Color Wavelet Cross Co-Occurrence Matrices for Endoscopy Image Classification

Roland Kwitt
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: rkwitt@cosy.sbg.ac.at

Andreas Uhl
Department of Computer Sciences
University of Salzburg
Salzburg, Austria
Email: uhl@cosy.sbg.ac.at

***Abstract*— In this work we propose a set of new wavelet-domain based color-texture features for the classification of zoom-endoscopy images in the field of medical imaging. We extend the concept of classic co-occurrence matrices to capture information between detail-subband pairs of different color channels. Our results show, that the proposed features outperform other wavelet-domain based color-texture features in terms of leave-one-out cross-validation accuracy.**

I. Introduction

Recent statistics of the American Cancer Society[1] reveal that colorectal cancer is the third most common cancer in men and women and the third most common cause of US cancer deaths. Since most colorectal cancers develop from polyps, a regular inspection of the colon is recommended in order to detect lesions with a malignant potential or early cancer. A common medical procedure to examine the inside of the colon is colonoscopy, which is usually carried out with a conventional video-endoscope. A diagnostic benefit can be achieved by employing so called zoom-endoscopes, which achieve a magnification factor of up to 150 by means of an individually adjustable lens. In combination with dye-spraying to enhance the visual appearance (chromo-endoscopy) of the colon mucosa, zoom-endoscopy reveals characteristic surface patterns, which can be interpreted by experienced physicians. Commonly used dyes are either methylene-blue, or indigo-carmine, which both lead to a plastic effect. In the research work of Kudo [1], [2], the macroscopic appearance of colorectal polyps is systematically described and is known as the so called pit-pattern classification scheme.

To allow computer-assisted pit-pattern classification, we search for a set of image descriptors, by means of which we can discriminate the various pit types (see Section II). Generally, commonly known texture descriptors are used to serve this purpose. These include color histograms [3] or first-order statistics from wavelet coefficients [4] of grayscale images. Color information in the wavelet-domain is exploited in [5] by means of classic co-occurrence matrices and second order statistics. Other approaches in the field of color-texture discrimination include wavelet energy correlation signatures [6] or Gabor opponent features [7], which are basically the same, but reside in different transform domains. In this work, we introduce a set of new color-texture descriptors, based on second-order statistics from cross co-occurrence matrices [8], [9], computed between wavelet subbands of different color channels.

The remainder of the paper is structured as follows: Section II briefly discusses the pit-pattern classification scheme and introduces the corresponding classification problems. In Section III, we review the concept of co-occurrence matrices and introduce our proposed extension. Section IV presents the experimental results of our work, including a comparative study to similar approaches. Finally, Section V concludes the paper with a short summary and an outlook on further research.

II. Pit-Pattern Classification

Polyps of the colon are a frequent finding and are usually divided into metaplastic, adenomatous and malignant. Since the resection of all polyps is rather time-consuming, it is imperative that those polyps which warrant resection can be distinguished. Furthermore, polypectomy[2] of metaplastic lesions is unnecessary and removal of invasive cancer may be hazardous. The classification scheme of Kudo divides the mucosal crypt patterns into five groups (pit-patterns I-V, see Figure 1).

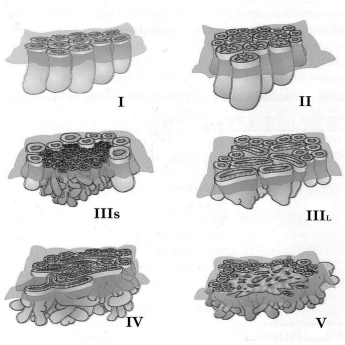

Fig. 1. Pit-Patterns I-V

While types I and II are characteristic of benign lesions and represent normal colon mucosa or hyperplastic polyps, types

[1] http://www.cancer.org

[2] the process of removing polyps

III-V represent neoplastic, adenomatous and carcinomatous structures. Our classification problem can be stated as follows: the problem to differentiate pit-types I and II from III-V will be denoted as the two-class problem, whereas the more detailed discrimination of pit-types I-V will be denoted as the six-class problem. Note, that pit-type III is subdivided into types III-S/III-L and thus accounts for two classes.

III. Feature Extraction & Classification

Since our proposed feature extraction approach will be based upon co-occurrence matrices, we briefly review this concept here. Lets establish some notation first. We consider our images as matrices, given by $\mathbf{C}^0 = \{c^0_{ik}\}_{0 \leq i,k \leq n-1}$, with c^0_{ik} denoting the intensity of a pixel at location (i,k). For simplicity, the location is often written as a lowercase, boldface variable, such as $\mathbf{x}$ or $\mathbf{y}$. All uppercase, boldface variables will denote matrices. The superscript 0 signifies that we are working with the original, unmodified image. In case of vector images (e.g. color images), we extend this notation by another superscript p or p' to signify the image plane. If we work on RGB images for instance, we have $p, p' \in \{R, G, B\}$. Any further notation will be introduced when necessary.

