IOP Publishing Phys. Med. Biol. 63 (2018) 035031 (11pp) https://doi.org/10.1088/1361-6560/aaa3af

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Automatic T1 bladder tumor detection by using wavelet analysis in cystoscopy images

Nuno R Freitas[1], Pedro M Vieira[1], Estevão Lima[2,3] and Carlos S Lima[1]

[1] CMEMS-UMinho Research Unit, University of Minho, Guimarães, Portugal
[2] Life and Health Sciences Research Institute, University of Minho, Campus Gualtar 4710-057, Braga, Portugal
[3] Department of Urology, Hospital of Braga, Braga, Portugal

E-mail: nrafeb@gmail.com

RECEIVED
27 July 2017

REVISED
29 November 2017

ACCEPTED FOR PUBLICATION
22 December 2017

PUBLISHED
2 February 2018

Keywords: bladder tumor, cystoscopy, discrete wavelet transform, multilayer perceptron, segmentation

Abstract

Correct classification of cystoscopy images depends on the interpreter's experience. Bladder cancer is a common lesion that can only be confirmed by biopsying the tissue, therefore, the automatic identification of tumors plays a significant role in early stage diagnosis and its accuracy. To our best knowledge, the use of white light cystoscopy images for bladder tumor diagnosis has not been reported so far. In this paper, a texture analysis based approach is proposed for bladder tumor diagnosis presuming that tumors change in tissue texture. As is well accepted by the scientific community, texture information is more present in the medium to high frequency range which can be selected by using a discrete wavelet transform (DWT). Tumor enhancement can be improved by using automatic segmentation, since a mixing with normal tissue is avoided under ideal conditions. The segmentation module proposed in this paper takes advantage of the wavelet decomposition tree to discard poor texture information in such a way that both steps of the proposed algorithm segmentation and classification share the same focus on texture. Multilayer perceptron and a support vector machine with a stratified ten-fold cross-validation procedure were used for classification purposes by using the hue-saturation-value (HSV), red-green-blue, and CIELab color spaces. Performances of 91% in sensitivity and 92.9% in specificity were obtained regarding HSV color by using both preprocessing and classification steps based on the DWT. The proposed method can achieve good performance on identifying bladder tumor frames. These promising results open the path towards a deeper study regarding the applicability of this algorithm in computer aided diagnosis.

1. Introduction

Bladder cancer is the second most common genitourinary malignancy and the fourth most common cancer in men, with 80 000 new cases estimated to have occurred in the U.S. during 2017 (Stein *et al* 2011, American Cancer Society 2017). More than 70% of primary bladder tumors are superficial transitional cell carcinoma, in which recurrence rates for patients with Ta and T1 tumors are in the range of 50%–75% at 20 years (Fielding *et al* 2002, Grossman *et al* 2007). About 9 out of 10 people with this type of cancer are over the age of 55, in which the chance of men developing this cancer during their life is about 1 in 26, and for women about 1 in 88 (American Cancer Society 2017).

The inside of the bladder can be visualized through cystoscopy. This exam consists of inserting a thin tube through the opening of the urethra that advances into the bladder. The equipment is called a cystoscope and comprises a light source and a lens that allows for the images to be seen in an exterior monitor, leading to a quick prediagnosis of this area of the urinary tract. There is also a side channel where various thin devices can pass through. The initial diagnosis is based only on a visual analysis by the physician, which obviously depends on the urologist's experience and expertise, among other subjective conditions. To confirm if it is cancer, the abnormal area identified during a cystoscopy is biopsied and classified according to its invasiveness and grade by a pathologist (Messer *et al* 2005).

According to the 2009 TNM (tumor, node, metastasis) classification system of malignant tumors, bladder tumors are classified as Ta when a non-invasive papillary carcinoma is present, CIS (carcinoma *in situ*), a 'flat', high-grade, non-invasive tumor which means that the lesion is only found on the bladder surface and T1 when a tumor invades the subepithelial connective tissue. These tumors can be treated by transurethral resection of the bladder (TURB). Tumors can also be classified as T2, when a tumor invades muscularis and T3 and T4 where the tumor invades adjacent fat or the adjacent organs, respectively (Rouprêt *et al* 2011, Babjuk *et al* 2017). Depending on the patient's age, cystectomy can be performed as a treatment for these types of tumors (Witjes *et al* 2014). Usually, papillary tumors, both high and low-grade, are easily identified but hard to classify. On the other side, flat tumors are harder to identify (Mitropoulos *et al* 2005). CIS is often invisible to the naked eye, but it can quite frequently be identified as a red, velvet area along the surface of the normally, pink, translucent bladder wall (Vining *et al* 1995). These visual features usually induce ambiguity in the diagnosis between CIS or inflammation. In those less aggressive tumor cases (Ta, T1, and CIS), after tumor removal, most patients are prone to experience recurrence at least once. Thus, every 3–6 months during 2–5 years, patients are submitted to controls by cystoscopy and urinary cytology (NCCN 2015). This whole process works as diagnosis; but the histological results can take several days, and due to their high prevalence, the need for early detection methods has become of huge importance. In this work, only T1 images were used because this is the first stage of tumor growth where the tumor has invaded the first layer of tissue and besides that, they are often of a higher grade than Ta tumors (Pasin *et al* 2008).

In previous works, the authors applied segmentation approaches to separate tumor from normal tissue in individual images taken from cystoscopy exams. This was done with a color-based analysis, considering that the distribution of the intensities of different color channels is modeled by a Gaussian mixture model, whose parameters can be found by using the EM algorithm (Freitas *et al* 2017a, 2017b). Although, due to the repetitive change in the gray values and structures of the textures, traditional detection methods based on intensity or edge detection are ineffective because textures are realizations of non-Gaussian stationary processes. Texture is an inherent property in an image, especially in medical images since the tissue itself carries a dominant textural appearance (Maroulis *et al* 2008). There are three main approaches to the task of texture feature extraction: the spectral processing approach, structural approach, and statistical approach (Kodogiannis *et al* 2007).

