

The Representation and Matching of Images Using Top Points

M. Fatih Demirci · Bram Platel · Ali Shokoufandeh ·
Luc L.M.J. Florack · Sven J. Dickinson

Published online: 28 May 2009
© Springer Science+Business Media, LLC 2009

Abstract In previous work, singular points (or top points) in the scale space representation of generic images have proven valuable for image matching. In this paper, we propose a novel construction that encodes the scale space description of top points in the form of a directed acyclic graph. This representation allows us to utilize coarse-to-fine graph matching algorithms for comparing images represented in terms of top point configurations instead of using solely the top points and their features in a point matching algorithm, as was done previously. The nodes of the graph represent the critical paths together with their top points. The edge set captures the neighborhood distribution of vertices in scale space, and is constructed through a hierarchical tessellation of scale space using a Delaunay triangulation of the top points. We present a coarse-to-fine many-to-many matching algorithm for comparing such graph-based representations. The algorithm is based on a metric-tree representation of labeled graphs and their low-distortion embeddings into normed vector spaces via spherical encoding. This is a two-step transformation that reduces the matching problem to that of computing a distribution-based dis-

tance measure between two such embeddings. To evaluate the quality of our representation, four sets of experiments are performed. First, the stability of this representation under Gaussian noise of increasing magnitude is examined. Second, a series of recognition experiments is run on a face database. Third, a set of clutter and occlusion experiments is performed to measure the robustness of the algorithm. Fourth, the algorithm is compared to a leading interest point-based framework in an object recognition experiment.

Keywords Top points · Catastrophe theory · Scale space · Graph matching · Object recognition

1 Introduction

Previous research has shown that top points (singular points in the scale space representation of generic images)¹ have proven to be valuable sparse image descriptors that can be used for image reconstruction [26, 44] and image matching [1, 27, 48]. In our previous work, images were compared using a point matching scheme which took into account the positions, scales, and differential properties of corresponding top points [3, 4, 26, 27, 47]. The underlying matching framework was based on the Earth Mover's Distance [19], a powerful, point matching framework. However, treating the points as an unstructured collection ignores the salient neighborhood relations that may exist within a given scale or across scales. Grouping certain top points together explicitly encodes the neighborhood structure of a point, effectively enriching the information encoded at a point—information

M.F. Demirci (✉)
TOBB University of Economics and Technology, Söğütözü,
Ankara, 06560, Turkey
e-mail: mfdemirci@etu.edu.tr

B. Platel · L.L.M.J. Florack
Eindhoven University of Technology, P.O. Box 513, 5600 MB
Eindhoven, The Netherlands

A. Shokoufandeh
Drexel University, Philadelphia, PA 19104, USA

S.J. Dickinson
University of Toronto, Toronto, Ontario, Canada M5S 3G4

¹The terminology is reminiscent of the 1D case, in which only annihilations occur generically.

that can be exploited during both indexing [55] and matching [6, 57].

In this paper, we take an unstructured set of top points and impose a neighborhood structure on them. Similar to the work of Lifshitz and Pizer [32], we will encode the scale space structure of a set of top points in a *directed acyclic graph* (DAG). Specifically, we combine the position-based grouping of the top points provided by a Delaunay triangulation with the scale space ordering of the top points to yield a directed acyclic graph. This new representation allows us to utilize powerful graph matching algorithms to compare images represented in terms of top point configurations, rather than using point matching algorithms to compare sets of isolated top points. Specifically, we draw on our recent work in many-to-many graph matching [12, 13, 28], which reduces the matching problem to that of computing a distribution-based distance measure between embeddings of labeled graphs. Moreover, we employ the matching framework in a coarse-to-fine manner, using coarse-level similarity to select candidate models for further (fine-level) analysis.

Following a review of related work, we describe our new construction by first elaborating on those basics of catastrophe theory required to introduce the concept of a top point in Sect. 3. Next, we formally define a top point, and introduce a measure for its stability that will be used to prune unstable top points. Section 5 describes the construction of the DAG through a Delaunay triangulation scheme. Section 6 reviews our many-to-many DAG matching algorithm, and describes how it will be applied in a coarse-to-fine search of the database. In the first experiment, we examine the stability of the framework under Gaussian noise of increasing magnitude applied to the original images. In the second experiment, we examine the invariance of the framework to within-class image deformation. Finally, in the third experiment, we examine the robustness of the framework to clutter and occlusion.

2 Related Work

There has been considerable effort devoted to multiscale image analysis. One of the earliest studies was due to Crowley and Parker [10], who proposed a representation framework in which peaks and ridges were identified in scale space and encoded in a tree. The resulting trees were then used to match images at different scales. Given a query and large database, a linear search was used to match the query tree to each database tree in order to determine the closest model to the query. In contrast, our matching algorithm achieves sub-linear search complexity through a coarse-to-fine search of the database. Moreover, our matcher accounts for the complete topological structure of a directed acyclic graph, as opposed to the *peak paths* of a tree.

Reconstruction of one-dimensional and two-dimensional signals has received significant interest from the scale space community. Hummel and Moniot [20] reconstructed signals and images from zero crossings in scale space. Johansen et al. [23] showed that it is possible to reconstruct a one-dimensional signal from its top points. Specifically, one-dimensional signals can be reconstructed using the locations of the top points for zero-crossing curves. To generalize one-dimensional signals to two-dimensional signals (images), one may replace zero-crossing curves by zero-crossing surfaces. As an alternative, the set of extrema and saddles for all values of the scale parameter (or, curves) may be taken into account. Johansen [22] studied the behavior of curves in scale space consisting of critical points, and showed how the partial derivatives at a top point could be used for top point classification. In contrast to the model proposed in this paper, Johansen's approach ignores the salient neighborhood structure of a top point, which can be useful for matching.

Previous work on scale-invariant detectors searches for local extrema in the 3D scale space representation of an image. After applying a difference of Gaussian filter to the image at different scales, a point is selected as a feature point if it is a local extremum (in both space and scale) and its response is greater than a predefined threshold. This idea originated in the work of Witkin [61] and Koenderink [29]. Existing techniques based on this paradigm differ in the differential expressions used for the construction of the scale space representation. The reader is referred to [43] for a comparison of state of the art affine region detectors.

A multiscale representation does not state what image structures are significant, what relations may exist between the structures, or what scales are appropriate for extracting the features. To address these problems, Lindeberg [34] studied the behavior of image structures over scales, measured the saliency of image structures from stability properties, and extracted relations between structures at different scales. In related work, Lindeberg utilized the Laplacian-of-Gaussian to search for 3D maxima of scale-normalized differential operators [37]. By successively smoothing the image with Gaussian kernels of different size, a scale space representation of the image was constructed. Additionally, Lindeberg [36] also presented a framework for edge and ridge detection by defining the concept of these features as one-dimensional curves in the three-dimensional scale space representation of the image. These approaches were empirically shown to be effective for selecting regions of interest with associated scale levels in a broad scope of image domains. As exemplified by our face recognition experiments on a publicly available face database, a set of occlusion experiments using real images of real objects, and a set of object retrieval experiments on a publicly available database, the proposed approach is also shown to be applicable to various image domains.

The Laplacian-of-Gaussian was shown to be suitable for automatic scale selection of image structures [35, 37]. However, the main disadvantage of this function is that it corresponds not only to blobs but to high gradients in one direction as well. As a result, localization of features may not be very accurate. One way of addressing this problem was given by Dufournoud et al. [15] who presented the scale-adapted Harris operator. Using this operator, a scale space is created in which local 3D maxima are selected as salient points. The scale-adapted Harris operator separates feature detection from scale selection. However, this operator rarely finds a maximum over scales, resulting in too few points with which to represent the image. To overcome this problem, Mikolajczyk and Schmid [42] merge the scale adapted Harris corner detector with the Laplacian-based scale selection. The resulting detector, called the Harris–Laplace detector, is extended to deal with affine transformations.

