Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 1318–1325
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

A color image segmentation approach for content-based image retrieval

Mustafa Ozden, Ediz Polat*

Electrical and Electronics Engineering Department, Kirikkale University, 71450 Kirikkale, Turkey

Received 28 October 2005; received in revised form 11 July 2006; accepted 15 August 2006

Abstract

This paper describes a new color image segmentation method based on low-level features including color, texture and spatial information. The mean-shift algorithm with color and spatial information in color image segmentation is in general successful, however, in some cases, the color and spatial information are not sufficient for superior segmentation. The proposed method addresses this problem and employs texture descriptors as an additional feature. The method uses wavelet frames that provide translation invariant texture analysis. The method integrates additional texture feature to the color and spatial space of standard mean-shift segmentation algorithm. The new algorithm with high dimensional extended feature space provides better results than standard mean-shift segmentation algorithm as shown in experimental results.

Keywords: Image segmentation; Mean-shift algorithm; Discrete wavelet frames; Texture segmentation

1. Introduction

Image segmentation is an important step for many image processing and computer vision algorithms. The interest is motivated by applications over a wide spectrum of topics. For example, analyzing different regions of an aerial photo is useful for understanding plant/land distribution. Extracting an object of interest from background of an image is important for building intelligent machines for factory automation systems. Segmenting and counting blood cells from cell images can help hematologists to improve diagnosis of diseases. Scene segmentation is also helpful to retrieve images from large image databases for content-based image retrieval systems [1,2]. Most of the methods require using image features that characterize the regions to be segmented. Particularly, texture and color have been independently and extensively used in the area. Similar image segmentation algorithms have been developed in Refs. [3–5], where texture is used as a main descriptor and wavelet frames are employed for feature extraction. However, it is also mentioned that using only texture descriptor cannot be sufficient for successful texture segmentation. In Ref. [6], color and spatial information are used in a nonparametric probabilistic framework where the mean-shift algorithm is employed for feature space analysis. The methods combining multiple image features in a probabilistic framework remain limited and active. Some of the work that employs different image features can be found in Refs. [7–10] where color and texture features are used together for segmentation. For texture segmentation, Bayes network-based algorithms provide better results comparing to other algorithms especially in case of having unknown texture types [11]. Other similar texture segmentation algorithms that employ multiple features can be found in Refs. [12–15]. All of the algorithms developed in these works consider relatively low-dimensional feature spaces which are computationally inexpensive, however, relatively less robust.

This paper considers the segmentation problem of image regions based on color, texture and spatial information in a nonparametric framework. The proposed method uses discrete wavelet frames (DWF) [3] to characterize textured

* Corresponding author. Tel.: +90 318 357 3571x1228; fax: +90 318 357 2459.

E-mail addresses: mozden@kku.edu.tr (M. Ozden), polat@kku.edu.tr (E. Polat).

0031-3203/$30.00 © 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2006.08.013

regions in images. DWF decomposition of a textured region provides a translation invariant texture description which results in better estimation and more detailed texture characterization at region boundaries. The color and spatial feature space of the mean-shift algorithm [6] is then extended using these texture characteristics to create higher dimensional feature space which results in improved segmentation.

The rest of the paper is organized as follows. In Section 2, the brief description of texture characterization using wavelet frames are presented. Subsequently in Section 3, building the higher dimensional feature space using spatial, color and texture information and the mean-shift filtering based on this feature space is described. The experimental results are given in Section 4. Finally Section 5 concludes the paper.

2. Wavelet frames for texture characterization

In this paper, the DWF decomposition, a variation of the discrete wavelet transform, is used for texture characterization. A filter bank is employed for decomposing the image into orthogonal components which simplify the classification problem. Unlike other decompositions, DWF is computationally inexpensive for the evaluation of low-frequency components. Dissimilar to other wavelet-based approaches, the output of the filter banks is not sub-sampled in DWF decomposition between levels. This provides translation invariant texture description of input signal. This property yields a better estimation of texture statistics and more detailed characterization at region boundaries. DWF decomposition can be calculated by successive 1-D processing along the rows and columns of the image using initial low-pass (h_i) and high-pass (g_i) filters which are expanded in every iteration i. The expansion can be achieved iteratively in the z and signal domains as follows [3]:

$$
\begin{aligned}
H_{i+1}(z) &= H(z^{2^i})H_i(z), \quad h_{i+1}(k) = [h]_{\uparrow 2^i} * h_i(k), \\
G_{i+1}(z) &= G(z^{2^i})H_i(z), \quad g_{i+1}(k) = [g]_{\uparrow 2^i} * h_i(k), \qquad (1)
\end{aligned}
$$

