Non-negative Matrix Factorization for Endoscopic Video Summarization

Spyros Tsevas[1], Dimitris Iakovidis[1], Dimitris Maroulis[1], Emmanuel Pavlakis[2],
and Andreas Polydorou[2]

[1] Dept. of Informatics and Telecommunications, University of Athens
Panepistimiopolis, GR-15784, Athens, Greece
{s.tsevas,dimitris.iakovidis}@ieee.org, dmarou@di.uoa.gr
[2] Department of Surgery, Aretaieion Hospital, V. Sofias 76 avenue, 115 27 Athens, Greece
egp@otenet.gr, apolyd@med.uoa.gr

Abstract. Wireless Capsule Endoscopy (WCE) has been introduced as a non-invasive colour imaging technique for the inspection of the small intestin along with the rest of the gastrointestinal tract. Each WCE examination results in a 50,000-frames video that has to be visually inspected frame-by-frame by the doctor and this may be a highly time-consuming task even for the experienced gastroenterologist. In this paper we propose a novel approach that leads to a summarized version of the original video enabling significant reduction in the video assessment time without losing any critical information. It is based on symmetric non-negative matrix factorisation initialized by the fuzzy c-means algorithm and it is supported by non-negative Lagrangian relaxation to extract a subset of video frames containing the most representative scenes from an entire examination. The experimental evaluation of the proposed approach was performed using previously annotated endoscopic videos from various sites of the small intestine.

Keywords: Non-negative matrix factorisation, wireless capsule endoscopy, video, summarisation.

1 Introduction

Wireless Capsule Endoscopy (WCE) [1] represents a major departure from conventional endoscopy which is inefficient for the examination of the small intestine and is usually uncomfortable for the patient. By using the WCE technique, the physician can efficiently diagnose a range of gastrointestinal disorders, including ulcer, unexplained bleeding, and polyps. One of the major challenges that WCE imposes is the size of the resulting video which results in a more than an hour of intense labour for the physician in order to examine the whole frame sequence [2] while this manual examination process does not guarantee that some abnormal regions are missed.

Several computational approaches coping with the analysis of the WCE video have been proposed [3-14]; however, to the best of our knowledge, no major contribution has been made to the reduction of the time required for visual inspection of the WCE video. To cope with this issue, we propose an effective computational approach that

J. Darzentas et al. (Eds.): SETN 2008, LNAI 5138, pp. 425–430, 2008.

drastically reduces the video frames to be inspected enabling this way faster inspection of the video sequence. The proposed approach [17] applies a methodology based on non-negative matrix factorisation (NMF) [18, 19] to summarize the WCE video by keeping the most representative scenes from the whole examination.

The rest of this paper consists of three sections. Section II provides a description of the proposed methodology. Section III, presents the results of its experimental application on WCE video data, and Section IV summarises the conclusions that can be derived from this study.

2 Methodology

The proposed approach for WCE video summarisation is based on the data reduction methodology described in [17] and it takes place in three steps. In the first step Fuzzy C-Means (FCM) is applied on the input video stream to group its frames into a predefined number of clusters, whereas in the second and in the third step two NMF algorithms are subsequently applied on the clustered frames so that they extract only some representative video frames from the whole video.

Given a non-negative $m \times n$ matrix **V**, the NMF algorithms seeks to find non-negative factors **W** and **H** of $\overline{\mathbf{V}}$ such that:

$$\mathbf{V} \approx \overline{\mathbf{V}} = \mathbf{W} \times \mathbf{H} \tag{1}$$

where $\mathbf{W} \in \Re^{m \times k}$ and $\mathbf{H} \in \Re^{k \times n}$.

The dimensionality and the initial values of **W** and **H** (or just **H** in certain algorithms) are determined by means of the FCM algorithm. FCM performs soft clustering of the video frames so that they belong to more than a single cluster. The memberships of each frame to the different clusters are stored in a $k \times n$ matrix $\mathbf{U_{FCM}}$.

The neighbouring frames in the original m-dimensional vector space, are determined by calculating the $n \times n$ matrix of the Euclidean distances which is used for the calculation of the geodesic distance matrix $\mathbf{D}_{\mathrm{G}}$ that contains the geodesic distances (shortest paths) between the vectorial representations of the frames. Next, $\mathbf{D}_{\mathrm{G}}$ is transformed into a pairwise similarity matrix according to the exponential weighting scheme in Eq. (2),

$$V = e^{-\frac{D_G}{r}} \tag{2}$$

V is going to be used as an input to the FCM.

