Color Texture Recognition in Video Sequences using Wavelet Covariance Features and Support Vector Machines

D. K. Iakovidis[1], D. E. Maroulis[1], S. A. Karkanis[2], I. N. Flaounas[1]
[1]RealTime Systems & Image Analysis Group, Department of Informatics and Telecommunications,
University of Athens, Panepistimiopolis, Illisia, 15784 Athens, Greece (rtsimage@di.uoa.gr)
[2]Technological Educational Institute of Lamia, Dept. of Informatics and Computer Technology,
3rd klm Old National Road, 35100 Lamia, Greece (sk@teilam.gr)

Abstract

This paper pertains to the recognition of textural regions for color video analysis. The proposed scheme uses the covariance of 2^{nd}-order statistics on the wavelet domain, between the different color channels of the video frames. These features, named as Color Wavelet Covariance (CWC), are used as color textural descriptors. A Support Vector Machine was chosen for the classification of the CWC feature vectors. Experiments were conducted using both animated Vistex texture mosaics and standard video clips. The estimated average accuracy ranged from 90% to 97%. The results show that the proposed methodology could efficiently be used in various multimedia applications as a complete supervised color texture recognition system.

1. Introduction

Video sequence analysis is an arising research area, which becomes essential as multimedia applications enter in our everyday life. The increase of the computational power of modern workstations has made feasible the application of complicated image analysis techniques on video frames. Such techniques usually exploit color and texture, both fundamental properties of the visible surfaces. Significant research effort has concentrated to the mathematical representation of color and texture for video sequence analysis. State of the art applications exploiting these properties include object tracking [1], face detection and recognition systems [2], tumor detection in endoscopic video [3] and content indexing [4].

Recent studies in color texture analysis have considered the use of perceptual approaches [5], the use of chromaticity moments [6], the derivation of textural information from luminance channel along with pure chrominance features as well as the processing of each color channel separately, by applying gray-level texture analysis techniques [7]. Other approaches exploit the interdependence of the existent textural information within the different channels of a color image, usually captured by means of correlation. On this direction Van de Wouwer et al [8] achieved high classification rates using correlation signatures estimated from the wavelet coefficients of color images. Paschos [9] proposed a set of discriminative and robust chromatic correlation features using directional histograms. Vandebroucke et al [10] exploited the correlation of 1st order statistical features between the different color channels for unsupervised soccer image segmentation and Al-Rawi et al [11] proposed Zernike moments of correlation and covariance functions for illumination invariant color texture recognition.

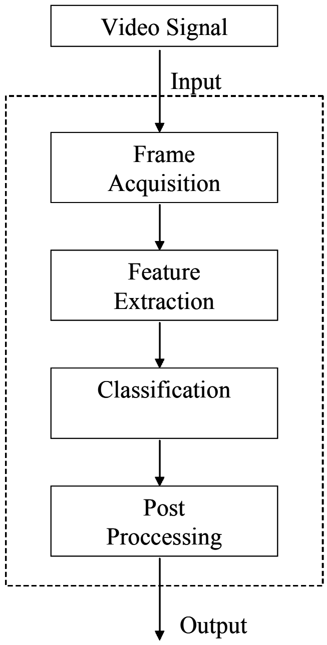


Figure 1. Color texture recognition scheme

In the same framework we have formulated a mathematical representation of color texture named as Color Wavelet Covariance (CWC) that exploit the covariance of 2nd-order textural measures in the wavelet domain of the color channels of the images. Considering that video sequences are series of sequentially ordered images in time, this methodology could also be applied for the representation of color texture in video frames. In this paper we propose the use of CWC features for the identification of regions characterized of different texture in video sequences. The scheme illustrated in Fig.1 outlines our approach.

For each acquired video frame a set of CWC feature vectors are calculated. These vectors constitute the input of a Support Vector Machine (SVM) classifier, which is responsible for the texture recognition task. SVMs are supervised machine learning algorithms that are based on statistical learning theory [12] and they have shown remarkably robust performance in several pattern recognition applications [13-16]. In the proposed framework the SVM is trained using features extracted from a single frame. After training it is capable of identifying different texture regions along the whole video sequence. The SVM's output are video frames depicting the identified texture regions, which are consecutively enhanced following a post-processing stage.

