A cluster-based wavelet feature extraction method and its application

Gang Yu a,⁎, Sagar V. Kamarthi b

a Department of Mechanical Engineering and Automation, Harbin Institute of Technology (HIT), Shenzhen Graduate School, HIT Campus Xili, Shenzhen University Town, Shenzhen, Guangdong 518055, PR China
b Mechanical and Industrial Engineering, Northeastern University, 360 Huntington Avenue, 334SN, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 8 November 2007
Received in revised form 26 November 2009
Accepted 27 November 2009
Available online 6 January 2010

Keywords:
Wavelet
Cluster based
Feature extraction
Texture classification
Surface defect

ABSTRACT

In this paper, a new cluster-based approach is proposed for extracting features from the coefficients of a two-dimensional discrete wavelet transform. The wavelet coefficients from the matrix of each frequency channel are segregated into non-overlapping clusters in an unsupervised mode using a set of application-specific representative images. In practical situations, this set of representative images can be the same as the ones kept aside for training a classifier. The proposed method divides the matrices of computed wavelet coefficients into disjoint clusters that are centered around the position of dominant coefficients. The features that can distinguish images of one class from those of other classes are obtained by computing energies of the clusters. The feature vectors so obtained are then presented as input patterns to an image classifier, such as a neural network. Experimental results based on the applications for texture classification and wood surface defect detection have shown that the proposed cluster-based wavelet feature extraction method is able to effectively extract important intrinsic information content from the test images, and increase the overall classification accuracy as compared with conventional feature extraction methods.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past few decades, we have witnessed a substantial interest in the development of feature extraction techniques, which has been motivated largely by the need for addressing pattern recognition and image processing problems. Selecting features that can effectively classify patterns is often a nontrivial task in many applications. Due to its excellent properties of time–frequency localization and adaptive multi-scale decomposition, the two-dimensional (2-D) discrete wavelet transform (DWT) proposed by Mallat (1989) has been applied extensively for feature extraction for image processing applications. A 2-D DWT can be viewed as an image filtered by a series of band-pass filters at different scales into a set of independent spatially oriented frequency channels. The image features are then extracted from the decomposed sub-images in different frequency channels and different resolution levels. Chen and Kuo (1993) proposed a tree-structured wavelet transform for texture classification. They used an energy criterion to select the sub-image for decomposition. A set of textural features was derived from the wavelet packet decomposition coefficients. Arivazhagan and Ganesan (2003) applied DWT for a set of texture images, and extracted both statistical features and co-occurrence features. It was found that the texture classification rate is significantly improved by combining wavelet-based statistical and co-occurrence features. The wavelet-based feature extraction methods were also applied to industrial material inspection such as LSI wafers (Lee et al., 1997; Maruo et al., 1999), cold rolled strips (Sari-Sarraf and Goddard, 1998), woven fabrics (Tsai and Huang, 2003), and copper clad laminate defects (Li, 2009).

It is obvious that all these wavelet-based image processing and analysis techniques benefit from the collective information extracted from the different frequency channels involved (Scheunders, 2003). Nevertheless, in most wavelet-based feature extraction methods the patterns within a frequency channel are assumed to be uniform. Such simplistic treatment of patterns ignores the details that lie within the boundaries of a frequency channel. Therefore, these classical feature extraction methods might not have enough discriminating ability to deal with the images with complex shapes, irregular texture, and a wide variability within the same class of patterns.

In previous work (Pittner and Kamarthi, 1999), we developed a new cluster-based feature extraction method for signals based on one-dimensional discrete wavelet transform, and verified the effectiveness of the method on a flank wear estimation problem in turning processes and on a lung sound classification problem.

In this paper, we extend the idea of cluster-based feature extraction from a one-dimensional signal to 2-D images. In this...
context, we develop a new clustering scheme that divides the matrices of computed 2-D wavelet coefficients into clusters at each scale. The energy of each of these clusters is treated as a feature that contains an useful piece of information about the image. The feature vector of an image thus computed serves as input pattern to a neural network for image classification. The results from the applications on texture classification and wood surface defect detection show that the discrimination performance obtained with the proposed cluster-based method is superior to that obtained using just channel energies from the standard matrices of 2-D wavelet coefficients.

