Proceedings of the 1998 IEEE International Conference
on Computer Vision, Bombay, India

A Metric for Distributions with Applications to Image Databases *

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

Abstract

We introduce a new distance between two distributions that we call the Earth Mover's Distance *(EMD), which reflects the minimal amount of work that must be performed to transform one distribution into the other by moving "distribution mass" around. This is a special case of the transportation problem from linear optimization, for which efficient algorithms are available. The EMD also allows for partial matching. When used to compare distributions that have the same overall mass, the EMD is a true metric, and has easy-to-compute lower bounds. In this paper we focus on applications to image databases, especially color and texture. We use the EMD to exhibit the structure of color-distribution and texture spaces by means of Multi-Dimensional Scaling displays. We also propose a novel approach to the problem of navigating through a collection of color images, which leads to a new paradigm for image database search.*

1 Introduction

Feature distributions are often used in computer vision to summarize the content of an image. Examples in image retrieval are the one-dimensional distribution of image intensities and the three-dimensional distribution of image colors. At another level, a single texture can itself be considered a distribution. In fact, texture can be envisioned as a distribution of signal energy over the domain of spatial frequencies. Consequently, in image retrieval, as well as in the study of texture discrimination and color perception, it becomes important to define a distance between two distributions. This requires in turn a notion of distance between the basic elements that appear in the distribution. For instance, in the case of color, a metric is needed for individual colors. For texture, one needs a measure of the dissimilarity between two periodic signals, that is, between two points in the frequency domain. Fortunately, these *ground distances* have been studied in psychophysics, and have led to the CIE-Lab space [19] for color perception and to quantitative models [17], [6] of spatial frequency discrimination.

In this paper, we define a consistent measure of distance, or dissimilarity, between two distributions of points in a space for which a ground distance is given. This new measure is itself a metric. For color, supplying a ground distance means defining distances between color distributions. For spatial frequency, it means defining distances between textures. Mathematically, it would be convenient if these distribution distances were true metrics. Practically, it is crucial that they correlate with human perception. In this paper we strive to achieve both goals. For the first we have proof, for the second we show experiments and some ties with psychophysical findings for the examples of color and texture. For image retrieval it is important that these distances allow for partial matches in cases when one distribution is similar to a subset of the other. For partial matches, our distances are not metric. This is consistent with Tversky [16] who shows that perceptual distances can be non-metric.

An image yields a distribution in color space by mapping each pixel of the image to its color. It is often advantageous to 'compress' or otherwise approximate the original distribution by another distribution with a more compact description. This yields important savings in storage and processing time, as well as a certain perceptual robustness to the result. Now multidimensional distributions are usually compressed by partitioning the underlying space into fixed-size bins: the resulting quantized structure is a histogram. However, often only a small fraction of the bins in a histogram contain significant information. For instance, a picture of a desert landscape contains mostly blue pixels in the sky region and yellow-brown pixels in the rest. A finely quantized histogram in this case is highly inefficient. On the other hand, multitude of colors is a characterizing feature for a picture of a carnival in Rio, and a coarsely quantized histogram would be inadequate. In brief, because histograms are fixed-size structures, they cannot achieve balance between expressiveness and efficiency.

*Research supported by DARPA grant DAAH04-94-G-0284 and by a grant from the Charles Lee Powell foundation.

In contrast, we propose *variable-size descriptions* of distributions, which we call *signatures*. Only the dominant clusters are extracted from the original distribution and are used to form its signature. Each cluster is represented by a single point in the underlying space (the cluster center), together with a weight that denotes the size of that cluster. Simple images have short signatures, complex images have long ones. Note that signatures are themselves distributions. Of course, in some applications, fixed-size histograms may still be adequate, and can be considered as special cases of signatures.

We introduce a new distance between two signatures that we call the *Earth Mover's Distance* (EMD). This reflects the minimal cost that must be paid to transform one signature into the other, in a sense that will be made precise in section 3. The EMD is based on a solution to a special case of the old *transportation problem* [1] from linear optimization, for which efficient algorithms are available. The EMD has many desirable properties. It is more robust in comparison to other histogram matching techniques, in that it suffers from no arbitrary quantization problems due to the fixed binning of the latter. It allows for partial matching, and it can be applied to signatures with different sizes. When used to compare distributions that have the same overall mass, the EMD is a true metric.

