Performance Evaluation of Texture Segmentation Methods

Igor V. GRIBKOV, Petr P. KOLTSOV, Nikolay V. KOTOVICH, Alexander A. KRAVCHENKO, Alexander S. KUTSAEV, Andrey S. OSIPOV, Alexey V. ZAKHAROV

Scientific Research Institute for System Studies, Russian Academy of Sciences (NIISI RAN)

Nakhimovskii pr. 36-1, Moscow, 117218, Russia

ABSTRACT

Various methods have been developed for texture segmentation. Since all of them have their merits and drawbacks, the choice of the most suitable method becomes a nontrivial task. We present a comparative study of texture segmentation methods based on four frequently used texture feature extraction techniques: gray level co-occurrence matrices, Gaussian Markov random fields, Gabor filtering, and Laws' masks. The methods have been tested on the large set of images synthesized especially for this investigation. The performance of these methods has been evaluated against ground truth image by using well-known performance measures – modified Pratt's figure of merit and correct pixel classification rate.

Keywords: Performance Evaluation, Segmentation, Texture, Feature Extraction, Testing.

1. INTRODUCTION

Texture analysis methods are used in a variety of computer vision applications. Texture segmentation is an important subarea of texture analysis. Its goal is to find the texture boundaries in the image being processed. During last decades a number of different methods have been developed for texture segmentation. Since all of them have their merits and demerits, a choice of the most suitable method becomes a nontrivial task for the user.

The comparative study of texture analysis methods is important for development of computer vision applications. In recent years, a growth of interest to this topic can be observed. However, most of the published papers deal with texture classification problem only, without considering full image segmentation.

There are relatively few papers focused entirely on comparing texture segmentation methods. Du Buf *et al.* [1] were among the first to evaluate these methods in a quantitative framework. Seven texture features extraction techniques (gray level co-occurrence matrix (GLCM), fractal, Laws', Unser's, Knutsson's, Mitchell's, Barba-Ronsin's) were compared using a set of 20 synthetic bipartite images. The GLCM, Laws and Unser methods gave best overall results.

An extensive investigation of major filtering approaches to texture feature extraction was made by Randen and Husoy [2]. They included in this study Laws masks, ring/wedge filters, dyadic Gabor filter banks, wavelet transforms, wavelet packets and wavelet frames, quadrature mirror filters (QMF), discrete cosine transform, eigenfilters, optimized Gabor filters, linear predictors, and optimized finite impulse response filters. In addition two non-filtering approaches (GLCM and multiresolution autoregressive model) were tested for reference. In the process of testing the authors used 12 mosaic images, consisting of several texture samples (from 2 to 16) taken from the Brodatz album [3], the MIT Vision Texture database, and the MeasTex Image Texture Database. The best average performance was achieved with the QMF approach. However, none of these methods showed the best results for all the images.

Chang *et al.* [4] considered three texture feature extraction methods: GLCM, Law's texture energy and Gabor multi-channel filtering. The test set of 35 real scene images and 5 Brodatz mosaic images was used. Gabor approach showed the best performance in this study.

Clausi and Yue [5] compared the discrimination ability of two methods: GLCM and Markov random fields (MRF). The testing was performed on a set of three synthetic, five Brodatz, and six SAR sea ice images. The role of window size in a process of feature extraction was especially investigated. While the window size was small, the GLCM produced a better segmentation than the MRF. In the case of large window size the performance of both methods was approximately equal.

Different evaluation techniques, different test image sets, different method parameters, and different software implementations can sometimes lead to the opposite results of comparative studies. Some years ago we designed the PICASSO framework intended especially for helping practical users to test image processing algorithms themselves. This paper illustrates using the current version of PICASSO for performance evaluation of texture segmentation methods.

