COLOR EIGEN-SUBBAND FEATURES FOR ENDOSCOPY-IMAGE CLASSIFICATION

Roland Kwitt and Andreas Uhl

Department of Computer Sciences
University of Salzburg, Austria
Email: {rkwitt,uhl}@cosy.sbg.ac.at

ABSTRACT

This paper presents a new image feature extraction approach in the wavelet domain. We incorporate color-channel information of the LAB color space into the feature extraction process by computing variances from decorrelated detail subbands of the stationary wavelet transform. We evaluate our approach on a medical image classification problem using a k-Nearest Neighbor classifier and sequential forward feature selection. Our experimental results, which include a comparative study to the popular color wavelet energy correlation signatures show that we can produce highly discriminative feature sets in terms of leave-one-out classification accuracy.

***Index Terms*—** Cancer, Wavelet Transforms, Biomedical Image Processing, Image classification, Pattern classification

1. INTRODUCTION

Recent statistics of the American Cancer Society reveal that colorectal cancer is the third most common cancer in men and women and the second most common cause of US cancer deaths. Since most colorectal cancers develop from polyps (usually divided into metaplastic, adenomatous and malignant), 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. Since the resection of all polyps is rather time-consuming and can be hazardous, it is imperative that those polyps which warrant resection can be distinguished. In the research work of Kudo et al. [1, 2], the macroscopic appearance of colorectal polyps is described systematically and results in the so called *pit-pattern* classification scheme, which differentiates the mucosal crypt patterns into five groups (pit-patterns I-V, see Figure 1).

In this work, we try to capture the characteristics of the various pit-pattern types by incorporating color channel information into a texture description framework. The primary aim is to allow computer-assisted pit-pattern classification and thus improve the quality of medical diagnosis. The remainder of the paper is structured as follows: In Section 2 we introduce our proposed feature extraction approach. Section 3 briefly introduces the classification step, followed by Section 4, where we present the experimental results of our work. Finally, Section 5 concludes the paper with a short summary and an outlook on further research.

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

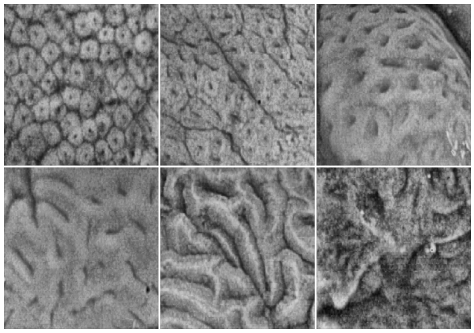


Fig. 1. Pit-Patterns I-V (top-left to botton-right)

2. COLOR EIGEN-SUBBAND FEATURES

In the research field of texture classification/discrimination, recent work has shown that the incorporation of color information into the feature extraction step can lead to an improvement in discrimination performance. Van de Wouwer et al. [3] for example, introduce so called color wavelet energy correlation signatures (CWECS) to capture both texture and color information in the wavelet domain. An almost equivalent approach was introduced by Jain et al. [4], combining the output of a set of Gabor filters applied on different color channels. In the spatial domain, Palm et al. [5, 6] propose to use second-order statistics computed from cross-cooccurrence matrices between color channels. In the research area of medical image processing, Karkanis et al. [7] exploit the dependency of second-order statistics computed from classic co-occurrence matrices in the wavelet domain to classify colonoscopy images. Another approach, based on the same image database we use here, is followed by Häfner et al. [8] using color histograms in the spatial domain.

In this paper we propose a new feature extraction approach, which is placed in the wavelet domain. The basic building elements are a variant of the discrete wavelet transform and principal component analysis (PCA). We do not use the classic, maximally decimated discrete wavelet transform (DWT), since it is not translation-invariant and leads to aliasing effects caused by downsampling the filter outputs. These particular drawbacks are especially important for image analysis problems. Since the process of colonoscopy is subjected to varying physical conditions, we will at least have to deal with shifted images showing the same mucosal surface patterns. Working with the 2-D DWT is thus suboptimal in our case. One way to overcome that lack of translation-invariance is to use a redundant representation that approximates the continuous wavelet transform. This is accomplished by a transform commonly known as the sta-