Lowe [40] proposed a framework transforming an image into a large collection of image feature vectors, which are largely invariant to changes in scale, illumination, and local affine distortions. The framework is based on local 3D extrema in the scale space pyramid obtained by difference of Gaussian filters. The algorithm uses a large number of dimensions (128) to encode features. Bay et al. [5] recently developed a scale and rotation invariant interest point detector and descriptor named *speeded up robust features* (SURF), which builds on the strengths of existing detectors and descriptors. Specifically, the SURF detector is based on the Hessian matrix, and uses Haar wavelet responses to construct the descriptor. To reduce the time required for feature computation and matching, this technique uses a lower dimensional feature descriptor (64). Given an image, our framework, on the other hand, uses only seven unary features of a point (x , y , t coordinates, dx_x , dy_y , dx_y second order derivatives at the point, and the stability) and adds hierarchical relational information to the representation. Despite the weak encoding of a point's feature, experimental evaluation of the framework, including a comparison with [40] in an object recognition experiment, demonstrates the power of explicitly encoding the relational information.

Matas et al. [41] introduced the maximally stable extremal regions, which are closed under the affine transformation of image coordinates and invariant to affine transformation of intensity. The authors show that the maximally stable extremal regions can produce large numbers of matching features with good stability. Kadir and Brady [24] proposed a different multiscale algorithm for the selection of salient regions of an image. The algorithm detects scale localized features with high entropy. Since the algorithm treats the scale as isotropic, it results in blob-like features. These techniques do not encode the relational information for the interest points. Similar to some existing approaches, such as

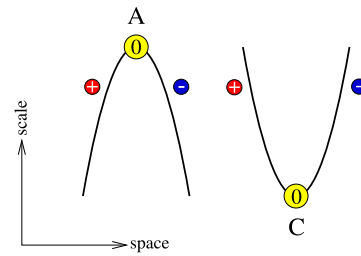


Fig. 1 The generic catastrophes in isotropic scale space. *Left*: an annihilation event. *Right*: a creation event. A positive charge \oplus denotes an extremum, a negative charge \ominus denotes a saddle, \odot indicates the singular point

[7, 53, 54, 56], one of the benefits of our framework comes from employing such information in the framework.

3 Catastrophe Theory

Critical points are points at any fixed scale in which the gradient vanishes, i.e., $\nabla u = 0$. The study of how these critical points change as certain control parameters change is called *catastrophe theory* [49, 58, 59]. A Morse critical point will move along a *critical path* when a control parameter is continuously varied. In principle, the single control parameter can be identified as the scale of the blurring filter. The only generic morsifications in Gaussian scale space are *creations* and *annihilations* of pairs of Morse hypersaddles of opposite Hessian signature² [11, 17]. An example of this is given in Fig. 1.

The movement of critical points through scale, together with their annihilations and creations, forms *critical paths* in scale space. In this article, we restrict ourselves to generic (non-symmetrical) 2D images, but the theory is easily adapted to higher dimensions. In the 2D case, the only generic morsification is an annihilation or creation where a saddle point and an extremum point meet [30, 33]. Critical paths in 2D therefore consist of an *extremum branch*, that describes the movement of an extremum through scale, and a *saddle branch*, that describes the movement of the saddle with which the extremum annihilates. Note that under mild conditions only one extremum branch continues up to infinite scale [38]. In Fig. 2, the critical paths and their top points are shown for a picture of a face.

4 Top Points

The points at which creation and annihilation events take place are often referred to as *top points*. A top point is a

²The Hessian signature is the sign of the determinant evaluated at the location of the critical point.

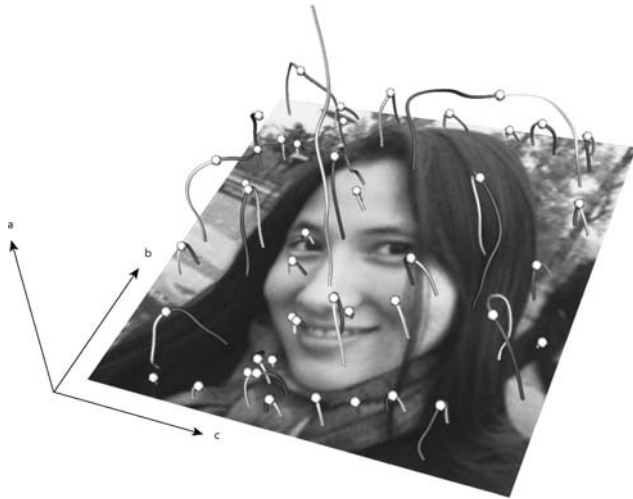


Fig. 2 Critical paths and top points of a face

critical point at which the determinant of the Hessian degenerates:

$$\begin{cases} \nabla u = 0 \\ \det(H) = 0. \end{cases} \quad (1)$$

An easy way to find these top points is by means of zero-crossings in scale space. This involves derivatives up to second order and yields sub-pixel results. Other, more elaborate methods can be used to find or refine the top point positions; for details, the reader is referred to [17, 22, 23].

It is obvious that the positions of extrema at very fine scales are sensitive to noise. This, in most cases, is not a problem. Most of these extrema are blurred away at fine scales and won't affect our matching scheme at slightly coarser scales. However, problems do arise in areas in the image that consist of almost constant intensity. One can imagine that the positions of the extrema (and thus the critical paths and top points) are very sensitive to small perturbations in these areas. These unstable critical paths and top points can continue up to very high scales since there is no structure in their vicinity to interact with. To account for these unstable top points, we need to have a measure of stability, so that we can either give unstable points a low weight in our matching scheme, or disregard them completely.

A top point is more stable in an area with a lot of structure. The amount of structure contained in a *spatial* area around a top point can be quantified by the *total quadratic variation* (TQV) norm over that area:

$$\text{TQV}(\Omega) \stackrel{\text{def}}{=} \frac{\sigma^2 \int_{\Omega} \|\nabla u(x)\|^2 dV}{\int_{\Omega} dV}. \quad (2)$$

We calculate the TQV norm in a circular area with radius $\lambda\sigma$ around a top point at position (x_c, t_c) . Note that the size

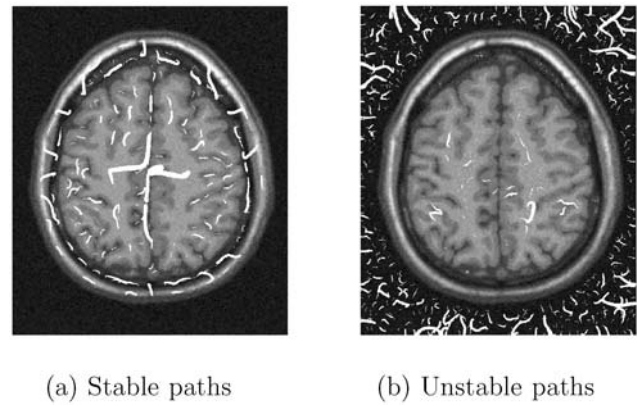


Fig. 3 Spatial projection of critical paths of a MR brain scan image. The paths are filtered by thresholding the stability norm of their top points. Most instabilities occur in flat regions, as expected

of the circle depends on the scale σ . The integration area of the TQV norm Ω is defined by:

$$\Omega : \|x - x_c\|^2 \leq \lambda^2 \sigma^2. \quad (3)$$

By using a spatial Taylor series expansion around the considered top point, and taking into account that the first order spatial derivatives at this point are zero, we can simplify the TQV-norm (2) to what we refer to as the *differential TQV-norm* by the following limiting procedure [48]:

