Texture Retrieval with Descriptors Based on Local Fourier Transform: Comparing the Rectangular and Circular Neighbourhoods

AHSAN AHMAD URSANI*, WAJIHA SHAH**, AND SYED ASIF ALI SHAH***

RECEIVED ON 01.10.2010 ACCEPTED 03.01.2011

ABSTRACT

The texture descriptors derived from 1-D DFT (Discrete Fourier Transform) of the pixel values of a local neighbourhood have been shown to perform better than the methods based on wavelets for image retrieval and recognition. These DFT-based texture descriptors were extracted from rectangular or circular neighbourhoods. This paper compares the texture descriptors extracted from rectangular and the circular neighbourhoods previously proposed in the literature. A database of images is constructed from Brodatz album and the texture descriptors extracted from the two types of neighbourhoods are compared for texture retrieval. This paper shows that extracting DFT-based features from circular neighbourhood is almost thrice as expensive as extracting the same from the rectangular neighbourhood. The results of image retrieval on a large image database show that the descriptor extracted from rectangular neighbourhoods performs better than the same extracted from the circular neighbourhoods.

Key Words: **Texture Description, DFT-Based Texture Features, Circular Neighbourhood, Rectangular Neighbourhoods, Content-Based Image Retrieval.**

1. INTRODUCTION

Texture is an important image-content processed in Computer graphics, image retrieval, artificial vision systems, bio-medical imaging, land-use classification in remote sensing and many more applications. There are two different ways in which the texture is processed, i.e. texture analysis and texture synthesis. The texture is analysed either for recognition or retrieval. This paper concerns with texture analysis for the sake of retrieval.

Varma and Zisserman [1] divided the texture features into two broad categories, i.e. ones which put emphasis on the local properties of the texture-image and others that are based on large filter banks. The first category included

* Associate Professor, Department of Telecommunication Engineering, Mehran University of Engineering and Technology, Jamshoro.

** Assistant Professor, Department of Electronic & Biomedical Engineering, Mehran University of Engineering and Technology, Jamshoro.

*** Assistant Professor, Department of Electrical Engineering, Mehran University of Engineering and Technology, Jamshoro.

the texture description based on Markov random fields, local binary pattern operator, local Fourier transform, and spatial grey level co-occurrence matrix, whereas second category included texture description based on wavelets, which use spatial-frequency representations. They [1] found that those texture descriptors from the first category perform better than those from the second category. They also argued that the texture description based on pixel neighbourhoods as small as 3x3 can prove even better than those emphasizing the spatial frequencies, because texture is the characteristic of the immediate pixel neighbourhood. Ursani, A.A., et. al. [2] also found that texture descriptor based on local Fourier transform outperform the descriptor based on Gabor wavelets.

The texture descriptors based on discrete Fourier transform using rectangular neighbourhood and circular neighbourhood and were introduced in [3-4], respectively. Since then, these texture features have been used for several applications including segmentation of multibeam echosounder data [5]. This paper compares the texture descriptors extracted from the two neighbourhoods. Results are presented on their overall retrieval accuracy in the form of precision vs. recall curve.

1.1 Rectangular Neighbourhood

DFT-based texture descriptor extracted from rectangular neighbourhoods was proposed in [3]. Fig. 1 shows the 8-pixel sequence whose grey levels form an 8-value vector to be transformed using the DFT. In an PxQ pixel image, there are (P-2)x(Q-2) central pixels. The method proposed in [3] computes DFT of the sequence of grey-levels of the 8 pixels around each central pixel in an image. Later, it computes histograms of the magnitudes of the complex values of X_0, X_1, X_2, X_3, and X_4 DFT coefficients. These histograms, called Local Fourier Histograms (LFH) were used to describe the texture. The texture descriptor consists of 40 values since each of the DFT coefficients is quantized into 8 bins. Ursani, A.A., et. al. [6] introduced phases of the DFT coefficients as rotation invariant texture features in addition to the absolute values.

1.2 Circular Neighbourhood

The absolute values of DFT coefficients are also susceptible to changes if the image is rotated, since the pixel values in the rectangular neighbourhood get altered in the rotated image. Therefore, the absolute values of the DFT coefficients are not fully rotation invariant.