A. Co-Occurrence Matrices

The co-occurrence matrix $\mathbf{M}^p_{\mathbf{d}}(i,j)$ at position (i,j) captures the joint occurrence of intensity values i and j separated by the displacement vector $\mathbf{d} \in \mathbb{N}^2$. The displacement vector thus implicitly defines the orientation and the distance of considered pixel pairs. Formally, $\mathbf{M}^p_{\mathbf{d}}$ is defined as

$$\mathbf{M}^p_{\mathbf{d}}(i,j) = \mathbb{P}\left(c^0_{\mathbf{x}} = i \wedge c^0_{\mathbf{y}} = j | \mathbf{x} - \mathbf{y} = \mathbf{d}\right) \tag{1}$$

This classic formulation of the co-occurrence matrix captures information of intensity images only and is usually used in combination with grayscale images. Depending on the type of texture in an image, we can observe characteristic patterns in the shape of $\mathbf{M}^p_{\mathbf{d}}$. A first extension of the classic co-occurrence matrix is, to capture the joint occurrence of intensity values between different image planes p and p' [10]. Formally, this can be written as

$$\mathbf{M}^{p,p'}_{\mathbf{d}}(i,j) = \mathbb{P}\left(c^{0,p}_{\mathbf{x}} = i \wedge c^{0,p'}_{\mathbf{y}} = j | \mathbf{x} - \mathbf{y} = \mathbf{d}\right) \tag{2}$$

According to the terminology in [10], co-occurrence matrices as defined by (1) will be denoted as *within* co-occurrence matrices and co-occurrence matrices as defined by (2) will be denoted as *cross* co-occurrence matrices. However, the latter concept can be generally applied to all kinds of vector images, including multiscale images. In [8] for example, images are analyzed on different scales, whereby the scale space is generated by repeatedly applying Gaussian filters with varying variance σ^2.

In our work, we leave the spatial domain and extend the concept of cross co-occurrence matrices to the wavelet-domain. We choose the maximally decimated, discrete wavelet transform (DWT) [11] in 2-D as our basic transformation (implemented by the Mallat algorithm). For convenience, we briefly summarize the computational steps: let $l^2(\mathbb{Z})$ be the space of square-summable sequences, then we first define the decomposition operators H and G on the sequence $\mathbf{c} = \{c_l | l \in \mathbb{Z}\}$ as

$$\begin{aligned} H: \quad l^2(\mathbb{Z}) &\to l^2(\mathbb{Z}) \\ \mathbf{c} &\mapsto H\mathbf{c} = \left\{ (Hc)_k = \sum_{l \in \mathbb{Z}} h_{l-2k} c_l \right\} \end{aligned} \tag{3}$$

$$\begin{aligned} G: \quad l^2(\mathbb{Z}) &\to l^2(\mathbb{Z}) \\ \mathbf{c} &\mapsto G\mathbf{c} = \left\{ (Gc)_k = \sum_{l \in \mathbb{Z}} g_{l-2k} c_l \right\} \end{aligned} \tag{4}$$

In the terminology of signal processing, $\{h_k | k \in \mathbb{Z}\}$ and $\{g_k | k \in \mathbb{Z}\}$ denote the filter coefficients of a low- and high-pass filter in a two-channel perfect-reconstruction filterbank. Since we work on images, we additionally define $\mathcal{H} := H_c H_r$, $\mathcal{H}_v := H_c G_r$, $\mathcal{H}_h := G_c H_r$ and $\mathcal{H}_d := G_c G_r$, where H_c, G_c, H_r, G_r operate on the column (c) or row (r) indices. In a J-scale 2-D DWT decomposition, our two-dimensional signal $\mathbf{C}^0$ is now be decomposed into a series of lower-dimensional signals $\mathbf{C}^J, \mathbf{D}^J_h, \mathbf{D}^J_v, \mathbf{D}^J_d, \ldots, \mathbf{D}^1_h, \mathbf{D}^1_v, \mathbf{D}^1_d$. The so called *approximation* subband is obtained by $\mathbf{C}^{j+1} = \mathcal{H}\mathbf{C}^j$ and the *detail subbands* are computed by

$$\mathbf{D}^{j+1}_k = \mathcal{H}_k \mathbf{C}^j, \quad k \in \{h, v, d\} \tag{5}$$