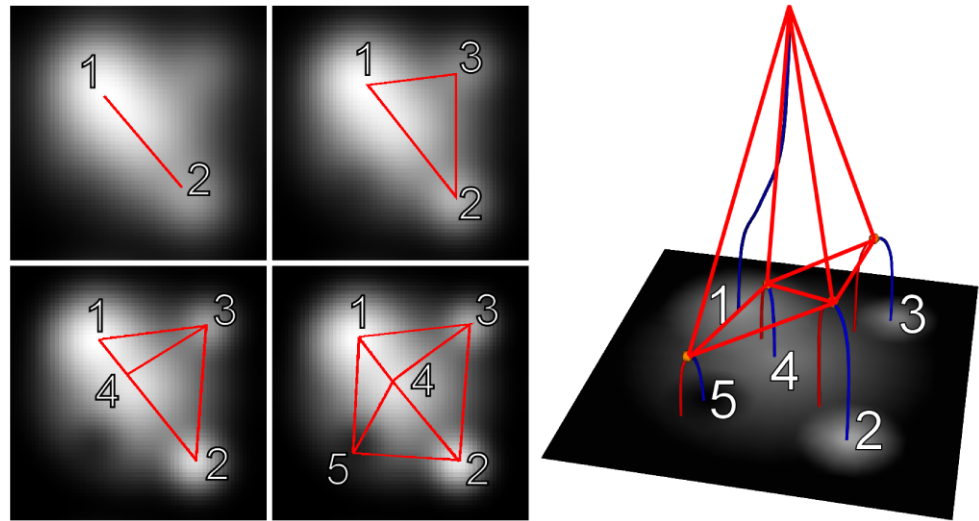
$$\text{tqv} \stackrel{\text{def}}{=} \lim_{\lambda \rightarrow 0} \frac{4}{\pi} \frac{1}{\lambda^2} \text{TQV}(\lambda) = \sigma^4 \text{Tr}(H^2). \quad (4)$$

The proportionality factor $\frac{4}{\pi}$ is irrelevant for our purposes and has been chosen for mere convenience. The normalization factor $\frac{1}{\lambda^2}$ is needed prior to evaluation of the limit since $\text{TQV}(\lambda) = \mathcal{O}(\lambda^2)$. Equation (4) has been referred to by Koenderink as *deviation from flatness*, which can indeed be seen to be the differential counterpart of (2). It enables us to calculate a stability measure for a top point *locally* by using only its second order derivatives. Alternative stability measures may also be used here, e.g., [2]. This stability norm can be used to weigh the importance of top points in our matching scheme, or to remove any unstable top points by thresholding them on their stability value. The latter is demonstrated in Fig. 3.

5 Construction of the Graph

The goal of our construction is two-fold. First, we want to encode the neighborhood structure of a set of points, explicitly relating nearby points to each other in a way that is invariant to minor perturbations in point location. Moreover, when local neighborhood structure does indeed change, it is essential that such changes will not affect the encoded

Fig. 4 Visualization of the DAG construction algorithm. *Left:* the Delaunay triangulations at the scales of the nodes. *Right:* the resulting DAG (edge directions not shown)



structure elsewhere in the graph. The Delaunay triangulation imposes a position-based neighborhood structure with exactly these properties [50]. It represents a triangulation of the points which is equivalent to the nerve of the cells in a Voronoi tessellation, i.e., that triangulation of the convex hull of the points in the diagram in which every circumcircle of a triangle is an empty interior [45]. The directed edge set of our resulting graph will be based on the edges of the triangulation. Our second goal is to capture the scale space ordering of the points to yield a directed acyclic graph, with coarser scale top points directed to nearby finer scale top points.

The first step in constructing our graph G is the detection of top points and critical paths using [25]. The root of G , denoted as v_1 , will correspond to the single critical path that continues up to infinity; note that there is no top point associated with this critical path, but simply its position at the coarsest scale. All other nodes in G , denoted as v_2, \dots, v_n , correspond to the detected top points and their corresponding critical paths. Let σ_k denote the scale of node v_k . The nodes, v_2, \dots, v_n , are ordered such that $\sigma_i > \sigma_j$, where $i < j$.

As we build the Delaunay triangulation of the points, we will simultaneously construct the DAG. Beginning with the root, v_1 , we have a singleton point in our Delaunay triangulation, and a corresponding single node in G . Next, at the scale corresponding to v_2 , we project v_1 's position down to v_2 's level, and recompute the triangulation. In this case, the triangulation yields an edge between v_1 and v_2 . Each new edge in the triangulation yields a new edge in G , directed from coarser top points to finer top points; in this case, we add a directed edge in G from v_1 to v_2 . We continue this process with each new top point, first projecting all previous top points to the new point's level, recomputing the triangulation, and using the triangulation to define new directed

edges in G . A summary of this procedure is presented in Algorithm 1.

Algorithm 1 Top point graph construction procedure

- 1: Detect the critical paths.
 - 2: Extract the top points from the critical paths.
 - 3: Label the extremum path continuing up to infinity as v_1 .
 - 4: Label the rest of the nodes (critical paths, together with their top points) according to the scale of their top points from high scale to low as v_2, \dots, v_n .
 - 5: For $i = 2$ to n evaluate node v_i :
 - 6: Project the previous extrema into the scale of the considered node v_i .
 - 7: Calculate the 2D Delaunay triangulation of all the extrema at that scale.
 - 8: All connections to v_i in the Delaunay triangulation are stored as directed edges in G .
-

The construction is illustrated for a simple image in Fig. 4. In the top two frames in the left figure, we show the transition in the triangulation from v_2 (point 2) to v_3 (point 3); the root is shown as point 1. In the upper right frame, the triangulation consists of three edges; correspondingly, G has three edges: $(1, 2), (1, 3), (2, 3)$, where (x, y) denotes an edge directed from node x to node y . In the lower left figure, point 4 is added to the triangulation, and the triangulation recomputed; correspondingly, we add edges $(1, 4), (2, 4), (3, 4)$ to G (note that $(1, 2)$ is no longer in the triangulation, but remains in G). Finally, in the lower right frame, point 5 is added, and the triangulation recomputed. The new edges in the triangulation yield new edges in G : $(2, 5), (4, 5), (1, 5)$. The right side of Fig. 4 illustrates the resulting graph (note that the directions of the edges are not shown). Figure 5 is the result of applying this construction to the face of Fig. 2.

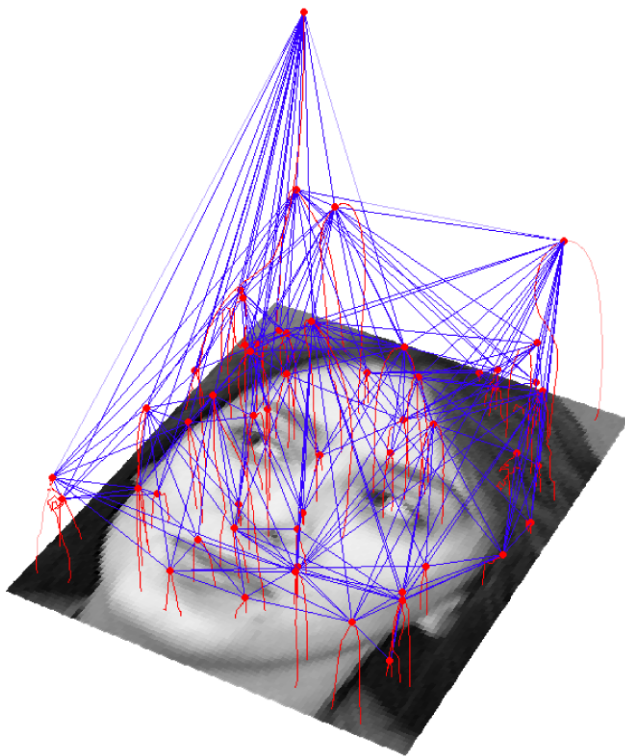


Fig. 5 The DAG obtained from applying Algorithm 1 to the critical paths and top points of a face image

6 Many-to-Many Matching Algorithm

Given a pair of DAGs, our next objective is to compute their similarity using both vertex and edge attributes. Our matching algorithm is based on the low-distortion embeddings of labeled graphs into normed vector spaces via spherical coding [13, 28]. The advantage of this embedding technique stems from the fact that it prescribes a single vector space into which both graphs are embedded. In order to achieve this embedding, we first have to choose a suitable metric for our graphs, i.e., we must define a distance between any two vertices. Since we are dealing with a scale space of a 2D image, our distance function should depend on both the scale of a point and its spatial position in scale space.

