ADAPTIVE HISTOGRAMS AND DISSIMILARITY MEASURE FOR TEXTURE RETRIEVAL AND CLASSIFICATION

Fun Siong Lim and Wee Kheng Leow

School of Computing, National University of Singapore,
3 Science Drive 2, Singapore 117543, Singapore.
limfs, leowwk@comp.nus.edu.sg

ABSTRACT

Histogram-based dissimilarity measures are extensively used for content-based image retrieval. In an earlier paper [1], we proposed an efficient weighted correlation dissimilarity measure for adaptive-binning color histograms. Compared to existing fixed-binning histograms and dissimilarity measures, adaptive histograms together with weighted correlation produce the best overall performance in terms of high accuracy, small number of bins, no empty bin, and efficient computation for image classification and retrieval.

This paper follows up on the study of adaptive histograms by applying them to texture classification, retrieval, and clustering. Adaptive histograms are generated from the amplitude of the discrete Fourier transform of images. Extensive comparisons with well-known texture features and dissimilarity measures show that, again, adaptive histograms and weighted correlation produce good overall performance.

1. INTRODUCTION

Histograms are extensively used for content-based image retrieval. There are two methods of generating histograms: *fixed binning* and *adaptive binning*. Typically, a fixed binning method induces histogram bins by partitioning the texture space into rectangular bins. Once the bins are derived, they are fixed and the same binning is applied to all images. On the other hand, adaptive binning adapts the bins to the actual distributions of the images. As a result, different binnings are induced for different images.

In an earlier paper [1], we proposed an adaptive clustering algorithm for generating color histograms that adapt to the color content of an image. In addition, an efficient weighted correlation (WC) dissimilarity measure was proposed for comparing histograms with different binnings. Comprehensive tests show that, compared to existing fixed-binning histograms and dissimilarity measures, adaptive histograms with WC produced that best overall performance in terms of high accuracy, small number of bins, no empty bin, and efficient computation for image retrieval and classification.

This paper follows up on the study of adaptive histograms by applying them to texture classification, retrieval, and clustering. Discrete Fourier transform (DFT) of texture images are used as the raw texture features, and adaptive histograms are generated from the DFT amplitude. Extensive tests were performed to compare its performance with commonly used texture features and dissimilarity measures. The test results confirm that adaptive histograms with WC produce good overall performance.

2. RELATED WORK

Commonly used texture features can be divided into two main categories: statistical and spectral. Statistical features characterize textures in terms of local statistical measures (such as coarseness, directionality, contrast [2]), multiresolutional simultaneous autoregressive model (MRSAR) [3], and Markov random field (MRF) [4]. Among them, the features of Tamura et al. [2] are used in QBIC [5] and MRSAR is used in PhotoBook [6].

The spectral approach is based on the response of a set of band-pass filters, typically 2D Gabor [7] and wavelets [6, 7]. Each filter responds most strongly to the patterns at a specific spatial-frequency and orientation band. These features have been used in NeTra [7], VisualSeek [8], etc. In addition, features derived directly from discrete Fourier transform (DFT) has also been used in PhotoBook [6].

Various dissimilarity measures have been defined for comparing texture features. Euclidean distance is the most widely used. Mahalanobis has been used with MRSAR [6] and weighted mean-variance with Gabor [7].

3. ADAPTIVE BINNING

The Fourier transform of an image completely captures the information of the image at all spatial frequencies. The original image can be recovered exactly by performing an inverse Fourier transform. Therefore, texture features based on Fourier transform are expected to give better performance.

Discrete Fourier transform (DFT) produces a discretization of the frequency axes, which is equivalent to a rectangular (or square) binning of the frequency domain. It can be regarded as a fixed-binning histogram with a large number of rectangular bins, and with DFT amplitude serving as the bin counts. Similarly, Gabor features can be regarded as a fixed-binning histogram with overlapping elliptical bins of varying sizes. Both of them have their bins fixed in advanced, and the histograms of all texture images are induced using the same set of bins.