The rest of this paper is organized in 4 sections. In section 2, the feature extraction methodology used for the representation of texture in video frames is described. Section 3 provides a short description of basic principles of the SVM classification scheme. In section 4, we present the results of the experimentation aiming to the assessment of the recognition performance of the proposed approach. In the last section the conclusions of this study are summarized.

2. Color Texture Features

In the proposed feature extraction methodology we assume that each video-frame I, is decomposed into three color channels Ci, where i = 1, 2, 3. Each channel is raster scanned with a fixed size sliding square window. On each window a K-level 2D-Discrete Wavelet Transform (DWT) is applied. The Daubechies wavelet bases were used due to their orthonormal properties, which are important for the preservation of the textural structure along the different scales of the transform [17]. This transform results in a new representation of the original window, which consists of

$$B = 3K + 1 \tag{1}$$

sub-windows, corresponding to different wavelet bands.

Each band is denoted as *Bj(k)*, where *k* is the current level of the transform and *j* = 0, 1, 2, 3 for *k* = *K*, or *j* = 1, 2, 3 for *k* < *K*. *B*$_0$*(k)* corresponds to the low frequency band.

The textural information contained in each window is captured with the use of cooccurrence matrices. Cooccurrence matrices encode the gray level spatial dependence based on the estimation of the 2nd order joint conditional probability density function f(*i*, *j*, *d*, *a*), which is computed by counting all pairs of pixels at distance d having gray levels *i* and *j* at a given direction *a*. The angular displacement of *d* = 1 is included in the range of the *a*-values {0, π/4, π/2, 3π/4}.

The proposed approach for the estimation of color textural features takes advantage of the covariance between statistical measures of the cooccurrence matrix corresponding to each color channel of the video frame. To investigate the performance of this approach we have considered four Haralick's measures, namely the angular second moment (f_1), the correlation (f_2), the inverse difference moment (f_3) and the entropy (f_4). These four features provide high discrimination accuracy which can only be marginally increased by adding more features in the feature vector [18].

The features f_1- f_4 are estimated over each sub-window *Bj(k)*, *j* ≠ 0, *k* = 1, 2, … *K*, of the color channels C_i, i = 1, 2, 3 of the frame and they are noted as:

$$F_{C_i}^{B_j(k)}(a), \tag{2}$$

$$j \neq 0,\ k = 1, 2, \ldots K,$$

where $F \in \{ f_1, f_2, f_3, f_4 \}$ and *a* corresponds to the angle considered in the estimation of the cooccurrence matrices, $a \in \{0,\ \pi/4,\ \pi/2,\ 3\pi/4\}$. We define *Color Wavelet Covariance of a feature F* (*CWC or CWC$_F$*), $F \in \{f_1, f_2, f_3, f_4\}$ at wavelet band $B_j(k)$, $j \neq 0$, $k = 1, 2, \ldots K$, between two color channels C_l and C_m as:

$$CWC^{B_j(k)}(C_l, C_m) = Cov\left(F_{C_l}^{B_j(k)}, F_{C_m}^{B_j(k)}\right) \tag{3}$$

estimated over the different angles *a*. For *K*=1, the corresponding feature vectors consist of 72 CWC features ((3 variances + 3 covariances) x 4 cooccurrence matrices x 3 wavelet bands).

The use of these features can lead to a reduced feature space compared to the original feature space defined by Eq.(2).

3. Support Vector Machines

Let Φ be a non-linear mapping from the input space $I \subseteq \Re^n$ to the feature space $F \subseteq \Re^m$. The SVM algorithm is capable of finding a hyperplane defined by the equation

$$w\Phi(x) + b = 0 \tag{4}$$

so that the *margin of separation* is maximized. It is easy to prove [12][19] that for the *maximal margin* hyperplane,

$$w=\sum_{i=1}^{N}\lambda_i y_i \Phi^{\mathrm{T}}(x_i) \qquad (5)$$

where the variables λ_i are Lagrange multipliers that can be estimated by maximizing the quantity

$$L_D=\sum_{i=1}^{N}\lambda_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j K(x_i,x_j) \qquad (6)$$

with respect to λ_i, where the following constraints should be satisfied: $\sum_{i=1}^{N}\lambda_i y_i=0$ and $0\le\lambda_i\le c$, for i = 1, 2, …, N, and a given value c. $K(x_i,x_j)$ is called kernel function and it is defined as the inner product

$$K(x_i,x_j)=\Phi^{\mathrm{T}}(x_i)\Phi(x_j)\,. \qquad (7)$$

Linear, polynomial, Radial Basis (RBF) and sigmoid are the most common functions used as SVM kernels. The one-against-one strategy is used for the classification of multiple classes [19].