Eberly [16] defined such a metric that depends on a parameter $\rho > 0$, weighting the relative importance of spatial and scale values. The metric is given below:

$$ds^2 = \sum_{j=1}^n \frac{dx_j^2}{\sigma^2} + \frac{1}{\rho^2} \frac{d\sigma^2}{\sigma^2}, \quad (5)$$

where σ represents the scale of a point and x_i is the spatial position of the point in scale space. After solving the differential equations as shown in [60], the distance S between two points (x_1, y_1, σ_1) and (x_2, y_2, σ_2) in scale space can be

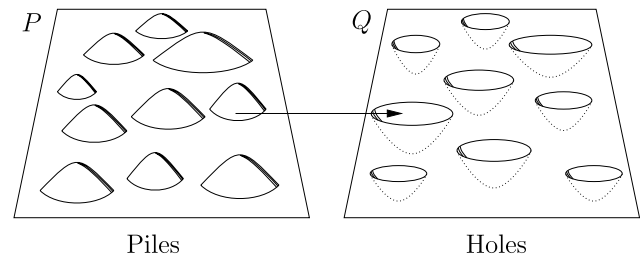


Fig. 6 Computing the minimum amount of work it takes to transform one distribution (P) into another (Q) by the Earth Mover's Distance algorithm

defined as:

$$\begin{aligned} S &= (r_2 - r_1) / \rho, \\ r_i &= \ln(\sqrt{b_i^2 + 1} - b_i), \quad i = 1, 2, \\ b_i &= \frac{\sigma_2^2 - \sigma_1^2 - (-1)^i \rho^2 R^2}{2\sigma_i \rho R}, \quad i = 1, 2, \end{aligned} \quad (6)$$

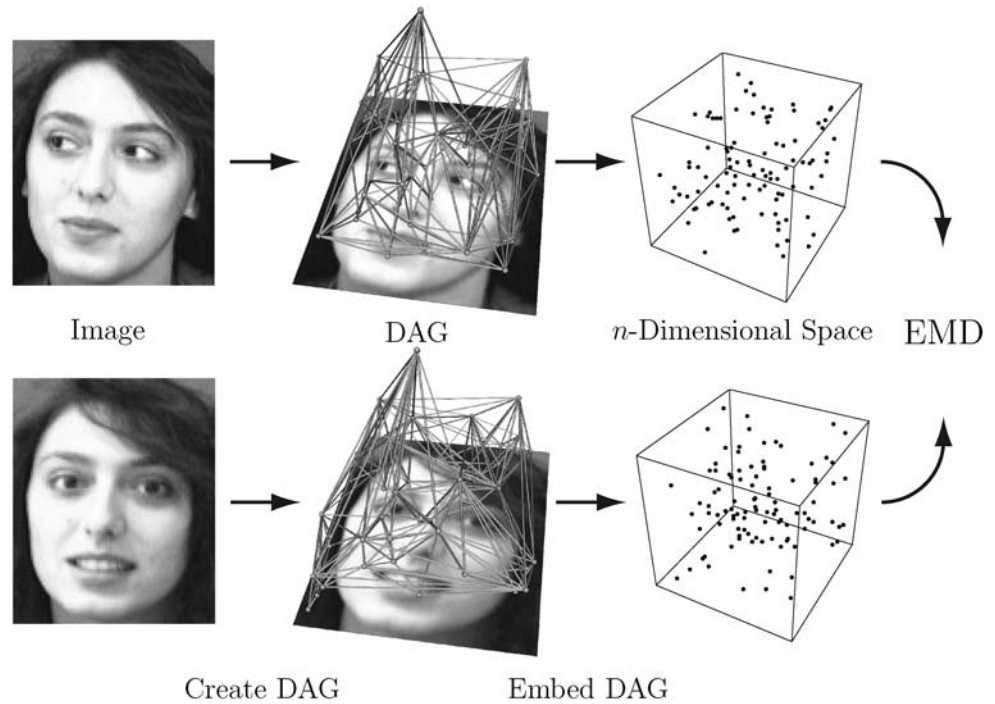
where R is the Euclidean distance between (x_1, y_1) and (x_2, y_2) .

Embedding graphs into normed vector spaces reduces the many-to-many matching problem in the graph domain to that of computing a distribution-based distance measure between two points sets. To compute such a distance, we use a variation of the Earth Mover's Distance (EMD) under transformation [9]. The EMD approach computes the minimum amount of work (defined in terms of displacements of the masses associated with points) it takes to transform one distribution into another (Fig. 6). For two given graphs, the algorithm provides an overall similarity (distance) measure.

Figure 7 presents an overview of the approach. For a given image, we first create its DAG according to Sect. 5, and embed each vertex of the DAG into a vector space of prescribed dimensionality using a deterministic spherical coding. Finally, we compute the distance between the two distributions by the modified Earth Mover's Distance under transformation. The dimensionality of the target space has a direct effect on the quality of the embedding. Specifically, as the dimensionality of the target space increases, the quality of the embedding will improve. Still, there exists an asymptotic bound beyond which increasing the dimensionality will no longer improve the quality of the embedding. Details on the many-to-many matching algorithm can be found in [14].

In the absence of an effective many-to-many graph indexing algorithm, the query graph must be matched to each database graph in order to determine the closest model to the query. As the database grows, such a linear search can become costly. This is further compounded if the sizes of the graphs also grow, since the many-to-many matching algorithm takes into consideration all nodes and edges (at

Fig. 7 Computing similarity between two given faces



all scales) of two graphs in order to compute the similarity between them. If we are indeed constrained to a linear search of the database, we can reduce the complexity of our search task by effectively reducing the complexity of the graphs through a coarse-to-fine search. Specifically, if we assume that a query and a particular model are dissimilar with respect to their coarser-scale features, then determining the similarity of their finer-scale features is fruitless. Distinguishing between the coarser and the finer features is facilitated by the hierarchical structure of the top point directed acyclic graph construction.

This coarse-to-fine search of the model database effectively represents a breadth-first search. Given a query DAG, we first match its root (coarsest node) with the root of each database DAG. The similarities between the query and each root can be sorted in decreasing similarity. If the difference in similarity between the top two most similar models (to the query) is large, the algorithm terminates by selecting the most similar model, and avoids the complexity of matching the finer scales of that or any other model. However, if the difference is small, then *all* models whose similarity exceeds a threshold are considered candidates for further examination at finer scales. We add a level (forming the union of the root plus the new level) to each of the surviving candidates (and to the query) and repeat the procedure. As a new level is added, the contribution of the new nodes to the distance function is increased relative to the previously matched nodes, in order to take maximum advantage of the disambiguating information. This iterative

process of pruning candidates is repeated until a “winner” emerges.

