PERCEPTUAL TEXTURE SPACE FOR CONTENT-BASED IMAGE RETRIEVAL

HUI ZHONG LONG, WEE KHENG LEOW
School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
longhuiz, leowwk@comp.nus.edu.sg

FOOK KEE CHUA
Dept. of Social Work and Psychology, National University of Singapore
Block AS 6, Level 4, 11 Law Link, Singapore 117570
swkcfk@nus.edu.sg

Texture is an important visual feature for content-based image retrieval. An ideal content-based retrieval system should compare images in its database with the query in a manner that is consistent with human's perception of visual similarity. In practice, however, texture similarity computed using computational texture features is not necessarily consistent with human's perception. Although several perceptual texture spaces that capture human's perception of texture exist, the mapping from computational features to these perceptual spaces has not been derived. Another difficulty with using these perceptual spaces is that their measurements of texture similarity are confounded by texture scale and orientation because the textures used to construct the spaces are not normalized to canonical scales and orientations. This paper presents a perceptual texture space constructed using textures with canonical scales, orientations, intensity, and contrast. To illustrates the application of this perceptual space, computational texture features are extracted using Gabor filters and a multi-layer feedforward neural network is trained to map the Gabor features to the perceptual space. Test results show that the neural network can perform the mapping quite accurately. Moreover, retrieval tests show that retrieval performance is improved by mapping computational features to perceptual texture space.

1 Introduction

An ideal content-based image retrieval system should compare images with the user-specified query in a manner that is consistent with human's perception. For example, color similarity is typically measured in the CIE L*u*v* space which is quite consistent with human's perception. Unfortunately, texture similarity computed using existing computational features is not necessarily consistent with human's perception. Although several perceptual texture spaces that capture human's perception of texture exist, the mapping from computational features to these perceptual spaces has not been derived.

Another difficulty with using these perceptual spaces is that their measurements of texture similarity are confounded by texture scale and orientation. The textures used to construct the spaces are not normalized to canonical scales and orientations. However, in retrieving images of natural scenes, a user typically requires images that contain specific textures of any scale and orientation. So, the measurement of texture similarity should be decoupled from the measurement of similarity in scale and orientation.

This paper presents the construction of a perceptual texture space using textures with canonical scale, orientation, intensity, and contrast (Section 3). To illustrate the application of this perceptual space in image retrieval, computational texture features are extracted using Gabor filters and a multi-layer feedforward neural network is trained to map the Gabor features to the perceptual space (Section 4). Tests were performed to assess the neural network's mapping accuracy. In addition, retrieval tests were conducted to compare the retrieval performance of the perceptually consistent method with existing methods (Section 5).

2 Related Work

2.1 Perceptual Texture Properties

The earliest study of human's perception of texture similarity was conducted by Tamura et al.[13] In their experiments, 48 human subjects were asked to judge the similarity of two textures according to six visual features, namely coarseness, contrast, directionality, line-likeness, regularity, and roughness. Due to the combinatorial nature of the task, only 16 textures were used. In the experiment of Benke et al.,[2] 8 subjects were asked to rank 14 texture images according to each of three visual features similar to those used by Tamura et al. Amadasun and King conducted similar ranking experiment with five visual properties including coarseness, contrast, busyness, complexity, and texture strength.[1]

The major difficulty with these studies is that the subjects were asked to judge texture similarity according to subjective visual properties. However, the subjects' interpretations of the meaning of these visual properties are expected to vary from one person to the next. Therefore, it is uncertain whether the individual ranking results can be combined into group ranking results that represent the perception of a typical person. The second difficulty is that the ranking results of the individual visual properties cannot be combined to form a texture space. The ranking results were measured according to individual visual properties. But, the relative scale between two visual properties is unknown. For example, one unit difference in coarseness may not be perceptually equal to one unit difference in regularity. So, the differences in the various visual properties cannot be combined into a single measurement of the overall perceptual texture similarity.