The lack of previous works regarding tumor detection in white light bladder images, lead us to focus on similar types of tissue, specifically intestinal tissue in images of wireless capsule endoscopy (WCE). When trying to detect intestinal tumors in WCE images, textural features lead to good results because of its irregular shape and heterogeneous appearance. One of the approaches is the use of Haralick descriptors, which are calculated with a gray-level co-occurrence matrix calculated in the four directions (Haralick *et al* 1973). These descriptors are considered state-of-the-art for texture codification, and have been used with good results in tumor detection in WCE images by Barbosa *et al* (2008a) and Lima *et al* (2009). Maroulis *et al* (2008) proposed a method based on the covariance of second-order textural measures in the wavelet domain for WCE images, namely, in the bands 4, 5, and 6, while Kodogiannis *et al* (2007) proposed statistical approaches where descriptors are calculated from the histograms of different color spaces. They also proposed a scheme to extract features from texture spectra in the chromatic and achromatic domains (Barbosa *et al* 2012). Intestinal and bladder tissue have a high degree of similarity, so it can be inferred that similar methods can be applied while probably leading to good results. The discrete wavelet transform (DWT) has been widely used and provides a powerful tool for multi-resolution analysis of images (Barbosa *et al* 2008b, Karkanis *et al* 2003, Maroulis *et al* 2008, Lima *et al* 2008, Barbosa *et al* 2012). The DWT is suitable for dealing with singularities, and extracting details in horizontal, vertical, and diagonal directions. The image or signal to be analyzed is passed through filters with different cut-off frequencies at different scales, allowing a better decomposition of the patterns within an image (Barbosa *et al* 2011).

The experimental evaluation of extracted features, i.e. classification methods, focus mainly on the statistical analysis of texture, and have been based on several classifiers such as multi-layer perceptron (MLP) networks, radial basis functions (RBFs), extreme learning machines (ELMs), wavelet neural networks, support vector machines (SVMs), the K-nearest-neighbor (KNN), or the adaptive fuzzy logic system (AFLS) (Kodogiannis *et al* 2007, Prasath 2016, Yahia *et al* 2017). MLP is a powerful 'expert' tool due to its remarkable ability to extract patterns and detect trends from imprecise data that are too complex to be noticed by either humans or other computer techniques. The backpropagation scheme combined with the parameter learning rate minimizes the function error. It has the capability to learn non-linear models and real-time models (Kuncheva 2014, Prasath 2016).

The proposed scheme to detect T1 tumor tissue in cystoscopy images contains three main steps: preprocessing, segmentation, and classification. All steps are based on high frequency texture information obtained from filtering the frame in the wavelet domain. Since previous attempts to segment tumor frames (Freitas *et al* 2017a, 2017b) did not perform optimally for T1 tumor aspects a preprocessing, which computes an image with only higher frequency content, allowed substantial improvements in segmentation. The segmentation module uses the maximum *a posteriori* (MAP) approach and the expectation-maximization (EM) algorithm. The classifica-

tion module is based on the MLP classifier where multi-scale texture descriptors are extracted from the wavelet domain of the image and used as features.

2. Methodology

This section is divided in three sub-sections: firstly, a brief explanation of texture characterization, specifically in the DWT domain, is given; secondly, an explanation of the segmentation method used is provided, and lastly, an overview of the overall method is presented.

2.1. Texture characterization

Tumors are lesions that are visually described as having a very specific texture, when compared to healthy bladder tissue. The proposed method relies on a color textural features extraction process based on complex transform analysis, specifically, the DWT.

The wavelet transform emerged as an alternative to Fourier transform (FT) and its related transforms, namely, the discrete sine transform (DST) and the discrete cosine transform (DCT), which are more appropriate for signals generally assumed as periodic (Toufik and Mokhtar 2012). The DWT introduced a significant change in the direction of the research done on biomedical signal/image processing and allows a spatial/frequency representation by decomposing the image at different scales with different frequency content. It has also been vastly explored for several problems, such as compression of medical images, image denoising, image enhancement, tomographic reconstruction, and feature extraction for image classification (Barbosa *et al* 2011). Large wavelet coefficients are located in the neighborhood of edges and irregular textures (Mallat 2008). Boundary extent and gradient direction are other important clues captured by the DWT and precisely located viewed at different scales along the image. Wavelet orthonormal bases of images can be constructed from wavelet orthonormal bases of 1D signals. The continuous wavelet transform (CWT) is a signal representation in a scale-time space, and its coefficients of signal $x(t)$ are given through:

$$X_{\psi}(\tau, s) = \int_{-\infty}^{+\infty} x(t)\psi^{*}\left(\frac{t-\tau}{s}\right) \mathrm{d}t, \tag{1}$$

where ψ is the mother wavelet function and * stands for the complex conjugate. Analogously, a signal can be recovered from its wavelet coefficients through the inverse transform (ICWT). By varying both the scale s and the translation shift parameter τ it is possible to obtain a family of daughter wavelets from the mother wavelet function ψ (Barbosa *et al* 2011):

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right). \tag{2}$$

One important property of the wavelet filter is that for a discrete set of scales, namely, dyadic scale $s = 2^{i}$, a quality factor of the filter is defined as the central frequency to bandwidth ratio. Sampled points in the time-scale are separated by a power of 2, inspired by the so-called dyadic grid. Under this constraint, the DWT is defined as (3).

$$\psi_{j,k}(t) = s_0^{\frac{-j}{2}}\psi(s_0^{-j}t - k\tau_0). \tag{3}$$

Since a dyadic scale is used, the scaling and translation factors are determined by $s = 2^{j}$ and $\tau = k2^{j}$, where j and k are integers, $s_0 = 2$, and $\tau_0 = 1$. Under this scale progression the signal can be sequentially half-band high-pass ($g[n]$) and low-pass ($h[n]$) filtered (Kociolek *et al* 2001). The process proceeds iteratively in a scheme known as a wavelet decomposition tree, which is illustrated in figure 1.

These properties can be extended to the 2D space and applied to image processing. Mallat (1989) introduced an extension of the concepts of multi-resolution decomposition to image processing. The first step is identical to the 1D approach, however, instead of keeping the low-level resolution and processing the high-level resolution, both are processed using two identical filter banks after the transposition of the incoming data. Thus, the image is scanned in both rows and columns (Toufik and Mokhtar 2012).

The A_n is obtained by low pass filtering (H) leading to a less detailed image, at scale n. D_{ni} are obtained by band pass filtering (G) in a specific direction, therefore, encoding details in different directions. This recursive filtering is the extension of the scheme represented in figure 1 to the 2D space represented in figure 2 (Lima *et al* 2010).

