Perceptual Texture Space Improves Perceptual Consistency of Computational Features

Huizhong Long and Wee Kheng Leow
School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
longhuiz, leowwk@comp.nus.edu.sg *

Abstract

Perceptual consistency is important in many computer vision applications. Unfortunately, except for color, computational features and similarity measurements for other visual features are not necessarily consistent with human's perception. This paper addresses three critical issues regarding perceptually consistent texture analysis: (1) development of perceptual texture space, (2) assessment of how consistent computational features are to human perception, and (3) mapping computational features to perceptual space. It demonstrates the construction of a reliable perceptual texture space, which can be used as a yardstick for assessing the perceptual consistency of computational features and similarity measurements. Moreover, it is found that commonly used computational texture features are not very consistent with human perception, and mapping them to the perceptual space improves their perceptual consistency.

1 Introduction

Perceptual consistency is important in many computer vision applications. For example, a computer image understanding system should analyze and interpret an image in a manner that is consistent with human's perception of the image. A content-based image retrieval system should compare images in its database with the query in a manner that is consistent with human's perception of image similarity. To this end, many systems and algorithms measure color similarity in the CIE L*u*v* space which has been shown to be quite consistent with human's perception [Meyer and Greenberg, 1987]. Unfortunately, computational features and similarity measurements for other visual features are not necessarily consistent with human's perception, with the possible exception of Santini and Jain's *fuzzy features contrast model* [Santini and Jain, 1999].

Several perceptual spaces that capture human's perception of texture exist. However, the mapping from computational features to these perceptual spaces has not been derived. Another difficulty with using these perceptual spaces is that the scales and orientations of the textures used to construct the spaces are not normalized. Therefore, texture similarity measured in these spaces are confounded by scale and orientation variations. This problem makes it very difficult to use existing texture spaces in practical applications.

This paper addresses three critical issues regarding perceptually consistent image analysis, with specific application to texture analysis:

1. Development of a *perceptual texture space* (PTS) based on texture images with canonical scales and orientations (Section 3). Once the reliability of PTS is established, it can be used as a yardstick for assessing the perceptual consistency of computational features and similarity measurements.
2. Assessment of how consistent commonly used computational texture features and similarity measurements are to human's perception (Section 4).
3. Assessment of the accuracy of mapping computational texture features to PTS (Section 5). Once an accurate mapping is accomplished, texture similarity can be measured in PTS to achieve perceptual consistency.

2 Perceptual Consistency

There are many ways of defining perceptual consistency. This section discusses some definitions which will be used throughout the paper. Let p_{ij} denote the perceptual distance between samples i and j, and d_{ij} denote the corresponding measured or computational distance. A simple notion of perceptual consistency is that d_{ij} is *proportional* to p_{ij}. That is, there exists a linear function f such that

$$p_{ij} = f(d_{ij}), \ \ \forall i, j. \tag{1}$$

Then, perceptual consistency can be measured in terms of the mean squared error (MSE) e of linear regression:

$$e = \frac{1}{N} \sum_{i,j} (p_{ij} - f(d_{ij}))^2 \tag{2}$$

where N is the number of sample pairs. The smaller the MSE, the better is the consistency. A perfect consistency has an MSE of 0.

*This research is supported by NUS ARF R-252-000-049-112 and joint NSTB and NUS ARF RP3999903.

A less stringent notion of perceptual consistency is to require that f be a monotonic function which can be nonlinear. The difficulty with this definition is that it is difficult to determine the best nonlinear function to use in practice.