To avoid these difficulties, Rao and Lohse[11] performed an experiment in which 20 subjects were asked to sort 56 textures into as many groups as the subjects wished such that the textures in each group were perceptually similar. The textures were sorted based on the subjects' perception of overall texture similarity without using subjective visual properties. A co-occurrence matrix of the data was computed and multidimensional scaling (MDS)[5] was performed to map the textures into a 3-dimensional space. Heaps and Handel[6] conducted further studies using the same methodology. They concluded that texture similarity perceived with attentive judgment depends on the contexts in which the textures are considered. That is, how similar two textures appears to humans depends not only on the two textures being judged, but also on the whole set of textures with which pairwise judgments are made. In our experiments, the influence of the context problem is reduced by (1) selecting all the texture patterns that are expected to be present in images of natural scenes, and (2) normalizing the intensity, contrast, scale, and orientation of the textures.

2.2 Computational Texture Features

Computational models of texture features can be broadly grouped into three categories, namely structural, statistical and spectral. Structural models characterize textures according to the spatial relationships between image pixels. Statistical models categorize textures according to statistical measurement of visual qualities such as coarseness, granularity, regularity, etc. Spectral models characterize textures based on the properties of their Fourier transforms or the results of filtering the textures by appropriate filters.

The statistical and spectral models are commonly used in existing texture-based image retrieval systems. In particular, the statistical features proposed by Tamura et al.[13] are used in IBM's QBIC system. The Wold model,[9] which is based on Fourier transform (spectral model) and Markov Random Field (statistical model) is used in MIT's PhotoBook. Gabor features (spectral model) are used in NeTra[10] and Leow and Lai's invariant texture matching method.[7,8]

Although the above features are easy to compute, the texture similarity computed based on these features do not necessarily correspond to human's perception. On the other hand, Tamura et al.,[13] Benke et al.,[2] and Amadasun and King[1] have attempted to derive statistical features that match the human ranking data obtained from the experiments. Unfortunately, their perceptual data do not constitute a texture space in which the overall perceptual similarity of two textures can be measured. In contrast, this paper demonstrates how a neural network can be trained to directly map Gabor texture features into the perceptual texture space that we construct.

3 Construction of Perceptual Texture Space

A rich set of textures are collected to construct the perceptual space. Fifty texture images are selected from the Brodatz album.[3] The remaining images in the album are omitted for various reasons. For example, some textures are scaled versions of each other. In this case, only one of them is needed because the texture images are normalized before they are used in the psychological experiments. Another reason is that the contents of some images (e.g., flowers in lace) are perceptually stronger than their texture patterns. It is very difficult for the subjects to avoid classifying the images based on their contents. In addition to these Brodatz textures, ten images of natural textures such as foliage and roof tiles are included because they are not collected in the Brodatz album. These 60 texture images should constitute a good set of representatives of the textures that appear in images of natural scenes. These images are first normalized before they are used in the psychological experiments.

3.1 Texture Image Normalization

Four texture features, namely intensity, contrast, scale, and orientation, that can confound the subjects judgment are normalized. The intensity μ and contrast σ of a texture image is measured as the mean and standard deviation of the intensity of the pixels in the image. The texture's intensity and contrast are normalized to a fixed canonical intensity μ' and contrast σ' by changing a pixel's intensity p_{ij} to p'_{ij} by a linear transformation:

$$p'_{ij} = \mu' + \frac{\sigma'}{\sigma}(p_{ij} - \mu) \ . \tag{1}$$

Our analysis of the textures reveal that there is no single scaling method that fits all the textures. However, by categorizing the textures into four different groups based on their 2D FFT patterns, each group of textures can be scaled in the same way. The four groups consists of structured, directional, granular, and random textures [8] (Fig. 1). Structured textures have very well defined FFT patterns. Their FFT patterns contain a small number of highly

