Affine invariant retrieval and classification of texture images

Jianguo Zhang, Tieniu Tan

Accepted by International Journal of Pattern Recognition

Affine invariant classification and retrieval of texture images☼

*Jianguo Zhang, Tieniu Tan**

National Laboratory of Pattern Recognition (NLPR),

Institute of Automation, Chinese Academy of Sciences, Beijing, 100080, P.R.China

Abstract

In this paper, we proposed a new method of extracting affine invariant texture signatures for content-based affine invariant image retrieval (CBAIR). The algorithm discussed in this paper exploits the spectral signatures of texture images. Based on spectral representation of affine transform, anisotropic scale invariant signatures of orientation spectrum distributions are extracted. Peaks distribution vector (PDV) obtained from signature distributions captures texture properties invariant to affine transform. The PDV is used to measure the similarity between textures. Extensive experimental results are included to demonstrate the performance of the method in texture classification and CBAIR

Keywords: Affine invariant texture analysis; Texture signature; Invariant retrieval

1. Introduction

Content-based image retrieval is a very active research topic in recent years. With the

☼ This work is founded by research grants from the NSFC (Grant No. 69825105 and 69790080) and the Chinese Academy of Sciences.
* Corresponding author: Tel.: +86-10-6264-7441; fax: +86-10-6255-1993 Email address: tnt@nlpr.ia.ac.cn, jgzhang@nlpr.ia.ac.cn;

development of the Internet, the number of images that Internet users could access increases almost exponentially. Accordingly, powerful retrieval tools of Web images are highly desirable.

A great deal of work on image retrieval has been done during the past years. Comprehensive surveys of these methods may be found in [1][14][15][16]. The features commonly used for content-based image retrieval include color, texture, sketch, shape, etc. [1]. Texture features play a very important role in computer vision and pattern recognition, especially in describing the content of images. Images in a digital database are usually subject to geometric distortions due to the change of imaging viewpoint. So the retrieval tools should ideally be invariant to such distortions. Invariant image descriptions deserve more attention [1][2][3], and a new challenge is content-based viewpoint invariant image retrieval. The majority of existing work on invariant texture description has been focused on obtaining translation and rotation invariance [4][9][10][12]. Texture descriptors found in the literature are hardly invariant to affine or perspective transform. Such higher-level invariance is however desirable in practical content-based image retrieval systems. This is a very difficult and challenging problem.

This paper attempts to tackle this hard problem focusing on affine invariant texture feature extraction, and its application in CBAIR. The algorithm discussed in this paper exploits the spectral signatures of texture images. Based on the spectral representation of affine transform, anisotropic scale invariant signatures of orientation spectrum distribution are extracted. Peaks distribution vector (PDV) obtained from signature distributions captures texture properties invariant to affine transform. The PDV is used to measure the similarity between textures. Extensive experimental results are included to demonstrate the performance of the method in texture classification and CBVIIR.

2.Normalization of texture spectrum representation under affine transform

It is well known that affine transform can be decomposed into scale, skew, rotation and translation. It can be viewed as the approximation of the perspective projection if the depth of an object is relatively large with respect to its dimension [5].

Let $f(x,y)$ be the original image and $f_a(x_a,y_a)$ its affine transformed version. Then the relationship of these two images is as follows:

$$f_a(x_a,y_a) = f(x,y);\ \begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \tag{1}$$

Let matrix A denote $\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$ and D $\begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$. D is the translation factor and A is the rotation and scaling factor. It is easy to see that when the original texture undergoes an affine transform, its frequency spectrum takes a similar affine transform.

Let $F(u,v)$ be the Fourier transform of the original texture $f(x,y)$, $F_a(u_a,v_a)$ the Fourier transform of $f_a(x_a,y_a)$. Thus we can obtain the following equations (here we assume that $\det(A^{-1}) \neq 0$)[5]:

$$\left|F_a(u_a,v_a)\right| = \det(A^{-1})\left|F(u,v)\right|, \begin{pmatrix} u_a \\ v_a \end{pmatrix} = A_T^{-1} \begin{pmatrix} u \\ v \end{pmatrix} \tag{2}$$

where A_T^{-1} is the inverse of the transpose of matrix A.