4. Results

The proposed color texture recognition approach was tested on different video sequences. The experiments presented in this paper are organized in two parts. In the first part we evaluate the recognition performance of the proposed approach using animated color texture mosaics constructed from Vistex texture images [20]. In the second part two standard video clips are used to demonstrate the recognition accuracy of the proposed approach in real-world scenes.

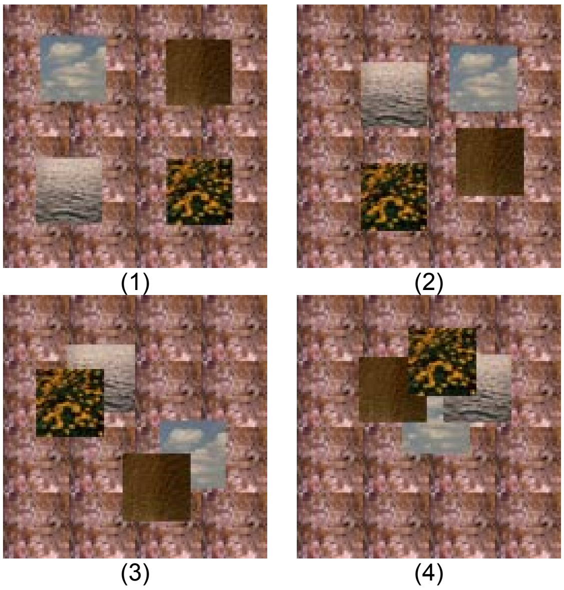
4.1. Texture recognition in animated mosaics

Five 10-frame animated mosaics of 5 textures were used to evaluate the recognition performance of the proposed approach. The video frames were 128x128 pixels in size and the color depth was 24 bit = 3×8 bit. Each video consists of the following Vistex textures:

Video 1: "bark", "clouds", "sand", "water", "flowers"
Video 2: "wood", "food", "fabric", "leaves", "grass"
Video 3: "sand", "flowers", "leaves", "grass", "bark"
Video 4: "clouds", "water", "metal", "stone", "brick"
Video 5: "metal", "fabric", "food", "stone", "wood"

Fig.2 illustrates four indicative frames of Video 1.

For each video, 1 frame was used for training and the rest 9 were used for testing. This procedure was repeated 10 times using a different frame for training each time in order to avoid bias. The window sizes tested were 8x8, 16x16 and 32x32 pixels. We have considered the use of three different color spaces namely RGB, L*a*b* and K-L (estimated as linear approximation of the Karhunen-Loeve transformation of the RGB image coordinates). These color spaces have been used in various texture recognition applications in the literature [8][21-23]. Among the four different SVM kernels mentioned in section 3, we have chosen the linear as the least computationally complex. The results, shown in Fig. 3, are estimated in terms of Mean Classification Error (MCE).



(1) (2) (3) (4)

Figure 2. Four frames of Video 1

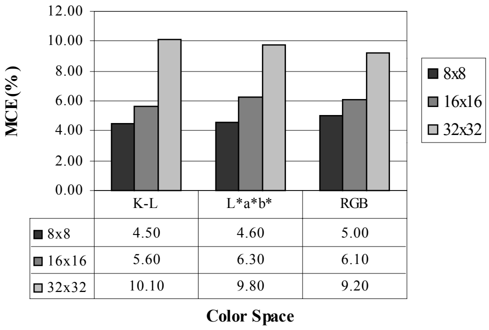


Figure 3. MCE in different color spaces and window sizes

Fig.3 shows that best results were obtained using the K-L color space and a window size of 8x8 pixels. The MCE obtained using L*a*b* color space is also low for the same window size but the RGB to L*a*b* transform involves non-linear computations which are more expensive than the linear computations involved in the RGB to K-L transform. The use of a small window size

8x8, which results to the lowest MCE, can be evaluated as an advantageous characteristic of the proposed methodology since it increases the output frames' detail.

To investigate the effect of the SVM kernel type in the texture recognition performance of the proposed approach, different kernel functions were tested including linear, 2nd-order polynomial, RBF and sigmoid. The results obtained using K-L color space and a window size of 8x8 pixels are illustrated in Fig.4.

