INVARIANT CONTENT-BASED IMAGE RETRIEVAL BY WAVELET ENERGY SIGNATURES

Chi-Man Pun

Department of Computer and Information Science
University of Macau, Macau S.A.R.
cmpun@umac.mo

ABSTRACT

An effective rotation and scale invariant log-polar wavelet texture feature for image retrieval was proposed. The feature extraction process involves a log-polar transform followed by an adaptive row shift invariant wavelet packet transform. The log-polar transform converts a given image into a rotation and scale invariant but row-shifted image, which is then passed to the adaptive row shift invariant wavelet packet transform to generate adaptively some subbands of rotation and scale invariant wavelet coefficients with respect to an information cost function. An energy signature is computed for each sub-band of these wavelet coefficients. In order to reduce feature dimensionality, only the most dominant log-polar wavelet energy signatures are selected as feature vector for image retrieval. The whole feature extraction process is quite efficient and involves only $O(n \cdot \log n)$ complexity. Experimental results show that this rotation and scale invariant texture feature is effective and outperforms the traditional wavelet packet signatures.

1. INTRODUCTION

With the advance of information technology, thousands of digitized images or photos are produced everyday, e.g. in World Wide Web. However, most of these collections of images are heterogeneous and poorly indexed. So there is a great demand on some effective content-based image retrieval (CBIR) systems to make use of these huge image archives. Traditionally, the most straightforward way to implement a image database management systems is by means of using the conventional database-management systems (DBMS) such as relational databases or object-oriented databases. The system of these kinds are usually called keyword-based, in which the image attributes (e.g., text, annotations) are extracted manually or partially computed and managed within the framework of a conventional DBMS, such as Chabot [1], Piction [2], Photobook [3], WebSeer [4], etc.. However, the keyword-based approach provides for limited capacity for retrieving visual information. In most of cases, the associated image attributes cannot fully describe the contents of the imagery by themselves. Since the images attributes are annotated manually or semi-automatically. So the process of feature extraction is extremely time-consuming and human-resource intensive. In the current research of CBIR systems [3][5][6][7][8][9], most of them focus on capability of visual search, i.e., images are retrieved by similarity for user provided sample images or sketch. Recently, texture has become popular for CBIR systems. Since most natural surfaces exhibit texture, a successful CBIR system must be able to deal with textured images in real world. For example, color alone cannot distinguish between tigers and cheetahs. Systems like QBIC [6], NETRA [7] and Photobook [3] employed different low-level image features such as color, shape and texture for content-based image retrieval, and demonstrated that texture is one of the important feature. The Photobook project utilizes a 2-D Wold ordering and represents periodic textures using the autocorrelation function and random textures with a Multiresolution Simultaneous Autoregressive (MR-SAR) model, to perform texture annotation. In the NETRA project, Gabor filters have also been used for retrieval purposes. In their work, Gabor is shown to outperform other texture features such as pyramid-structured wavelet, tree-structured wavelet and MR-SAR. In addition, texture feature has also been used in the VisualSeek project by Smith and Chang [5], who extract binary texture features from the Quadrature Mirror Filterbank methods. However, most of the proposed methods assumed the images having the same orientation and/or scale. This assumption is, however, not realistic for most practical applications. The performance of these methods becomes worse when this underlying assumption is no longer valid.

In this paper, we propose an effective rotation and scale invariant log-polar wavelet texture feature for image retrieval. The feature extraction process involves a log-polar transform followed by an adaptive row shift invariant wavelet packet transform. To reduce feature

dimensionality, only the most dominant log-polar wavelet energy signatures are selected as feature vector for image retrieval.

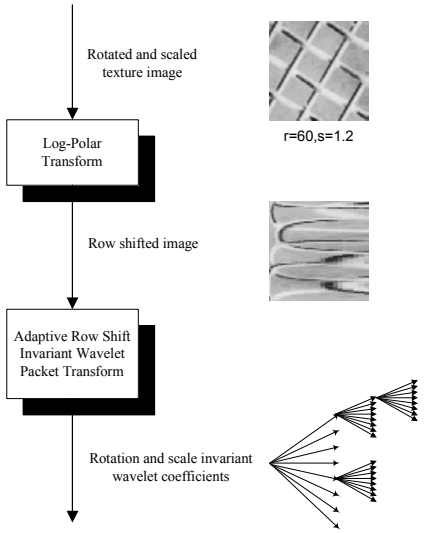

Fig. 1. Schematic diagram of extracting rotation and scale invariant log-polar wavelet feature.