An alternative definition is to require that d_{ij} be *statistically correlated* to p_{ij}. In this case, it is useful to transform the populations $\{d_{ij}\}$ and $\{p_{ij}\}$ to equivalent zero-mean unit-variance populations $\{d'_{ij}\}$ and $\{p'_{ij}\}$:

$$p'_{ij} = \frac{p_{ij} - \bar{p}}{\sigma_p}, \quad d'_{ij} = \frac{d_{ij} - \bar{d}}{\sigma_d} \tag{3}$$

where $\bar{p}$ and $\bar{d}$ are the means and σ_p and σ_d are the standard deviations of the populations. Then, perceptual consistency can be measured in terms of the correlation r:

$$r = \frac{1}{N} \sum_{i,j} p'_{ij}\, d'_{ij} \, . \tag{4}$$

Substituting Eq. 3 into Eq. 4 yields the Pearson's correlation coefficient:

$$r = \frac{\displaystyle\sum_{i,j} (p_{ij} - \bar{p})(d_{ij} - \bar{d})}{\left[\displaystyle\sum_{i,j} (p_{ij} - \bar{p})^2 \sum_{i,j} (d_{ij} - \bar{d})^2\right]^{1/2}} \, . \tag{5}$$

The coefficient r ranges from -1 to $+1$ (for perfect correlation). In the following sections, we will use both MSE and Pearson's correlation coefficient to measure two different aspects of perceptual consistency. Note that, with perfect consistency ($e = 0$ or $r = 1$), we obtain the following condition:

$$d_{ij} \leq d_{kl} \Rightarrow p_{ij} \leq p_{kl} \quad \text{for any textures } t_i, t_j, t_k, t_l. \tag{6}$$

That is, if perfect consistency is achieved, computational similarity would imply perceptual similarity. This condition is especially useful in practical applications (Section 6).

3 Development of Perceptual Texture Space

This section addresses the first critical issue put forth in Section 1: development of perceptual texture space (PTS). First, let us briefly review existing perceptual texture models.

3.1 Existing Perceptual Texture Models

The earliest study of human's perception of texture similarity was conducted by Tamura et al. [Tamura *et al.*, 1978] In their experiments, 48 human subjects were asked to judge the similarity of texture pairs according to six visual properties, namely, coarseness, contrast, directionality, line-likeness, regularity, and roughness. Similarity judgments were measured and each texture was assigned a perceptual rating value along each of the six visual scales. Due to the combinatorial nature of the task, only 16 textures were used. Amadasun and King [Amadasun and King, 1989] conducted similar ranking experiments to measure similarity judgments according to various visual properties.

The major difficulty with these studies is that the subjects were asked to judge texture similarity according to subjective visual properties. Unfortunately, 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 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 different visual dimensions cannot be easily combined to form a perceptual texture space.

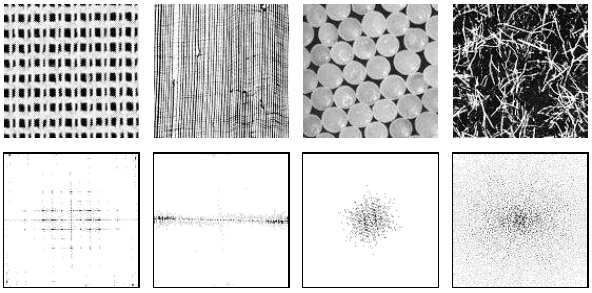
To avoid these difficulties, Rao and Lohse [Rao and Lohse., 1993] performed an experiment in which 20 subjects were asked to sort 56 textures into as many groups as the subjects wished such that the textures in each group were perceptually similar. The textures were sorted based on the subjects' perception of overall texture similarity instead of individual visual properties. A co-occurrence matrix of the sorting results was computed and multidimensional scaling [Hair *et al.*, 1998] was performed to derive a 3D perceptual space. Rao and Lohse concluded that the 3 dimensions of the space strongly correlate with the visual properties of repetitiveness, orientation, and complexity.

Heaps and Handel [Heaps and Handel, 1999] conducted further studies using the same methodology. However, they arrived at different conclusions than those of Rao and Lohse. They concluded that it is not possible to reliably associate a single visual property to each dimension of the texture space because each dimension appears to be correlated with a combination of more than one visual property. In addition, perception of texture similarity depends on the context in which the similarity is judged. That is, how similar two textures appear 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 development of perceptual texture space, we do not assign visual properties to the dimensions of the space. Moreover, the influence of the context problem is reduced by (1) selecting most, if not all, 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 used in the psychological experiment. It is well known, for example, that human's sensitivity of perceiving spatial frequency (i.e., spatial patterns) is dependent on contrast [Schiffman, 1996]. Therefore, these confounding factors should be removed from the experiment.

