Texture Metrics

Yossi Rubner and Carlo Tomasi
Computer Science Department, Stanford University
Stanford, CA 94305
[rubner,tomasi]@cs.stanford.edu

ABSTRACT

We introduce a class of *metric* perceptual distances between textures. The first metric is sensitive to both rotation and scale differences, and provides a basis for two other metrics, one invariant to rotation, and the other invariant to both rotation and scale. Our metrics are based on Earth Mover's Distance computations on log-polar distributions of spatial frequency computed from Gabor filters. We show consistency of our metrics with psychophysical findings on texture discrimination and classification.

1 INTRODUCTION

Similarity measures between textures are important for image understanding applications such as content-based image retrieval, texture segmentation, and texture classification. In order to be useful, it is important that these similarity measures correspond to human texture perception. In addition, in image retrieval it is often crucial that the similarity distances be *metric*, so that efficient data structures and search algorithms [2, 3] can be used.

In this paper we define a class of texture metrics based on texture features close to the model of simple cells in the primary visual cortex[11]. For the distance between texture feature histograms we use the Earth Mover's Distance (EMD), an effective and efficient measure of histogram differences [18]. We evaluate our metrics both quantitatively, by examining the actual distances between different textures, and qualitatively, by using multidimensional scaling techniques [21] to find what are the texture properties that affect our metrics the most, and to "visualize" the metrics. We obtain similar results to those found by psychophysical experiments [20, 17].

2 PREVIOUS WORK

Many similarity measures for textures exists. They can be divided into measures that are defined in the image domain [15], and measures that are defined in the (local) frequency domain, mostly by using Gabor filters [12, 14, 16]. In principle, most of the methods compute histograms of a predefined feature set. This is done either by taking a single feature set or by gathering statistics over some neighborhood. Then a similarity measure is defined over the histograms. Common similarity measures includes the L_1-norm [14], the L_2-norm [12], and statistical tests [16].

Other methods proposed for applications such as image retrieval and texture segmentation rephrase distances between textures in probabilistic terms. Given a prototype texture, the other textures are ranked by their posterior probabilities of matching the prototype texture [13].

Common problems with most texture similarity measures are that they are usually not metric. The techniques used for histogram matching often lead to false negatives (L_1-norm), or false positives (L_2-norm) [19]. This happens because only corresponding bins in the histograms are compared, ignoring the neighboring bins. Also, some texture similarity measures are not shown to correspond well to visual similarity. Lastly, many measures cannot be extended naturally to handle image transformations such as rotation and scale as required in some applications.

3 TEXTURE REPRESENTATION

Gabor functions [9] are commonly used in texture analysis (e.g. [1, 8, 14]). There is strong evidence that simple cells in the primary visual cortex can be modeled by Gabor functions tuned to detect different orientations and scales on a log-polar grid [7]. When applied to images, these functions produce features which are the basis for many definitions of texture.

In [11] Lee derives the following Gabor wavelets with parameters properly constrained by neurophysiological data on simple cells and by the theory of wavelets. A discretization of a two-dimensional wavelet is given by [6]

$$W_{lmpq} = \tag{1}$$
$$a^m \iint I(x,y)\psi_{l\Delta\theta}\left(a^m(x-p\Delta x), a^m(y-q\Delta x)\right) dxdy \ ,$$

where a is a scale parameter, $\Delta x, \Delta y$ is the spatial sampling rectangle, $\Delta\theta = \pi/L$ is the orientation sampling interval with $l = 0, \ldots, 2L-1$, and $m = 0, \ldots, M-1$ are the scales. p, q are image position. We used $a = 1/2$, $\Delta x = \Delta y = 1$, $M = 5$ scales, and $2L = 16$ orientations.

The notation $\psi_\theta(x,y) = \psi(\tilde{x}, \tilde{y})$ denotes a rotation of the *mother wavelet* $\psi(x,y)$ by θ. We chose the following Gabor function as our mother wavelet [11]:

$$\psi(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{8}(4x^2+y^2)} \left(e^{ikx} - e^{-\frac{k^2}{2}}\right) \ .$$

The constant k determines the frequency bandwidth of the filters. We use $k = 3.14$ which corresponds to a half-amplitude bandwidth of 1 octave, consistently with neurophysiological findings. See [11] for details on Gabor wavelets and the choice of parameters.

Since for this paper our images are homogeneous textures, we take as our texture features the spatial mean of the energies of the Gabor responses:

$$E_{lm} = \frac{\sum_{p,q} |W_{lmpq}|^2}{\sum_{l,m,p,q} |W_{lmpq}|^2} \ .$$

The denominator normalizes E_{lm} so that $\sum_{l,m} E_{lm} = 1$. In general when the image has more than one texture, the energy mean can be computed for smaller regions or at every pixel.

