Performance Evaluation of Tree Object Matching

Somchaipeng, Kerawit; Sporring, Jon; Kreiborg, Sven; Johansen, Peter

Publication date:
2005

Document version
Early version, also known as pre-print

Deep Structure, Singularities, and Computer Vision

DSSCV

IST-2001-35443

Deliverable

Deliverable No.:	33
Title:	Performance Evaluation of Tree Object Matching
Author(s):	K. Somchaipeng, J. Sporring, S. Kreiborg, and P. Johansen
Institution:	3DLab, School of Dentistry, University of Copenhagen
Classification:	Public
Date:	September 15, 2005

Abstract

Multi-Scale Singularity Trees (MSSTs) represents the deep structure of images in scale-space and provide both the connections between image features at different scales and their strengths. In this report we present and evaluate an algorithm that exploits the MSSTs for image matching. Two versions of the algorithm is presented: an exact and an approximation. Several experiments are conducted to empirically evaluate the MSST matching algorithm under image distortions. Further, the performance of the MSST matching algorithm is measured on three databases: the ORL face database, magazine covers, and the COIL database. Finally the performance is compared with algorithms based on the Scale Invariant Feature Transform (SIFT), and the Position of Catastrophes (CAT).

1 Introduction

The quantification of the differences or distances between images, and choosing the two closest images is called image matching, and image matching is a fundamental task in a content-based image retrieval system. The typical application is that the user presents the system with an image, and the system returns a ranked list of some image from a database that are similar. Such application is becoming increasingly popular, and examples of usages are: paper and television news archives, security systems, and home image databases.

This report has chosen to focus on two related works, algorithms based on the Position of Catastrophes (CAT) [1] and on the Scale Invariant Feature Transform (SIFT) [2]. These are all based on scale invariant features, where SIFT is the most developed on successful.

In [1], a set of catastrophes in scale-space and their reconstruction coefficients were used as an image representation, which we call Position of Catastrophes (CAT). In this report we only discuss Gaussian scale-space, which we just call scale-space, and the history of which is discussed in [3]. The image matching problem was then translated into comparing sets of points in high dimensional space. The distances between point sets were calculated using the Earth Mover Distance (EMD) [4]. As the authors realized, some catastrophes are more stable than others, hence in [5] unstable catastrophes were discarded from the image representations. They argued that catastrophes are more stable in an area with a lot of structure. The amount of structure contained in a spatial area around catastrophes can be estimated by the total variation norm. More in depth discussion of the stability of catastrophes based on perturbation theory and noise propagation was presented in [6]. Such stability measures are an integral part of the EMD algorithm.

In [2] and accompanying articles, a set of Scale Invariant Feature Transform (SIFT) were presented for image matching. The SIFT features are calculated in scale-space as extremal points of differences of Gaussian blurs. These are approximations of the scale normalized (spatial) Laplacian, and their extremal points correspond to the points of blob-detection [7]. Out of all detected points, the SIFT algorithm selects robust points by eliminating low-contrast points and edge points. Then local histogram of the gradient vector is sampled non-linearly in a small number of orientation and magnitude bins. The SIFT is a set of features for which an accompanying matching algorithm has been proposed [8]: Best-Bin-First (BBF). In this work we will use the Earth Movers Distance instead, since it appears to have comparable performance on the SIFT features.

None of the above mentioned image matching algorithms includes the information on relations or linking between catastrophes in their image representations as opposed to the Multi-Scale Singularity Trees (MSSTs) introduced in [9]. These novel trees and powerful multi-scale image descriptors represent the deep structures of images and the relations of image features at different scales. Two kinds of MSSTs have been proposed: Extrema-Based MSSTs and Saddle-Based MSSTs. The difference between the two MSSTs are that in the extrema-based, the catastrophes are linked with extrema, while in the saddle-based, the catastrophes are connected with saddles. A study of the transitions of MSSTs under image perturbations [10]

suggested that Saddle-Based MSSTs are more stable and potentially powerful as multi-scale image descriptors.

In this report we will investigate the usefulness of Saddle-Based MSSTs for image matching. The image matching algorithm, which will be described in detail in this report, uses only the information contained in the energy matrices of MSSTs, which are where the strengths of the linking between catastrophes are stored. The topology of MSSTs may be derived from the energy matrix, and therefore the matching results using the algorithm indicates the amount of image information that is captured by the linking of MSSTs.

2 Multi-Scale Singularity Trees

A Saddle-Based MSST is fully described by the saddles at the first-scale image, the catastrophes, and the invariant energy matrix. Each element in an energy matrix is the energy calculated between a pair of saddles at the first-scale image. The energy matrix of an MSST then contains the energies between all pairs of saddles in the image.

In scale-space, increasing the scale parameter simplifies the image. Saddles and extrema disappear or appear at annihilation and creation catastrophes, respectively, and the only generic catastrophe is pairwise annihilation or creation [11]. Since all saddles and extrema except one extremum in an image eventually disappear at annihilation catastrophes, saddles may all be uniquely associated with annihilation catastrophes in scale-space.

The probability of two annihilation occurring at same scale is zero, and we may thus rank saddles according to the scales of their associated annihilation catastrophes. This we do in a coarse to fine ordering. The linking for each catastrophe c_i in an MSST can be decided by looking for the saddle s_j, which is presents at the scale of the catastrophe c_i, and calculate an energy measure. The catastrophe c_i is then linked to the catastrophe c_j that is associated with the saddle s_j with lowest energy.

The energy between a pair of saddle s_i and saddle s_j at the first scale image, ranked according to the scales of their associated catastrophes, is the element E_{ij} of the energy matrix.

