Selecting Texture Discriminative Descriptors of Capsule Endpscopy Images

Piotr Szczypiński Artur Klepaczko

Medical Electronics Division, Institute of Electronics, Technical University of Lodz
ul. Wolczanska 211/215, PL-90924 Lodz, Poland
E-mail: {aklepaczko, kocinski, materka}@p.lodz.pl

Abstract

In supervised data classification one of the problems is to reduce dimensionality of feature vectors. It is important to find such features which have high ability for discrimination of diverse classes and to get rid of features which are useless for such discrimination. In this paper we propose a new method for feature subset selection utilizing a convex hull (or convex polytope). The method searches for feature space subspaces in which vectors of one class cluster and they are surrounded by vectors of the other class. The method is applied for selection of color and texture descriptors of capsule endoscope images. The study aims at finding a small set of descriptors for detection of pathological changes in the gastrointestinal tract. The results are compared with results produced by a Support Vector Machine with the radial basis function kernel.

1 Introduction

Wireless capsule endoscopy [5, 10] (WCE) is a technique for visualization of internal lumen of gastrointestinal tract, including a small intestine. Interpretation of the WCE video sequence involves human expert, is monotonous and time-consuming task. It requires a high level of concentration, so as not to miss lesions that might be present in only a few frames. Therefore, there is a need for automatic methods which would aid in the investigation by focusing the attention of the clinician on medically relevant video fragments.

Our approach to the problem of aiding the WCE interpretation utilizes image texture analysis to numerically describe anatomical structures viewed in the endoscopic images. We presume there exist texture descriptors which enable automatic discrimination between normal and pathologically altered tissues [13, 11]. There are programs which compute image textural features. The MaZda program [13, 11], which was used in this study, computes several hundred features per arbitrarily selected region of interest. The features are texture and color descriptors, and some of them have ability to discriminate different classes of WCE images. The main problem is a high dimensionality of feature space which is difficult for further analysis.

In this study it is proposed to select relevant texture parameters using a measure which respects a possibility of encapsulating all vectors from one chosen pathology type by a convex hull. Simultaneously, vectors representing other classes should remain outside the hull. The main motivation for such an approach is to minimize the rate of false negative errors committed by the classifier. Since the process of convex hull construction largely depends on data vectors lying closest to the decision boundary, the proposed method hereafter shall be referred to as Vector Supported Convex Hull (VSCH). The method not only identifies significant parameters, but it also determines a classification rule based on the mathematical definition of the best found convex hull.

2 Capsule Endoscopy

The technique of capsule endoscopy facilitates the imaging of the human gastrointestinal system including small intestine [5, 10]. 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. When the capsule goes through the small bowel it is propelled by peristaltic movements. The capsule transmits video data at a rate of two frames per second for approximately 8 hours. Investigation of the recorded material requires substantial effort even from a trained clinician. The diagnostic procedure involves viewing the video and searching for pathological changes. It is a tedious task that usually takes more than an hour.

It arises that a method which would automate the investigation process would provide significant support to a diagnostician. Studies described in the literature (e.g. in [2, 3, 8, 7, 16]) aim at the segmentation of the gastrointestinal tract into segments and then denoting the most relevant ones. For that purpose variety of image features are used, including color, intensity or selected geometrical descriptors.

Another method leads to obtaining an image of the bowel surface [12] by preprocessing the WCE video. Such an image, a bowel map, enables quick examination of the entire recording in terms of completeness and quality. The map also facilitates the identification of abnormal areas and helps focusing attention on relevant ones.

In the presented approach we presume that image regions containing different pathologies and various aspects of normal mucosal appearance also differ in terms of color and texture parameters. It is postulated here to compute such features and then use them for differentiation of image contents.

3 Texture Analysis

A texture is a visualization of complex patterns composed of spatially organized, repeated subpatterns, which have a characteristic, somewhat uniform appearance. The local subpatterns within an image are perceived to demonstrate specific brightness, color size, roughness, directivity, randomness, smoothness, granulation, etc. A texture may carry substantial information about the structure of physical objects. In medical images it may characterize the structure of human tissues or organs. Consequently, textural image analysis is an important issue in image processing and understanding in medical applications. To perform such analysis, mathematically defined texture properties are computed.