The dimension k of $\mathbf{U_{FCM}}$ is set equal to the predefined number of clusters c, whereas **W** and **H** are initialized with the m-dimensional cluster centroids and the values of the membership matrix of the converged FCM, respectively.

The symmetric NMF (SymNMF) which for a square matrix is:

$$\mathbf{V} \approx \mathbf{H} \times \mathbf{H}^{\mathrm{T}} \tag{3}$$

SymNMF is applied on **V** so that it “unfolds” the clusters and makes them more transparent. For the calculation of **H** we followed the iterative approach described in [17, 22] initializing **H** with $\mathbf{U}_{\mathbf{FCM}}$. Iteration takes place until the objective function of the SymNMF converges to a small positive value close to zero.

The final step of the methodology imposes orthogonality constraints on the output of the SymNMF so as to extract the most representative members of a given cluster. It is implemented by means of an NMF multiplicative update iterative algorithm known as Non-negative Lagrangian Relaxation (NLR) [17, 23].

In NLR the entries are viewed as cluster indicators and as a result the interpretation of the results at convergence is straightforward allowing this way a relatively easy interpretation of the cluster structure.

3 Results

In order to illustrate the performance of the proposed summarization approach, a number of experiments were conducted on a controlled dataset comprising of annotated video frames with ground truth information provided by expert endoscopists who visually inspected and annotated each frame. A total of 281,000 WCE frames were obtained with identical imaging settings from different patients and two kinds of abnormal findings were identified; ulcers and bleeding. As each finding was visible in more than a single frame, neighbourhoods of frames were extracted for each finding. This process led to the extraction of a total of eight neighbourhoods of frames with abnormal findings which were further balanced at 40 frames per category by random sampling of the larger set to avoid bias. The final dataset dataset consists of 4 ulcer neighbourhoods that contain 11, 8, 12 and 9 frames and 4 bleeding neighbourhoods that consists of 12, 19, 5 and 4 frames; summing up to 40 and 58 frames of ulcers and bleedings respectively. Aiming to investigate the discrimination between abnormal and normal tissues, a total of 40 frames of normal tissues, was appended. The normal frames were extracted from randomly sampled sites of normal tissues over the whole dataset leading to the formation of a 7 neighbourhoods of normal frames. The resulting dataset consists of 120 frames (n=120).

In order to reduce the computational cost and the detail of each frame the video frames were rescaled from 260×260 pixels to 91×91 pixels (m=8281). Experimentation showed that the use of smaller frames was not beneficial for the overall results.

The images were converted to greyscale and used to form the initial dataset matrix of $m \times n$ dimensions. By following the process described in the previous section we calculated the similarity matrix V according Eq. (2), with r=100 [17], so as to proceed with the FCM calculations. FCM was executed for 3 clusters.

In the following, we subsequently applied SymNMF and NLR to V. It can be observed that still after SymNMF the cluster structure is not clear. After the application of NLR the clusters are not really separated, though NLR enforces orthogonality. This lack of strict orthogonality is due to the fact that the number of iterations of the SymNMF and NLR accordingly are finite. Actually, only a part of the examples are strictly 'orthogonal' to the members of other clusters. These members form the Most Representative Examples (MREs) of the cluster.