$$E_{ij} = \inf_{\gamma \in \Gamma_{s_i s_j}} \int_0^P \sqrt{(1-\alpha) \mid \frac{d\gamma(p)}{dp} \mid^2 + \alpha \mid \frac{\partial I(\gamma(p))}{\partial p} \mid^2} \, dp \,, \tag{1}$$

where $I : \Omega \to R^+$ is an intensity image, $\gamma : [0, P] \to \Omega$ is a path in the image parameterized by p, such that $\gamma(0) = s_i$ and $\gamma(P) = s_j$, $\Gamma_{s_i s_j}$ is the set of all possible paths the two saddles s_i and s_j, and α is a weighting factor between space and image intensity. Note that (1) is independent of the parameterization, e.g. integrating with respect to $q : [0, P] \to [0, P]$ we find:

$$\begin{aligned}
E_{ij} &= \inf_{\gamma \in \Gamma_{s_i s_j}} \int_{p^{-1}(0)}^{p^{-1}(x)} \sqrt{(1-\alpha) \mid \frac{d\gamma(p(q))}{dp} \mid^2 + \alpha \mid \frac{\partial I(\gamma(p(q)))}{\partial p} \mid^2} \, dq \\
&= \inf_{\gamma \in \Gamma_{s_i s_j}} \int_{p^{-1}(0)}^{p^{-1}(x)} \sqrt{(1-\alpha) \mid \frac{d\gamma(p)}{dp} \cdot \frac{dp}{dq} \mid^2 + \alpha \mid \nabla I \cdot \frac{d\gamma(p)}{dp} \cdot \frac{dp}{dq} \mid^2} \, dq \\
&= \inf_{\gamma \in \Gamma_{s_i s_j}} \int_{p^{-1}(0)}^{p^{-1}(x)} \sqrt{(1-\alpha) \mid \frac{d\gamma(p)}{dp} \mid^2 + \alpha \mid \frac{\partial I(\gamma(p))}{\partial p} \mid^2} \, \frac{dp}{dq} \cdot dq \\
&= \inf_{\gamma \in \Gamma_{s_i s_j}} \int_0^P \sqrt{(1-\alpha) \mid \frac{d\gamma(p)}{dp} \mid^2 + \alpha \mid \frac{\partial I(\gamma(p))}{\partial p} \mid^2} \, dp \,.
\end{aligned} \tag{2}$$

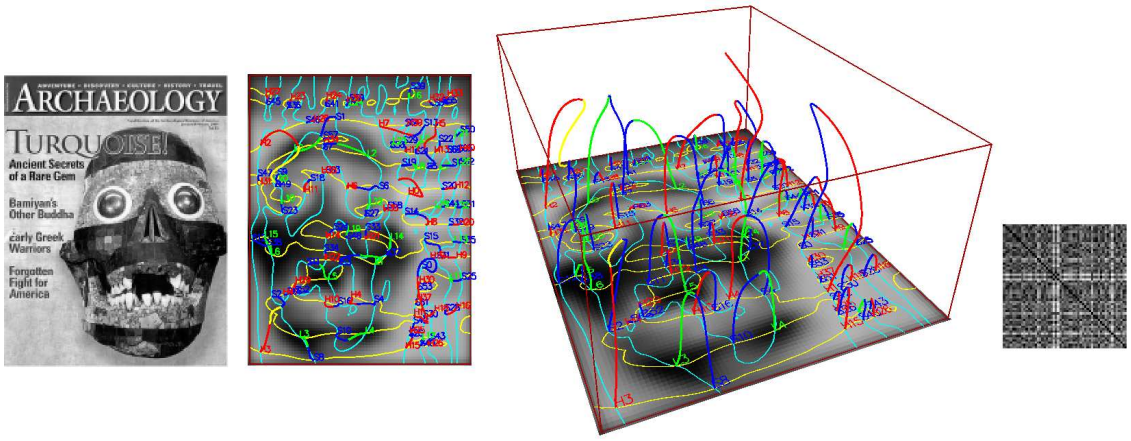

Figure 1: A magazine cover, its deep structure, and it's energy matrix. Maximal-paths, minimal-paths, and saddle-paths are displayed in red, green, and blue, respectively. The zero-crossings of the first derivatives I_x and I_y at the first-scale image are shown in light blue and yellow, respectively.

In this report the $\alpha = 1$ is set so that the energy matrix only depends on the intensity of the image and is theoretically invariant under all image transformations that act on space, but not on the intensity of I. Substituting $\alpha = 1$ in Eq. (1) gets

$$E_{ij} = \inf_{\gamma \in \Gamma_{s_i s_j}} \int_0^P \left| \frac{\partial I(\gamma(p))}{\partial p} \right| dp \,. \tag{3}$$

In plain English, the energy E_{ij}, where $\alpha = 1$ is the minimum sum of image intensity differences along any possible paths from s_i to s_j. This is demonstrated in Fig. 1, which shows a magazine cover, the zero-crossings of the first derivatives overlayed on top of the first scale image, the deep structure, and the energy matrix. The first scale image appears flipped because it is viewed from behind so that the coordinate axises agree with the right hand rule. The energy matrix is symmetric with zeros along its diagonal, and all its elements are non-negative.

3 Coping with Creations

In 2D, creations are generic but the tree structure as discussed above assume that only annihilations occurs. In order to preserve the tree structure of the MSSTs and to simplify the matching algorithm, creations and loops in scale-space images are systematically removed. It has been shown in [12] that although creations are generic, they might not be easily detected. Creations that form loops are short-lived, hence if we do not sample scale-space fine enough along scale, then they may pass undetected. On the other hand, most creations that do not form loops are followed closely by annihilations and slight perturbations will merge them. Therefore we handle creations as follows:

1. Creations that occurs in critical-paths, where those paths can be traced down to the first-scale image, are pairwise removed with the next annihilations on that path in the direction that moves to the first-scale image. The top-most annihilations catastrophes on those paths are the catastrophes associated with the saddles at the first-scale image.

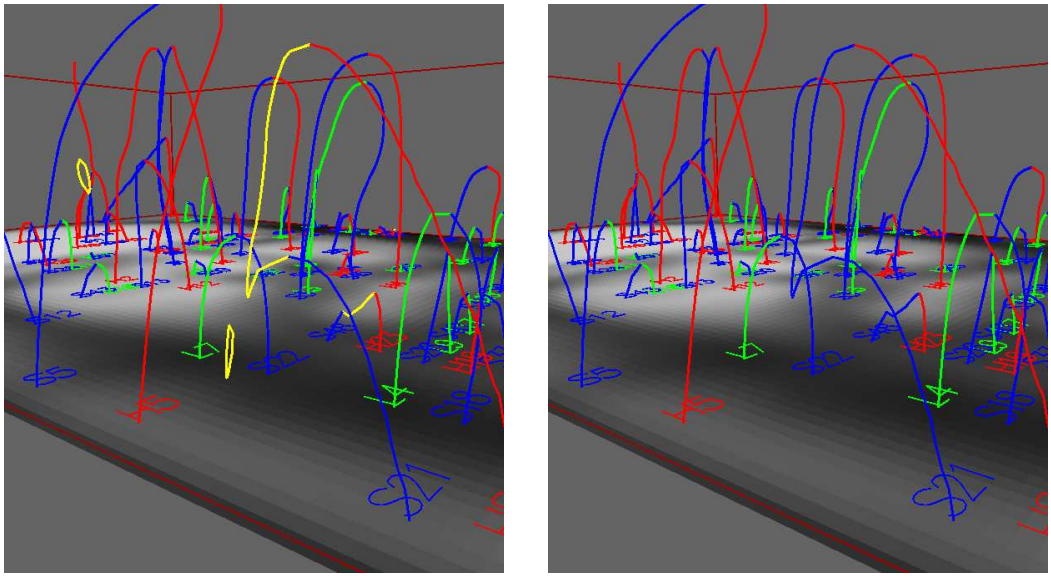

Figure 2: Creations are systematically ignored. Critical-paths that originate from creations are indicated by yellow. Left and right panels show part of a deep structure before and after removals. Pairwise removals of creation-annihilation pairs are indicated by color changes of the paths from yellow to blue.