In Section 2, we briefly review the basics of the 2-D discrete wavelet transform for image analysis and then introduce the proposed wavelet-based clustering method in Section 3. Sections 4 and 5 demonstrate the effectiveness of the proposed method by applications on texture classification and wood surface defect detection, respectively.

2. 2-D discrete wavelet transform

Let an image be viewed as a finite energy function \( f(x,y) \). Let \( Z \) denote the set of integers and \( R \) the real numbers. Let \( \{V_j\}_{j=-2}^\infty \) be a separable multi-resolution approximation of \( L^2(\mathbb{R}^2) \), and let \( \phi(x,y)=\phi(x)\phi(y) \) be the associated 2-D scaling function. If \( \psi(x) \) is the one-dimensional wavelet associated with the scaling function \( \phi(x) \), then the 2-D separable wavelets can be defined as \( \psi^1(x,y)=\psi(x)\psi(y) \), \( \psi^2(x,y)=\psi(x)\psi(y) \), and \( \psi^3(x,y)=\psi(x)\psi(y) \). For any wavelet decomposition scale \( J \geq 0 \), an image \( f(x,y) \) can be completely represented by the hierarchy of 3\( J+1 \) discrete images \( \{A^2_j,f,D^1_j,f,D^2_j,f,D^3_j,f\} \), \( 1 \leq j \leq J \), where

\[
A^2_j(f,m,n) := (f(x,y)=\phi_2(-x,-y)(2^{-j}m,2^{-j}n)) \quad (1)
\]

and

\[
D^p_j(f,m,n) := (f(x,y)=\phi^p(-x,-y)(2^{-j}m,2^{-j}n)) \quad p = 1, 2, 3 \quad (2)
\]

In (1) and (2), \( (m,n) \in Z^2 \). This set of images is called an orthogonal wavelet representation in two dimensions (Mallat, 1989). The wavelet decomposition of an image can then be interpreted as a set of independent, spatially oriented frequency channels. The schematic of 2-D wavelet decomposition for an original image is shown in Fig. 1, where the maximum decomposition scale \( J \) is taken to be 3 for simplicity in this illustration. It can be seen that the original image is decomposed into a pyramid structure composed of a smooth image \( A^2_J \), and the images \( D^1_j,f,D^2_j,f, \) and \( D^3_j,f \), which represent high frequency details in the vertical, horizontal, and diagonal directions, respectively, at resolution \( 2^j \) for \( j=1,2,3 \). Due to the orthogonality of this representation, the total number of coefficients in this new representation is equal to the number of coefficients in the original image.

3. A new clustering scheme for feature extraction

In this section, a novel clustering method for extracting features from the 2-D wavelet coefficient matrix of each frequency channel is introduced. The 2-D wavelet coefficients of each matrix are segregated into clusters in unsupervised mode by using a set of representative images that are specific to a given application. In a practical situation, this set of representative images can be the same as the ones kept aside for training the image classifier (e.g., a neural network). The following subsections describe the cluster formation method. The procedure proposed in this section divides each of the wavelet transform coefficient matrices into disjoint clusters \( U_1, U_2, \ldots, U_c \) from each of these clusters a single robust feature \( u_i \) \( (i=1,2,\ldots,c) \) is computed. The so obtained feature vector \( u_1, u_2, \ldots, u_c \) is considered the compact representation of the image from which these features are extracted; it serves as input patterns to an image classifier, such as a neural network.

3.1. Binary matrices for cluster determination

To determine the clusters, the proposed method uses a set of \( K \) images that represent the whole spectrum of possible types of images specific to an application.