2. PICASSO – EVALUATION FRAMEWORK

The PICASSO is a software framework for quantitative performance evaluation of image processing algorithms. It exploits the so-called empirical discrepancy evaluation technique, which uses a ground truth data for assessing the results of testing the algorithms under investigation. The framework contains a test image database, a set of program implementations of image processing algorithms and performance measures, and a special image editor, which is meant for generation of ground truth and test images. The PICASSO has a modular design that allows a user to easily update the framework with his test suites.

Originally the PICASSO was designed for testing edge detectors [6]. The practical usage of this evaluation framework demonstrated the validity of the applied approach. In further versions of PICASSO, some other types of image processing algorithms, test images, and performance measures have been added. We used them in our studies of such problems as image restoration, boundary improvement and image segmentation [7-9].

The current version of PICASSO has been supplemented with texture test suite including more than 5000 test images, software realization of four feature extraction algorithms and two performance measures.

3. TEXTURE METHODS

The texture segmentation process involves two basic stages: texture features extraction and partition of obtained feature set into clusters in feature space. Generally texture features are divided into four major classes: statistical, model-based, signal processing, and geometrical (including structural) features [10] (the first three feature classes are most often used). In this paper we consider four feature extraction techniques: gray level co-occurrence matrices (statistical), gaussian Markov random fields (model-based), Gabor filtering, and Laws' masks (signal processing). On the second segmentation stage the standard *k*-means method [18] has been used.

Gray level co-occurrence matrices

The gray level cooccurence matrix (GLCM) is constructed by estimating the pairwise statistics of pixel gray levels. Each element g_{ij} of this matrix represents an estimate of the probability that two pixels at certain relative displacement have gray levels i and j. The displacement is determined by a distance d and an angle θ. Dimension of the matrix depends on image gray level number as a square of this number.

Many features derived from the GLCM have been proposed for using in texture analisys. We have computed six features in our experiments: four of Haralick's fourteen features [11] (energy E, entropy H, local homogeneity L, and inertia I) and two features suggested by Conners *et al.* [12] (cluster shade CS and cluster prominence CP):

$$E = \sum_{i,j} g_{ij}^2 , \quad H = -\sum_{i,j} g_{ij} \log g_{ij} , \tag{1}$$

$$L = \sum_{i,j} \frac{1}{1+(i-j)^2} g_{ij} , \quad I = \sum_{i,j} (i-j)^2 g_{ij} , \tag{2}$$

$$CS = \sum_{i,j} ((i-\mu_i)+(j-\mu_j))^3 g_{ij} , \tag{3}$$

$$CP = \sum_{i,j} ((i-\mu_i)+(j-\mu_j))^4 g_{ij} , \tag{4}$$

where $\mu_i = \sum_i i \sum_j g_{ij} , \quad \mu_j = \sum_j j \sum_i g_{ij}$ (5)

Gaussian Markov random fields

In the case of Gaussian Markov random fields (GMRF) approach some image model is assumed to exist [13]. The relation between gray level value $g(s)$ of pixel s and gray level values $g(s+r)$ of its neighbours can be expressed as

$$g(s) = \sum_{r \in D} \theta(r) g(s+r) + \varepsilon(s) \tag{6}$$

where D is the set of neighbours of pixel s, $\theta(r), r \in D$ are the model parameters, $\varepsilon(s)$ is independent Gaussian random variable with zero mean and variance σ^2.

The model parameters $\theta(r)$ are estimated in overlapping windows using least squares error method and are used as texture features. In order to overcome problems concerning too high feature variances, the features are smoothed by a Gaussian low-pass filter.

Gabor filtering

Gabor multi-channel filtering technique suggests the creation of a bank of Gabor filters (Gaussian shaped band-pass filters), with dyadic coverage of the radial spatial frequency range and multiple orientations. The filters can be generated for varying frequency ω, orientation θ, phase shift φ, and Gaussian width σ [14]:

$$G(x, y \mid \omega, \theta, \varphi, X, Y) = \exp\left\{ \frac{-((x-X)^2 + (y-Y)^2)}{2\sigma^2} \right\} \sin(2\pi\omega a + \varphi) , \tag{7}$$

where X, Y are the coordinates of the filter center, $a = x\cos\theta - y\sin\theta$.

