Feature Detection and Description for Endoscopy feature extraction: A Survey

Hoang Khoi Do[1]

[1]School of Electrical and Electronics, University of Science and Technology, Hanoi 10000, Vietnam

October 25, 2022

Abstract

Automated analysis of the gastric lesions in endoscopy videos is a challenging task and dynamics of the gastrointestinal environment make it even more difficult. In computer-aided diagnosis, gastric images are analyzed by visual descriptors. Various Deep Convolutional Neural Network (DCNN) models are available for representation learning and classification. Besides, using feature descriptor is also a great performance method. Some state-of-the-art method includes GLCM, BDIP, BVLC, etc. In this report, a various researches will be discussed. After all, a summary table is presented with research factors including: proposed method, performance results, etc.

1 Introduction

2 Feature Detection and Descriptor

2.1 Image Processing

2.1.1 Image Registration

Image registration is defined as a process that overlays two or more images from various imaging equipment or sensors taken at different times and angles, or from the same scene to geometrically align the images for analysis.

2.2 Vector Representation

2.2.1 Histogram-based feature extraction

Since, Histogram-based feature extraction (HBFE) have been used in various researches, in this survey, 5 popular ways of using histogram are stated.

- Luminance histogram: The RGB color channels are combined into a 8 bpp luminance channel Y in the YUV color space model. Subsequently a 1-dimensional, 256 bin histogram is constructed using the luminance values.

- Color channel histograms: a 1-dimensional, 256 bin histogram is constructed for each of the three color channels RGB independently.

- Color histogram: a 3-dimensional histogram is constructed where the three color channels represent the three dimensions, each dimension using 256 bins. For each pixel, the entry in the histogram is incremented according to its respective R, G, and B value.

- Co-occurrence histogram (CH): The co-occurrence histogram [1] counts the number of pairs of pixels exhibiting specific color or luminance values that occur at certain separation distances in image space. Therefore, the CH adds geometric information to the classical color histogram, which abstracts away all geometry. By adjusting the distances over which we check cooccurrences, we can adjust the sensitivity of the algorithm. We use horizontal and vertical adjacency as separation distance and construct the CH out of the luminance channel Y. This results in a 2-dimensional histogram where the luminance values of the two adjacent pixels represent the two dimensions, each with 256 bins.

- Using CH to compute further information of a picture (see Equation).

$$Uniformity = \sum_{i,j} p_{i,j}^2 \tag{1}$$

$$Constrast = \sum_{i,j} |i-j|^k (p_{ij})^2 \tag{4}$$

$$Entropy = \sum_{i,j} p_{i,j} \log P(i,j) \tag{2}$$

$$Inverse\ difference\ moment = \sum_{i,j} \frac{(p_{ij})^2}{|i-j|^k} \tag{5}$$

$$Maximum porbability = max(p_{i,j}) \tag{3}$$

$$Correlation = \sum_{i,j} \frac{(i-u)(j-u)p_{ij}}{\sigma^2} \tag{6}$$

The problem of using histogram vector is that histogram feature vector is not a vector in Euclidean Space. Therefore a histogram distance formula is proposed as the intersection between two distribution. Consider two distribution histogram $H(I)$ and $H(I\prime)$ calculated from two images I and $I\prime$. Since $H(I)$ and $(I\prime)$ have the same bin n, the normalized histogram is calculated in the Equation 7:

$$H(I) \cap H(I\prime) = \sum_{j=1}^{n} min(H_j(I), H_j(I\prime)) \tag{7}$$

2.2.2 Wavelet-based Feature Extraction

The (one-dimensional) DWT operates on a real-valued vector x of length 2^n, $n \in 2, 3, ...$, and results in a transformed vector w of equal length. Figure 1(a) and 1(b) illustrate the first two steps of the DWT for a vector of length 16. First, the vector x is filtered with some discrete-time, low-pass filter (LPF) h of given length at intervals of two, and the resulting values are stored in the first eight elements of w (see Figure 1). Otherwise, the low-pass filtered part of w (first eight elements for this example) can be further transformed using the identical procedure as outlined above. (see Figure 2) Filter coefficients for the low, high pass filter vector denoted h, g respectively, should be related in Equation 8

$$g_k = (-1)^k h_{n-k-1}, k \in 0, \ldots, n-1 \tag{8}$$

where n denotes the length of the filter (see Equation 9 and 10).

$$h = \begin{bmatrix} c_0 & c_1 \end{bmatrix} \Rightarrow g = \begin{bmatrix} c_1 & -c_0 \end{bmatrix} \tag{9}$$

$$h = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 \end{bmatrix} \Rightarrow g = \begin{bmatrix} c_3 & -c_2 & c_1 & -c_0 \end{bmatrix} \tag{10}$$