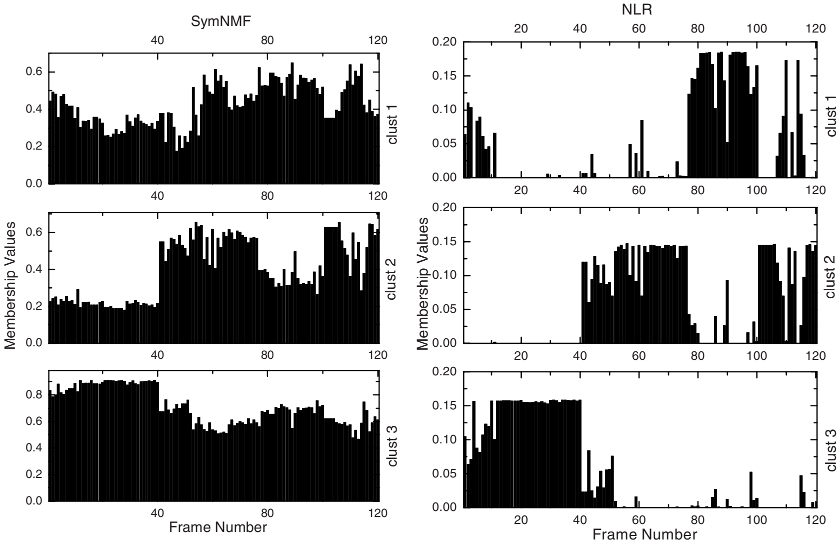

Fig. 1. Results of the SymNMF (left) and NLR (right)

In order to extract the MREs of each cluster, we apply the orthogonality condition with a mild deviation from the strict orthogonality according to [17]. Thus, we apply a threshold *T* to the entries of **X**. The value of *T* controls the degree of summarization of the WCE video. Large values of T lead to more examples (frames) in the resulting set of MREs. Figure 2 illustrates how the total number of frames in the resulting video varies with *T*, as well as the percentage reduction in the total number of frames of the original video. From these figures it is obvious that for threshold values close to 1E-5 the total number of frames per cluster is substantially reduced. Moreover, the total number of frames may be reduced down to 10% of the number of frames of the original video, and since the number of frames is proportional to the visual inspection time, a 90% reduction in this time is feasible.

Such a frame reduction would actually be worthless if the remaining frames would not contain representatives from all the possible abnormal findings, since missing even a sing abnormality could be critical for the patient. A thorough examination of the frames comprising the summarised video validated that the proposed approach did not miss any abnormal finding. The summarized video was containing at least one representative frame from each neighbourhood of frames of abnormal findings. For certain values of *T* the proposed approach missed some neighbourhoods of normal frames, but this is insignificant considering that it does not have any implications for the patient. For *T* below 1E-5 the number of frames became too small and many neighbourhoods didn't have any representative frames in the final video.

By integrating a time stamp to each representative frame we can offer the physician the ability to return to the corresponding frame of initial video so as to further examine the area of interest.

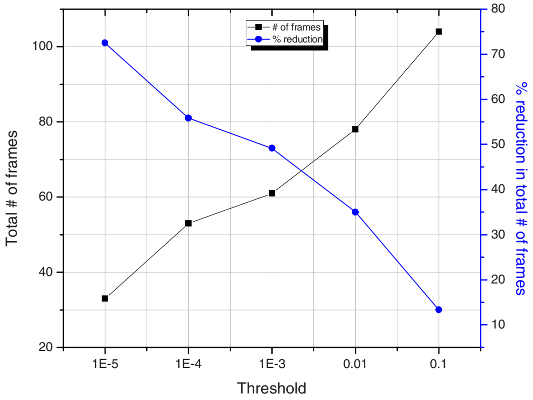

Fig. 2. Total number of frames and percentage reduction in the total number of frames for different values of threshold *T*

4 Conclusions

The novel approach to WCE video summarisation that we presented is based on the application of the NMF on the video frames according to the methodology proposed in [17]. The results of its experimental evaluation on annotated WCE videos with multiple findings showed that a significant reduction in the total number of frames of the original video without any loss of patient-critical information is feasible, leading to a significant reduction of the visual inspection time required per endoscopic examination.

Our future work includes utilization of various image features for the discrimination of other types of abnormal findings such as polyps and cancer, as well as further experimentation with many annotated WCE videos and investigation of memory-efficient techniques to perform NMF on large WCE video streams.

References

1. Iddan, G., Meron, G., Glukhovsky, A., Swain, P.: Wireless capsule endoscopy. Nature 405 (6785), 417–418 (2000)
2. Maieron, A., Hubner, D., Blaha, B., Deutsch, C., Schickmair, T., Ziachehabi, A., Kerstan, E., Knoflach, P., Schoefl, R.: Multicenter retrospective evaluation of capsule endoscopy in clinical routine. Endoscopy 36(10), 864–868 (2004)
3. Coimbra, M.T., Cunha, J.P.S.: MPEG-7 visual descriptors - contributions for automated feature extraction in capsule endoscopy. IEEE Transactions on Circuits and Systems for Video Technology 16(5), 628–636 (2006)
4. Li, B., Meng, M.Q.-H.: Analysis of the gastrointestinal status from wireless capsule endoscopy images using local color feature. In: ICIA 2007 International Conference on Information Acquisition, July 8-11, 2007, pp. 553–557 (2007)
5. Mackiewicz, M., Berens, J., Fisher, M., Bell, D.: Colour and texture based gastrointestinal tissue discrimination. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2, pp. II597-II600 (2006)