First, the complete 2-D discrete wavelet transform is computed for all \( K \) representative images. As shown in Fig. 1, an image can be completely represented by 3\( J+1 \) coefficient matrices \( \{A^2_j,f,D^1_j,f,D^2_j,f,D^3_j,f\} \). The matrices

\[
\text{SD}^1_j := \sum_{k=1}^{K} D^1_j \quad (3)
\]

\[
\text{SD}^2_j := \sum_{k=1}^{K} D^2_j \quad (4)
\]

\[
\text{SD}^3_j := \sum_{k=1}^{K} D^3_j \quad (j=1, \ldots, J) \quad (5)
\]

are constructed as the sum of the corresponding wavelet coefficient matrices \( D^1_j,f,D^2_j,f, \) and \( D^3_j,f \) for the \( K \) representative images. For any matrix \( A_k \), we denote the sample mean and standard deviation of the elements of the matrix by \( \mu(A) \) and \( \sigma(A) \), respectively. Then, we construct the matrices

\[
\text{SD}^p_j := \frac{1}{\sigma(\text{SD}^p_j)} (SD^p_j - \mu(\text{SD}^p_j)) \quad (p=1,2,3; j=1,2,\ldots,J) \quad (6)
\]

and expect the elements of each matrix \( \text{SD}^p_j \) to be nearly \( N(0,1) \) distributed by the central limit theorem (Pittner and Kamath, 1999). Applying a threshold of the form \( T^p_j : = \sqrt{2\ln L^p_j / n^p_j} \), with \( \gamma = \varepsilon \) (is the Euler’s number), to the elements of the matrices \( \text{SD}^p_j \), \( L^p_j \) being the number of computed detail coefficients at each scale, we get the corresponding binary matrices

\[
\theta^p_j := (\theta(G^p_j - T^p_j)) \quad (p=1,2,3; j=1,2,\ldots,J) \quad (7)
\]

with the Heaviside function: \( \theta(x)=1 \) for \( x \geq 0 \) and \( \theta(x)=0 \) for \( x < 0 \). We expect \( \theta^p_j \) to contain at least \( \gamma / 2 \sqrt{\ln L^p_j} \) by Lemma 1 (Pittner and Kamath, 1999). Fig. 2 sketches the binary matrices
obtained by applying the thresholds $T_{2j}$ to the matrices $G_{2j}$ at each scale.

After we computed the binary matrices $B_{2j}$ from the wavelet coefficient matrices of $K$ representative images, we pass the binary matrices $B_{2j}$ through a clustering procedure to form the boundaries of the clusters. The clusters $U_1, U_2, ..., U_c$ as shown in Fig. 3 are determined once all the computed wavelet coefficients from each and every scale are labeled according to a scheme described in the next section. Here the clusters thus formed neither overlap across different scales nor cross the boundaries of different frequency channels.

3.2. Cluster boundary formation

The clustering procedure for the cluster boundary formation in each fixed binary matrix is illustrated in Fig. 4. It has the following steps.

Step 1: Label each occurrence of 1 in the binary matrix with a unique label. For example, if there are three 1 scattered in the matrix, label them, say, as A, B, and C; in this example the matrix is eventually divided into three clusters, but the cluster boundaries are determined using the process described in steps 2 and 3.

Step 2: Use the labeled 1 (e.g., A–C) as the “seed centers” to grow the clusters. In the first growth iteration, each seed center label is expanded by one cell to the left, one to the right, one to the top, and one to the bottom. Expansion in any direction is carried out only if the neighboring cell is 0; no expansion is carried out if the cell is either at the matrix boundary or if the neighboring cell is already labeled.

Step 3: Repeat the same growth pattern of labels until matrix has no more 0.

Step 4: Once all the 0 in the matrix are labeled, cluster boundaries are drawn respecting the homogeneity of the labels in each cluster. In other words, the boundaries are drawn at the interface of differently labeled clusters.

3.3. Feature extraction method

Through the procedure described in the previous subsections, the boundaries of the clusters $U_1, U_2, ..., U_c$ are demarked. Using this boundary template the feature vector of an image is computed in three steps. In the first step, the 2-D DWT is applied to the image to obtain the matrix of wavelet coefficients. In the second step, it is divided into clusters $U_1, U_2, ..., U_c$ according to the cluster boundary template. In the third step, the image features $u_1, u_2, ..., u_c$ are determined by simply computing the Euclidean norm of the clusters

$$u_i = \sqrt{\sum_{v \in U_i} v^2}$$

In other words, each feature $u_i$ is determined as the square root of the energy of the wavelet coefficients in cluster $U_i$. Consequently, the number of features, $c$, of an image is equal to the number of clusters determined with the method described in the previous section.