The simplest wavelet filter is the Haar filter (see Equation 11)

$$h = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{11}$$

Another very popular set of wavelet filters is due to Daubechies. The most compact of these has four coefficients (Daubechies-4), where h is given by:

$$h = \begin{bmatrix} \frac{1+\sqrt{3}}{4\sqrt{2}} & \frac{3+\sqrt{3}}{4\sqrt{2}} & \frac{3-\sqrt{3}}{4\sqrt{2}} & \frac{1-\sqrt{3}}{4\sqrt{2}} \end{bmatrix} \tag{12}$$

Computing the two-dimensional DWT is conducted through repeated application of the one-dimensional DWT. Figure 3 illustrates the basic, one-level, two-dimensional DWT procedure. First, a one-level, one-dimensional DWT is applied along the rows of the image. Second, another one-level, one-dimensional DWT is applied along the columns of the transformed image from the first step. As depicted in Figure 4 (left), the result of these two sets of operations is a transformed image with four distinct bands: (1) LL, (2) LH, (3) HL and (4) HH. Here, L stands for low-pass filtering, and H stands for high-pass filtering. The LL band corresponds roughly to a down-sampled (by a factor of two) version of the original image. The LH band tends to preserve localized horizontal features, while the HL band tends to preserve localized vertical features in the original image. Finally, the HH band tends to isolate localized high-frequency point features in the image.

2.2.3 Gray Level Co-occurence Matrix

Features extraction technique of GLCM is employed for textural information. It is a technique that allows for the extraction of statistical information from the image regarding the pixel pair distributions. GLCM is the most popular method for texture analysis and were demonstrated to feature a potential for effective texture discrimination. GLCM is defined by

$$P_d[i,j] = n_{i}j \tag{13}$$

where n_{ij} is the number of pair pixel occurrences (see Figure 5). It is computed by defining a direction, θ and a distance, d. Values (i,j) lying at distance d in the image. Then a pairs of pixels separated by d, computed across the defined θ are analyzed. As suggested in [2], directions of 0°, 45°, 90°, and 135° and the distance of 1 pixel is used in this research. The directions are chosen by assuming that the matrix is symmetrical. Symmetrical means the pixel pair is separated by d and –d. So, the count will consider the pixel in the direction opposite to d. Each pixel pair is computed to get a co-occurrence matrix. There are fourteen textural feature can be gathered from GLCM.

$$\text{Angular Second Moment: } f_1 = \sum_i \sum_j \{p(i,j)\}^2 \tag{14}$$

$$\text{Contrast: } f_2 = \sum_{n=0}^{N_g-1} \{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\} \, | \, p(i,j) \, (|i-j|=1) \tag{15}$$

$$\text{Correlation: } f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{16}$$

$$\text{Sum of Squares - Variance: } f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j) \tag{17}$$

$$\text{Inverse Difference Moment: } f_5 = \sum_i \sum_j \frac{1}{1+(i+j)^2}p(i,j) \tag{18}$$

$$\text{Sum Average: } f_6 = \sum_{i=2}^{2N_g} ip_{x+y}(i) \text{ where } p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \, D[k] = \{a \in N^* | 2 \le a \le 2N_g\}, \, i+j=k \tag{19}$$

$$\text{Sum Variance: } f_7 = \sum_{i=2}^{2N_g} (i-f_8)^2 p_{x+y}(i) \tag{20}$$

$$\text{Sum Entropy: } f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i)log\{p_{x+y}(i)\} \tag{21}$$

$$\text{Entropy: } f_9 = -\sum_i \sum_j p(i,j)log\{p(i,j)\} \tag{22}$$

$$\text{Difference Variance: } f_{10} = Var[p_{x-y}], where p_{x-y}(k) = p_{x+y}(k) \, for \, D[k] = \{a \in N^* | 0 \le a \le N_g-1\} |i-j|=k \tag{23}$$

$$\text{Difference Entropy: } f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i)log\{p_{x-y}(i)\} \tag{24}$$

$$\text{Difference Entropy: } f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i)log\{p_{x-y}(i)\} \tag{25}$$

$$\text{Information Measure of Correlation (1): } f_{12} = \frac{H[XY] - H[XY1]}{max\{H[X], H[Y]\}} \tag{26}$$

$$\text{Information Measure of Correlation (2): } f_{13} = (1 - \exp[-2.0(H[XY2] - H[XY])])^{1/2} \tag{27}$$

$$\text{Maximal Correlation Coefficient: } f_{14} = (Second largest eigenvalue of Q)^{1/2} \tag{28}$$

where $p(i,j)$ is $(i,j)th$ entry in a normalized gray-tone spatial dependence matrix, $= P(i,j)/R$. $p_x(i)$ is ith entry in the marginal-probability matrix obtained by summing the row of $p(i,j)$, $= \sum_{j=1}^{N_g} P(i,j)$. N_g is the number of distinct gray levels in the quantized image. In Equation 26 and 27, $H[X,Y] = -\sum_i \sum_j p(i,j)log(p(i,j))$, $H[XY1] = -\sum_i \sum_j p(i,j)log(p_x(i)p_y(j))$, and $H[XY2] = -\sum_i \sum_j p_x(i)p_y(j)log(p_x(i)p_y(j))$. in Equation 28, $Q(i,j)) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(j)}$.