This 2D implementation is therefore a recursive 1D convolution of the low and band pass filters with the rows and columns of the image, followed by the respective subsampling. The 2D DWT decomposition, as shown in figure 3, is the result of each considered scale, in sub-bands of different frequency content or detail, in the different orientations. It is a compact representation of the original image, in which the key information will be sparsely represented. The inverse discrete wavelet transform (DWT) is possible through the application of the filter bank in the opposite direction (Mallat 1989, Lima *et al* 2010).

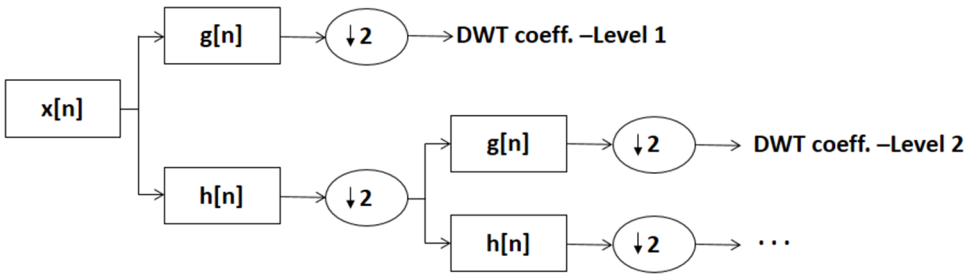

Figure 1. Wavelet decomposition tree.

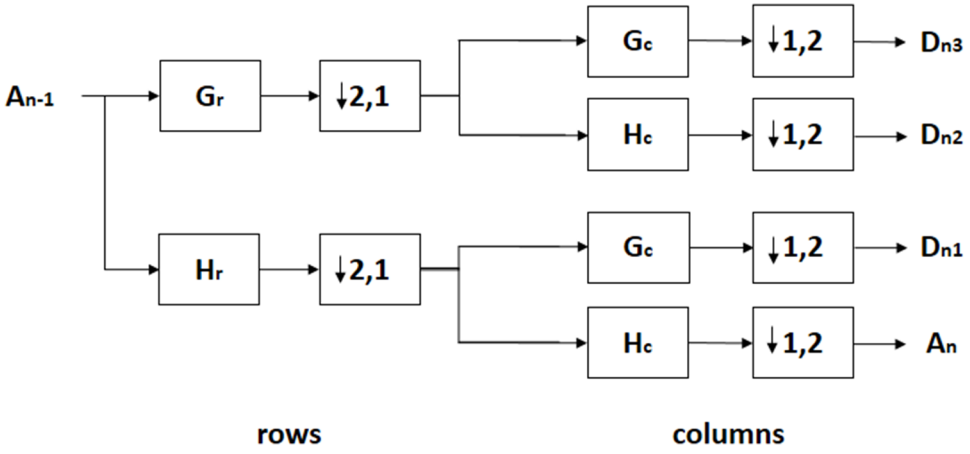

Figure 2. Wavelet 2D decomposition tree.

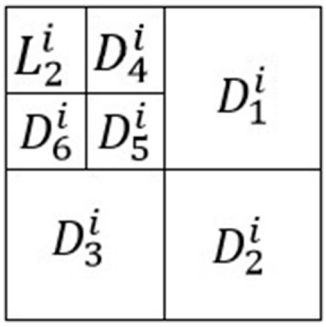

Figure 3. Two-level discrete wavelet decomposition scheme of the original image for color channel i.

Since in similar modalities (e.g. CE images) low frequency components of the images do not contain major texture information, the most important bands are those with higher frequency (lower scales). Good features should be invariant to the transformations which do not change object class. This multi-scale analysis requires that multiple images be synthesized from different levels of focus selected in the wavelet domain. This function allows a decomposition of texture information in three different directions for all frequencies and scales considered: vertical, horizontal, and diagonal; preserving both global and local information. Interesting texture information is presented in the sub-bands $\{1, 2, 3\}$ shown in figure 3, corresponding to high frequency. Each cystoscopy frame I can be decomposed in three color channels, originally in the red-green-blue (RGB) color space, but with the possibility of transforming it into other color spaces; I_i where $i = 1, 2, 3$ stands for the correspondent color channel.

Once an image is a 2D representation, a two-level DWT is applied to each color channel, I_i, resulting in a new representation of the original image by a low-resolution image and the detail images, $W^i = \{L_n^i, D_l^i\}$, where $l = 1, 2, 3$ stands for the wavelet band and n is the decomposition level. The image representation consists of the detail images produced from W^i for the values $l = 1, 2, 3$ to a two-level wavelet decomposition as shown in figure 3. This results in a set of nine sub-images, once three sub-images are generated per each one of the three-color channel components.

2.2. Segmentation

The main goal of the segmentation process is to locate lesion tissue to be processed separately from normal tissue. The segmentation algorithm used in these experiments was a multivariate MAP approach, based on Bayes' rule (4) (Freitas *et al* 2017a), that indicates how the *a posteriori* probability, $p(C_i|x)$, of each class is calculated. Once the objective is to separate abnormal from normal tissue, a two Gaussian model is appropriate.

$$P(C_i|x) = \frac{p(x|C_i) \cdot P(C_i)}{p(x)}. \tag{4}$$

In (4), $P(C_i)$ refers to the *a priori* probability of class C_i, and $p(x|C_i)$ is the class conditional probability density function (pdf). The term $p(x)$ is a scaling factor that can be ignored for class comparison. This manipulation represents the MAP estimation, with the purpose of maximizing the likelihood of a given value x to one of the k classes (Sprager and Zazula 2014). The pdf used is the Gaussian function and observations are modeled as a Gaussian mixture whose parameters can be iteratively estimated by using the EM algorithm under the maximum likelihood (ML) criterion. This process is repeated until convergence is reached, i.e. when from one iteration to the next one the increase in likelihood is negligible (below an established threshold). To do this, the usual way is to consider that a local maximum is reached when the content of each cluster does not change in consecutive iterations. The pdf of a multivariate normal distribution modeling each class can be computed by equation (5) (Vieira *et al* 2012, Freitas *et al* 2017a).

$$p(x_j|C_i, \Phi_i) = \frac{1}{(2\pi)^{D/2}\left|\sum_i\right|^{1/2}} \exp\left\{-\frac{1}{2}(x_j - \mu_i)^T \sum_i^{-1} (x_j - \mu_i)\right\}, \tag{5}$$

where D is the dimension of the distribution, i.e. the number of different random variables (channels) considered as an input of the multivariate model (in this case $D = 3$).