3.2 Texture Image Preparation

A rich set of texture images are collected to construct the perceptual texture space (PTS). Fifty texture images are selected from the Brodatz album. The remaining images in the album are omitted for various reasons. For example, some textures are scaled versions of each others, and only one of them is required. Another reason is that the contents of some images (e.g., flowers in lace) appear very striking but their texture patterns are perceptually weak. It is very difficult for the subjects to avoid classifying the images based on the content instead of the texture alone. 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



(a) (b) (c) (d)

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

album. These 60 texture images constitute a good set of representative textures that appear in images of natural scenes. These images are first normalized before they are used in the psychological experiments.

Four image features, namely, intensity, contrast, scale, and orientation, that can confound the subject's judgment are normalized. The intensity and contrast of a texture image is measured as the mean and standard deviation of the intensity of the pixels in the image. A texture's intensity and contrast are normalized by a simple linear transformation to a canonical mean intensity and standard deviation.

Our analysis reveals 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 discrete Fourier transform (DFT) patterns, each group of textures can be scaled in the same way. The four groups consist of structured, directional, granular, and random textures (Fig. 1).

Structured textures have very well defined DFT patterns which contain a small number of highly localized, prominent frequency components with large amplitudes (Fig. 1a). The textures in this group are scaled and rotated so that the frequencies of the 8 most prominent components in the DFTs of various textures are identical. The DFT 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 DFT patterns have the same orientations.

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

3.3 Measurement of Human Perception

A direct way of measuring the dissimilarity δ_{ij} between textures t_i and t_j is to perform an exhaustive comparison of all $n(n-1)/2$ sample pairs. When the number of samples n is large (e.g. 60), the number of pairs becomes very large (in this case, 1770). It is practically impossible for a subject to perform such a large number of judgments. Another commonly used method is to sort the samples into as many groups as the subjects wish such that the samples in each group are perceptually similar. This method sorts n samples into g groups, requiring at most ng judgments, which is far smaller than $n(n-1)/2$. This method is also adopted in [Rao and Lohse., 1993] and [Heaps and Handel, 1999].

In our experiment, sixty subjects were asked to freely sort the images according to texture similarity. The sorting results were used to construct distance matrices based on two similarity measurements: co-occurrence matrix and information measurement [Donderi, 1988].

Let c_{ij} denote the number of times textures t_i and t_j are grouped into the same group. Then, the distance d_{ij} measured by the co-occurrence method is:

$$d_{ij} = n - c_{ij} \tag{7}$$

where n is the number of subjects in the experiment.

Donderi's distance measurement d_{ij} is based on Shannon's information theory :

$$d_{ij} = \sqrt{H(t_i|t_j) + H(t_j|t_i)} \tag{8}$$

where $H(t_i|t_j)$ is the *conditional entropy* or *independent information* of t_i after knowing the grouping of t_j. Let G denote the set of textures used in the experiment and $G(t_i)$ denote the group in which t_i belongs. Then, the independent information $H(t_i|t_j)$ is computed as follows:

$$H(t_i|t_j) = \left\{ \begin{array}{cl} 0 & \text{if } G(t_i) = G(t_j) \\ \log_2 |G| - \log_2 |G(t_i)| & \text{if } G(t_i) \neq G(t_j) \end{array} \right. . \tag{9}$$

According to Eq. 8, informationally independent stimuli are dissimilar and informationally redundant stimuli are similar. It has been demonstrated that information measurements derived from the free sorting task are empirically and theoretically equivalent to the distance measurements generated by direct pairwise comparison [Donderi, 1988].

3.4 Construction of Perceptual Space

After obtaining the distance matrices, multidimensional scaling (MDS) was applied to derive a perceptual texture space (PTS). MDS, also known as *perceptual mapping*, maps the original dissimilarity judgments δ_{ij} into points in a multidimensional space such that each point corresponds to a texture, and the Euclidean distance p_{ij} between two textures t_i and t_j matches as well as possible the original dissimilarity δ_{ij}. The multidimensional space would correspond to a perceptual space, and the Euclidean distance measured in the space would correspond to perceptual distance.