Arof and Deravi [4] suggested a circular neighbourhood instead of the rectangular for extraction of similar texture features based on 1-D DFT. It argues that because rotating the image alters the pixel values in the rectangular 9-pixel neighbourhood, and therefore, the features extracted from the rotated and non-rotated image are dissimilar. On the other hand, the pixel values in the corresponding 9-pixel circular neighbourhood remain unaltered in the wake of image rotation.

Fig. 2 shows the circular neighbourhood with the neighbours 1, 3, 5 and 7 (shown using empty circles) having non-integer pixel coordinates. Arof and Deravi [4] interpolate these neighbours of non-integer coordinates using inverse Euclidean distance as explained in Equations (1 and 2). For example, the pixel value at x_1 (having the coordinates X=0.707 and Y=0.707) is interpolated between the four closest neighbours of integer coordinates, i.e. P0(1,0), P1(1,1), P2(0,-1), and C(0,0) in the rectangular neighbourhood.

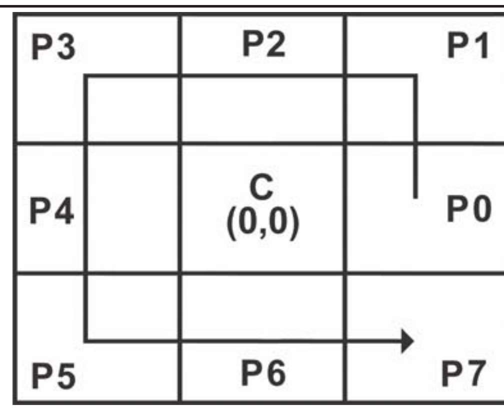

FIG. 1. THE 9-PIXEL RECTANGULAR NEIGHBOURHOOD IN THE SPACE DOMAIN

$$x_m = \frac{1}{\sum_{n=1}^{4} d_{m,n}} \sum_{n=1}^{4} d_{m,n} x_n \qquad (1)$$

Where m is a 2-D non-integer coordinate in the circular neighbourhood, n is a 2-D integer coordinate in the rectangular neighbourhood, xm is the interpolated value at a non-integer pixel coordinate m, xn is value of one of the four closest neighbours in the rectangular neighbourhood, dm,n is the inverse Euclidean distance between the coordinates m and n, given as in Equation (2).

$$d_{m,n} = \frac{1}{\|m-n\|} \qquad (2)$$

In addition, Arof and Deravi [4] suggest using mean and variance of the feature images of the DFT coefficients X_0 through X_4 instead of computing their histograms and using bins values as features as in [3]. However, since this paper aims at comparing only the rectangular and the circular neighbourhoods, we extract exactly the same descriptor as the one explained in Section 1.1 from the two types of neighbourhood.

The rest of the paper comprises Section 2 that explains the datasets used and the experimental setup, Section 3 compares the computational cost of the methods involved, Section 4 presents the results, and Section 5 concludes the findings.

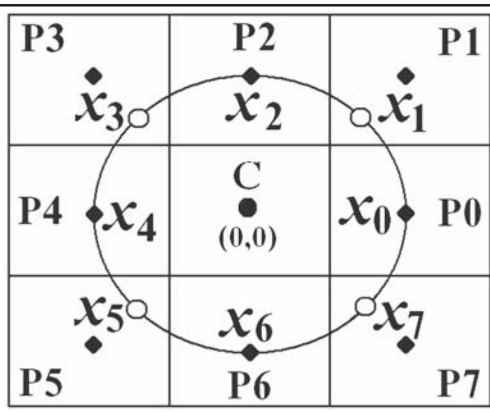

FIG. 2. THE CIRCULAR NEIGHBOURHOOD AS SUGGESTED IN [4]

Evaluating a texture descriptor requires two separate datasets; one for training the classifier, and other for testing its performance. The following subsections explain separately the image database to be searched and the query image set used the results reported in this paper. The retrieval experiments use the query images to train the classifier and then search the database of images for the closest matches.

2.1 The Training and the Query Image-Set

We downloaded texture images of Brodatz album from a webpage belonging to the University of Stavanger. The downloaded images measured 640x640 pixels each. Realising that the Brodatz album in fact contains several images of a given kind (woven wire, paper, canvas, cloth, stone, brick wall, water, flower, skin, etc) with varying zoom and lighting conditions etc., only a single image from each kind is included in the query image set, resulting in an image-set comprising 32 texture classes shown in Fig. 3.