Although the EMD is a general method for matching multidimensional distributions, in this paper we focus on applications to color and texture. In the next section, we introduce histograms and signatures, and briefly survey some of the existing measures of dissimilarity. Then, in section 3, we introduce the Earth Mover's Distance (EMD), which we apply to color and texture in section 4. In particular, we use the tool of Multi-Dimensional Scaling (MDS, [5]) to display collections of images in a two-dimensional space in a way that preserves EMDs as much as possible. Section 5 addresses computational issues, and section 6 concludes with a summary and plans for future work.

2 Histograms vs. Signatures

A *histogram* $\{h_{\mathbf{i}}\}$ is a mapping from a set of d-dimensional integer vectors $\mathbf{i}$ to the set of nonnegative integers. These vectors typically represent bins (or their centers) in a fixed-size partitioning of the relevant region of the underlying space, and the associated integers are a measure of the mass of the distribution that falls into the corresponding bin. For instance, in a grey-level histogram, d is equal to one, the set of possible grey values is split into N intervals, and $h_{\mathbf{i}}$ is the number of pixels in an image that have a grey value in the interval indexed by $\mathbf{i}$ (a scalar in this case).

A *signature* $\{\mathbf{s}_j = (\mathbf{m}_j, w_j)\}$, on the other hand, represents a set of clusters. Each cluster is represented by its d-dimensional mean (or mode) $\mathbf{m}_j$, and by the number w_j of pixels that belong to that cluster. The integer subscript j ranges from one to a value that varies with the signature. As a special case, a histogram $\{h_{\mathbf{i}}\}$ can be viewed as a signature $\{\mathbf{s}_j = (\mathbf{m}_j, w_j)\}$ in which the vectors $\mathbf{i}$ index a set of clusters defined by a fixed *a priori* partitioning of the underlying space. If vector $\mathbf{i}$ maps to cluster j, the point $\mathbf{m}_j$ is the central value in bin $\mathbf{i}$ of the histogram, and w_j is equal to $h_{\mathbf{i}}$.

Several distance measures have been defined for histograms. If $H = \{h_{\mathbf{i}}\}$ and $K = \{k_{\mathbf{i}}\}$ are two histograms, the L_1-distance is used for color [14] as

$$d(H, K) = \sum_i |h_{\mathbf{i}} - k_{\mathbf{i}}| \ . \tag{1}$$

Related distances have been defined by replacing the L_1 metric with L_2 or L_∞ [13].

A weighted version of the L_2 norm was used, also for color, in [8]:

$$d^2(H, K) = (\mathbf{h} - \mathbf{k})^t \mathbf{A} (\mathbf{h} - \mathbf{k}) \ , \tag{2}$$

where $\mathbf{h}$ and $\mathbf{k}$ are vectors that list all the entries in H and K, and where $\mathbf{A}$ is a matrix of the ground dissimilarities for every pair of color bins.

Both distances (1) and (2) transform a *ground distance* between the constituent elements (colors) into a distance between *distributions* of those elements. For (1), the ground distance between two colors is zero if they fall in the same bin, and two otherwise, and this ground distance is made into a distribution distance by simple addition. For (2), the ground distances are the entries of $\mathbf{A}$, derived from psychophysics in [8], and the corresponding distribution distance is their quadratic mean (assuming proper normalization).

The metric (1) overestimates distances because neighboring bins are not considered when there is no match between the exact corresponding bins in the two histograms [13]. The metric (2) underestimates distances because it tends to accentuate the similarity of color distributions without a pronounced mode [13].

Cumulative histograms were proposed [13] to overcome these limitations:

$$\hat{h}_{\mathbf{i}} = \sum_{\mathbf{j} \leq \mathbf{i}} h_{\mathbf{j}} \ .$$

Although measuring distances between cumulative histograms cures the problems mentioned above in one dimension, the relation $\mathbf{j} \leq \mathbf{i}$ is not a total ordering in more dimensions, and the resulting arbitrariness is likely to cause problems. Another method proposed by [13] is to compute the distance between distributions as the sum of the weighted distance of the distributions' first three moments. However, it is unclear how to tune the weights of the different moments, and the resulting measure is not a metric distance.

