M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 237
Appeared: IEEE ICASSP Adelaide, Apr. 1994, V:129-132

A new Wold ordering for image similarity

Rosalind W. Picard and Fang Liu *

Abstract

The problem of measuring perceptual similarity between images is addressed using a new image model based on the Wold decomposition. The model permits separate treatment of image components which correspond approximately to periodicity, directionality, and randomness. We compare its performance in an image search application to two other methods – one based on shift-invariant principle components and one based on a multiscale simultaneous auto-regressive model. When textured images are ordered by distances between their Wold components, the results appear to be much closer to the human perception of similarity. We discuss how decoupling the three components can increase flexibility for measuring image similarity and can save computation, permitting the "quickest" matches when the features are the most perceptually "salient."

1 Introduction

Measuring similarity in images is an important problem in image processing, particularly in new applications such as image search and retrieval. Although our look at the problem is in a relatively new context, it is not a new problem, as the need for measures of perceptual similarity has been present in the image coding community for decades – i.e., the need to find a more perceptual measure than signal-to-noise ratios. There has also been emphasis in extracting textural features which correspond to human perception of similarity, notably the work of Tamura *et. al.* [1], which has focused on discrimination, without the goal of being able to synthesize the data. Here, the general problem is one of identifying a model capable of synthesis, and having a "perceptual" parameter space so that distances between images in the parameter space are close when humans perceive the images to be similar, and are not close otherwise. The model for the database search application is successful if ordered distances between its parameters correspond to a human ordering of the images by perceptual similarity.

In pursuit of a more perceptual similarity metric, this work investigates a relatively new image model based on the Wold decomposition for regular stationary stochastic processes [2]. If an image is assumed to be a homogeneous 2-D discrete random field, then the 2-D Wold-like decomposition is a superposition of three mutually orthogonal components: a purely-indeterministic field, a generalized-evanescent field and a harmonic field [3]. The background theory as well as some applications of this 2-D Wold-like decomposition to spectral estimation and modeling of homogeneous textures can be found in [3] and [4].

It is necessary to note that the Wold-based theory and applications presented to date in the literature assume the random field is stationary; this important assumption is usually violated in natural images. Inhomogeneities may even arise in homogeneous data simply by viewing it with perspective or affine transformations, or by viewing it on the surface of a 3D non-planar object. The work of Francos *et. al.* did not attempt to apply this model to inhomogeneous data, nor is their implementation designed to be applied to such data [5]. This paper addresses the problem of adapting the model for use in natural texture data.

In addition to its significance in random field theory, we have found that the Wold-based model has an interesting relationship to independent psychophysical findings of perceptual similarity. Noteworthy is a recent study by Rao and Lohse where humans grouped patterns according to perceived similarity [6]. The three orthogonal dimensions identified were repetitiveness, directionality, and complexity. These dimensions might be considered the perceptual equivalents of the harmonic, evanescent, and indeterministic components in the Wold-based model.

2 New Wold implementation

The new model implementation consists of three stages. The first stage determines if there is strong periodic structure. Although highly structured textures may contain all three Wold components, their harmonic components are usually prominent and provide good features for comparison. Not only is this component more salient than the random component (agreeing with Rao and Lohse's ordering of the three texture dimensions) but it is the quickest to compute.

In the first stage, the autocorrelation function of the image is computed by the inverse Fourier transform of the image power spectrum density function. For periodic patterns, the energy concentrated regions of their autocorrelation functions are also periodic and spread over the displacement plane while the random-looking textures have most of their energy in a small displacement region. Examples are shown in Figure 1, where D3 was chosen as highly structured. By computing the ratio between the small displacement energy and the total energy for a training set of images, a decision boundary is established and subsequently applied as a threshold. The threshold found and used in this work is 18%.

*This work was supported in part by BT and by NEC.

The second stage of processing occurs for periodic images on the peaks of their Fourier transform magnitudes. An algorithm is implemented first to estimate the location of large local maxima and then to extract the fundamental frequencies of all harmonic peaks. The direction of the harmonic frequency which is closest to the origin is regarded as the main orientation angle of the texture. Picking the frequency with the largest amount of energy was found not to be robust since the energy is a result of many factors such as lighting and contrast. Instead, the structural arrangement of the peaks appears to be important for perceptual comparisons. Rotations and other transformations may be applied to the peaks to align them before comparison. Applying the transformations to the peaks can be a lot less computational than applying them to the entire image.

