# Region Graph Spectra as Geometric Global Image Features

Qirong Ho, Weimiao Yu, and Hwee Kuan Lee

Bioinformatics Institute, A*STAR, Singapore

**Abstract.** In quantitative biology studies such as drug and siRNA screens, robotic systems automatically acquire thousands of images from cell assays. Because these images are large in quantity and high in content, detecting specific patterns (phenotypes) in them requires accurate and fast computational methods. To this end, we have developed a geometric global image feature for pattern retrieval on large bio-image data sets. This feature is derived by applying spectral graph theory to local feature detectors such as the Scale Invariant Feature Transform, and is effective on patterns with as few as 20 keypoints. We demonstrate successful pattern detection on synthetic shape data and fluorescence microscopy images of GFP-Keratin-14-expressing human skin cells.

## 1 Introduction

Sophisticated microscopy and cell culture systems have enabled high-throughput and high-content screens that yield thousands of images. Timely biological discovery using these images requires computational methods that rapidly exploit the voluminous amount of information they contain. In particular, many biological studies involve identification of cells with specific phenotypes (visual appearances), either to identify compounds with a specific biological effect [1] or to perform further data mining on the cell images [2]. The need for computerized phenotype detection becomes evident when one considers the time required for visual inspection of each cell.

Because traditional approaches to medical image analysis have proven less than ideal for this task, the biomedical imaging community has begun to adopt data mining and machine learning methods [3]. Pioneering work by Boland *et al.* [4] demonstrated successful classification of cells into a fixed number of categories, using a neural network classifier trained on global features extracted from manually-labeled cells. If automatic cell detection is provided, then their method becomes applicable to large-scale phenotype detection. Recently, Jones *et. al* [5] have combined cell segmentation, gentle boosting [6] on cell features, and iterative learning into a framework for supervised detection of arbitrary cell phenotypes. Because the features are cell-specific, their framework is reliant on accurate cell segmentation, which may be challenging under certain imaging conditions.

On the other hand, the trend in the computer vision community has been to represent visual objects as unordered distributions of local descriptors produced

G. Bebis et al. (Eds.): ISVC 2009, Part I, LNCS 5875, pp. 253–264, 2009.

by some detector [7]; examples of detectors include Harris detector specializations [8], the Scale-Invariant Feature Transform (SIFT) [9] and the derivative Rotation-Invariant Feature Transform (RIFT) [10], as well as Speeded-Up Robust Features (SURF) [11]. Such unordered distributions, also known as *bag-of-words* representations, have been popular due to their successful application to natural (real-world) scenes [12]. Because natural scene objects are complex, containing hundreds or thousands of local descriptors, bag-of-words methods usually employ clustering to obtain representative descriptors [12]. Conversely, some bio-image phenotypes are low in complexity and local feature count — for instance, the biological image set in this paper contains ∼25 SIFT descriptors per exemplar. This reduces the number of meaningful clusters that can be extracted, which may limit the discriminative capabilities of bag-of-words approaches.

To overcome this problem, we propose a phenotype detection framework whose key contribution is a *global feature* for visual objects derived from *relationships* between local descriptors, as opposed to the *distribution* of descriptors that characterizes bag-of-words approaches. This global feature is a vector in $\mathbb{R}^k$, hence we describe it as *geometric* because it naturally admits the Euclidean norm as one notion of inter-feature distance. Furthermore, it can be made both scale- and rotation-invariant for local feature detectors with those properties, such as SIFT. Additionally, our framework does not rely on segmentation; we employ a general region-sampling technique that performs well for objects that fit within cell-shaped regions. We demonstrate our framework's capabilities on synthetic and biological images, where in both cases the patterns of interest express few (i.e. tens of) SIFT descriptors.

## 2 Global Features from Local Feature Detectors

In this section, we develop a geometric global feature for arbitrary closed regions in an image. This global feature, which we call a *region spectrum*, is derived from local feature descriptors such as SIFT [9], SURF [11] and RIFT [10]; these descriptors characterize visually significant points of the image. The general idea is to find all descriptors within a region of interest, then construct multiple graphs whose vertices represent descriptors, and whose edge weights are proportional to similarities between descriptor feature vectors. Collectively, the eigenvalues of every graph's Laplacian matrix make up the region spectrum, a vector describing the region's geometric properties.

### 2.1 Regions and Local Feature Descriptors

Formally, let $\mathcal{D}$ be the set of descriptors generated by a local feature detector (e.g. SIFT) on an image domain $\Omega$. Each descriptor $\mathbf{d} \in \mathcal{D}$ is associated with a position $\mathbf{x} = (x, y)$ on the image. In addition, $\mathbf{d}$ also contains features such as local patch information. Generally, $\mathbf{d}$ is a real valued vector. For any closed region $R \subseteq \Omega$ in the image, let $\mathcal{D}_R \subseteq \mathcal{D}$ be the set of descriptors whose associated positions fall within $R$. This is illustrated in Fig. 1(a).