Our choice of the different parameters also makes the Gabor representation close to a tight frame [11]. A tight frame allows for complete reconstruction of an image from its wavelet responses. In our case however, reconstruction is not possible since we keep only the energies of the Gabor responses. Nevertheless, tight frame is a desirable property since it guarantees that the energy of the image is preserved by the E_{lm}'s up to a constant factor [6].

In order to make the frame even tighter, each scale can be furthermore divided into *voices* by:

$$\psi^{\eta}_{l\Delta\theta}(x,y) = 2^{-2\eta/N}\psi_{l\Delta\theta}\left(-2^{\eta/N}x, -2^{\eta/N}y\right) \ ,$$

where N is the number of frequency sampling steps per scale, and $\eta = 0, \ldots, N-1$ is the index. We use $N = 2$ voices per scale.

Exploiting symmetry, computing only L orientations is sufficient, so our texture representation is a 5 by 8 array of energies which we call *texture signature*. Figure 1 shows two examples of textures signatures. [1]

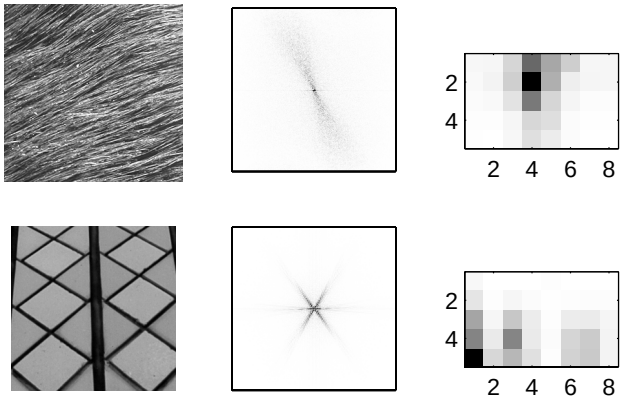


Figure 1: Texture signatures. *Left:* Patches from fabric (top) and tile (bottom) textures. *Middle:* DFT magnitude. *Right*: Texture signatures.

4 THE EARTH MOVER'S DISTANCE

The texture signature derived in section 3 is a two-dimensional histogram. To compare textures we need a measure of similarity between these histograms.

In [18] the concept of the *Earth Mover's Distance* (EMD) is introduced as a flexible similarity measure between multidimensional histograms or, more generally, distributions. Intuitively, given two histograms, one can be seen as a mass of earth properly spread in space, the other as a collection of holes in that same space. Then, the EMD measures the least amount of work needed to fill the holes with earth. Here, a unit of work corresponds to transporting a unit of earth by a unit of distance.

[1] The textures used in this paper are mostly from the MIT Media Lab's VisTex texture collection, and from the Brodatz album.

The EMD is based on the *transportation problem* [5]: Let P be a set of points called *suppliers*, Q a set of points called *consumers*, and c_{ij} the cost to ship a unit of supply from p_i to q_j. In general, P and Q does not have to be of the same size. We want to find a set of flows f_{ij} that minimize the overall cost

$$\sum_{\substack{i\,=\,1\,\ldots\,|P| \\ j\,=\,1\,\ldots\,|Q|}} f_{ij}c_{ij} \ , \tag{2}$$

subject to the following constraints:

$$f_{ij} \geq 0 \quad i = 1\ldots|P|,\ j = 1\ldots|Q| \tag{3}$$

$$\sum_{i=1\ldots|P|} f_{ij} = w_{q_j} \qquad j = 1\ldots|Q| \tag{4}$$

$$\sum_{j=1\ldots|Q|} f_{ij} \leq w_{p_i} \qquad i = 1\ldots|P| \ , \tag{5}$$

where w_{p_i} is the total supply (weight) of supplier i and w_{q_j} is the total capacity (weight) of consumer j. Constraint 3 allows shipping from a supplier to a consumer and not vice versa. Constraint 4 forces the consumers to fill up all of their capacities and constraint 5 limits the supply that a supplier can send to its total amount. A feasibility condition is that the total demand does not exceed the total supply.