Finally, the comparison of periodic texture images is carried out by matching their harmonic peaks. Let $m_c(s)$ and $m_t(r)$ be the feature sets of the class image and a test image respectively, where $s = (s_1, s_2), r = (r_1, r_2) \in \mathcal{T}$. Region $\mathcal{T}$ is half of the discrete frequency plane. By the "class image", we mean the image selected by a user as a representation of what kind of "class" the system should look for, and by a "test image" we mean one of the database images whose similarity to the class image is to be measured. The similarity measure between these two images is defined as

$$M_{ct} = \sum_{s \in \mathcal{T}} m_c(s) \sum_{r \in \mathcal{T}} w_m(r-s) \frac{m_c(s) m_t(r)}{[m_c(s) + m_t(r)]^2}, \qquad (1)$$

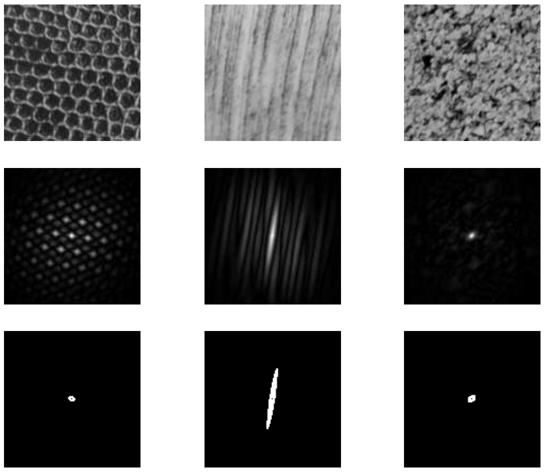
where $w_m(\cdot)$ is a weighting function for the frequency deviation of two peaks, implemented here as a 5×5 mask with unity at the center and decaying values as a function of distance from the center. The ratio term measures the relative value of the peaks since

$$\left[\frac{m_c(s)}{m_c(s) + m_t(r)} \cdot \frac{m_t(r)}{m_c(s) + m_t(r)} \right]$$

reaches its maximum when $m_c(s) = m_t(r)$. For each peak in the class feature set, this measure looks for peaks in the test feature set within a neighborhood in the frequency plane. Peak matching within a small neighborhood is necessary due to the image inhomogeneities and the frequency sampling effects of the DFT. Note that the larger the value M_{ct} is, the more similar the two images are.

The third stage of processing is applied when an image is not highly structural. This is the most computationally costly part of the procedure, and can be omitted to result in substantial savings if the previous stage indicates that the harmonic information is sufficient for a given task. In this stage we approximate the indeterministic component and the evanescent "directional" component, which can be thought of as corresponding to the two less salient dimensions identified in the study of Rao and Lohse.

The most general model for the purely-indeterministic component is the moving average (MA) model. However, under certain assumptions, an auto-regressive (AR) representation of this part of the random field exists [4]. Various implementations of auto-regressive models have been used successfully for segmenting 4-8 textures in an image [7]. In this work we use the simultaneous auto-regressive (SAR) model of Mao and Jain [7] for the purely-indeterministic component, as well as by itself for comparison to the Wold



(a) (b) (c)

Figure 1: From the top row to the bottom: the original Brodatz textures; their autocorrelation functions; and their regions used for computing the small displacement energy. (a) D3: Reptile skin. The energy concentrated regions are spread over the entire displacement plane. (b) D69: Wood grain. The small displacement region contains more energy. (c) D4: Beach sand. Most of the energy is concentrated in the small displacement region.

model in the image search application. This implementation cascades second-order AR parameters estimated over three different scales in the image, for a total of 15 parameters per texture.

It is important to note that theoretically in the Wold model, the AR parameters are estimated after removal of the harmonic component. The result is the application of the AR model to a more "continuous" spectrum. (In practice of course, the implementation is discrete.) This stands in contrast to the typical application of an AR model directly to texture, where the low-order model may be insufficient to appropriately capture higher-order behavior contributed by spectral discontinuities.

The evanescent component can be estimated by fitting a line in the 2D spectral domain. This is similar to the computer vision problem of estimating global dominant orientations in an image. Here, we approximate the evanescent information with an estimate of the texture's dominant orientations. These are found by using a basis set of oriented bandpass filters and a decision process based on thresholding orientation histograms [8]. To conclude the third stage of processing, only the textures which possess the same number of main orientations are compared by examining the Mahalanobis distances between their SAR parameters.

3 Similarity experiments

In this section we illustrate the performance of the new Wold model on the Brodatz database, which consists of 1008 texture patches cropped from all 112 images in the Brodatz Album [9]. Each Brodatz texture provides nine non-overlapping 8-bit 128×128 subimages. This collection of natural textures exhibits large variety, including many inhomogeneous patterns; therefore, it provides a challenge to traditionally homogeneous image models.