for $j = 0, \ldots, J-1$. The 2-D DWT is thus obtained by recursively applying convolution-decimation operations to the approximation subband. Now, that we can decompose an image into a set of detail subbands at each scale, we can go on to define our proposed color wavelet cross co-occurrence (CWCC) matrices, which aim at capturing joint occurrences of wavelet coefficients at equal subbands but different color channels (or image planes in general). The color wavelet cross co-occurrence matrix $\mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}(i,j)$ at position (i,j) between two arbitrary subbands $\mathbf{D}^{s,p}_k, \mathbf{D}^{s,p'}_{k'}$ at scale s is now defined as

$$\begin{aligned} &\mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}(i,j) = \\ &\mathbb{P}\left(c^{s,k,p}_{\mathbf{x}} = i \wedge c^{s,k',p'}_{\mathbf{y}} = j | \mathbf{x} - \mathbf{y} = \mathbf{d}\right) \end{aligned} \tag{6}$$

The additional superscripts for the transform coefficients are necessary to completely specify their position in the decomposition structure. For our experiments, we will make the restriction $k = k'$, which means that only pairs of subbands at equal positions in the decomposition are considered. To make (6) computationally feasible, it is further necessary that the transform coefficients are quantized. Therefore we use three common quantization factors, set to $64, 128$ and 256. We point out, that when using (2) and (6) it is now possible to have a zero-displacement vector $\mathbf{d} = \mathbf{0} = (0,0)^T$, which bears a relation to two-dimensional histograms [9].

B. Image Features

Until now, we have only concentrated on how to compute co-occurrence matrices. Unfortunately, the elements of these matrices cannot be directly used as input features to a

classifier. Even with a quantization factor of 64, we would end up with 64^2-dimensional feature vectors. According to [12], the number of samples needed to train a classifier grows exponentially with the number of input dimensions (known as the *curse of dimensionality*). To remedy this problem, we compute a subset of the popular Haralick [13] second-order statistics from our co-occurrence matrices, which are then assembled into feature vectors. Given that Q denotes the quantization factor used to compute $\mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}$, the features *Contrast*, *Correlation*, *Homogeneity* and *Energy* are defined by

Contrast

$$F_1 = \sum_{i=0}^{Q-1}\sum_{j=0}^{Q-1} |i-j|^2 \mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}(i,j) \tag{7}$$

Correlation

$$F_2 = \frac{\sum_{i=0}^{Q-1}\sum_{j=0}^{Q-1}(i-\mu_i)(j-\mu_j)\mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}(i,j)}{\sigma_i\sigma_j} \tag{8}$$

Homogeneity

$$F_3 = \sum_{i=0}^{Q-1}\sum_{j=0}^{Q-1} \frac{\mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}(i,j)}{1+|i-j|} \tag{9}$$

Energy

$$F_4 = \sum_{i=0}^{Q-1}\sum_{j=0}^{Q-1} \left(\mathbf{M}^{p,p'}_{\mathbf{d},s,k,k'}(i,j)\right)^2 \tag{10}$$

where μ_i, σ_i denote the horizontal mean and variance and μ_j, σ_j denote the vertical mean and variance. We will use the notation $F_i(\cdot)$ (e.g. $F_1(\mathbf{M}^p_{\mathbf{d}})$) in order to signify that a feature is computed from a specific kind of co-occurrence matrix. Since we evaluate the discriminative power of all features separately, a J-scale 2-D DWT decomposition produces a $J\times 3$-dimensional feature vector $\mathbf{v}^{p,p'}$ for a given combination p, p'. The final feature vector $\mathbf{v}$ for an image is a concatenation of all possible combinations. For a concrete example, consider the case of RGB images. We have $\mathbf{v}^{R,G}, \mathbf{v}^{R,B}$ and $\mathbf{v}^{B,G}$, which leads to the final $J \times 9$-dimensional feature vector

$$\mathbf{v} = [\mathbf{v}^{R,G}, \mathbf{v}^{R,B}, \mathbf{v}^{B,G}] \tag{11}$$

C. Classification & Feature Selection

To compare the discriminative power of our features, we employ a simple k-Nearest Neighbor (k-NN) classifier [12] using the euclidian formula as a distance metric. Each element of a feature vector is first normalized by subtracting the mean and dividing by the standard deviation. This ensures equal weights when calculating the euclidian distance. We further use sequential forward feature selection (SFFS) [12] to select a feature subset using a 1-NN classifier as a criterion function. The method of leave-one-out cross-validation (LOOCV) [12] is used to estimate the classification error.