In contrast, adaptive binning derives the bins adaptively from the DFT of each texture image. Therefore, it can allocate bins according to the energy distribution reflected in the DFT. Let $A(p)$ denote the amplitude of the DFT channel p with frequency $\mathbf{f}_p = (u_p, v_p)$. Adaptive binning is achieved by applying a variant of k-means clustering that can automatically determine the appropriate number of clusters required [1]. The algorithm can be summarized as follows:

Adaptive Clustering

Repeat
  For each frequency channel p in DFT,
    Find the nearest cluster k to p.
    If no cluster is found or distance $d_{kp} \geq S$,
      create a new cluster with p in it;
    Else, if $d_{kp} \leq R$,
      add p to the cluster k.
  For each cluster i,
    If cluster i has a bin count of at least N_m,
      compute centroid of cluster i.
    Else, remove cluster i.

The distance d_{kp} between the centroids of cluster k and channel p is defined as the Euclidean distance:

$$d_{kp} = \left[(u_k - u_p)^2 + (v_k - v_p)^2 \right]^{1/2} \tag{1}$$

where $\mathbf{f}_k = (u_k, v_k)$ is the centroid of cluster k. Note that, in computing cluster centroids, it is necessary to take into account the channel amplitude $A(p)$ so that the algorithm can position cluster centroids near local amplitude peaks:

$$\mathbf{f}_k = \sum_p A(p)\, \mathbf{f}_p \tag{2}$$

where the summation is performed over all frequency channels p in cluster k.

The adaptive clustering algorithm groups a frequency channel p into its nearest cluster if it is near enough ($d_{kp} \leq R$). On the other hand, if p is far enough ($d_{kp} \geq S$) from its nearest cluster, then a new cluster is created. Otherwise, it is left unclustered and will be considered again in the next iteration. This clustering algorithm, thus, ensures that each cluster has a *maximum radius* of R and that the clusters are separated by the distance of approximately S called the *nominal cluster separation*. The value of S is defined as a multiple γ of R, i.e., $S = \gamma R$. Reasonable values of γ range from 0 (for complete overlapping of the clusters) to 2 (for non-overlapping of clusters).

Since the algorithm creates a cluster only when a frequency channel is far from all existing clusters, it can determine the number of clusters required to effectively represent the DFT. It also ensures that each cluster has a significant bin count (i.e., amplitude) of at least N_m; otherwise, the cluster is removed. In the current implementation, N_m is fixed at $1/2500$ of the total amplitude of the DFT $\sum_p A(p)$.

This adaptive clustering algorithm is similar to that of Gong et al. [9] Both algorithms ensure that the clusters are not too large in volume and not too close to each other. However, our adaptive algorithm is simpler than that in [9]. Moreover, it does not require seed initialization, and can automatically determine the appropriate number of clusters.

In the current implementation, the algorithm is repeated for 20 iterations. When the algorithm terminates, some channels may still be unclustered. Experimental results show that these channels do not significantly affect the overall performance because they have very low energy levels.

4. WEIGHTED CORRELATION

An adaptive histogram $H = (n, \mathcal{C}, \mathcal{H})$ is defined mathematically as a 3-tuple consisting of a set $\mathcal{C}$ of n bins $\mathbf{c}_i$, $i = 1, \ldots, n$, and a set $\mathcal{H}$ of corresponding bin counts $h_i \geq 0$. The similarity $w(\mathbf{b}, \mathbf{c})$ between bins $\mathbf{b}$ and $\mathbf{c}$ is given by a monotonic function inversely related to the distance $d(\mathbf{b}, \mathbf{c})$ between them. For color histograms, the weight $w(\mathbf{b}, \mathbf{c})$ is defined in terms of the volume of intersection V_s between the bins [1]. For texture DFT, which is 2D, it can be defined in terms of the area of intersection between circular bins. ¿From elementary geometry, the similarity $w(\alpha)$ between two circular bins of radius R separated by a distance $d = \alpha R$ between their centroids can be derived as:

$$w(\alpha) = \begin{cases} \dfrac{2}{\pi} \cos^{-1} \dfrac{\alpha}{2} - \dfrac{\alpha}{\pi} \sqrt{1 - \dfrac{\alpha^2}{4}} & \text{if } 0 \leq \alpha \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

The weighted correlation $G \cdot H$ between histograms $G = (m, \{\mathbf{b}_i\}, \{g_i\})$ and $H = (n, \{\mathbf{c}_i\}, \{h_i\})$ is defined as:

$$G \cdot H = \sum_{i=1}^{m} \sum_{j=1}^{n} w(\mathbf{b}_i, \mathbf{c}_j)\, g_i h_j \,. \tag{3}$$

For a histogram H, its norm $\|H\| = \sqrt{H \cdot H}$, and its normalized form $\overline{H} = H / \|H\|$. The similarity $s(G, H)$ between histograms G and H is $s(G, H) = \overline{G} \cdot \overline{H}$, and the dissimilarity $d(G, H) = 1 - s(G, H)$.