In [11], the histogram unfolding method is introduced for grey-level images and is extended to more dimensions in [18]. An "unfolded histogram" is simply the image itself reshaped to a vector and with its pixels sorted in increasing order of value. The distance between two unfolded histograms is then defined as the L_1 norm of their vector difference. However, the computational complexity of this method is very high, because unfolded histograms are as large as the original pictures they come from. Even more importantly, unfolded histograms, as the other methods described above, cannot be used for partial matching, an essential requirement for image retrieval.

3 The Earth Mover's Distance

In this section we propose the *Earth Mover's Distance* (EMD) between distributions in order to address the difficulties discussed above. Intuitively, given two distributions, one can be seen as a mass of earth properly spread in space, the other as a collection of holes in that same space. We can always assume that there is at least as much earth as needed to fill all the holes to capacity by switching what we call earth and what we call holes if necessary. 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 (ground) distance.

Computing the EMD is based on a solution to the old *transportation problem* [1]. This is a bipartite network flow problem which can be formalized as the following linear programming problem: Let $\mathcal{I}$ be a set of suppliers, $\mathcal{J}$ a set of consumers, and c_{ij} the cost to ship a unit of supply from $i \in \mathcal{I}$ to $j \in \mathcal{J}$. Figure 1 shows an example with three suppliers and two consumers. We want to find a set of flows f_{ij} that minimize the overall cost

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} c_{ij} f_{ij} \ , \tag{3}$$

subject to the following constraints:

$$f_{ij} \geq 0 \qquad i \in \mathcal{I},\ j \in \mathcal{J} \tag{4}$$

$$\sum_{i \in \mathcal{I}} f_{ij} = y_j \qquad j \in \mathcal{J} \tag{5}$$

$$\sum_{j \in \mathcal{J}} f_{ij} \leq x_i \qquad i \in \mathcal{I} \ , \tag{6}$$

where x_i is the total supply of supplier i and y_j is the total capacity of consumer j. Constraint 4 allows shipping of supplies from a supplier to a consumer and not vice versa. Constraint 5 forces the consumers to fill up all of their capacities and constraint 6 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

$$\sum_{j \in \mathcal{J}} y_j \leq \sum_{i \in \mathcal{I}} x_i \ .$$

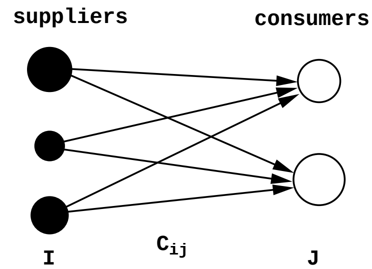

Figure 1: An example of a transportation problem with three suppliers and two consumers.

The transportation problem can be naturally used for signature matching by defining one signature as the supplier and the other as the consumer, and solving the transportation problem where the cost c_{ij} is the ground distance between element i in the first signature and element j in the second. When the total weights of the signatures are not equal (partial matches), the smaller signature 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

$$\text{EMD}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} c_{ij} f_{ij}}{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} f_{ij}} = \frac{\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} c_{ij} f_{ij}}{\sum_{j \in \mathcal{J}} y_j}$$

where the denominator is a normalization factor that avoids favoring signatures with smaller total weights. In general, the ground distance c_{ij} can be any distance and it will be chosen according to the problem at hand. Examples are given in section 4.

Thus, the EMD naturally extends the notion of distance between single elements to distance between sets of elements, or distributions. The advantages of the EMD over previous definitions of distribution distances should now be apparent. First, the EMD applies to signatures, which subsume histograms. The greater compactness and flexibility of signatures is in itself an advantage, and having a distance measure that can handle these variable-size structures is important. Second, the costs of moving "earth" reflect the notion of nearness properly, without the quantization problems of most current measures. Even for histograms, in fact, items from neighboring bins contribute similar costs. Third, the EMD allows for partial matches in a natural way. This is important in order to deal with occlusions and clutter in image retrieval. Fourth, if the ground distance is a metric and the total weights of two signatures are equal, the EMD is a true metric. Computational advantages are discussed in section 5.

4 Applications to Image Databases

In this section we show a few examples of application of the earth mover's distance in the areas of color and texture analysis. Because of how the human vision system is built, color lives naturally in a three dimensional space. Color distributions, then, can describe the color contents of entire images.