Each of these images is a texture class. The features vectors extracted from these 32 images are used for training the NN (Nearest Neighbour) classifier to produce the results presented in this paper. NN is a highly flexible classification process that does not involve any parametric modelling of the training data. This can offer both space and speed advantages in very large problems [7]. Size of the original downloaded images was 6402 pixels each. These were divided into 9 images of 2102 pixels each. One feature vector was extracted from each of the 9 sub-images from each of the training images. The mean of the 9 feature vectors was used as the 10th feature vector representing a class. This forms 32 LUTs (Lookup Tables), each representing a texture class and was used as a query in the process of retrieval. Each LUT contains 10 feature vectors (i.e. representatives) of each texture class. Two separate LUTs are formed for each training image, one from each i.e. rectangular neighbourhood and the circular neighbourhood.

The image database to be searched is formed by dividing each of the 32 Brodatz images into 16 sub-images each measuring 1602 pixels. This gives a total of 16x32, i.e. 512 sub-images. Each of these 512 images is then rotated to 0, 15, 30, 45, 60, 75, and 90° using bilinear interpolation and the central part from each measuring 1022 pixels is cropped. This resulted in a database comprising 3584 images, containing 112 siblings from each of the 32 classes and oriented at seven different angles. Each of these database images is represented by a single feature vector.

2.3 Image Retrieval

Retrieval is the process of sorting the objects in the order of relevance. In the image retrieval, an image database is searched for the N closest matches of a query image. Performance of the retrieval process is measured in terms of precision and recall [8-9] computed as in Equations (3 and 4).

$$recall = 100 * \frac{R}{T} \tag{3}$$

where R is the number of relevant images retrieved and T is the total number of relevant images in the database. Equation (4) gives the precision.

$$precision = 100 * \frac{R}{N} \tag{4}$$

where N is the total number of images retrieved.

The similarity criterion used for finding the matches is the cross correlation coefficient. Since each query image is represented by a set of 10 feature vectors, the coefficient of correlation is computed between each of the 3584 feature vectors representing the database images and each of the 10 feature vectors representing a query image. Hence 10 similarity values are computed between each database image and the query image.

