Visualizing Image Collections Using High-Entropy Layout Distributions

Ruixuan Wang, Stephen J. McKenna, *Senior Member, IEEE*, Junwei Han, and Annette A. Ward

***Abstract*—Mechanisms for visualizing image collections are essential for browsing and exploring their content. This is especially true when metadata are ineffective in retrieving items due to the sparsity or esoteric nature of text. An obvious approach is to automatically lay out sets of images in ways that reflect relationships between the items. However, dimensionality reduction methods that map from high-dimensional content-based feature distributions to low-dimensional layout spaces for visualization often result in displays in which many items are occluded whilst large regions are empty or only sparsely populated. Furthermore, such methods do not consider the shape of the region of layout space to be populated. This paper proposes a method, high-entropy layout distributions. that addresses these limitations. Layout distributions with low differential entropy are penalized. An optimization strategy is presented that finds layouts that have high differential entropy and that reflect inter-image similarities. Efficient optimization is obtained using a step-size constraint and an approximation to quadratic (Renyi) entropy. Two image archives of cultural and commercial importance are used to illustrate and evaluate the method. A comparison with related methods demonstrates its effectiveness.**

***Index Terms*—Content-based browsing, high-entropy layout distribution (HELD), image layouts, manifold learning, Renyi entropy.**

I. Introduction

BROWSING and exploring image collections require mechanisms for arranging items for visualization to make clear both the content of individual items and any relationships between these items. One approach is to automatically arrange images in low-dimensional (2-D or 3-D) spaces so that they can then be rendered on displays under user control.

Content-based retrieval systems often lay out images as thumbnails ordered by similarity to a query [1]; such displays do not portray the mutual relationships between items. Browsing systems often categorize images and lay them out according to class [2] or based on metadata; a *time quilt*, for example, orders representative thumbnail images by time and then wraps them into vertical columns of some maximum height [3]. Rectangle packing has been used to arrange images from each of a number of precomputed clusters, but the packing algorithm does not take into account mutual relationships between the images within each cluster of images [4]. Similarly, Bederson describes methods for arranging images within an area or on a grid but without accounting for image relationships within precomputed clusters of images [5]. Alternatively, 2-D map-based visualizations [6]–[9] lay out items so similar items appear close to one another while very different items will be further apart. These differ in how they extract high-dimensional feature vectors, measure pairwise item similarity, and perform dimensionality reduction to map the distribution of items from the high-dimensional space to a 2-D space [10]. For example, Rubner *et al.* [9] used color and texture features, earth mover's distance, and multidimensional scaling (MDS) [11]. The layouts that result can in certain cases reflect aspects of perceptual organization such as lightness and chroma for color images, directionality and coarseness for textured images [9], or lightness and pose for face images [12].

The literature on dimensionality reduction is extensive, and several survey papers are available [13]–[16]. Linear methods include principle component analysis (PCA) [17] and MDS [11]. More generally, manifold learning methods estimate the dimensionality and geometry of nonlinear data manifolds and can be broadly categorized into global methods and local methods according to whether they emphasize preservation of global or local properties of the data distribution. Global methods include Isomap [12], stochastic neighbor embedding [18], maximum variance unfolding [19], and diffusion maps [20]. Local methods include locally linear embedding (LLE) [21], Laplacian eigenmaps [22], Hessian LLE [23], and local tangent space alignment (LTSA) [24]. All of these dimensionality reduction methods were formulated with the goal of approximating high-dimensional data in spaces of lower dimensionality. In particular, they can be used to provide visualizations of high-dimensional data sets as points in 2-D or 3-D displays. The data content is often visualized by rendering images centered at each of these points. Fig. 1 shows an example in which a set of 1000 images represented using 64-D color histograms is visualized in two dimensions. While this does provide a useful insight into the data set, there are limitations. Such methods often result in displays in which many images occlude other images whilst large areas of the layout space are empty or only sparsely populated with images. Additionally, such methods do not account for the shape of the region in the layout space that will be populated, the proportions of a computer screen or the desired shape of the display. An application for a multi-user, tabletop touch display could be

Manuscript received March 18, 2010; revised June 20, 2010; accepted June 26, 2010. Date of publication July 15, 2010; date of current version November 17, 2010. This work was supported by the U.K. Technology Strategy Board under Grant FABRIC in collaboration with Liberty Art Fabrics, System Simulation, and the Victoria and Albert Museum. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jia Li.

The authors are with the School of Computing, University of Dundee, Angus DD1 4HN, U.K. (e-mail: ruixuanwang@computing.dundee.ac.uk; stephen@computing.dundee.ac.uk; junweihan@hotmail.com; award@computing.dundee.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2010.2057411

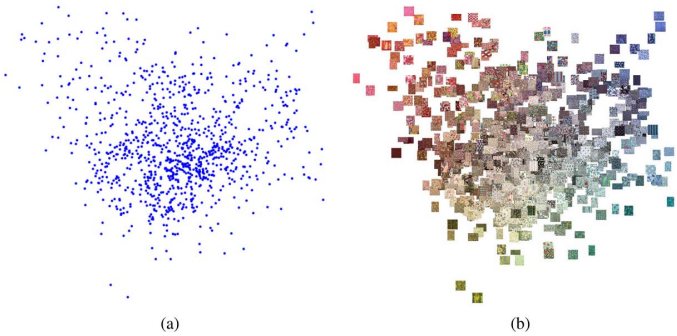


Fig. 1. Two-dimensional Isomap visualizations of a set of 1000 images from a textile design archive. (a) Isomap distribution. (b) Images rendered according to the Isomap result. Images were represented using histograms in the a^*b^*-subspace of the CIE $L^*a^*b^*$ color space. Images courtesy of Liberty Art Fabrics.

designed with an annular layout region, leaving the center of the display available for functional menus, for example.

The high-entropy layout distribution (HELD) method described in this paper addresses these shortcomings by generating layouts that conform to the shape of the available layout region, approximate the high-dimensional data distributions and result in rendered displays that are populated evenly with images. This is achieved by optimization of an objective function that combines manifold learning with a layout entropy measure. The images are taken to form a distribution in the low-dimensional layout space, and distributions with low entropy are penalized since they result in layouts in which some regions are overpopulated (i.e., many images are occluded) and other regions are sparsely populated or empty. High-entropy layouts, on the other hand, arrange the images more evenly in the layout space. The proposed method can be applied to visualize collections of images on layout regions of various shapes. An example HELD visualization using a circular layout region is shown in Fig. 2. This paper demonstrates that the layout entropy can be suitably approximated as a pairwise summation over images within a local neighborhood. It also describes how conjugate gradients descent can be used to perform the optimization provided that a mechanism for limiting step size is employed. A related method was published as an earlier conference paper [25]. It is superceded by the method described in this paper, which yields superior layouts, has reduced computational expense, and is easier to use. The most closely related work elsewhere in the literature addressed the related aim of reducing image overlap when visualizing collections of images. Basalaj [26] and Liu *et al.* [27] used an analog of MDS in a discrete domain to display each image within a single cell of a grid. Moghaddam *et al.* [7] and Nguyen *et al.* [8] used gradient descent methods to move overlapped images towards unoccupied areas of a 2-D layout without constraining image positions to be within a layout region. Empirical comparisons are made with these two methods in Section IV. Image sets used for testing are from Liberty Art Fabrics (LAF) and the Victoria & Albert Museum (VAM). Both collections are used for design inspiration and other educational and commercial applications. Thus, it is imperative that users be able to effectively and efficiently browse the images. The LAF set is comprised of images of textiles and textile-related pattern repeats. The VAM set includes images of textiles, wallpaper, mosaics, carpet, and other related objects.