2.3 Feature Map Representation

2.3.1 Higher order Local Auto-correlation (HLAC)

The HLAC features [3], F, represent the expressed characteristics for a complete image, derived from the product-sum operations of the following formula that represents auto-correlations(see Equation 29).

$$F_n = \sum_r I(r) \cdot I(r + a_i) \ldots I(r + a_n) \tag{29}$$

where N is the order of auto-correlation, r is the $X - Y$ coordinate vector of the colonoscopy image, $I(r)$ is the pixel value at r, and $a_i (i = 1 \ldots M)$ are displacement vectors around r. 25 masks are formed by the configurations of r and a_i, restricted by the second orders of autocorrelation ($N = 0, 1, 2$) within a 3x3 area around r.

2.3.2 Integral image techniques

In integral image techniques [4], The value G at any $X - Y$ coordinate $r(x, y)$ in the integral feature table is the sum of all the auto-correlations f at r (see Equation 30)

$$G(r(x,y)) = \sum_{x\prime \leq x} \sum_{y\prime \leq y} f(r(x\prime, y\prime)) \tag{30}$$

2.3.3 Local binary pattern (LBP)

Local binary pattern (LBP) texture operator, which was first proposed by Ojala et al. The LBP feature is calculated as shown in Equation 31.

$$L(x, y) = \big[I(x + p_i, y + p_i) \big] \cdot D \tag{31}$$

where $L(x, y)$ represent the LBP value at position (x, y), $\big[I(x + p_i, y + p_i) \big]$ is the intensity vector of size 8, representing the intensity of 8 cell around the center. D is the binomial weights vector ($\begin{bmatrix} 2 & 4 & 6 & 8 & \ldots \end{bmatrix}^T$)

2.3.4 Block Difference of Inverse Probabilities (BDIP)

2.3.5 Block Variation of Local correlation Coefficients (BVLC)

3 Related Works

There has been many researches using feature descriptor over the last decade. Hirokazu et al proposes a method of retrieving multi-scale objects from optical colonoscopy images based on image recognition techniques [5]. The proposed method is a method of content-based image retrieval (CBIR). [5] improves the geometric feature extraction component in order to handle the retrieval of multi-scale objects by utilizing the integral image technique [4]. In this research, both object size and object color are considered as points of contrast, because the illumination conditions for each image are a function of the focal distance. For preprocessing, we adopt the saturation element converted from the original colonoscopy images for the RGB color space. In addition, bright blobs within colonic images are interpolated by a method of image inpainting [6] prior to conversion, because the saturation element cannot be converted from bright blobs. In this study, a feature extraction method formed by the combination of HALC [3] and Integral Image Technique [6] are proposed.

Using the same technique, Sergio et al proposed e a minimally invasive technique based on multispectral imaging and a methodology to identify malignancies in the stomach [7]. First, in order to collect the image, the authors decide to create a new gastroendoscopic system that use six different wavelength including 440, 480, 520, 560, 600 and 640nm. The sequence of the monoband images are extracted from the gastroendoscopic video deinterlaced by Yadif algorithm [8]. Then, the captured image at each wavelength is extracted semiautomatically. Second, these monoband images need to be put into a registration process due to the shifting of observed tissue. Before applying image registration, an image contrast enhancement is used to improve the quality of image by using CLAHE algorithm [9]. Then, hierarchical motion-based estimation is used [10] for the registration. Third, the image is extracted and filtered by using t-distribution. Finally, the image data is fed into three model including K nearest neighbor classifier (k-nn), Support vector machine classifier (SVM), and The neural networks using generalized relevance learning vector quantization (NN). In order to evaluate the performance of each classifier leave one patient out cross-validation (LOPOCV) is applied. The SVM, NN, and k-nn achieved the accuracy of 0.77, 0.72, 0.62, sensitivity of 0.91, 0.85, 0.61, and Specificity of 0.62, 0.59, 0.63, respectively.