2. Creations that eventually end at annihilations and form loops in scale-space with no connection to saddles at the first-scale image are ignored.

In Fig. 2 is shown an example of loop removals and pairwise removals of creations and annihilations on paths that can tracked back to the first-scale image. The figure shows examples of pairwise removals of multiple creations and annihilations in critical-paths such that the highest annihilation in each saddle-path is kept as the catastrophe associated with the saddle at the first-scale image. Likewise the figure shows that creations resulting in loops are ignored.

4 Matching Algorithm

The distance between two images is calculated as a sum of the squared differences of the corresponding elements in the normalized MSSTs' energy matrices,

$$D(I,J) = \sum_{i,j} (E_I(i,j) - E_J(i,j))^2, \tag{4}$$

where I and J are two images to be matched, and E_I and E_J are the corresponding energy matrices. Since D is a linear combination of E, then D observes the same invariance as E, i.e. D is invariant under all transformation of space, such as scaling, translation, and rotation. Because the ordering of catastrophes in scale effects the location of rows and columns in energy matrices. A swap of ordering between catastrophe c_i and catastrophe c_j in scale-space, corresponds to a swap of row i and row j, and column i and column j in the energy matrix. In order to compute the minimal differences of these matrices taking into account also the possible catastrophe reordering, the matching algorithm is based on the minimum under possible swapping. In the following we will present two algorithms: the exact and the approximative matching algorithm.

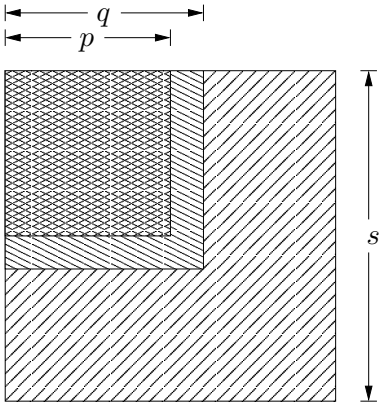

Figure 3: The schematic diagram of the exact matching algorithm.

4.1 Exact Matching

The *exact matching algorithm*, shown schematically in Fig. 3, looks, among all permutations of the ordering, for the catastrophe ordering that minimizes the distance in (4). Catastrophes corresponding to large scale structures in images are usually located high in scales. Because the saddles are ranked according to the scales of their associated catastrophes, part of the energy matrix that corresponds to large scale image structures are located on the top left of the energy matrix.

We cannot assume that all catastrophes in one image will exist in an image to be matched. Hence, we must be able to discarded or delete catastrophes in the matching process. This is achieved in the exact matching algorithm by permutes a slightly larger sub-matrix than used in the distance calculation. For an energy matrix with s catastrophes, the exact matching algorithm permutes the top-left sub-matrix with q catastrophes, $q < s$. Then, the computation of the distance is performed only on the p catastrophes, $p < q$, allowing catastrophes that are not fit very well to be discarded.

Because the searching space grows exponentially fast, only a small numbers of top catastrophes can be used. The experiment in the next section shows that 6 catastrophe is the most appropriate and 8 catastrophe sets the practical limit.

Catastrophes, which are far apart in scale are less likely to have come form compatible image structure. However, the scale-space of a flat image with i.i.d. noise will show a rich catastrophe structure, which is completely unstable, in the sense that drawing another noise example from the i.i.d. noise will result in a completely different set of catastrophes. Luckily, given a noise-level of the image, the statistical variance of the catastrophe localization may be estimated [5, 6] and this may be used as a weighted penalty in the catastrophe swapping.

We thus propose a matching cost based on Bayes' Maximum A Posteriori,

$$P(I|J) = \frac{P(J|I)P(I)}{P(J)} \tag{5}$$

where P denotes the a posteriori, the error, the prior, and the evidence probability distributions. When only interested in the maximum of (5), then the evidence may be ignored to give,

$$I^* = \arg\max_J P(J|I)P(I). \tag{6}$$

Further, since the logarithm is a strictly monotonic function, the maximum of (6) is equivalent to minimum of

$$I^* = \arg\min_J -\log P(J|I) - \log P(I) \tag{7}$$

Since the distance measure between energy matrices is a squared measure, we may rewrite it as a logarithm of a Gibbs distribution,

$$-\log P(J|I) = \frac{D(I,J)}{2\tau} + \log k, \tag{8}$$

where $k = \sum_K \exp \frac{D(I,K)}{2\tau}$ is the normalization factor calculated from all possible images K, and τ is some temperature variable. The variable k is constant for constant τ. In a similar manner we will design a prior probability distribution. One possibility is to use

$$-\log P(I) = \sum_{i,j \in S} \frac{w_i |\log \sigma_i - \log \sigma_j|}{\mu} + \log c \tag{9}$$

where we constantly keep track of the original scale σ_i of every catastrophe, and relate this to the scale of the catastrophe it has been swapped with, σ_j. The weighting factor w_i may be used to control the movement of catastrophes, such that setting w_i high will imply that a catastrophe is unlikely to move. Finally the constant c is a normalizing constant. Writing the cost of swapping catastrophe i and j as,

$$F(i \leftrightarrow j) = 2|\log \sigma_i - \log \sigma_j|, \tag{10}$$

we find that F has two nice properties. Firstly it is scale invariant, i.e. scaling σ with some constant η implies that

$$F(i \leftrightarrow j) = 2|\log \eta\sigma_i - \log \eta\sigma_j| = 2|\log \sigma_i - \log \sigma_j|. \tag{11}$$

Secondly, we only need not use the sequence of swappings performed for catastrophe i to have reached place of catastrophe j, we need only the absolute logarithmic difference to its starting point. This is an advantage algorithmically. The current implementation sets $w_i = 1$ meaning that all catastrophes can move arbitrarily far in scale.