In our study we use MaZda 4.7 software [13, 11] for textural feature computation. The software is capable of conducting a quantitative analysis of texture within arbitrarily selected regions of interest (ROI) and can provide an interpretation of the computed results. There are three categories of feature computation approaches that MaZda utilizes: statistical (based on image histogram, gradient, co-ocurrence matrix, run-length matrix), model-based (implementation of the autoregressive model) and image transform (based on the Haar wavelet). MaZda may be used to compute textural descriptors based on color components of a color image, such as Y, R, G, B, U, V, I, Q, color saturation and hue. The textural features computed for different color components can be combined to obtain a comprehensive characterization of a colored texture. Therefore, feature vectors computed by MaZda may include over a thousand elements per individual region of interest. Such a large number of features, creating several-hundred-dimensional spaces, are not easy to handle by statistical analysis or by classifiers.

4 Vector Supported Convex Hull Method

Since the main problem is to find a way of discriminating between various image classes, the Vector Supported Convex Hull Method aims at two objectives. The first is to reduce the dimensionality of the vector space by optimizing the number of vector features. This goal is achieved by selection of such subsets of features, which present best discrimination ability among other feature subsets. Usually only a limited number of features carry relevant information needed for discrimination and other features are redundant for classification. The profit of such selection is that redundant features are not calculated, which saves computation time.

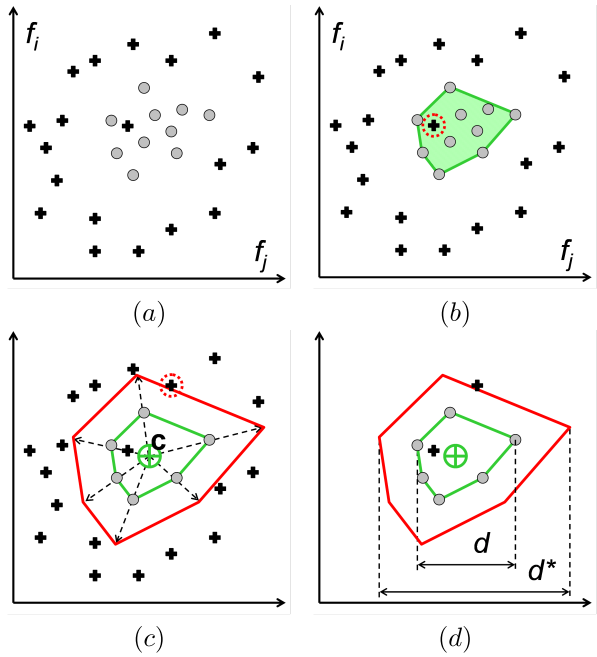


Figure 1. Illustration of convex hull method in 2D space (k=2).

The second objective of the VSCH method is to propose a way for vector classification. The method produces a number of conditions - inequalities, which define a region of vectors as belonging to the class of interest. Thus, rules for classification are formulated throughout these inequalities.

VSCH is a discriminant analysis method of supervised learning for reduction of vectors dimensionality and data classification. It aims at finding a subspace in feature vector space and produces a classification rule to separate the two classes. To explain the concept of VSCH let us assume input data consist of two sets (classes) of feature vectors in an n-dimensional space.All the features are real numbers and the feature vector space is also real. Moreover, there exists a k-dimensional subspace ($k < n$) such that vectors of the set number one form a cluster surrounded by vectors of the set number two (cf. Fig. 1).

$$\Theta \subset \Omega; \Theta \in \mathcal{R}^k; \Omega \in \mathcal{R}^n; k < n. \tag{1}$$

Let us consider a convex hull of set one in a k-dimensional subspace of feature vectors space ($m < n$). The convex hull can be found by solving a system of equations (2) and inequality conditions (3).