IV. Experimental Results

In this section, we present the experimental results of our work. Our image database consists of 484 images, acquired in 2005/2006 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-endoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150. To enhance visual appearance, dye-spraying with indigo-carmine was applied and biopsies or mucosal resections were taken to obtain a histopathological diagnosis (our ground truth). For pit-pattern types I,II and V, biopsies were taken, since these types need not be removed. Lesions of pit-pattern types III-S/III-L and IV have been removed endoscopically. Table I lists the number of image samples per class. For all experiments, we use the RGB color

TABLE I
NUMBER OF IMAGES PER PIT-PATTERN CLASS (GROUND TRUTH)

I	II	III-L	III-S	IV	V
126	72	62	18	146	60

model and a 4-scale 2-D DWT with symlets of 4-th order. The extend, to which the decomposition depth and the choice of filter affect the classification results, will have to be studied in the future. Two preprocessing steps are implemented before the feature extraction. First, we employ histogram equalization on each color channel using the CLAHE [14] (contrast limited adaptive histogram equalization) algorithm with 8×8 tiles and an uniform distribution for constructing the contrast transfer function. Second, we blur the images using a Gaussian 3×3 mask with standard deviation $\sigma = 0.5$. Regarding our proposed approach, the displacement vector setup is as follows: $\mathbf{d}_0 = (0,0)$ (zero-displacement), $\mathbf{d}_{1,i} = (0,i), \mathbf{d}_{2,i} = (-i,i), \mathbf{d}_{3,i} = (-i,0)$ and $\mathbf{d}_{4,i} = (-i,-i)$ for $i = 1,\ldots,10$ (denoted as the standard displacement vectors). Table II lists the best LOOCV results obtained over all displacement vectors and quantization levels. We note, that throughout our experiments

TABLE II
COLOR WAVELET CROSS CO-OCCURRENCE LOOCV RESULTS

Problem	F_1	F_2	F_3	F_4
$\mathbf{d}_{1,1},\ldots,\mathbf{d}_{4,10}$				
2-cls.	81.61	86.16	79.34	81.61
6-cls.	50.00	66.12	49.79	51.45
$\mathbf{d}_0$				
2-cls.	75.00	**90.91**	73.55	77.27
6-cls.	49.17	**77.69**	45.25	50.21

it turned out, that the quantization factors only slightly affect the classification results, which is why we omit quantization level specific results here. From Table II we observe, that the the setup $(F_2, \mathbf{d}_0)$ outperforms all other combinations in terms of LOOCV accuracy. Another interesting fact is, that for almost all 40 variations of the standard displacement

vectors, $\mathbf{d}_{1,1}$ always performed best. Furthermore, increasing the distance i covered by $\mathbf{d}_{1,i}$ lead to a decrease in LOOCV accuracy for both classification problems, which is consistent with the observations in [9].

For our comparative study, we choose three other feature extraction approaches to which we compare our results. First, the so called *wavelet energy correlation signatures* [6] (WECS) between subbands of different color channels are computed. Given that $\mathrm{en}(\cdot)$ computes the energy of an arbitrary subband with $N_s \times N_s$ coefficients, the WECS $\gamma_k^s(p,p')$ between $\mathbf{D}_k^{s,p}$ and $\mathbf{D}_k^{s,p'}$ is defined by

$$\gamma_k^s(p,p') = \sum_{i=0}^{N_s-1} \sum_{j=0}^{N_s-1} \frac{c_{ij}^{s,k,p} \cdot c_{ij}^{s,k,p'}}{\mathrm{en}(\mathbf{D}_k^{s,p'}) \cdot \mathrm{en}(\mathbf{D}_k^{s,p'})} \tag{12}$$

for $p \neq p'$. Concatenation of the features from all possible color channel combinations leads to a 36-dimensional feature vector for our 4-scale 2-D DWT decomposition.

Second, we calculate the *color wavelet covariance* (CWC) features, introduced in [5]. There, the classic co-occurrence matrix defined in (1) is calculated over the second-scale detail subbands of each color channel. Using four displacement vectors covering directions $0°, 45°, 90°, 135°$ leads to 36 features per image for each Haralick measure. To incorporate information between color channels into the final feature vector, covariances between equal features are calculated according to [5]. By extending (1) by another subscript q to denote an arbitrary detail subband, the CWC features can be defined as

$$\begin{aligned} &\mathrm{CWC}_i\left(\mathbf{M}_{\mathbf{d},q}^{p}, \mathbf{M}_{\mathbf{d},q}^{p'}\right) = \\ &\mathrm{Cov}\left(F_i\left(\mathbf{M}_{\mathbf{d},q}^{p}\right), F_i\left(\mathbf{M}_{\mathbf{d},q}^{p'}\right)\right) \quad \forall i = 1,\ldots,4 \end{aligned} \tag{13}$$

The scale index s is omitted here, since $s = 2$ like in the original setup [5]. Hence, we obtain a final 18-dimensional CWC feature vector per Haralick measure F_i.