In our experiments we used SIFT [9] to obtain local feature descriptors. SIFT detects gradient maxima and minima in an image's scale-space representation, generating a descriptor $\mathbf{d} = (\mathbf{x}, \sigma, \theta, \mathbf{h})$ for each detected extremum. Here, $\mathbf{x}$ is the keypoint location associating the descriptor to a position in the original image and $\sigma$ is the extremum's scale coordinate[1]. $\theta$ is the dominant gradient orientation of the region surrounding $(\mathbf{x}, \sigma)$ in scale-space, and $\mathbf{h}$ is a 128-bin histogram of gradients in a $\theta$-oriented window centered at $(\mathbf{x}, \sigma)$. We chose SIFT for its robustness and utility in image registration tasks [9,13], though our method works with any feature detector for which meaningful inter-descriptor distances can be taken [11,10].

### 2.2 Region Graphs and Region Spectra

A region's descriptor set $\mathcal{D}_R$ can be used to construct one or more graphs $G_{R,C}$, where $C : \mathcal{D}_R \times \mathcal{D}_R \mapsto [0, 1]$ is a "connectivity" function between two descriptors. The vertex set of $G_{R,C}$ is $\mathcal{D}_R$, while its edge weights are $w_{\mathbf{d}_i,\mathbf{d}_j} = C(\mathbf{d}_i, \mathbf{d}_j)$, $\mathbf{d}_i, \mathbf{d}_j \in \mathcal{D}_R$. Refer to Fig. 1(b) for an illustration. For SIFT descriptors, we define three connectivity functions $C_\mathbf{x}$, $C_\sigma$ and $C_\theta$:

$$C_\mathbf{x}(\mathbf{d}_i, \mathbf{d}_j) = \exp(-\alpha_\mathbf{x} \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (1)$$

$$C_\sigma(\mathbf{d}_i, \mathbf{d}_j) = \exp(-\alpha_\sigma |\sigma_i - \sigma_j|^2) \quad (2)$$

$$C_\theta(\mathbf{d}_i, \mathbf{d}_j) = \exp(-\alpha_\theta \min[|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|]^2) \quad (3)$$

where $\mathbf{x}_i$, $\sigma_i$ and $\theta_i$ denote the $\mathbf{x}$, $\sigma$ and $\theta$ components of $\mathbf{d}_i$ respectively. $\alpha_\mathbf{x}$, $\alpha_\sigma$ and $\alpha_\theta$ are scaling coefficients. These connectivity functions indicate descriptor similarity in terms of image Euclidean distance $\mathbf{x}$, scale $\sigma$ and angle $\theta$. We did not use histogram $\mathbf{h}$ similarities in our experiments, as they do not improve performance on our data sets, yet are relatively expensive to compute. Regarding the coefficients $\alpha$, our primary concern was avoiding numerical underflow; we set $\alpha_\sigma = \alpha_\theta = 1$ and $\alpha_\mathbf{x} = 10/(\text{mean training exemplar area})$.

For each graph $G_{R,C}$, we can generate its Laplacian matrix

$$\begin{aligned} \mathcal{L} &= I - DAD \quad (4) \\ A_{i,j} &= w_{\mathbf{d}_i,\mathbf{d}_j} \\ D_{i,j} &= \begin{cases} (\sum_k w_{\mathbf{d}_i,\mathbf{d}_k})^{-1/2} & i = j \text{ and } \sum_k w_{\mathbf{d}_i,\mathbf{d}_k} \neq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where $I$ is the identity matrix [14]. Applying an eigendecomposition algorithm to $\mathcal{L}$ yields $\boldsymbol{\lambda}$, its vector of eigenvalues in ascending order. The number of eigenvalues is equal to the number of descriptors in $\mathcal{D}_R$, which we denote by $|\mathcal{D}_R|$. Since each connectivity function $C_\mathbf{x}, C_\sigma$ and $C_\theta$ gives rise to one graph, we may concatenate the $\boldsymbol{\lambda}$'s resulting from each graph, $\boldsymbol{\lambda}_\mathbf{x}$, $\boldsymbol{\lambda}_\sigma$ and $\boldsymbol{\lambda}_\theta$, to obtain the *region spectrum*

[1] The SIFT computes $\sigma$ as $2^{o_i + o_f}$, where $o_i$ and $o_f$ are the integer "octave" and fractional "interval" described in [9].