7 Experiments

To evaluate our construction, we explore the invariance of the construction to three types of perturbations. The first is the sensitivity of the construction to noise in the image, the second is within-class deformation resulting in displacements of top points both within and across scales, and the third is robustness to clutter and occlusion. We conduct the first two experiments using a subset of the Olivetti Research Laboratory face database [46], while the third experiment contains cluttered, occluded scenes imaged in our laboratory. In a fourth set of experiments, our approach is compared to a leading interest point-based framework in an object recognition experiment using a subset of a publicly available database.

Invariance of a graph to noise, within-class deformation, or clutter/occlusion requires a measure of graph distance, so that the *distance* between the original and perturbed graphs can be computed. For the experiments reported in this paper, we compute this distance using our many-to-many graph matching algorithm. Note that we have developed a general algorithm that is in no way specifically designed for face recognition. The proposed algorithm is more widely applicable than dedicated algorithms specifically constructed for face recognition. Therefore we have not compared our method to state-of-the-art face recognition algorithms. We present this experiment only as a proof of concept.

Fig. 8 Sample faces from 20 people

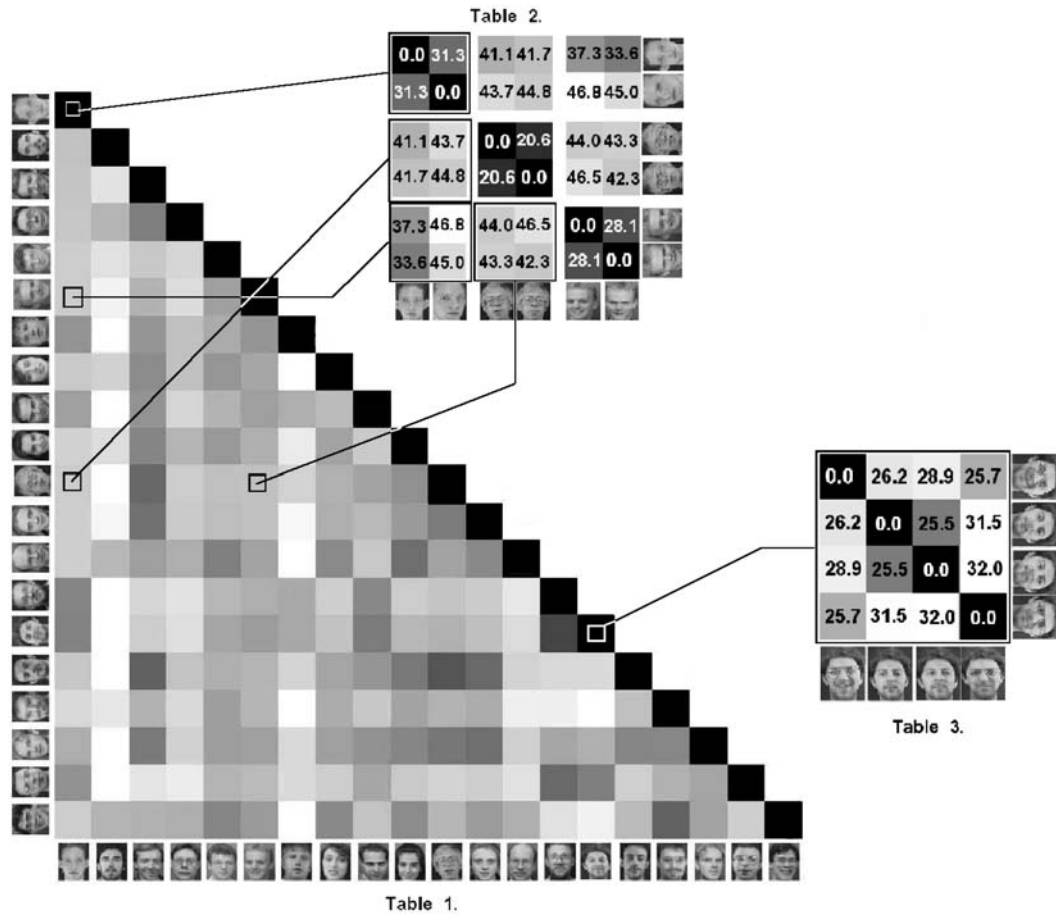


Fig. 9 *Table 1*: Matching results of 20 people. The rows represent the queries and the columns represent the database faces (query and database sets are non-intersecting). Each row represents the matching results for the set of 10 query faces corresponding to a single individual matched against the entire database. The intensity of the table entries indicates matching results, with black representing maximum similarity between two faces and white representing minimum similarity. *Table 2*: Subset of the matching results with the pairwise distances shown. *Table 3*: Effect of presence or absence of glasses in the matching for the same person. The results clearly indicate that the graph perturbation due to within-class deformation, including facial expression changes, illumination change, and the presence/absence of glasses is small compared to the graph distance between different classes

7.1 Graph Stability under Additive Noise

The face database consists of faces of 20 people with 10 faces per person, for a total of 200 images; each image in the database is 112×92 pixels. The face images are in frontal view and differ by various factors such as gender, facial expression, hair style, and presence or absence of glasses. A representative view of each face is shown in Fig. 8. While the top-point graphs (DAGs) constructed for the face data-

base have, on average, 75 levels, the coarse-to-fine matching algorithm converges in the first 30 levels, representing a very significant reduction in matching complexity.

To test the robustness of our graph construction, we first examine the stability of our graphs under additive Gaussian noise at different signal levels applied to the original face images. For this experiment, the database consists of the original 200 unperturbed images, while the query set consists of noise-perturbed versions of the database images.

Specifically, for each of the 200 images in the database, we create a set of query images by adding 1%, 2%, 4%, 8%, and 16% of Gaussian noise. Next, we compute the similarity between each query (perturbed database image) and each image in the database, and score the trial as correct if its distance to the face from which it was perturbed is minimal across all database images. This amounts to 40,000 similarity measurements for each noise level, for a total of 200,000 similarity measurements. Our results show that the recognition rate decreases down to 96.5%, 93%, 87%, 83.5%, and 74% for 1%, 2%, 4%, 8%, and 16% of Gaussian noise, respectively. These results indicate a graceful degradation of graph structure with increasing noise.

7.2 Graph Stability under Within-Class Variation

To test the stability of the graph construction to within-class variation (e.g., different views of the same face), we first group the faces in the database by individual; these will represent our categories. Next, we remove the first image (face) from each group and compare it (the query) to all remaining database images. The image is then put back in the database, and the procedure is repeated with the second image from each group, etc., until all 10 face images of each of the 20 individuals have been used as a query. If the graph representation is invariant to within-class deformation, resulting from different viewpoints, illumination conditions, presence/absence of glasses, etc., then a query from one individual should match closest to another image from the same individual, rather than an image from another individual. The results are summarized in Table 1, Fig. 9.

The magnitudes of the distances are denoted by shades of gray, with black and white representing the smallest and largest distances, respectively. Due to symmetry, only the lower half of distance matrix is presented. Intra-object distances, shown along the main diagonal, are very close to zero. In addition, based on the overall matching statistics, we observed that in all but 6.5% of the experiments, the closest match selected by our algorithm belonged to the same category.

To better understand the differences in the recognition rates for different people, we randomly selected a subset of the matching results among three people in the database, as shown in Table 2, Fig. 9. Here, the (i, j) -th entry shows the actual distance between face i and face j . It is important to note that the distance between two faces of the same person is smaller than that of different people, as is the case for all query faces. In our experiments, one of our objectives was to see how various factors, such as the presence or absence of glasses, affects the matching results for a single person. Accordingly, we took a set of images from the database of one person, half with the same factor (with the presence of glasses), and computed the distances between each image

pair. Our results show that images with the same factors are more similar to each other than to others. Table 3 of Fig. 9 presents a subset of our results. As can be seen from the table, images of the same person with glasses are more similar than those of the same person with and without glasses. Still, in terms of categorical matching, the closest face always belongs to the same person.