The likelihood of the whole dataset of size N (frame) with sample values X, considered the set of every feature vectors x is given by equation (6). In this equation, Φ_i is a tuple containing the estimated parameters for class i (mean vector μ_i, covariance matrix Σ_i, and class weight/coefficient).

$$p(X|\Phi) = \prod_{j=1}^{N} p(x_j|\Phi) = \prod_{j=1}^{N} \sum_{i}^{k} P(C_i) \cdot p(x_j|C_i, \Phi_i). \tag{6}$$

Likelihood maximization is achieved by maximizing the log-likelihood, since the logarithm is a crescent function. The update of the parameters (Freitas *et al* 2017a), is given by (7.1), (7.2) and (7.3).

$$\hat{\pi}_i = \frac{1}{N} \sum_{j=1}^{N} p(x_j|C_i, \Phi_i) \tag{7.1}$$

$$\hat{\mu}_i = \frac{\sum_{j=1}^{N} p(x_j|C_i, \Phi_i).x_j}{\sum_{j=1}^{N} p(x_j|C_i, \Phi_i)} \tag{7.2}$$

$$\hat{\sum}_i = \frac{\sum_{j=1}^{N} p(x_j|C_i, \Phi_i)(x_j - \mu_i)(x_j - \mu_i)^T}{\sum_{j=1}^{N} p(x_j|C_i, \Phi_i)}. \tag{7.3}$$

Initial estimates can be random or established according to some data distribution using a partition algorithm such as the *K-means* that divides the pixels into k clusters usually in accordance with the Euclidean distance. A partition algorithm improves initial estimates leading to local maxima closer to the global maximum being a usual procedure in practice. The entire algorithm is explained more in-depth in Freitas *et al* (2017a, 2017b).

Some advantages of the HSV color space domain are low correlation and perceptual motivation. While perceptual motivation can be advantageous to mimic human findings, low correlation allows using the IID (independents and identically distributed) criterion in such a way that the covariance matrix in equation (5) can be considered diagonal saving computational resources at runtime and the avoidance of inaccurate estimates of the off diagonal elements.

2.3. Overall method

Figure 4 shows the entire pipeline of how the cystoscopy frames were analyzed to detect tumors in bladder tissue. The preprocessing using the DWT emphasizes high texture areas (tumors are characterized by high texture variations) by discarding the more constant ones. The use of this technique as a preprocessing step produced better visible results than the standard segmentation process as shown in figure 6. Since a database with manual segmentations does not exist, only a qualitative evaluation was performed.

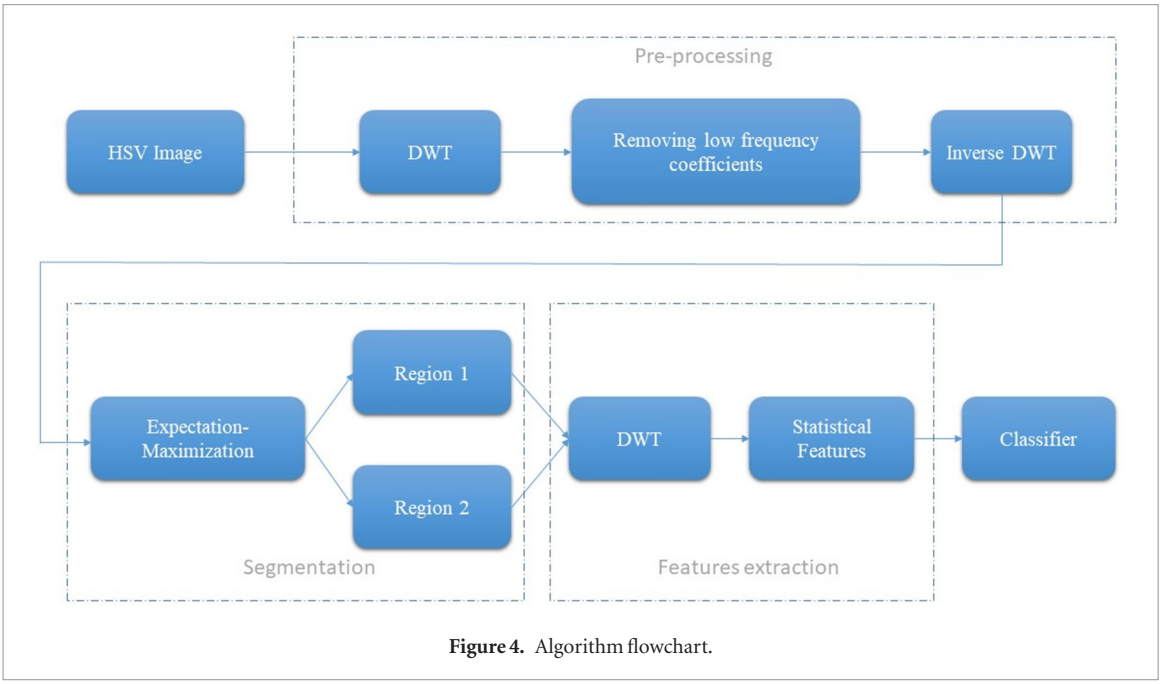

Figure 4. Algorithm flowchart.

Once texture information is mainly *encoded* in the higher frequency bands, low frequency bands can be removed replacing by zeros the lower frequencies wavelet coefficients ($L_2^i, D_4^i, D_5^i, D_6^i$, as shown in figure 1), and applying the inverse DWT to the resulting matrix. Unfortunately, the contribution of each of the details of the scale regarding textures cannot be visually compared since the lack of the coarse scale (low frequency) removes most of the visual information. This step synthesizes one image containing only the most relevant texture information from the source image. The resulting image is then segmented into two regions.

Feature value is defined as a real number which encodes some information about a property of an object. In the current case properties are texture information that can be extracted from the high frequency wavelet coefficients, D_1^i, D_2^i, D_3^i, as shown in figure 3. This representation comes from the three HSV color channels, resulting in a set of nine sub-images. The result of the segmentation procedure allows accessing suspected tumor tissue directly in this set of nine sub-images.