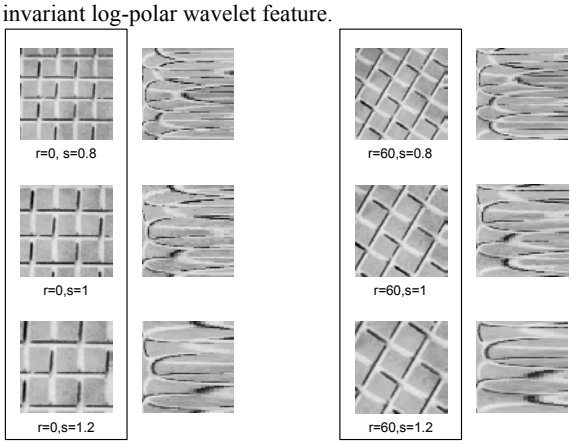

Fig. 2. A sample texture (D1) from the Brodatz album in different rotation angles (r in degrees) and scales (s) and their corresponding log-polar images.

2. ROTATION AND SCALE INVARIANT LOG-POLAR WAVELET TEXTURE FEATURES

It is well known that one of the major drawbacks of DWT is their lack of invariance to the shifting of the input signal, and for two dimensional input signals / images, 2D-DWT are also sensitive to orientation / rotational changes; that is, the same images with different orientations may have different wavelet coefficients. In this section, we propose an approach to extract the rotation and scale invariant log-polar wavelet energy signatures for a given query image, which can be obtained by applying a log-polar transform on the image, followed by adaptive row shift invariant wavelet packet transform (as shown in Fig. 1).

2.1. Log-Polar Transform

As a first step to extract the rotation and scale invariant wavelet signatures, the log-polar transform is used to eliminate the rotation and scale effects in the input image by converting the image into a corresponding log-polar image. Such log-polar image is rotation invariant and nearly scale invariant; however, it is row-shifted. The log-polar transform algorithm is divided into two major steps. In the first step, the radius of the largest circle inside the given $N \times N$ image is used as a scan line to sample S times from 0° to 360° to produce its equivalent $S \times \lfloor N/2 \rfloor$ polar form. Formally, the polar form $p(a,r)$ of the input $N \times N$ image $f(x,y)$ can be computed as follows:

$$p(i,j) = f\left(\left\lfloor \frac{N}{2} \right\rfloor + \left\lfloor j\cos\left(\frac{2\pi i}{S}\right)\right\rfloor, \left\lfloor \frac{N}{2} \right\rfloor - \left\lfloor j\sin\frac{2\pi i}{S}\right\rfloor\right) \quad (2.1)$$

for $i = 0,\cdots,S-1$, and $j = 0,\cdots,\lfloor N/2 \rfloor - 1$.

In the second step, logarithm functions are applied to all radii values in the polar form and their outputs are then quantized into S bins. Hence, a $S \times S$ log-polar image for the given $N \times N$ image is produced. The procedure can be formally defined as follow:

$$lp(i,j) = p\left(i, \left\lfloor \frac{\log_2(j+2)}{\log_2(S+2)} \cdot \left\lfloor \frac{N}{2} \right\rfloor \right\rfloor\right) \quad (2.2)$$

for $i, j = 0,\cdots,S-1$.

As shown in Fig 2, the log-polar images of texture images with different rotation angles and scales seem having only row shifts when compared with the log-polar image of the original texture.