II. Formulation

The goal is to arrange a set of images $\{I_i\}, i = 1, \ldots, N$, on a prespecified bounded region $\mathbf{R}$ of a layout space. The content of each image is represented as a high-dimensional feature vector $\mathbf{x}_i$. The layout produced must trade off two requirements: distances between images in the layout should depend on their content similarity, and images should spread out so as to make good use of $\mathbf{R}$. The first requirement, referred to as content structure preservation, can be met by dimensionality reduction based on the assumption that the data are distributed in a low-dimensional nonlinear manifold embedded in the feature space. As noted earlier, many manifold learning techniques have been proposed and, in principal, any such technique can be used. In this paper, Isomap is used. Isomap first constructs a sparse graph based on $\{\mathbf{x}_i\}$ with one vertex for each image. Edges are constructed between similar images' vertices by the K-nearest neighbor method. Each edge is assigned a weight w_{ij} that is the dissimilarity between the two images. An approximation, D_{ij}, to the geodesic distance between any two images is then obtained as the shortest path between their corresponding vertices. Without loss of generality, $\{D_{ij}\}$ are normalized such that the maximum D_{ij} is limited by the layout region size. Isomap determines image positions $\{\mathbf{y}_i\}$ in the low-dimensional space by minimizing E_s as follows:

$$E_s = \frac{1}{\phi_1} \sum_{i=1}^{N} \sum_{j=1}^{N} (d_{ij} - D_{ij})^2 \tag{1}$$

where d_{ij} is the Euclidean distance between $\mathbf{y}_i$ and $\mathbf{y}_j$, and $\phi_1 = \sum_i \sum_j D_{ij}^2$ is the normalization factor. When two images I_i and I_j are similar in content, the distance D_{ij} between them will be small and accordingly the two images in the low-dimensional space will tend to appear close to each other.

The second requirement is met by layouts that have high entropy when the images are considered to be samples from a distribution in the layout space. Given an image position $\mathbf{y}_i$ in the low-dimensional layout space, a Gaussian $G(\mathbf{y}_i, \mathbf{\Sigma}_i)$ is used to approximate the region occupied by this image in the space, where the covariance matrix $\mathbf{\Sigma}_i$ is determined by the image's size and shape and the number of images. The Gaussians for all of the images can be combined in a Gaussian mixture with equal weight for each Gaussian component, i.e.,

$$p(\mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} G(\mathbf{y} - \mathbf{y}_i, \mathbf{\Sigma}_i). \tag{2}$$

Renyi described a family of entropy measures of which the Shannon entropy is a special case [28], [29]. In particular, the differential version of Renyi's quadratic entropy measure can be obtained as

$$H = -\log \int p(\mathbf{y})^2 d\mathbf{y}. \tag{3}$$

This is the measure of entropy that is used here. The main reason for this choice is that, in the case of a Gaussian mixture, it can

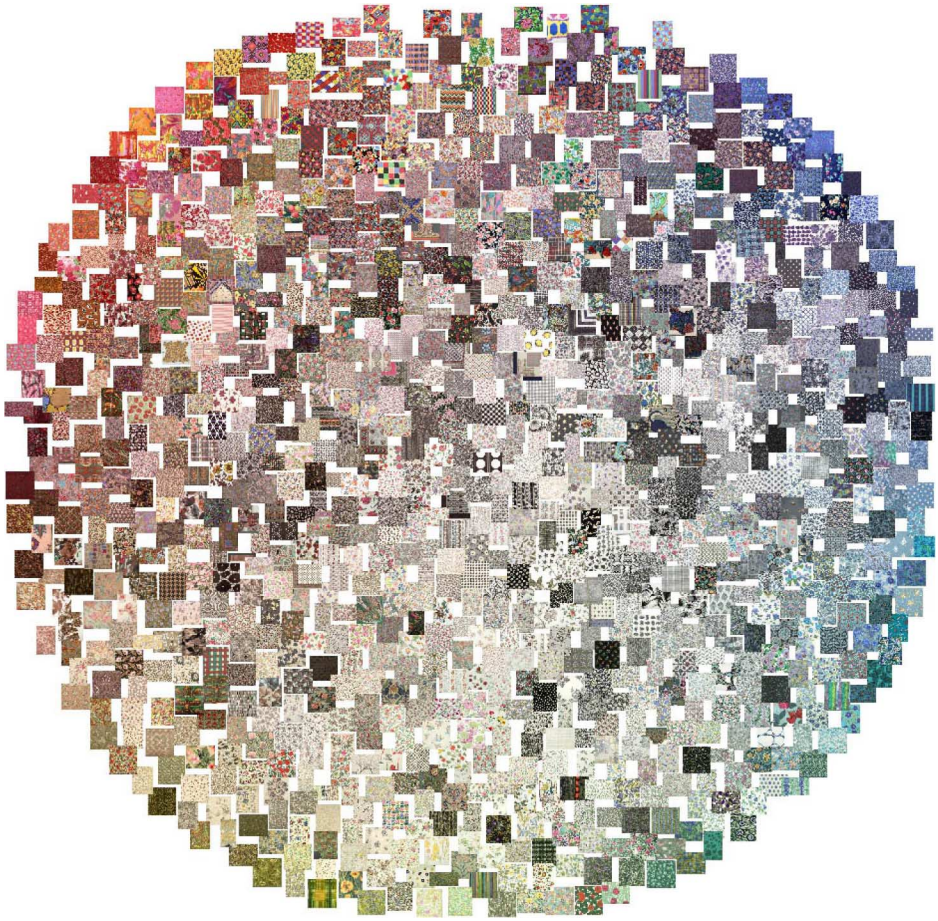


Fig. 2. Two-dimensional HELD visualization of the 1000 images used in Fig. 1. (This figure is best viewed in color, found in the online version of this paper.)

be efficiently estimated as a sum of pair-wise measures between Gaussian components [28], [29], i.e.,

$$H = -\log\left\{\frac{1}{N^2} \times \int\left(\sum_{i=1}^{N}\sum_{j=1}^{N} G(\mathbf{y}-\mathbf{y}_i,\boldsymbol{\Sigma}_i)G(\mathbf{y}-\mathbf{y}_j,\boldsymbol{\Sigma}_j)\right)\right\} = -\log\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(\mathbf{y}_i-\mathbf{y}_j,\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j)\right\}. \quad (4)$$

Furthermore, this can be approximated as

$$\hat{H} = -\log\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\in\Gamma(i)} G(\mathbf{y}_i-\mathbf{y}_j,\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j)\right\} \quad (5)$$

where the inner summation is only over each image's nearest neighbors. The set of nearest neighbors of image I_i is denoted $\Gamma(i)$. This approximation provides a good compromise between reducing computational expense and maintaining accuracy, as will be demonstrated empirically in Section IV-D.

A tradeoff between layout entropy and content structure preservation is obtained by minimizing E_λ as

$$E_\lambda = (1-\lambda)E_s - \lambda\hat{H} \quad (6)$$

subject to the constraint that each image should stay within the layout region $\mathbf{R}$, where $\lambda \in [0,1]$ is a tradeoff parameter. The value of λ should be determined with an application-dependent approach. When λ is close to 0, preservation of the manifold structure is emphasized. When λ is close to 1, spreading the images to maximize entropy is emphasized.