Textures, on the other hand can be seen as mixtures of sinusoidal signals. If phase information is ignored, a single texture can be seen as a distribution of signal energy in the frequency domain. Thus, we compute EMDs between single textures, rather than between images with many textures. This may be expanded to EMDs between images with many textures by considering the EMD between single textures as the ground distance in a space of textures. We do not pursue this idea in this paper.

In order to evaluate the meaningfulness of our new metric when applied to color and texture, we use *Multidimensional Scaling* (MDS) [12, 5] in order to embed the images in a two-dimensional Euclidean space so that distances in the embedding are as close as possible to the true EMDs between the images.

The MDS is introduced in the next section and examples for color are given in section 4.2. MDS layouts also lead to a new method for navigating in a database of images, as shown in section 4.3. Section 4.4 applies both EMD and MDS to textures.

4.1 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 [12, 5] 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

$$d_{ij} = ||p_i - p_j||$$

between the points in $\mathbf{R}^d$ match as well as possible the original distances δ_{ij} between the corresponding objects. Kruskal's [5] 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. Embedding methods such as SVD and principal components analysis are not appropriate here because our signatures do not form a linear space, and we do not have points in any space, but only non-Euclidean distances between points.

As we show in sections 4.2 and 4.4, performing MDS on a set of images using the proper EMD automatically reveals important perceptual features of color and texture without the need for an explicit definition of the features themselves.

4.2 Color Distributions

For the computation of the earth mover's distance between color images, we use Euclidean distance in the CIE-Lab color space [19] as the underlying ground distance between individual colors. In this color space, short Euclidean distances correlate strongly with human color discrimination performance [19].

In order to produce figure 2 (in the color plates), colors in each of 1,000 color images from a rather diverse set were first clustered with an algorithm based on k-d-trees [9]. As a result, the color distribution of each image was summarized by a handful of clusters, eight on the average. Clusters were collected into signatures, as described in section 2, and the $\binom{1000}{2}$ earth mover's distances between every pair of signatures were computed. The corresponding distances for a two-dimensional embedding were then computed by the MDS algorithm (section 4.1), with a STRESS value of 0.146, to produce figure 2. Loosely speaking, images end up being arranged according to their dominant lightness and chroma values, as suggested by the arrows in figure 2, but a trend from more saturated colors at the periphery of the display to less saturated and more complex color distributions close to the center can also be noticed. Images with similar color distributions are near one another, while dissimilar images are far apart.

When images are closely related to each other, the MDS display exposes more subtle similarities and differences between images. This can be seen in figure 4 (c) (in the color plates), which displays 20 images with large expanses of blue. Here the interpretation of the axes of the MDS map (STRESS = 0.2) is different. One axis distinguishes between images with or without plants (green vs. yellow/brown), and the other axis corresponds to the amount of saturation of the (blue) sky.

4.3 Navigation

In image retrieval, we can use the MDS of the image thumbnails as a new display technique. This display makes it easy for the user to grasp a large set of returned images at a glance, and decide where to go next. When the user clicks the mouse on a location of the displayed MDS, a new, more specific query is automatically generated, and returns a smaller set of images. These are again displayed by a new MDS, which now reflects the new dominant axes of variation. Thus, the embeddings are *adaptive*, in the sense that they use the screen's real estate to emphasize whatever happen to be the main differences and similarities among the particular images at hand. By iterating this process, the user is able to quickly navigate to the portion of the image space of interest. Rather than following a thin path of images from query to query, as in the traditional approach, the user now *zooms in* to the images of interest. Precision is added incrementally in subsequent query refinements, and fewer and fewer images are displayed as the desired images are approached.

An example of navigation using the color signatures described in section 4.2 is given in figure 4 (in the color plates). We are looking for images of deserts. Since we don't know which colors can describe the desert ground (yellow? brown?) we only specify 20 percent blue for the sky and the rest 80 percent as "don't care", and ask the system to return 500 images. The MDS embedding of the returned images is shown in part (a) of the figure. By glancing at the results, we see that images are sorted from right to left by the amount of blue they contain, and from bottom to top by the amount of illumination. We see immediately that images of deserts will probably be located in the upper-center part of the map so we click there (the location of the mouse click is indicated by a black blob in the figure) and ask for 100 images. In the new MDS map (b), images at the bottom have more desert ground, images at the top have more plants (green), images at the top-right have more sky. Clicking again and asking for only 20 images results in a new MDS display, shown in (c).