4. Application on texture classification

Texture analysis is a fundamental problem in computer vision and pattern recognition with a wide variety of applications such as remote sensing, medical image analysis, and quality inspection. A critical issue in texture analysis is how to characterize textural structure and define a perceptually meaningful distance/similarity measure between textures. Despite extensive research work that has been in progress for a long time, this issue remains unresolved. In this section, we apply the proposed cluster-based wavelet feature extraction method for a texture classification problem.

4.1. Experimental data set

The 12 textures from the Brodatz album (Brodatz, 1966) that are shown in Fig. 5 are equalized to 256 × 256 pixels and 256 gray
levels. Each image is divided into 16 disjoint $64 \times 64$ blocks, and each block is independently histogram-equalized to abolish luminance differences among textures. In addition, each original texture block is transformed into one additional block: a $64 \times 64$ scaled block obtained from the $45 \times 45$ pixels in the middle. The 12 texture categories included 384 blocks altogether. On half of these texture blocks are randomly selected for training the neural networks and the remaining blocks are used for evaluating the classification accuracy. The classification accuracy is evaluated on 10 different random selections of training data sets and tested on their complementary testing data sets.

4.2. Texture features and classification algorithms

Two sets of features, one based on the new clustering scheme and the other based just on the DWT, are extracted using the Haar wavelet transform at the maximum decomposition scale $J=6$. The texture images meant for training the neural networks are used to determine the cluster boundaries to form feature vectors $u_1, u_2, \ldots, u_c$. In this application, we obtained 28 clusters and hence 28 features using the proposed clustering scheme. We call this feature set F1. Using the traditional DWT, we obtained $19 (=3 \times 6 + 1)$ features that are computed by taking square roots of energy contents of frequency channels of the DWT. We call this feature set F2.

Two different neural network classifiers are used for generating the classification results. One is the Multilayer Perceptron (MLP) developed by Rumelhart and McClelland (1986), which is the most popular network architecture in use today. The other one is the Probabilistic Neural Network (PNN) proposed by Specht (1990). The original PNN is a direct neural network implementation of the Parzen nonparametric probability density function estimation (Parzen, 1962) and Bayes classification rule; it has the advantages of rapid training and guaranteed convergence (Wesserman, 1993).

4.3. Results

The evaluation of the classification accuracy based on the proposed clustering scheme and the standard DWT feature extraction methods is presented in Table 1. The results show that in most cases the classification accuracy based on feature set F1 is better than that based on feature set F2. These results indicate that the proposed wavelet-based clustering scheme provides additional discriminatory information for texture classification. In the texture classification experiments, the results again indicate that the PNN generally gives better classification accuracy than the MLP.

5. Application on wood surface defect classification

In machine vision-based classification of wood defects, feature extraction is the main problem. Geometric features (Alapuranen and Westman, 1992), texture-based features (Kim and Koivo, 1994), and color-based features (Lampinen, 1994) have been applied with limited success for classifying wood knots. In this section, we investigate the effectiveness of the proposed feature extraction method described in the previous section for a problem of distinguishing wood knots of spruce wood lumber.

<table>
<thead>
<tr>
<th>Training images</th>
<th>Testing images</th>
<th>Classification Accuracy (F1)</th>
<th>Classification Accuracy (F2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PNN</td>
<td>MLP</td>
</tr>
<tr>
<td>Bark</td>
<td>16</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>BeachSand</td>
<td>16</td>
<td>100.0%</td>
<td>75.0%</td>
</tr>
<tr>
<td>Beans</td>
<td>16</td>
<td>93.8%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Burlap</td>
<td>16</td>
<td>93.8%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Ice</td>
<td>16</td>
<td>81.3%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Paper</td>
<td>16</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Peb54</td>
<td>16</td>
<td>75.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Pigskin</td>
<td>16</td>
<td>87.5%</td>
<td>68.8%</td>
</tr>
<tr>
<td>Raffia</td>
<td>16</td>
<td>100.0%</td>
<td>93.8%</td>
</tr>
<tr>
<td>Reptile</td>
<td>16</td>
<td>93.8%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Watre</td>
<td>16</td>
<td>100.0%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Woodgrain</td>
<td>16</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Overall</td>
<td>192</td>
<td>93.2%</td>
<td>90.6%</td>
</tr>
</tbody>
</table>