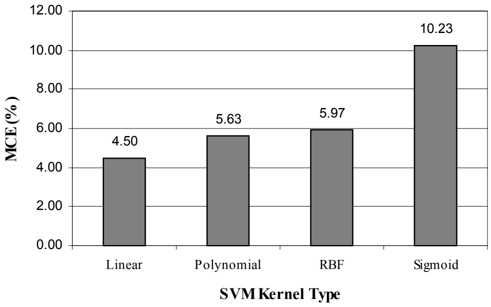


Figure 4. MCE using different SVM kernel types

From the above diagram (Fig. 4), it can be concluded that the linear kernel results in the lowest MCE. Fig.5 illustrates the output frames corresponding to Fig.2 using the linear kernel. The different shades of gray in the output images correspond to the different classes. These frames validate the high accuracy achieved using the proposed approach.

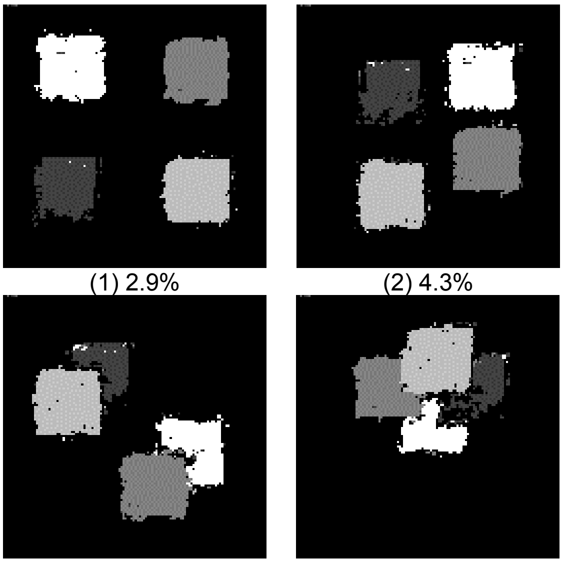

(1) 2.9%
(2) 4.3%
(3) 4.2%
(4) 5.0%

Figure 5. Output frames

Further improvement of the results can be achieved using a post-processing stage, which involves the application of a noise reduction scheme. We tried indicatively a median filter with a kernel size of 5x5 pixels on the output frames [24]. The application of this technique resulted to a decrease of the MCE, from 4.50% to 3.03%. Fig.6 illustrates the output frames of Fig.5 after post-processing stage.

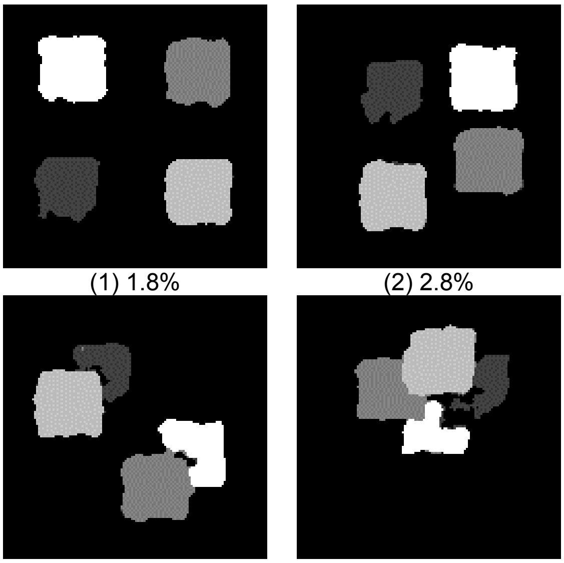

(1) 1.8%
(2) 2.8%
(3) 2.5%
(4) 3.8%

Figure 6. Output frames after post-processing

4.2. Texture recognition in standard video clips

In the second part of the experimentation we tested our methodology on two standard uncompressed color video sequences, named "silent" and "container". They both consist of 300 frames and 5-class scene segmentation was considered. For each sequence the first frame was used for training and the rest were used for testing. The estimated MCE over each sequence in K-L color space using windows of 8x8 pixels in size and linear SVM kernel was 10.5% for "silent" and 12.1% for "container". Fig.7 and Fig.8 illustrate indicative results for different frames of these sequences. The first, the second and the third column of these figures, correspond to the original, the reference and the output frames respectively. The different shades of gray in the reference and in the output images correspond to the different classes. It can be observed that the output frames are comparable to the reference frames. Some misclassified regions belong to the classes comprised of fewer samples and correspond to small texture areas. The dominant classes are well defined in both video clips (e.g. the background painting in "silent" and the sea in "container").