Statistics calculated from the image's histogram are also widely used for the representation of color features. The histogram features of endoscopic images are used for representing the texture of images and combined with

wavelet-based texture features to classify frames with celiac disease [11]. T. F. A. et al uses histogram-based feature extraction which includes three methods: 1-dimensional single color channel histograms, 2-dimensional Co-occurrence histograms and a 3-dimensional color histogram combining information from all three color channels. Otherwise, a feature vector consisting of 10 descriptive statistical texture measures of the intensity distributions is also proposed. 5 first-order measures (Mean, Variance, Standard Deviation, Skewness, Kurtosis) determined from each of the 1-dimensional color-channel histograms and 5 second-order measures (Entropy, Energy, Inverse Difference Moment, Contrast, Covariance as defined in [12], also see Equations 1, 5, 2, 4, and 6) calculated from the corresponding co-occurrence histograms. Wavelet-based feature extraction, otherwise is also applied for feature extraction. Three methods include features based on Pyramidal wavelet decomposition, Best-Basis decomposition, and Local Discriminant Bases [13].

Gabor filters can be used for the multi-resolution analysis of gastric images and rotation invariant texture descriptors extracted by exploiting the auto-correlation property of Gabor filters (GF). Moreover, naïve Bayes (NB) and SVM classifiers used for classification [14]. Farhan et al proposed a novel features based on GF which are invariant to rotation, scaling, and illumination. Invariance properties of GF is theoretically explained. A region-based descriptor (auto-correlation homogeneous texture—AHT), otherwise, is also proposed. For classification, novel AGF texture features in a classification framework which is based on textons: texton-AGF is used. After conducting the experiment, AHT takes the highest position in classifying Chromoendoscopy image while Texon-AGF hits a peak of 0.88 accuracy, 0.05 false positive, 0.872 precision, 0.875 recall, and 0.945 AUC.

On the other hand, Wavelet-based LBP texture obtained from WCE images for detection of cancer frames [15]. Baopu et al proposed a method that classify Tumor in Capsule Endoscopy (CE) Image using SVM-based Feature Selection. In the feature extraction phase, two main methods are proposed including Local binary pattern (LBP) texture operator and discrete wavelet transform (DWT). Otherwise, feature extraction methods in RGB space, HSI space and YCbCr space are also proposed to gain more feature information for further analysis. Then the feature representation of CE image is the combination of feature extracted from the three color space. For classification task, the authors used Support Vector Machine (SVM). Another research that used Wavelet-based LBP texture for Wireless Capsule Endoscopy is [16].

Similarly, the color wavelet covariance computed for texture representation of WCE images in [17]. B. Li et al proposed method of using curve wavelet instead of traditional wavelet transform. Local Binary Pattern (LBP), otherwise is also applied for feature extraction. The Wireless Capsule Endoscopy images, on the other hand are transformed into YCbCr color space which is a color space widely used in video and digital photography systems in order to gather further information for analysis. Finally Multi-layer Perceptron and Support Vector Machine algorithm are applied for classification task.

Pit-patterns of mucosal surface analyzed by computing Gabor wavelet and dual-tree complex wavelet transforms for the detection of cancer is discussed in [18]. Each image of the database is decomposed by the Gabor wavelet transform (GWT) [19] and Kingsbury's dual-tree complex wavelet transform (DT-CWT) [20]. Before actually transforming the images, image-quality enhancing pre-processing procedures are also applied. First CLAHE [9] with 8×8 tiles and an uniform distribution is applied for constructing the contrast transfer function. Second, a Gaussian blur with $r = 0.5$ using a 3×3 mask is applied. Finally, Leave only one cross validation (LOOCV) strategy is applied for the classification tasks.

A completely different approach is presented in [19], [21], where the authors compute a set of texture-descriptors from the outputs of the discrete cosine transform (DCT) and the discrete Fourier transform (DFT). Features are either computed from non-overlapping pixel blocks in the DCT domain or from adaptively sized rings in the Fourier domain. Concatenation of the feature vectors of each color channel is then used to incorporate color information. The authors employ a Bayes normal classifier together with feature subset selection to classify the endoscopy images. The best reported LOOCV accuracy in the two-class problem is 97.7 and 86.36 in the six-class problem.

Gray-level and global texture features are extracted for the early detection of carcinoma from endoscopy frames by using a higher-order graph matching kernel SVM classifier in [17]. Zhihong et al proposed a method of representing texture features of early esophageal carcinoma in Endoscopic ultrasonography (EUS) images as a graph by expressing pixels as nodes and similarity between the gray-level or local features. Then, similarity measurements such as a high-order graph matching kernel can be constructed so as to provide an objective quantification of the properties of the texture features of early esophageal carcinoma in EUS images. First the Region of Interest (ROI) is extract from each image, then a h-layer graph representation is created from each ROI where nodes are found by SIFT Algorithm [22]. Second, a first order feature point matching is conducted as course of matching results. These results are refined by higher order graph matching. Finally, a kernel computation based on high order graph matching results is conducted and play as a kernel of Support Vector Machine (SVM) classifer. Overall, the new kernel SVM achieved an accuracy of 0.93.