7.3 Graph Stability under Clutter and Occlusion

To demonstrate the framework's robustness to clutter and occlusion, we turn to a set of real images of real objects under occlusion. The database images were obtained by placing six unoccluded objects, in turn, on a table and taking their pictures from a single viewpoint. Cluttered query scenes were generated by arranging one or more database objects (along with possible non-database objects) in occluded configurations, such that the viewpoint of any database object in the query scene is similar to that of its database (model) image. Given a query image, the goal is to retrieve the database objects that are at least partially visible in the scene. Table 1 shows the results of matching every query image (leftmost column) to every database image (topmost row). The entries in the table reflect distance, with the closest matching database image marked with a box.

Examination of Table 1 reveals that for each query, the closest matching database object correctly identifies the single object in the query that is unoccluded (when one exists). The exceptions are rows 9 and 12, in which all query objects are occluded to some degree; in these two cases, the closest matching object is one of the less-occluded objects. These results clearly demonstrate the framework's robustness to scene clutter. If we examine the second closest matching object for each query, we find that it represents one of the occluded objects in the query. And in most cases, the third closest matching object also represents an occluded object in the query. These results clearly demonstrate the framework's robustness to occlusion. While the top-point graphs (DAGs) constructed for the database images have, on average, 2600 levels, the coarse-to-fine matching algorithm converges in the first 500 levels, representing an even more dramatic reduction in matching complexity.

7.4 Comparison to Other Approaches

The above experiments have clearly established the framework's stability under noise, within-class variation, clutter, and occlusion. In this section, we compare our framework to SIFT, a leading interest point-based framework presented in [40]. For the comparison, we use a subset of the Amsterdam Library of Object Images (ALOI) [18]. The subset consists of 5 objects with 36 views per object. In the original ALOI database, images of the objects were taken at pose

Table 1 Occlusion results. The distance between each query and database image is shown in the table. The closest database object for each query is indicated by a box around its minimum distance



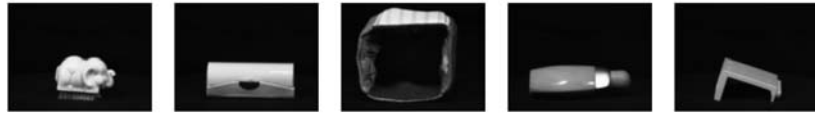
Query	Database					
						
	183.70	36.86	228.65	364.63	372.83	331.62
	201.40	56.23	210.28	344.81	352.97	323.71
	250.17	142.13	133.45	184.60	328.78	158.79
	120.67	112.43	135.44	78.32	225.91	147.86
	157.65	135.27	169.08	203.62	116.50	174.25
	163.01	157.20	143.15	209.08	130.71	197.58
	150.08	143.28	167.17	170.04	123.27	187.10
	213.01	152.17	189.07	167.18	129.76	183.75
	172.11	155.18	197.84	181.28	169.80	193.27
	185.27	178.58	201.04	192.17	139.26	128.13
	216.60	221.81	161.66	246.37	147.15	137.33
	178.13	153.10	183.78	112.22	130.57	181.73

Fig. 10 Sample images from the Amsterdam library of object images



intervals of 5 degrees. To provide a greater degree of variation between query and database image, we removed every second image from the database and used the remaining image set for our experiment; sample views from each object are shown in Fig. 10. We follow the same experimental procedure presented in Sect. 7.2. Specifically, the first shape (query) is removed from the database and compared to all remaining database shapes. The query is then put back in the database and the procedure repeated for all other database shapes. This process results in a similarity matrix in which the entry for the i -th row and the j -th column represents the similarity between the i -th and j -th shapes from the database. Using this matrix, we examined the top matches to see how many of the within-category shapes belong to the same class as the query. Our objective here is to evaluate the retrieval performance of each algorithm.

Based on the overall matching statistics, we observe that in 72.2% and 73.5% of the experiments, the closest matching shape, i.e., the nearest neighbor, retrieved by SIFT and our algorithm, respectively, belongs to the same class as the query. In a second experiment, we compute how many of the shapes in the query's class appear within the top $n - 1$ matches, where n is the size of the query's class (first tier [52]). This experiment examines the stability of the representation to within-class variation for content-based retrieval tasks. Using SIFT, this number is 91.4%, while our algorithm achieves 94.2%. Finally, we repeat the second experiment but consider the top $2 \times n - 1$ matches (second tier [52]). Here our goal is to identify the spread of objects representing a category with respect to other classes, i.e., we want to make sure the views are not too widely spread. Both SIFT and our approach yield %100 recognition rates for this experiment.

It is important to note that despite using only 7 dimensions to encode our unary features, our framework outperforms or achieves comparable rates such as SIFT, in which 128 dimensions are used to encode features. Clearly, the important hierarchical (relational) information exploited in our framework has made up for the decrease in unary information. One might expect that if we used a higher dimensional encoding of unary features, even more improvement over SIFT could be achieved at the expense of greater feature complexity.

7.5 Discussion

Typical interest point encodings used in some frameworks, such as [8, 40, 51], are highly specific and can have many dimensions (e.g., SIFT has 128 dimensions). With such strong

unary features to minimize feature ambiguity, the need for relational (e.g., binary) features to help disambiguate unary features goes down. At the other extreme, classical model-based recognition frameworks (e.g., [21, 31, 39]) based on highly ambiguous corner or contour features require strong relational (typically geometrical) constraints to help disambiguate highly ambiguous unary features.

Our experiments are aimed more at this latter end of the spectrum, in which the dimensionality of our unary features is quite low (7), with more emphasis placed on not only encoding relational information, but encoding it hierarchically to facilitate coarse-to-fine matching. The power of the relational encoding reflected in the experiments is clear, despite the much weaker encoding of a node's feature. However, the framework is flexible, for the node specificity can easily be increased (e.g., replace our 7-dimensional feature with a 128-dimensional SIFT feature) and the relational constraints relaxed (i.e., node difference induces a greater EMD penalty than embedded distance) in a more exemplar-oriented matching task. Conversely, for a more category-oriented matching task, node dimensionality can be decreased, with greater emphasis put on relational information. The power of our framework is its ability to support such a spectrum of tasks.

8 Conclusions

Imposing neighborhood structure on a set of points yields a graph, for which powerful indexing and matching algorithms exist. In this paper, we present a method for imposing neighborhood structure on a set of scale space top points. Drawing on the Delaunay triangulation of a set of points, we generate a graph whose edges are directed from top points at coarser scales to nearby top points at finer scales. The resulting construction is stable to noise, within-class variability, clutter, and occlusion, as reflected in a set of directed acyclic graph matching experiments. More importantly, the hierarchical structure can be exploited in a coarse-to-fine matching framework that can help manage the complexity of a linear search of a database.

Acknowledgements The work of Fatih Demirci is supported, in part, by TÜBİTAK grant No. 107E208. Ali Shokoufandeh acknowledges the partial support from NSF grant IIS-0456001 and ONR grant ONR-N000140410363. Luc Florack acknowledges the Netherlands Organization for Scientific Research (NWO) for financial support. Sven Dickinson acknowledges the support of NSERC, PREA, OCE, and IRIS.