Third, all images are converted to the LUV color model and the luminance (L) channel is retained to obtain a grayscale image. Then, the 2-D DWT is computed and the popular statistics mean and standard deviation (see [15]) of the wavelet coefficients of each detail subband are used to compose a 24-dimensional feature vector (abbreviated by *Classic*).

To compare the discussed approaches, we select the best results of Table II and compare them to the best results obtained from the three aforementioned approaches. Table III presents the corresponding LOOCV rates.

TABLE III
LOOCV-RATE COMPARISON TO OTHER APPROACHES

Problem	CWCC	WECS	CWC	Classic
2-cls.	**90.91**	90.50	72.73	86.78
6-cls.	**77.69**	75.83	44.12	70.66

As we can see, the CWCC features lead to superior results in terms of LOOCV accuracy. Only the WECS features perform at a comparable level for both problems.

V. Conclusion & Further Work

In this paper we proposed a new approach to capture color channel information for texture discrimination in the field of medical imaging. Our experimental results show, that the presented features perform quite well with our dataset. We conclude, that the results strongly confirm the presumption (see [5], [6], [9], [10], etc.) that color channels are an important source of discriminative information in the field of texture-image classification. Future research will include an evaluation on how a combination of the presented Haralick measures can improve the LOOCV rates. In addition to that, the impact of different color models has to be studied and further tests on other texture databases have to be conducted.

Acknowledgment

This work is funded by the Austrian Science Fund (FWF) under Project No. L366-N15.

References

[1] S. Kudo, S. Hirota, T. Nakajima, S. Hosobe, H. Kusaka, T. Kobayashi, M. Himori, and A. Yagyuu, "Colorectal tumorous and pit pattern," *Journal of Clinical Pathology*, vol. 47, pp. 880–885, 1994.

[2] S. Kudo, S. Tamura, T. Nakajima, H. Yamano, H. Kusaka, and H. Watanabe, "Diagnosis of colorectal tumorous lesions by magnifying endoscopy," *Gastrointestinal Endoscopy*, vol. 44, no. 1, pp. 8–14, July 1996.

[3] M. Häfner, C. Kendlbacher, W. Mann, W. Taferl, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl, "Pit pattern classification of zoom-endoscopic colon images using histogram techniques," in *Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG 2006)*, J. R. Sveinsson, Ed. Reykavik, Iceland: IEEE, June 2006, pp. 58–61.

[4] M. Häfner, M. Liedlgruber, F. Wrba, A. Gangl, A. Vécsei, and A. Uhl, "Pit pattern classification of zoom-endoscopic colon images using wavelet texture features," in *Proceedings of the International Conference on Advances in Medical Signal and Image Processing (MEDSIP'06)*, W. Sandham, D. Hamilton, and C. James, Eds., Glasgow, Scotland, UK, July 2006, paper no. 0038.

[5] S. A. Karkanis, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 3, pp. 141–152, Sept. 2003.

[6] G. V. de Wouwer, S. Livens, P. Scheunders, and D. V. Dyck, "Color Texture Classification by Wavelet Energy Correlation Signatures," in *Proceedings of the 9th International Conference on Image Analysis and Processing (ICIAP'97)*. Florence, Italy: Springer, 1997, pp. 327–334.

[7] A. Jain and G. G. Healey, "A Multiscale Representation including Opponent Color Features for Texture Recognition," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 124–128, Jan. 1998.

[8] C. Palm, V. Metzler, B. M. O. Dieker, T. Lehmann, and K. Spitzer, *Bildverarbeitung für die Medizin*, 1999, ch. Co–Occurrence Matrizen zur Texturklassifikation in Vektorbildern, pp. 367–371.

[9] C. Palm, "Color texture classification by integrative co–occurrence matrices," *Pattern Recognition*, vol. 37, no. 5, pp. 965–976, May 2004.

[10] K. S. C. Palm T.M. Lehmann, "Color texture analysis of moving vocal cords using approaches from statistics and signal theory," in *Proceedings of the 4th International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research*, 2000, pp. 49–56.

[11] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1997.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley & Sons, Nov. 2000.

[13] R. M. Haralick, Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, pp. 610–621, November 1973.

[14] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. Morgan Kaufmann, 1994, pp. 474–485.

[15] W. M. B.S. Manjunath, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, Aug. 1996.