The appropriate number of dimensions of the perceptual space is determined by the *stress* measure which indicates the proportion of the variance of the data that is not accounted for [Hair *et al.*, 1998]:

$$\text{stress} = \sqrt{\sum_{i,j} (p_{ij} - \delta_{ij})^2 \Big/ \sum_{i,j} p_{ij}^2} \tag{10}$$

The larger the number of dimensions, the better is the fitting of the data and the lower is the stress. However, a space

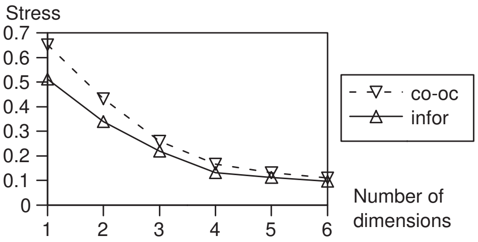


Figure 2: Stress plot of the MDS fitting results. Optimal trade-off between accurate fitting and small number of dimensions is obtained with four dimensions. The space constructed with Donderi's information measurement (infor) has a better fit than that constructed using co-occurrence (co-oc).

Table 1: Comparison of Heaps and Handels's space and our perceptual texture space (PTS) with Rao and Lohse's space. Pearson's correlation coefficients show that the spaces are consistent with each others.

perceptual space	3D	4D	5D
Heaps & Handel	0.790	–	–
PTS (co-occurrence)	0.722	0.732	0.713
PTS (info. measure)	0.726	0.694	0.695

with too many dimensions is difficult to visualize and to use. So, a trade-off is made to find a small number of dimensions that offer good data fitting. Rao and Lohse [Rao and Lohse., 1993], as well as Heaps and Handel [Heaps and Handel, 1999], concluded that a texture space with three dimensions is appropriate. Our MDS fitting results show that increasing the number of dimensions beyond 4 does not significantly reduce the stress. Thus, a four-dimensional texture space provides a better trade-off (Fig. 2).

The constructed spaces are verified by comparing them with existing perceptual spaces. This comparison is performed by computing the Pearson's correlation (Eq. 5) between the distances measured in the spaces. Table 1 summarizes the comparison results. Heaps and Handel reported a good correlation ($r = 0.790$) with Rao and Lohse's data [Heaps and Handel, 1999]. Our PTS constructed using co-occurrence has a better correlation with Rao and Lohse's space compared to that using Donderi's information measurement. This is expected because Rao and Lohse's space was developed using co-occurrence as well. We do not have Heaps and Handel's data for direct comparison. Nevertheless, Table 1 shows that the spaces are mutually consistent but not the same. The difference between the space of Rao and Lohse and our PTS can be attributed to the normalization of texture images. Thus, the PTS can be regarded as a reliable measurement of human's perception of texture similarity. In the following sections, the 4D PTS constructed based on information measurement will be used as the standard perceptual space because it has a better fit to the psychological data than that developed using co-occurrence.

4 Consistency of Computational Features

This section addresses the second critical issue: assessment of the perceptual consistency of computational texture features and similarity measurements. First, let us review commonly used texture features and similarity measurements.

4.1 Computational Texture Features

Statistical and spectral texture models are commonly used in existing image retrieval systems. Statistical models categorize textures according to statistical measurements of visual qualities such as coarseness and granularity. Spectral models characterize textures based on Fourier spectrum or filtering results. The statistical features proposed by Tamura et al. [Tamura *et al.*, 1978] are used in IBM's QBIC system. The Wold model [Liu and Picard, 1996] based on Fourier transform (spectral model) and Multiresolution Simultaneous Autoregressive model (MRSAR, statistical model) is used in MIT's PhotoBook. Gabor features (spectral model) are used in Ma and Manjunath's NeTra [Ma and Manjunath, 1996] and Leow and Lai's invariant texture space [Leow and Lai, 2000].