$$\mathbf{B}^T \mathbf{x}_\Theta + b_0 = 0, \tag{2a}$$
$$\|\mathbf{B}\| = 1 \tag{2b}$$

$$\mathbf{B}^T \mathbf{x}_\Theta + b_0 \leq 0. \tag{3}$$

Equation (2a) defines a hyperplane in k-dimensional space. Equation (2b) reduces a number of possible solutions to two. The equations (2a) and (2b) are solved for $k+1$ number of linearly independent vectors belonging to class number one. Vector $\mathbf{B}$ and parameter b_0 are unknowns. There are two solutions of (2) per each subset of $k+1$ vectors. The boundary of the convex hull is then defined by such solutions, which in addition satisfy inequality (3) for all the vectors belonging to class one.

Now we define a coefficient Q_1. It is the number of vectors belonging to the second class, which also belong to the convex hull built on class number one. It is the number of vectors of the second class satisfying inequality (3) defining the convex hull. The example in 1b shows one such vector. Therefore, in the case presented in the figure the $Q_1 = 1$. Generally, the lower the value of Q_1, the better class separation for the analyzed Θ subspace. The next step is to find a centroid $\mathbf{c}$ of the convex hull. Then the convex hull is isotropically scaled up (cf. Fig. 1c) around the fixed centroid $\mathbf{c}$ of the convex hull. The scaling is an affine transformation given by Eq. 4.

$$\mathbf{X}^*_Q = a(\mathbf{x}_\Theta - \mathbf{c}) + \mathbf{c}. \tag{4}$$

The parameter a defines a space enlargement. Parameter a is a maximum scaling factor for which Q_1 does not increase. Now we define a Q_2 coefficient which is equal to reciprocal of the parameter a (cf. Fig. 1d also).

$$Q_2 = a^{-1} = \frac{d}{d^*}. \tag{5}$$

Since, the a parameter is larger than 1, the coefficient Q_2 is a fraction. On the other hand coefficient Q_1 is an integer number equal or higher than 1. Now, we combine the two coefficients and define a comprehensive Q coefficient as:

$$Q = Q_1 + Q_2. \tag{6}$$

The Q specifies discriminative power of k-dimensional feature space. The lower value of the Q indicates the analyzed Θ subspace has better class separability. The algorithm for feature space reduction based on VSCH method was implemented. The algorithm searches 1D, 2D and 3D feature subsets (Θ subspaces) and computes Q coefficient for each subset. For further analysis and classification purpose such subset is chosen, which exposes the lowest Q coefficient. The algorithm also produces rules of classification. The rules are given in form of inequalities (3). Inequalities define boundaries obtained by scaling-up (4) the convex hull by factor of $a/2$.

In many medical applications it is crucial not to overlook any indications of pathology. Such indications usually are later verified by medical experts and may be rejected. If they are mistakenly rejected by an automatic method, an expert may never notice the pathology. Therefore, it is important to find methods characterized by a minimal false negative error. The VSCH method reveals a property which is particularly useful in biomedical image analysis.. The method produces classification rules, for which (for the training set vectors) the false negative error is equal to zero. The minimization of false positive errors is a secondary goal, and is achieved directly by minimization of the Q_1 coefficient.

5 Support Vector Machines

The proposed VSCH method presumes specific distribution of vectors. Similar concept underlies a well established classification algorithm – Support Vector Machine (SVM) with the radial basis function (RBF) kernel [15], which assumes spherical shape of a decision boundary between two different classes. Thus, it is reasonable to evaluate VSCH performance in comparison with the SVM-RBF classifier. For the need of the comparative study reported below, SVM-RBF was employed for both feature selection and classification tasks.

The SVM itself is a linear classification algorithm. The constructed decision hyperplane is defined as

$$y(\mathbf{x}) = b + \sum_{\alpha_i \neq 0} \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x}, \tag{7}$$

where parameters α_i and b together with the support vectors $\mathbf{x}_i$ determine location and orientation of the separating hyperplane. The learning procedure involves solving a constrained quadratic optimization problem which leads to determination of α_i coefficients, whose values are nonzero only for those vectors in a training sample which lie closest (on either side) to a decision boundary. It must be noted, that SVM algorithm constructs a hyperplane which defines the largest margin between different data vectors classes in a particular feature subspace. In this aspect, the VSCH method behaves similarly to SVM.