The transportation problem is a special case of linear optimization which can be computed efficiently [18]. It can be naturally used for histogram matching by defining one histogram as the supplier and the other as the consumer, and solving the transportation problem where $c_{ij} = d(p_i, q_j)$ is the *ground distance* which should be chosen according to the problem at hand. When the total weights of the histograms are not equal (partial matches), the smaller histogram will be the consumer in order to satisfy the feasibility condition. Once the transportation problem is solved, and we have found the optimal flow F, the earth mover's distance is defined as

$$\mathrm{EMD}(\mathbf{x},\mathbf{y}) = \frac{\sum_{i,j} f_{ij}d(p_i,q_j)}{\sum_{i,j} f_{ij}}$$

where the denominator is a normalization factor that avoids favoring histograms with smaller total weights. For texture signatures however, by construction $\sum_{i,j} f_{ij} = 1$.

Thus, the EMD naturally extends distance between single elements to distance between sets of elements, or histograms, where items from neighboring bins contribute similar costs, thereby eliminating histogram binning artifacts. If the ground distance is a metric and the total weights of two signatures are equal, the EMD is a true metric [18].

5 TEXTURE METRICS

The notion of distance between textures varies with athe task at hand. While for texture classification one may want the distance to be invariant to rotations of the texture and perhaps also to changes in scale, for texture segmentation these invariants may be inappropriate. In this section we show how to use the EMD to define different distances between textures. In section 5.2 we define a distance with no invariance. In section 5.3 we define a rotation invariant distance and in section 5.4 we add also scale invariance.

In order to evaluate the meaningfulness of our new texture metrics, we use *MultiDimensional Scaling* (MDS) [21] to embed the textures in the Euclidean plane so that distances in the plane are as close as possible to the EMDs between them. Such embeddings allow us to see all distances at once, albeit approximately, and to evaluate whether the EMD is a perceptually natural metric. The MDS is introduced in the next section.

5.1 Multidimensional Scaling as a Perceptual Evaluation Tool

Given a set of n objects together with the distances δ_{ij} between them, the Multi-Dimensional Scaling (MDS) technique [21] computes a configuration of points $\{p_i\}$ in a low-dimensional Euclidean space $\mathbf{R}^d$, (in our experiments we use $d = 2$) so that the Euclidean distances $d_{ij} = \|p_i - p_j\|_2$ between the points in $\mathbf{R}^d$ match as well as possible the original distances δ_{ij} between the corresponding objects. Kruskal's [10] formulation of this problem requires minimizing the following quantity

$$\text{STRESS} = \left[\frac{\sum_{i,j} (d_{ij} - \delta_{ij})^2}{\sum_{i,j} \delta_{ij}^2} \right]^{1/2}$$

with the additional constraint that the d_{ij}s are in the same rank ordering as the corresponding δ_{ij}s. STRESS is a nonnegative number that indicates how well distances are preserved in the embedding. Zero STRESS indicates a perfect fit. Rigid transformations and reflections can be applied to the MDS result without changing the STRESS.

As we show in sections 5.2, 5.3 and 5.4, performing MDS on a set of texture patches using different metrics automatically reveals the properties of texture that predominantly affect our distance, without the need for an explicit definition of the properties themselves. However, it is important to remember that since the textures do not "live" in two-dimensional space (or in any Euclidean space), the MDS gives only an approximation to the distance matrix. While the MDS gives the general "feel" for the main discriminative factors of a metric, it should not be used for exact measures of distance between pairs of points.

5.2 No Invariance

In section 3 we represented a texture by its energy samples on a log-polar frequency grid. We can now use the EMD as the distance measure between these distribution samples. To this end, we define our ground distance to be the L_1-distance in the log-polar space. We obtained similar results with the Euclidean distance. Since the log-polar space is actually cylindrical, we have two possible distances between a pair of points. We define the ground distance to be the shorter of the two distances. The ground distance between the points (l_1, m_1) and (l_2, m_2) is therefore:

$$d\left((l_1, m_1), (l_2, m_2)\right) = |\Delta l| + \alpha |\Delta m| \,, \qquad (6)$$

where

$$\Delta l = \min(|l_1 - l_2|,\ L - |l_1 - l_2|) \quad , \quad \Delta m = m_1 - m_2 \,.$$

The parameter α controls the relative importance of scale and orientation. We used $\alpha = 1$ with good results. Other choices of α can result from the applications or from psychophysical experiments. Our ground distance is metric (see proof in Appendix A) and therefore the defined EMD is metric as well [18].

We performed a similar experiment as in [14] on a small texture database of 62 images of texture of size 256×256 pixels. Sets of textures which we perceptually considered as the same contributed one representative to the database. Each image was divided into 4 non-overlapping 128×128 subimages. Each of the resulting 248 subimages was used as a query, asking for the top 4 matches. The average percentage of number of matches which belongs to the same images as the query, was 89.6%. When the top 8 matches where returned, the performance increased to 96.8%.

