A Study on Texture Segmentation Towards Content-based Image Retrieval

Md. Khayrul Bashar

Department of Information Engineering, Nagoya University, Japan

Extended Abstract:

Texture segmentation is an important but challenging task in image analysis or computer vision applications. Among various cues, texture plays a vital role towards object recognition. Recent studies reveal the two popular methods for texture analysis: filter bank methods and Gray level cooccurrence matrices (GLCM). In this work, we have proposed several texture features in the spatial and transform domain as well as some approaches for texture segmentation and applications based on multi-channel filtering technique. Among them, the wavelet intermittency based salient points and wavelet transform-based locally orderless images (WLOIs) are remarkable. The later approach is versatile one, which may combine the filter bank methods with cooccurrence matrices for many applications. Contributions are elaborated below.

In the **first approach**, we are motivated by the behavior of human visual system (HVS) and proposed a "Block Processing Approach (BPA) of Cortex Transform", where filtering operation is performed inside a small block of data. The average energy in the frequency domain is calculated for each filter and assigned them as representative texture features of the center pixel of the block. The block size is fixed as 16 x 16 by a boundary verification experiment on several images. The sliding block operation can be performed by using overlapping or disjoint blocks. In our experiment, we performed overlapping sliding block operation for every pixel and hence we obtained a set of feature images. These feature images are applied to the classifier for the supervised and the unsupervised segmentation of input texture images. Note that the within block filtering operation and the octave scale for the filter placement reduces the feature space dimension in the proposed scheme. Again the feature computation is performed in the frequency domain, which eliminates the inverse Fourier transformation. An experiment with 12 images, natural scene images (Camera, Satellite, Brodatz's album) in the real environment as well as in the standard lighting conditions demonstrates the superior performance of the proposed BPA compared to the GLCM and DWF approaches. Confusion matrix analysis on the segmented images shows the average OA of 97.63 % for our BPA, while the same for GLCM and discrete wavelet frame (DWF) approaches are 76.1 % and 92.8% respectively. Another experiment on 16 images is performed for the visual comparison of the supervised (using minimum distance classifier) and unsupervised (using K-means clustering) schemes. Both the schemes use the same feature set generated by the BPA. Results highlight no remarkable differences between the two schemes. However, the proposed BPA performs quite well on the natural images compared to standard mosaic texture images, where the boundary and noise performance is relatively inferior.

In the **second method**, we have proposed three intensity contrast features, namely directional surface density (DSD), normalized sharpness index (NSI) and the normalized frequency index (NFI). DSD characterizes intensity variability in various directions, while NSI and NFI characterize the sharpness and frequency of this variation. An experiment on the standard and natural texture images shows that they are quite good for texture boundary extraction. However, they are less efficient for low frequency dominant images (some natural scenes). As we mentioned the BPA approach is superior for natural images, while they are inferior to reliable

boundary preservation of standard texture images. This symmetrically inverse behavior of the two descriptors above is combined through i) stacked vector technique and ii) correlation based technique for better segmentation. In the stacked vector approach, the feature vectors of the two descriptors are normalized before applying to the classifier. In correlation-based method, all feature images (from two descriptors) are divided into three groups designated as "similar", "dissimilar" and "between" respectively. All candidates in the "similar" and "dissimilar" groups are fused to two features through logical (AND, OR) operation. Finally, the reduced feature set is used with the classifier. Experiments on the **Brodatz** and **VisTex** images show better performance of the integration method with respect to accuracy and boundary performance. Confusion matrix analysis shows the following average OA:

1. **For 20 images**: 94.9 % by cortex features, 79.8 % by contrast features (DSD, NSI, NFI), and 97.7 % by combined feature.
2. **For 50 images**: 90.7 % by cortex features, 62.5 % by contrast features (DSD, NSI, NFI), and 93.4 % by combined features.

In the **third method**, we proposed a versatile framework of wavelet transform based Locally Orderless Images (WLOIs), which take the advantage of the discrete wavelet transform to reduce parametric redundancy of the existing Gaussian or Gaussian derivative based LOIs. WLOIs probabilistically better represent textural features. In this approach, we replaced the scaled or scaled derivative images of existing LOIs by wavelet sub-bands and allowed the system to vary the inner scale in a dyadic manner with inherent sub-band directions. This integration reduces three explicit parameters (derivative order (n), direction (θ) and inner scale (σ)) of the existing LOIs and allows user to effectively control the system by tuning only two explicit parameters: tonal scale (β) and outer scale (α). The isophote images corresponding to equally spaced coefficients are obtained from wavelet sub-bands using a non-linear transformation by Gaussian with bin-width β. Each isophote image is convolved with a Gaussian aperture having an extent α to obtain a WLOI. The direct WLOIs or the moments derived from them can be used as texture features. Experiments using Brodatz and VisTex images show an excellent performance of the direct WLOIs compared to existing LOIs, WLOI-based moments, conventional wavelet energy, and Gabor energy features. Confusion matrix analysis shows the following quantitative results: The average OA over **8 mosaic images** from Brodatz album is 1.99.44% for WLOIs, 2. 94.14% for WLOI based moments, 3. 97.39% for wavelet energy, and 4. 97.47 % for Gabor energy. While the same for **8 VisTex images** is 1.99.39% for WLOIs, 2. 93.74% for WLOI based moments, 3. 98.61% for wavelet energy, and 4. 95.24 % for Gabor energy. Over the combined data set, the performance order achieved is 1.WLOIs (99.41%), 2. wavelet energy (98%), 3. Gabor energy (96.35%), and 4.WLOI-based moments (93.97%). Another experiment on 14 separate mosaic images shows that the proposed WLOIs produce an average OA of 95.71% (using 3 scaled images (i.e., sub-bands)) , while the existing LOIs produce 94.79% (using 9 scaled images) for the same values of α and β parameters. A classification test is also performed over 5 different textures from Brodatz album using disjoint training (640 samples, 16x16 size) and test data set (also 640 samples, 16x16). An error counting on the test samples shows that the proposed WLOIs produces low (14.2%) misclassification error compared to the existing LOIs (15.8%). However, the proposed WLOIs achieve this performance at the cost of a little more computational time compared to wavelet or Gabor energy.