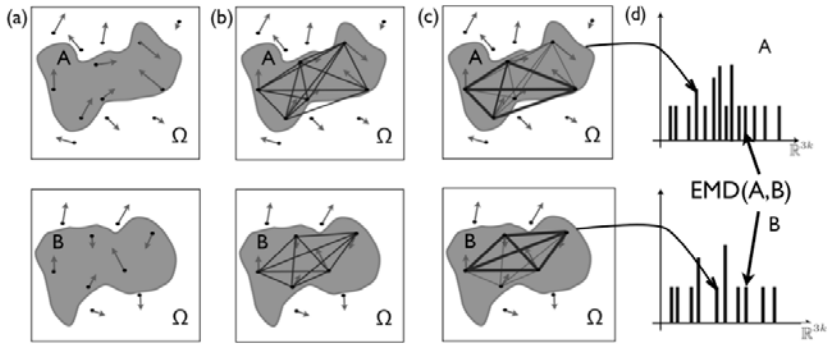


**Fig. 1.** Example illustration of the steps involved in generating geometric global features. (a) Descriptors of two regions A and B are generated (e.g. using SIFT), (b) Graphs $G_{A,C}$ and $G_{B,C}$ are formed by connecting descriptors within the regions A and B. In this example, $G_{A,C}$ has six vertices while $G_{B,C}$ has five vertices. (c) Sample $k=4$-vertex subsets from each graph to form subgraphs. (d) Each sample of 4 vertices contributes one bin to its region's EMD signature. The distance between regions A and B is the EMD between their signatures.

$\mathbf{s}_R = (\boldsymbol{\lambda}_\mathbf{x}, \boldsymbol{\lambda}_\sigma, \boldsymbol{\lambda}_\theta)$, which serves as a global feature vector for regions. Some properties of $\boldsymbol{\lambda}$ bear mentioning. First, $\boldsymbol{\lambda}$ is bounded in $[0,2]^{|\mathcal{D}_R|}$, and the smallest eigenvalue $\lambda_1 = 0$. Moreover, $\boldsymbol{\lambda}$ bounds invariants of $G_{R,C}$ such as the graph diameter, distances between subgraphs, and random walk stationary distribution convergence times [14]. These properties capture the collective behavior of the set of local features in an image region $R$, hence turning local features into a geometric global feature $\mathbf{s}_R \in [0,2]^{3|\mathcal{D}_R|}$. This feature's notion of region similarity is near-cospectrality, or similarity in Laplacian eigenspectra. Region spectra $\mathbf{s}_R$ have several notable properties:

1. $\mathbf{s}_R$ *is rotation invariant if the local feature detector is rotation invariant, and the connectivity function* $C(\mathbf{d}_i, \mathbf{d}_j)$ *remains invariant when the same rotation is applied to* $\mathbf{d}_i$ *and* $\mathbf{d}_j$*.*
   This is because $C$'s satisfying the latter condition generate isomorphic graphs for any rotation of $R$, while isomorphic graphs have identical eigenspectra. The three SIFT connectivity functions $C_\sigma$, $C_\mathbf{x}$ and $C_\theta$ are rotation invariant.
2. $\mathbf{s}_R$ *is scale invariant for* $C$*'s and local feature detectors exhibiting scale-invariance.*
   The function $C_\theta$ is scale invariant, while $C_\sigma$ becomes scale-invariant if $\sigma_i$ and $\sigma_j$ are substituted with $o_i = \log_2 \sigma_i$ and $o_j = \log_2 \sigma_j$. $C_\mathbf{x}$ can also be made scale invariant by substituting $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ with $\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_i \sigma_j}$. In practice however, the non-scale-invariant versions of $C_\mathbf{x}$ and $C_\sigma$ (Eq. (1),(2)) yielded better results on our data sets.
3. *The equivalence class of graphs with the same (or similar) eigenspectra is not limited to isomorphisms.*

Region spectra thus enable generalization in ways that are meaningful, yet not immediately obvious from a visual standpoint.

## 3 Comparing Region Spectra

Despite their advantages, region spectra require further modifications to make them a suitable feature for machine learning. The chief difficulty arises when trying to compare $\mathbf{s}_R$'s with differing $|\mathcal{D}_R|$ and consequently with different vector lengths. This is an issue because many local feature detectors (SIFT included) do not produce a fixed number of descriptors per region.

### 3.1 Region Signatures

The eigenvalues are related to distances between subgraph partitions of $G_{R,C}$ [14], which are incomparable for graphs with small vertex sets of different sizes due to discretization effects. Hence the eigenvalues of two regions should not be compared directly, except when they contain the same number of descriptors.