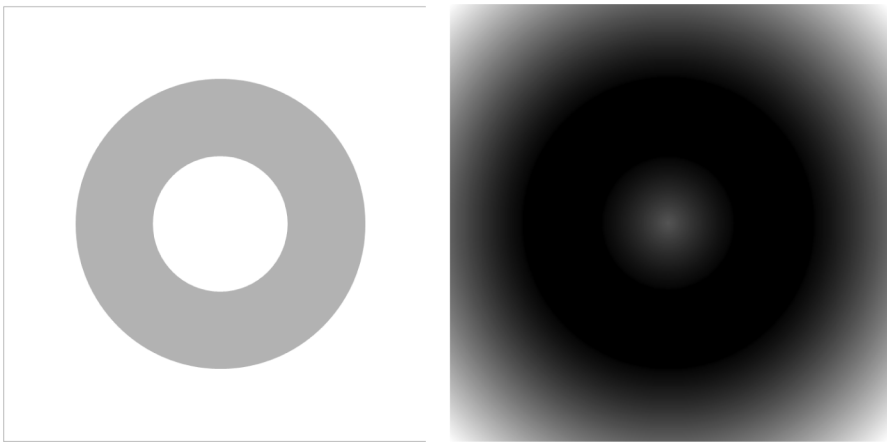
The constrained optimization problem in (6) can be solved using a penalty function method to penalize image positions outside $\mathbf{R}$. Denote by E_b the total penalty incurred by all image positions, i.e.,

$$E_b = \sum_{i=1}^{N} f(\mathbf{y}_i) \quad (7)$$

where $f(\mathbf{y}_i)$ is a monotonically increasing nonnegative function of the Euclidean distance from $\mathbf{y}_i$ to the layout region $\mathbf{R}$ (i.e., $\min_{\mathbf{y}\in\mathbf{R}} \|\mathbf{y}-\mathbf{y}_i\|$). In other words, $f(\mathbf{y}_i)$ will be zero (i.e., no penalty) if $\mathbf{y}_i$ is inside the layout region $\mathbf{R}$, and $f(\mathbf{y}_i)$ will increase when $\mathbf{y}_i$ becomes further away from $\mathbf{R}$.

The proposed method is not limited by the dimensionality of the layout space, nor by the shapes of the images. However, it is often the case that images are aligned with the axes of the layout space and covariance matrices $\boldsymbol{\Sigma}_i$ are thus diagonal. In the case of 2-D layouts

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_i^2 & 0 \\ 0 & (a\sigma_i)^2 \end{pmatrix} \quad (8)$$



(a) (b)

Fig. 3. (a) Annular layout region $\mathbf{R}$. (b) Distance transform of this layout region; all points within $\mathbf{R}$ have transform value zero.

where $a \equiv h_i/w_i$ is the aspect ratio, and h_i and w_i are the height and width of the ith image. The value of σ_i can be chosen based on image sizes, layout region size, and the number of images. Specifically, we will set it as

$$\sigma_i = \frac{w_i}{4}\sqrt{\frac{|\mathbf{R}|}{N\overline{w}\overline{h}}} \tag{9}$$

where $|\mathbf{R}|$ is the area of the layout region, and $\overline{w}$ and $\overline{h}$ are the average width and height of the images.

III. Optimization

The goal is to minimize E, where

$$E = E_\lambda + \gamma E_b \tag{10}$$

and γ is a constant to balance E_λ and E_b. Gradient-based methods can be used to find local minima of E. From (6) and (10), we have

$$\frac{\partial E}{\partial \mathbf{y}_j} = (1-\lambda)\frac{\partial E_s}{\partial \mathbf{y}_j} - \lambda\frac{\partial H}{\partial \mathbf{y}_j} + \gamma\frac{\partial E_b}{\partial \mathbf{y}_j}. \tag{11}$$

The gradient of E_s with respect to $\mathbf{y}_j$ is [11]

$$\frac{\partial E_s}{\partial \mathbf{y}_j} = -2\sum_{i\neq j}\frac{(d_{ij}-D_{ij})}{d_{ij}}\cdot(\mathbf{y}_i-\mathbf{y}_j). \tag{12}$$

From (4), we can derive the gradient of H with respect to $\mathbf{y}_j$ as

$$\frac{\partial H}{\partial \mathbf{y}_j} = -\frac{1}{\phi_2}\sum_i\{G(\mathbf{y}_i-\mathbf{y}_j,\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j) \times(\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j)^{-1}\cdot(\mathbf{y}_i-\mathbf{y}_j)\} \tag{13}$$

where $\phi_2 = \sum_i\sum_j G(\mathbf{y}_i-\mathbf{y}_j,\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j)$.

For the gradient of E_b with respect to $\mathbf{y}_j$, a discrete approximation is adopted because, in general, it is difficult to parametrically represent the function $f(\mathbf{y}_i)$ due to the freeform shape of the layout region. The layout space is discretized into a grid with unit δ and the Euclidean distance from each cross point on the grid to the layout region $\mathbf{R}$ is efficiently computed using the distance transform of the layout region [30] (Fig. 3). The function value $f(\mathbf{y}_i)$ at any point $\mathbf{y}_i$ in the layout space can then be approximated by linear interpolation of the function values at the corresponding four neighboring cross points on the grid. Therefore, the kth component of the gradient of E_b with respect to $\mathbf{y}_j$ can be approximated as

$$\frac{\partial E_b}{\partial y_{jk}} \approx \sum_{i=1}^{N}\frac{f(\mathbf{y}_j+\delta\mathbf{u}_k)-f(\mathbf{y}_j)}{\delta} \tag{14}$$

where δ is the discrete unit scale and $\mathbf{u}_k$ is the basis vector for the kth dimension of the layout space.

Optimization is an iterative process that updates image positions $\{\mathbf{y}'_j\}$ to new positions $\{\mathbf{y}_j\}$ in discrete steps as follows:

$$\mathbf{y}_j = \mathbf{y}'_j + \alpha\mathbf{p}_j \tag{15}$$

where $\mathbf{p}_j$ depends on the gradient $\partial E/\partial \mathbf{y}_j$ [see (11)] and α is the step size. Good initial positions $\{\mathbf{y}_i\}$ can be obtained by minimizing E_s using the Isomap method.

Conjugate gradients descent can be used provided that a mechanism to limit the step size is employed. Conjugate gradient methods adapt α automatically. Care must be taken that these discrete steps are not too large, particularly when λ is close to 1, as then the structure term E_s has little effect and structure can be unnecessarily lost. Therefore, the step size was limited by setting it to $\alpha = \min\{\hat{\alpha}, \tau/\max\{\|\mathbf{p}_j\|\}\}$, where $\hat{\alpha}$ was the step size determined by the conjugate gradient method, $\max\{\|\mathbf{p}_j\|\}$ was the maximum over j of all $\|\mathbf{p}_j\|$, and τ was a free parameter. This helped to limit the quantity of change of each image's position in any one iteration.

IV. Empirical Evaluation

The HELD method was evaluated using two image databases: 1000 images of textile designs from a commercial archive (LAF) and 1000 art images from a museum collection (VAM).