6. Berens, J., Mackiewicz, M., Bell, D.: Stomach, intestine and colon tissue discriminators for wireless capsule endoscopy images. In: Progress in Biomedical Optics and Imaging, Proceedings of SPIE, vol. 5747, pp. (I): 283–290 (2005)
7. Lee, J., Oh, J., Shah, S.K., Yuan, X., Tang, S.J.: Automatic classification of digestive organs in wireless capsule endoscopy videos. In: Proceedings of the ACM, Symposium on Applied Computing, pp. 1041–1045 (2007)
8. Bourbakis, N.: Detecting abnormal patterns in WCE images. In: Proceedings - BIBE 2005, 5th IEEE Symposium on Bioinformatics and Bioengineering, pp. 232–238 (2005)
9. Hwang, S., Oh, J., Cox, J., Tang, S.J., Tibbals, H.F.: Blood detection in wireless capsule endoscopy using expectation maximization clustering. In: Progress in Biomedical Optics and Imaging, Proceedings of SPIE, vol. 6144 I (2006)
10. Kodogiannis, V.S., Boulougoura, M.: Neural network-based approach for the classification of wireless-capsule endoscopic images. In: Proceedings of the International Joint Conference on Neural Networks, vol. 4, pp. 2423–2428 (2005)
11. Kodogiannis, V.S., Boulougoura, M., Lygouras, J.N., Petrounias, I.: A neuro-fuzzy-based system for detecting abnormal patterns in wireless-capsule endoscopic images. Neurocomputing 70(4-6), 704–717 (2007)
12. Li, B., Meng, M.Q.-H.: Wireless capsule endoscopy images enhancement by tensor based diffusion. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Proceedings, pp. 4861–4864 (2006)
13. Vilariño, F., Kuncheva, L.I., Radeva, P.: ROC curves and video analysis optimization in intestinal capsule endoscopy. Pattern Recognition Letters 27(8), 875–881 (2006)
14. Vilariño, F., Spyridonos, P., Pujol, O., Vitrià, J., Radeva, P., De Iorio, F.: Automatic detection of intestinal juices in wireless capsule video endoscopy. In: Proceedings - International Conference on Pattern Recognition, vol. 4, pp. 719–722 (2006)
15. Vilariño, F., Spyridonos, P., Vitrià, J., Malagelada, C., Radeva, P.: A machine learning framework using SOMs: Applications in the intestinal motility assessment. In: Martínez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS(LNAI), vol. 4225, pp. 188–197. Springer, Heidelberg (2006)
16. Wadge, E., Boulougoura, M., Kodogiannis, V.: Computer-assisted diagnosis of wireless-capsule endoscopic images using neural network based techniques. In: Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA 2005, pp. 328–333 (2005)
17. Okun, O., Priisalu, H.: Unsupervised data reduction. Signal Processing 87(9), 2260–2267 (2007)
18. Lee, D.D., Seung, H.S.: Unsupervised learning by convex and conic coding. Adv. Neural Inf. Process. Systems 9, 515–521 (1997)
19. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)
20. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. Adv. Neural Inf. Process. Systems 13, 556–562 (2000)
21. Saul, L.K., Lee, D.D.: Multiplicative updates for classification by mixture models. Adv. Neural Inf. Process. Systems 14, 897–904 (2002)
22. Ding, C., He, X., Simon, H.D.: On the equivalence of nonnegative matrix factorization and spectral clustering. In: Proceedings of the SIAM International Conference on Data Mining, Newport Beach, CA, April 2005, pp. 606–610 (2005)
23. Ding, C., He, X., Simon, H.D.: Nonnegative Lagrangian relaxation of K-means and spectral clustering. In: Proceedings of the Sixteenth European Conference on Machine Learning, Porto, Portugal, October 3–7, 2005, pp. 530–538 (2005)