From Equation (2) and (1), we can see that the two expressions are very similar except that the spectrum in Equation (2) is scaled by det(A^{-1})and the translation factor D is removed. So that if Equation (2) is normalized by the sum of $\left|F(u,v)\right|$, the following equation can be derived:

$$\frac{\left|F_a(u_a,v_a)\right|}{\sum\left|F_a(u_a,v_a)\right|} = \frac{\left|F(u,v)\right|}{\sum\left|F(u,v)\right|} \tag{3}$$

Let $F_a^{'}(u_a, v_a)$ denote the left part of Equation (3) and $F^{'}(u,v)$ the right part. By rewriting Equation (3), we get

$$\left|F^{'}{}_a(u_a, v_a)\right| = \left|F^{'}(u,v)\right|, \begin{pmatrix} u_a \\ v_a \end{pmatrix} = A_T^{-1} \begin{pmatrix} u \\ v \end{pmatrix} \tag{4}$$

Thus the normalized spectrum also satisfies the affine model. The relationship between textures and their spectral representations under affine transform is illustrated in Figure 1.

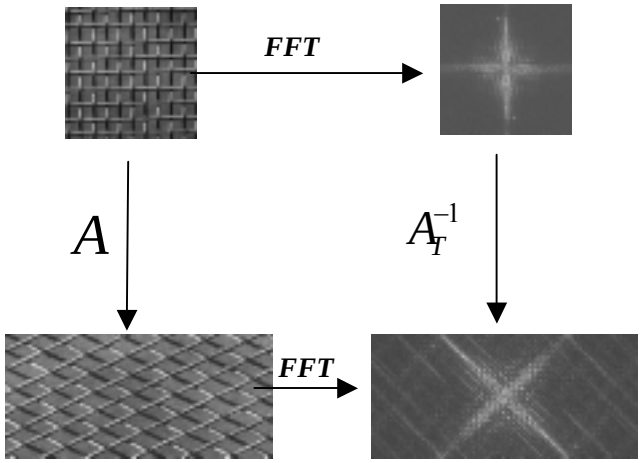

Figure 1 Spectral representation of textures under affine transform

3. Affine invariant texture signatures

In Section 2, we have shown that affine transform in the spatial domain results in similar affine transform in the spectrum. We have also derived a spectral representation (Equation (4)) that mathematically satisfies affine model. This indicates that we can derive texture properties from its spectrum that is invariant to affine transforms (notice that spectrum contains most of the useful texture information of the original image). This section discusses whether and how we can use the new representation for affine invariant texture analysis.

In order to simplify notations, here we simply use $F_a(u_a, v_a)$ ($F(u,v)$) to represent

$\left|F'_a(u_a, v_a)\right|$ ($|F'(u,v)|$), that is

$$F_a(u_a, v_a) = F(u,v);\ \begin{pmatrix} u_a \\ v_a \end{pmatrix} = \hat{A}\begin{pmatrix} u \\ v \end{pmatrix} \tag{5}$$

where $\hat{A}$ denotes the matrix $A_T^{-1} = \begin{bmatrix} \hat{c}_{11} & \hat{c}_{12} \\ \hat{c}_{21} & \hat{c}_{22} \end{bmatrix}$.

Let

$$\rho_a = \sqrt{u^2{}_a + v^2{}_a}\ , \theta_a = arctg(u_a / v_a)$$

$$\rho = \sqrt{u^2 + v^2}\ ,\ \theta = arctg(u / v)$$

Equation (5) can then be expressed in its polar version

$$F_a(\rho_a, \theta_a) = F(\rho, \theta) \tag{6}$$

where $F_a(\rho_a, \theta_a)$ and $F(\rho, \theta)$ are called the line-spread functions [17] along the lines at angle θ_a and θ. We have investigated the changes between ρ and ρ_a in our previous work [13] by using their ratio s ($s = \frac{\rho_a}{\rho}$). Our study indicates that s is only dependent on θ for a given affine transform. Although s varies with θ, s is a fixed value for a given orientation θ. By applying log operation on ρ and ρ_a along the orientation θ and θ_a, we have

$$F_a(w_1, \theta_a) = F(w_2, \theta) \tag{7}$$

where

$$w_1 = \log \rho_a = \log s + \log \rho$$

$$w_2 = \log \rho$$

It is important to note that in Equation (7) scaling in the line-spread spectrum function has been transformed into a shift by the log operation. Since the frequency distribution (here we consider the spectrum magnitude as the probability of the corresponding frequency) can give a description of texture periodicity, we calculate the central moment of the line spread function as