where the notation $[.]_{\uparrow m}$ indicates the up-sampling by a factor of m. These expanded filters can be used to decompose a signal in sub-bands approximately one octave each. In order to construct orthogonal wavelet decomposition, the following discrete normalized wavelet basis functions can be used:

$$
\begin{aligned}
\phi_{i,t}(k) &= 2^{i/2}h_i(k - 2^i t), \\
\varphi_{i,t}(k) &= 2^{i/2}g_i(k - 2^i t), \qquad (2)
\end{aligned}
$$

where i and t are the scale and translation indices, respectively. A fast and iterative implementation of the decomposition can be achieved as follows:

$$
\begin{aligned}
s_{i+1}(k) &= [h]_{\uparrow 2^i} * s_i(k), \\
d_{i+1}(k) &= [g]_{\uparrow 2^i} * s_i(k), \qquad (3)
\end{aligned}
$$

where s_i and d_i are the signal expansion coefficients and wavelet coefficients, respectively. A block diagram of iterative DWF decomposition of a 1-D signal is presented in Fig. 1(a). A block diagram of one-level DWF decomposition of 2-D image is also shown in Fig. 1(b) where L and H correspond to low-pass and high-pass filters, respectively. Rows and columns of the image are processed separately using filters. The output of the filter bank is organized into the N-component vector where N is the number of sub-bands ($N = 4$ in Fig. 1(b) for one-level decomposition). Each vector contains coefficients which represent approximate, horizontal, vertical and diagonal characteristics of input image.

A texture is characterized by a set of median values of energy estimated in a local window at the output of the corresponding filter bank. The energy in a local window can be calculated using coefficients of DWF decompositions (LL, LH, HL, and HH) where the energy is defined as the square of the coefficients. The advantage of using median filter is that it preserves the energy associated with texture between regions. The sub-bands at the output of filter bank in Fig. 1(b) correspond to approximate, horizontal, vertical and diagonal components of the input image signal [7]. Due to the fact that most of the texture information are contained in LH and HL sub-bands, we used only these decomposition coefficients to obtain texture features. A pixel in textured region can be classified into one of four texture categories based on texture orientation. These are *smooth* (not enough energy in any orientation), *vertical* (dominant energy in vertical direction), *horizontal* (dominant energy in horizontal direction), and *complex* (no dominant orientation). Texture feature extraction consists of two steps. First, the energy of LH and HL sub-bands are classified into two categories (0 and 1) using K-means clustering algorithm. Second, a further classification is made using combination of two categories in each sub-band LH and HL. A pixel is classified as *smooth* if its category is 0 in both LH and HL sub-bands. A pixel is classified *vertical* if its category is 0 in LH, and 1 in HL sub-bands. Similarly, a pixel is classified *horizontal* if its category is 1 in LH, and 0 in HL sub-bands. Finally, a pixel is classified as *complex* if its category is 1 in both LH and HL sub-bands. An input image with different texture regions as well as the classification results is illustrated in Fig. 2. In Fig. 2(a), four regions with different Brodatz textures are shown. In Fig. 2(b), the regions are classified based on their energy and orientation. The leftmost region in the original image is classified as *horizontal*, the rightmost region is classified as *vertical*, and the middle top region is classified as *smooth*, finally the middle bottom region is classified as *complex* based on their texture characteristics. The spatial accuracy in region boundaries depends on the texture window size. The window size should be large enough to capture texture characteristics. On the other hand, excessively large window size can cause border effects. The goal is to characterize image pixels using these texture features and use them to extent the mean-shift feature space to obtain better segmentation. The details of

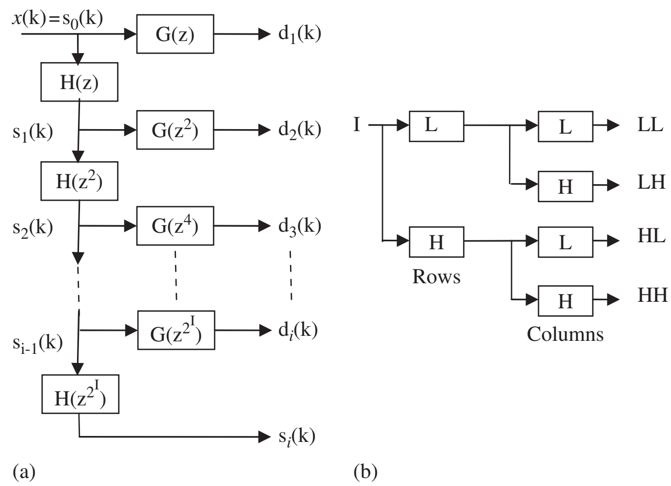


Fig. 1. Illustration of discrete wavelet frame decomposition: (a) iterative implementation for 1-D signal; (b) one-level decomposition of 2-D image using low-pass (L) and high-pass (H) filters. The input image is decomposed into four sub-bands.