References

1. Balmashnova, E., Florack, L., ter Haar Romeny, B.: Feature vector similarity based on local structure. In: Sgallari, F., Murli, A., Paragios, N. (eds.) *Scale Space and Variational Methods in Computer Vision: Proceedings of the First International Conference, SSVM 2007, Ischia, Italy. Lecture Notes in Computer Science*, vol. 4485, pp. 386–393. Springer, Berlin (2007)
2. Balmashnova, E., Florack, L.M.J., Platel, B., Kanters, F.M.W., ter Haar Romeny, B.M.: Stability of top-points in scale space. In: Kimmel, R., Sochen, N., Weickert, J. (eds.) *Scale Space and PDE Methods in Computer Vision: Proceedings of the Fifth International Conference, Scale-Space 2005, Hofgeismar, Germany. Lecture Notes in Computer Science*, vol. 3459, pp. 62–72. Springer, Berlin (2005)
3. Balmashnova, E., Platel, B., Florack, L., ter Haar Romeny, B.M.: Content-based image retrieval by means of scale-space top-points and differential invariants. In: Greenspan, H., Lehmann, T. (eds.) *Proceedings of the MICCAI Workshop on Medical Content-Based Image Retrieval for Biomedical Image Archives: Achievements, Problems, and Prospects (Brisbane, Australia, October 29, 2007)*, pp. 83–92 (2007)
4. Balmashnova, E., Platel, B., Florack, L.M.J., ter Haar Romeny, B.M.: Object matching in the presence of non-rigid deformations close to similarities. In: *Proceedings of the 8th IEEE Computer Society Workshop on Non-Rigid Registration and Tracking through Learning, held in Conjunction with the IEEE International Conference on Computer Vision (Rio de Janeiro, Brazil, October 14–20, 2007. IEEE Computer Society Press, Los Alamitos (2007)*
5. Bay, H., Ess, A., Tuytelaars, T., Gool, L.: Speeded-up robust features (surf). *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008)
6. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(24), 509–522 (2002)
7. Bretzner, L., Lindeberg, T.: Qualitative multiscale feature hierarchies for object tracking. *J. Vis. Commun. Image Represent.* **11**(2), 115–129 (2000)
8. Carneiro, G., Jepson, A.D.: Phase-based local features. In: *Proceedings of the 7th European Conference on Computer Vision, Part I*, pp. 282–296. Springer, London (2002)
9. Cohen, S., Guibas, L.: The earth mover's distance under transformation sets. In: *ICCV'99: Proceedings of the International Conference on Computer Vision*, vol. 2, p. 1076. IEEE Computer Society, Washington (1999)
10. Crowley, J., Parker, A.: A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(2), 156–170 (1984)
11. Damon, J.: Local Morse theory for solutions to the heat equation and Gaussian blurring. *J. Differ. Equ.* **115**(2), 368–401 (1995)
12. Demirci, M.F., Shokoufandeh, A., Dickinson, S., Keselman, Y., Bretzner, L.: Many-to-many matching of scale-space feature hierarchies using metric embedding. In: *Proceedings, Scale Space Methods in Computer Vision, 4th International Conference*, pp. 17–32, June 2003
13. Demirci, M.F., Shokoufandeh, A., Dickinson, S., Keselman, Y., Bretzner, L.: Many-to-many feature matching using spherical coding of directed graphs. In: *Proceedings, 8th European Conference on Computer Vision*, pp. 332–335, May 2004
14. Demirci, M.F., Shokoufandeh, A., Keselman, Y., Bretzner, L., Dickinson, S.: Object recognition as many-to-many feature matching. *Int. J. Comput. Vis.* **69**(2), 203–222 (2006)
15. Dufournaud, Y., Schmid, C., Horaud, R.: Matching images with different resolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 612–618. IEEE Computer Society Press, Los Alamitos (2000)
16. Eberly, D., Gardner, R., Morse, B., Pizer, S., Scharlach, C.: Ridges for image analysis. *J. Math. Imaging Vis.* **4**(4), 353–373 (1994)
17. Florack, L., Kuijper, A.: The topological structure of scale-space images. *J. Math. Imaging Vis.* **12**(1), 65–79 (2000)
18. Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.M.: The Amsterdam library of object images. *Int. J. Comput. Vis.* **61**(1), 103–112 (2005)
19. Giannopoulos, P., Veltkamp, R.: A pseudo-metric for weighted point sets. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *Proceedings of the Seventh European Conference on Computer Vision (Copenhagen, Denmark, May–June 2002). Lecture Notes in Computer Science*, vols. 2350–2353, pp. 715–730. Springer, Berlin (2002)
20. Hummel, R., Moniot, R.: Reconstructions from zero-crossings in scale-space. *IEEE Trans. Acoust. Speech Signal Proces.* **37**(12), 2111–2130 (1989)
21. Huttenlocher, D., Ullman, S.: Recognizing solid objects by alignment with an image. *Int. J. Comput. Vis.* **5**(2), 195–212 (1990)
22. Johansen, P.: On the classification of topoints in scale space. *J. Math. Imaging Vis.* **4**, 57–67 (1994)
23. Johansen, P., Skelboe, S., Grue, K., Andersen, J.: Representing signals by their topoints in scale space. In: *Proceedings of the International Conference on Image Analysis and Pattern Recognition*, pp. 215–217, 1986
24. Kadir, T., Brady, M.: Saliency, scale and image description. *Int. J. Comput. Vis.* **45**(2), 83–105 (2001)
25. Kanters, F.M.W.: Scalespaceviz. <http://www.bmi2.bmt.tue.nl/image-analysis/people/FKanters/Software/ScaleSpaceViz.html/> (2004)
26. Kanters, F.M.W., Florack, L.M.J., Platel, B., ter Haar Romeny, B.M.: Image reconstruction from multiscale critical points. In: *Proceedings of the 4th International Conference on Scale Space Methods in Computer Vision*, pp. 464–478. Isle of Skye, UK (2003)
27. Kanters, F.M.W., Platel, B., Florack, L.M.J., ter Haar Romeny, B.M.: Content based image retrieval using multiscale top points. In: *Proceedings of the 4th International Conference on Scale Space Methods in Computer Vision*, pp. 33–43. Isle of Skye, UK (2003)
28. Keselman, Y., Shokoufandeh, A., Demirci, M.F., Dickinson, S.: Many-to-many graph matching via low-distortion embedding. In: *Proceedings, Computer Vision and Pattern Recognition*, pp. 850–857, 2003
29. Koenderink, J.J.: The structure of images. *Biol. Cybern.* **50**(5), 363–370 (1984)
30. Koenderink, J.J., van Doorn, A.J.: Dynamic shape. *Biol. Cybern.* **53**(6), 383–396 (1986)
31. Lamdan, Y., Schwartz, J., Wolfson, H.: Affine invariant model-based object recognition. *IEEE Trans. Robot. Automat.* **6**(5), 578–589 (1990)
32. Lifshitz, L.M., Pizer, S.M.: A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(6), 529–541 (1990)
33. Lindeberg, T.: Scale-space behaviour of local extrema and blobs. *J. Math. Imaging Vis.* **1**(1), 65–99 (1992)
34. Lindeberg, T.: Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. J. Comput. Vis.* **11**(3), 283–318 (1993)
35. Lindeberg, T.: *Scale-Space Theory in Computer Vision*. Kluwer Academic, Norwell (1994)
36. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. *Int. J. Comput. Vis.* **30**(2), 117–156 (1998)
37. Lindeberg, T.: Feature detection with automatic scale selection. *Int. J. Comput. Vis.* **30**(2), 77–116 (1998)
38. Loog, M., Duistermaat, J.J., Florack, L.: On the behavior of spatial critical points under Gaussian blurring. A folklore theorem and scale-space constraints. In: *Scale-Space'01: Proceedings of the Third International Conference on Scale-Space and Morphology in Computer Vision*, pp. 183–192. Springer, London (2001)