tionary wavelet transform (SWT) [9]. Since the SWT is less popular and widespread than the DWT, we will briefly review the basic concept (commonly known as the undecimated á trous algorithm [10]), following a similar approach to [11]. Recalling a picture of the the classic two-channel filterbank implementation of the DWT, the main idea of the SWT can be summarized in just one sentence: instead of decimating the filter outputs at each decomposition stage, the corresponding low- and bandpass filters are upsampled by two. To formalize this, let $\mathbf{c} = \{c_l | l \in \mathbb{Z}\}$ be an arbitrary sequence in the space of square summable sequences $l^2(\mathbb{Z})$ and let $\mathbf{h}^0$ and $\mathbf{g}^0$ be two filter coefficient sequences (e.g.: Haar filters). The superscript 0 denotes that we are working with the original, unmodified filters. Further, let $\mathcal{Z}^{[2^r]} : l^2(\mathbb{Z}) \rightarrow l^2(\mathbb{Z})$ denote a dyadic-upsampling operator, defined by

$$\mathbf{c} \mapsto \mathcal{Z}^{[2^r]}\mathbf{c} = \left\{ \left(\mathcal{Z}^{[2^r]}c\right)_k = \begin{cases} c_{k/2^r}, & \text{if } k/2^r \in \mathbb{Z} \\ 0, & \text{else} \end{cases} \right\} \quad (1)$$

with $r \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ representing the upsampling factor. The new filter coefficients for the low- and bandpass filter at an arbitrary decomposition level r of the SWT are then given by $\mathbf{h}^{[r]} = \mathcal{Z}^{[2^r]}\mathbf{h}^0$ and $\mathbf{g}^{[r]} = \mathcal{Z}^{[2^r]}\mathbf{g}^0$. Depending on r, the upsampling operation thus inserts 2^r zeros between the filter taps. By adhering to this notation, we can define two decomposition operators $H : l^2(\mathbb{Z}) \rightarrow l^2(\mathbb{Z})$

$$\mathbf{c} \mapsto H^{[r]}\mathbf{c} = \left\{ \left(H^{[r]}c\right)_k = \sum_{l \in \mathbb{Z}} c_l h^{[r]}_{k-l} \right\} \quad (2)$$

and G, where G is defined equivalent to Eq. (2), but using $\mathbf{g}^{[r]}$ instead. Both operators simply implement a convolution operation, where a given sequence is convolved with the filter coefficients at level r. The extension to 2-D is straightforward by separate row and column filtering. We define four operators $\mathcal{H}^{[r]} := H^{[r]}_c H^{[r]}_r$, $\mathcal{H}^{[r]}_v := H^{[r]}_c G^{[r]}_r$, $\mathcal{H}^{[r]}_h := G^{[r]}_c H^{[r]}_r$ and $\mathcal{H}^{[r]}_d := G^{[r]}_c G^{[r]}_r$, where $H^{[r]}_r$ and $H^{[r]}_c$ operate on the (r)ow and (c)olumn indices only. The same holds for $G^{[r]}_r$ and $G^{[r]}_c$. When we write our original image in matrix notation $\mathbf{C}^0 = \{c^0_{ij}\}_{1 \leq i,j \leq n}$, a J-scale 2-D SWT now decomposes $\mathbf{C}^0$ into one approximation image and a series of detail images (detail subbands). The approximation image at level $j+1$ is given by

$$\mathbf{C}^{j+1} = \mathcal{H}^{[j]}\mathbf{C}^j, \quad j = 0, \ldots, J \quad (3)$$

The wavelet detail subbands at level $j+1$, which mainly capture (h)orizontal, (v)ertical and (d)iagonal details, are computed from the approximation image $\mathbf{C}^j$ by

$$\mathbf{D}^{j+1}_k = \mathcal{H}^{[j]}_k \mathbf{C}^j, \quad k \in \{h, v, d\}, j = 1, \ldots, J \quad (4)$$

In matrix notation, an arbitray detail subband in the decomposition structure is given by $\mathbf{D}^j_k = \{w^{j,k}_{uv}\}_{1 \leq u,v \leq n}$. All resulting subbands are of the same size as the original image, which leads to a total redundancy of $J \times 3$. However, this is acceptable for image analysis purposes. In order to signify that a subband belongs to a specific color channel, we add an additional subscript p to identify a subband $\mathbf{D}^j_{k,p}$. In case of LAB images for example, we have $p \in \{L, A, B\}$. A subband is then unambiguously identifiable by the triple (j, k, p).

A straightforward approach to feature extraction would now be to decompose each color channel separately by the 2-D DWT/SWT and compute some statistics over the detail subbands. Commonly used statistics in this context are the mean, variance or subband energy for example. However, this approach would simply neglect possible correlations between the color channel subbands, which might lead to a loss of discriminative information.