The image search environment used here, as well as the performance of a Karhunen-Loéve (principal components) based model applied to the database search problem, were previously discussed in [10]. Here, the performance of the Wold-based model is illustrated on two examples and compared to the performance of both the shift-invariant principal components method of [10] (based on a pooled covariance of 100 training images, and a subspace of the 20 principal components with largest eigenvalues) and the 15 parameter multiscale SAR model of [7]. A benchmark comparison of these two models was recently conducted [11], where the SAR model outperformed the principal components using traditional pattern recognition criteria. However, in "playing with the system" we found that the principal components were more useful than the SAR for *navigating* through the database. An example which illustrates this can be seen in Figure 2, where we also show new results of experiments with the Wold-based model.

Figure 2 illustrates the three different methods applied to two different images selected by the user. From top to bottom the methods are the (a) principal components, (b) multiscale SAR, and (c) new Wold-based model. Six displays are shown. The image selected by the user is shown in the upper left corner of each of the displays, and the "next 26 closest" images found by the computer appear in raster-scan order after the selected one.

There are two key performance criteria we consider. The first is quantitative – there are nine samples from the same original Brodatz image in the database, so "perfect" pattern recognition performance implies that all nine patterns are found in the first row. The second criterion is more qualitative – of the other kinds of images found near the selected one, how many of them "look similar" to it? In other words, how many of them might be selected by the user while trying to "navigate" to the image in the upper left corner?

In the left column of Figure 2 (a) the principal components fill the screen with perceptually similar patterns, but do not find all nine brick images as closest. In contrast, the SAR finds all nine brick images, but then fills the screen with images of water which are not useful in navigating toward the bricks. In (c) the new Wold model succeeds not only in finding all nine brick images, but also in filling the screen with perceptually similar images. The Wold ordering of the images can be said to be more perceptual than that of the other two methods.

The second column shows a second case where both the principal components and SAR models fail, while the Wold model perfectly finds all nine most similar textures. In this case, only the structural "most salient" component is needed, so the computation of the Wold parameters is less than that of the SAR. With pre-computation of features, all three models can be used to search the database in real-time on a DEC 5000 workstation.

4 Summary

We present a Wold-based model which decomposes textures into three mutually orthogonal components, corresponding approximately to periodicity, directionality, and randomness. The separate treatment by the Wold decomposition of continuous and discontinuous spectral components avoids the breakdowns usually associated with fitting auto-regressive models to multidimensional data with discontinuous spectral components. A comparison of the new Wold model to auto-regressive and principal components models indicates that the Wold-based parameters may be more relevant for measuring image similarity in a perceptual sense.

5 Acknowledgments

The authors wish to thank J. Francos for many stimulating discussions on the Wold decomposition theory and applications, M. Gorkani for her orientation detection software, and T. P. Minka for his X wizardry.

References

[1] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE T. Sys., Man and Cyber.*, SMC-8(6):460–473, 1978.

[2] T. W. Anderson. *The Statistical Analysis of Time Series.* John Wiley & Sons, 1971.

[3] J. M. Francos. Signal processing and its applications. pages 207–227. North Holland, 1993.

[4] J. M. Francos, A. Zvi Meiri, and B. Porat. A unified texture model based on a 2-D Wold like decomposition. *IEEE T. Sig. Proc.*, pages 2665–2678, August 1993.

[5] J. M. Francos. *Personal Communication.* 1992.

[6] A. R. Rao and J. Lohse. Identifying high level features of texture perception. Computer Science RC17629 #77673, IBM, 1992.

[7] J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Patt. Rec.*, 25(2):173–188, 1992.

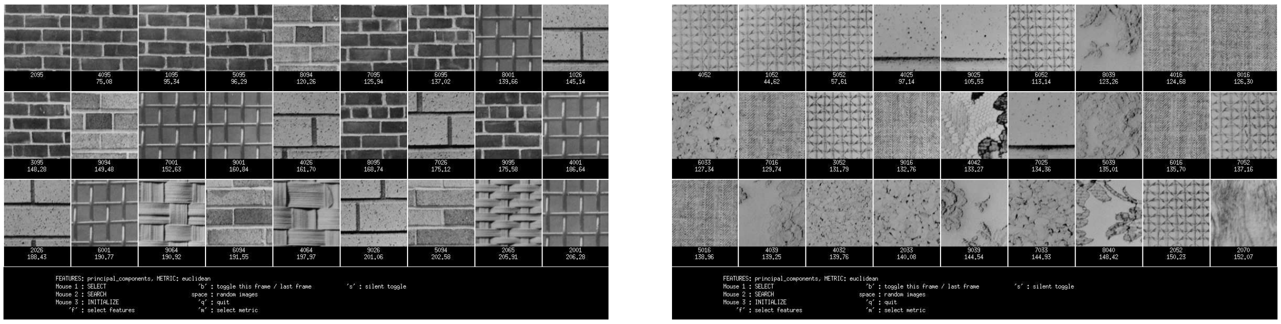
[8] R. W. Picard and M. Gorkani. Finding perceptually dominant orientations in natural textures. *Spatial Vision, special Julesz birthday issue.* To Appear; also available as Percep. Comp. TR #229, MIT Media Laboratory, 1993.