Similarly, the GLCM with a fusion of color features is extracted from endoscopic frames in [23] for the detection of stomach gastritis and used SVM for training and classification of endoscopic-frames. Before classifying, image enhancement, image cropping, and color conversion are applied for pre-processing procedure. In feature extraction, two-level DWT is used before applying Gray Level Co-occurence Matrix (GLCM) to calculate six features (contrast,

correlation, dissimilarity, angular second moment, entropy and inverse different moment). Then the feature is fed into SVM with various kernel for classification task. To the end, the method of using the combination wavelet transform, Gray Level Co-occurence Matrix, and Color Moment hits a peak of 0.87 accuracy.

GLCM texture features combined with temporal features in [24] for retrieval of images with similar clinical conditions from gastric images database. Color co-occurrence can compute the color-texture features from the wireless capsule endoscopy (WCE) image for bleeding detection by assuming the blood has a texture. In this research, Barbara et al proposed a method of using SIFT algorithm, a scale invariant feature descriptor to extract interest points from WCE images. Then the GLCM and K Mean Clustering is applied to find K visual words in Bag-of-Visual-Word (BVW) method. Finally, K-nearest Neighbor is applied for classification task. The KNN classifier whose number of neighbor is 10 outperform all other methods with the accuracy of 0.76.

Moreover, the dominant color features computed from hue, saturation, and value (HSV) frequency of images [25]. In this research, Balathasan proposed a method of blood detection from capsule endoscopy videos. As part of preprocessing the bounding black region in CE frames is removed, which is outside the field-ofview of the camera. Then an averaging filter was used to remove random noise, which were present in some of the videos. And finally, frames which are poorly illuminated or over-exposed are removed as these frames do not provide any useful information. The second feature extracted from images is the dominant color using dominant color descriptor [26]. A third image feature extracted is the co-occurrence of the dominant colors. For co-occurrence matrix, the cooccurrence for a window of 5×5, for the top 8 dominant colors colors were computed. The top 8 dominant colors were picked from the positive frames in the training set. Finally, the Support Vector Machine (SVM) is used for classification.

In the same way, LBP texture with the color histogram features used to classify the endoscopy frames having cancerous regions in [27]. SU et al proposed a method of abnormal region detection in gastroscopic images. For feature extraction of gastroscopic image, the color histogram is used to represent the color feature and the uniform Local Binary Pattern (LBP) [28] histograms to represent the structural (textural) feature of gastroscopic images. Three kinds of regions are considered as the invalid regions in gastroscopic image besides the non-imaging region: the dark region, the text-stained region and the specular reflectance region. The dark region cannot provide the reliable information which is defined as the region with intensity less than a threshold (40 in this paper). The text-stained regions can be simply removed by a predefined mask since the positions of texts are fixed. the reflection regions can be enhanced by multiplying intensity with the saturation stated in [27]. For abnormal region detection task, boosted dicision stump is used with a modified GentleBoost algorithm [29].

In [30], two new approach to GLCM is proposed including Deep Gray-Level Co-occurrence Matrix (DeepGLCM) and Local-Gray-Level Co-occurrence Matrix (L-GLCM). Instead of using the whole image and feed into the AlexNet, the statistical feature including Energy, Homogeneity, Contrast, and Correlation computed from GLCM is used. In this research, the Kvasir data set is used, which includes 7 classes: Esophagitis, Dyed and Lifted Polyp, Dyed Resection Margin, Cecum, Polyrus, Z-line, Polyps, and Ulcerative colitis. For the whole process, AlexNet is used for feature extraction, then the feature extracted is put into GLCM to calculate statistical feature. Finally, SVM is used for classification task.

References

[1] P. Chang and J. Krumm., "Object recognition with color cooccurrence histogram.," *Proceedings of CVPR*, 1999.

[2] K. S. R. M. Haralick, "Textural features for image classification," *IEEE Trans. On Systems, Man and Cybernatics*, 1973.

[3] N. Otsu and T. Kurita, "A new scheme for practical flexible and intelligent vision systems," *" Proc. of IAPR Workshop on Computer Vision*, 1988.

[4] C. Franklin, "Summed-area tables for texture mapping," *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 1984.

[5] E. T. Hirokazu Nosato, Hidenori Sakanashi and M. Murakawa, "Method of retrieving multi-scale objects from optical colonoscopy images based on imagerecognition techniques," *IEEE Xplore*, 2015.

[6] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, 2004.

[7] M. B. J.-F. E. F. M. D. L. F. G. Sergio E. Martinez-Herrera, Yannick Benezeth, "Identification of precancerous lesions by multispectral gastroendoscopy," *SIViP*, 2015.

[8] U.-A. W. G. V. A. Hegenbart, S., "On the effects of de-interlacing on the classification accuracy of interlaced endoscopic videos with indication for celiac disease," *IEEE*, 2013.