To remedy this problem, we propose to compute the variance of decorrelated detail subbands as image features. Decorrelation is accomplished by PCA. We explain this procedure on a concrete example using the LAB color model. Certainly, almost any other color model is suitable as well (e.g.: RGB, HSV, etc.). First, we need to illustrate how to construct the input data matrix. Each color channel is subjected to a J-scale 2-D SWT/DWT, resulting in a total of $J \times 9$ detail subbands. We then construct a $n^2 \times 3$ wavelet coefficient matrix $\mathbf{X}^j_k$ according to the illustration presented in Figure 2. The detail subbands at position $(j, k, \cdot)$ in the decomposition structure of each color channel $p \in \{L, A, B\}$ are stacked so that we obtain three wavelet coefficients for every given pair of coordinates (i, j). In the following, these coefficients are used to form a row vector of the data matrix.

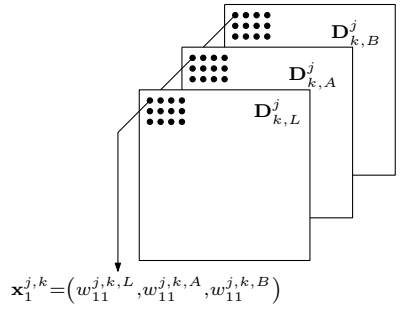


Fig. 2. Construction of the row vectors of the data matrix $\mathbf{X}^j_k$

Given that $M := \{(i, j) | 1 \leq i, j \leq n\}$ denotes the set of matrix indices for the detail subbands specifying the coefficient position, we thus obtain $|M|$ three-dimensional row-vectors $\mathbf{x}_i$, which constitute the data matrix $\mathbf{X}^j_k$. Principal component analysis is now used to compute a new coordinate system for our data. This coordinate system is designed so that most of the variation in the original data lies along the first coordinate axis (principal component axis). Variation then decreases with increasing coordinate axis index. Given an arbitrary $n \times 3$ data matrix $\mathbf{X}$, PCA works by diagonalizing the covariance matrix $\mathbf{S}(\mathbf{X})$ with help of its transposed eigenvector matrix $\mathbf{\Phi}^T$. Since PCA is often used for dimensionality reduction as well, it is important that $\mathbf{\Phi} = [\phi_1, \phi_2, \phi_3]$, with $\mathbf{S}\phi_i = \phi_i \lambda_i$ is arranged so that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The covariance matrix $\mathbf{S}$ can then be written as $\mathbf{S} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T$, with $\mathbf{\Lambda}$ denoting the diagonal matrix of eigenvalues. The transformation of the original data vectors then follows from $\mathbf{y}_i = (\mathbf{x}_i - \overline{\mathbf{x}})\mathbf{\Phi}^T$, which now constitute the new data matrix $\mathbf{Y}$. Consequently, the transformed covariance matrix $\tilde{\mathbf{S}} = \mathbf{\Phi}^T\mathbf{S}\mathbf{\Phi} = \mathbf{\Phi}^T\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{\Lambda}$ has diagonal form, with diagonal elements $\lambda_1 \geq \lambda_2 \geq \lambda_3$. This means that the variances along the new coordinate axis are given by the eigenvalues. In case the column vectors of $\mathbf{Y}$ are reshaped back to $n \times n$ matrices, we obtain our *color eigen-subbands (CES)*. Since the λ_i are the variances of the color eigen-subbands, we can directly use them as image features. Again, the indices j and k are added to $\lambda^{j,k}_i$ for completeness. In our experiments, we will compare the discriminative power of the CES features to the classic variance features computed from the original subbands. The feature vector construction is straightforward by concatenating the CES features. Given that N denotes the number of images in our database, a J-scale 2-D DWT/SWT leads

to $(J \times 9)$-dimensional feature vectors $\mathbf{f}_i, i = 1, \ldots, N$ which are given by

$$\mathbf{f}_i = [\lambda_1^{1,h}, \lambda_1^{1,v}, \lambda_1^{1,d}, \ldots, \lambda_3^{J,h}, \lambda_3^{J,v}, \lambda_3^{J,d}] \quad (5)$$

In case of the detail-subband variance features, we replace the eigenvalues by the corresponding subband variances and obtain $(J \times 9)$-dimensional feature vectors as well.