[9] P. Brodatz. *Textures: A Photographic Album for Artists and Designers.* Dover, New York, 1966.

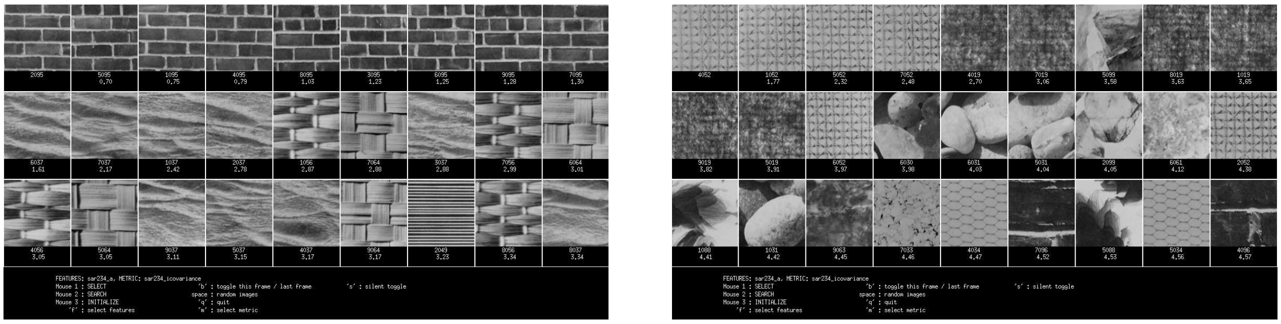
[10] R. W. Picard and T. Kabir. Finding similar patterns in large image databases. In *Proc. ICASSP*, pages V–161–V–164, Minneapolis, MN, 1993.

[11] R. W. Picard, T. Kabir, and F. Liu. Real-Time Recognition with the Entire Brodatz Texture Database. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 638–639, New York, June 1993.

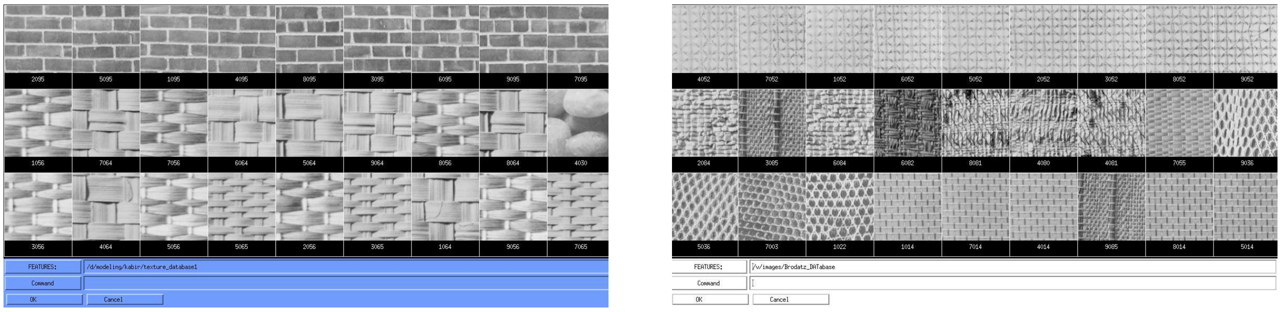
Bricks, Brodatz image D95 | Oriental straw cloth, Brodatz image D52



(a) Shift-invariant principal components



(b) Multiscale simultaneous auto-regressive (SAR)



(c) New model based on Wold decomposition

Figure 2: Comparison of ordering textures from the entire Brodatz database, using three methods: (a) shift-invariant principal components (b) multiscale SAR, and (c) new Wold model. The user selected bricks for the displays shown on the left, and oriental straw cloth for the displays on the right. In each display the images are raster-scan ordered by their similarities to the image in the upper left.