[9] K. Zuiderveld, "Contrast limited adaptive histograph equalization," *Graphic Gems IV*, 1994.

[10] A.-P. H. K. H. R. Bergen, J., "Hierarchical model-based motion estimation," *Computer Science*, 1992.

[11] T. F. A. V´ecsei1 and A. Uhl2, "Towards automated diagnosis of celiac disease by computer-assisted classification of duodenal imagery," *International Conference on Advances in Medical, Signal and Information Processing*, 2008.

[12] R. M. Haralick., "Statistical and structural approaches to texture.," *Proceedings of the IEEE*, 1979.

[13] M. Wickerhauser., "Adapted wavelet analysis from theory to software," *Wellesley*, 1994.

[14] F. R. F. B. S. M. D. R. M. T. Coimbra, "Invariant gabor texture descriptors for classification of gastroenterology images," *IEEE Xplore*, 2021.

[15] B. L. M. Q.-H. Meng, "Tumor ce image classification using svm-based feature selection," *IEEE Xplore*, 2010.

[16] B. L. M. Q.-H. Meng, "Small bowel tumor detection for wireless capsule endoscopy images using textural features and support vector machine," *IEEE Xplore*, 2009.

[17] M.-H. M. B. Li, "Texture analysis for ulcer detection in capsule endoscopy images," *Image and Vision Computing 27*, 2009.

[18] A. U. A. G. F. W. A. V. Michael Hafner, Roland Kwitt, "Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images," *Elsevier*, 2009.

[19] M. BS and M. WY, "Texture features for browsing and retrieval of image data.," *IEEE Trans Pattern Anal Mach Intell*, 1996.

[20] K. N, "The dual-tree complex wavelet transform: a new technique for shift-invariance and directional filters.," *Proceedings of the 8th IEEE DSP Workshop*, 1998.

[21] P. H. R. R. W. F. G. A. V. A. U. A. Ha¨fner M, Brunauer L, "Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images," *IEEE International Symposium on Computer-Based Medical Systems*, 2007.

[22] P. R. E. R. H. Zhihong Zhang, Lu Bai, "Distinctive image features from scale-invariant key points," *Int J Comput Vis*, 2008.

[23] Z. Sobri and H. A. M. Sakim, "Texture color fusion based features extraction for endoscopic gastritis images classification," *International Journal of Computer and Electrical Engineering*, 2012.

[24] A. P. A. M. B. M. B. W. . N. A. Barbara André, Tom Vercauteren, "Introducing space and time in local feature-based endomicroscopic image retrieval," *MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support*, 2009.

[25] J. L. B. B. J. O. S. J. T. Balathasan Giritharan, Xiaohui Yuan, "Bleeding detection from capsule endoscopy videos," *30th Annual International IEEE EMBS Conference*, 2008.

[26] J. L. B. B. J. O. S. J. T. Balathasan Giritharan, Xiaohui Yuan, "Circuits and systems for video technology," *IEEE Transactions*, 2001.

[27] Y.-L. W. R. Y. S.-D. C. SU ZHANG, WEI YANG, "Abnormal region detection in gastroscopic images by combining classifiers on neighboring patches," *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics*, 2009.

[28] A. A. S. Zidonghua Xuebao, "Research and perspective on local binary pattern," *Elsevier*, 2013.

[29] T. H. J. Friedman and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Annals of Statistics*, 2000.

[30] M. Y. M. H. R. Hussam Ali, Muhammad Sharif, "A shallow extraction of texture features for classification of abnormal video endoscopy frames," *Biomedical Signal Processing and Control*, 2022.

A   Image Appendix

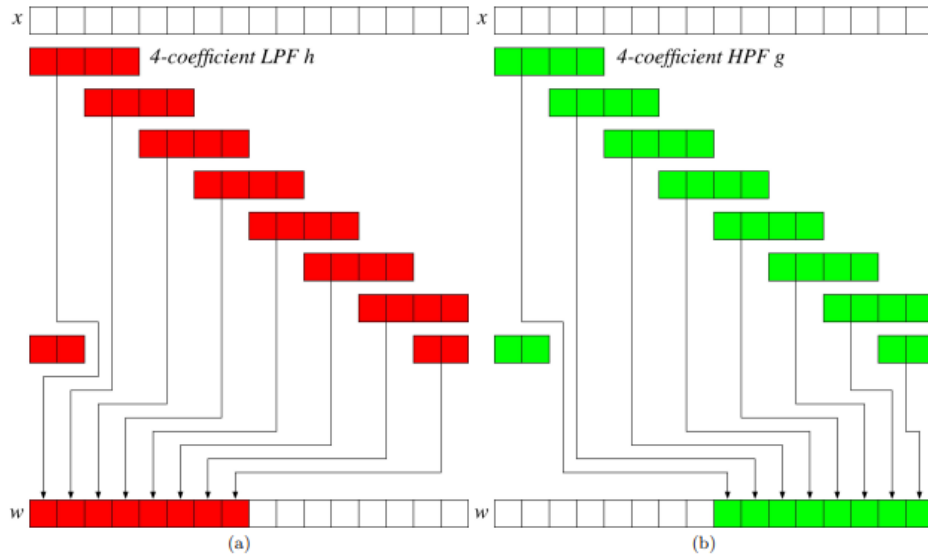Figure 1: Wavelet Transform step. (a) First step of the DWT for a signal of length 16. (b) Second step of the DWT.