4.2 Approximate Matching

To increase the number of catastrophes used in the distance calculations and to reduce the computational time, we have devised the *approximated matching algorithm*, which is shown schematically in Fig. 4. Th approximate matching algorithm uses a *moving window* strategy: starting from the top-left of the energy matrix, the approximate algorithm places a moving window on top of q catastrophes. Then, using the exact matching algorithm, the approximate matching algorithm looks for the best catastrophe ordering among all possible permutations within the window. Once the ordering has been found, the moving window is moved to the right and down to the next position. The top most catastrophe in the window is assumed that it locates at the correct position. The approximate matching algorithm continues in the same manner until all or a given number of catastrophes are processed. By keeping the window size small, more of catastrophes can be used in the distance calculation than in the exact matching algorithm.

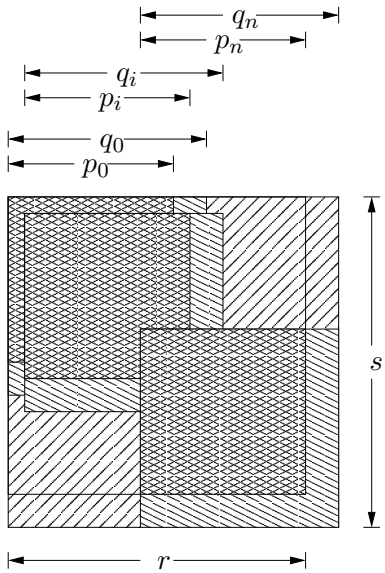

Figure 4: The schematic diagram of the approximate matching algorithm.

5 Experiments

Several experiments have been performed on three image databases to investigate various aspects of the two proposed matching algorithms and compare these with matching algorithms from the literature. The image databases used are the ORL face database [13], the magazine cover database, and the Columbia Object Image Library (COIL) database [14]. The experiments are grouped as follows:

1. The matching of images of faces in the ORL face database. We investigate the effect of each tunable algorithm parameters namely the number of catastrophes and the number of deletion ($q - p$). The comparison between the exact matching algorithm and the approximate algorithm using different number of catastrophes. The matching results using randoms points instead of the saddles are also presented to demonstrate the richness of saddles. Finally, the performance of the matching method on the database is presented and compared with two other methods.

2. The magazine cover database is used in the matching of transformed and noisy images. A magazine cover is scaled, rotated, and added with random noise and the distance between the distorted image and the original image is computed. This experiment demonstrates the effect of each image transformations and noise on the distance measured. Finally, the performance of the matching method on the database is presented and compared with two other methods.

3. The matching of images selected from the COIL database. Images of real-life objects viewed at different angles are used in the experiment. The experiment demonstrates how robust the matching method against scaling and 3D view point changes. The performance of the matching method on the database is presented and compared with two other methods.

5.1 ORL Face Database

In Fig. 5 is shown images of 10 individuals selected from the ORL face database. For each person, 10

Figure 5: Selected 10 individuals from the ORL face database.

Figure 6: There are 10 images for each individual. Images are captured for each individual with various facial expressions, different hair styles, with/with out glasses, etc. The red box indicates the key image for the set.

images are captured as shown in Fig. 6 for person number four in the selected group.

The matching results of the exact matching algorithm using different number of catastrophe are shown in Fig. 7. The matching performance increases with the number of catastrophes used increases. For this particular image database the matching performance starts to level out when 8 catastrophes are used. The matching results using random points instead of saddles is also presented for comparison. The deletion number of 1 was used in the experiment.

In Fig. 8 is shown the impacts of deletions. The positive effects of deletions only start to be visible when at least 6 catastrophes are used in the distance calculation. With lower number of catastrophes, the deletions might actually worsen the matching result. This is because, for a very small number of catastrophe, the amount of image information contained in each catastrophe is proportionally large and it offsets the positive effects of letting go unfitted catastrophes.

When more than 6 catastrophes are used in the distance calculation, the exact matching algorithm becomes very slow. Using the approximate matching algorithm, more catastrophes can be included in the distance calculation. In Fig. 9 is shown the matching performance of the exact matching algorithm compared with the approximate matching algorithm using different number of catastrophes. The approximate matching algorithm uses window size of 6 and 1 deletion.

In Fig. 10 is shown the time used in second, plotted in logarithmic scale, for the exact matching algorithm and the approximate matching algorithm to complete the distance calculations on the ORL face database in order to produce the results in Fig. 9. The computation time used for exact matching algorithm increases significantly when more than 6 catastrophes is used. The computation time used for the approximate algorithm on the other hands grows linearly with the number of catastrophes. The approximate matching algorithm uses window size of 6 and 1 deletion. Note that using 10 catastrophes, the performance of the approximate matching algorithms already beats that of the exact matching algorithm using 8 catastrophes, and spends less than $\frac{1}{20}$ the computation time in comparison to the exact matching algorithm.

The matching results of our algorithm compared with the SIFT key-points (SIFT) and the positions of catastrophes (CAT) are shown in Fig. 11. The SIFT key-points and catastrophe positions in CAT are matched using the Earth Mover Distance (EMD). The mass of each feature point is set equally to $\frac{1}{n}$, where n is the number of feature points. In this particular experiment, our algorithm performs relatively poor compared with the other two methods.

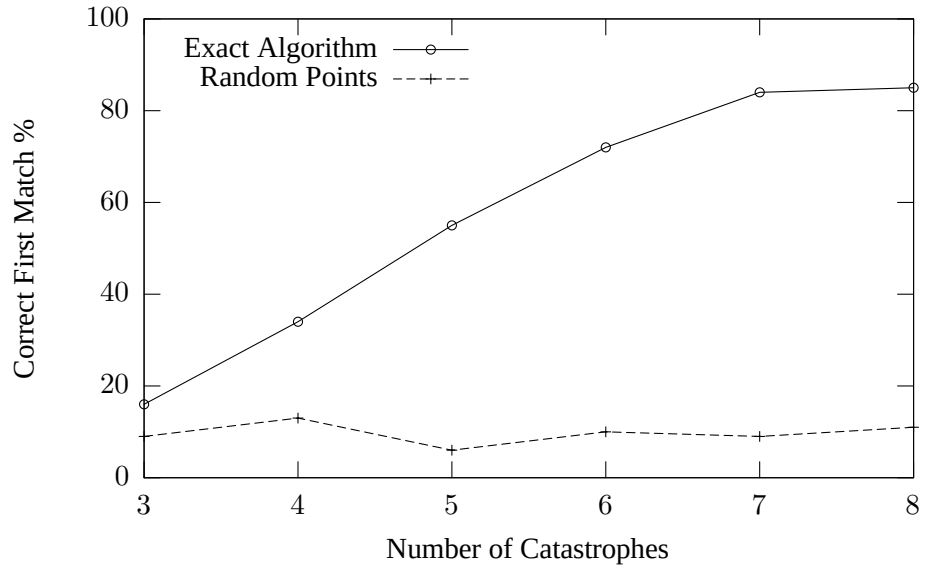

Figure 7: Matching results of the exact matching algorithm on the ORL face database using different number of catastrophes.