To understand the metric qualitatively, we picked the 16 textures shown above Table 1, which shows their (symmetric) EMD distance matrix. Figure 2 shows the results of applying a two-dimensional MDS to 16 textures using the described EMD.[2] In this figure, the arrows we superimposed by hand suggest that one axis reflects the coarseness of the texture, from fine to coarse. The other (curved) axis reflects the dominant orientation of the texture. On the left we see horizontal textures, on the right vertical textures, and as we move from left to right on the lower half circle, the orientation changes counter-clockwise. The other textures have no dominant orientation. STRESS in this figure is 0.061, so on average, distances in the picture are close to the real EMDs. Adding more oriented textures with different scales and orientations will complete also the top half of the orientation circle, at the expense of the coarseness axis. This will increase the STRESS, making two dimensions insufficient. In this case, a three-dimensional MDS would use two dimensions for the orientation circle and the third for coarseness.

5.3 Rotation Invariance

In our log-polar array, rotation translates to cyclic shifts along the orientation axis. Although inefficient, we can achieve rotation invariance by an exhaustive search for the minimal distance over all possible shifts in orientation. In section 6 we mention an efficient algorithm which avoids this exhaustive search.

Let $\mathbf{t}_1$ and $\mathbf{t}_2$ be two texture signatures. An EMD that is invariant to texture rotation is

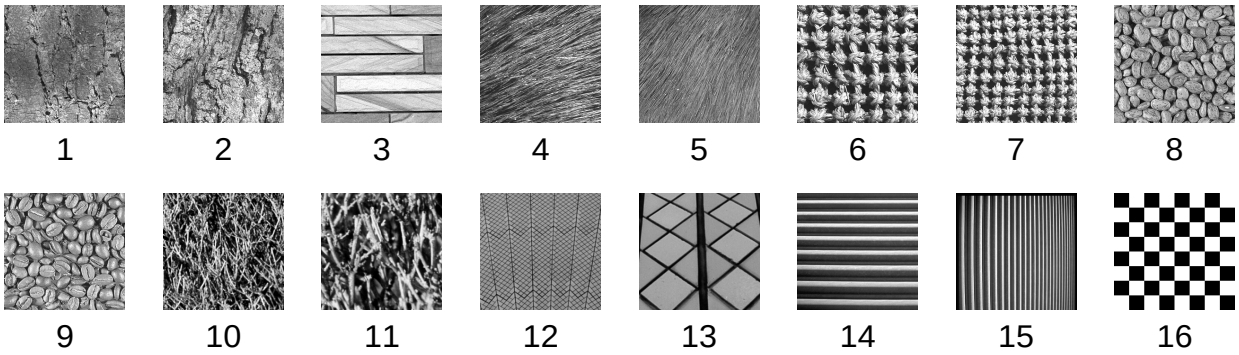
$$\text{EMD}(\mathbf{t}_1, \mathbf{t}_2) = \min_{l_s = 0, \ldots, L-1} \text{EMD}(\mathbf{t}_1, \mathbf{t}_2, l_s) \,,$$

where $\text{EMD}(\mathbf{t}_1, \mathbf{t}_2, l_s)$ is the EMD with orientation shift l_s. The ground distance in equation (6) uses the same Δm but

$$\begin{aligned} \Delta l \quad = \quad & \min(\ |l_1 - l_2 + l_s \pmod L|, \\ & \quad L - |l_1 - l_2 + l_s \pmod L|) \,. \end{aligned}$$

A two-dimensional MDS using the rotation-invariant EMD on the texture signatures in shown in figure 3 (with a STRESS value of 0.044). One axis emphasizes the directionality of the texture, where textures with one dominant orientation (any orientation) are at the top, and textures without a dominant orientation (no orientation at all, or more than one orientation) are at the bottom. The other axis is coarseness, similarly to the previous experiment. For example, the two oriented textures of fabrics on the right are close together although they have different orientations.