5. Conclusions

In this paper we presented a novel methodology for color texture recognition in video sequences. The feature extraction scheme was based on the CWC features which produce statistical color descriptors for texture on the wavelet domain of the video frame sequences. The recognition task was assigned to SVMs due to their reliable performance.

The proposed methodology was tested on both animated Vistex texture mosaics and standard video clips, reaching to an average recognition accuracy of 97% and 90%, respectively.

The different texture classes identified in a video clip could be utilized in several video-processing tasks including object detection and tracking. A general approach to refine these classes was to use median filtering, but depending on the application, different image processing algorithms could be applied. A future perspective for the extension of this work could be the integration of the heavy computational procedures involved, on dedicated hardware, in order to reduce the overall response time of the system for a potential real-time multimedia application.

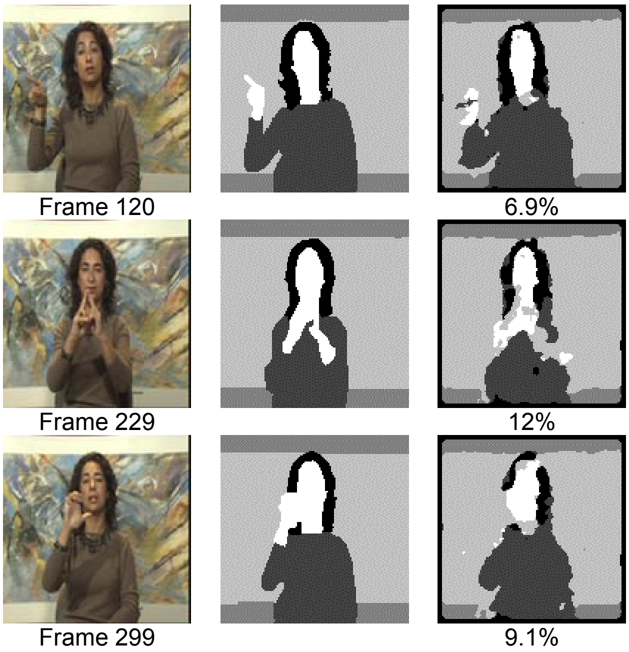


Figure 7. Classification results for the "silent" video clip

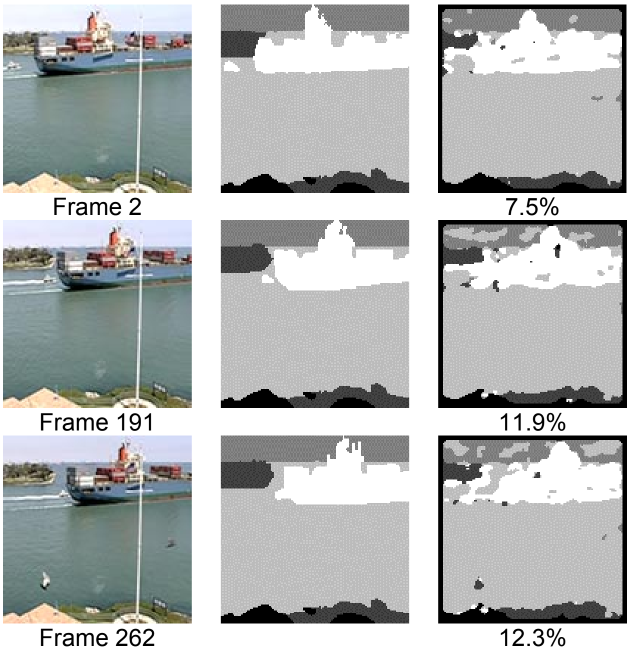


Figure 8. Classification results for the "container" video clip

6. References

[1] E. Ozyildiz, N. Krahnstöver, R. Sharma, "Adaptive Texture and Color Segmentation for Tracking Moving Objects", *Pattern Recognition*, vol. 35, no. 10, pp. 2013-2029. Oct. 2002.

[2] E. Acosta, L. Torres, A. Albiol, E. Delp, "An Automatic Face Detection and Recognition System for Video Indexing Applications", in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Orlando, USA, 2002.