follows:

$$c_a(\theta_a) = \int (w_1 - \overline{w_1}) F_a(w_1, \theta_a) dw_1$$

$$c(\theta) = \int (w_2 - \overline{w_2}) F(w_2, \theta) dw_2 \tag{8}$$

where $\overline{w_1}$ and $\overline{w_2}$ are the mean value of w_1 and w_2. Since the shift factor has been removed in Equation (8), it is obviously that $c_a(\theta_a)$ equals to $c(\theta)$. Notice that the power spectrum also plays a very important role in measuring texture properties. We incorporate power spectrum that is invariant to scaling and compute the spectrum signatures at angle θ and θ_a as follows:

$$\begin{aligned} T(\theta_a) &= c_a(\theta_a) \int F_a(w_1, \theta_a) dw_1 \\ &= c(\theta) \int F(w_2, \theta) dw_2 = T(\theta) \end{aligned} \tag{9}$$

We have now obtained the orientation spectrum signature $T(\theta)$ which is invariant to anisotropic scale (notice that affine transform results in different scaling along different orientations. It is often overlooked in most of the literature on invariant texture analysis). This is the key problem of affine invariant texture analysis.

When we investigate the spectrum signatures along all the orientations throughout the whole

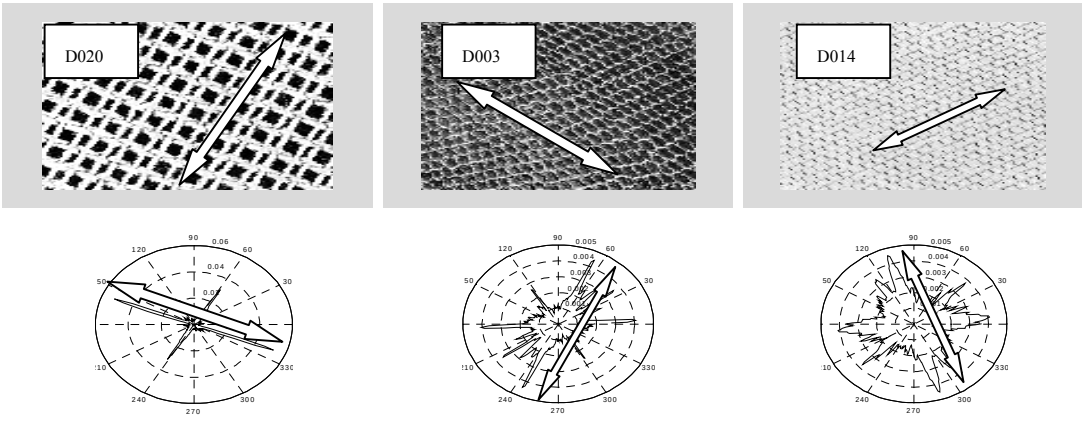

Figure 2 Examples of spectrum signatures of different textures from the Brodatz album. Arrows show the main texture direction indicated by texture signatures

image, $T(\theta)$ becomes the distribution function of spectrum signatures at different angles (Here

we call this function as signature distribution function). The major peaks of $T(\theta)$ indicate those specific, dominating directions within the spectrum domain, which in turn reflect strong texture directionality as illustrated in Figure 2.

Figure 3 gives an example of the effect of affine transform on texture signatures. From this figure, we can see that although the signature distribution function changes under affine transform, the order of the major peaks, the number of the peaks, and the magnitude of the peaks tend to be quite stable. That is the PDV almost remains the same. Similar observation can also be found in [6]. The PDV is constructed based on those peaks. A peak p_i of $T(\theta)$ is defined as follows:

$$p_i = T(\theta_i) \tag{10}$$

when $T(\theta_i)$ satisfies the following conditions

1. $T(\theta_i) = \max_{m \in ws}(T(\theta_m)), m \in \{i - \frac{ws}{2}, i + \frac{ws}{2}\}$; (11)
2. $\int_m (T(\theta_m) - \frac{1}{ws}\int_m T(\theta_m)d\theta)^2 d\theta \geq t$ (12)

where t is a given threshold to select the peaks and ws is the size of the local window of θ_m. Since the spacing between peaks can be small, the size of the estimation window cannot be very large. Here we use the local window size of 1x7. Condition 1(Equation 11) implies that the value of the central point of the peak support window must be a local maximum. Condition 2 (Equation 12) suggests that a local maximum qualifies for a peak only when the deviation of its adjacent area is no less than a given threshold. In fact it gives a measurement of smoothness of the peak profile. Note that the noise of the signature must be first removed since it maybe affects the peaks detection. Here Gaussian filter is first applied to denoise the signature. The process of the PDV extraction is illustrated in Figure 4.