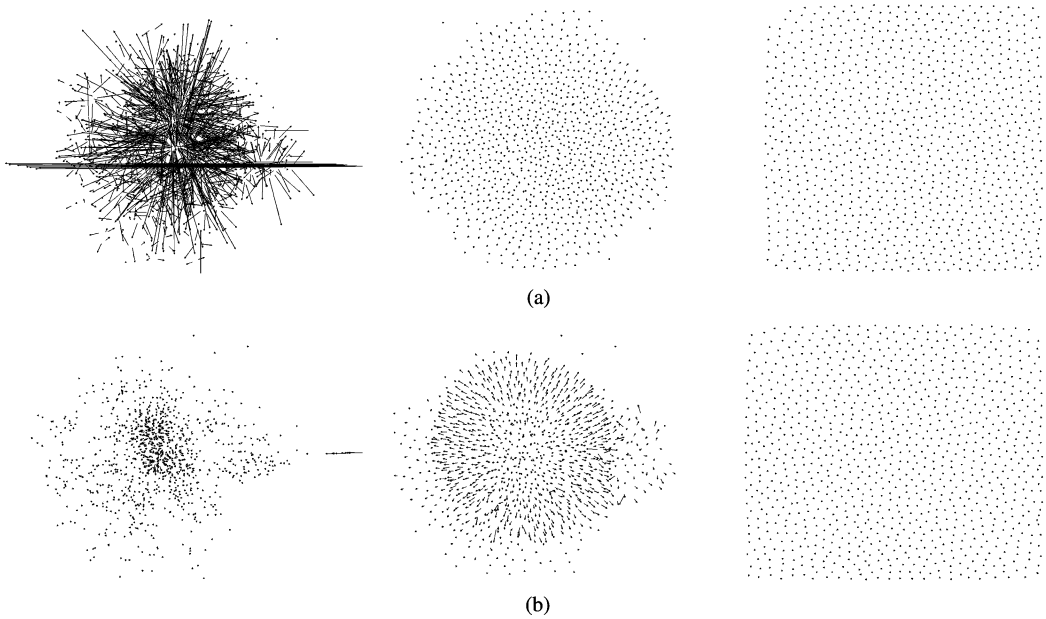


Fig. 4. Position changes of 1000 images with and without an adaptive step-size threshold after the first, tenth, and 50th iterations. The two ends of each arrow correspond to an image's positions in two consecutive iterations. (a) No threshold on step size. (b) $\tau = 0.05$.

Two kinds of features were used to represent images. Color histograms with 512 bins were extracted by regularly quantizing hue into 32 values and saturation into 16 values in the HSV color space. Texture features were extracted by performing multiscale Gabor filtering and then computing the means and variances of the normalized magnitude responses at each scale and orientation, giving 108 texture features. For both kinds of features, Euclidean distance was used to determine the nearest neighbors for constructing the manifold (see Section II). Each image was resized such that the maximum of its height and width was $0.08\sqrt{|\mathbf{R}|}$. For each test, the optimization process was terminated either when the change of cost in consecutive iterations was less than an experimentally set threshold (10^{-3}) or when the number of iterations exceeded a preset maximum number (500). All of the tests were performed using a MATLAB R2007a implementation running on an Intel Core 2 Quad 2.4-GHz PC with 3.5-GB RAM.

A. Effect of Step Size

In order to investigate the effect of the step size, α, during optimization, the HELD method was applied to the 1000 art images with $\lambda = 1.0$ and $\gamma = 1.0$. Each image was represented by a color histogram, and the function $f(\mathbf{y}_i)$ in (7) was the square of the Euclidean distance from $\mathbf{y}_i$ to the layout region $\mathbf{R}$. Fig. 4 illustrates the change of image positions in several iterations with no threshold τ (first column) and with $\tau = 0.05$ (second column). When the threshold was not used (i.e., $\alpha = \hat{\alpha}$), a large number of image positions changed abruptly in the first iteration, resulting in images "jumping over" each other and significant loss of structure. In comparison, when $\tau = 0.05$, the changes in image positions were limited initially; the largest movements being those of the outlying images that can be observed on the right hand side. Subsequently, the dense central part of the layout underwent divergence. In both cases, the resulting point distribution was similar (Fig. 4).

In addition, we have found that the use of the threshold τ can make the optimization insensitive to a large range of γ (e.g., 0.1 to 100) and to different types of function $f(\mathbf{y}_i)$ (i.e., linear and square of the Euclidean distance from $\mathbf{y}_i$ to $\mathbf{R}$). In the subsequent experiments reported in this paper, $\tau = 0.01$, $\gamma = 1.0$, and $f(\mathbf{y}_i)$ was the square of the Euclidean distance from $\mathbf{y}_i$ to the layout region $\mathbf{R}$, unless stated otherwise.

B. Methods for Comparison

The HELD method was compared with methods used by Moghaddam *et al.* [7] and Nguyen *et al.* [8]. Although not actually included in the original methods of Nguyen *et al.* and Moghaddam *et al.*, E_b in (10) was used in our implementations of their algorithms in order to facilitate fair comparison. For similar reasons, the structure preservation term E_s was used in our implementation of Nguyen *et al.*'s method such that their cost function [8, eq. (9)]) can be written

$$E = (1-\lambda)E_s + \lambda \cdot E_V + \gamma E_b \tag{16}$$

where E_V is the cost of the overall overlap. In [8], every image was modeled as circular with the same radius r, and the overlap between two images I_i and I_j was measured as the area of intersection of two circles:

$$O_{ij} = \begin{cases} r^2\left(2\arccos\left(\frac{d_{ij}}{2r}\right) - \sin\left(2\arccos\left(\frac{d_{ij}}{2r}\right)\right)\right), \\ \qquad \text{if } d_{ij} < 2r \\ 0, \qquad \text{otherwise} \end{cases} \tag{17}$$

where d_{ij} is the Euclidean distance between the image centers.

Similarly, the cost function of Moghaddam *et al.*'s method [7, eq. (1)] was reformulated as

$$E = (1-\lambda)\cdot S \cdot G + \lambda \cdot F + \gamma E_b \tag{18}$$

where F is the cost of the overall overlap and G is the cost of the overall deviation of estimated image positions from the initial

image positions. S is a scaling factor. In [7], each image I_i was modeled as circular with radius r_i, and the overlap between two images I_i and I_j was measured by

$$O_{ij} = \begin{cases} 1 - e^{-u_{ij}^2/\sigma_f}, & \text{if } u_{ij} > 0, \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where $u_{ij} = r_i + r_j - d_{ij}$, and σ_f was predetermined [7, eq. (4)].

C. Overlap Versus Structure Preservation

Since the explicit aim of the methods of Moghaddam *et al.* [7] and Nguyen *et al.* [8] was image overlap reduction, HELD was compared with those methods using a similar measure of overlap. Specifically, total image overlap was measured as the sum of all pairwise image overlaps, $\sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{z_{ij}}$, where the area z_{ij} of the overlap region can be directly computed from the image positions $\mathbf{y}_i$ and $\mathbf{y}_j$ and the image sizes (w_i, h_i) and (w_j, h_j).

Structure preservation error was measured as

$$\min_{\beta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} (\beta \cdot d_{ij} - D_{ij})^2 \right\}^{1/2} \tag{20}$$

where the value of the normalization factor β at this minimum can be analytically computed as

$$\beta^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij} \cdot D_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}^2}. \tag{21}$$

It is necessary to use β because the structure of the image distribution remains the same if all d_{ij} are scaled by the same amount.

A set of 500 images was randomly sampled from the LAF data set. Each image was represented using the Gabor features. The tradeoff parameter λ was varied between 0.0 to 1.0. For each λ value, the structure error and the overlap error were measured based on the convergent result of each method. Furthermore, a relative structure error was computed as the ratio of the structure error at the chosen value of λ to the structure error when $\lambda = 0.0$. Conjugate gradients optimization was used for all three methods with $\tau = 0.01$. Each method took between 20–100 iterations to converge. An iteration took approximately 2.8, 3.9, and 4.6 s for the HELD method, Nguyen's method, and Moghaddam's method, respectively. Experiments indicate that computation time in all three methods increases roughly quadratically with respect to the number of images. Fig. 5 plots relative structure error against overlap error. When relative structure error was low (i.e., the solution was close to the Isomap initialization), the methods were comparable. However, the lowest overlap error was considerably lower for the HELD method. The overlap error was 140 compared with 380 and 180 for the other methods. This behavior can be perceptually verified by observing the visualizations of 100 images shown in Fig. 6. The visualization obtained by HELD [Fig. 6(a)] shows almost no image overlap, as compared to results from the methods of Moghaddam *et al.* [Fig. 6(b)] and Nguyen *et al.* [Fig. 6(c)].