There are several texture descriptors extracted from the wavelet domain as statistical features, being the most common the mean (8) and the standard deviation (9) given as:

$$\mu = E\{P(i,j)\} = \frac{1}{NM}\sum_i\sum_j P(i,j) \tag{8}$$

$$\sigma = \sqrt{E\left\{(P(i,j)-\mu)^2\right\}} = \sqrt{\frac{1}{NM}\sum_i\sum_j (P(i,j)-\mu)^2}, \tag{9}$$

where N is the number of rows and M is the number of columns of the frame, and $P(i, j)$ is the pixel intensity at position (i, j). Features are used as inputs of the classifier that assign them to the class they represent. This process emphasizes the difference between what characterizes a tumor and what characterizes normal tissue by extracting features from both regions. With this quantification the dimensionality of the problem is doubled, however, it is a way to measure the relative values in terms of texture between both regions; normal and tumor. Relative measures may be more robust to environmental conditions changing such as device related (light variations among different devices) and subject related than absolute measures usually obtained without segmentation facilities. Experimental results confirm this suspicion.

In order to compare the strength of these descriptors, a state-of-the-art methodology already used in WCE images was accessed (Barbosa *et al* 2012) regarding the HSV image. Thus, co-occurrence matrices were computed and the 4 most significant measures among the 14 statistical Haralick measures are considered, namely, the angular moment (F_1), the correlation (F_2), the inverse difference moment (F_3), and entropy (F_4), representing the homogeneity directional linearity, smoothness, and randomness of the matrix (Haralick *et al* 1973). These measures are those which are mostly related to the human perception and discrimination of textures (Lima *et al* 2008).

3. Classification process

Data classification was performed by using MLP and SVM classifiers. These are the most fundamental classification methods based on statistical pattern recognition and have been widely used for decades due to their

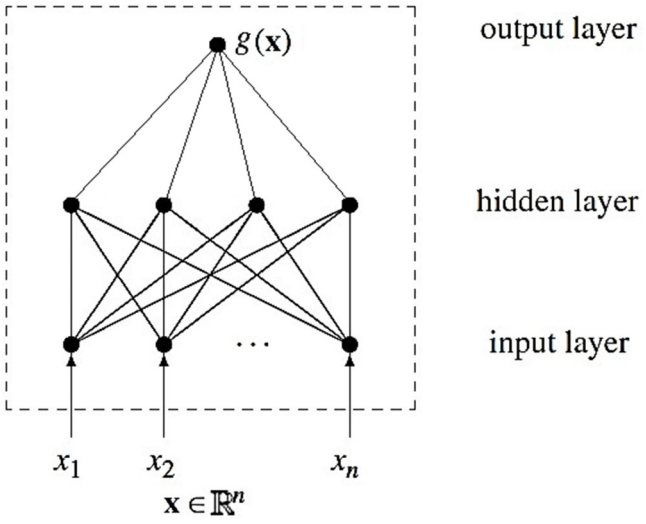

Figure 5. A MLP with one hidden layer (Kuncheva 2014. Copyright © 2014 John Wiley & Sons, Inc. All rights reserved.).

effectiveness and robustness. Unlike other statistical techniques, they can model highly non-linear functions and can be trained to accurately generalize when presented with new, unseen data (Li and Meng 2009).

The architecture of a MLP, as shown in figure 5, consists of one hidden layer of simple interconnected neurons (nodes) in a feed-forward way and utilizes a supervised learning technique called backpropagation to classify instances.

In this example, x is the feature vector and is submitted to an input layer, and $g(x)$ is the output layer. The inputs are fed simultaneously into the units making up the input layer. These inputs pass through the input layer and are then weighted and fed simultaneously to a second layer, until they reach the final output layer. The number of hidden layers is arbitrary.

The backpropagation neural network is essentially a network of simple processing elements working together to produce a complex output. Here the output values are compared with the correct answer to compute the value of some predefined error. The error updates the network weights iteratively until the output error drops below the predefined threshold or until the limit of the number of iterations is reached. To adjust weights properly, it is applied a non-linear optimization method that is called gradient descend, which converges to a set of weights corresponding to a local minimum of the output error (Gardner and Dorling 1998, Arora and Suman 2012, Kuncheva 2014).

Since its conception, the SVM classifier has been a prominent landmark in statistical learning theory and its success is attributed to two main premises: the original space is transformed into a very high-dimensional new space and a large margin can be found in this new space. Thus, it has the objective to define a hyperplane, which can be used for good separation of the training data. This good separation is achieved by the hyperplane that has the largest distance (margin) to the nearest training data point of any class. The points which are at the exact minimum distance corresponding to the minimum margin are called the support vectors (Kuncheva 2014, Marsland 2015). More elaborate schemes for classification purposes such as ensemble systems or deep neural networks are current lines of research of wide interest around the world, however, they require large datasets currently not available in the bladder tumor field.

4. Results and discussion

The proposed algorithm was developed in MATLAB on a 2.60 GHz Intel Core i7-4510U processor with 8 GB of RAM. The experimental dataset contains a total of 353 frames from which 246 are normal and 107 are indexed as T1 bladder tumors. Some of the normal frames were extracted from cystoscopy videos of 8 healthy patients, and others from clean areas of 6 sick patients. The abnormal frames are from 6 different patients. All patients were between 52 and 90 years old. These images are of various resolutions due to the recording performed using three different devices, therefore, it was important to do a normalization step of the image resolutions. They were obtained and analyzed by several experienced urologists of the Hospital of Braga and the final classification of each lesion was obtained after the analysis of the biopsy done by experienced pathologists. Figure 6(a) shows some examples of frames belonging to the dataset.

The features were imported into the open source machine learning package WEKA (available at www.cs.waikato.ac.nz/ml/weka/). A stratified ten-fold cross-validation procedure was chosen to train both classifiers.