The affine invariant feature vector is constructed by computing statistics of PDV. The statistics we used here are the largest value, the average value, the standard deviation, and the peak density.

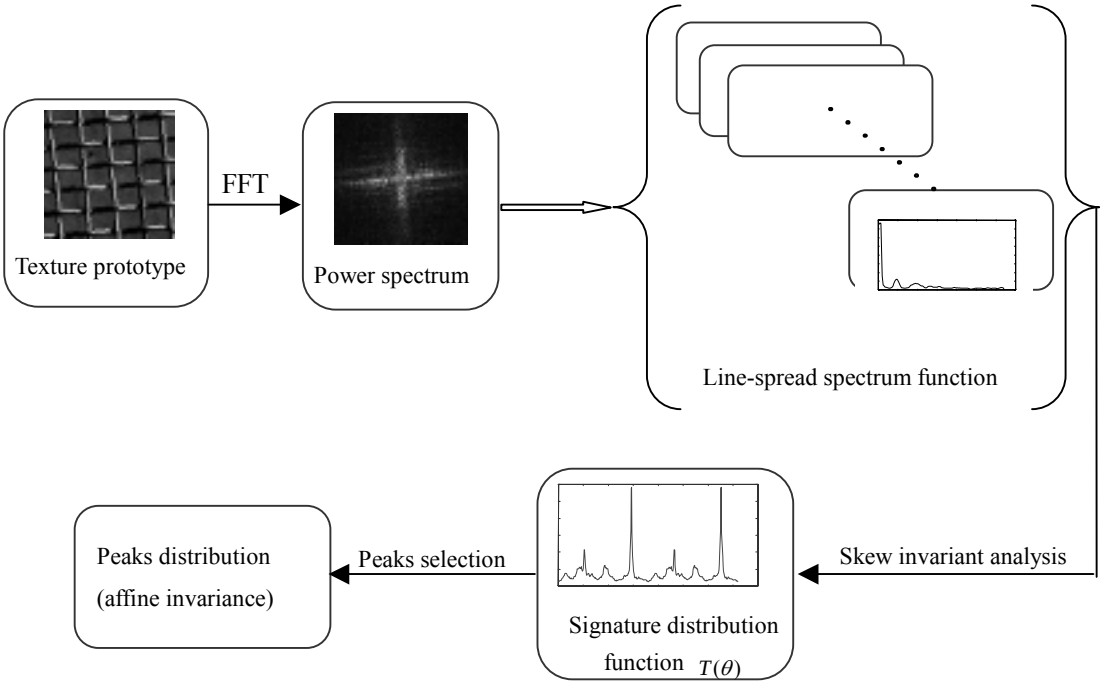

Figure 4 Schematic diagram of PDV extraction

The peak density is calculated as the ratio between the number of peaks and the length of θ.

The following outlines the major steps of computing the affine invairiant texture features:

1. Compute the Fourier transform of an original texture image, and then normalize its magnitude by its 1-norm as described in Equation 3.
2. Extract the line-spread function along each orientation. This step needs to sample the magnitude domain along its orientation axis. Usually the sample interval is 1^o
3. Apply the log operation on every line-spread function and compute $T(\theta)$ according to Equation 9. Thus the signature distribution is obtained.

4. Detect the peaks according to the criteria described in Equation 10 and compute the the four statistics as described above. Thus the affine invarinat textue features are derived.