2.2. Adaptive Row Shift Invariant Wavelet Packet Transform

In our adaptive row shift-invariant wavelet packet decomposition, we employ a pair of quadrature mirror filters (*QMF*) to obtain orthonormal representation. In order to achieve row shift invariance, we build a redundant set of wavelet packet coefficients for one additional row circular shift. That is, on each level *p+1*, we compute four periodic images with no shift: $\{C^{p+1}_{8k,(i,j)}\}^{i=I,j=J}_{i,j=0}$, $\{C^{p+1}_{8k+1,(i,j)}\}^{i=I,j=J}_{i,j=0}$, $\{C^{p+1}_{8k+2,(i,j)}\}^{i=I,j=J}_{i,j=0}$, $\{C^{p+1}_{8k+3,(i,j)}\}^{i=I,j=J}_{i,j=0}$ as follows:

$$C^{p+1}_{8k,(i,j)} = \sum_m \sum_n h(m)h(n)C^p_{k,(m+2i,n+2j)} \quad (2.3)$$

$$C^{p+1}_{8k+1,(i,j)} = \sum_m \sum_n h(m)g(n)C^p_{k,(m+2i,n+2j)} \quad (2.4)$$

$$C^{p+1}_{8k+2,(i,j)} = \sum_m \sum_n g(m)h(n)C^p_{k,(m+2i,n+2j)} \quad (2.5)$$

$$C^{p+1}_{8k+3,(i,j)} = \sum_m \sum_n g(m)g(n)C^p_{k,(m+2i,n+2j)} \quad (2.6)$$

where $I = \left\lfloor N/2^{p+1} \right\rfloor - 1$, $J = \left\lfloor M/2^{p+1} \right\rfloor - 1$ and $C^0_{0,(i,j)} = x_{(i,j)}$ is given by the intensity levels of the image x at row i and column j.

Since we just keep one out of two rows, these coefficients appear the same if $C^{p}_{k,(i,j)}$ is circulantly shifted by 0, 2, 4, ..., 2^n rows. In order to have row shift-invariance, we need to compute another four periodic images each with one row shift: $\{C^{p+1}_{8k+4,(i,j)}\}^{i=I,j=J}_{i,j=0}$, $\{C^{p+1}_{8k+5,(i,j)}\}^{i=I,j=J}_{i,j=0}$, $\{C^{p+1}_{8k+6,(i,j)}\}^{i=I,j=J}_{i,j=0}$, $\{C^{p+1}_{8k+7,(i,j)}\}^{i=I,j=J}_{i,j=0}$ as follows:

$$C^{p+1}_{8k+4,(i,j)} = \sum_m \sum_n h(m)h(n)C^{p}_{k,(m+2i+1,n+2j)} \quad (2.7)$$

$$C^{p+1}_{8k+5,(i,j)} = \sum_m \sum_n h(m)g(n)C^{p}_{k,(m+2i+1,n+2j)} \quad (2.8)$$

$$C^{p+1}_{8k+6,(i,j)} = \sum_m \sum_n g(m)h(n)C^{p}_{k,(m+2i+1,n+2j)} \quad (2.9)$$

$$C^{p+1}_{8k+7,(i,j)} = \sum_m \sum_n g(m)g(n)C^{p}_{k,(m+2i+1,n+2j)} \quad (2.10)$$

In a similar manner, these coefficients appear the same if $C^{p}_{k,(i,j)}$ is circulantly shifted by 1, 3, 5, ..., 2^{n+1} rows respectively. At each step, we decompose the image C^{p}_{k} into eight quarter-size images C^{p+1}_{8k}, C^{p+1}_{8k+1}, C^{p+1}_{8k+2}, C^{p+1}_{8k+3}, C^{p+1}_{8k+4}, C^{p+1}_{8k+5}, C^{p+1}_{8k+6} and C^{p+1}_{8k+7}.

In order to have a more effective and concise representation, we need to select the best basis representation for the image. Similar to the approach proposed by Coifman and Wickerhauser [17], we can adaptively select some subbands to decompose further, instead of decomposing every subbands. The basic idea is to compute the information cost $\mathcal{M}$ of each subband, and compare it with that of the sum of all next level subbands. If the information cost of the current subband is less than that of the sum of all next level subbands, then the current subband will not be decomposed; otherwise we decompose the current subband further and do comparison again until maximum level is reached. Hence, the best basis representation can be obtained by an efficient recursive selection process, which determines the best decomposition of the image based exclusively on the local minimization of the information cost function. Let the best basis representation for coarse resolution j (level j) be A^{j}_{k}. Then the best basis A^{0}_{0} for the image x can be computed recursively by:

$$A^{p}_{k} = \begin{cases} C^{p}_{k}, & \text{if } \mathcal{M}(C^{p}_{k}) \le \frac{1}{2}\sum_{i=0}^{7}\mathcal{M}(A^{p+1}_{8k+i}) \\ \bigoplus_{i=0}^{7} A^{p+1}_{8k+i}, & \text{otherwise} \end{cases} \quad (2.11)$$