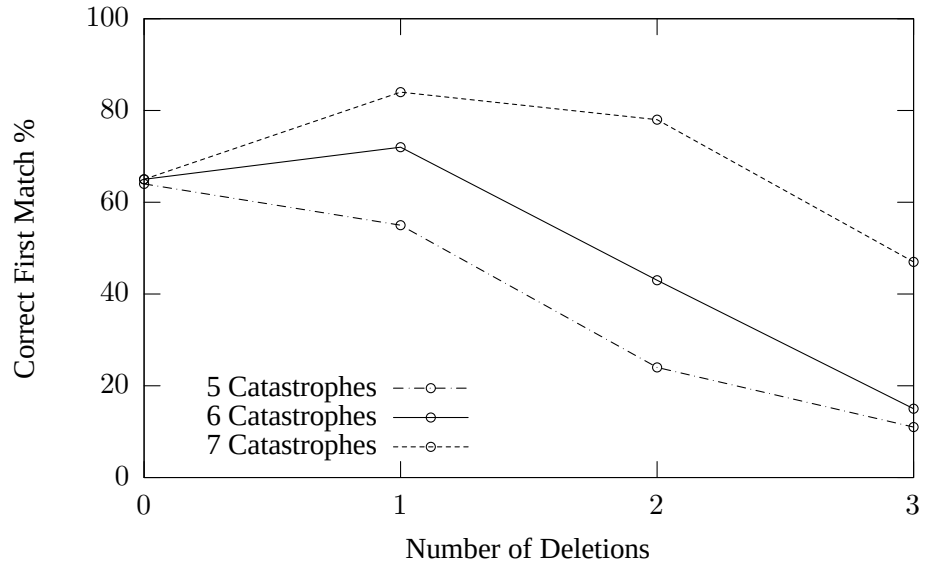

Figure 8: The effects of the deletion on the matching results. The results are produced using exact matching algorithm.

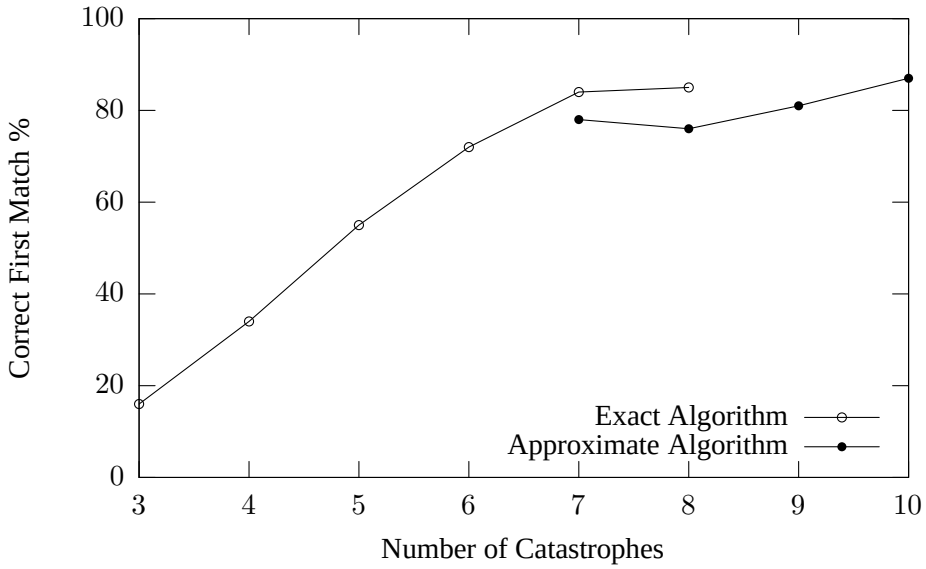

Figure 9: Matching results of the approximate matching algorithm compared with that of the exact matching algorithm using different number of catastrophes. For the approximate matching algorithm, the window size of 6 catastrophes is used and both use 1 deletion.

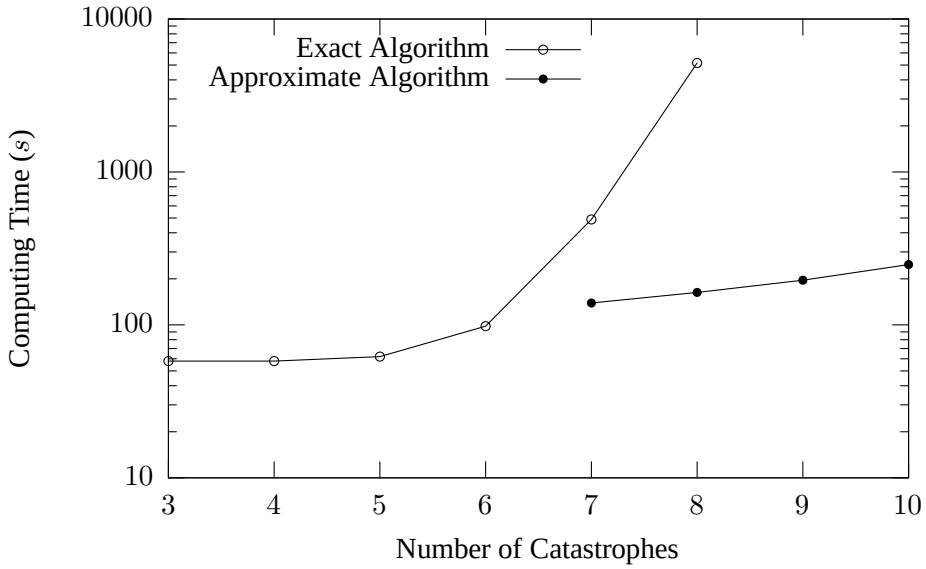

Figure 10: Computation time used by the exact matching algorithm compared with that used by the approximate matching algorithm to produce the results in Fig. 9.