39. Lowe, D.: *Perceptual Organization and Visual Recognition*. Kluwer Academic, Norwell (1985)
40. Lowe, D.: Object recognition from local scale-invariant features. In: *Proceedings of the International Conference on Computer Vision*, vol. 2, pp. 1150–1157. IEEE Computer Society, Washington (1999)
41. Matas, J., Chum, O., Martin, U., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: *Proceedings of the British Machine Vision Conference*, vol. 1, pp. 384–393, London (2002)
42. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. *Int. J. Comput. Vis.* **60**(1), 63–86 (2004)
43. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. *Int. J. Comput. Vis.* **65**(1–2), 43–72 (2005)
44. Nielsen, M., Lillholm, M.: What do features tell about images? In: *Scale-Space '01: Proceedings of the Third International Conference on Scale-Space and Morphology in Computer Vision*, pp. 39–50. Springer, London (2001)
45. Okabe, A., Boots, B.: *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams*. Wiley, New York (1992)
46. ORL. The ORL face database at the AT&T (Olivetti) research laboratory. <http://www.cl.cam.ac.uk/Research/DTG/attarchive/face/database.html> (1992)
47. Platel, B., Balmashnova, E., Florack, L., ter Haar Romeny, B.M.: Top-points as interest points for image matching. In: *Proceedings of the 9th European Conference on Computer Vision*, pp. 418–429 (2006)
48. Platel, B., Kanters, F.M.W., Florack, L.M.J., Balmashnova, E.G.: Using multiscale top points in image matching. In: *Proceedings of the 11th International Conference on Image Processing*, Singapore, October 2004
49. Poston, T., Stewart, I.N.: *Catastrophe Theory and its Applications*. Pitman, London (1978)
50. Preparata, F., Shamos, M.: *Computational Geometry*. Springer, New York (1985)
51. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(5), 530–535 (1997)
52. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The Princeton shape benchmark. In: *Proceedings of the Shape Modeling International 2004*, pp. 167–178. IEEE Computer Society, Washington (2004)
53. Shokoufandeh, A., Bretzner, L., Macrini, D., Demirci, M.F., Jönsson, C., Dickinson, S.: The representation and matching of categorical shape. *Comput. Vis. Image Understand.* **103**(2), 139–154 (2006)
54. Shokoufandeh, A., Dickinson, S., Jönsson, C., Bretzner, L., Lindeberg, T.: On the representation and matching of qualitative shape at multiple scales. In: *Proceedings of the 7th European Conference on Computer Vision, Part III*, pp. 759–775. Springer, London (2002)
55. Shokoufandeh, A., Dickinson, S., Siddiqi, K., Zucker, S.: Indexing using a spectral encoding of topological structure. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 491–497. Fort Collins, Glasgow (1999)
56. Shokoufandeh, A., Marsic, I., Dickinson, S.: View-based object recognition using saliency maps. *Image Vis. Comput.* **17**(5/6), 445–460 (1999)
57. Siddiqi, K., Shokoufandeh, A., Dickinson, S., Zucker, S.: Shock graphs and shape matching. *Int. J. Comput. Vis.* **30**, 1–24 (1999)
58. Thom, R.: *Stabilité Structurelle et Morphogénèse*. Benjamin, Paris (1972)
59. Thom, R.: *Structural Stability and Morphogenesis*. Benjamin-Addison Wesley, New York (1975) (translated by D.H. Fowler)
60. van Wijk, J.J., Nuij, W.A.A.: A model for smooth viewing and navigation of large 2d information spaces. *IEEE Trans. Vis. Comput. Graph.* **10**(4), 447–458 (2004)
61. Witkin, A.: Scale-space filtering. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1019–1022, 1983



M. Fatih Demirci received his Ph.D. in computer science from Drexel University in 2005. After working as a postdoctoral research scientist at Utrecht University, The Netherlands, he joined TOBB University of Economics and Technology, Turkey as an assistant professor at computer engineering department. His research interests include structural pattern recognition in computer vision, shape indexing, and graph theory. He was the recipient of the best Ph.D. dissertation award in engineering and physical sciences, Drexel University in 2006.



working on image analysis and surgery applications.

Bram Platel received his Master's degree (cum laude) in Biomedical Engineering at the Eindhoven University of Technology (TU/e), The Netherlands, in 2002. He continued his work on image registration and object retrieval based on scale-space and graph theory and earned his Ph.D. degree in Biomedical Engineering at the TU/e in 2007. Currently, Bram Platel is employed as an assistant professor at the Maastricht University Medical Center where he is leading the Image Guided Surgery group. His group is visualizing techniques for neurosurgery applications.



Ali Shokoufandeh received the B.Sc. degree in computer science from the University of Tehran and the M.Sc. and Ph.D. degrees in computer science from Rutgers in 1996 and 1999, respectively. He is an associate professor of computer science at Drexel University. His research focuses on computer vision, pattern recognition, extremal graph theory and geometry, and combinatorial optimization. He currently serves as Associate Editor for the journals: *Pattern Recognition Letters* and *IET Computer Vision*.



Luc L.M.J. Florack received his M.Sc. degree in theoretical physics in 1989, and his Ph.D. degree cum laude in 1993 with a thesis on image structure, both from Utrecht University, The Netherlands. During the period 1994–1995 he was an ERCIM/HCM research fellow at INRIA Sophia-Antipolis, France, and INESC Aveiro, Portugal. In 1996 he was an assistant research professor at DIKU, Copenhagen, Denmark, on a grant from the Danish Research Council. In 1997 he re-

turned to Utrecht University, where he became an assistant research professor at the Department of Mathematics and Computer Science. In 2001 he moved to Eindhoven University of Technology, Department of Biomedical Engineering, where he became an associate professor in 2002. In 2007 he was appointed full professor at the Department of Mathematics and Computer Science, retaining a parttime professor position at the former department. His research covers mathematical models of structural aspects of signals, images, and movies, particularly multiscale and differential geometric representations, and their applications to imaging and vision, with a focus on cardiac cine magnetic resonance imaging, high angular resolution diffusion imaging, and diffusion tensor imaging, and on biologically motivated models of “early vision”.



Sven J. Dickinson received the B.A.Sc. degree in Systems Design Engineering from the University of Waterloo, in 1983, and the M.S. and Ph.D. degrees in Computer Science from the University of Maryland, in 1988 and 1991, respectively. He is currently Professor of Computer Science at the University of Toronto. From 1995–2000, he was an Assistant Professor of Computer Science at Rutgers University, where he also held a joint appointment in the Rutgers Center for Cognitive Science (RuCCS). From 1991–1994, he was a Research Associate at the Artificial Intelligence Laboratory, University of Toronto. Prior to his academic career, he worked in the computer vision industry, designing image processing systems for Grinnell Systems Inc., San Jose, CA, 1983–1984, and optical character recognition systems for DEST, Inc., Milpitas, CA, 1984–1985. His research interests revolve around the problem of object recognition, in general, and generic object recognition, in particular. In 1996, he received the NSF CAREER award for his work in generic object recognition, and in 2002, received the Government of Ontario Premier’s Research Excellence Award (PREA), also for his work in generic object recognition. He was co-chair of the 1997, 1999, 2004, and 2007 IEEE International Workshops on Generic Object Recognition, and co-chaired both the First and Second International Workshops on Shape Perception in Human and Computer Vision in 2008 and 2009, respectively. He serves or has served on the editorial boards of a number of journals, including the IEEE Transactions on Pattern Analysis and Machine Intelligence, the International Journal of Computer Vision, Image and Vision Computing, Pattern Recognition Letters, IET Computer Vision, and the Journal of Electronic Imaging. He is also co-editor of the Synthesis Lectures on Computer Vision from Morgan and Claypool Publishers.