4 Experimental results

In our experiments, we demonstrate the capability of the proposed features in affine invariant texture classification in Section 4.2. Then the efficacy of this method is tested in the context of affine invariant image retrieval in Section 4.3

4.1 Image database

A test database is constructed for our experiment. The database consists of 20 structural

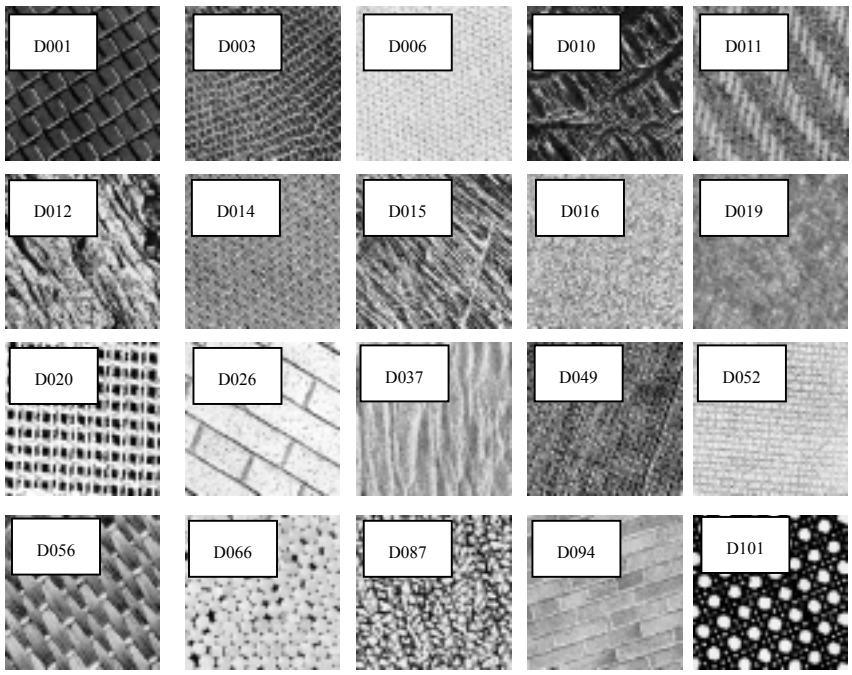

Figure 5 20 textures from Brodatz database used for affine invariant texture classification

texture images from the Brodatz texture album as shown in Figure 5 (if textures are isotropic or random, any texture descriptors will be invariant to affine transform [11]). Each texture image of size 512x512 is randomly affine transformed into 50 versions, from which subimages of size 128x128 are extracted. Thus a database of 1000 (50 for each texture class) images is constructed

for this experiment.

4.2 Affine invariant texture classification

The performance of the proposed features is first tested in affine invariant texture classification. 140 images from the above database (7 for each texture class) are used for training and the remaining 860 images (43 for each texture) for testing. The PDV is extracted from each image and the feature vector is constructed by computing statistics of PDV as discussed in Section 3. Euclidean distance is calculated to measure the similarity between textures. Suppose that f_i represents the i[th] component of the feature vector, $i = 1,2,3,4$. The distance D between textures is calculated as follows:

$$D = \sqrt{\sum_{i=1}^{4} (f_i - f_i^{'})^2}$$

The K-nearest neighborhood classifier is employed. An average correct recognition rate of 87.91% is obtained. The classification confusion matrix in Table 1 shows the classification results of each texture. It is interesting to notice that major misclassifications occur between D016 and D019. The two textures are indeed the most similar ones in that both are more random than others.

4.3 Image Retrieval

In this section, we further test the proposed texture features in the context of affine invariant image retrieval. It should be pointed out that although much has been done on content-based image retrieval [1], little attention has been paid to affine invariant image retrieval despite of its

scientific and practical value. We do this in a designed image retrieval system. A query texture prototype is presented to the query system. 50 texture images are returned as the most similar images to the prototype per search. The retrieval database contains 1000 images described above.

In this experiment, the performance criterion is quantitative. The perfect retrieval results for a query texture should be an image subset whose elements are of the same class as the query image. The criterion lies on the fact that a texture image and its affine transformed versions are always perceived as the same textures by human observers. That is they are viewpoint independently and perceptually similar.

The retrieval system is evaluated by the precision and recall measurement [1]. Suppose the