In the **fourth approach**, we introduced a wavelet domain technique which characterizes image texture based on the density and distribution of salient energy points in the wavelet sub-bands. The proposed features are designated as salient point density (SPD) and salient point distribution non-uniformity (SPDN). SPD approximately characterizes texture coarseness, while SPDN indicates the distribution of texture primitives. In this approach, we first obtain salient point images (binary) from the wavelet sub-bands using intermittency threshold of wavelet coefficients.

A small moving window is then applied at each pixel of the salient point images to compute SPD and SPDN features. SPD is obtained by counting average salient points in the window, while SPDN is extracted by computing chi-square statistic from the probability of salient points in the sub-blocks of window. Thus the obtained feature images (SPD, SPDN) is applied to the K-means clustering for unsupervised texture segmentation. This method produces flexible segmentation results with high computational efficacy. Experiments on Brodatz and natural images show the potentiality of the proposed features over conventional local extrema density or wavelet energy features. The performance order in terms of error rate, computed over 12 mosaic texture and 8 natural images, is (i) 3.86% for SPD, (ii) 6.24% for SPDN, (iii) 13.64% for wavelet energy and (iv) 20.6 % for local extrema density feature.

In fine, we have proposed three wavelet domain perceptual features, namely directionality, regularity and symmetry, which are integrated with supervised learning vector quantization technique (LVQ) for indexing and content-based retrieval of image database. The directionality is computed from the cross-correlation of wavelet coefficients across columns or rows. The regularity feature is computed using auto-correlation function on the region-based correlation sequence of sub-band coefficients. On the other hand, symmetry feature is extracted from the multiresolution edge images (obtained from detail sub-bands) by using a soft symmetric measure. The database is then categorized based on the above features using supervised learning vector quantization (LVQ). The retrieval is performed on the categorized sub-set of the entire database, which apparently reduces the query processing time for the large database. There is also a provision for the user feedback until he/she is not satisfied with the retrieved results. Currently we have applied the method on a small textile-curtain database (150 images) obtained from the SANGETSU Company, Japan. The primary experiment shows impressive results of the proposed system.

For better efficiency of our scheme, accurate categorization is necessary. Usually, the high classification accuracy for all categories ensures the minimum requirement of user feedback. For a total of 6 categories (150 images total, 25 per category), the order of the average classification accuracy is (1) 52.67 % for wavelet energy, (2) 60.67% for symmetry (S), (3) 72.67% for regularity (R), (4) 89.33% for directionality (D), and (5) 90% for the combined (D+R+E) feature, respectively.   Clearly, the combined feature indicates the highest individual classification accuracy.

In our experiment, the performance of the retrieval system is evaluated by analyzing   precision-recall graphs, which are constructed by retrieving images from each categorized set. Note that each categorized set per feature contains maximum number of relevant images for that class ensured by the minimal user feedback. Since, we do not consider the misclassified images in the present study, we cannot obtain 100% recall rate for the features. Thus the average (over 6 queries) interpolated recall rates for various features are:
1. Combined (D+R+E) (80%), 2. Directionality (D) (60%), 3. Regularity (R) (50%), and 4. Symmetry (S) (40%) and 5. Wavelet energy (E) (30%). Clearly, the (interpolated) precision rate exists for all features when recall rate is <= 30%, while we see it only for the combined and directionality features when recall rate is >50%. Precision-recall graph also shows that the directionality achieves the highest interpolated precision in the range of lower recall rate (i.e., <50%), while the same for the combined feature take a lead over the directionality for the recall rates >50%. However, the combined feature shows better precision-recall performance on an average.

In the future study, we will search for the misclassified images too, to obtain the standard 11 levels of recall rates from 0% to 100% with 10 intervals. We are currently modifying the approach to include shape and color information in our retrieval system with the design of necessary user interface. Incorporating machine learning techniques for relevance feedback and more semantic retrieval is also the desired future goal. We would also like to develop more practical multimedia system for various applications in future.