To compare regions with different descriptor counts, we make use of $k$-descriptor combinations from the $|\mathcal{D}_R|$ descriptors in $R$. There are $\binom{|\mathcal{D}_R|}{k}$ combinations, and for each combination we generate the induced subgraph $S_i \subseteq G_{R,C}, i = 1, \cdots, \binom{|\mathcal{D}_R|}{k}$, followed by the three $k$-dimensional eigenvalue vectors $\mathbf{s}_R^{k,i} = (\boldsymbol{\lambda}_{\mathbf{x}}^i, \boldsymbol{\lambda}_{\sigma}^i, \boldsymbol{\lambda}_{\theta}^i),\ i = 1, \cdots, \binom{|\mathcal{D}_R|}{k}$ corresponding to connectivity functions $C_{\mathbf{x}}, C_{\sigma}$ and $C_{\theta}$. Next, we construct an Earth Mover's Distance (EMD) [15] signature $\mathcal{S}_R^k : [0,2]^{3k} \mapsto \mathbb{R}$ from the $\binom{|\mathcal{D}_R|}{k}$ spectra $\mathbf{s}_R^{k,i}$. Each spectrum $\mathbf{s}_R^{k,i}$ gives rise to one bin in $[0,2]^{3k}$, and the squared Euclidean distance is used as the ground distance between bins. We refer to the EMD signatures $\mathcal{S}_R^k$ as *region signatures*.

In practice, calculating all $\binom{|\mathcal{D}_R|}{k}$ region spectra will often be computationally prohibitive. We alleviate this problem by estimating the EMD signatures $\mathcal{S}_R^k$, i.e. we randomly sample some number of subsets as a function[2] of $|\mathcal{D}_R|$ and $k$; random sampling has been shown to be effective in estimating true distributions [16]. Subset sampling and region signature creation are illustrated in Fig. 1(c,d).

### 3.2 Restoring Information with Neumann Eigenvalues

The eigenvalues of a $k$-descriptor subset's induced subgraph $S \subseteq G_{R,C}$ do not capture all information encoded in $G_{R,C}$. However, we can restore some information from $G_{R,C}$ by considering the *Neumann eigenvalues* of $S$. These eigenvalues arise from boundary conditions analogous to a "Neumann random walk", in which agents moving to a vertex $v \notin S$ immediately move to some neighbor of $v$ in $S$ — in other words, they "reflect" off the subgraph boundary [14]. Neumann eigenvalues therefore incorporate information from descriptors near to the $k$-descriptor subset. Since our graphs $G_{R,C}$ are fully connected, using the

[2] $10|\mathcal{D}_R|/k$ subsets for a given $R$ in our experiments.

Neumann eigenvalues of $S$ restores the information encoded in $G_{R,C}$ to a significant extent. Intuitively, few subset samples will be needed to approximate the full signature on all $\binom{|\mathcal{D}_R|}{k}$ subsets. The Neumann eigenvalues are obtained via eigendecomposition of the modified Laplacian [14]:

$$\begin{aligned}
\mathcal{L}_N &= D_{k\times k} I_{k\times n} (T - A) \mathcal{N} D_{k\times k} \qquad (5)\\
T_{i,j} &= \begin{cases} \sum_{h=1}^{n} w_{\mathbf{d}_i,\mathbf{d}_h} & i = j \\ 0 & \text{otherwise} \end{cases}\\
\mathcal{N}_{i,j} &= \begin{cases} 1 & i = j \\ 0 & i \le k \text{ and } i \ne j \\ w_{\mathbf{d}_i,\mathbf{d}_j} / \sum_{h=1}^{k} w_{\mathbf{d}_i,\mathbf{d}_h} & \text{otherwise, i.e. } i > k \end{cases}
\end{aligned}$$

where vertices have been relabeled so that the $k$ chosen descriptors correspond to matrix indices $\{1, \ldots, k\}$. $T$ is an $n \times n$ diagonal matrix of vertex degrees, $\mathcal{N}$ is an $n \times k$ matrix that redistributes edge weights according to the Neumann random walk, and $A$ is as defined in (4) (taking into account the relabeling). $D_{k\times k}$ is $D$ from (4) (relabeled) but truncated to the upper left $k \times k$ block, while $I_{k\times n}$ denotes a $k \times n$ matrix with 1's on the main diagonal and 0's everywhere else.

## 4 Pattern Detection Using Region Signatures

We now describe our machine learning framework for pattern detection using region signatures. Given a set of images $\mathcal{I}$, our framework ranks them according to its confidence that the pattern of interest is present, and also provides the approximate center of each detection.

A brief summary of our method follows. We begin by training a $\nu$-SVM classifier [17] on user-provided training *exemplars* — image regions with the pattern of interest — as well as non-exemplars, which can be provided or obtained from a bootstrapping procedure that will be discussed shortly. For each image $\Omega \in \mathcal{I}$, we generate a series of random region masks and subject the mask regions to the $\nu$-SVM classifier, which outputs 1 for the exemplar class and 0 for the non-exemplar class. The classification outcomes are averaged for each pixel to generate a *score landscape* (Figs. 3,4), whose local maxima are detected and sorted in descending score order. By ranking the local maxima lists for all images in lexical descending order, we obtain the required detection confidence ranking for $\mathcal{I}$. Moreover, the local maxima coordinates locate detected patterns of interest.