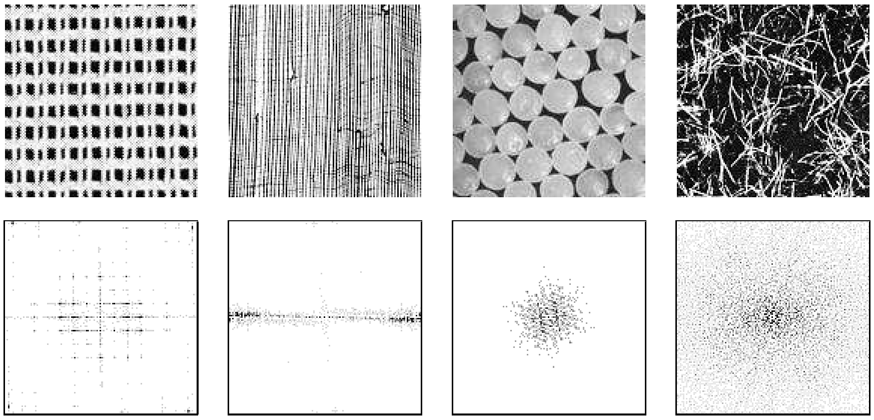


Figure 1. Four types of FFT patterns (bottom) and the corresponding representative textures (top): (a) structured, (b) directional, (c) granular, (d) random.

localized, prominent frequency components with large amplitudes (Fig. 1a). The textures in this group are scaled so that the 8 most prominent components have the same frequencies in the FFTs of the various textures.

The FFT patterns of oriented textures have many frequency components and well defined orientations (Fig. 1b). They lack prominent frequency components which can be used for scale normalization. However, their orientation can be normalized so that their FFT patterns have the same orientations.

Granular textures have well defined frequencies and many orientations. Their FFT patterns appear in the form of small disks (Fig. 1c). The scales of these textures are normalized so that their FFT patterns have approximately the same size. Random textures have many frequencies and orientations (Fig. 1d). They lack well defined scales and orientations for normalization. So, they are not subjected to scale and orientation normalization.

3.2 Measurement of Human Perception

The procedure of the psychological experiment is similar to that of Rao and Lohse.[11] Each of the 60 normalized texture images were printed on a photo-quality paper of 4×5 inches. Twenty subjects were asked to freely sort the texture images according to texture similarity into as many groups as they wished such that the textures in each group were perceptually similar. The sorting results were used to construct a distance matrix based on Donderi's information measurement method.[4] This distance measure is preferred to the more commonly used co-occurrence matrix because it correlates well with direct estimation of human's pairwise similarity judgment.[4]

Donderi's information measurement is based on Shanon's information distance measure $d(IK)$ between two stimuli I and K:

$$d(IK) = H_K(I) + H_I(K) \tag{2}$$

where $H_I(K)$ is the *independent information* of K relative to I, and similarly for $H_K(I)$. Independent information is the information gained by classifying two stimuli into separate groups. Donderi modified Eq. 2 to

$$d(IK) = \sqrt{H_K(I) + H_I(K)} \tag{3}$$

so that, according to the distance measurement of Eq. 3, informationally independent stimuli are dissimilar and informationally redundant stimuli are similar.[4]

Given the subjects' sorting results, the distance matrix is computed as follows. Let $n(K)$ denote the number of textures of the group to which texture K belongs, and N denote the total number of textures used in the experiment. Then, the independent information $H_I(K)$ (and similarly for $H_K(I)$) of each subject's classification is computed as:[4]

$$H_I(K) = \begin{cases} 0 & \text{if } I \text{ and } K \text{ belong to the same group} \\ \log_2 n(K) - \log_2 N & \text{otherwise.} \end{cases} \tag{4}$$

The overall $H_I(K)$ is computed as the average of the individual $H_I(K)$ of all the subjects. Finally, the overall $H_I(K)$ and $H_K(I)$ for all the textures are substituted into Eq. 3 to compute the distance measure $d(IK)$ which constitute the components of the distance matrix.

After obtaining the distance matrix, multidimensional scaling (MDS)[5] is applied to the distance matrix to compute the coordinates on the textures in a multi-dimensional space. Given a particular number of dimensions, MDS will compute the coordinates so that the Euclidean distances between the textures in the space best fit the measurements in the distance matrix. The appropriate number of dimensions is determined by the *stress* factor which is defined as the proportion of the variance of the data that is not accounted for.[5] The larger the number of dimensions, the better is the fit and the lower is the stress. However, a space with too many dimensions is difficult to visualize and to use. So, a trade-off is made between the number of dimensions and the goodness of fit. Rao and Lohse concluded that a texture space with three

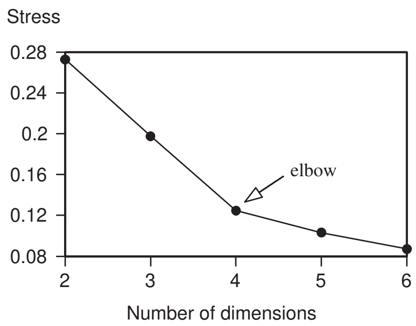


Figure 2. Stress plot of the MDS fitting results. The most appropriate number of dimensions, in this case 4, is given be the so-called *elbow point*.

dimensions is appropriate.[11] On the other hand, our MDS fitting results show that a four dimensional texture space provides a better trade-off (Fig. 2).