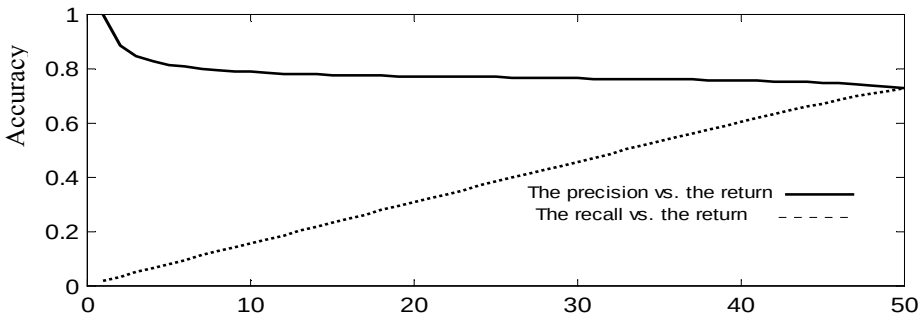

Figure 6 Performance of affine invariant texture retrieval

image database is divided into two sets: the set of images $R(q)$ predetermined to be in the same class as the query q, and the set of irrelevant images. Assume that the query image q is presented to the retrieval system and the system returns a set of images $A(q)$ as the answer. The precision of the answer is the fraction of the returned images that are relevant to the query:

$$p = \frac{A(q) \cap R(q)}{A(q)}$$

While the recall is the fraction of relevant images that are returned by the query:

$$r = \frac{A(q) \cap R(q)}{R(q)}$$

We explore the performance of our retrieval system by computing p and r via

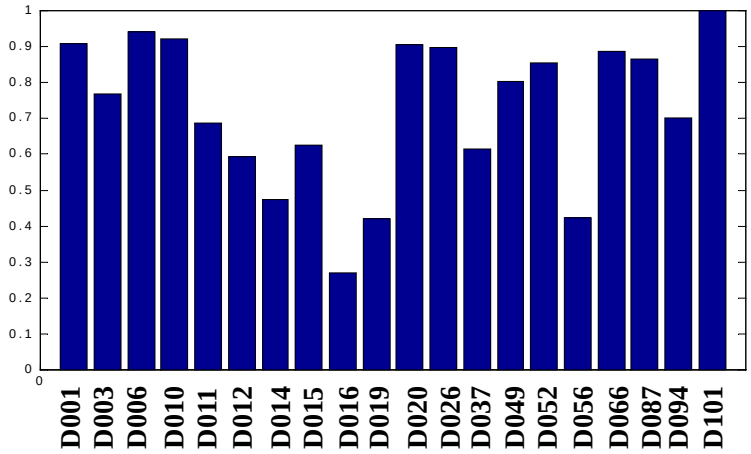

Figure 7 Retrieval results of each texture class as 50 images returned per search

changing the number of the returned images per search. The evaluation results are illustrated in Figure 5. An average of 38 out of 50 images returned is of the same texture class. The retrieval performance is shown in Figure 6 as the number of returned images increases. We can see that the precision is robust to the returned images. The recall rate increases fast versus the number of the returned images (notice that the ideal recall should reach 100% when the number of returned images increases to 50). Figure 6 illustrates that our method gives a very encouraging precision while keeping good recall rate. This clearly shows the efficacy of the proposed method for affine invariant texture based retrieval. The retrieval result of each texture class is illustrated in Figure 7. Experimental results in Figure 7 and Table 1 show that a texture that gives good classification also gives good retrieval results.

5 Discussion and conclusions

It has been reported that Gabor features and Wold features can characterize both ordered and random textures [7][11]. To the best of our knowledge [4][9], these texture models have not been developed for affine invariant texture classification. The spectrum signatures described in this paper are comparable to these features in measuring texture periodicity. When Gabor masks are tuned to the orientations indicated by larger spectrum signatures, the Gabor channels often produce larger output. For Wold features, the harmonic peaks that can measure texture structural properties are often centered on those regions around the orientations indicated by the peaks in $T(\theta)$. This suggests that our features can capture texture properties as many other texture models do. Furthermore they are invariant to affine transforms.

In conclusion, we have presented a novel algorithm of extracting affine invariant texture features for CBAIR. The new proposed signatures are computed from the texture spectrum domain. The PDV extracted from the signature carries texture properties invariant to affine transform. Our experimental study has clearly shown the efficacy of the proposed features in both invariant texture classification and CBAIR. This is a good start in affine invariant texture analysis. This approach may be extended to perspective invariant texture classification. Future work will investigate affine invariant texture segmentation.

References

[1] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Gain, Content-based image retrieval at the end of the early years, IEEE Transaction on Pattern Analysis and

Machine Intelligence, Vol. 22, No. 12, December 2000.

[2] S.R. Fountain, T.N. Tan, Efficient rotation invariant texture features for content-based image retrieval, Pattern Recognition, Vol. 31,No. 11 1998, pp1725-1732.