Fig. 5. Twelve Brodatz textures for classification experiments.
Wood knots are the most commonly encountered defects in wood industry and have a crucial role in sorting lumber. A knot is the hard, cross-grained mass of wood at the place where a branch joins the trunk of a tree. Knots of the same kind may occur in different sizes, shapes, and orientations. Some examples of five wood knots are shown in Fig. 6.

5.1. Experimental data set

The color images of five types of wood knots are extracted from an image database (Machine Vision Group, 2004). We transformed all the images into the format of 64 x 64 pixels with 8 bit/pixel gray level. From a total of 329 wood knot images, we used 264 images for training the neural network classifiers and 65 for evaluating the classification accuracy.

5.2. Wood knot image features and classification algorithms

One of the main objectives of this study is to demonstrate how the proposed feature extraction method improves the texture classification accuracy relative to the methods that use little more than the DWT. Two sets of features, one based on the new clustering scheme and the other based just on the DWT, are extracted using the Haar wavelet transform at the maximum decomposition scale J=6. The wood knot images from the training set are used to determine the cluster boundaries to form feature vectors $u_1, u_2, \ldots, u_c$. Coincidentally, we obtained 28 features in feature set F1 and 19 features in feature set F2 just like in the previous application. The PNN and the MLP are again used for the wood knot image classification experiments.

5.3. Results

The evaluation of the classification accuracy for the wood knot detection experiments is presented in Table 2. Both feature sets, F1 and F2, based on a standard DWT feature extraction method, are used as input vectors to the MLP and PNN neural networks for classification. The results show that the overall classification accuracy based on feature set F1 is better than that based on feature set F2. The results also indicate that the PNN generally gives better classification accuracy than the MLP in the wood knot detection experiments as well.

6. Conclusions

In this paper, we presented a new clustering scheme based on the 2-D discrete wavelet transform, which especially deals with the classification problems using the features extracted from the wavelet coefficients of the images. The results from the applications on texture classification and wood knot classification have shown that the proposed wavelet-based clustering scheme can efficiently extract most of the problem-specific information content intrinsic in input images. The classification performance obtained with the proposed clustering scheme is superior to the performance obtained using just the energies of the standard 2-D discrete wavelet transform coefficients. The proposed method has a great potential for other image processing applications such as medical image analysis.
Appendix

Pseudocode for the clustering procedure:

```plaintext
procedure cluster_formation (var input_matrix: array [1.. N, 1.. M] of inregers);
    var
        cluster_matrix: array [1.. N, 1.. M] of zeros;
        border_matrix: array [1.. N, 1.. M] of boolean;
        delta_i: array [-1 0 1 0];
        delta_j: array [0 1 0 -1];
    begin
        cluster_no:=1;

        for i := 1 to N do
            for j := 1 to M do
                if input_matrix[i, j] = 1
                    cluster_matrix[i, j] := cluster_no;
                    border_matrix[i, j] := true;
                    cluster_no := cluster_no+1;
                endif
            end
        end

        total_cells := N*M - cluster_no+1;

        while total_cells > 0 do
            for i := 1 to N do
                for j := 1 to M do
                    if cluster_matrix[i, j] > 0 and border_matrix[i, j] = true
                        for k := 1 to 4 do
                            if cluster_matrix[i+delta_i(k), j+delta_j(k)] = 0
                                cluster_matrix[i+delta_i(k), j+delta_j(k)] := cluster_matrix[i, j];
                                border_matrix[i+delta_i(k), j+delta_j(k)] := true;
                                total_cells := total_cells - 1;
                            endif
                        end
                        border_matrix[i, j] := false;
                    endif
                end
            end
        end
end; {cluster formation}
```
References