A perspective view of the perceptual texture space obtained is shown in Fig. 3. Since it is impossible to visualize 4D space, a 3D version of the texture space is displayed. This 3D version has a higher stress than the 4D space but their structures are similar. It can be observed that there is an empty region at the origin of the texture space that has no texture. This characteristic is shared by the texture space of Rao and Lohse. One probable reason is that the region near the origin corresponds to textures that appear equally different from other textures used in the experiment, and that such textures do not exist in the real world.

4 Texture Space Mapping

Gabor texture features have been proved to be favorable compared with other texture features for image retrieval.[8] They are obtained by convolving an image $I(x, y)$ with a set of Gabor filters G with different spatial frequency f and orientation θ:

$$H_{c,f\theta}(x, y) = I(x, y) * G_{c,f\theta}(x, y) \tag{5}$$

$$H_{s,f\theta}(x, y) = I(x, y) * G_{s,f\theta}(x, y) \tag{6}$$

where $G_{c,f\theta}$ and $G_{s,f\theta}$ are the real and the imaginary components of the Gabor function. The Gabor filters' output energy is given by:

$$E_{f\theta}(x, y) = H^2_{c,f\theta}(x, y) + H^2_{s,f\theta}(x, y) \ . \tag{7}$$

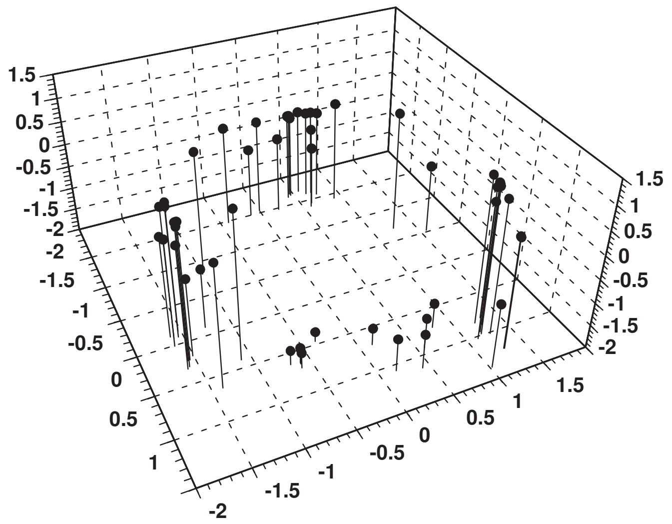


Figure 3. 3D version of the perceptual texture space. Each dot represents the location of a texture in the space. The vertical lines indicate the coordinates of the textures in the horizontal plane.

After Gabor filtering, the filters' outputs $E_{f\theta}$ are smoothed by Gaussian filters to remove local variations introduced by the sinusoidal terms in the Gabor functions. In the current implementation, 48 filters comprising 6 frequencies and 8 orientations are used.

In general, the mapping from an arbitrary computational space to a perceptual space can be very complex and difficult to derive. However, in the case of mapping Gabor features to the perceptual space, we found that a multilayer feedforward neural network can be trained to perform the mapping with good accuracy.

The neural network had 48 inputs for the Gabor features and 4 outputs for each dimension of the perceptual texture space. By experimenting with various network configurations, it was determined that a network with two hidden layers consistently performed most accurately. The network had 25 units in the first hidden layer and 20 units in the second hidden layer. The Resilient Propagation algorithm, which is a variant of back-propagation, of the SNNS neural network software was used to train the network. It has the advantage of using weight decay method to remove insignificant connection weights in the network. Such a weight pruning operation has been found to

Table 1. Mean squared errors (MSE) of the two experiments.