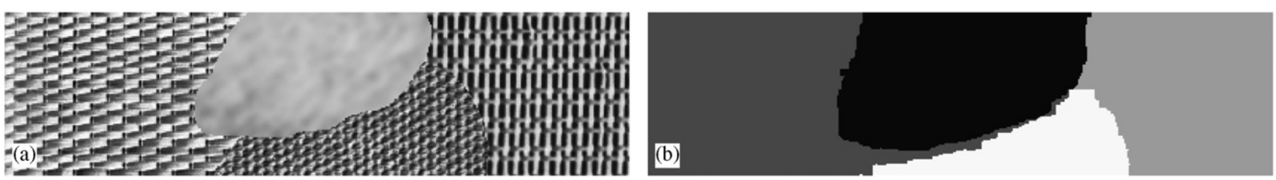


Fig. 2. Illustration of classifying different textured regions: (a) a textured image containing vertical, horizontal, smooth and complex textures created using Brodatz textures; (b) classification result using wavelet frame decomposition with median energy in a local window.

extending mean shift feature space is given in the following section.

3. Mean-shift filtering in higher dimensional space and segmentation

Given n data points x_i, $i=1,\ldots,n$ in the d-dimensional space R^d, the kernel density estimation at the location x can be calculated by

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^d}\,k\left(\left\|\frac{x-x_i}{h_i}\right\|^2\right), \tag{4}$$

with the bandwidth parameter $h_i>0$. The kernel K is a spherically symmetric kernel with bounded support satisfying [16]

$$K(x) = c_{k,d}\,k(\|x\|^2) > 0, \quad \|x\| \leqslant 1, \tag{5}$$

where the normalization constant $c_{k,d}$ assures that $K(x)$ integrates to one. The function $k(x)$ is called the profile of the kernel. Assuming derivative of the kernel profile $k(x)$ exists, using $g(x)=-k'(x)$ as the profile, the kernel $G(x)$ is defined as $G(x)=c_{g,d}\,g(\|x\|^2)$. The following property can be proven by taking the gradient of Eq. (4) as follows:

$$m_G(x) = C\,\frac{\hat{\nabla} f_K(x)}{\hat{f}_G(x)}, \tag{6}$$

where C is a positive constant and, it shows that, at location x, the mean-shift vector computed with kernel G is proportional to the normalized density gradient estimate obtained with kernel K. The mean-shift vector is defined as follows:

$$m_G(x) = \frac{\sum_{i=1}^{n}\frac{1}{h_i^{d+2}}\,x_i\,g\left(\left\|\frac{x-x_i}{h_i}\right\|^2\right)}{\sum_{i=1}^{n}\frac{1}{h_i^{d+2}}\,g\left(\left\|\frac{x-x_i}{h_i}\right\|^2\right)} - x. \tag{7}$$

The mean-shift vector thus points toward the direction of maximum increase in the density. The mean-shift procedure is obtained by successive computation of the mean-shift vector and translation of the kernel $G(x)$ by the mean-shift vector. At the end of the procedure, it is guaranteed to converge at a nearby point where the estimate has zero gradient [17]. In other words, it is a hill climbing technique to the nearest stationary point of the density. The iterative equation is