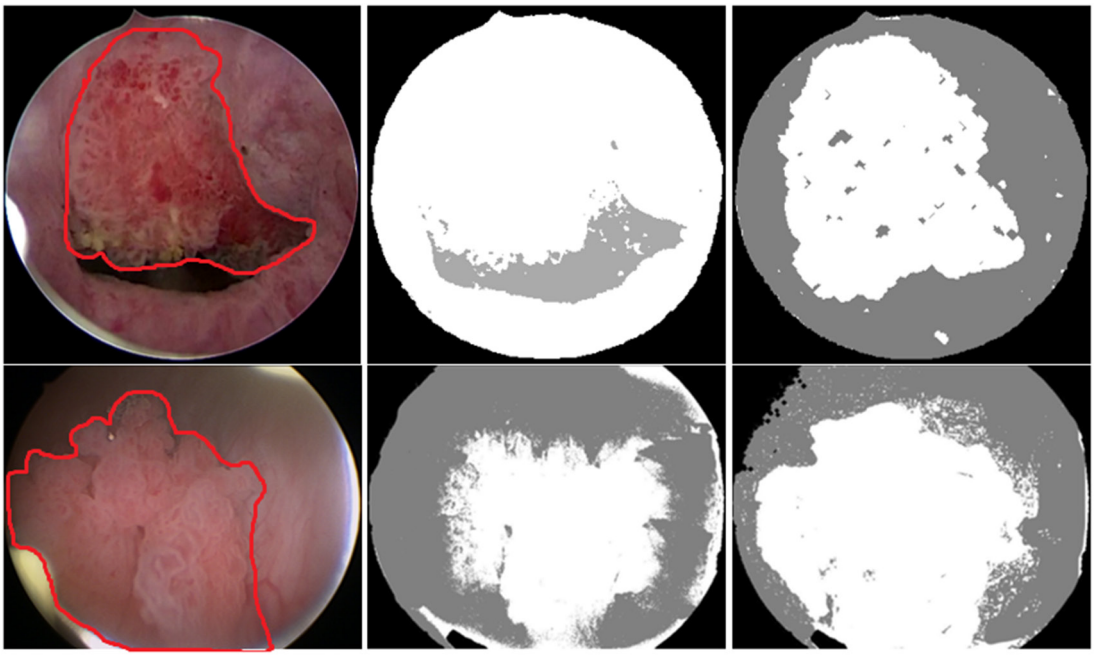

Figure 6. Resulting segmentation of (a) two tumoral frames, (b) without preprocessing and (c) with preprocessing regarding the HSV color space.

The MLP model consisted of one hidden layer with five neurons, which was considered an appropriate number for the size of the dataset. The input, in both cases, was normalized in order to avoid biased network weighting optimization. The activation function used in the MLP was the unipolar *Sigmoid*. All the other parameters were kept at the default. The ten-fold cross-validation algorithm divides the data into ten partitions, where the proportion of both normal and abnormal frames in each partition is similar to the entire dataset. The training and classification process is then repeated ten times, where 9 partitions are used to train and 1 partition is used to assess the classification process. This way, each frame will be used exactly once as test data, allowing efficient use of the available dataset. The test results are collected and averaged over all folds, which gives the cross-validation estimate of the accuracy. The variance of the resulting estimate is reduced as the number of partitions is increased (Alpaydın 2014).

In this experiment imbalanced data was used, i.e. the number of tumor frames is much smaller than the number of normal frames. This type of data usually induces high accuracy immediately, because the chosen model cleverly decides that the best thing to do is to always predict the class that allows a high accuracy. Due to these more reliable measures, the evaluation of the classification performance can be achieved by using the sensitivity (true positive rate), specificity (true negative rate), accuracy, and receiver operating characteristic curve (ROC) (Harrington 2012). The classification performance is high when both sensitivity and specificity are high.

The selected color space evaluated was the HSV, since it is more similar to the physiological perception of the human eye, and the components have a low correlation and can separate the image intensity from the color information. Since the results obtained regarding the CIELab and RGB color spaces were substantially worse, only the classification performances regarding this color space are presented in this work.

Figure 6 shows the segmentation results of two images with and without preprocessing. It is possible to infer that the segmented region with the preprocessing in both examples is more similar to the tumor area.

The results of this paper show that textural information can be adequate to classify images from cystoscopy. This textural information was obtained from the mean and variances calculated over the wavelet frame transformation of the different color bands. A two-level DWT was applied to each color channel, selecting medium and higher frequencies, medium frequencies, and a one-level discrete wavelet transformation selecting only the higher frequencies. This process was performed to verify if the higher frequencies are the ones that contain more textural information about bladder tumors. Table 1 indicates that the first scale captures more texture information than the second scale, which is reflected by a better classification performance in both the MLP and SVM classifiers. It is also possible to verify that MLP allows obtaining higher performances.

Since the initial assumption based on works done in WCE images is confirmed, only the DWT lower scale is considered. We already showed that the preprocessing step is important to improve the segmentation results, as

Table 1. Classification performance of the wavelet transform applying the preprocessing step for the MLP and SVM classifiers using the HSV color space. Performances achieved regarding scales of medium frequencies (M-F), medium and high frequencies (M + H-F) and higher frequencies (H-F) can be observed.

	MLP			SVM		
Classification vector (mean and standard deviation)	M-F	M + H-F	H-F	M-F	M + H-F	H-F
Sensitivity ($\mu \pm \sigma$%)	69.7 ± 13.1	79 ± 14.5	**91 ± 7.8**	69.4 ± 10.1	71 ± 14.0	85.1 ± 8.8
Specificity ($\mu \pm \sigma$%)	91.9 ± 9.1	93.7 ± 3.4	**92.9 ± 6.0**	89 ± 9.8	90.1 ± 4.1	82.5 ± 5.7
Accuracy ($\mu \pm \sigma$%)	87 ± 8.2	90.5 ± 2.0	**92.4 ± 5.1**	83.4 ± 9.0	85.6 ± 2.9	85.1 ± 5.3
ROC ($\mu \pm \sigma$%)	90.7 ± 8.9	91.8 ± 5.6	**96.5 ± 2.9**	80 ± 11.3	79.7 ± 6.1	83.8 ± 3.5

Table 2. Performances obtained for MLP and SVM classifiers in order to evaluate the influence of the preprocessing procedure using the HSV color space.

	MLP		SVM	
Classification vector (mean and standard deviation)	Without preprocessing	With preprocessing	Without preprocessing	With preprocessing
Sensitivity ($\mu \pm \sigma$%)	81.4 ± 12.2	**91 ± 7.8**	94 ± 4.2	85.1 ± 8.8
Specificity ($\mu \pm \sigma$%)	91.6 ± 7.8	**92.9 ± 6.0**	85.7 ± 5.2	82.5 ± 5.7
Accuracy ($\mu \pm \sigma$%)	89.8 ± 5.5	**92.4 ± 5.1**	87.6 ± 7.1	85.1 ± 5.3
ROC ($\mu \pm \sigma$%)	93.3 ± 8.0	**96.5 ± 2.9**	89.8 ± 6.9	83.8 ± 3.5

Table 3. Performances obtained for MLP and SVM classifiers to evaluate the proposed segmentation algorithm using the HSV color space.