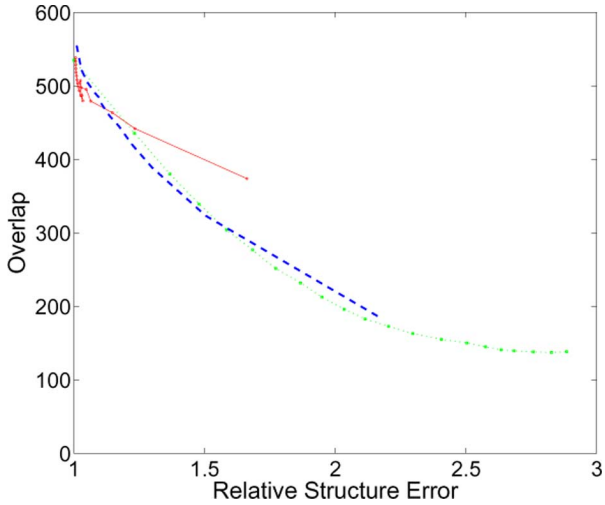


Fig. 5. Relationships between structure error and image overlap obtained by Moghaddam *et al.*'s method (red solid line), Nguyen *et al.*'s method (blue dashed line), and the proposed HELD (green dotted line) method for 500 textile images.

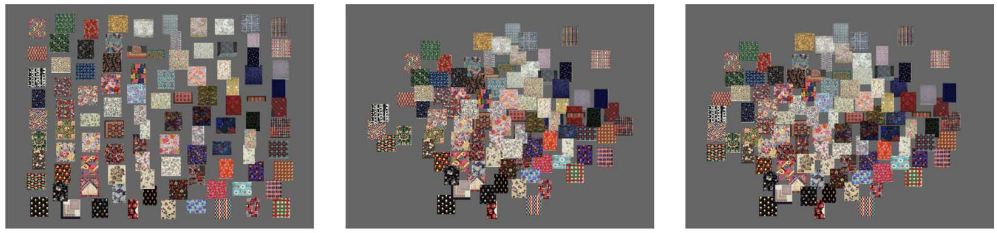
D. Entropy Approximation

The Renyi entropy can be efficiently approximated by summing only over images within each image's neighborhood as indicated in (5). Fig. 7 explores the effect of making such an approximation on computation time and accuracy. The sets of nearest neighbors included in the entropy approximation were those within $3\sigma_i$ of each image's center. A set of 500 textile images was arranged with and without the approximation and the results compared. (Note that results reported elsewhere in this paper were computed without the approximation). Fig. 7(a) and (b) shows plots of the time per iteration and the accuracy of the approximation obtained, respectively. Fig. 7(c) shows a plot of the time taken against the number of images in the layout and indicates the improved scaling. Fig. 7(d) shows a layout of 100 images obtained using the approximation and should be compared with Fig. 6(a), which shows the same image set arranged without the use of the approximation. Although the approximation produces a slightly inferior layout, this is superior to those obtained using the competing methods (Fig. 6).

E. Ability to Spread Out Images

A further experiment was performed to compare the ability of HELD to spread out images when the number of images was large relative to the available layout region. A set of 1000 VAM images was automatically arranged based on color histograms. Fig. 8 shows the resulting locations of the 1000 images. Results based on Moghaddam *et al.* [Fig. 8(a)] showed approximately 100 very tight clusters with most images seriously overlapped when rendered; whereas Nguyen *et al.* [Fig. 8(b)] showed a more even spread, although image density still varied. HELD (see rightmost column of Fig. 4] yielded the most evenly distributed result in which the image density was approximately constant.

The following synthetic experiment helps to explain why the methods performed as they did. Consider a situation in which



(a) (b) (c)

Fig. 6. Visualizations of 100 textile images. Visualizations based on the image positions obtained by (a) HELD, (b) Moghaddam *et al.*, and (c) Nguyen *et al.*. Images courtesy of LAF.

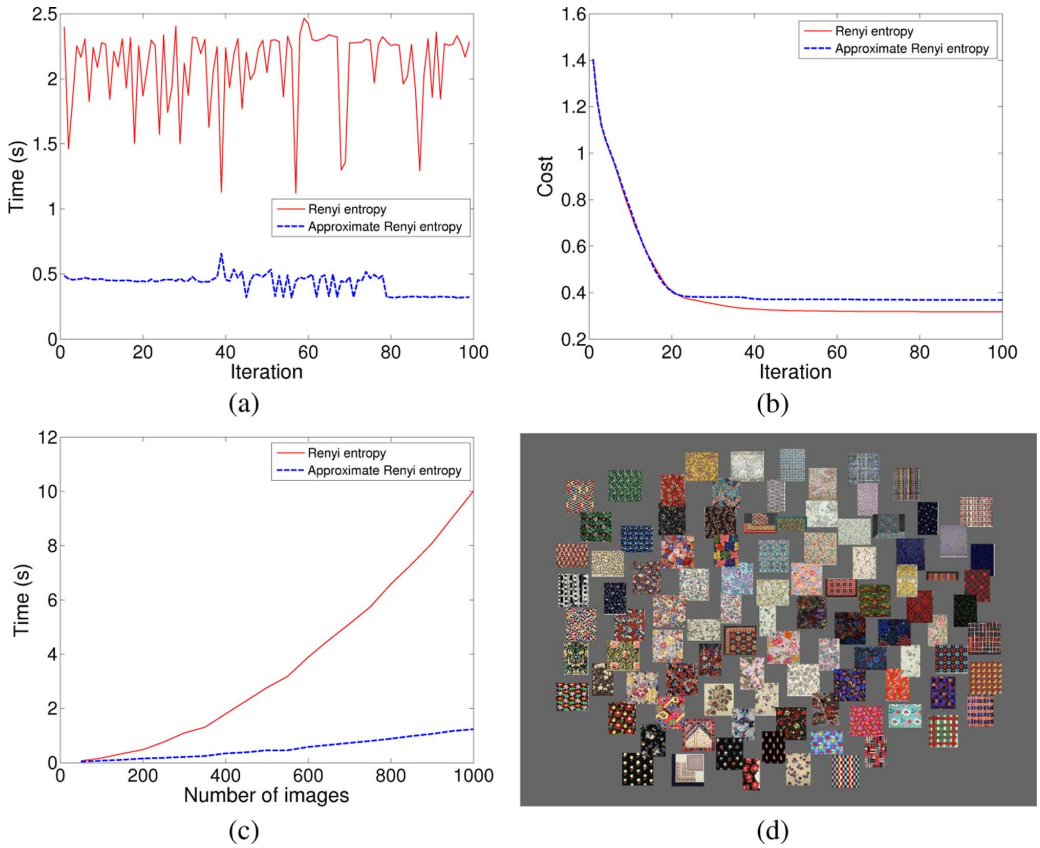
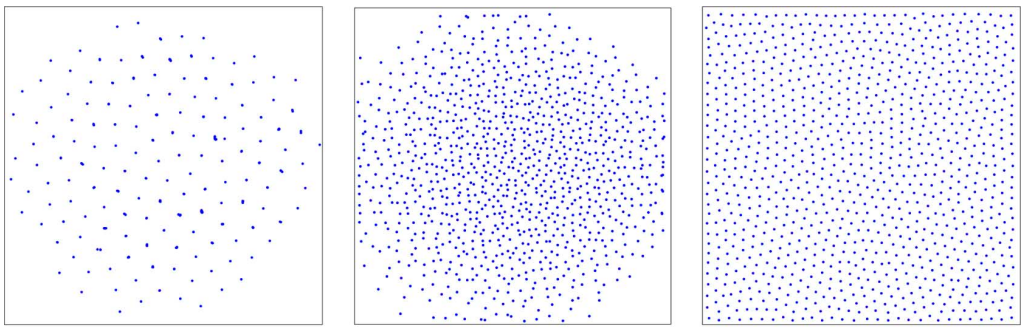