[2] In order to better see the fine details of the textures, only one quarter of the textures that was used to compute the texture signature are displayed.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00	0.29	2.07	1.46	1.26	1.25	1.11	0.71	0.72	0.74	1.42	0.72	1.56	2.86	1.78	1.65
2	0.29	0.00	2.13	1.50	1.24	1.11	1.01	0.55	0.54	0.57	1.21	0.77	1.38	2.89	1.62	1.54
3	2.07	2.13	0.00	1.63	2.72	1.85	1.73	1.90	1.92	2.33	2.32	2.14	1.86	0.90	3.49	1.69
4	1.46	1.50	1.63	0.00	1.42	2.10	1.92	1.77	1.73	1.98	2.40	1.29	2.13	2.45	2.76	2.23
5	1.26	1.24	2.72	1.42	0.00	2.13	1.98	1.64	1.58	1.52	2.12	1.22	2.31	3.48	1.64	2.47
6	1.25	1.11	1.85	2.10	2.13	0.00	0.26	0.70	0.72	1.10	0.87	1.64	0.83	2.35	1.98	1.16
7	1.11	1.01	1.73	1.92	1.98	0.26	0.00	0.60	0.63	1.04	0.94	1.49	0.91	2.30	1.96	1.30
8	0.71	0.55	1.90	1.77	1.64	0.70	0.60	0.00	0.11	0.51	0.78	1.04	0.91	2.53	1.64	1.12
9	0.72	0.54	1.92	1.73	1.58	0.72	0.63	0.11	0.00	0.50	0.79	1.00	0.89	2.55	1.62	1.12
10	0.74	0.57	2.33	1.98	1.52	1.10	1.04	0.51	0.50	0.00	0.78	0.98	1.17	2.96	1.20	1.37
11	1.42	1.21	2.32	2.40	2.12	0.87	0.94	0.78	0.79	0.78	0.00	1.63	0.70	2.75	1.47	1.04
12	0.72	0.77	2.14	1.29	1.22	1.64	1.49	1.04	1.00	0.98	1.63	0.00	1.67	2.88	1.84	1.78
13	1.56	1.38	1.86	2.13	2.31	0.83	0.91	0.91	0.89	1.17	0.70	1.67	0.00	2.24	2.03	0.82
14	2.86	2.89	0.90	2.45	3.48	2.35	2.30	2.53	2.55	2.96	2.75	2.88	2.24	0.00	4.07	2.02
15	1.78	1.62	3.49	2.76	1.64	1.98	1.96	1.64	1.62	1.20	1.47	1.84	2.03	4.07	0.00	2.24
16	1.65	1.54	1.69	2.23	2.47	1.16	1.30	1.12	1.12	1.37	1.04	1.78	0.82	2.02	2.24	0.00

Table 1: *Top:* The 16 textures used in this paper. *Bottom:* Distance matrix for the no invariance case.

Coarseness and directionality were found by psychophysical experiments by Tamura et al. [20] to be the two most discriminant texture properties for human perception.

5.4 Scale Invariance

Scale invariance can be obtained in a similar manner. In the log-polar array, scale invariance can be seen as invariance to shifts in the scale axis. An EMD that is invariant to both rotation and scale is

$$\mathrm{EMD}(\mathbf{t}_1, \mathbf{t}_2) = \min_{\substack{l_s = 0, \ldots, L-1 \\ m_s = -(M-1), \ldots, M-1}} \mathrm{EMD}(\mathbf{t}_1, \mathbf{t}_2, l_s, m_s) \ ,$$

where $\mathrm{EMD}(\mathbf{t}_1, \mathbf{t}_2, l_s, m_s)$ is the EMD with orientation shift l_s and scale shift m_s. The ground distance is similar to the scale invariance case with

$$\Delta m = m_1 - m_2 + m_s \ .$$

Now the 2D MDS, shown in figure 4 (with STRESS equal to 0.074), can be interpreted as follows. One axis is again the directionality, while the other shows what we call the "scality" of the texture, a measure that distinguishes between textures with one dominant scale and textures with more than one, or no dominant scale. For example, the two textures of oriented bars which have different orientations and scales are close to each other. Also, the two textures of tiles on the right are very close to each other even if they differ by more than three octaves in scale!

Although not the same, this measure is correlated with the regularity [17] of the textures. While the intuitive interpretation of this second axis deserves further investigation, we point out that our results, based on EMD and MDS, are close to those of Rao and Lohse [17] who used psychophysical experiments instead, and concluded that regularity and directionality are perceptually the two most important properties for texture classification.

6 COMPUTATION

A single EMD can be computed quite efficiently [18]. In addition, since the two signatures do not have to be of the same size, the histograms can be reduced by pruning bins with very small weights which contribute little to the EMD result. This significantly reduces the size of the EMD problem which can be computed much faster. In practice, for most textures most of the bins can be pruned with negligible change to the EMD result.

The invariant texture distances in sections 5.3 and 5.4 can be computed without an exhaustive search by using an EM type optimization algorithm [4].