Figure 2: Three-level wavelet transform on signal x of length 16. Note that from w_1 to w_2, coefficients H_1 remain unchanged, while from w_2 to w_3, coefficients H_1 and H_2 remain unchanged.
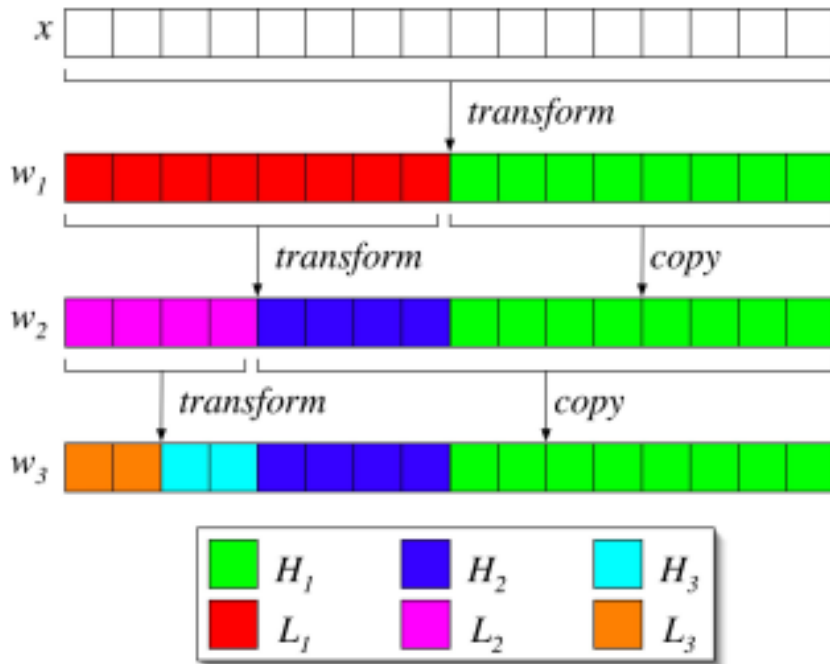
Figure 3: One-level, two-dimensional DWT. First, the one-dimensional DWT is applied along the rows; second, the one-dimensional DWT is applied along the columns of the first-stage result, generating four sub-band regions in the transformed space: LL, LH, HL and HH.
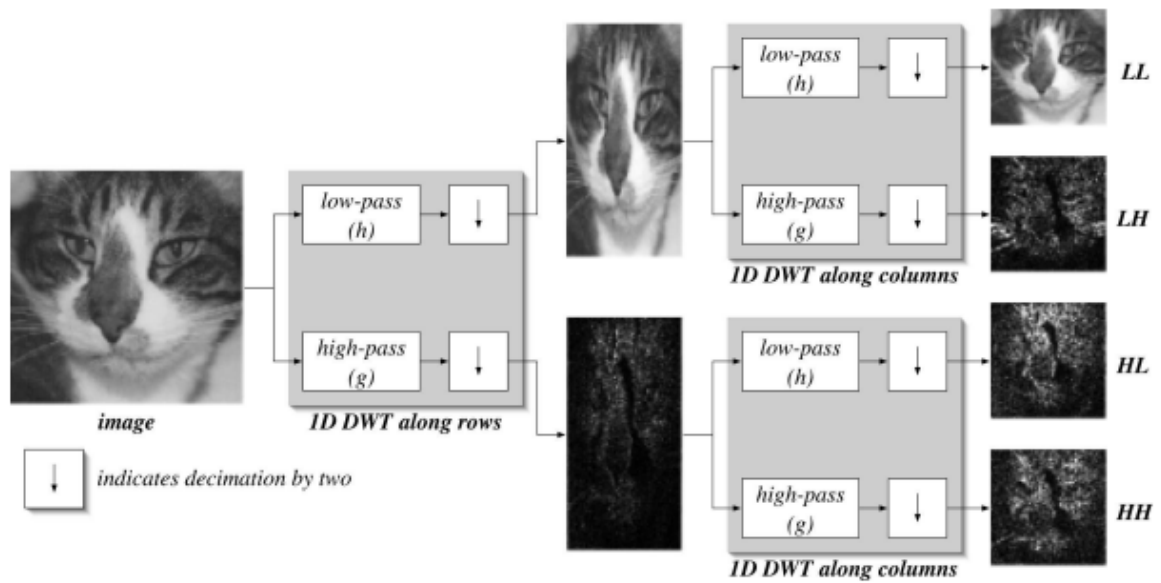
Figure 4: Two-dimensional wavelet transform: (left) one-level 2D DWT of sample image, and (right) threelevel 2D