Experiment	Training MSE	Testing MSE
1. unknown instance	0.02108	0.01680
2. unknown texture	0.02211	0.13143

improve generalization accuracy in practice.

The data set consists of the 60 normalized texture images. Each image is cropped into 9 subimages of size 128×128 pixels, resulting in a total of 540 texture samples. To better assess the performance of the neural network, two different training and testing experiments were conducted:

1. Unknown instance of known texture.
   In this experiment, 7 (i.e. 80%) of the 9 samples of each texture are used for training the network, 1 sample (10%) is used as validation data for determining when to stop training, and the remaining 1 (10%) sample is used as testing data. Although the testing samples are different from the training and validation samples, they are perceptually similar to some of the training and validation samples. Therefore, the testing error is expected to be small.

2. Unknown texture.
   This experiment tests the neural network's performance on unknown texture. It is an important test because new and unknown textures can appear in real applications. In this experiment, the training set consists of all the samples of 48 (80%) of the 60 textures. The validation and test sets each consists of all the samples of the other 6 (10%) of the 60 textures. Since the testing samples are perceptually different from the training and validation samples, the testing error is expected to be larger than that of the first experiment.

The mean squared errors of the two experiments are listed in Table 1. As expected, the neural network generalized well to new instances of known textures in Experiment 1. In Experiment 2, the network can be trained to the same accuracy level as in Experiment 1. But, the testing error of Experiment 2 is much larger than that of Experiment 1. Nevertheless, as we shall see in the next section, even though the mapping error of unknown textures is larger than that of unknown instances of known textures, the neural network method can still perform better than existing methods on texture-based image retrieval.

5 Texture-Based Image Retrieval

Based on the above mapping method, any texture can be mapped into a location in the perceptual space. Subsequently, texture similarity can be measured in terms of the Euclidean distance between the textures in the perceptual space. The texture similarity measured in the perceptual space would be more consistent with human's perception than would be existing matching methods.

Ideally, different instances of the same texture should map to the same location in the texture space. However, due to the presence of image noise and the limited resolutions of the discretized Gabor features used in the computation, the instances are mapped into a small region instead. Consequently, ranking based on Euclidean distance in the perceptual space alone may not be accurate enough to reflect minute differences between certain texture patterns.

To improve the ranking of the textures, the rank refinement process of Leow and Lai [8] is performed. A texture t's original Gabor feature vector is compared with those of its nearest N neighbors down the ranked list. The comparison is performed using normalized vector dot product of the Gabor feature vectors. Neighbors that are more similar to texture t are moved up the list whereas those that are less similar are moved down. This procedure is repeated for every texture in the ranked list.

In order to take into account perceptual similarity in evaluating image retrieval performance, the textures were categorized into 38 groups in the same manner as the work of Ma and Manjunath [10] and Leow and Lai [8]. All texture images in the same group as the query texture are regarded as relevant images in a query.

Retrieval experiments were conducted to compare the performance of the perceptually consistent texture matching method with existing methods. Three existing methods were considered:

1. The matching method based on the texture model of Tamura et al. [13] because it is used in many existing image retrieval systems.

2. The matching method based on the Wold model because it has been reported to outperform many existing methods for retrieving Brodatz texture images.[9]

3. The invariant matching method of Leow and Lai [8] which was reported to outperform the Tamura and the Wold methods.

For the perceptually consistent method, the neural network trained in Experiment 2 (Section 4) was used. That is, this network has learned only 80% of the textures during training.

A sample retrieval result is shown in Fig. 4. The images are displayed in decreasing order of similarity to the query texture. The images retrieved by the perceptually consistent method are perceptually more consistent to the query texture (the first image in the list) than those retrieved by other methods. For example, the Wold method retrieved marble textures which are not very similar to the query texture consisting of pellets. The perceptually consistent method also retrieved coffee bean textures which are similar to the query texture, but these textures were left out by the invariant method.