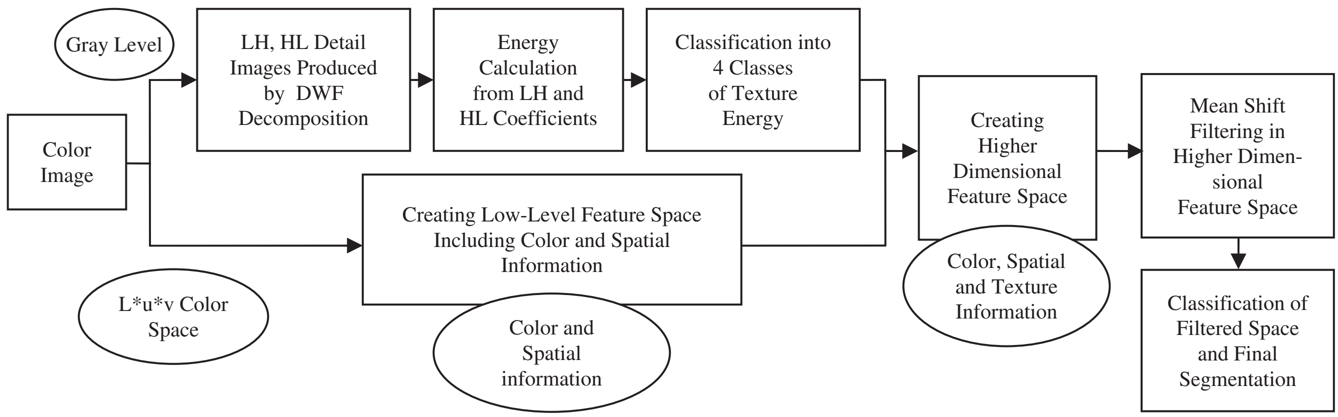


Fig. 3. Overview of the proposed approach.

given by

$$y_{j+1} = \frac{\sum_{i=1}^{n} \frac{x_i}{h_i^{d+2}} g\left(\left\| \frac{y_j - x_i}{h_i} \right\|^2\right)}{\sum_{i=1}^{n} \frac{1}{h_i^{d+2}} g\left(\left\| \frac{y_j - x_i}{h_i} \right\|^2\right)}, \quad j = 1, 2, \ldots . \tag{8}$$

The initial position of the kernel (starting point to calculate y_1) can be chosen as one of the data points x_i. Usually, the modes (local maxima) of the density are the convergence points of the iterative procedure.

The mean-shift image segmentation algorithm [6] considers a joint domain representation that includes spatial and range (color) domains. An image is represented as a two-dimensional lattice where the space of the lattice is known as spatial domain, and the gray-level or color information is represented in the range domain. Every pixel in the image can be considered as a p-dimensional vectors where $p = 1$ in gray-level case and $p = 3$ for color images. The dimension of joint domain representation becomes $d = p + 2$. Using this representation, the mean-shift filtering is performed to smooth the image and to preserve the region boundaries based on color and spatial information. However, in cases where colors in region boundaries are similar, this representation will not be sufficient and additional features are needed for more robust segmentation. In this paper, we addressed this problem and thus we extended the mean-shift feature space by integrating texture descriptors to improve segmentation. The details of obtaining texture descriptors are explained in the previous section. The block diagram of proposed method is shown in Fig. 3. The steps can be explained as follows:

1. Use wavelet transformation to decompose the image into sub-bands (LL, LH, HL, and HH). The DWF decompositions are the same as these sub-bands except that there is no sub-sampling. Most of the texture information are in the LH and HL sub-bands.
2. Calculate the median energy using coefficients of LH and HL sub-bands in a local window. The size of the window should be large enough to capture the local texture characteristics. The energy is defined as the square of the coefficients, and used for texture characterization.
3. Use K-means clustering algorithm to classify the energy values in two classes for each sub-band. There will be four texture classes based on energy: smooth (not enough energy in any orientation), vertical (dominant energy in vertical orientation), horizontal (dominant energy in horizontal orientation), and complex (no dominant orientation).
4. Generate the feature vector such that every pixel in the image has p-dimensional feature vector which includes spatial (x, y), color (gray-level or $L * u * v$ values) and texture (smooth, vertical, horizontal or complex) information.
5. Filter the image using mean-shift algorithm in higher dimensional feature space which includes spatial, color and texture information. The filtering operation can be controlled by setting the spatial window radius (h_s) and color range (h_r). The filter output (convergence point in mean shift algorithm) is determined by color as well as texture information unlike in standard mean-shift filtering. This provides better discrimination between regions where colors are similar but texture is different.
6. The output image can be segmented using K-means clustering algorithm.

The results will be given in the following section.

4. Results

The objective of the proposed algorithm is to achieve robust and accurate segmentation in images. The standard mean-shift filtering algorithm with color and spatial information are not sufficient for superior segmentation in cases

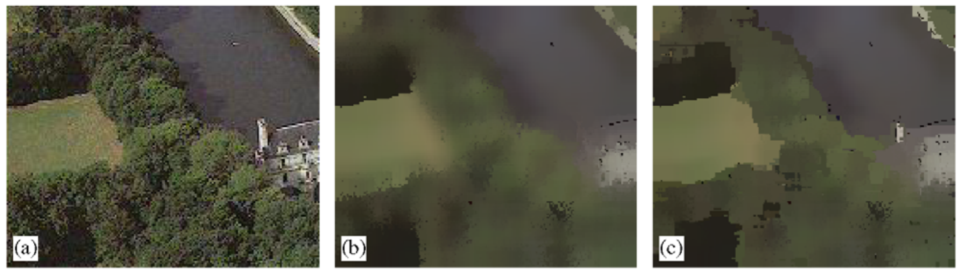


Fig. 4. The comparison of filtering results: (a) original image; (b) mean shift filtering result. Note that the region transitions are blurred; (c) texture supported filtering result. Note that the region boundaries are well-separated.

