysis for intestinal contraction detection. *CARS*, 9–10 (2006)
9. Black, M.J., Anandan, P.: The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Comput. Vis. Image Underst.* 63, 75–104 (1996)
10. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley-Interscience Publication, New York, NY, USA (2000)
11. Roth, S., Black, M.J.: On the spatial statistics of optical flow (2005)
12. Fleet, D.J., Black, M.J., Yacoob, Y., Jepson, A.D.: Design and use of linear models for image motion analysis. *Int. J. Comput. Vision* 36(3), 171–193 (2000)
13. Skocaj, D., Bischof, H., Leonardis, A.: A robust pca algorithm for building representations from panoramic images. In: Tistarelli, M., Bigun, J., Jain, A.K. (eds.) *ECCV 2002*. LNCS, vol. 2359, pp. 761–775. Springer, Heidelberg (2002)
14. Lapedriza, A., Seguí, S., Masip, D., Vitrià, J.: A sparse bayesian approach for joint feature selection and classifier learning. *Technical Report CVC* (2007)
15. Zhao, P., Yu, B.: Boosted lasso (2004)
16. Tipping, M.: The relevance vector machine. In: *Advances in Neural Information Processing Systems*, Morgan Kaufmann, Seattle, Washington, USA (2000)
17. Given Imaging, L.: (2007) <http://www.givenimaging.com>