Fig. 7. Approximation of Renyi entropy [(4)]. (a) Time per iteration when optimizing a layout of 500 images using nearest neighbors within 3σ. (b) Accuracy of the approximation thus obtained. (The cost plotted is $-H+\gamma E_b$.) (c) Average computation time as a function of the number of images. (d) Visualization of 100 textile images based on the approximate Renyi entropy using nearest neighbors within 3σ. (All of these results were obtained with $\lambda = 1.0$.)



(a) (b) (c)

Fig. 8. Image positions of 1000 art images obtained by methods of (a) Moghaddam *et al.*, (b) Nguyen *et al.*, and (c) HELD.

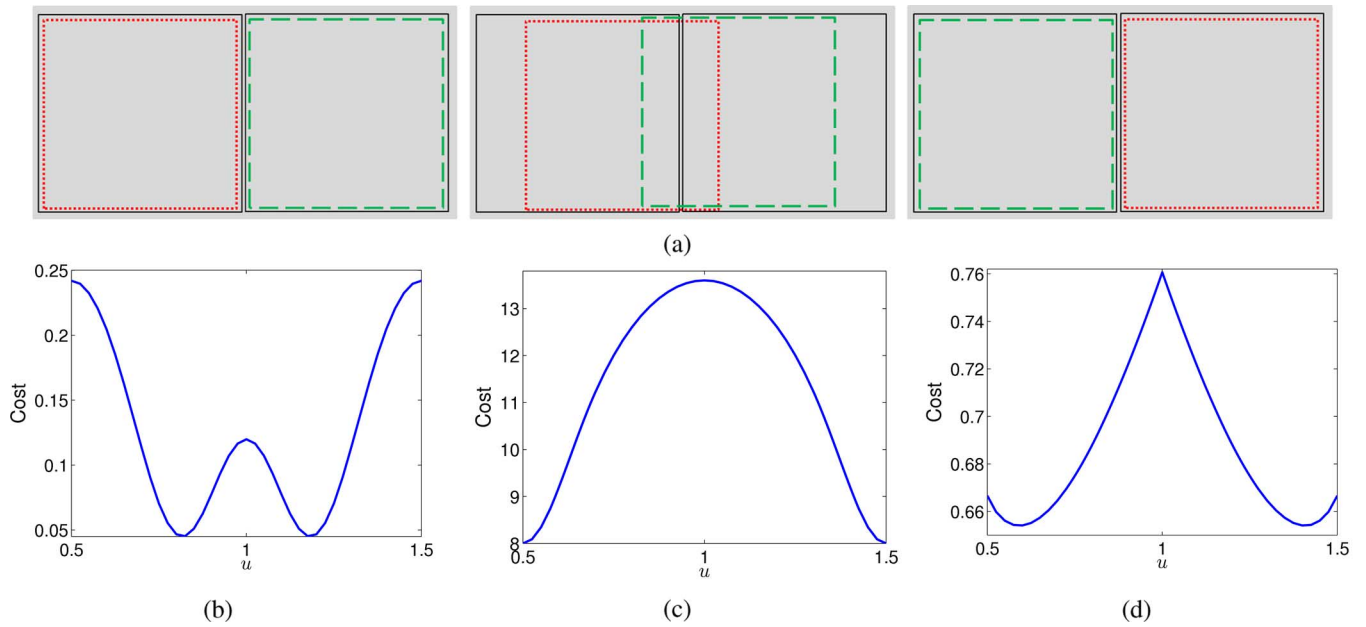


Fig. 9. (a) Layout region (gray) with two fixed images (black squares) and two movable images (red dotted and green dashed (online version) squares). Left to right: three of the many possible layouts. (b-d) Values of overlap cost terms obtained by antisymmetrically moving the two movable images: (b) HELD, (c) Moghaddam *et al.*'s method, and (d) Nguyen *et al.*'s method.

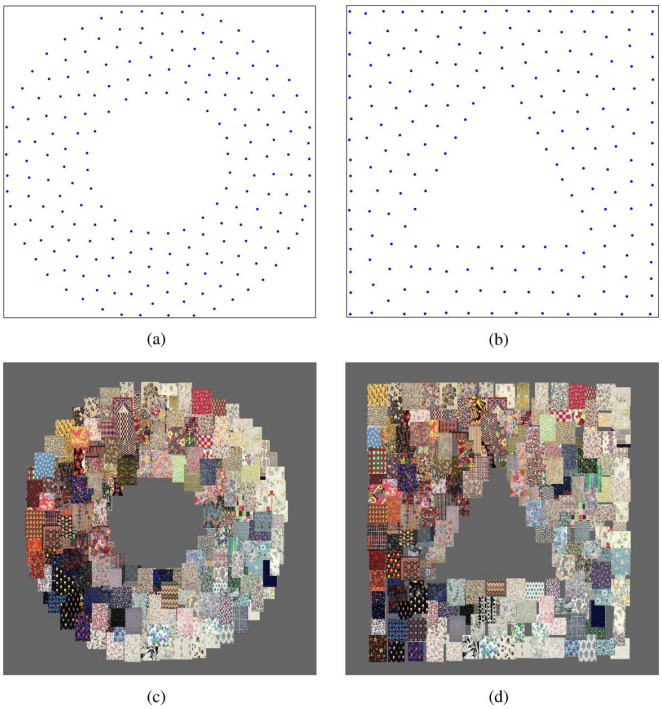


Fig. 10. Visualization of 200 textile images on two different layout regions. Positions of image centers in layouts obtained for: (a) an annular region and (b) a rectangular region with a triangular hole. (c,d) Corresponding visualizations of the images based on the layouts.

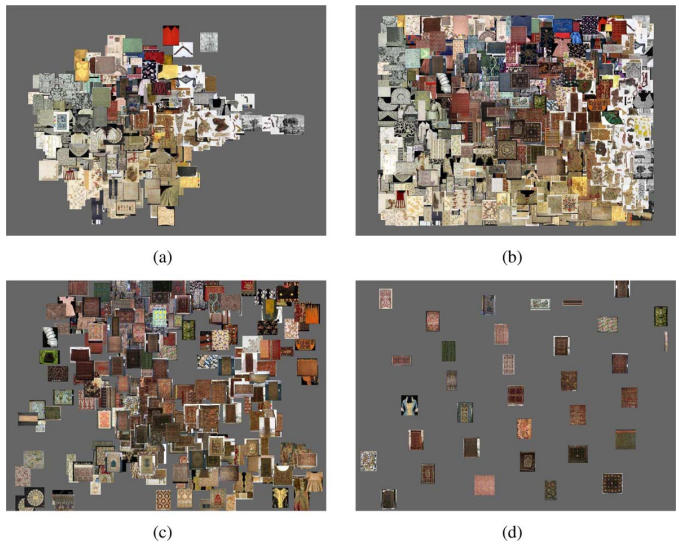


Fig. 11. Visualization of 1000 art images using color features. (a) Isomap. (b) HELD (emphasizing entropy maximization). (c) Isomap visualization after rescaling the image positions around an image at the layout center. (d) HELD visualization after similarly rescaling the image positions around an image at the layout center. Fig. 4 illustrates image positions obtained during optimization of this data set. Images courtesy of VAM.