The MDS can be computed quite efficiently. On a SGI Indigo 2, MDS of 100 images, including the computation of the full distance matrix (4950 EMD's), takes about four seconds on average.

4.4 Texture

While color is a point-wise property of images, texture involves a notion of spatial extent: a single point has no texture. The Fourier spectrum of an image filled with a single texture can be seen as the distribution of signal energy over the two-dimensional domain of spatial frequencies. Because Gabor filters are localized on the spatial frequency plane, the energies of their outputs can be seen as samples of the Fourier spectrum of the given texture. In this section, we use Gabor filters [3, 2, 7] to compute texture signatures. More details about the derivation of the Gabor wavelets and the choice of parameters can be found in [6]. We use a log-polar coordinates with $M = 5$ scales, and $L = 8$ orientations. Since our images are homogeneous textures, we take as our texture features the spatial mean of the energies of the Gabor responses, which we call E_{lm}, where l is the orientation and m is the scale.

We can now use the EMD as the distance measure between textures. To this end, we define our ground distance in the frequency domain to be the Euclidean distance in log-polar space, with cyclic permutation being allowed on the orientation axis. In order to have rotation-invariant representations of texture, we use the fact that rotation invariance reduces to cyclic shifts along the orientation axis in the log-polar space, to perform an exhaustive search for the minimal distance over all shifts in orientation. Formally, 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 is

$$\|E_{l_1,m_1} - E_{l_2,m_2}\|^2 = (\Delta l)^2 + (\Delta m)^2 \ , \qquad (7)$$

where

$$\begin{aligned} \Delta l &= \min\big(|l_1 - l_2 + l_s \pmod L|, \\ &\qquad L - |l_1 - l_2 + l_s \pmod L|\big) \ , \\ \Delta m &= m_1 - m_2 \ . \end{aligned}$$

A 2D MDS (with a STRESS value of 0.077) using the rotation-invariant EMD on the texture-signatures is shown in figure 3 (in the color plates). One axis emphasizes the directionality of the texture, where textures with one dominant orientation (any orientation) are at the bottom-right, and textures without a dominant orientation are at the top-left. The other axis is the texture coarseness, from coarse textures at the bottom-left to fine textures at the top-right. It is interesting to notice that coarseness and directionality were found by psychophysical experiments by Tamura et al. [15] to be the two most significant perceived texture features.

5 Efficiency considerations

It is important that the EMD be computed efficiently, especially for image retrieval systems, where a large number of distances is computed for every query. There are two ways to achieve this. The first is to have efficient algorithms for the transportation problem. The second is to have easy to compute lower bounds which can significantly reduce the number of EMD's that actually need to be computed for an image retrieval query by ignoring "unpromising" signatures.

Fortunately, efficient algorithms exist. We implemented the transportation-simplex method which is a streamlined algorithm based on the simplex method [4]. For the initial basic feasible solution we used Russell's approximation method [10]. On a SGI Indigo 2, for random pairs of color images, each represented by a color-signature with eight clusters on average, as described in section 4.2, slightly more than 1000 EMD's were computed per second. When a user defined color signature with only three clusters was matched against color signatures of real images, more that 5000 EMD's were computed per second.

An easy-to-compute lower bound between two signatures with the same total weight is the distance between their centroids. Let p_i and q_j be the coordinates of cluster i in the first signature, and cluster j in the second signature respectively. Then, using the notations of Eq. 3,

$$\begin{aligned}
\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} c_{ij} f_{ij} &= \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \|p_i - q_j\| f_{ij} \\
&= \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \|f_{ij}(p_i - q_j)\| \qquad (f_{ij} \geq 0) \\
&\geq \left\| \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} f_{ij}(p_i - q_j) \right\| \\
&= \left\| \sum_{i\in\mathcal{I}} \Big(\sum_{j\in\mathcal{J}} f_{ij} \Big) p_i - \sum_{j\in\mathcal{J}} \Big(\sum_{i\in\mathcal{I}} f_{ij} \Big) q_j \right\| \\
&= \left\| \sum_{i\in\mathcal{I}} x_i p_i - \sum_{j\in\mathcal{J}} y_j q_j \right\| .
\end{aligned}$$

Using this lower bound in our color-based, 20,000-image database retrieval system reduces the number of EMD computations on average by a factor of six.