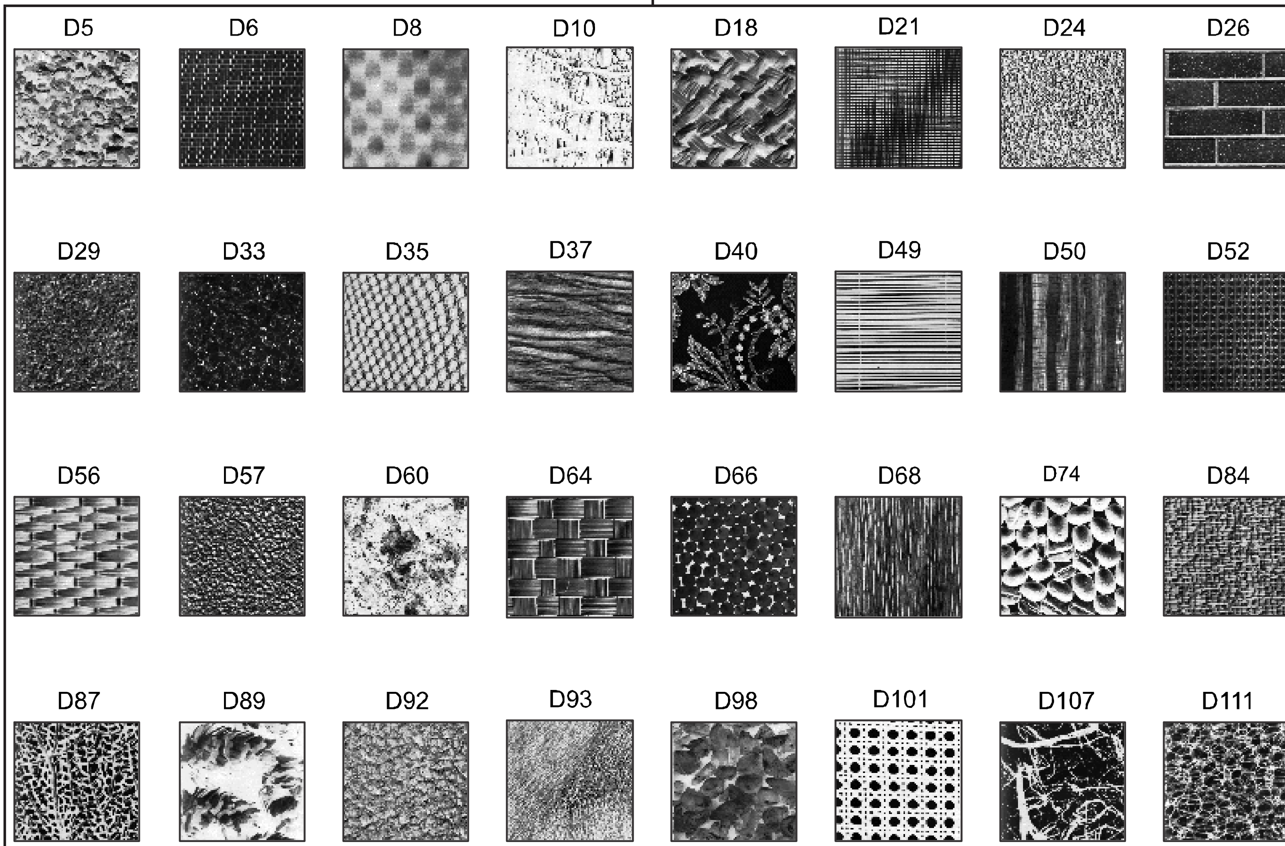

FIG. 3. THE 32 IMAGES FROM BRODATZ ALBUM SELECTED TO FORM THE IMAGE DATABASE

Retrieval is performed using the NN classification. The highest correlation coefficient is used for sorting the relevant images from the database.

3. COMPUTATIONAL COST ANALYSIS

Since the method of extraction from the two types of neighbourhoods basically remains the same, the only difference is the additional cost of interpolation that is required in the use of circular neighbourhood. The cost of computing 8-point FFT (Fast Fourier Transform) is given by:

$$C_{FFT} = O(N \log_2 N) \tag{5}$$

Since N=8, in our application, the complexity becomes:

$$C_{FFT} = O(8 \log_2 8) = O(24) \tag{6}$$

which shows that one requires 24 multiplications and 24 addition operations to compute DFT of 8-pixel values around a central pixel.

Interpolating pixel values at a non-integer coordinate requires 4 multiplications (i.e. M=4), 8 additions (i.e. A=8), one division (i.e. D=1), and a Euclidean distance calculation. One Euclidean distance calculations requires two additions (A=2), two multiplications (M=2), and a square root operation (S=1). This gives the total of 6 multiplications (M=6), 10 additions (A=10), one division, and one square root operation. Although the square root and division operations are more complex than the multiplication, but if one approximates each square root and each division by two multiplications, we have M=10 and A=10. Since one requires interpolating pixel values at 4 non-integer coordinates for computing DFT around each central pixel. Therefore, the complexity of four interpolation operations around a single central pixel is given by:

$$C_{INTP} = 0(40) \tag{7}$$

This shows that interpolating the pixel values at non-integer coordinates is twice as expensive as extracting the DFT-based features itself. Therefore, extraction of DFT-based features from circular neighbourhoods is thrice as expensive as extraction of DFT-based features from rectangular neighbourhoods.

4. RESULTS

The results of image retrieval performed on the dataset containing 3584 images with each of the 32 training images used as a query image. N closest matches of a query image were sought from the dataset containing a total of 3584 images, out of which only 112 were relevant to each query and the rest were irrelevant. In the reported experiments, N takes integer values from 220212 separated by 10. The resulting 22 pairs of precision-recall values are plotted in the form of PR (Precision vs. Recall) curve.

Fig. 4 presents the results of the retrieval using the NN classification. This PR curve shows the retrieval results averaged over all the 32 queries. It can be seen that the LFH features extracted from the rectangular neighbourhood perform better than those extracted from the circular neighbourhood. The curve representing rectangular neighbourhood shows greater recall for a given precision and a higher precision for a given recall.

5. CONCLUSION

This paper presented the results of image retrieval on a large image database containing 3584 images from 32 classes. Despite being computationally much more expensive, the features extracted from interpolated circular neighbourhood prove less efficient than those extracted from rectangular neighbourhood. The reason of this failure is that there is a pitfall in the hypothesis of the circular neighbourhood itself. The interpolated neighbours also come from the same pixel values of the 9-pixel rectangular neighbourhood that is unstable to the rotation. This paper however used only amplitudes of the DFT coefficients. Future work may investigate the performance of the phases of the DFT coefficients extracted from rectangular and the circular neighbourhoods as the texture features.