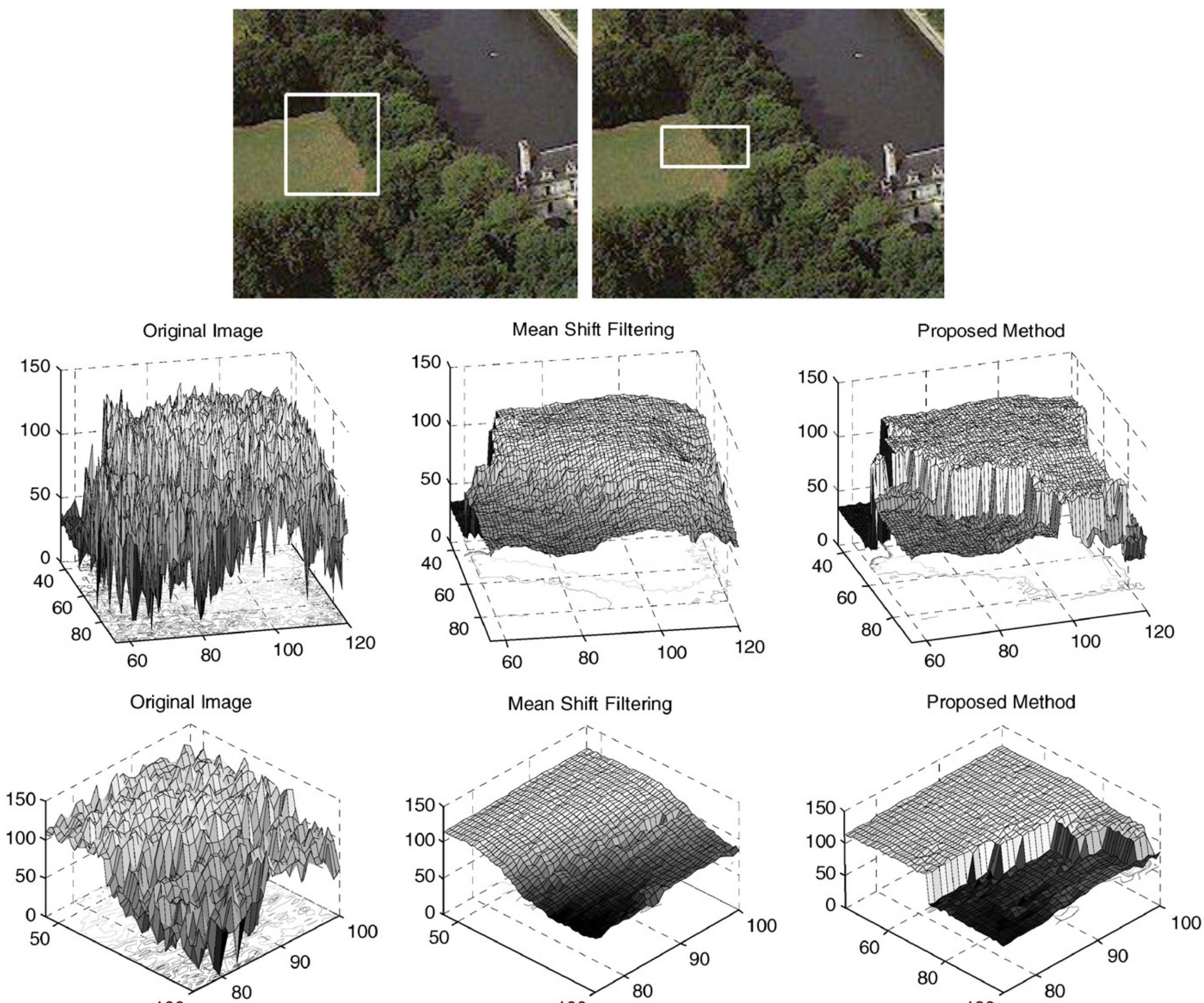


Fig. 5. Three-dimensional visualization of results. *First row*: original images with two different windows that the both standard mean shift and proposed filtering are applied. *Second and third row*: the proposed method provides well-separated regions which result in better segmentation.

where only color feature does not provide essential discriminative information. The proposed method addresses this problem and employs texture descriptors as an additional feature in a higher-dimensional feature space using mean-shift framework. The method uses wavelet frames that provide translation invariant texture analysis. This property