6 Conclusions and Future Work

The earth mover's distance is a general and flexible metric, and does not suffer from the binning problems of most extant definitions of distribution distance. It allows for partial matches, and it can be applied to variable-length representations of distributions. It can be computed efficiently, and lower bounds are readily available for it. Because of these advantages, we believe that the EMD can be of use both for understanding distributions related to vision problems, as exemplified by our case studies with color and texture, and as a fundamental element of image retrieval systems.

More general definitions of a signature are possible, and can lead to more general EMD's. For example, including positional and geometric information to record the spatial location and extent in the image of the cluster pixels.

Our analysis of texture similarity in particular has brought forth a number of interesting open problems. For instance, how can the shortest distance between two signatures be computed if either of them is allowed to undergo a transformation from a predefined group at no cost? An answer to this question would lead to a more direct approach to the issue of invariance when comparing textures or other features. Also, can distances between images with many textures be computed efficiently by using the earth mover's metric itself as a ground distance for a higher level comparisons between distributions of textures? We plan to pursue these issues in our future research.

Finally, it would be interesting to apply the earth mover's distance to other vision problems such as classification and recognition based on other types of visual cues. In fact, we surmise that the EMD may be a useful metric also for problems outside the realm of computer vision.

Acknowledgments

We thank Ron Kimmel for the Gabor filter code, and Serge Plotkin, Walter Murray, Meredith Goldsmith and Scott Cohen for useful discussions about computational aspects of the EMD.

References

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

[2] F. Farrokhnia and A. K. Jain. A multi-channel filtering approach to texture segmentation. *CVPR*, 364–370, 1991.

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

[4] F. S. Hillier and G. J. Liberman. *Introduction to Mathematical Programming*. McGraw-Hill, 1990.

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

[6] T. S. Lee. Image representation using 2D Gabor wavelets. *IEEE PAMI*, 18(10):959–971, 1996.

[7] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE PAMI*, 18(8):837–842, 1996.

[8] W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, G. Taubin, and Y. Heights. Querying images by content, using color, texture, and shape. *SPIE*, 1908:173–187, 1993.

[9] Y. Rubner, L. J. Guibas, and C. Tomasi. The earth mover's distance, multidimensional scaling, and color-based image retrieval. *DARPA IUW*, 1997.

[10] E. J. Russell. Extension of Dantzig's algorithm to finding an initial near-optimal basis for the transportation problem. *Operations Research*, 17:187–191, 1969.

[11] H. C. Shen and A. K. C. Wong. Generalized texture representation and metric. *CVGIP*, 23:187–206, 1983.

[12] R. N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function, i and ii. *Psychometrika*, 27:125–140,219–246, 1962.

[13] M. Stricker and M. Orengo. Similarity of color images. *SPIE*, 2420:381–392, 1995.

[14] M. J. Swain and D. H. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.

[15] H. Tamura, T. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE SMC*, 8:460–473, 1978.

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

[17] R. R. De Valois and K. K. De Valois. *Spatial Vision*. Oxford University Press, 1988.

[18] M. Werman, S. Peleg, and A. Rosenfeld. A distance metric for multi-dimensional histograms. *CVGIP*, 32:328–336, 1985.

[19] G. Wyszecki and W. S. Styles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, 1982.