	2	3	4	5	6	7	8	9	10	Key
10-6-1	87%	76.5%	66.67%	60%	56.8%	52.67%	47%	42.5%	39.22%	54%
10-7-1	89%	80%	69.67%	63.5%	58.4%	53%	47.43%	43.63%	39.89%	54%
CAT	97%	96%	92.33%	90.5%	85.4%	81.83%	77.57%	73.63%	69.22%	82%
SIFT	100%	100%	100%	99.75%	97%	94.5%	92.86%	91.38%	89.33%	91%

Figure 11: Matching results of three different methods on the ORL face database. Our method uses 10 catastrophes, 1 deletion, and with the window size of 6 and 7 respectively. The first column are the percentages of the second image being matched correctly, the second column are the percentages of the second and the third images being matched correctly, etc. The first image is the inquiry image and is always matched correctly. The last column gives the results of matching all faces only to the key image of each image set (the first sample). The matching results of SIFT and CAT are also provided for comparison.

Figure 12: The original 11 magazine covers in the magazine cover database. Various kinds of objects can be found including typed letters, man-made objects, natural scenes, and artificial graphics.

5.2 Magazine Database

In Fig. 12 is shown the 11 magazine covers in the Magazine cover database. Various kinds of objects can be found including typed letters, man-made objects, natural scenes, and artificial graphics. We will use these images to evaluate the proposed algorithm's behavior under various image distortions. Three kinds of image distortions will be discussed: rotation, scaling, and noise.

In Fig. 13 is shown the Piggy magazine cover rotated at different in plane rotation angles starting from 0° to 90° counterclockwise. In Fig. 14 is shown the Piggy magazine cover scaled at different scaling factor starting from 0.5 to 1.0. Fig. 15 shows the Piggy magazine cover with noise added at different levels starting from 1% to 10%. 1% noise means that each pixel in the image is added with random number drawn from $[-0.01, 0.01]$, when the pixels in the image are valued between $[0, 1]$

The results of matching distorted images of different kinds to their original images are shown in Fig. 16, Fig. 17, and Fig. 18. It is noteworthy to mention that the impact of image rotation is periodic. The error is at the highest point, when the image is rotated at approximately 45°. In contrast, the error is very small, when the image is rotated at multiples of right angles. This is due to the fact that digital images are sampled on rectangular grids, and when the energy map calculation is implemented using rectangular grids, then the result inevitably will be most inaccurately at rotation angles near 45°.

Total of 10 images are produced from each of every magazine covers in the magazine cover database by performing the mixed uniform scaling and rotation. The original image is rotated at various angles then the rotated images are scaled so that the width of the images remain constant. In Fig. 19 is shown an example of images produced from the Archeology magazine cover. The set of 11×10 transformed magazine covers are then used in the matching experiment. The Matching results of our method using different number of top catastrophes compared with the other two methods are shown in Fig. 20. SIFT and our method perform very well with SIFT performing slightly better. CAT performs poorly because positions of catastrophes are not invariant to severe translation, rotation, and scaling.

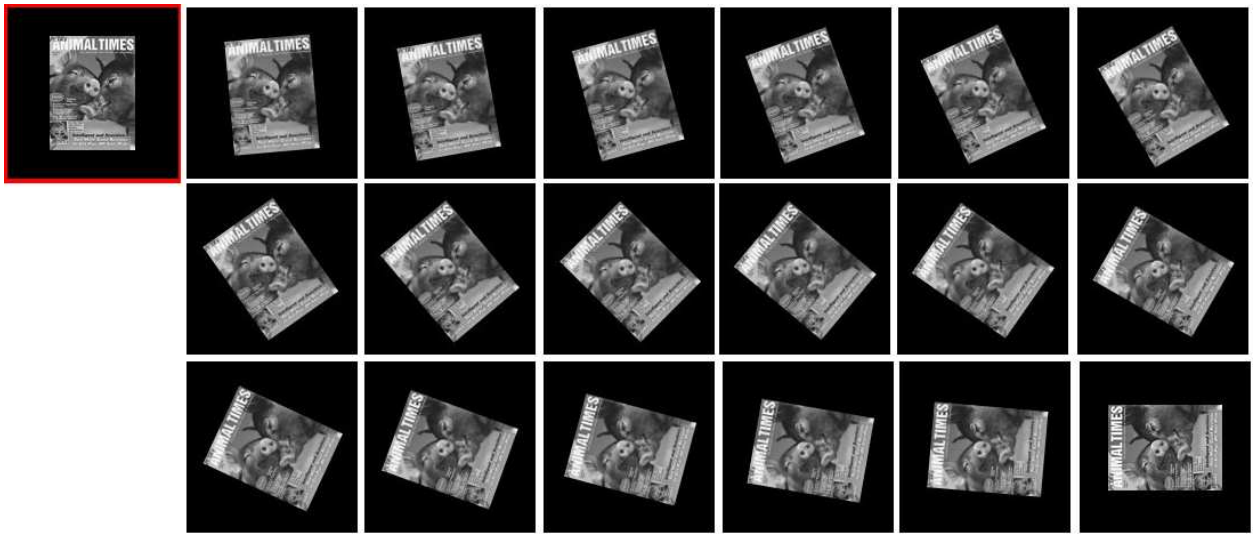

Figure 13: The Piggy magazine cover at different rotation angles from 0° to 90°. The original image is indicated by its red border.

Figure 14: The Piggy magazine cover at various levels of scaling from 0.5 to 1.0. The original image is indicated by its red border.

Figure 15: The Piggy magazine cover at various levels of noise added. The original image is indicated by its red border.

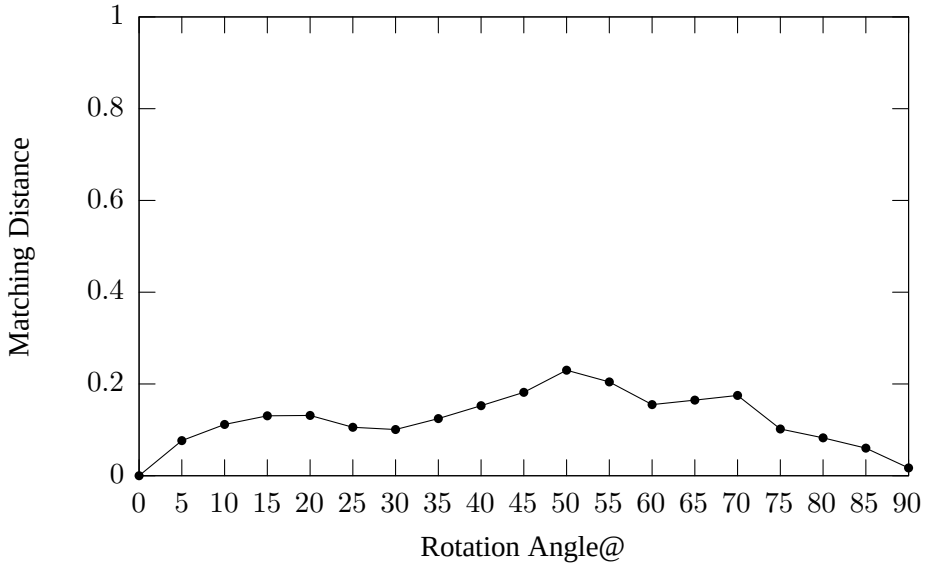

Figure 16: The effect of rotation. The distances between rotated images at different rotation angles and the original image.