	MLP		SVM	
Classification vector (mean and standard deviation)	Without segmentation	With segmentation	Without segmentation	With segmentation
Sensitivity ($\mu \pm \sigma$%)	69.4 ± 8.1	91 ± 7.8	79 ± 6.3	85.1 ± 8.8
Specificity ($\mu \pm \sigma$%)	90.1 ± 4.5	92.9 ± 6.0	93.7 ± 4.1	82.5 ± 5.7
Accuracy ($\mu \pm \sigma$%)	85.6 ± 6.6	92.4 ± 5.1	86.5 ± 6.0	85.1 ± 5.3
ROC ($\mu \pm \sigma$%)	79.7 ± 9.3	96.5 ± 2.9	91.8 ± 2.8	83.8 ± 3.5

can be observed in figure 6. To prove that this preprocessing step can indeed improve the detection of tumors, we also compare the results from the classification system when the preprocessing is not used.

Table 2 summarizes the efficiency of the preprocessing procedure, by extracting the mean and standard deviation, from both the normal and abnormal areas.

In this case, when the SVM is used to classify the data an increase in sensitivity was noted when preprocessing is not employed. However, neither specificity nor accuracy follow the same behavior. When the MLP is used instead, a decrease in all measures can be observed in the absence of preprocessing. Relative to the SVM, the MLP allows achieving a higher accuracy rate.

Looking at table 3, we can see that the segmentation process applied leads to better results than those obtained when features are extracted directly from the DWT without any previous steps.

It is clear from the results shown in table 3 that relative measures are more effective than absolute measures in detecting tumor tissue as heuristically expected since relative measures are more robust to environmental variations such as device or subject related. Moreover, by not only comparing the segmentation of two tumoral frames, with and without preprocessing, as shown in figures 6(b) and (c), but also their performances, as presented in tables 2 and 3, we can infer that preprocessing does indeed increase the detection rates.

Haralick descriptors are considered state-of-the-art texture descriptors, so we conducted some tests in order to compare Haralick and wavelet based texture descriptors. Table 4 shows that all metrics present a decrease when Haralick descriptors are used.

In spite of the different dimensionality in both cases, wavelet based texture descriptors seem to be significantly superior to Haralick based descriptors for the purpose of bladder tumor characterization. In order to evaluate which color space allows achieving the best results with this implementation, the performances of the HSV, RGB, and CIELab color spaces were computed, as table 5 shows. With respect to the CIELab color space the influence of the lightness in the classification task was also studied by removing the L channel.

This comparison allows to deducing that the HSV is the color space with higher capacity to detect tumor frames. Both the RGB and CIELab performances are substantially lower than those obtained with the HSV. Dis-

Table 4. MLP and SVM performances to evaluate the strength of features extracted from the HSV color space.

	MLP		SVM	
Classification vector	Haralick descriptors	Mean and standard deviation	Haralick descriptors	Mean and standard deviation
Sensitivity ($\mu \pm \sigma$%)	88.3 ± 5.9	**91 ± 7.8**	83.9 ± 3.7	85.1 ± 8.8
Specificity ($\mu \pm \sigma$%)	86.6 ± 7.1	**92.9 ± 6.0**	95.6 ± 8.4	82.5 ± 5.7
Accuracy ($\mu \pm \sigma$%)	87.3 ± 4.3	**92.4 ± 5.1**	89.8 ± 5.3	85.1 ± 5.3
ROC ($\mu \pm \sigma$%)	88.5 ± 6.8	**96.5 ± 2.9**	97.4 ± 10.8	83.8 ± 3.5

Table 5. Comparison of the classification performances between several color spaces. Detection rates were obtained for the HSV, RGB, *Lab*, and *ab* configurations, and the features were extracted from the high-frequency sub-bands of the wavelet transform with the preprocessing.

Classification vector (mean and standard deviation)	MLP				SVM			
	HSV	RGB	*Lab*	*ab*	HSV	RGB	*Lab*	*ab*
Sensitivity ($\mu \pm \sigma$%)	**91 ± 7.8**	66.1 ± 8.5	80.6 ± 6.9	64.2 ± 9.8	85.1 ± 8.8	65.7 ± 9.0	73.1 ± 8.7	61.6 ± 11.8
Specificity ($\mu \pm \sigma$%)	**92.9 ± 6.0**	93.5 ± 4.3	87 ± 5.3	85.6 ± 5.7	82.5 ± 5.7	95.5 ± 3.1	85.2 ± 4.1	69.8 ± 10.7
Accuracy ($\mu \pm \sigma$%)	**92.4 ± 5.1**	87 ± 5.0	85.5 ± 6.1	77.7 ± 7.3	85.1 ± 5.3	84.6 ± 2.5	82.4 ± 3.9	76.9 ± 8.3
ROC ($\mu \pm \sigma$%)	**96.5 ± 2.9**	91 ± 4.6	89.6 ± 3.9	87 ± 3.1	83.8 ± 3.5	77.6 ± 6.9	79.2 ± 6.6	75.3 ± 7.5

carding the lightness did not lead to better results than those achieved when all three color channels are used. As mentioned above, the MLP classifier provided higher performances.

To confirm that the best result was obtained, a confusion matrix was used in which it could be observed that 97 out of 107 were correctly predicted as abnormal frames and 228 out of 246 were correctly predicted as normal frames.

5. Conclusion

Automatic tumor detection in cystoscopy exams could constitute an important diagnostic tool that could potentially be used by urologists. In this work the proposed module takes advantage of a wavelet decomposition tree to discard low frequency information in such a way that both steps of the algorithm, segmentation, and classification regarding the HSV, RGB, and CIELab color spaces, share the same focus on texture. It was possible to conclude that the preprocessing step increased the detection rate, as did the use of the HSV color space. The experimental results show that high accuracies can be achieved making the system reliable enough to be used in clinical practice. The obtained accuracies are in the range of the accuracies of systems used in other fields in clinical practice. More effective texture descriptors such as the those based on the curvelet transform, improvements in the segmentation module, and more powerful classification approaches such as ensemble learning strategies that however require a large database that is currently under construction could increase the performance of the system.

Acknowledgments

This work is supported by FCT under Project No. UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020—Programa Operacional Competitividade e Internacionalização (POCI) under Project No. POCI-01-0145-FEDER-006941.