The recursive computation proceeds down to the specified level J, where

$$A^{J}_{k} = C^{J}_{k},\ 0 \le k < 8^{J} \quad (2.12)$$

For the issue of computational complexity, since we produce eight 2-D periodic images from one level to the next higher level in order to achieve row shift-invariance, we have at most $8^{l} = 2^{l} \cdot 4^{l}$ 2-D periodic images for a decomposition up to level l, which are at most 2^{l} times of those from standard wavelet packet decomposition. However, this decomposition can still be performed efficiently. Stepping from one level to the next higher level we double the number of 2-D periodic images and quarter the size of each of them. By repeating this procedure recursively to all levels, we can get the wavelet packet coefficients for all circular row shifts in $\log N$ steps with only $O(n \cdot \log n)$ complexity (n is the number of pixels in the image).

2.3. Extraction of Wavelet Energy Signatures

With the row shifted log-polar image obtained from the log-polar transform as the input to the adaptive row shift invariant wavelet packet transform, the row shift problem of the log-polar image is properly solved. So the generated wavelet coefficients are rotation and scale invariant now. However, the large number of wavelet coefficients is not suitable for image retrieval. So we reduce the feature dimensionality of the wavelet coefficients by computing energy signature for each sub-band. In this way, the number of energy signatures is equal to the number of sub-bands generated by the adaptive row shift wavelet packet transform. However, the number of energy signatures for image retrieval is not fixed and can be still very large. As suggested by Chang and Kou [11], the most dominant frequency channels provide very useful information for discriminating textures. Therefore, we sort all energy signatures and choose only K most dominant energy signatures (with highest energy values) as feature vector.

3. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our proposed log-polar wavelet feature for texture image retrieval, we use twenty-five natural textures, as shown in Fig. 3, from the Brodatz's texture album [19] as class images. Each texture is scanned with 150 dpi resolution, and each image is of size 512x512 pixels with 256 grey levels. And each texture image is divided into four 256x256 non-overlapping regions. We extract 72 sub-samples of size 128x128 with different orientations (0^o to 180^o with 15^o interval) from each region. So, a dataset of 1200 (25x12x 4) images was prepared for texture image retrieval experiments and each of the 48 images can be treated as a single class. The similarity between the query image q and the nth image in database is defined by the Euclidean distance[20] as follows:

$$d(f^{(q)}, f^{(n)}) = \sum_{i=1}^{M} \left\| f_i^{(q)} - f_i^{(n)} \right\|$$

where $f^{(q)}$ is the feature vector of query image q, $f^{(n)}$ is the feature vector of the nth image in the feature databases.

In Fig. 4, the average precision and recall curves of our proposed method for different number of retrieved images using different number of dominant polar-wavelet energy features are plotted. It can be seen that our proposed method achieved the best retrieval accuracy with only 128

features, outperforming the traditional wavelet packet signature (WPS) image retrieval method. The experiments results also show that the more number of dominant energy signatures do not imply a higher accuracy rate; for instance, as shown in the above mentioned table and figures, retrievals using 128 dominant energy signatures could result in a higher accuracy rate than those using 160 dominant energy signatures.

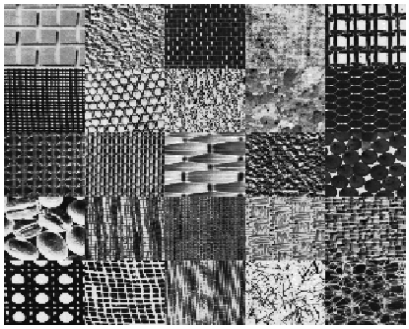

Fig. 3. Twenty-five classes of textures from the Brodatz album. Row 1: D1, D4, D6, D19, D20, Row 2: D21, D22, D24, D28, D34, Row 3: D52, D53, D56, D57, D66, Row 4: D74, D76, D78, D82, D84, Row 5: D102, D103, D105, D110, D111.