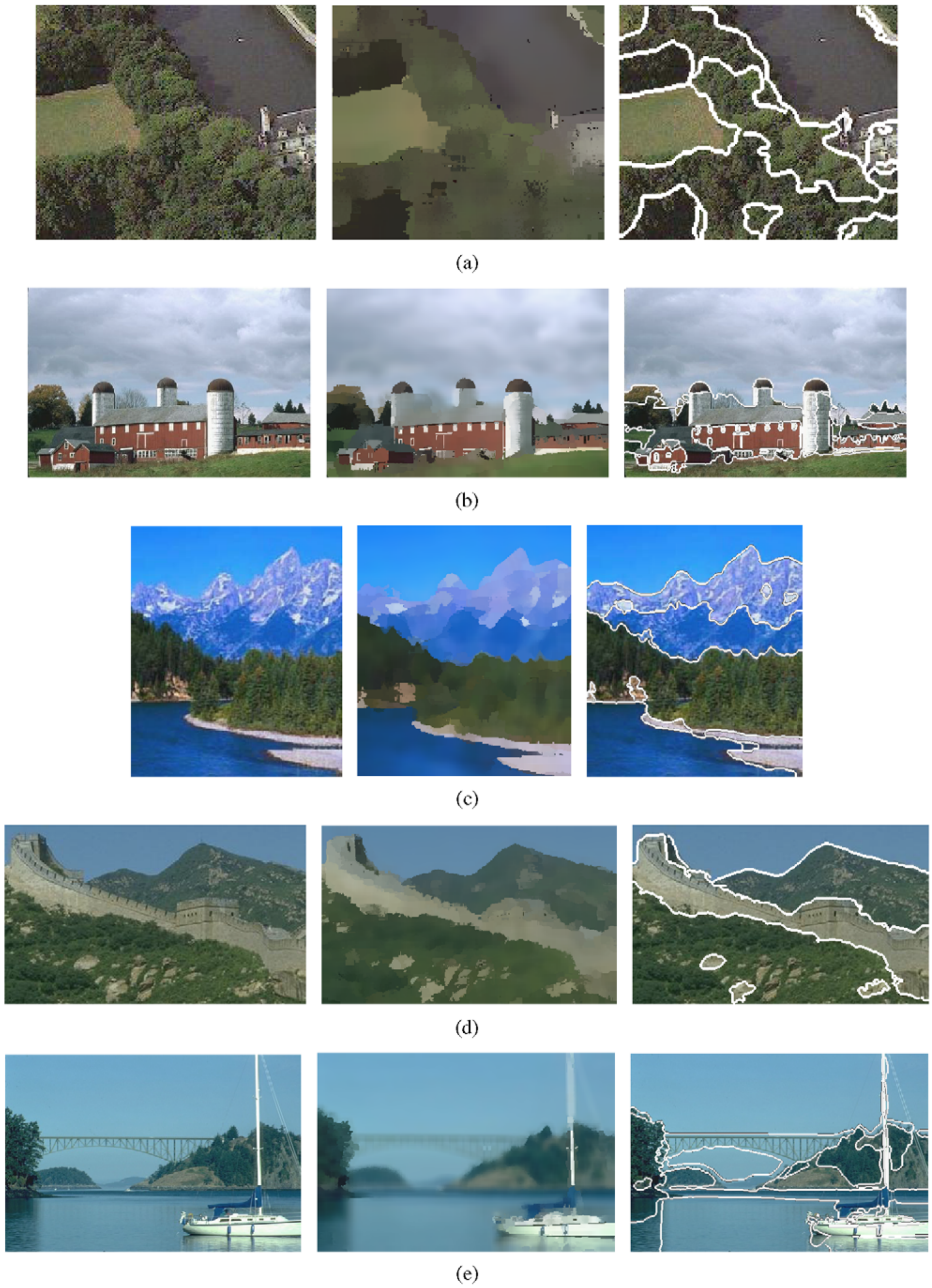
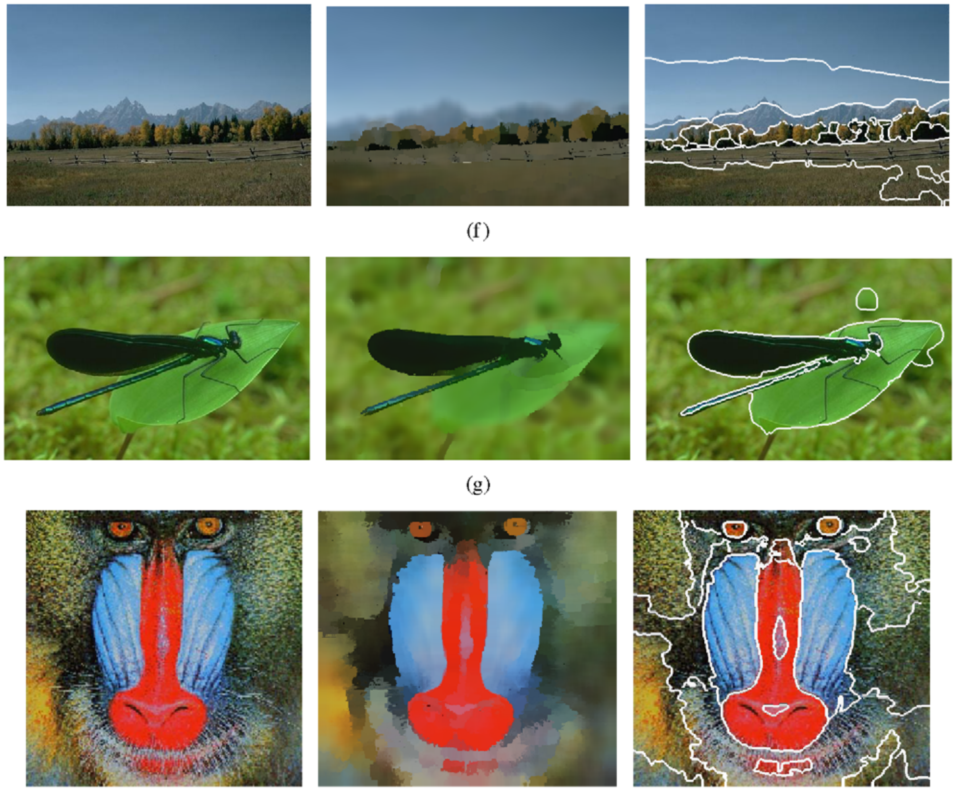