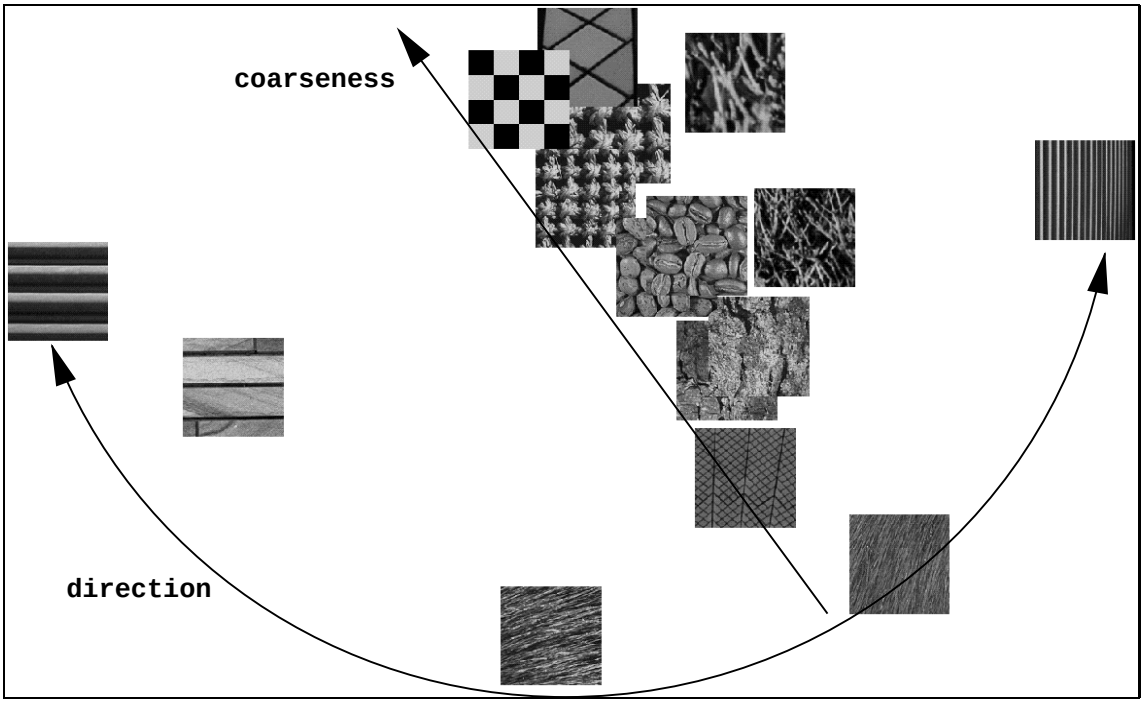


Figure 2: 2D MDS of 16 textures. The two dominant axes of orientation and coarseness emerge.

7 CONCLUSIONS

The texture distances introduced in this paper enjoy two crucial properties for image retrieval and other texture discrimination and classification tasks: they are perceputally natural and they are metric. In addition, they can be computed very fast. Presumably, our metrics can also improve the performance of texture segmentation algorithms, thereby allowing application of our metrics to entire, realistic images, rather than to patches of homogeneous texture. This is a major thread of our current research.

ACKNOWLEDGMENTS

We thank Leonidas Guibas and Scott Cohen for helpful discussions about the EMD, and Ron Kimmel for the Gabor filter code.

REFERENCES

[1] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12(12):55–73, 1990.

[2] T. Bozkaya and M. Ozsoyoglu. Distance-based indexing for high-dimensional metric spaces. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 26(2):357–368, May 1997.

[3] Kenneth L. Clarkson. Nearest neighbor queries in metric spaces. In *ACM Symposium on Theory of Computing*, pages 609–617, 1997.

[4] S. D. Cohen and L. J. Guibas. The earth mover's distance: Lower bounds and invariance under translation. Technical Report STAN-CS-TR-97-1597, Stanford University, November 1997.

[5] G. B. Dantzig. Application of the simplex method to a transportation problem. In *Activity Analysis of Production and Allocation*, pages 359–373, New York, NY, 1951. John Wiley and Sons.

[6] I. Daubechies. *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[7] J. D. Daugman. Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36:1169–1179, 1988.

[8] F. Farrokhnia and A. K. Jain. A multi-channel filtering approach to texture segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 364–370, June 1991.

[9] D. Gabor. Theory of communication. *The Journal of the Institute of Electrical Engineers, Part III*, 93(21):429–457, January 1946.

[10] J. B. Kruskal. Multi-dimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27, 1964.

[11] T. S. Lee. Image representation using 2d Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):959–971, October 1996.