3. CLASSIFICATION

To compare the discriminative power of the proposed CES features, we employ a simple k-Nearest Neighbor (k-NN) classifier [12] using the euclidean formula as a distance metric. Each element of the feature vectors is first normalized by subtracting the feature mean and dividing by the feature standard deviation to ensure equal weights when calculating the euclidean distance. Further, we use sequential forward feature selection (SFFS) [13] to select a subset of features using the leave-one-out cross-validation (LOOCV) [12] accuracy of a 1-NN classifier as a criterion function. We note that we do not limit the number of features selected by SFFS. The algorithm simply returns the very subset, which achieved the lowest error. The final classification error is then estimated LOOCV as well. The comparative measure we choose is accuracy, which is defined as the number of correctly classified samples divided by the total number of samples. In a medical context, accuracy would represent the probability of a correct diagnosis.

4. EXPERIMENTAL RESULTS

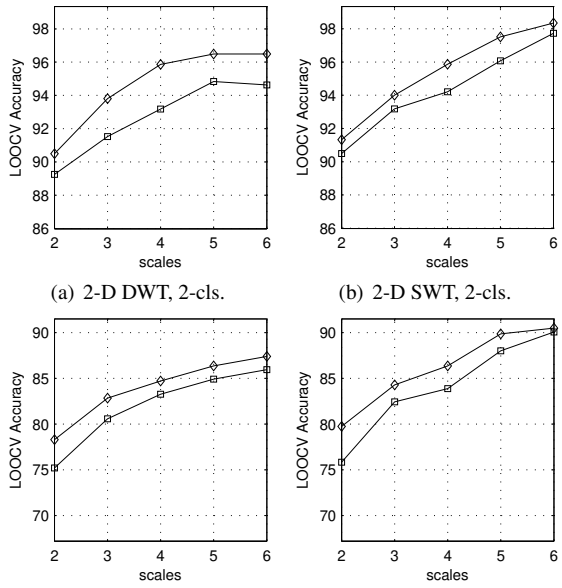
Our image database consists of 484 images, acquired in 2005/2006 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-endoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150. To enhance visual appearance, dye-spraying with indigo-carmine was applied and biopsies or mucosal resections were taken to obtain a histopathological diagnosis. 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 1 lists the number of image samples per class. We emphasize the point that our ground truth is based on histophatological findings (which is the most reliable source) and not on visual inspection.

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

Table 1. Number of images per pit-pattern class (ground truth)

Our experimental setup is as follows: regarding the choice of color space, we use the perceptually uniform CIE LAB [14] color space, which is derived from the CIE XYZ color space. In LAB, color information is separated into lightness (L), color information on a red/green (A) and yellow/blue (B) axis. All color channels are subjected to two preprocessing steps before the wavelet decomposition. First, we employ histogram equalization using the CLAHE [15] (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 with a Gaussian 3×3 mask and $\sigma = 0.5$. We compare the classic, maximally decimated 2-D DWT to the 2-D SWT in order to justify our reasoning from Section 2. For both transforms, the maximum decomposition depth J is varied from two to six. Additionally, we use four different Daubechies filters with two to eight taps (denoted as 'db2' to 'db8') to check the impact of the wavelet filters. The parameter k of the k-NN classifier is fixed to one, since this setup produced the best results throughout almost all of our experiments.

First, we check the main argument of our paper that the CES features lead to higher classification accuracies than the simple subband variances. Figure 3 shows a comparison of both feature sets for both wavelet transforms. The maximum LOOCV accuracy obtained over all filters is plotted against the decomposition depth.

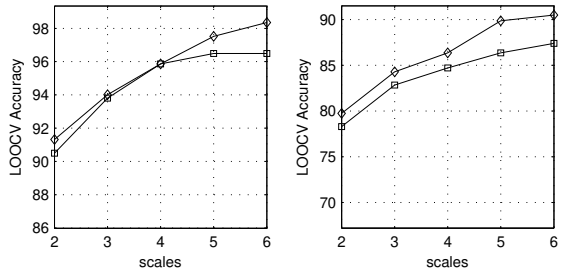


(a) 2-D DWT, 2-cls. (b) 2-D SWT, 2-cls.

(c) 2-D DWT, 6-cls. (d) 2-D SWT, 6-cls.

Fig. 3. LOOCV accuracy comparison between CES features ($\diamond$) and subband variance features ($\square$)

As we can see, the CES features perform consistently better in all cases. The impact of the filters is neglectable, although the highest rates were consistently achieved using the longer Daubechies filters (db6, db8). In Figure 4, both transforms are now directly compared using only the CES features. As before, the maximum LOOCV accuracies over all filters are plotted against the decomposition depth.



(a) 2-D DWT/SWT, 2-cls. (b) 2-D DWT/SWT, 6-cls.