Fig. 6. Image segmentation results using proposed approach. Left images: original images. Middle images: filtered images. Right images: edges are superimposed on original images.

yields a better estimation of texture statistics and more detailed characterization at region boundaries. The method integrates additional texture feature to the color and spatial space of standard mean-shift filtering algorithm. The new algorithm with high dimensional extended feature space provides better results than standard algorithm as shown in experimental results. To demonstrate the performance of the algorithm, we experimented with a number of natural images. To illustrate the accuracy of the proposed algorithm, an image with trees, grass, river and a house is used as shown in Fig. 4(a). The standard (color-based) mean-shift filtering result is shown in Fig. 4(b) as a comparison. The filtering result using proposed algorithm is also shown in Fig. 4(c). It can be easily seen in Fig. 4(b) that the region transitions between trees and grass are smoothed and blurred which negatively affect robust segmentation of regions. More improved and separated regions can be obtained using proposed algorithm as shown in Fig. 4(c). The accuracy of the filtering output can be adjusted by changing the size of the texture window, color range and spatial window. For better visualization of



(f)

(g)

(h)

Fig. 6. (*continued*).

the results, a three-dimensional graphical demonstration is also shown in Fig. 5. The first row shows the original images marked with two different region of interest. The data in these windows are represented with (x, y) image coordinates and gray-level values in three-dimension as shown in second and third rows of Fig. 5. The integration of texture information to the mean-shift feature space provided more improved results particularly between regions due to the fact that the proposed approach contributes better estimation and more detailed texture characterization at region boundaries. More results are shown in Fig. 6. The original images are shown on the left, filtered images are shown in the middle and the segmented images are shown on the right columns. The regions smaller then a threshold were removed for better visualization.

Table 1
Comparison of number of misclassified pixels using mean shift and proposed algorithms

Image	Number of misclassified pixels		Error (%)	
	Mean shift	Proposed algorithm	Mean shift	Proposed algorithm
Fig. 2(a)	3988	2891	13.00	9.60
Fig. 6(a)	542	340	9.40	5.90
Fig. 6(b)	250	123	9.16	4.50
Fig. 6(c)	360	270	17.02	12.70
Fig. 6(d)	55	50	0.85	0.77
Fig. 6(e)	5548	1360	33.40	8.20
Fig. 6(g)	155	155	3.67	3.67
Fig. 6(h)	592	30	7.50	0.68

4.1. Quantitative analysis of results

In order to provide quantitative support and evaluate the performance of the proposed algorithm, the number of misclassified pixels in a region of interest is used. A misclassified pixel in a region can be defined as a pixel which does not belong to region of interest. The number of such pixels provides a performance measure for the algorithm. The segmentation/classification process depends on both the quality of the filtered image and the segmentation/classification algorithm itself. In this paper, the goal is to improve the filtered image such that better segmentation results can be obtained using well-known classification algorithms. On the other hand, instead of using well-known data classification methods, it is also possible to develop a specific segmentation algorithm for this purpose. Table 1 shows the quantitative comparison of results between mean shift and proposed algorithm in terms of number of misclassified pixels. The error is calculated using the following equation:

$$\text{Error} = \frac{\text{Number of misclassified pixels}}{\text{Total number of pixels in region}}. \quad (9)$$

The quantitative results indicate that the inclusion of texture feature to the mean-shift filtering algorithm reduces the number of misclassified pixels and thus helps to improve the segmentation process.