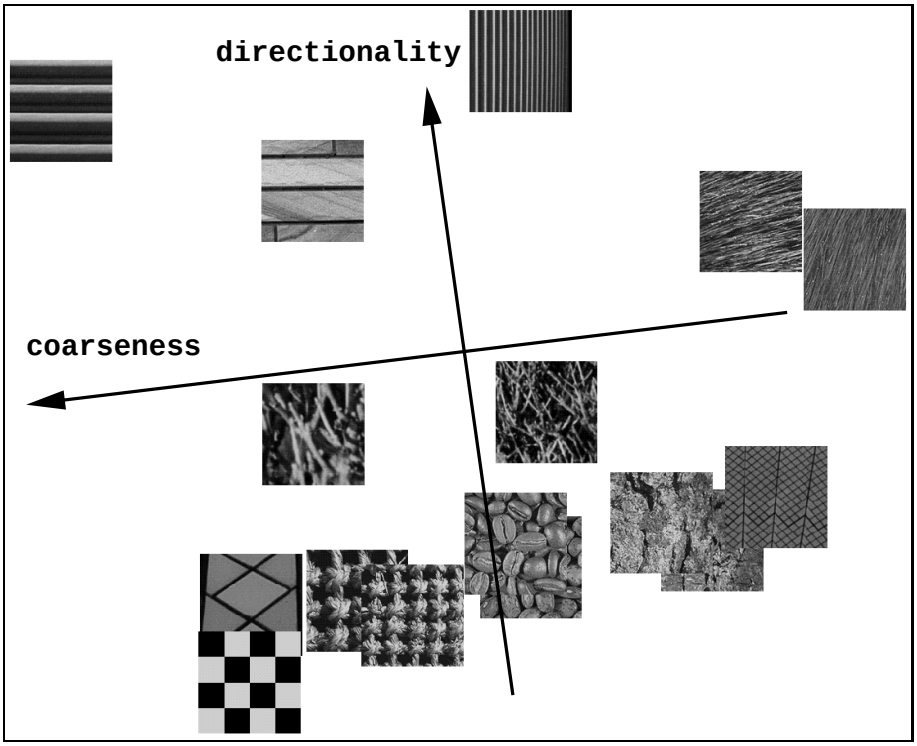


Figure 3: 2D MDS of 16 textures with rotation invariance. Coarseness vs. directionality.

[12] T. S. Lee, D. Mumford, and A. Yuille. Texture segmentation by minimizing vector–valued energy functionals: The coupled–membrane model. In *Proceedings of Computer Vision (ECCV '92)*, volume 588 of *LNCS*, pages 165–173, Berlin, Germany, may 1992. Springer.

[13] F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):722–733, July 1996.

[14] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[15] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29:51–59, 1996.

[16] J. Puzicha, T. Hofmann, and J. M. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 1997.

[17] A. Ravishankar Rao and G. L. Lohse. Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing*, 55:218–233, May 1991.

[18] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *IEEE International Conference on Computer Vision*, pages 59–66, Bombay, India, January 1998.

[19] M. Stricker and M. Orengo. Similarity of color images. In *SPIE Conference on Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 381–392, February 1995.

[20] H. Tamura, T. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8:460–473, June 1978.

[21] W. S. Torgerson. *Theory and Methods of Scaling*. John Wiley and Sons, New York, NY, 1958.

A PROVING THAT THE DISTANCES ARE METRIC

Here we prove that all distances defined in this paper are indeed metric. Nonnegativity and symmetry hold trivially in all cases, so we only need to prove that the triangle inequality holds. In [18] it is proven that if the ground distance is metric, the EMD is metric. First, since our log-polar space is cylindrical, we need to show that the ground distance (L_1 in our case) on a cylinder is metric as well.

The projection of two points A and C on the cylinder onto its circular base divides it into two circular arcs (in the degenerate case one arc can be a point). The projection of a third point B can be in only one of these two arcs. Now the cylinder can be unfolded into a rectangle by a vertical cut anywhere through the other arc. Since in this two-dimensional plane $\overline{AC} \leq \overline{AB} + \overline{BC}$, the triangle inequality holds also on the cylinder (the other possible distance from A to C through the cut is chosen only if it makes $\overline{AC}$ shorter). Therefore, our ground distance is metric.