Two commonly used performance measurements of image retrieval were computed, namely normalized precision P_n and normalized recall R_n:[12]

$$P_n = 1 - \frac{\sum_{i=1}^{R} (\log r_i - \log i)}{\log N! - \log((N-R)!R!)} \tag{8}$$

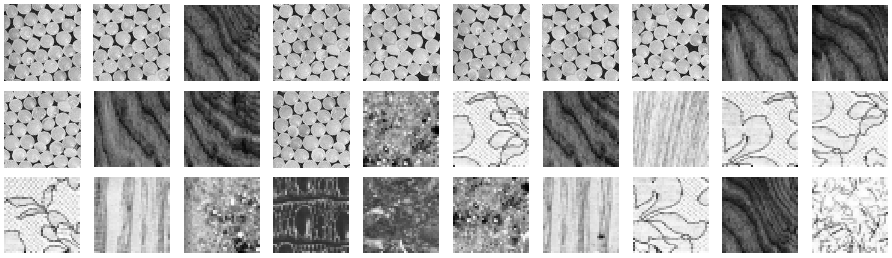
$$R_n = 1 - \frac{\sum_{i=1}^{R} (r_i - i)}{(N-R)R} \tag{9}$$

where N is the number of images in the database, R is the number of relevant images in the database, and r_i is the rank order of a relevant image i retrieved.

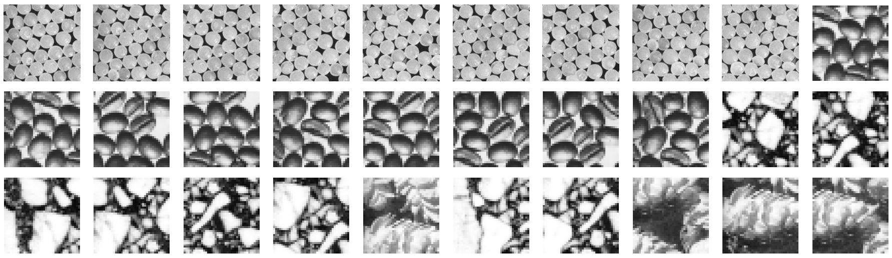
Table 2 summarizes the retrieval performance of various retrieval methods. To compare with existing methods, the same query textures used in the tests of Leow and Lai [8] are used in this test. In comparison, the perceptually consistent method performs slightly better than the invariant method and much better than the Wold and the Tamura methods.

6 Conclusion

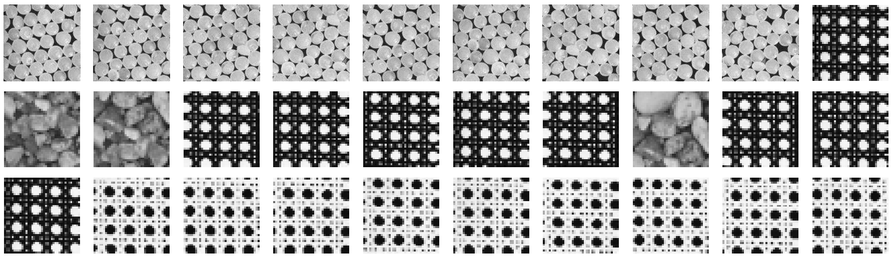
This paper presented a new perceptual texture space constructed based on human's perception of texture similarity obtained from psychological experiments. In contrast to existing perceptual texture space, this new space is constructed using texture images whose intensities, contrasts, scales, and orientations have been normalized to remove these confounding factors in human's perception. To illustrate the application of this texture space in image retrieval, a neural network has been trained to map Gabor texture features into the perceptual space. During image retrieval, images in the database are compared with the query texture in the perceptual space and are ranked in decreasing order of similarity to the query texture. The ranked order is fur-



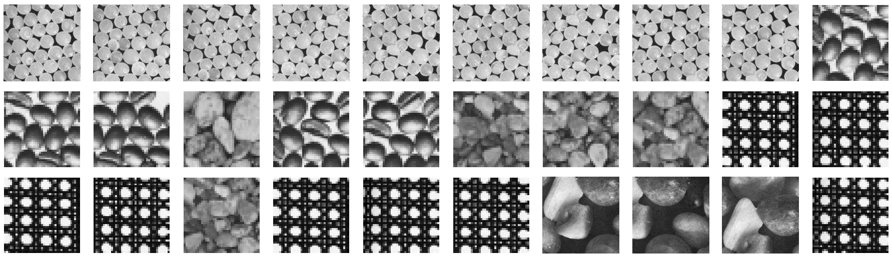
(a)



(b)



(c)



(d)

Figure 4. Sample retrieval results using (a) Tamura method, (b) Wold method, (c) invariant matching method, and (d) perceptually consistent method. The first texture is the query texture. The other textures are displayed in decreasing order of similarity to the query texture.