[3] S.R Fountain, T.N. Tan, RAIDER: rotation invariant retrieval and annotation of image database, Proc. of BMVC, Vol.2. pp. 390-399, 1997.

[4] T.N. Tan, Geometric transform invariant texture analysis, SPIE. Vol. 2488, pp475-485, 1995.

[5] Jezekiel Ben-Arie, Zhiqian Wang, Pictorial recognition of objects employing affine invariance in the frequency domain, IEEE Trans. on PAMI, Vol. 20, No.6, June 1998.

[6] D. Chetnerikov, Pattern regularity as a visual key, Image and Vision Computing 18 pp. 975-985, 2000.

[7] Fang Liu and Rosalind W. Picard, Periodicity, directionality, and randomness: Wold features for image modeling and retrieval, IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 18, No. 7, July 1996.

[8] J. G. Daugman and C. J. Downing, Demodulation, predictive coding, and spatial vision, Opt. Soc. Am. A, Vol. 12,No. 4, pp. 641-660, April 1995.

[9] J.G. Zhang and T.N. Tan, Brief review of invariant texture analysis methods, Pattern Recognition, 2001 (in press).

[10] S. Chang et al., Texture discrimination by projective invariants, Pattern Recognition Letters 5 pp. 337-342, 1987.

[11] T.N.Tan, Rotation invariant texture features and their use in automatic script identification, IEEE Trans. Pattern Analysis and Machine Intelligence Vol.20, No.7, pp.

751-756, July 1998.

[12] Coloma Ballester and Manuel González, Affine invariant texture segmentation and shape from texture by variational methods, Journal of Mathematical Imaging and Vision 9, 141-171, 1998.

[13] J.G. Zhang, Tieniu Tan, Affine invariant texture signatures, Proc. of IEEE International Conference on Image Processing (ICIP 2001), October 2001.

[14] Yong Rui and Thomas S. Huang, Image retrieval: current techniques, promising directions, and open issues, Journal of Visual Communication and Image Representation, 10, 39-62 (1999).

[15] V. Alp Aslandogan and Clement T. Yu, Techniques and systems for image and video retrieval, IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No 1, 56-63, January 1999.

[16] J.M. Zachary, Jr., S. Iyengar, Content based image retrieval systems, Application-Specific Systems and Software Engineering and Technology,, 1999. ASSET '99, pp. 136-143.

[17] Kenneth R. Castleman, Digital image processing, pp.171-206, Prentice Hall,Inc, 1996.

About the author—Jianguo Zhang received his B.Sc.(1996) and M.Sc.(1998) in automation from Shandong University of Technology. He is currently a Ph.D. candidate in the National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences, Beijing China. His research interests include invariant perception analysis, and pattern recognition.

About the author—Tieniu Tan received his B.Sc. (1984) in electronic engineering from Xi'an Jiaotong University, China, and M.Sc. (1986), DIC (1986) and Ph.D. (1989) in electronic engineering from Imperial College of Science, Technology and Medicine, London, England. In October 1989, he joined the Computational Vision Group at the Department of Computer Science, The University of Reading, England, where he worked as Research Fellow, Senior Research Fellow and Lecturer. In January 1998, he returned to China to join the National Laboratory of Pattern Recognition, the Institute of Automation of the Chinese Academy of Sciences, Beijing, China. He is currently Professor and Director of the National Laboratory of Pattern Recognition and Assistant Director of the Institute of Automation. Dr. Tan has published widely on image processing, computer vision and pattern recognition. He is a Senior Member of the IEEE and was an elected member of the Executive Committee of the British Machine Vision Association and Society for Pattern Recognition (1996-1997). He serves as referee for many major national and international journals and conferences. He is an Associate Editor of the International Journal of Pattern Recognition, the Asia Editor of the International Journal of Image and Vision Computing and is a founding co-chair of the IEEE International Workshop on Visual Surveillance. His current research interests include speech and image processing, machine and computer vision, pattern recognition, multimedia, and robotics.