### 4.1 Pattern Detection without Segmentation

In order to detect patterns, we could segment each image $\Omega \in \mathcal{I}$ into regions and classify them using the SVM. However, this requires a segmentation algorithm that *a priori* separates the pattern of interest from other image regions. Finding such an algorithm can be a difficult task onto itself.

Instead of segmentation, we employ a general technique we term *masking* that is similar to sliding-window techniques. Given an image $\Omega \in \mathcal{I}$, we start by generating a set of *masks* $\mathcal{M}$, where each mask $M \in \mathcal{M}$ is a set of randomly-generated regions $R \subseteq \Omega$. For each mask $M$, we compute region signatures $\mathbf{s}_R$ for every $R \in M$. The $\mathbf{s}_R$ are then classified in $\{0, 1\}$ via the trained SVM, where 1 indicates the class of exemplars and 0 otherwise. In our experiments, we generated enough masks to cover each pixel $\geq 50$ times. Finally, we generate $\Omega$'s *score landscape* $\Phi_\Omega(\mathbf{x})$ (Figs. 3,4), the average classification of all random mask regions $R$ covering pixel $\mathbf{x}$. Local maxima $\mathbf{x}_m$ of $\Phi_\Omega$ then correspond to detected patterns of interest, where $\Phi_\Omega(\mathbf{x}_m)$ is the detection confidence, a value in $[0, 1]$ with 1 representing perfect confidence and 0 representing no confidence.

Since the score landscape may have numerous local maxima, we employ a DBSCAN-like [18] algorithm to group local maxima into significant clusters. Each cluster is assigned a score equal to the greatest maxima inside it, and clusters are sorted in descending score order to get a *maxima list* for $\Omega$. By sorting all image maxima lists in descending lexical order, we obtain a ranking for all images in $\mathcal{I}$. Comparing this ranking with the ground truth then gives a Receiver Operating Characteristic (ROC) curve.

Ideally, the randomly-generated regions should have shapes that fully contain the pattern of interest, yet contain little in the way of other regions. In our experiments, the patterns of interest are either cell-shaped or fit reasonably well into cell-shaped regions. Thus, we employed the following procedure to generate masks with cell-shaped regions:

1. Initialize a blank image with larger dimensions than $\Omega$.
2. Place random seeds on the image, and convolve with a Gaussian filter.

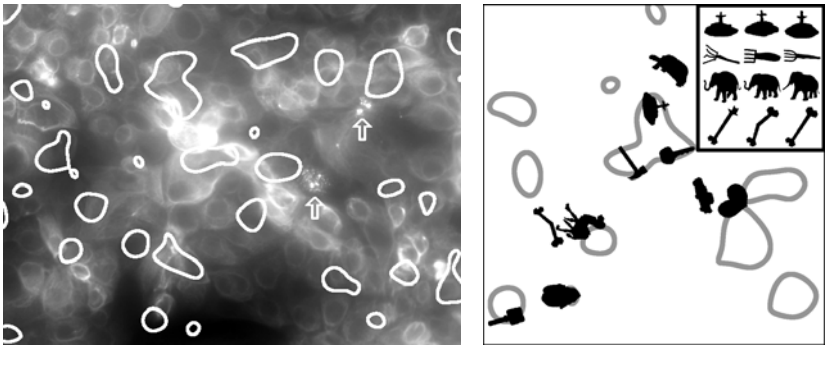

**Fig. 2.** Example images and masks from the MPEG-7 and skin cell image sets. LEFT: Skin cell image (resolution 696x520) with example mask region overlaid (parameters $(m_\sigma, m_p, m_t) = (16.0, 10^{-3}, 0.9)$). Two exemplars of the phenotype of interest (Keratin-14 aggregates) are indicated by arrows. RIGHT: MPEG-7 montage image (resolution 1000x1000), with example mask region overlaid (parameters $(m_\sigma, m_p, m_t) = (40.0, 2.5^{-4}, 0.9)$). The upper right insert shows the 4 classes to be detected, with 3 exemplars each.

3. Threshold at some (high) percentile, which generates connected components.
4. Place an $\Omega$-sized window at the image center, and extract all connected components in the window. By overlaying these connected components onto $\Omega$, we obtain a mask $M$.

Three parameters are involved: Gaussian filter standard deviation in pixels $m_\sigma$, seed distribution defined as the independent probability $m_p$ that a given pixel will become a seed, and threshold percentile $m_t$. In our experiments, we adjust them so that the mean region size approximately matches training exemplar region sizes; refer to Fig. 2 for example masks used in each experiment.