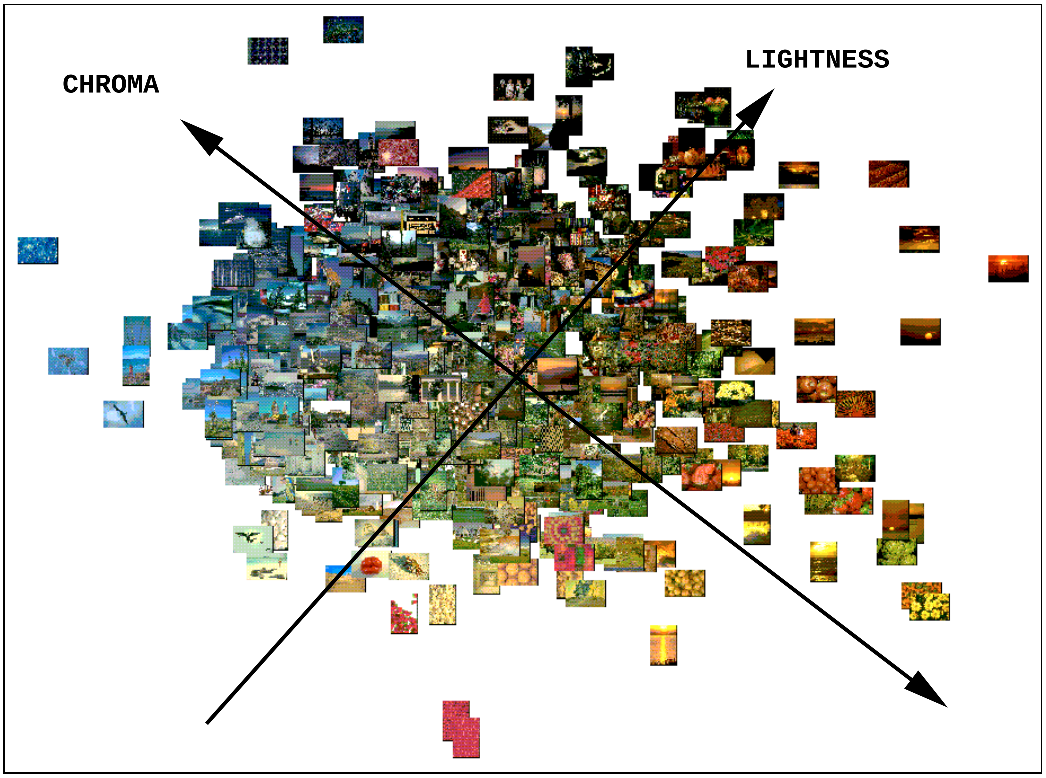

Figure 2: 2D MDS of 1000 random color images using EMD between color signatures.

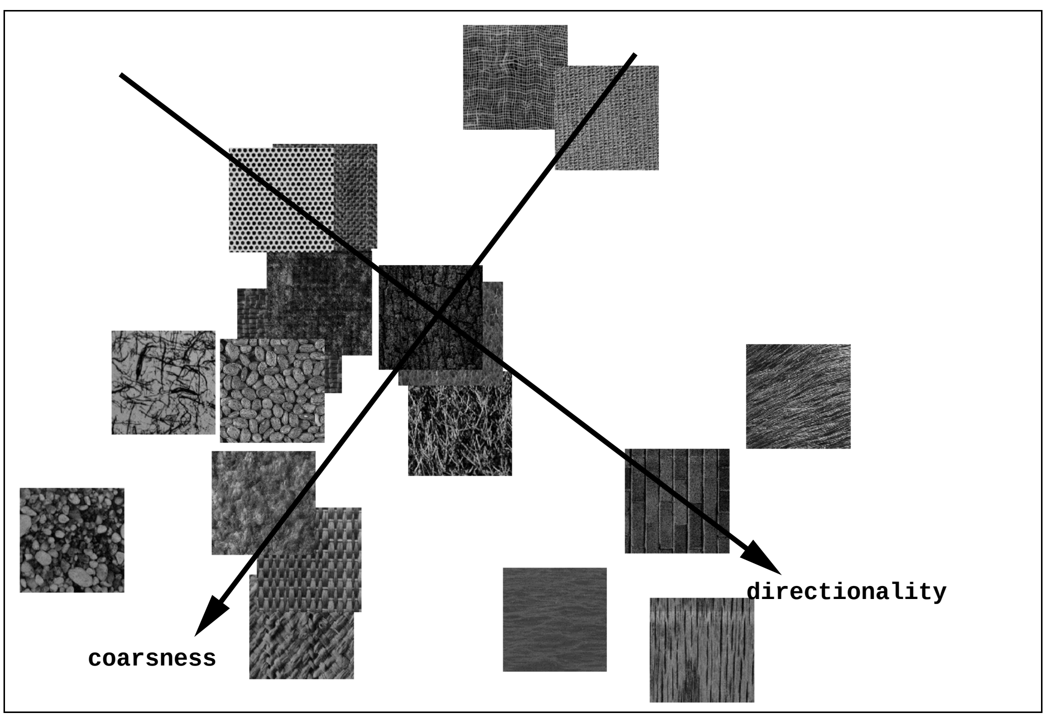

Figure 3: 2D MDS of 20 textures with rotation invariance.