We have computed Gabor energy S^2 at each pixel of the image $I(x,y)$ and have used it as a texture feature:

$$S^2(X,Y \mid \omega,\theta) = \left[\sum_{x,y} G(x,y \mid \omega,\theta,0,X,Y) I(x,y) \right]^2$$

$$+ \left[\sum_{x,y} G(x,y \mid \omega,\theta,\frac{\pi}{2},X,Y) I(x,y) \right]^2 \qquad (8)$$

Laws' masks

One of the pioneering techniques is the approach by Laws [15], where predetermined one-dimensional kernels are combined into various convolution masks. We have used the set of kernels of length five:

$$\begin{aligned} &L5 = [\ 1\ 4\ 6\ 4\ 1\ ], \quad E5 = [\ -1\ -2\ 0\ 2\ 1\ ] \\ &S5 = [\ -1\ 0\ 2\ 0\ -1\ ], \quad W5 = [\ -1\ 2\ 0\ -2\ 1\ ] \qquad (9) \\ &R5 = [\ 1\ -4\ 6\ -4\ 1\ ] \end{aligned}$$

From these five 1-D kernels one can generate 25 different 2-D masks by convolving a vertical 1-D kernel with a horizontal 1-D kernel. At first, each of these 2-D masks is applied to the input image. Then, the output images *I(x,y)* of the convolution process are taken for the computation of Laws' texture energy measure *E*:

$$E(x,y) = \frac{1}{(2W+1)^2} \sum_{i=-W}^{i=W} \sum_{j=-W}^{j=W} \left| I(x+i, y+j) \right|, \quad (10)$$

where W is half-size of local window used for computing $E(x,y)$.

Software implementation

We have used fragments of MeasTex's source codes [21] in our software implementation of feature extraction algorithms.

4. TESTING METHODOLOGY AND RESULTS

Test image set

The above mentioned feature extraction algorithms have been applied to the texture segmentation task by many researchers. Texture samples from Brodatz's album are traditionally used for test images creating. But all of these algorithms show good enough performance under testing on the set of "mosaic" images composed from Brodatz's album samples. This makes the choice of the best algorithm difficult.

In order to increase the discriminative power of our evaluation method, we have generated a special set of test images. We have extracted feature vectors for texture samples from Brodatz's album (about 700 features per texture). Then for each pair of samples we have synthesized a set of texture subimages, whose feature values change with small steps from one original texture to another (the software package *textureSynth* designed by J.Portilla and E.Simoncelli [16] have been used at this stage of our investigation). Finally we have generated test images. Each of them is composed of original texture pair and intermediate textures (see Fig. 1 as the example). These images contain not only sharp transitions between strongly different textures but also more smooth transitions. It allows one to reveal more profound distinctions in performance of texture segmentation methods under testing.

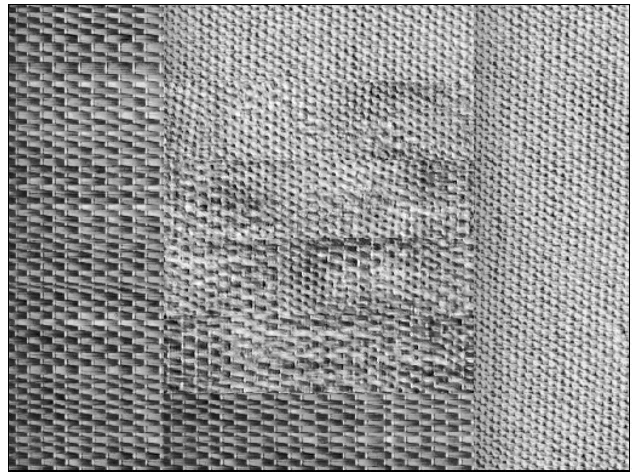


Fig. 1. The test image D55_D77.