ORCID iDs

Nuno R Freitas https://orcid.org/0000-0003-3784-8817
Pedro M Vieira https://orcid.org/0000-0001-5458-2307

References

Alpaydın E 2014 *Introduction to Machine Learning* (Cambridge, MA: MIT)

American Cancer Society 2017 *Cancer Facts & Figures* (Atlanta, GA: American Cancer Society) (https://old.cancer.org/acs/groups/content/@editorial/documents/document/acspc-048738.pdf)
Arora R and Suman S 2012 Comparative analysis of classification algorithms on different datasets using WEKA *Int. J. Comput. Appl.* **54** 21–5
Babjuk M *et al* 2017 *EAU Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and CIS)* (Amsterdam: Elsevier) pp 1–45
Barbosa D, Ramos J and Lima C S 2008a Wireless capsule endoscopic frame classification scheme based on higher order statistics of multi-scale texture descriptors *4th European Conf. Int. Federation for Medical and Biological Engineering. IFMBE Proc.* vol 22, ed J Vander Sloten, P Verdonck, M Nyssen, J Haueisen (Heidelberg: Springer)
Barbosa D, Roupar D and Lima C S 2011 Multiscale texture descriptors for automatic small bowel tumors detection in capsule endoscopy *Discrete Wavelet Transforms—Biomedical Applications* ed H Olkkonen (Rijeka: InTech) (https://doi.org/10.5772/19391)
Barbosa D C *et al* 2012 Automatic small bowel tumor diagnosis by using multi-scale wavelet-based analysis in wireless capsule endoscopy images *Biomed. Eng. Online* **11** 1–17
Barbosa D J C, Ramos J and Lima C S 2008b Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform *30th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (Vancouver, BC)* pp 3012–5
Fielding J R *et al* 2002 Tumor detection by virtual cystoscopy with color mapping of bladder wall thickness *J. Urol.* **167** 559–62
Freitas N R *et al* 2017a Segmentation of bladder tumors in cystoscopy images using a MAP approach in different color spaces *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)* (IEEE) pp 1–4
Freitas N R *et al* 2017b Using cystoscopy to segment bladder tumors with a multivariate approach in different color spaces *Proc. of the Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, EMBS*
Gardner M and Dorling S 1998 Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences *Atmos. Environ.* **32** 2627–36
Grossman H B *et al* 2007 A phase III, multicenter comparison of hexaminolevulinate fluorescence cystoscopy and white light cystoscopy for the detection of superficial papillary lesions in patients with bladder cancer *J. Urol.* **178** 62–7
Haralick R M, Shanmugam K and Dinstein I 1973 Textural features for image classification *IEEE Trans. Syst. Man Cybern.* **smc 3** 610–21
Harrington P 2012 *Machine Learning in Action* ed J Bleil (Shelter Island, NY: Manning Publications Co.)
Karkanis S A *et al* 2003 Computer-aided tumor detection in endoscopic video using color wavelet features *IEEE Trans. Inf. Technol. Biomed.* **7** 141–52
Kociolek M *et al* 2001 Discrete wavelet transform—derived features for digital image texture analysis *Int. Conf. on Signals and Electronic Systems* (*September 2001*) pp 163–8
Kodogiannis V S *et al* 2007 The usage of soft-computing methodologies in interpreting capsule endoscopy *Eng. Appl. Artif. Intell.* **20** 539–53
Kuncheva L I 2014 *Combining Pattern Classifiers: Methods and Algorithms* 2nd edn (Hoboken, NJ: Wiley)
Li B and Meng M Q H 2009 Computer-aided detection of bleeding regions for capsule endoscopy images *IEEE Trans. Biomed. Eng.* **56** 1032–9
Lima C S *et al* 2008 Classification of endoscopic capsule images by using color wavelet features, higher order statistics and radial basis functions *Proc. 30th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC2008)* pp 1242–5
Lima C S *et al* 2010 Non-stationary biosignal modelling *New Developments in Biomedical Engineering* (InTech) pp 37–72
Lima C S *et al* 2009 Texture classification of images from endoscopic capsule by using MLP and SVM: a comparative approach *Proc. of the World Congress 2009 for Medical Physics and Biomedical Engineering* (Munich: Springer) pp 271–4
Mallat S 2008 *A Wavelet Tour of Signal Processing* 3rd edn (Cambridge, MA: Academic)
Mallat S 1989 Multifrequency channel decompositions of images and wavelet models *IEEE Trans. Acousti. Speech Signal Process.* **37** 2091–110
Maroulis D E *et al* 2008 CoLD: a versatile detection system for colorectal lesion endoscopy video-frames *J. Ethnopharmacol.* **70** 2006–9
Marsland S 2015 *Machine Learning an Algorithmic Perspective* 2nd edn, ed R Herbrich and T Graepel (Boca Raton, FL: CRC Press)
Messer J S, Chew D J and McLoughlin M A 2005 Cystoscopy: techniques and clinical applications *Clinical Techniques in Small Animal Practice* **20** 52–64
Mitropoulos D *et al* 2005 Accuracy of cystoscopy in predicting histologic features of bladder lesions *J. Endourol.* **19** 861–4
NCCN 2015 *National Comprehensive Cancer Network. Bladder Cancer* (Washington, DC: NCCN)
Pasin E *et al* 2008 Superficial bladder cancer: an update on etiology, molecular development, classification, and natural history *Rev. Urol.* **10** 31–43
Prasath V B S 2016 Polyp detection and segmentation from video capsule endoscopy: a review *J. Imaging* **3** 1–21
Rouprêt M *et al* 2011 European guidelines for the diagnosis and management of upper urinary tract urothelial cell carcinomas: 2011 update *Eur. Urol.* **59** 584–94
Sprager S and Zazula D 2014 Optimization of heartbeat detection in fiber-optic unobtrusive measurements by using maximum *a posteriori* probability estimation *J. Biomed. Inform.* **18** 1161–8
Stein B J P *et al* 2011 Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1054 patients *J. Clinical Oncology* **19** 666–75
Toufik B and Mokhtar N 2012 The wavelet transform for image processing applications *Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology* ed D Baleanu (Rijeka: InTech) (https://doi.org/10.5772/35982)
Vieira P *et al* 2012 Segmentation of small bowel tumor tissue in capsule endoscopy images by using the MAP algorithm *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 4010–3
Vining J *et al* 1995 CT Cystoscopy: an innovation in bladder imaging *Am. Roentgen Ray Soc.* **166** 409–10
Witjes J A *et al* 2014 Guidelines on muscle-invasive and metastatic bladder cancer *European Urology* **65** 778–92
Yahia S *et al* 2017 Comparison between extreme learning machine and wavelet neural networks in data classification *Proc. SPIE* **10341** 103412K