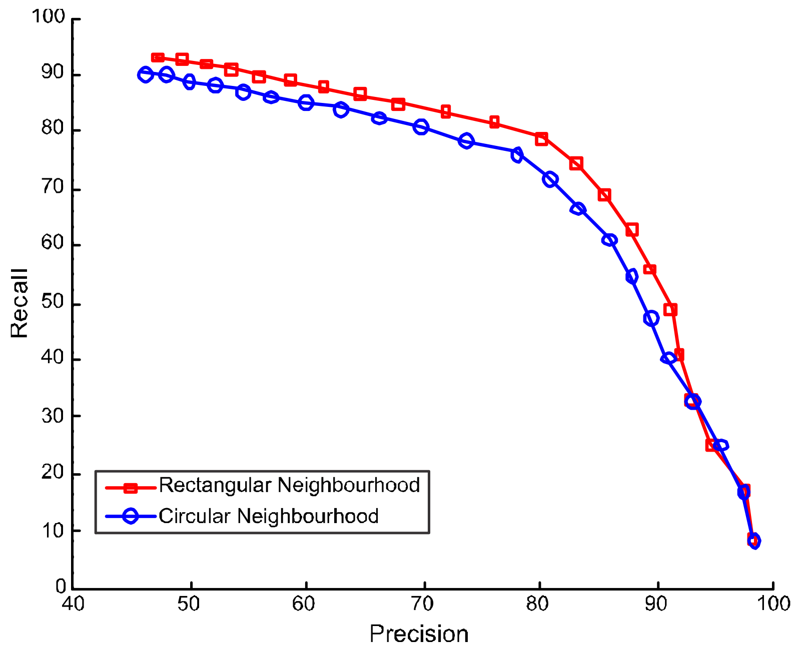

FIG. 4. RETRIEVAL VS. RECALL CURVES

ACKNOWLEDGEMENT

The authors are thankful to Mr. Trygve Randen, University of Stavanger, for providing the Brodatz Images for this work. The images were downloaded from his web page: http://www.ux.uis.no/~tranden/

REFERENCES

[1] Varma, M., and Zisserman, A., "Texture Classification: Are Filter Banks Necessary?", Proceedings of Conference on Computer Vision and Pattern Recognition, Volume 2, pp. 691-698, 2003.

[2] Ursani A.A., Kpalma, K., and Ronsin, J., "Improved Texture Features Based on Local Fourier Histogram", Wireless Networks, Information Processing and Systems, Volume 20, pp. 19-28, 2008.

[3] Zhou F., Feng J.F., Shi, Q.Y., "Texture Feature Based on Local Fourier Transform", IEEE Transactions on Image Processing, Volume 2, pp. 610-613, 2001.

[4] Arof, H., and Deravi, F., "Circular Neighbourhood and 1-D DFT Features for Texture Classification and Segmentation", IEE Proceedings on Vision, Signal Processing, Volume 145, pp. 167-172, 1998.

[5] Intelmann, S.S., Cutter, G.R., and Beaudoin, J.D., "Automated, Objective Texture Segmentation of Multibeam Echosounder Data-Seafloor Survey and Substrate Maps from James Island to Ozette Lake, Washington Outer Coast", Marine Sanctuaries Conservation Series MSD-07-05, US Department of Commerce, National Oceanic & Atmospheric Administration, MD.31, November, 2007.

[6] Ursani, A.A., Kpalma, K., and Ronsin, J., "Texture Features Based on Local Fourier Histogram: Self-Compensation Against Rotation", Journal of Electronic Imaging, Volume 17, No. 3, pp. 030503-1-030503-3, 2008.

[7] Hastie, T., and Tibshirani, R., "Discriminant Adaptive Nearest Neighbor Classification", IEEE Transactions on Pattern Analysis And Machine Intelligence, Volume 18, No. 6, June, 1996.

[8] Deng, Y., and Yang, S., "An Image Retrieval Approach Based on Annular Color Moments and Energy Moments Features", Proceedings of 2nd International Conference on Future Computer and Communication, Wuhan, May 21-24, 2010.

[9] Kekre, H.B., and Mishra D., "Digital Image Search & Retrieval Using FFT Sectors of Color Images", International Journal on Computer Science and Engineering, Volume 2, No. 2, pp. 368-372, 2010.