Table 2. Comparison of retrieval performance of various methods. Q lists the query textures, P_n and R_n are the normalized precision and recall rates, m denotes the mean performance, and Δ denotes the percentage improvement of the perceptually consistent method over the other methods.

	Tamura		Wold		Invariant		Perceptual	
Q	P_n	R_n	P_n	R_n	P_n	R_n	P_n	R_n
D15	0.995	0.999	1.000	1.000	0.981	0.998	1.000	1.000
D31	0.646	0.853	0.702	0.808	0.581	0.747	0.626	0.754
D49	1.000	1.000	0.986	0.999	1.000	1.000	1.000	1.000
D53	0.616	0.820	0.554	0.782	0.945	0.991	0.684	0.982
D66	0.427	0.609	0.853	0.963	0.615	0.855	0.895	0.921
D82	0.510	0.702	0.886	0.966	0.756	0.939	0.868	0.957
D96	0.702	0.871	0.724	0.934	0.836	0.942	0.847	0.986
D100	0.826	0.982	0.842	0.985	0.962	0.994	0.882	0.898
D112	0.872	0.974	0.915	0.981	0.829	0.965	0.787	0.954
m	0.733	0.868	0.829	0.935	0.834	0.937	0.843	0.939
Δ	15.0	8.3	1.7	0.4	1.1	0.2	–	–

ther refined to improve the ranking results. Experimental results show that the perceptually consistent method performs better than existing methods in retrieving texture images.

The method presented in this paper handles texture images with canonical scales and orientations. To apply the method to retrieving images of natural scenes which contain multiple textured regions, it is necessary to first segment the images into regions containing single textures. Since the textures in these images can have varying scales and orientations, a scale- and orientation-invariant texture matching method is required. We are now combining the invariant method of Lai and Leow [7,8] and the current method to achieve both invariant and perceptually consistent texture matching.

References

1. M. Amadasun and R. King. Textural features corresponding to textural properties. *IEEE Trans. Sys., Man, and Cyber.*, 19(5):1264-1274, 1989.
2. K. K. Benke, D. R. Skinner, and C. J. Woodruff. Convolution operators as a basis for objective correlates of texture perception *IEEE Trans. Sys., Man, and Cyber.*, 18(1):158-163, 1988.
3. P. Brodatz. *Textures: A Photographic Album for Artists and Designers.* Dover, New York, 1966

4. D. C. Donderi. Information measurement of distinctiveness and similarity. *Perception and Psychophysics*, 44(6), 576-584, 1988.
5. Joseph. F. Hair, R. E. Anderson and R. L. Tatham. *Multivariate Data Analysis*. Prentice Hall, 1998.
6. C. Heaps and S. Handel. Similarity and features of natural textures. *J. Experimental Psychology: Human Perception and Performance*, 25(2):299-320, 1999.
7. S. Y. Lai and W. K. Leow Invariant texture matching for content-based image retrieval. In *Proc. Int. Conf. on Multimedia Modeling MMM '97*, 53–68.
8. W. K. Leow. and S. Y. Lai Invariant matching of texture for content-based image retrieval. In *Texture Analysis in Machine Vision*, M. Pietikainen (Ed.), 2000.
9. F. Liu and R.W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. PAMI*, 18(7):722-733, 1996.
10. W. Y. Ma and B. S. Manjunath. Texture features and learning similarity. In *Proc. of IEEE Conf. CVPR*, 1160-1169, 1996.
11. A. R. Rao and G. L. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. *IEEE Conf. on Visualization*, 220-227, 1993.
12. G. Salton. *Introduction to Modern Information Retrieval.* McGraw Hill, New York, 1983.
13. H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Trans. Sys., Man, and Cyber.*, 8(6):460-473, 1978.