The above features are easy to compute, but the texture similarity computed based on these features are not necessarily consistent to human's perception. On the other hand, Santini and Jain developed the *fuzzy features contrast model* (FFC) [Santini and Jain, 1999] which has the potential of being perceptually consistent. Their similarity measurement is based on Tversky's feature contrast model [Tversky, 1977] which can account for various peculiarities of human's perceptual similarity. They have applied FFC to measure texture similarity based on Gabor features, and have obtained encouraging results. Unfortunately, due to combinatorial problem, they have compared FFC's measurements with only a small number of human ranking data, and are not able to perform more thorough comparison with human perception.

FFC defines a fuzzy membership function μ_{ki} for each Gabor feature G_{ki} of texture t_k by a sigmoid function:

$$\mu_{ki} = \left(1 + \exp\left(-\frac{G_{ki} - \bar{G}_i}{\gamma\sigma_i}\right)\right)^{-1} \qquad (11)$$

where $\bar{G}_i$ is the mean feature of the textures along the i-th dimension, σ_i is the corresponding standard deviation, and γ is a constant parameter (which is fixed at 1 in [Santini and Jain, 1999]). The similarity between two Gabor features G_{ki} and G_{li} is then defined as

$$\begin{aligned} S(G_{ki}, G_{li}) &= \sum_i \min\{\mu_{ki}, \mu_{li}\} \\ &- \alpha \sum_i \max\{\mu_{ki} - \mu_{li}, 0\} - \beta \sum_i \max\{\mu_{li} - \mu_{ki}, 0\} \end{aligned} \qquad (12)$$

where α and β are parameters. Various ranges of parameter values were tried. The values that produced the best match between FFC and PTS are $\alpha = \beta = 50$, $\gamma = 0.001$.

The major difficulty with using FFC is that the values of the parameters must be carefully selected to match human's perception. The authors acknowledge that the problem of determining the parameter values has not been adequately addressed [Santini and Jain, 1999].

4.2 Perceptual Consistency Tests

Perceptual consistency is assessed by comparing the texture distances measured by various features with that measured

Table 2: Comparison of various computational features and similarity measurements with PTS. Mahalanobis distance is used for MRSAR in the Wold model. r = Pearson's correlation coefficient, e = mean squared error.

feature	distance	r	e
Tamura	Euclidean	0.251	0.132
Gabor	Euclidean	0.273	0.131
Gabor	FFC	0.330	0.115
MRSAR	Euclidean	0.144	0.139
MRSAR	Mahalanobis	0.061	0.152
invariant	Euclidean	0.222	0.136

in the 4D PTS. Both Pearson's correlation coefficient (Eq. 5) and the mean squared error (MSE) of linear regression (Eq. 2) are used as the measurement of consistency. The larger the Pearson's coefficient, or the smaller the MSE, the more consistent computational features are to PTS (Section 2).

The following features are assessed: Tamura's features, Gabor, MRSAR, and invariant texture space. We could not assess the Wold model directly because it defines only a ranking order of the textures without explicit measurement of texture similarity. Therefore, only the MRSAR features used in the Wold model are assessed. The invariant texture space method maps Gabor features to a space that is invariant to texture scales and orientations [Leow and Lai, 2000], whereas the other features are not invariant. So, it is interesting to see how consistent is the invariant space to PTS.

Table 2 summarizes the results, and shows that Gabor feature and Gabor with FFC are most consistent with PTS. In particular, measuring Gabor similarity with FFC does improve Gabor feature's consistency. Measuring MRSAR similarity with Euclidean distance is perceptually more consistent than measuring with Mahalanobis distance. The degrees of consistency of computational features ($r \approx 0.3$) are, however, not very high compared to those between various perceptual spaces (Table 1, $r \approx 0.7$). Therefore, it can be concluded that these computational features and similarity measurements are not very consistent with human's perception.

5 Mapping Features to Perceptual Space

The previous section demonstrated that computational texture features are not very consistent with human's perception as measured in the 4D PTS. This section addresses the third critical issue: Assessment of the accuracy of mapping computational texture features to PTS.