Some regions may have too few descriptors to take $k$-descriptor samples. These regions do not contribute to the score landscape $\Phi_\Omega$, except for regions with exactly zero descriptors (such as blank or low-detail regions of the image) — these are assumed to have classification 0. If the proportion of non-contributing regions for some location $\mathbf{x}$ exceeds 0.8, we set $\Phi_\Omega(\mathbf{x})$ to 0. Such locations are not being sampled adequately, hence they are ambiguous and should not be considered during maxima detection.

### 4.2 Classifying Regions Using Support Vector Machines

We employ a $\nu$-SVM classifier [17] on EMD signatures with $K(\mathcal{S}_{R_1}, \mathcal{S}_{R_2}) = \exp(-\gamma E(\mathcal{S}_{R_1}, \mathcal{S}_{R_2}))$ as the kernel function, where $\mathcal{S}_{R_1}$ and $\mathcal{S}_{R_2}$ are region signatures, $\gamma > 0$ is a scale parameter, and $E$ is the EMD function with the squared Euclidean distance $\|\mathbf{b}_1 - \mathbf{b}_2\|^2$ as the ground distance[3] between signature bins $\mathbf{b}_1$ and $\mathbf{b}_2$. Although we cannot prove that $K(\mathcal{S}_{R_1}, \mathcal{S}_{R_2})$ is positive semidefinite, we did compute the spectra for a large sample of kernel matrices from our experiments, and did not find any negative eigenvalues. We also note that a similar kernel was used in [7], but with the unmodified Euclidean distance (rather than its square) as the ground distance.

The SVM is trained with exemplar regions containing the pattern of interest labeled as class 1, and non-exemplar regions labeled as class 0. If the user does not provide non-exemplar regions, they may be generated by *bootstrapping*, in which images known to lack the pattern of interest are *masked* (Section 4.1).

## 5 Experiments

We demonstrate our framework's performance on two image sets: the 216-image MPEG-7 CE Shape-1 Part-B database subset used by Sebastian *et al.* in [19], and a 304-image subset of the GFP-Keratin-14-expressing human skin cell fluorescence microscopy images used by Law *et al.* in [20]. Both image sets are single-channel. These choices reflect two distinct applications, namely shape retrieval and bio-image phenotype detection.

[3] To be specific, we use the squared *normalized* Euclidean distance, in which every bin dimension is rescaled to have standard deviation 1.0 (over all training data). This allows us to use $\gamma = 1.0$ as a reasonable starting point for parameter tuning. Another possibility would be to use the squared Mahalanobis distance.

Note that we did not use the MPEG-7 image subset as-is; instead we generated 29 montage images with 5-10 randomly placed shape images in each (Fig. 2). None of the 216 images were used more than once. Furthermore, some of the shapes overlap in the montage images. Using montages rather than invididual shapes demonstrates the functionality of our masking technique.

The common experimental setup was as follows:

1. Annotate exemplar regions with the pattern of interest, in the form of closed polygons. Skin cell image exemplars required the consensus of 5 individuals.
2. Divide the images for $k$-fold cross-validation — 2 folds for MPEG-7 montages, 5 for skin cell images.
3. Train the $\nu$-SVM classifier using the exemplar regions as class 1, while bootstrapping (Section 4.2) to obtain non-exemplar regions as class 0.
4. Apply masking (Section 4.1) to score and rank images. We adjusted masking parameters independently for both experiments (Fig. 2).
5. Use the rankings to compute ROC curves and Area Under Curves (AUCs).

### 5.1 MPEG-7 Results

The MPEG-7 subset contains 18 classes of 12 shapes each, which we randomly placed into 29 1000x1000 montage images. We chose 4 shape classes for retrieval: fountains, forks, elephants and bones (Fig. 2). For each class 11-12 training