We now prove that the EMD is metric in the rotation and scale

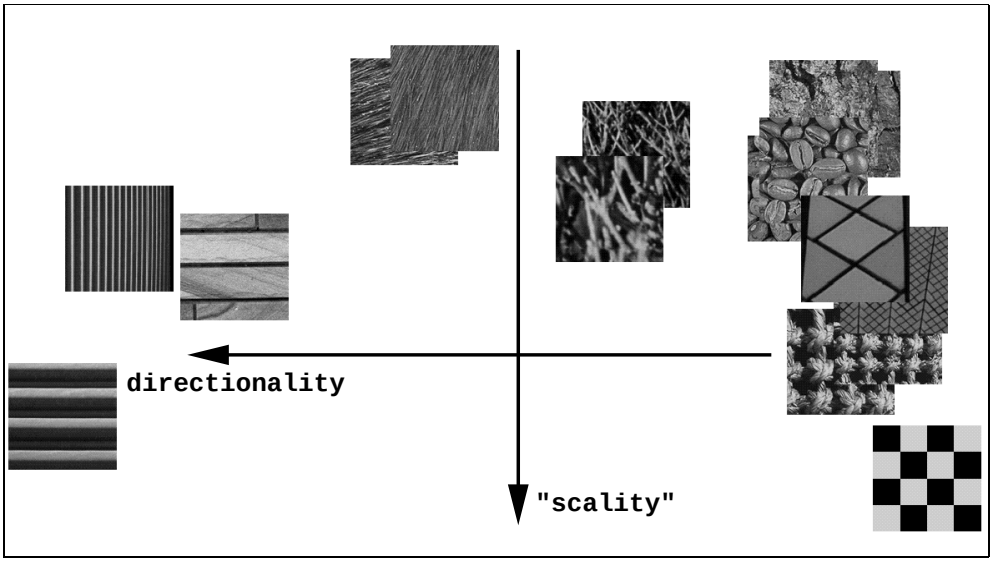


Figure 4: 2D MDS of 16 textures with rotation and scale invariance. "Scality" vs. directionality.

invariant cases. In the log-polar space rotation and scale shifts are reduced to translations, and we now prove in general that EMD's under translations are metric.

Translation invariant EMD between signatures P and Q can be written as

$$\widetilde{EMD}(P,Q) = \sum_{i,j} f_{ij}\, d(p_i, q_j - T_{PQ}) \ ,$$

where f_{ij} and T_{PQ} are the flow and translation that minimize the sum. Similar formulas hold for $\widetilde{EMD}(Q,R)$ and $\widetilde{EMD}(P,R)$, and we need to prove that

$$\widetilde{EMD}(P,R) \leq \widetilde{EMD}(P,Q) + \widetilde{EMD}(Q,R) \ .$$

Without loss of generality we assume here that the total sum of the flows is 1. Consider the flow $P \mapsto Q \mapsto R$. The largest unit of weight that moves together from P to Q and from Q to R defines a flow which we call b_{ijk} where i, j and k correspond to p_i, q_j and r_k respectively. Clearly $\sum_k b_{ijk} = f_{ij}$ and $\sum_i b_{ijk} = g_{ij}$. We define

$$h_{ik} \triangleq \sum_j b_{ijk}$$

which is a legal flow from i to k since

$$\sum_i h_{ik} = \sum_{i,j} b_{ijk} = \sum_j g_{jk} = w_{r_k} \ ,$$

and

$$\sum_k h_{ik} = \sum_{j,k} b_{ijk} = \sum_j f_{ij} = w_{p_i} \ .$$

Since $\widetilde{EMD}(P,R)$ is the minimal flow from P to R, and h_{ik} is some legal flow from P to R,

$$\begin{aligned}
\widetilde{EMD}(P,R) &\leq \sum_{i,k} h_{ik}\, d\left(p_i, r_k - (T_{PQ} + T_{QR})\right) \\
&= \sum_{i,j,k} b_{ijk}\, d\left(p_i, r_k - (T_{PQ} + T_{QR})\right) \\
&\leq \sum_{i,j,k} b_{ijk}\, d(p_i, q_j - T_{PQ}) + \\
&\qquad \sum_{i,j,k} b_{ijk}\, d\left(q_j - T_{PQ}, r_k - (T_{PQ} + T_{QR})\right) \\
&= \sum_{i,j,k} b_{ijk}\, d(p_i, q_j - T_{PQ}) + \sum_{i,j,k} b_{ijk}\, d(q_j, r_k - T_{QR}) \\
&= \sum_{i,j} f_{ij} d(p_i, q_j - T_{PQ}) + \sum_{j,k} g_{jk}\, d(q_j, r_k - T_{QR}) \\
&= \widetilde{EMD}(P,Q) + \widetilde{EMD}(Q,R) \ .
\end{aligned}$$