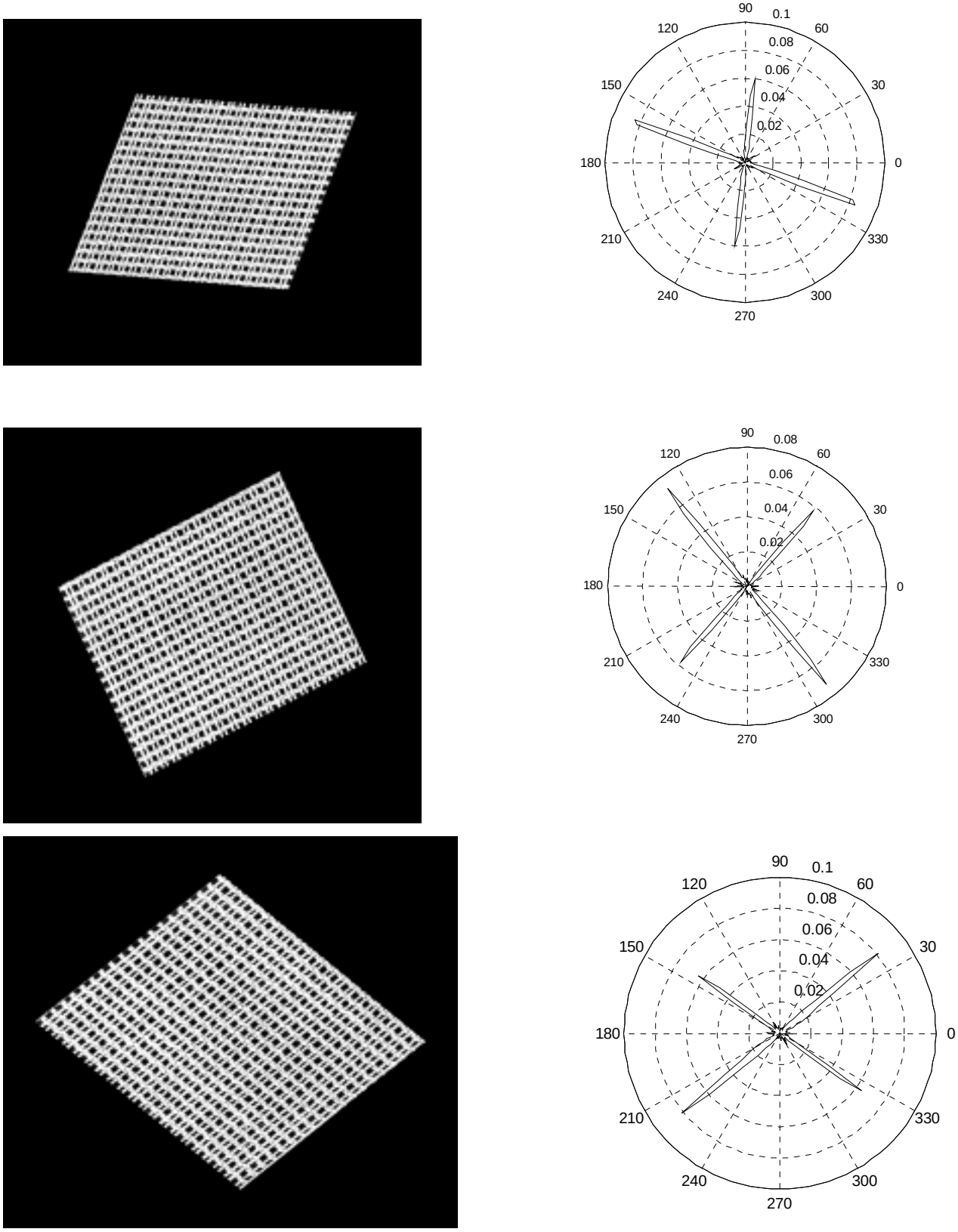

Figure 3, Affine transformed versions of texture D001 and their corresponding signatures

Table 1 Confusion matrix of affine invariant texture classification

Reference texture	No. of matches with																			
	D001	D003	D006	D010	D011	D012	D014	D015	D016	D019	D020	D026	D037	D049	D052	D056	D066	D087	D094	D101
D001	41			2																
D003		38												4						1
D006			43																	
D010	2			41																
D011					42											1				
D012						35				4									4	
D014							37		5				1							
D015				2				40								1				
D016					1		4		18	14						6				
D019									17	26										
D020											42								1	
D026			3									40								
D037							1		1				39			2				
D049		4												39						
D052															40		3			
D056					8			2					1			32				
D066															2		41			
D087							2						1					40		
D094											4								39	
D101																				43
Average accuracy=87.91%																				