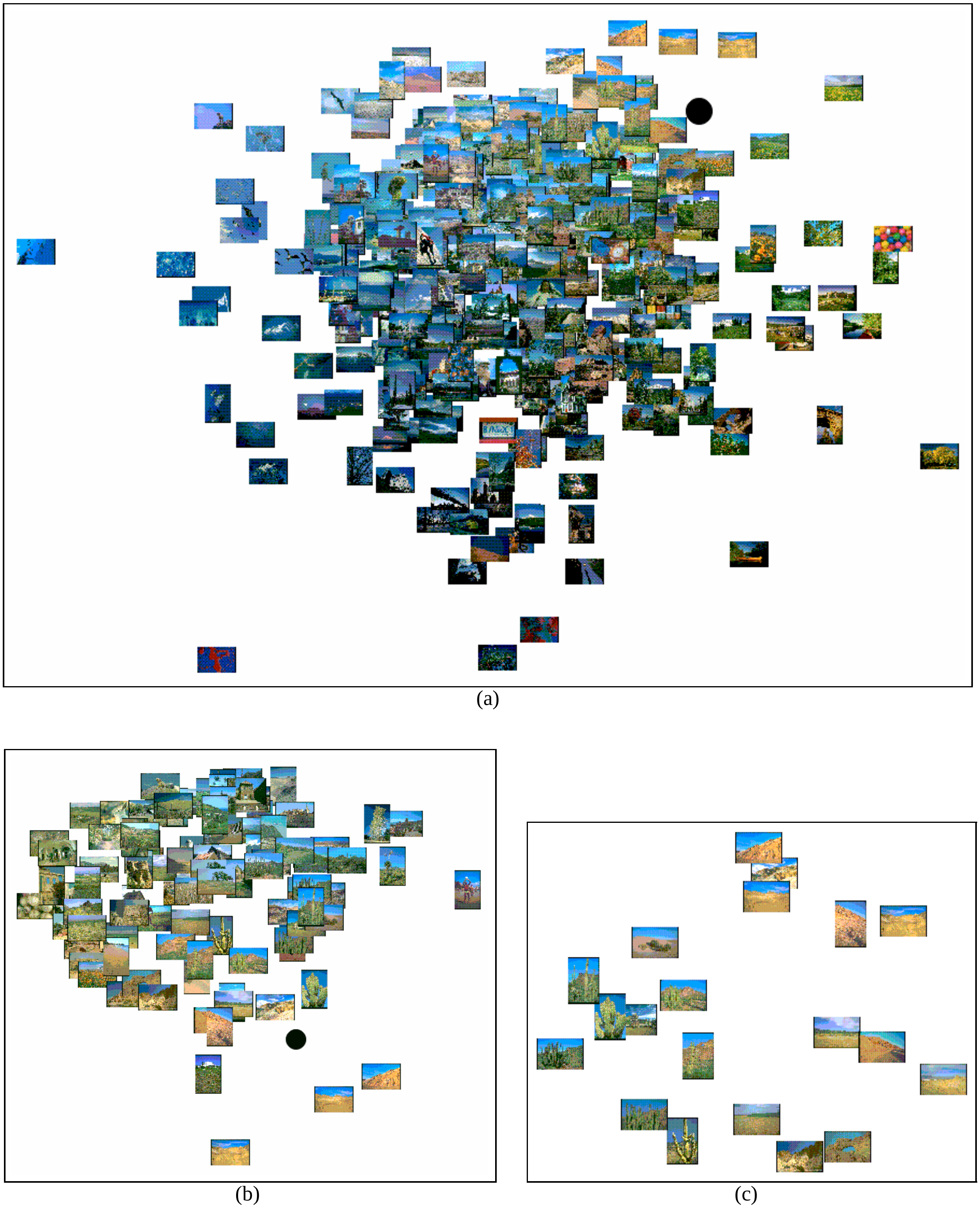

Figure 4: Looking for a desert landscape. (a) 500 images. (b) 100 images. (c) 20 images. The black dots show where the user clicked for the subsequent queries.