four square images with width $w = 1.0$ are arranged in a rectangular region of width 2.0 and height 1.0. Two of the images' positions are fixed so that together they fill the region. Let the origin be at the bottom-left corner of the layout region so that these two images have horizontal coordinates 0.5 and 1.5. The other two images are moved antisymmetrically between the two fixed ones such that when one of them is at u $(0.5 \leq u \leq 1.5)$, the other is at $2.0-u$. Fig. 9(a) shows a schematic of this experiment for three values of u. Fig. 9(b-d) shows the values of the overlap cost terms [based on (4), (17) and (19)] as u varies from 0.5 to 1.5. The overlap cost using HELD is minimized when the four images are positioned such that the distance between any image and its nearest neighbor is approximately the same [Fig. 9(b)]. Fig. 9(c) shows that the overlap cost for Moghaddam *et al.*'s method is minimized when the two movable images are positioned directly over the two fixed images. Fig. 9(d) shows

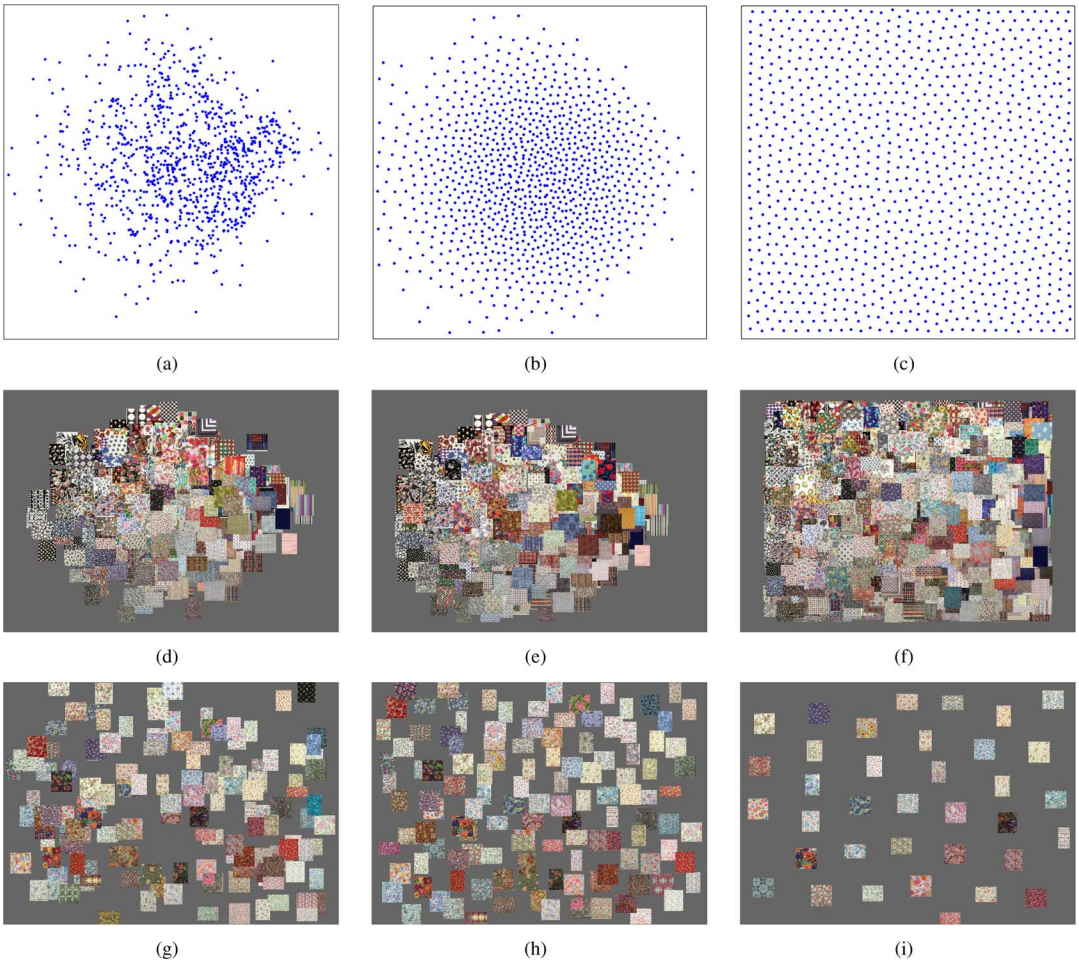


Fig. 12. Visualization of 1000 textile images by HELD using Gabor texture features. Image positions obtained using: (a) Isomap, (b) a tradeoff between structure preservation and image overlap ($\lambda = 0.1$), and (c) an emphasis on maximizing entropy. (d)–(f) The corresponding visualizations of the image collection. (g)–(i) Corresponding visualizations after rescaling the image positions around an image at the layout center. Images courtesy of LAF.

that the overlap cost for Nguyen *et al.*'s method is minimized when the movable images overlap each other only slightly. This is consistent with the result observed in Fig. 8.

F. Layout Region Shape

The proposed method can be applied to visualize collections of images on layout regions of various shapes. Here, an annular layout region [Fig. 10(a)] and a rectangular layout region with a triangular hole [Fig. 10(b)] were used to visualize 200 images of textile designs. Each image was represented by its color histogram. The algorithm was initialized using Isomap and then run with $\lambda = 0.9$ followed by further iterations with $\lambda = 1.0$ in order to spread out images in each layout region. Fig. 10(a) and (b) shows that all images are spread out in the layout region, while Fig. 10(c) and (d) qualitatively confirms that images similar in color are still positioned close to one another.

G. Visualizations for Browsing

Fig. 11 shows visualizations of 1000 VAM images based on color histograms. The γ parameter was set to 10 and the method took approximately 10 seconds per iteration (without the neighbourhood approximation). In the initial distribution obtained by Isomap [Fig. 11(a)], most images were clustered around the center of the layout with fewer images irregularly distributed near the boundaries. When entropy was emphasized ($\lambda = 1$), the image density became approximately constant [Fig. 11(b)]. Obviously, total image overlap is always large when visualizing 1000 images on a small 2-D display. In order to better show the effect of the method, the visualizations were zoomed in around one image near the layout center. The resulting visualizations are shown in Fig. 11(c) and (d). Note that the image positions were rescaled by this zoom operation but the images themselves were not rescaled. Image overlap was effectively reduced by the proposed method. This zoom operation can provide an effective way for users to focus on parts of a large collection during browsing.

Fig. 12 shows results using 1000 LAF images and Gabor texture features. Again, similar images remain close to one another as the requirement of structure preservation is relaxed. In Fig. 12(d-f), the roughness of the image texture changes smoothly from the upper left to the lower right in the display, for example. This gradual change of texture should help users to browse collections of images.

V. Conclusion

The HELD method described in this paper arranges collections of images for display with dependencies both on the content of the images and on the relative size and shape of the layout region to be populated. The images were taken to form a spatial Gaussian-mixture distribution in the layout space. An objective function was specified that rewards spatial distributions with high quadratic Renyi entropy that also preserves content-based image relationships. An efficient approximation to the entropy was described. The method was demonstrated using two image collections. It was effective provided that the step sizes used in optimization were controlled. While the optimization method is not guaranteed to find the global minimum of the cost function, the method never terminated in obviously poor local minima given layout regions such as those reported. It compared favorably with two methods previously proposed for reducing image overlap.