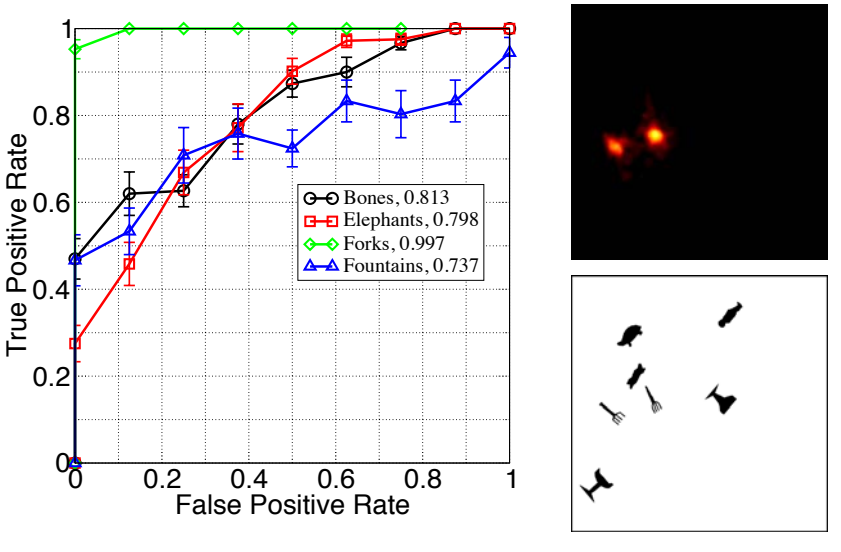


**Fig. 3.** LEFT: MPEG-7 database subset — average ROC curves and AUCs for detecting fountains, forks, elephants and bones. The curves may be non-monotonic due to binning effects on the small image set. RIGHT: Fork detection example score landscape (top) and original image (bottom). In the score landscape, black represents score 0, white represents score 1.

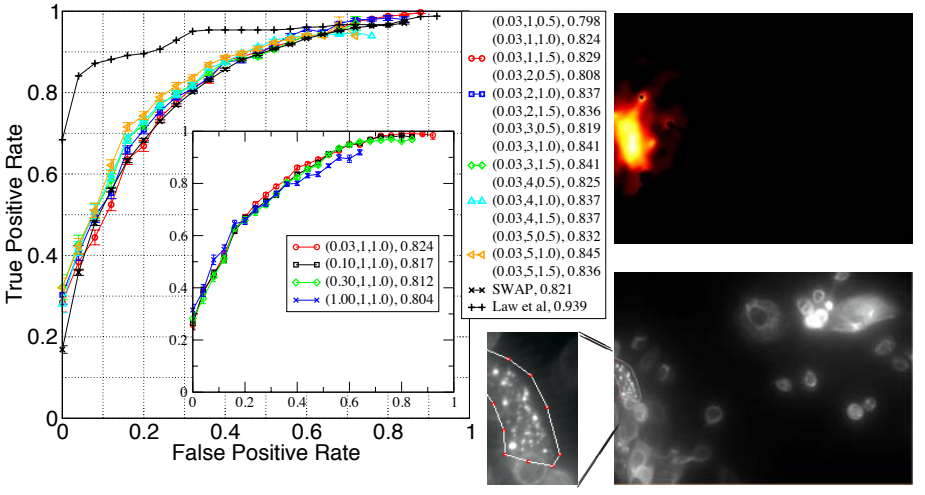


**Fig. 4.** LEFT: Skin Cell image set — average ROC curves and AUCs for detecting Keratin-14 aggregates. Legends for each ROC curve are given as "$(\gamma, r, \alpha)$, AUC". **GRAPH INSET:** Tuning over $\gamma = \{1.00, 0.30, 0.10, 0.03\}$, $r = 1$ and $\nu = 0.5$. **MAIN GRAPH:** Tuning over $\gamma = 0.03$, $r = \{1, 2, 3, 4, 5\}$ and $\nu = \alpha/(1 + r)$, $\alpha \in \{0.5, 1.0, 1.5\}$. Also shown are "SWAP", a training/validation-set-exchanged 5-fold cross-validation averaged over 5 trials with $\gamma = 0.03$, $r = 5$, and $\nu = 1/6$, and "Law et al.", the results of Law *et al.*'s deterministic aggregate-specific detector [20]. RIGHT: Example score landscape (top) and original image (bottom) with annotated exemplar. In the score landscape, black represents score 0, white represents score 1.

exemplars were annotated[4], and the montages divided into 2 folds with 5-6 exemplars each. As all exemplars contained 20 to ∼50 descriptors each, we set the number of descriptors per subset $k = 20$. The SVM was trained with a 1 : 5 ratio of exemplars to non-exemplars (i.e. we bootstrapped 5 non-exemplars for every training exemplar), $\nu = 1/6$ and $\gamma = 0.03$.[5] Masking was carried out with $(m_\sigma, m_p, m_t) = (40.0, 2.5 \times 10^{-4}, 0.9)$ (Fig. 2).

Fig. 3 shows the ROC curves and AUC for each class, averaged over 10 trials of 2-fold cross validation. We took timings using one core of a 3.0GHz Intel Core 2 system running x86-64 Linux; training and bootstrapping took ∼10 seconds per fold, while each image's score landscape took 4-5 minutes to generate. The time spent on maxima finding was negligible in comparison to score landscape generation.

### 5.2 Skin Cell Results

The phenotype (pattern) of interest was Keratin-14 aggregates, which manifest as fields of bright dots (Fig. 2). The image set contained 304 single-channel

[4] We omitted occluded shapes from the exemplar set.
[5] These parameters were selected based on Section 5.2's tuning.

images of 696x520 resolution, of which 152 had at least one exemplar. There were 193 exemplar regions in total, containing $25.3 \pm 17.4$ descriptors on average with significant right skew in the distribution. Based on this, we set the number of descriptors per subset $k = 10$, which required us to discard 19 exemplar regions containing $\leq 10$ descriptors. The masking parameters were $(m_\sigma, m_p, m_t) = (16.0, 10^{-3}, 0.9)$ (Fig. 2). ROC curves and AUCs are averaged over 10 trials of 5-fold cross-validation, unless otherwise indicated.