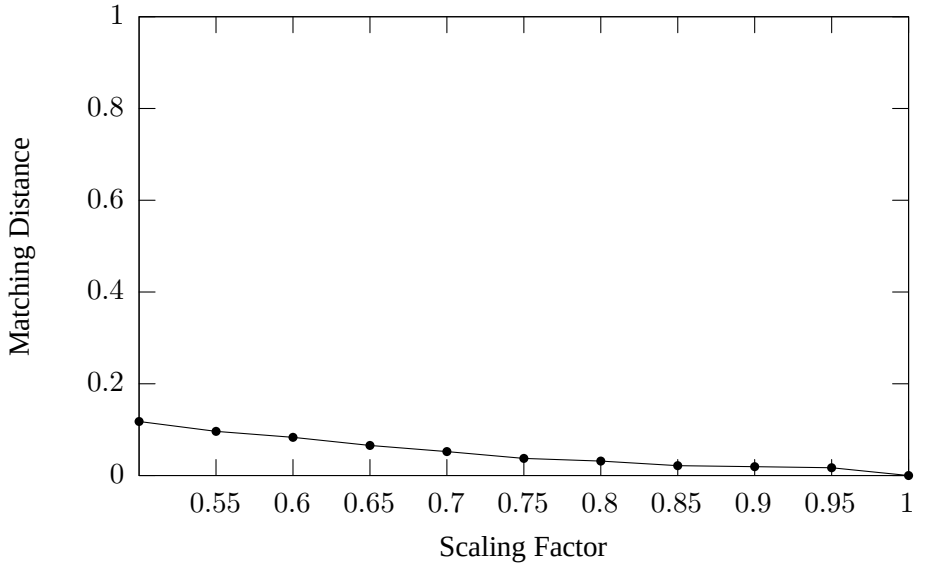

Figure 17: The effect of uniform scaling. The distances between scaled images at different scaling factors and the original image.

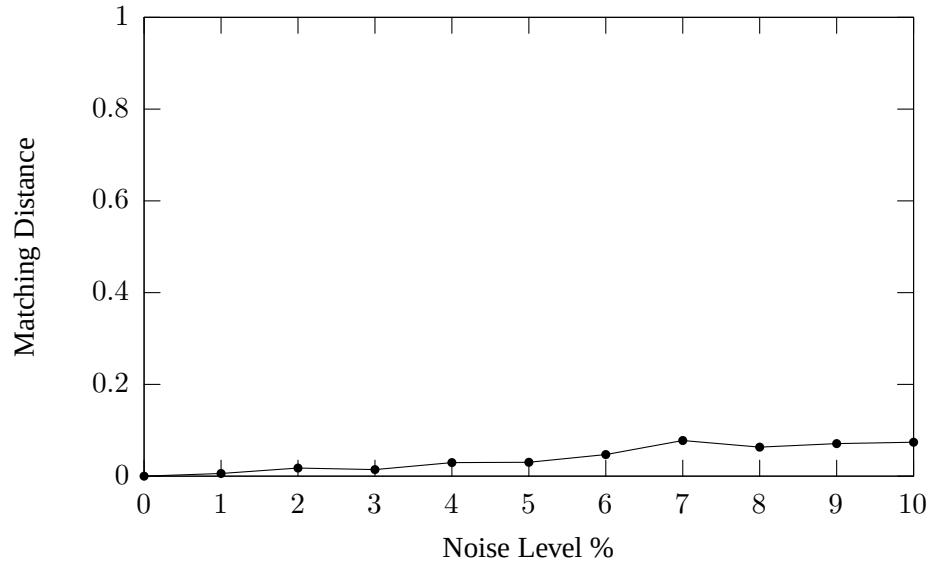

Figure 18: The effect of random noise. The distance between the images with random noise added at different levels and the original image.

Figure 19: For each and every 11 original magazine covers, 10 image are generated by performing mixed rotation and scaling at various levels. The original magazine cover is first rotated then it is scaled so that the width of the image remains constant.

	2	3	4	5	6	7	8	9	10	Key
8-6-1	98.18%	97.73%	96.36%	92.5%	90%	85.91%	81.56%	77.05%	72.42%	76.36%
10-6-1	100%	99.55%	98.18%	96.82%	95.09%	93.18%	90.13%	85.9%	80.51%	81.82%
CAT	57.27%	49.09%	39.09%	31.82%	28.73%	27.12%	24.81%	22.61%	20.71%	24.55%
SIFT	100%	100%	99.70%	99.77%	99.45%	99.24%	99.22%	97.05%	95.15%	100%

Figure 20: Matching results of three different methods on the transformed magazine cover database. Our method uses 8, and 10 top catastrophes, 1 deletion, with the window size of 6 catastrophes. The first column are the percentages of the second image being matched correctly, the second column are the percentages of the second and the third images being matched correctly, etc. The first image is the inquiry image and is always matched correctly. The last column gives the results of matching all distorted magazine covers only to the key image of each set (the original image). The matching results of SIFT and CAT are also provided for comparison.

Figure 21: The selected 10 objects from the COIL database.

Figure 22: The 10 selected views of an object from the COIL database. The key view is the view of 20° and is indicated by its red border.

5.3 COIL database

The Columbia Object Image Library (COIL) database consists of images of various real-life objects captured at different 3D view points. The objects are rotated at intervals of 5°. The images are also subjected to scaling.

Fig. 21 shows the 10 selected objects from the COIL database. For each object, images of 10 different 3D view points between 0° and 45° are drawn from the database. In Fig. 22 is shown the 10 selected views of an object in the COIL database. The key image is indicated by its red border.