We have used 108 of 112 Brodatz's samples for synthesizing test images. Thus our test set consists of 5778 grayscale images, each 512×384 pixels in size. We have excluded four Brodatz's samples (D43, D44, D45, D59) from our consideration. In texture analysis literature a texture image is usually defined to have variations of intensities which form certain repeated patterns [10]. Following this definition, we regard these samples as non-texture images. Consequently, they are not suitable for testing texture methods.

Performance measures

The performance of compared methods has been evaluated against ground truth using two well-known performance measures – correct pixel classification rate (*CR*) and modified Pratt's figure of merit (*FOM*) [17].

$$FOM = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{1+\gamma d_i^2}, \qquad (11)$$

where N – number of pixels in the image,
d_i – Euclidean distance between the ith pixel of segmented image and its correct class,
γ – scaling constant which can be used to change a contribution of errors to the *FOM*.

FOM and *CR* are equal to 1 if the segmentation is absolutely correct.

Parameters selection

In this paper, no optimization for parameters of the tested methods has been done. We have tried to follow standard recommendations found in image processing literature.

Four values of the angle θ (0°,45°,90°,135°) and one value of the distance $d=1$ have been used for calculating GLCM. The quantized number of gray levels Q is another important parameter. The requantization of the original image is usually implemented for reducing dimensions of co-occurance matrices and consequently for accelerating calculations. Clausi [19] confirms that large values of Q (Q>64) do not improve the classification accuracy. However, our experiments have shown that for our test images the finer quantization improves the performance of GLCM method (see Fig. 2.). Because of this, we have used $Q=256$.

The number of the GMRF parameters depends on selecting the model order (see [13]). We have used a fourth order symmetric model characterized by ten parameters.

Gabor filters have been implemented with the parameter set recommended in [20]. We have used six orientation θ (0°,30°,60°,90°,120°,150°) and three frequencies ω ($\sqrt{2}/W$, $2\sqrt{2}/W$, $4\sqrt{2}/W$ cycles per pixel, W=17 is a window size used for filtering).

The full set of 25 Laws' masks has been applied in our research. We have used 15×15 window for computing Laws' texture energy measure.

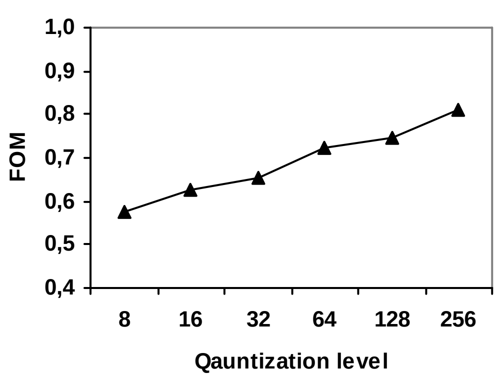


Fig. 2. The performance of GLCM method as a function of quantization level (average value for the set of 50 images).

Segmentation results

Each test image has been processed by four texture segmentation methods. A summary diagram is presented in Fig. 3. The results are shown for the FOM performance measure only because the ones for *CR* measure differ by no more than one percent.

The best average performance has been achieved with the GLCM method. Gabor filtering has demonstrated slightly worse performance. The results showed by GMRF and Laws' methods were significantly poorer than those of the GLCM.

Note that the segmentation of our test images has proved to be a challenge for methods being tested. Image regions containing original Brodatz's samples have been segmented quite well by all the methods. A middle part of test image contains smooth transitions between synthesized textures, and as one can see in Fig. 4 and Fig. 5 the segmentation of this image region is not an easy task for the studied texture methods.