5. QUANTITATIVE EVALUATION

Comprehensive tests were conducted to evaluate the performance of adaptive histogram and weighted correlation in texture retrieval, classification and clustering.

5.1. Parameter Setting

To determine the appropriate values for the adaptive clustering parameters R and $\gamma = S/R$, an exhaustive test was performed with various combinations of R and γ values on all 112 of the texture images in the Brodatz album. Test results show that the larger the bin radius R and bin separation ratio γ, the smaller is the number of bins but the larger is the mean frequency error. Mean error in frequency quantization does not improve significantly for number of bins larger than about 100. This point corresponds to a cluster radius R of 10 and separation ratio γ of 1.4.

5.2. Texture Retrieval

This test assessed the performance of various texture features and dissimilarity measures in image retrieval. Each of the 112 Brodatz images was partitioned (with overlapping) into 25 sub-images of size 128×128 pixels, giving a total of 2800 sub-images. Each sub-image was used, in turn, as a query image and other sub-images from the same class were regarded as relevant.

The test was performed with the following feature and dissimilarity combinations:

- L2: 30-channel Gabor features with Euclidean distance (L2). This combination served as the base case.
- WMV: 30-channel Gabor with weighted mean-variance. Each channel has two values: mean and variance.
- MRSAR: MRSAR with Mahalanobis distance.
- TWT: Tree-structured wavelet transform with L2.
- WC: Adaptive histograms of DFT with weighted correlation dissimilarity.

We also tested MRSAR with Euclidean distance and obtained very poor results. Therefore, this combination is omitted in the following discussion.

Figure 1 plots the precision-recall curves of the texture retrieval results. TWT and L2 had the worst performance. MRSAR performed slightly better, followed by WC and WMV showed the best performance. The performance of WC was very close to the performance of WMV.

The test results are consistent with those obtained by Manjunath and Ma [7] and by Liu and Picard [6]. In particular, [7] shows that Gabor and MRSAR have comparable retrieval performance and are better than those of pyramid and tree-structured wavelet transforms. [6] shows that the Wold model (MRSAR + DFT) performs better than MRSAR alone, and both have significantly better retrieval performance than TWT and Tamura features.

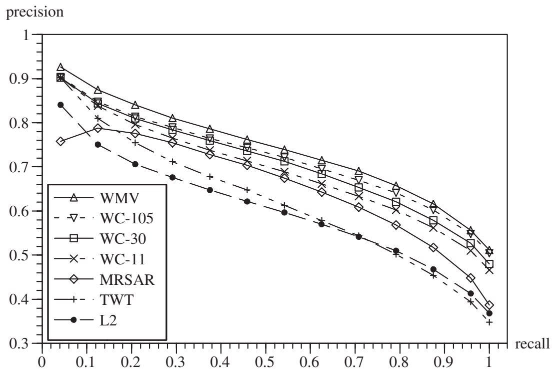

Fig. 1. Retrieval performance of various texture features and dissimilarity measures.

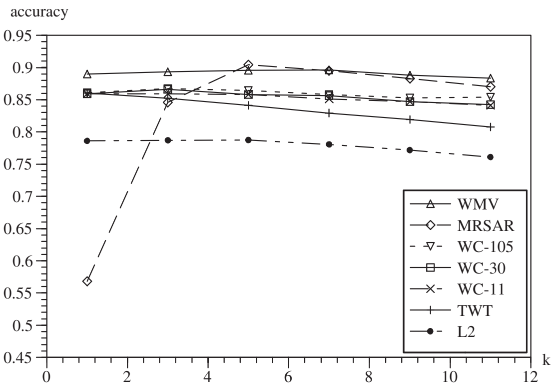

Fig. 2. Classification accuracy for various texture features and dissimilarity measures.

5.3. Image Classification

This test assessed the performance of various texture features and dissimilarity measures in classification. The k-nearest-neighbor classification with leave-one-out procedure was applied on the set of 2800 texture sub-images. Odd values of k ranging from 1 to 11 were chosen to reduce the possibility of ties. Classification error, averaged over all sub-images, were computed for each texture feature and dissimilarity combination described in Section 5.2.