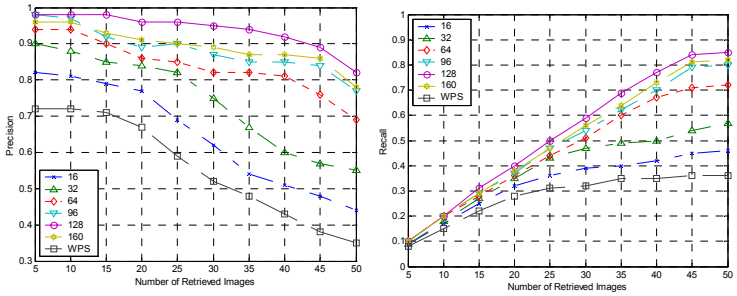

Fig. 4. Average precision and recall versus number of retrieved images for different number dominant energy signatures.

4. CONCLUDING REMARKS

A rotation and scale invariant log-polar wavelet texture feature for image retrieval was proposed. The feature extraction process is quite efficient with only $O(n \cdot \log n)$ complexity. Experimental results show that the proposed method achieve high retrieval accuracy and outperform the traditional wavelet packet signature method. However, the emphasis of this paper has been on rotation and scale invariant texture feature. Our current implementation is based on sequential search. Further improvement could be on investigating methods for efficient multidimensional indexing.

REFERENCES

[1] Ogle V. and Stonebraker M., Chabot, "Retrieval from a relational database images," *IEEE Computer*, September, 1995.

[2] Srihari R. K., "Automatic indexing and content-based retrieval of captioned images," *IEEE Computer*, September, 1995.

[3] Pentland A., Picard R. W. and Sclaroff S., "Photobook: Content-Based Manipulation of Image Databases," *International Journal of Computer Vision*, 18(3), 233-254, 1996.

[4] Swain M. J., Freankel C. and Athitsos V., "WebSeer: An Image Search Engine for the World Wide Web," *Proceeding of CVPR97*, 1997.

[5] Smith J. and Chang S. F., "VisualSEEk: a fully automated content-based image query system," *Proceeding of ACM Multimedia 96*, 87-98, 1996.

[6] Flickner M., Sawhney H., Niblack W., et al., "Query by Image and Video Content: The QBIC System," *IEEE Computer*, September, 1995.

[7] W. Y. Ma, "NETRA: A Toolbox for navigating large image databases," *Ph.D. Dissertation,* Dept. of Electrical and Computer Engineering, University of California at Santa Barbara, June 1997.

[8] Gupta A., "Visual Information Retrieval Technology: A Virage Perspective," *Virage Image Engine API Specification*, 1997.

[9] Belongie S., Carson C., et al., "Color- and Texture-Based Image Segmentation using EM and Its Application to Content-Based Image Retrieval," *Proc. 8th International conference on Computer Vision*, 1998.

[10] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet decomposition", *IEEE Trans. PAMI*, Vol. 11, No. 7, July 1989, pp. 674-693.

[11] T. Chang, and C.C.J. Kuo. "Texture analysis and classification with tree-structured wavelet transform", *IEEE Trans. Image Processing*, vol. 2, p.429-441, April 1993

[12] G. Beylkin, "On the representation of operators in bases of compactly supported wavelets", *SIAM J. Numer. Anal.,* Vol. 6, No. 6, 1992, pp. 1716-1740.

[13] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics, Lecture Notes in Statistics*, Springer-Verlag, 1995, pp. 125-150.

[14] Jie Liang and Thomas W. Parks, "A Translation-Invariant Wavelet Representation Algorithm with Applications", *IEEE Trans. Signal Processing*, vol. 44, Feb.1996

[15] J.C. Pesquet, H. Hrim and H. Carfantan, "Time-Invariant Orthonormal Wavelet Representations", *IEEE Trans. Signal Processing*, vol. 44, Aug 1996.

[16] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant wavelet packet decomposition and representation", *Signal Processing*, Vol. 57, No. 3, Mar. 1997, pp. 251-270.

[17] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection", *IEEE Trans. Inform. Theory*, Vol. 38, No. 2, Mar. 1992, pp. 713-718.

[18] I. Daubechies, Ten Lectures on Wavelets, *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM Press, Philadelphia, Pennsylvania, 1992

[19] P. Brodatz, *Texture: A Photographic Album for Artists and Designers*. Dover, 1966.

[20] R. J. Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches. New York: John Wiley and Sons, 1992.