The methods have shown substantially different performance for different images. No method is the best for all tests, but in the most cases the GLCM and Gabor filtering have yielded a satisfactory segmentation. The approach of Laws has shown the greatest variation in performance. For some images it has performed well, whereas for the others it has completely failed.

Further improvements in texture segmentation quality may be achieved in two ways. The first is a fusion of features extracted by different texture techniques followed by optimizing the obtained feature set. The second is a selection of the most suitable algorithm for clustering data in the feature space.

5. CONCLUSIONS

From the four texture segmentation methods tested, the GLCM and Gabor methods have shown the best overall results. The GLCM method has achieved slightly better performance, but it is more computationally costly than the Gabor filtering. If the computer time consumption is critically important for the image processing application, then the Gabor method may be a better choice.

No method is the best for all test images. This fact leads to the idea that a fused feature set containing features extracted by different texture techniques has to provide improved segmentation results. Our further studies will be aimed at optimizing such fusion texture feature sets and at selecting the most suitable clustering algorithm.

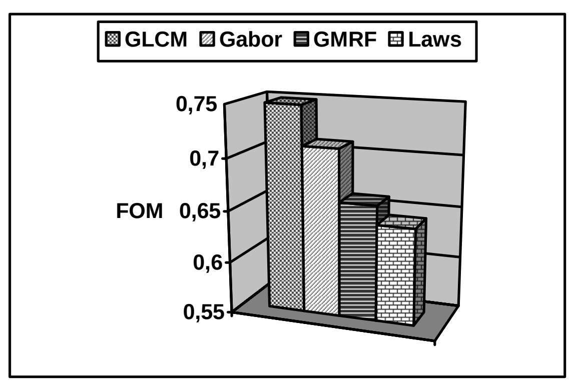
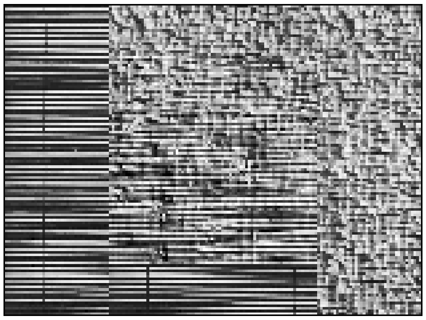
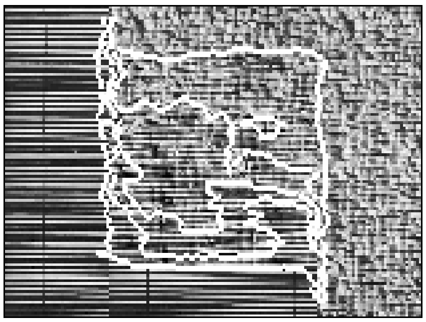


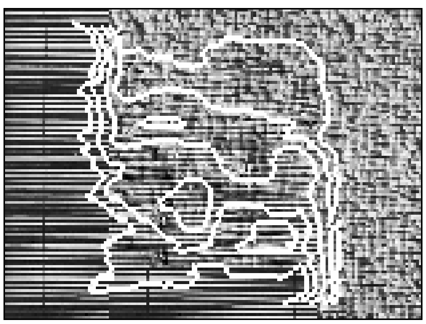
Fig. 3. Summary of testing results.



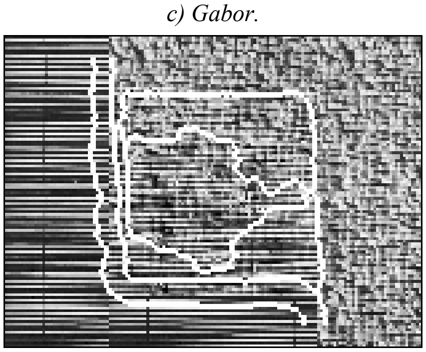
a) Original image.