The HELD method is integrated into a content-based browsing and retrieval system. In order to preserve content-based structure, the Isomap cost function based on a graph-based geodesic distance approximation was used. Future work will investigate incorporating other dimensionality-reduction methods. Importantly, HELD is not limited to 2-D displays, and we are exploring its use for 3-D vizualization of image collections. Finally, the approach could also be used for multimedia collections in which visual icons or thumbnails are used to denote diverse items.

Acknowledgment

The authors would like to thank I. Ricketts and M. Stapleton for helpful discussions and A. Buruma and J. Stevenson for providing image collections and insight into how these are used.

References

[1] J. Smith and S. Chang, "VisualSEEk: A fully automated content-based image query system," in *Proc. ACM Multimedia Conf.*, Boston, MA, 1997, pp. 87–98.

[2] H. Kang and B. Shneiderman, "Visualization methods for personal photo collections browsing and searching in the photofinder," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2000, pp. 1539–1542.

[3] D. F. Huynh, S. M. Drucker, P. Baudisch, and C. Wong, "Time quilt: Scaling up zoomable photo browsers for large, unstructured photo collections," in *Proc. CHI '05 Extended Abstracts on Human Factors in Comput. Syst.*, 2005, pp. 1937–1940.

[4] A. Gomi, R. Miyazaki, T. Itoh, and J. Li, "Cat: A hierarchical image browser using a rectangle packing technique," in *Proc. Int. Conf. Inf. Visualisation*, 2008, vol. 3, pp. 82–87.

[5] B. B. Bederson, "Photomesa: A zoomable image browser using quantum treemaps and bubblemaps," in *Proc. ACM Symp. User Interface Software and Technol.*, 2001, vol. 3, pp. 71–80.

[6] J. Fan, Y. Gao, H. Luo, D. A. Keim, and Z. Li, "A novel approach to enable semantic and visual image summarization for exploratory image search," in *Proc. ACM Conf. Multimedia Inf. Retrieval*, Vancouver, BC, Canada, 2008, pp. 358–365.

[7] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, and T. Huang, "Visualization and user-modeling for browsing personal photo libraries," *Int. J. Comput. Vis.*, vol. 56, no. 1, pp. 109–130, 2004.

[8] G. Nguyen and M. Worring, "Interactive access to large image collections using similarity-based visualization," *J. Visual Languages Comput.*, vol. 19, no. 2, pp. 203–224, 2006.

[9] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *Proc. Int. Conf. Comput. Vis.*, Bombay, India, 1998, pp. 59–66.

[10] K. Rodden, "Evaluating similarity-based visualisations as interfaces for image browsing," Ph.D. dissertation, Computer Laboratory, Univ. of Cambridge, Cambridge, U.K., 2002.

[11] M. Cox and M. Cox, *Multidimensional Scaling*. London, U.K.: Chapman and Hall, 2001.

[12] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[13] L. Cayton, "Algorithms for manifold learning," uCSD Tech. Rep. CS2008–0923, 2005.

[14] I. Fodor, "A survey of dimension reduction techniques," Lawrence Livermore Nat. Lab., 2002, Tech. Rep. UCRL-ID-148494.

[15] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee, *Spectral Methods for Dimensionality Reduction*. Cambridge, MA: MIT Press, 2006, ch. 16.

[16] L. van der Maaten, E. Postma, and H. van den Herik, "Dimensionality reduction: A comparative review," Tilburg Univ., Tech. Rep. TiCC-TR 2009–005, 2009.

[17] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. CVPR*, 1991, pp. 586–591.

[18] G. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Proc. Neural Inf. Process. Syst. (NIPS) Conf.*, 2002, vol. 15, pp. 857–864.

[19] K. Q. Weinberger, B. D. Packer, and L. K. Saul, "Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization," in *Proc. Int. Workshop AI and Statistics*, 2005.

[20] B. Nadler, S. Lafon, R. Coifman, and I. Kevrekidis, "Diffusion maps, spectral clustering and the reaction coordinates of dynamical systems," *Appl. Comput. Harmonic Anal.*, vol. 21, pp. 113–127, 2006.

[21] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[22] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Neural Inf. Process. Syste. (NIPS) Conf.*, 2002, vol. 14, pp. 585–591.

[23] D. Donoho and C. Grimes, "Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data," in *Proc. Nat. Acad. Sci.*, 2005, vol. 102, pp. 7426–7431.

[24] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment," *SIAM J. Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.

[25] R. Wang, S. J. McKenna, and J. Han, "High-entropy layouts for content-based browsing and retrieval," in *Proc. ACM Int. Conf. Image and Video Retrieval (CIVR)*, 2009, Article no. 16.

[26] W. Basalaj, "Proximity visualisation of abstract data," Ph.D. dissertation, Computer Laboratory, Univ. of Cambridge, Cambridge, U.K., 2000.

[27] H. Liu, X. Xie, X. Tang, Z. Li, and W. Ma, "Effective browsing of web image search results," in *Proc. ACM SIGMM Workshop Multimedia Inf. Retrieval*, 2004, pp. 84–90.

[28] A. Renyi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, 1961, vol. 1, pp. 547–561.

[29] J. C. Principe, D. Xu, Q. Zhao, and J. W. Fisher, "Learning from examples with information theoretic criteria," *J. VLSI Signal Process.*, vol. 26, no. 1–2, pp. 61–77, 2000.

[30] R. Fabbri, L. D. F. Costa, J. C. Torelli, and O. M. Bruno, "2D Euclidean distance transforms: A comparative survey," *ACM Comput. Surveys*, vol. 40, no. 1, pp. 1–44, 2008.

Ruixuan Wang received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China, in 1999 and 2002, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2008.

He is currently a Postdoctoral Research Assistant with the School of Computing, University of Dundee, Angus, U.K. His research interests include computer vision, image and video analysis, and machine learning.

Stephen J. McKenna (SM'08) received the B.Sc. (Hons) degree from the University of Edinburgh, Edinburgh, U.K., in 1990, and the Ph.D. degree from the University of Dundee, Angus, U.K., in 1994.

Following a European Union research fellowship, he was an EPSRC Postdoctoral Researcher with Queen Mary, University of London. He has held visiting research positions with BT and George Mason University. He was appointed to a Personal Chair in Computer Vision at the University of Dundee, Angus, U.K., in 2008. He is an associate editor for the journals *Machine Vision and Applications* and *Pattern Analysis and Applications*.

Junwei Han received the B.S. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 1999 and 2003, respectively.

He has been a Research Fellow with Nanyang Technological University, Singapore, the Chinese University of Hong Kong, and Dublin City University. He was a Visiting Student with Microsoft Research Asia, Beijing, China, and a Visiting Researcher with the University of Surrey. His research interests include image/video retrieval, object segmentation, video communication, and gesture recognition.

Annette A. Ward received the B.S. degree from the University of Nevada, Reno, the M.S. degree from Colorado State University, Fort Collins, and the Ph.D. degree from New Mexico State University, Las Cruces.

She holds the Scottish Power Fellowship in the School of Computing, University of Dundee, Angus, U.K. She has conducted research on user applications of content-based image retrieval at Liberty Art Fabrics, Victoria and Albert Museum, the BBC, London Guildhall Library and Art Gallery, and The British Library. She has held academic positions in textiles, clothing and merchandising, and marketing fields in the United States and India. Previously, she conducted research at the Institute for Image Data Research, University of Northumbria.