Another attractive property, which SVM possesses, allows its easy extension to non-linearly separable cases. The dot product $(\cdot)$ in (7) can be replaced by the kernel function which corresponds to the dot product of data vectors non-linearly transformed into higher dimensional feature space. It is expected, that this hypothetical multidimensional space already allows linear discrimination of different classes. The main problem which arises here is to find appropriate transformation of the input data set. This reduces to choosing a kernel function for calculation of the dot product. From the reasons outlined above, in this research the radial basis function was chosen. It is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2). \tag{8}$$

However, even if the general form of the kernel function is known, it still needs to be adjusted to the specific properties of a given data set. The value of γ coefficient cannot be determined automatically and several trials must be made before a trained classifier gains its discriminative power. In principle the larger γ is, the better accuracy on a training set is observed. On the other hand, there appears a risk of overfitting when the value of γ becomes too large. Hence, in every experiment one must find a good trade-off between error rate obtained for a training set and generalization capabilities of a trained SVM.

Figure 2. Example of class one (ulceration).

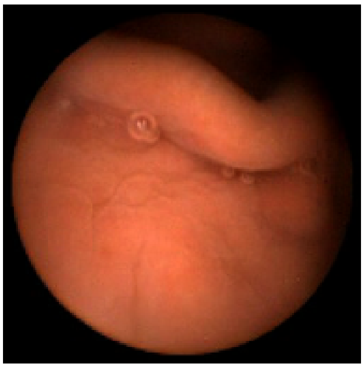
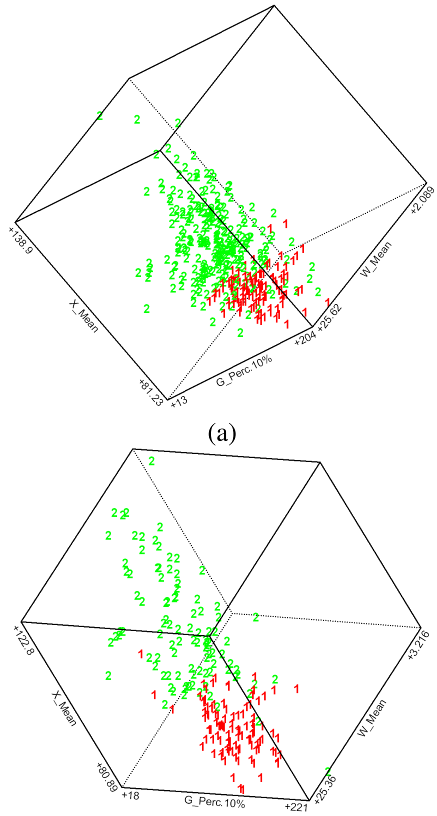


Figure 3. Examples of class two (normal appearance of mucosal surface).



(a)

(b)

Figure 4. Distributions of the training (a) and the testing (b) vectors in the feature subspace found by VSCH.

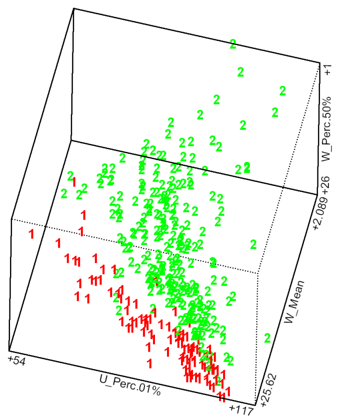
6 Experiment

To asses effectiveness of the VCSH method the following experiment was devised. Fifty images showing case of excessive ulceration were selected out of three video files obtained for three different patients (cf. Fig. 2). Regions of ulceration (regions of interest) were manually depicted within the images. For reference, 200 images showing normal appearance (cf. Fig. 3) of mucosal surface were randomly chosen from other ten videos. Then all the selected images were divided into circular overlapping subregions, each of 2009 pixels area. For images showing ulcerations, textural features were computed within circular subregions enclosed within the depicted regions of ulceration. For other images textural features were computed within circular subregions enclosed within the image field of view. Features were computed by means of MaZda program. Feature vectors included histogram descriptors computed for 14 various color channels as well as gradient, co-ocurrence matrix, run-length matrix, autoregressive model and Haar wavelet transform descriptors computed for image brightness (together over 300 features per region).