The feature mapping problem is to find a set of functions f_k, $k = 1, \ldots, 4$, that map the computational features $\mathbf{x}$ of a texture into the coordinates y_k in PTS. That is, $y_k = f_k(\mathbf{x})$. Our study reveals that the functions f_k are typically nonlinear, and the form of the nonlinearity cannot be readily determined. Therefore, both multilayer neural networks and Support Vector Machine (SVM) regression were tested. SVM regression was found to give more accurate results. In addition, it has fewer parameters to tune than do neural networks. So, only the results of SVM regression are reported here.

SVM regression maps a multidimensional input to a single output value. Thus, four separate SVMs are required to map computational features $\mathbf{x}$ to the four coordinates y_k, $k = 1, \ldots, 4$, of PTS:

$$y_k = \sum_i (\alpha_{ki} - \alpha^*_{ki})\, g(\mathbf{v}_{ki}, \mathbf{x}) + b_k \tag{13}$$

where α_{ki} and α^*_{ki} are the Lagrange multipliers (analogous to the weights of neural networks), g is a kernel function, $\mathbf{v}_{ki}$ are the support vectors, and b_k is the bias term. We experimented with various kernel functions and found that the Gaussian kernel consistently produced more accurate results. Various parameter values of the Gaussian kernel were tried, and it was found that different computational features require different parameter values for optimal mapping.

The assessment of perceptual consistency used the 60 normalized texture images described in Section 3.3 as the data set. Each image was cropped into 9 subimages of size 128×128 pixels, resulting in a total of 540 texture samples. To better assess the performance of SVM, two different training and testing experiments were conducted:

1. Unknown instance of known texture.
   In this experiment, 8 of the 9 samples of each texture were used for training, and the remaining 1 sample was used for testing. Although the testing samples were different from the training samples, they were perceptually similar to some of the training samples. Therefore, the testing error was expected to be small. 9-fold cross-validation tests were conducted.
2. Unknown texture.
   This experiment tests SVM's performance on unknown texture. It is an important test because new textures can appear in real applications. In this experiment, all the samples of 54 (90%) of the 60 textures were used for training, and all the samples of the other 6 (10%) textures were used for testing. Since the testing samples were perceptually different from the training samples, the testing error was expected to be larger than that of the first experiment. 10-fold cross-validation tests were conducted.

Table 3 illustrates the mapping results obtained by SVM regression. As expected, for all the features, testing errors for unknown instance is smaller than those for unknown texture.

Table 3: Testing errors of feature mapping. Columns 2 to 5 are the average testing errors of n-fold cross-validation. 1D denotes the mean squared error (MSE) along one of the four dimensions of PTS. 4D denotes the overall MSE of the 4D coordinates in PTS. Columns 6 and 7 are the results of comparing the distances in PTS with those measured by the features *after* they are mapped to PTS.

	unknown instance		unknown texture		distance comparison	
features	1D	4D	1D	4D	r	e
Tamura	0.036	0.144	0.064	0.256	0.857	0.020
Gabor	0.0013	0.0052	0.054	0.216	0.862	0.018
MRSAR	0.024	0.096	0.061	0.244	0.859	0.019
invariant	0.031	0.124	0.067	0.268	0.632	0.031

Moreover, being most consistent with PTS (Table 2), Gabor features can be mapped to PTS more accurately than other features can. In particular, Gabor features' testing error for unknown instance is much smaller than those of other features and those for unknown texture tests. Therefore, it can be concluded that accurate mapping to PTS can be achieved, at least for Gabor features and under the condition that some samples of all texture types appear in the training set.

After mapping computational features to PTS, one would expect the mapped coordinates to be more consistent with PTS. A comparison is performed between PTS and the computational features mapped by SVM models trained under the second condition (unknown texture). The distance correlation results are shown in Table 3. Comparing Table 3 with Table 2 shows that mapping computational features to PTS does improve the perceptual consistency of the features.

6 Application Examples

Given an accurate mapping, it is now possible to use PTS for various applications, such as, retrieval, classification, and classification with uncertainty.