We tuned the SVM kernel width $\gamma$ for values $\{1.00, 0.30, 0.10, 0.03\}$, using $\nu = 0.5$ and a $1:1$ ratio of exemplars to bootstrapped non-exemplars. Having identified $\gamma = 0.03$ as yielding the optimal AUC, we proceeded to tune the bootstrap ratio and $\nu$. First, we selected ratios of the form $1:r$, $r \in \{1, 2, 3, 4, 5\}$. Then, since $\nu$ upper-bounds the fraction of training set outliers [17], for each $r$ we selected $\nu = \alpha/(1 + r)$ for $\alpha \in \{0.5, 1.0, 1.5\}$. Choosing $\nu$ in this manner protects the smaller exemplar set from over-penalization by the $\nu$-SVM's regularization.

All AUCs and some ROC curves from our tuning are shown in Fig. 4. To demonstrate the effect of a small training set, we include a 5-fold cross-validation with training and validation sets exchanged. We also include results for Law *et al.*'s spot detector [20]. We emphasize that their detector is application-specific, whereas our method can be trained to recognize arbitrary patterns. Our worst runtimes are from the $1:5$ exemplar-to-non-exemplar ratio: training and bootstrapping took $\sim$4 min/fold, while score landscapes took $\sim$4.5 minutes each. Again, maxima finding takes negligible time compared to score landscape generation. In comparison, the method of Jones *et al.* requires $\sim$2.5 minutes to preprocess each 3-channel 512x512 image, on a 2.4GHz Intel CPU [5].

## 6 Conclusion

We have described a global geometric image feature for pattern retrieval called a *region signature*. This feature derives the collective behavior of local image descriptors from graphs of their differences. By utilizing the EMD as a distance measure between region signatures, the latter can be used with $\nu$-SVM classifiers and image *masking* to perform pattern detection without segmentation. Our framework demonstrates good performance on synthetic shapes and real biological images — in particular, it retrieves patterns with only tens of local descriptors, a quantity far smaller than typically used [12,7]. We also note that the EMD distance allows region signatures to be employed in image clustering.

## Acknowledgements

We would like to thank Ivy Yan Nei Law, Yudistira Mulyadi, Boyang Zheng and Peili Yu for assistance related to this work.

## References


1. Yarrow, J.C., Feng, Y., Perlman, Z.E., Kirchhausen, T., Mitchison, T.J.: Phenotypic Screening of Small Molecule Libraries by High Throughput Cell Imaging. Combinatorial Chemistry & High Throughput Screening 6, 279–286 (2003)
2. Bakal, C., Aach, J., Church, G., Perrimon, N.: Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology. Science 316, 1753–1756 (2007)
3. Peng, H.: Bioimage informatics: a new area of engineering biology. Bioinformatics 24, 1827–1836 (2008)
4. Boland, M.V., Markey, M.K., Murphy, R.F.: Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. Cytometry 33, 366–375 (1998)
5. Jones, T.R., Carpenter, A.E., Lamprecht, M.R., et al.: Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. PNAS 106, 1826–1831 (2009)
6. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Annals of Statistics 28, 337–374 (2000)
7. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. IJCV 73, 213–238 (2007)
8. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. PAMI 19, 530–535 (1997)
9. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60, 91–110 (2004)
10. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. PAMI 27(8), 1265–1278 (2005)
11. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
12. Everingham, M., Zisserman, A., Williams, C.K.I., Gool, L.V.: The 2006 pascal visual object classes challenge (voc2006) results. Technical report, University of Oxford (2007)
13. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. PAMI 27, 1615–1630 (2005)
14. Chung, F.R.K.: Spectral Graph Theory. American Mathematical Society, Providence (1997)
15. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. IJCV 40, 99–121 (2000)
16. Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer, Heidelberg (2002)
17. Schlkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. Neural Computation 12, 1207–1245 (2000)
18. Ester, M., Kriegel, H.-p., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: International Conference on Knowledge Discovery and Data Mining, pp. 226–231. AAAI Press, Menlo Park (1996)
19. Sebastian, T., Klein, P., Kimia, B.: Recognition of shapes by editing their shock graphs. PAMI 26, 550–571 (2004)
20. Law, Y.N., Ogg, S., Common, J., Tan, D., Lane, E.B., Yip, A.M., Lee, H.K.: Automated protein distribution detection in high-throughput image-based sirna library screens. Journal of Signal Processing Systems 55, 1–13 (2009)