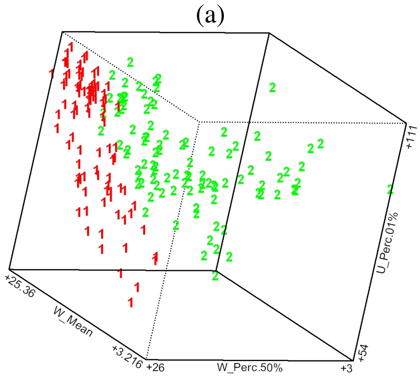
The number of vectors obtained was over 400 for class one (ulceration) and over 4800 for class two (normal). After that, training and testing sets were assembled. Training set was composed of 109 vectors of class one and 258 vectors of class two. Testing set was composed of 100 vectors of class one and 100 vectors of class two. In both cases vectors were picked randomly from the set of all the produced vectors. Then, VSCH and SVM methods were applied for attribute subset selection and data classification purpose. In both applications the goal of feature selection was to find a pair of features with the highest discrimination ability given the methods criteria. Feature space exploration was performed using exhaustive search. This eliminates the impact of local optima of criterion function or randomness associated with heuristic strategies such as genetic algorithm or sequential search methods [1, 6, 9].

Based on the training set, the VSCH method selected a pair of features computed from hue and green color components of the image. They are the mean value of the hue component (X_Mean) and a tenth percentile of the green component (G_Perc.10%) computed within the image region. Fig. 6a presents distribution of the training set vectors within feature space of X_Mean, G_Perc.10% and additional W_Mean (mean of brightness normalized U component). Fig 4b presents distribution of the testing set vectors within the feature space.

In the case of SVM-based analysis the resulting attribute subspace also consisted of two first-order-histogram



(a)



(b)

Figure 5. Distributions of the training (a) and the testing (b) vectors in the feature subspace found by SVM-RBF.

features calculated for color components of the images (W_Perc.50% and U_Perc.01%). Scatter plots of the training and testing data vectors in the reduced feature space are depicted in Fig. 5. The classification specificity, sensitivity, false positive rate and false negative rate computed for both selection methods are presented in Table 1.

7 Results discussion and conclusions

Analysis of the obtained results leads to the following conclusions. First of all, the performed experiments confirm that texture analasysis provides a practical numerical description of the WCE images. It is possible to accurately classify different types of visualized tissues basing on the selected, most relevant texture parameters. Among the calculated attributes, color component features appear to be the best at discriminating ulceration and normal regions.

Table 1. Classification results

		FPR [%]	Specificity	FNR [%]	Sensitivity
VSCH	Training set	9.3	0.907	0.0	1.000
	Testing set	7.0	0.930	6.0	0.940
SVM	Training set	4.3	0.958	6.4	0.936
	Testing set	5.0	0.950	9.0	0.910

Secondly, the error rates as well as accuracy measures viewed in Table 1 are comparable for both tested approaches to feature selection. The VSCH method appears to be overoptimistic when predicting the False Negative Ratio (FNR) on the training set. This results directly from the very nature of the algorithm which aims at construction of a convex hull arround all vectors from a chosen pathology class. However, despite the observed increase in FNR calculated for the testing set, it is still lower than the score obtained for the SVM-based method. In the case of the latter, cost-sensitive learning should be applied to improve its performance with respect to positive class vectors misclassified as negatives. As it has already been mentioned, missing an image that contains important diagnostic information implies consequences that are potentially more dangerous for a patient. The False Positive Ratio (FPR) is not as important – a diagnostician allways has a chance to disregard images wrongly marked as containing pathologies. The proposed VSCH method ensures the desired behaviour without any explicit weighting of error types.