Fig. 4. Direct LOOCV accuracy comparison between the 2-D DWT ($\square$) and 2-D SWT ($\diamond$) using the CES features

The results confirm the assumption of Section 2 that the 2-D SWT is more suitable for our purposes than the 2-D DWT. This is particularly evident for higher $(J \geq 5)$ decomposition scales. The

best numeric results from Figure 4 are listed in Table 2. The maximum classification rates are marked bold.

	2-D DWT	**2-D SWT**
2-class	96.49	**98.35**
6-class	87.40	**90.50**

Table 2. Best LOOCV accuracies for CES features

Looking at all the presented results so far, we conclude that although the choice of wavelet filter does not have a severe impact on the classification results, the decomposition depth surely does. Throughout all plots, we can see a clear upward trend in LOOCV accuracy with increasing decomposition depth. A careful inspection of the selected features, which lead to the maximum rates of Table 2 revealed that approximately 50% of the features originate from scales $j \geq 5$ for $J = 5, 6$. This is an evidence that finer frequency resolutions contain a lot of discriminative information. Of course, the trend towards higher rates for greater J is also related to the fact that increasing the decomposition depth equally increases the dimensonality of the feature vectors, which subsequently increases the search space of the SFFS algorithm.

Next, we compare our CES features to the popular color wavelet energy correlation signatures (CWECS), which were originally proposed by de Wouwer et al. [3] to capture characteristics of color textures. Given that $\mathrm{en}(\cdot)$ computes the energy of an arbitrary subband with $n \times n$ coefficients, the CWECS $\gamma_k^j(p, p')$ between $\mathbf{D}_{k,p}^j$ and $\mathbf{D}_{k,p'}^j$ is defined by

$$\gamma_k^j(p, p') = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{w_{ij}^{j,k,p} \cdot w_{ij}^{j,k,p'}}{\mathrm{en}(\mathbf{D}_{k,p'}^j) \cdot \mathrm{en}(\mathbf{D}_{k,p'}^j)} \tag{6}$$

for $p \neq p'$. In case of $p = p'$, the CWECS is equal to computing the subband energy [3]. However, we only consider the case $p \neq p'$ here, since this work is bout capturing coherencies between color channels. Concatenation of all CWECS features leads to $(J \times 9)$-dimensional feature vectors for a J-scale 2-D DWT/SWT decomposition. Table 3 list the LOOCV accuracies for the comparison between CES and CWECS features. Of course, the experimental setup remained the same for the CWECS features. We selected the best results over all filters, decomposition depth and wavelet transforms. It is noteworthy, that even in case of the CWECS features, the 2-D SWT with $J \geq 5$ leads to the best results, which is consistent with the CES findings.

	CWECS [3]	**CES**
2-class	97.50	**98.35**
6-class	89.60	**90.50**

Table 3. Comparison between CES and CWECS features

5. CONCLUSION

We have presented a set of new image features computed in the wavelet domain for a medical image classification problem. We have concentrated on the incorporation of color information and the decision for the most suitable wavelet transform variant. By using an orthogonal coordinate transformation to remove correlations between subbands of different color channels, we packed all discriminative information into the eigen-subbands. Our results show that this is a reasonable approach, since the classification rates are consistently higher than using subband variances alone. Another noteworthy result is, that the 2-D SWT is not only suitable for our CES features but even outperforms the classic 2-D DWT when relying on the CWECS features. In future work, we will examine the impact of other color spaces than LAB and test our approach on a number of well known texture databases to see if it is applicable in classic texture classification problems as well. From the pattern classification point of view, we will also evaluate the performance of other classifiers, which allow more complex decision boundaries.

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] G. Van de Wouwer, S. Livens, P. Scheunders, and D. Van 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, 1997, pp. 327–334, Springer.

[4] 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.

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

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

[7] Stavros 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.

[8] M. Häfner, Ch. 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)*, Johannes R. Sveinsson, Ed., Reykavik, Iceland, June 2006, pp. 58–61, IEEE.

[9] J.C. Pesquet, H. Krim, and H. Carfantan, "Time invariant orthonormal wavelet representations," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1964–1970, Aug. 1996.

[10] M.J. Shensa, "Wedding the à trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.

[11] G.P. Nason and B.W. Silverman, "The stationary wavelet transform and some statistical applications," *Lecture Notes in Statistics*, vol. 103, pp. 281–300, 1995.

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

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Morgan Kaufmann, 2nd edition, 1990.

[14] R. Lukac and K.N. Plataniotis, *Color Image Processing - Methods and Applications*, CRC Press, 2007.

[15] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics Gems IV*, Paul S. Heckbert, Ed., pp. 474–485. Morgan Kaufmann, 1994.