In total, 10×10 images selected from the COIL database are used in the experiment. The matching results of our algorithm, SIFT and CAT are shown together for comparison in Fig. 23. All three methods performs very well especially SIFT, with our method slightly lagging behind.

6 Discussions and Conclusions

For our method, the matching of ORL face database is slightly harder than the matching of the magazine cover database and the COIL database. This is due to the high similarity between sets of objects. Faces all

	2	3	4	5	6	7	8	9	10	Key
8-6-1	96%	92.5%	90%	84%	78.8%	72.67%	68.14%	62.38%	58.56%	70%
10-6-1	96%	92%	88.67%	83.25%	77.2%	70.83%	64.28%	59.38%	55.11%	72%
10-8-1	97%	93%	87%	83.25%	78%	73.17%	67.71%	63.25%	58.33%	76%
CAT	100%	99%	97%	95%	90.6%	85.67%	81.29%	76.13%	71.78%	89%
SIFT	100%	100%	100%	100%	100%	99.83%	99%	98.13%	96.33%	100%

Figure 23: Matching results of different methods on the COIL database. Our method uses 8, 10, and 10 top catastrophes with window size of 6, 6 and 8 catastrophes, respectively. All of which use 1 deletion. The first column are the percentages of the second image being matched correctly, the second column are the percentages of the second and the third images being matched correctly, etc. The first image is the inquiry image and is always matched correctly. The last column gives the results of matching all views of all objects only to the key view of each object. The matching results of SIFT and CAT are also provided for comparison.

look quite similar, and they become even more similar, when they are blurred.

The possibility of trading accuracy for performance is naturally provided for our method. Catastrophe located at lower scales can be simply discarded, if response time is the most critical factor with the price of lower recognition rates.

For small database, using only a few top catastrophes are enough for our matching method to produce good matching results. As the number of catastrophes used in the calculation of the distance increases, the method will have a better chance of discriminate in a large database, but it will also be more sensitive to image distortions and occlusions.

The algorithm is currently in its very first steps. There are many possibilities of improvements. One possibility is the *adaptive approximate algorithm*, where the window size can grow or shrink adaptively, while keeping the scale difference between the highest and the lowest catastrophes in the window under a tunable level. The adaptive approximate algorithm will not only reduce the computational time but also likely to improve the matching results, since it will give options to the algorithm to delete catastrophes located high in scale, when appropriate.

The accuracy of the energy map generation is a very crucial part for good matching results. As Fig. 16 and Fig. 17 show, the accuracy of the current implementation, can still be largely improved, and we expect this to substantially improve the matching results.

It is interesting to note that, in every experiments performed in this report, SIFT and CAT use every piece of information they can collect form the images for their calculation of the matching distances. SIFT uses in general more than 100×128 numbers to represent an image. CAT uses in general more than than 50×3 numbers. Due to the limitations of our relatively naive searching in a large space, our method uses only up to 10×10 numbers in the experiments. In spite, we find that the matching results are comparable.

Earth Mover Distance (EMD) [4] is a very powerful and flexible method for point-set matching, as we hopefully demonstrated by applying it for the matching of SIFT key-points and the positions of catastrophes. Using Multidimensional Scaling (MDS) [15], it is theoretically possible to embedded the MSSTs' energy matrices into sets of points in a high-dimensional euclidean space, where the distances between those points approximates the energies. We then can use EMD to compute the distance just like what we did on SIFT key-points and catastrophes' positions. This will allow the utilization of the whole energy matrix and the matching results should be improved considerably.

For image matching and similar applications, the knowledge of only the connections of image features at different scales is insignificant. What is significant, is the knowledge of the strength of these connections. Fortunately, MSSTs provide us with both. With these unique advantages, we believe that MSSTs will also be found useful for many other applications. Currently, we are investigating the usefulness of MSSTs for detecting and locating sub-objects in images

7 Acknowledgments

This work is part of the DSSCV project sponsored by the IST Programme of the European Union (IST-2001-35443)

References

[1] F. Kanters, B. Platel, L. Florack, and B.M. ter Haar Romeny. Content Based Image Retrieval Using Multiscale Top Points, A Feasibility Study. In *Proceedings of the 4th Intl Conference on Scale-Space 2003*, pages 33 – 43, June 2003.

[2] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2005.

[3] J. Weickert, S. Ishikawa, and A. Imiya. Om the History of Gaussian Scale-Space Axiomatics. In Sporring et al. [16], chapter 4, pages 45–59.

[4] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 4(2):99–121, 2000.

[5] B. Platel, F. Kanters, L. Florack, and E. Balmachnova. Using multiscale top points in image matching. In *Proceedings of the Intl Conference on Image Processing 2004*, pages 389 – 392, 2004.

[6] E. Balmachnova, L.M.J. Florack, B. Platel, F.M.W. Kanters, and B.M. ter Haar Romeny. Stability of Top-points in Scale Space. In *Proceedings of the 5th Intl Conference on Scale-Space 2005*, pages 62 – 72, April 2005.

[7] T. Lindeberg. *Scale-Space Theory in Computer Vision*. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Boston, USA, 1994.

[8] J. Beis and D.G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Conference on Computer Vision and Pattern Recognition*, pages 1000–1006, Puerto rico, 1997.

[9] K. Somchaipeng, J. Sporring, S. Kreiborg, and P. Johansen. Multi-Scale Singularity Trees: Soft-linked Scale-Space Hierarchies. In *Proceedings of the 5th Intl Conference on Scale-Space 2005*, pages 97 – 106, April 2005.

[10] K. Somchaipeng, J. Sporring, S. Kreiborg, and P. Johansen. Transitions of Multi-Scale Singularity Trees. In *Proceedings of the Intl Workshop on Deep Structure, Singularity, and Computer Vision 2005*, June 2005.

[11] J. Damon. Local Morse Theory for Gaussian Blurred Functions. In Sporring et al. [16], chapter 11, pages 147–163.

[12] Arjan Kuijper. *The deep structure of Gaussian scale space images*. PhD thesis, Image Sciences Institute, Institute of Information and Computing Sciences, Faculty of Mathematics and Computer Science, Utrecht University, 2002.

[13] F. Samaria and A. Harter. Parameterisation of a Stochastic Model for human Face Identification. In *IEEE Workshop on Applications of Computer Vision*, Sarasota (Florida), December 1994.

[14] S. A. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (COIL-100). Technical report, CUCS-006-96, February 1996.

[15] F.W. Young and T.M. Hamer. *Multidimensional Scaling: History, Theory and Applications*. Erlbaum, New York, 1987.

[16] J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors. *Gaussian Scale-Space Theory*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.