Eventually, usage of SVM involves problem-specific parameterization of a kernel function. Frequently, one must experiment with several values of power exponents (both in polynomial or radial basis functions) before a final choice can be made. On the other hand, VSCH is a non-parametric method and does not require any fine-tuning to solve particular tasks. Moreover, it does not require any feature space standardization. Also any other linear transformation of feature space has no influence on the result produced by the method.

The presented results constitute a preliminary study on classification of WCE images basing on their texture paramaters. This research shall be continued in order to further validate the proposed approach to feature selection. Furhter experiments are planned with the use of new sample images, possibly representing more than two classes.

Acknowledgments.

This work was supported by the Polish Ministry of Science and Higher Education grant no. 3263/B/T02/2008/35. The second author is a scholarship holder of the project entitled "Innovative education…" supported by the European Social Fund.

References

[1] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning", *Artificial Intelligence*, 97, 1997, pp. 245–271.

[2] M. Coimbra, P. Campos, and J. Cunha, "Extracting clinical information from endoscopic capsule exams using mpeg-7 visual descriptors", In *Integration of Knowledge, Semantics and Digital Media Technology, 2005. EWIMT 2005. The 2nd European Workshop on the*, pp. 105–110, 2005.

[3] M. Coimbra and J. Cunha, "Mpeg-7 visual descriptorscontributions for automated feature extraction in capsule endoscopy", *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(5), May 2006, pp. 628–637.

[4] R. Haralick, "Statistical and structural approaches to texture", *IEEE Proceedings*, 67(5), May 1979, pp. 768–804.

[5] G. Iddan, G. Meron, A. Glukhowsky, and P. Swain, "Wireless capsule endoscopy", *Nature*, 405(6785), 2000, pp. 417–418.

[6] R. Kohavi and G. H. John, "Wrappers for feature subset selection", *Artificial Intelligence*, 97, 1997, pp. 273–324.

[7] M. Mackiewicz, J. Berens, and M. Fisher, "Wireless capsule endoscopy video segmentation using support vector classifiers and hidden markov models", In *Proceedings of the International Conference on Medical Image Understanding and Analyses*, June 2006.

[8] M. Mackiewicz, J. Berens, M. Fisher, and G. Bell, "Colour and texture based gastrointestinal tissue discrimination", In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, volume 2, pp. 597–600, May 2006.

[9] P. Pudil and P. Somol, "Current feature selection techniques in statistical pattern recognition", In M. Kurzynski, E. Puchala, M. Wozniak, and A. Zolnierek, editors, *Computer Recognition Systems*, volume 30 of *Advances in Sof Computing*. Springer-Verlag, 2005.

[10] P. Swain and A. Fritscher-Ravens, "Role of video endoscopy in managing small bowel disease", *GUT*, 53, 2004, pp. 1866–1875.

[11] P. Szczypinski. http://www.eletel.p.lodz.pl/MaZda, 2009. Visited: April 2009.

[12] P. Szczypinski, R.D.Sriram, P. Sriram, and D. Reddy, "A model of deformable rings for interpretation of wireless capsule endoscopic videos", *Medical Image Analysis*, 13(2), April 2009, pp. 312–324.

[13] P. Szczypinski, M. Strzelecki, A. Materka, and A. Klepaczko, "Mazda - a software package for image texture analysis", *Computer Methods and Programs in Biomedicine*, 94, 2009, pp. 66–76.

[14] M. Tuceryan and A. Jain, *The Handbook of Pattern Recogntion and Computer Vision*, chapter Texture Analysis, pp. 207–248. World Scientific Publishing Co., 1998.

[15] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[16] F. Vilarinao, L. I. Kuncheva, and P. Radeva, "Roc curves and video analysis optimization in intestinal capsule endoscopy", *Pattern Recogn. Lett.*, 27(8), 2006, pp. 875–881.