Figure 2 shows that L2 and TWT had the worst performance. The performance of WMV was the best and WC (for all three cases) followed closely. The 1-nn classification of MRSAR was very inaccurate. However, with larger k values, its performance improved significantly. This shows that Mahalanobis distance of MRSAR is less reliable than other measures in finding the nearest neighbors of a texture.

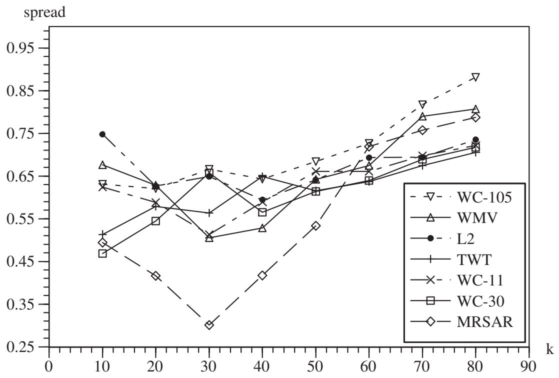

Fig. 3. Spread of clusters of various texture classes

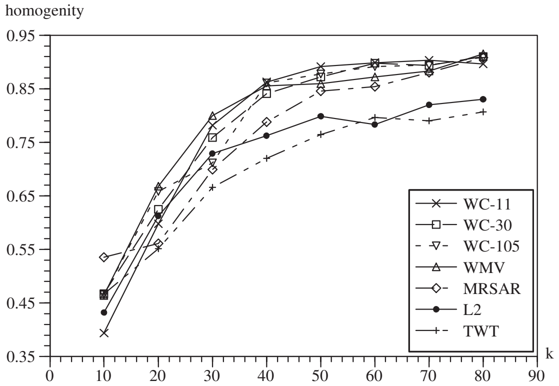

Fig. 4. Homogeneity of clusters various texture classes

5.4. Image Clustering

This test assessed the performance of various texture features and dissimilarity measures in clustering. 40 texture types of 20 images each were used. Clustering was performed on these 800 texture images with varying number of clusters k. The spread and homogeneity of each cluster was computed.

Spread measures how well clustered are the clusters. It is given by the average intracluster distance over the distance to the nearest cluster. The lower the spread ratio, the more compact the clusters are. Homogeneity measures the proportion of textures in a cluster that belong to the majority class of the cluster.

Figure 3 shows that WC-105 had the worst performance in terms of spread. This was probably due to the large number of bins. WMV was the second worst while MRSAR did well for small number of clusters but poorly for large number of clusters. WC-30 and WC-11, TWT, and L2 were more consistent than the rest and remained relatively well-clustered with increasing number of clusters.

Figure 4 shows that the performance of WC and WMV were almost similar in terms of homogeneity. MRSAR performed slightly worse while TWT and L2 were least homogeneous.

6. CONCLUSION

This paper presented the application of adaptive histograms for texture retrieval, classification, and clustering. It showed that, adaptive histograms and weighted correlation (WC) dissimilarity consistently perform well for classification, retrieval, and clustering. Together with our previous paper [1], we have shown that adaptive histograms have good performance in terms of high accuracy, small number of bins, no empty bins, and efficient computation for both color and texture classification and retrieval. Therefore, this method has a great potential for measuring the dissimilarity of other features. We are now pursuing these other applications.

7. REFERENCES

[1] W. K. Leow and R. Li, "Adaptive binning and dissimilarity measure for image retrieval and classification," in *Proc. IEEE CVPR 2001*, 2001.

[2] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. SMC*, vol. 8, no. 6, pp. 460–47, 1978.

[3] J. C. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognition*, vol. 25, pp. 173–188, 1992.

[4] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. PAMI*, vol. 5, pp. 25–39, 1983.

[5] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC project: Querying images by content using color, texture, and shape," in *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases*, 1993, vol. 1908, pp. 173–181.

[6] F. Liu and R.W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. PAMI*, vol. 18, no. 7, pp. 722–733, 1996.

[7] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. PAMI*, vol. 8, no. 18, pp. 837–842, 1996.

[8] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *Proc. ICIP '95*, 1995.

[9] Y. Gong, G. Proietti, and C. Faloutsos, "Image indexing and retrieval based on human perceptual color clustering," in *Proc. CVPR '98*, 1998.