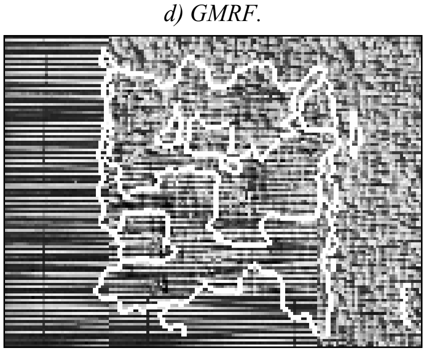
b) GLCM.



c) Gabor.

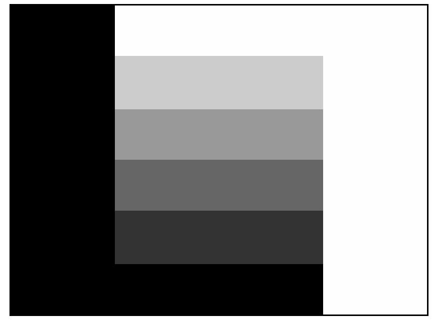


d) GMRF.

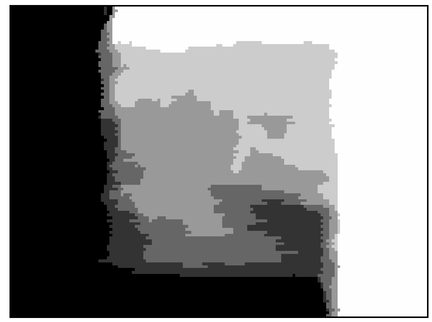


e) Laws.

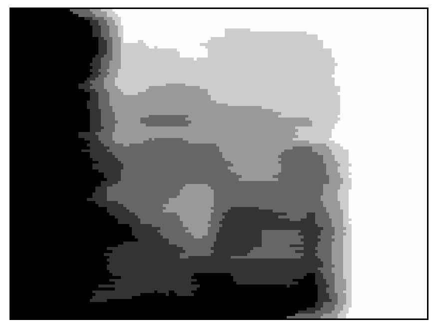
Fig. 4. *Test image D49_D84 and its segmentations yielded by four methods under testing.*



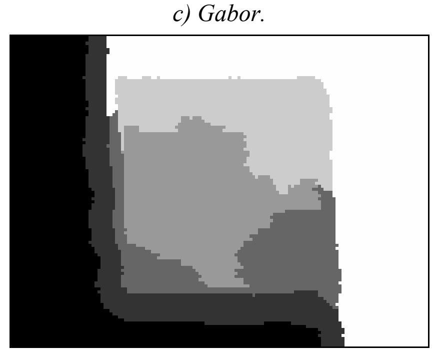
a) Ground truth.



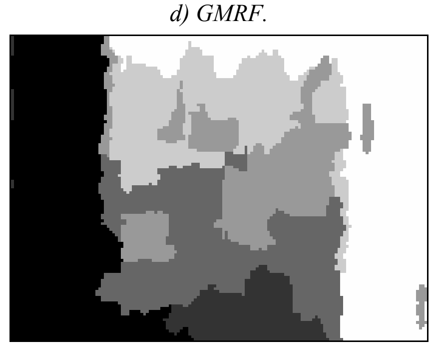
b) GLCM.



c) Gabor.



d) GMRF.



e) Laws.

Fig. 5. *Ground truth and segmentation maps for test image D49_D84.*

6. REFERENCES

[1] J.M.H. du Buf, M. Kardan, and M. Spann, “Texture feature performance for image segmentation”, **Pattern Recognition,** Vol. 23, No. 3/4, 1990, pp. 291-309.

[2] T. Randen and J.H. Husoy, “Filtering for Texture Classification: A Comparative Study”, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 21, No. 4, 1999, pp. 291-310.

[3] P. Brodatz, **Textures: A Photographic Album for Artists and Designers,** New York, 1966.

[4] K.I. Chang, K.W. Bowyer, and M. Sivagurunath, “Evaluation of Texture Segmentation Algorithms”, **Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition (CVPR'99)**, 1999, pp. 294-299.