5. Conclusions

In this paper, we presented a new approach for image segmentation based on low-level features including color, texture and spatial information. The proposed approach is based on extending the feature space for filtering in mean-shift algorithm. The proposed method uses discrete wavelet frames (DWF) to characterize textured regions into pre-defined texture classes based on texture energy in different orientations. DWF decomposition of a textured region provides a translation invariant texture description which results in better estimation and more detailed texture characterization at region boundaries. This texture description is then fused into the mean-shift filtering framework to create higher dimensional feature space. The performance of the proposed approach has been demonstrated using natural images where color and texture information are available. The filtering with extended feature space provides satisfactory results particularly between regions. This results in more accurate segmentation process. The algorithm has also been compared to the standard mean-shift filtering. The results indicate that the proposed approach is more robust and accurate than standard mean-shift filtering.

References

[1] Y. Rui, T.S. Huang, S.F. Chang, Image retrieval: current techniques, promising directions and open issues, J. Visual Commun. Image Representation 10 (1999) 39–62.

[2] W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of early years, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1349–1379.

[3] M. Unser, Texture classification and segmentation using wavelet frames, IEEE Trans. Image Process. 4 (1995) 1549–1560.

[4] T. Randen, J.H. Husoy, Texture segmentation using filters with optimized energy separation, IEEE Trans. Image Process. 8 (1999) 571–582.

[5] G.V. Wouver, P. Scheunders, D.V. Dyck, Statistical texture characterization from discrete wavelet representations, IEEE Trans. Image Process. 8 (1999) 592–598.

[6] D. Comanicui, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002).

[7] J. Chen, T.N. Pappas, A. Mojsilovic, B. Rogowitz, Image segmentation by spatially adaptive color and texture features, in: Proceedings of IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain, September 2003.

[8] Y. Deng, B.S. Manjunath, Unsupervised segmentation of color texture regions in images and video, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 800–810.

[9] S. Belongie, C. Carson, H. Greenspan, J. Malik, Color and texture based image segmentation using EM and its application to content based image retrieval, International Conference on Computer Vision, 1998, pp. 675–682.

[10] S. Liapis, G. Tziritas, Color and texture image retrieval using chromaticity histograms and wavelet frames, IEEE Trans. Multimedia 6 (5) (2004).

[11] M. Haindl, Advances in Pattern Recognition, Lecture Notes in Computer Science, vol. 1451, Springer, Berlin, 1998, pp. 1021–1028.

[12] D.K. Panjwani, G. Healey, Markov random field models for unsupervised segmentation of textured color images, IEEE Trans. Pattern Anal. Mach. Intell. 17 (1995) 939–954.

[13] J. Wang, Stochastic relaxation on partitions with connected components and its applications to image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 619–639.

[14] L. Shafarenko, M. Petrou, J. Kittler, Automatic watershed segmentation of randomly textured color images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1997) 1530–1544.

[15] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 888–905.

[16] M.P. Wand, M. Jones, Kernel Smoothing, Chapman & Hall, London, 1995.

[17] Y. Cheng, Mean shift, mode seeking and clustering, IEEE Trans. Pattern Anal. Machine Intell. 17 (8) (1995) 790–799.

About the Author—EDIZ POLAT received the Ph.D. degree from Pennsylvania State University, University Park, PA in 2002. He was a member of the Penn State Computer Vision group where he worked in developing vision-based human–computer interaction systems funded by National Science Foundation and Department of Defence. He is currently an Assistant Professor in Electrical and Electronics Engineering Department at Kirikkale University, Kirikkale, Turkey. His research interests include computer vision particularly detection and visual tracking of objects in image sequences.

About the Author—MUSTAFA OZDEN received the B.S. and M.S. degree from Electrical and Electronics Engineering Department at Kirikkale University, Turkey in 2001 and in 2005, respectively. He is currently a Ph.D. student in Electrical and Electronics Engineering Department at Gazi University. His main research area consists of computer vision and image processing particularly segmentation and tracking of objects in video sequences.