1. In retrieval, we like to sort a set of textures into an ordered list $\{t_i\}$ such that, with respect to the query texture t_0, $p_{0i} \leq p_{0j}$ for any $i < j$, i.e., textures that are perceptually similar to t_0 are placed at the beginning of the list. This problem can be solved by mapping t_0 to PTS and ordering the textures t_i according to d_{0i}.
2. In classification, we like to classify a texture t_0 to a class C such that for any $t_i \in C$ and any $t_j \notin C$, $p_{0i} \leq p_{0j}$. This can be achieved by classifying t_0 to the class, say, C of its nearest neighbor, say, t_k, i.e., $d_{0k} \leq d_{0i}$ for any $t_i \neq t_k$. Since the distance between textures in the same class is smaller than the distance between textures in different classes, by perceptual consistency, $p_{0i} \leq p_{0j}$ for any $t_i \in C$ and any $t_j \notin C$.
3. The distance measurement d_{ij} in PTS can be used to represent uncertainty of classification. It is computed in terms of the independent information $H(t_i|t_j)$ and $H(t_j|t_i)$ (Eq. 8), which represents the uncertainty of classifying t_i given t_j and vice versa.

7 Conclusion

This paper has addressed three critical issues about perceptually consistent texture analysis:

1. A 4D perceptual texture space (PTS) has been developed. By comparing with existing perceptual spaces, it is verified that the PTS provides a reliable measurement of human's perception of texture similarity.
2. It is shown, by correlating with the distances measured in PTS, that commonly used computational texture features and similarity measurements are not very consistent with PTS. Among the various features, Gabor features and Gabor with FFC give the highest correlation.
3. It is shown that mapping computational features to PTS improves the features' perceptual consistency. In particular, Gabor features can be most accurately mapped to PTS, especially when some samples of all texture types appear in the training set.

This paper thus establishes that PTS can be used as a yardstick for assessing the perceptual consistency of computational features and similarity measurements.

For practical applications, it might be necessary to find a method that can map computational features to PTS accurately under the more stringent condition of unknown texture types. This would be necessary if some rare textures that are distinct from the known textures used in the training process appear during actual run. In addition, it would be necessary to perform the mapping in a manner that is invariant to texture scale and orientation.

References

[Amadasun and King, 1989] M. Amadasun and R. King. Textural features corresponding to textural properties. *IEEE Trans. SMC*, 19(5):1264–1274, 1989.

[Donderi, 1988] D. C. Donderi. Information measurement of distinctiveness and similarity. *Perception and Psychophysics*, 44(6):576–584, 1988.

[Hair *et al.*, 1998] Joseph. F. Hair, R. E. Anderson, and R. L. Tatham. *Multivariate Data Analysis*. Prentice Hall, 1998.

[Heaps and Handel, 1999] C. Heaps and S. Handel. Similarity and features of natural textures. *J. Expt. Psycho.: Human Perception and Performance*, 25(2):299–320, 1999.

[Leow and Lai, 2000] W. K. Leow and S. Y. Lai. Scale and orientation-invariant texture matching for image retrieval. In M. K. Pietikäinen, editor, *Texture Analysis in Machine Vision*. World Scientific, 2000.

[Liu and Picard, 1996] 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.

[Ma and Manjunath, 1996] W. Y. Ma and B. S. Manjunath. Texture features and learning similarity. In *Proc. of IEEE Conf. CVPR*, pages 1160–1169, 1996.

[Meyer and Greenberg, 1987] G. W. Meyer and D. P. Greenberg. Perceptual color spaces for computer graphics. In H. J. Durett, editor, *Color and the Computer*. Academic Press, Lodon, 1987.

[Rao and Lohse., 1993] A. R. Rao and G. L. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. In *IEEE Conf. on Visualization*, pages 220–227, 1993.

[Santini and Jain, 1999] S. Santini and R. Jain. Similarity measures. *IEEE Trans. PAMI*, 21(9):871–883, 1999.

[Schiffman, 1996] H. R. Schiffman. *Sensation and Perception: An Integrated Approach*. John Wiley & Sons, 4 edition, 1996.

[Tamura *et al.*, 1978] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Trans. SMC*, 8(6):460–47, 1978.

[Tversky, 1977] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.