[3] S.A. Karkanis, D.K. Iakovidis, D.E. Maroulis, D.A. Karras, M. Tzivras, "Computer Aided Tumor Detection in Endoscopic Video using Color Wavelet Features", Accepted for Publication, *IEEE Transactions on Information Technology in Biomedicine*, IEEE, 2003.

[4] S. -F. Chang, "Compressed domain techniques for image/video indexing and manipulation", in *Proc. IEEE International Conference on Image Processing,* Oct. 1995, pp. 314-317

[5] M. Mirmehdi and M. Petrou, "Segmentation of Color Textures", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 142-159, 2000.

[6] G. Paschos, "Fast Color Texture Recognition Using Chromaticity Moments", *Pattern Recognition Letters*, vol. 21, pp. 847-841, 2001.

[7] A. Drimbarean, and P.F. Whelan, "Experiments in Colour Texture Analysis", *Pattern Recognition Letters*, vol. 22, pp. 1161-1167, 2001.

[8] G. Van de Wouwer, P. Scheunders, S. Livens, and D. Van Dyck, "Wavelet Correlation Signatures for Color Texture Characterization", *Pattern Recognition*, vol. 32, pp. 443-451, 1999.

[9] G. Paschos, "Chromatic Correlation Features for Texture Recognition", *Pattern Recognition Letters*, vol. 19, pp. 643-650, 1998.

[10] N. Vandenbroucke, L. Macaire, and J.-G. Postaire, "Unsupervised Color Texture Feature Extraction and Selection for Soccer Image Segmentation", in *Proc. IEEE ICIP*, Vancouver, Canada, vol. 2, pp. 800-803, 2000.

[11] M.Al-Rawi, and Y. Jie, "Illumination Invariant Recognition of Color Texture Using Correlation and Covariance Functions", in *Proc. EMMCVPR 2001*, LNCS 2134, 2001, pp. 216-231.

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

[13] E. Osuna, R. Freund, F. Girosi, "Training support vector machines: an application to face detection", in: *Proc. Computer Vision and Pattern Recognition*, 1997, pp. 130-136.

[14] Y. LeCun, L.D. Jackel, L. Bottou, A. Brunot, C. Cortes, J.S.Denker, H. Drucker, I. Guinon, U.A. Muller, E. Sackinger, P. Simard, V. Vapnik, "Comparison of learning algorithms for handwritten digit recognition", in: *Int. Conference on Artificial Neural Networks*, F. Fogelman and P. Gallinari (Ed.), 1995, pp.53-60.

[15] D.A. Karras, S.A. Karkanis, D. Iakovidis, D.E. Maroulis, B.G. Mertzios, "Support vector machines for improved defect detection using novel multidimensional wavelet feature extraction involving vector quantization and PCA techniques", in: *NIMIA Adnvanced Study Institute on Neural Networks for Instrumentation, Measurement and Related Industrial Applications*, Crema, Italy, 2001, pp. 139-144.

[16] D.K. Iakovidis, D.E. Maroulis, S.A. Karkanis, P. Papageorgas, M. Tzivras, "Texture Multichannel Measurements for Cancer Precursors' Identification using Support Vector Machines", Accepted for Publication, *Measurement*, Elsevier Science, 2002.

[17] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.

[18] R.M. Haralick, "Texture Measures for Carpet Wear Assessment", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 1, pp. 92-104, 1988.

[19] C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Kluwer Academic Publishers, Boston, 1998.

[20] Vistex, Color Image Database. http://www-hite.media.mit.edu/vismod/imagery/VisionTexture. MIT Media Lab, 2000.

[21] S.A Karkanis, G.D Magoulas, D.K Iakovidis, D.A Karras, and D.E Maroulis, "Evaluation of Textural Feature Extraction Schemes for Neural Network-based Interpretation of Regions in Medical Images", in *Proc. ICIP 2001*, Thessaloniki, Greece, pp. 281-284, 2001.

[22] G. Wyszecki, and W.S. Styles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, John Wiley & Sons, New York, 1982.

[23] Y.I. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation", in *Proc. CGIP, 1980*, vol. 13, pp. 222-241.

[24] K. In Kim, K. Jung, S. Hyun Park, H. Joon Kim, "Support Vector Machines for Texture Classification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542-1550, 2002.