[5] D.A. Clausi and B. Yue, “Comparing co-occurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery”, **IEEE Transactions on Geoscience and Remote Sensing**, Vol. 42, No. 1, 2004, pp. 215-228.

[6] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Kutsaev, V.K. Nikolaev, A.V. Zakharov, “PICASSO – Edge Detectors Evaluating System Based on the Comprehensive Set of Artificial Images”, **Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics**, Vol. 9, 2002, pp. 88-93.

[7] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Kutsaev, V.K. Nikolaev, A.V. Zakharov, “Testing of Energy Minimizing Methods in Image Preprocessing Using the PICASSO System”, **Proceedings of the 8th World Multiconference on Systemics, Cybernetics and Informatics**, Vol. 6, 2004, pp. 233-238.

[8] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Kutsaev, A.S. Osipov, A.V. Zakharov, “Empirical evaluation of Image Processing Methods Using PICASSO 2 System”, **WSEAS Transactions on Systems**, Vol. 4, Issue 11, 2005, pp. 1923-1930.

[9] I.V. Gribkov, P.P. Koltsov, N.V. Kotovich, A.A. Kravchenko, A.S. Kutsaev, A.S. Osipov, A.V. Zakharov, “Edge Detection under Affine Transformations: Comparative Study by PICASSO 2 System”, **WSEAS Transactions on Signal Processing**, Vol. 2, Issue 9, 2006, pp. 1215-1221.

[10] M. Tuceryan and A.K. Jain, “Texture Analysis”, in **The Handbook of Pattern Recognition and Computer Vision (2nd Edition)**, C.H. Chen, L.F. Pau, P.S.P. Wang (eds.), World Scientific Publishing Co., 1998, pp. 207-248.

[11] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification”, **IEEE Transactions on Systems, Man and Cybernetics**, Vol. 3, Nov.1973, pp. 610-621.

[12] R.W. Conners, M.M. Trivedi, and C.A. Harlow, “Segmentation of a High-Resolution Urban Scene using Texture Operators”, **Computer Vision, Graphics and Image Processing**, Vol. 25, 1984, pp. 273-310.

[13] R. Chellappa and S. Chatterjee, “Classification of Textures using Gaussian Markov Random Fields”, **IEEE Transactions on Acoustics, Speech, and Signal Processing**, Vol. 33, Aug. 1985, pp. 959-963.

[14] I. Fogel and D. Sagi, “Gabor Filters as Texture Discriminator”, **Journal of Biological Cybernetics**, Vol. 61, 1989, pp. 103-113.

[15] K.I. Laws, “Rapid Texture Identification”, **Proceedings of SPIE Conference on Image Processing for Missile Guidance**, 1980, pp. 376-380.

[16] J. Portilla and E.P. Simoncelli, “A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients”, **International Journal of Computer Vision**, Vol. 40, No. 1, 2000, pp. 49-71.

[17] K.C. Strasters and J.J. Gerbrands, “Three-Dimensional Image Segmentation Using a Split, Merge and Group Approach”, **Pattern Recognition Letters**, Vol. 12, No. 5, 1991, pp. 307-325.

[18] R.O. Duda, P.E. Hart, and D.G. Stork, **Pattern Classification**, New York: John Wiley & Sons, 2nd edition, 2001.

[19] D.A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization”, **Canadian Journal of Remote Sensing,** Vol. 28, No. 1, 2002, pp. 45-62.

[20] D.A. Clausi. and M.E. Jernigan, “Designing Gabor filters for optimal texture separability”, **Pattern Recognition**, Vol. 33, No. 11, 2000, pp. 1835-1849.

[21] G. Smith and I. Burns, “Measuring texture classification algorithms”, **Pattern Recognition Letters**, Vol. 18, No. 14, 1997, pp. 1495-1501.