DWT of the same image. Note that the LH bands tend to isolate horizontal features, while the HL band tend to isolate vertical features in the image.
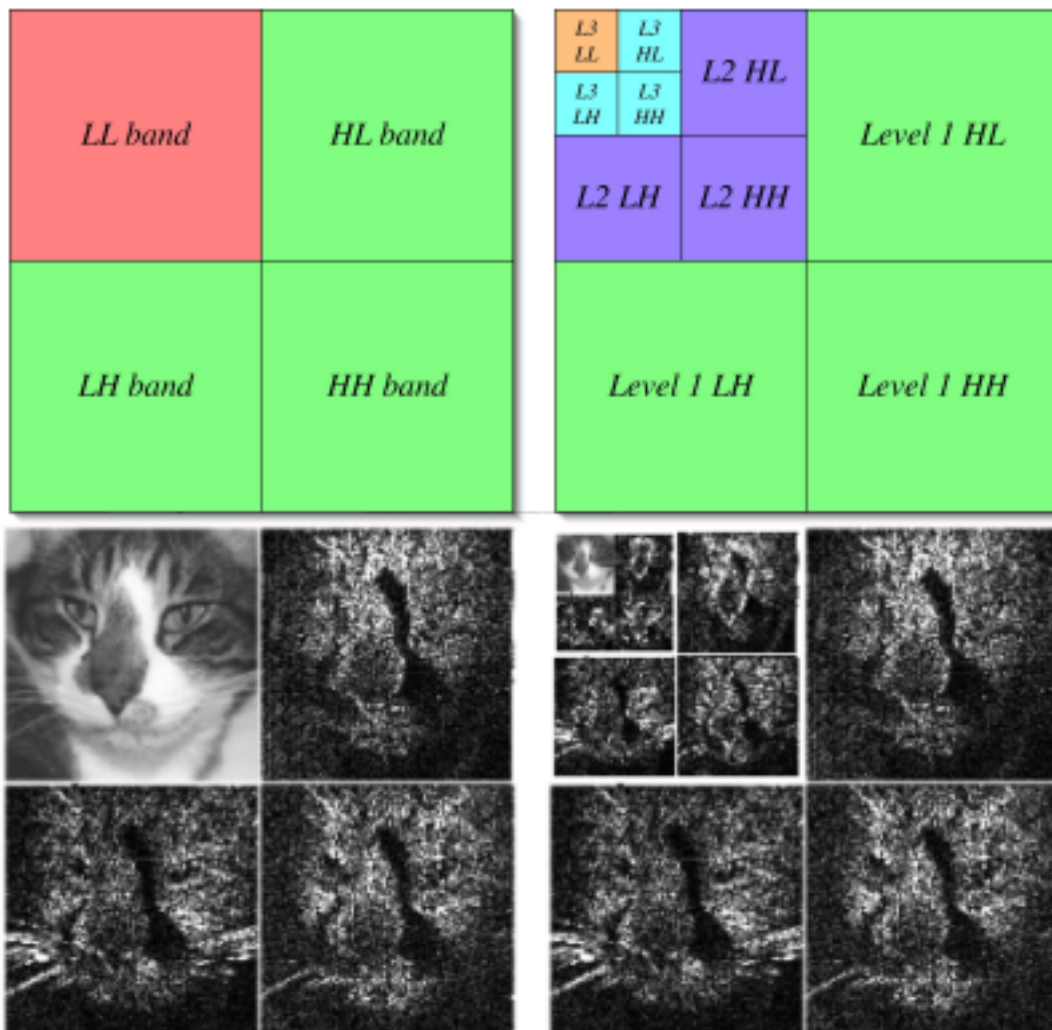
Figure 5: (a) Image example, (b) construction of co-occurrence matrix (c) matrix framework for 0°, (d) matrix framework for 45°, (e) matrix framework for 90°, (f) matrix framework for 135°

0	0	1	2
0	2	2	3
2	3	3	2
1	1	2	3

(a)

i/j	0	1	2	3
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(b)

2	1	1	0
1	2	2	0
1	2	2	4
0	0	4	2

(c)

2	0	0	0
0	0	1	2
0	1	6	1
0	2	1	2

(d)

2	0	2	0
0	0	2	1
2	2	0	6
0	1	6	0

(e)

0	0	2	1
0	0	1	1
2	1	2	2
1	1	2	2

(f)

B   Table Appendix