USING MULTISCALE TOP POINTS IN IMAGE MATCHING

B. Platel, F.M.W. Kanters, L.M.J. Florack, E.G. Balmachnova

Eindhoven University of Technology
Department of Biomedical Image Analysis
PO Box 513
NL-5600 MB Eindhoven, The Netherlands
{B.Platel, F.M.W.Kanters, L.M.J.Florack, E.Balmashnova}@tue.nl

ABSTRACT

In this paper we discuss the feasibility of using singular points in a scale space representation (referred to as *top points*) for image matching purposes. These points are easily extracted from the scale space of an image and they form a compact description of the image. The image matching problem thus becomes a point cloud matching problem. This is related to the transportation problem known from linear optimization and we solve it by using an earth movers distance algorithm. To match points in scale space a distance measure is needed as Euclidean distance no longer applies. In this article we suggest a metric that can be used in scale space and show that it indeed performs better than a Euclidean distance measure. To distinguish between stable and unstable top points we derive a stability norm based on the total variation norm which only depends on the second order derivatives at the top point. To further improve matching results we show that other features at the top points can also increase the accuracy of matching.

1. INTRODUCTION

In this paper we demonstrate the use of scale space singular points for image matching. Nowadays many methods for image matching exist. We however conjecture that the use of special points obtained from the scale space of an image may well contribute to better matching algorithms. The top point representation, which is explained in sections 2 and 3, describes an image with a small point set in scale space [1]. Thus image matching reduces to point cloud matching. This can be tackled like a transportation problem known from linear optimization, for which a number of algorithms already exist. As our point matching scheme we use the earth movers distance algorithm, explained in section 4. The algorithm relies on a distance measure. Standard Euclidean distance does not apply to points in scale space, but is nevertheless often used to measure distances in scale space. In section 4.2 we suggest a different metric that does take into account that scale space is not a Euclidean 3D space.

Not all top points are stable, as discussed in section 4.3. To distinguish between stable and unstable top points an expression is derived based on the total variation norm.

We conclude the paper by showing some preliminary results, obtained from a simple image retrieval experiment on a small face database from the Olivetti Reseach Laboratory.

We stress that our algorithm merely provides a "proof of concept". Optimization is not a concern at this stage.

2. CATASTROPHE THEORY

Critical points are points at any fixed scale in which the gradient vanishes ($\nabla u = 0$). Catastrophe theory is the study of how these critical points change as certain control parameters change. When a control parameter is continuously varied, a Morse critical point will move along a *critical path*. In principle the single control parameter in the models of this article can be identified as the scale of the blurring filter [2].

The only generic morsifications in Gaussian scale space are *creations* and *annihilations* of pairs of Morse hypersaddles of opposite Hessian signature[1] [3]. An example of this is given in Fig. 1.

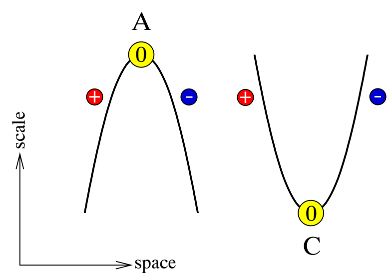


Fig. 1. The generic catastrophes in isotropic scale space. Left: an annihilation event. Right: a creation event.

Thus the movement of critical points through scale, together with their annihilations and creations form *critical paths* in scale space. In this article we will restrict ourselves to generic 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. 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 for generic images there is always one extremum branch continuing up to infinite scale [4]. In Fig. 2 the critical paths and their top points are shown for a picture of a face.

[1]The Hessian signature is the sign of the determinant evaluated at the location of the critical point.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

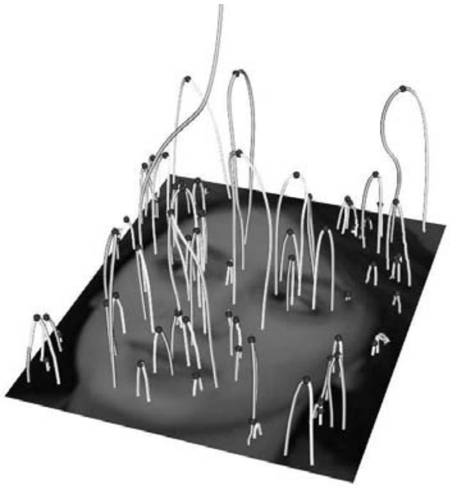


Fig. 2. Critical paths and top points of a face.

3. TOP POINTS

The points at which creation and annihilation events take place are often referred to as *top points*. A top point is a critical point at which the determinant of the Hessian degenerates. A top point therefore has the following properties:

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

Where H denotes the Hessian matrix.

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 subpixel results. Other, more elaborate methods, can be used to find or refine the top point positions. For details the reader is referred to [3].

4. TOP POINT MATCHING

Previous studies have shown that top points contain information about the image [5, 1] and can be used for content based image retrieval [6].

We want to reduce the problem of image matching to top point matching. Instead of comparing two images directly we can now compare them by comparing their sets of top points.

The top point matching problem can be written as a transportation problem and solved using an algorithm like the Earth Movers Distance (EMD).

4.1. Using EMD for matching

Earth Mover's Distance (EMD) reflects the minimal amount of work that must be performed to transform one distribution into the other by moving "distribution mass" around. Very efficient algorithms for calculating the EMD exist. In our experiments we have used an algorithm proposed by Rubner et al. [7].

Without going into details about the EMD and its implementation an intuitive view of the EMD is as follows: Given two distributions, one can be seen as a mass of earth properly spread in space, the other as a set of holes in that same space. Assuming that there is always at least as much earth as needed to fill all the holes (this can be achieved by switching what is called earth or holes if necessary), the EMD gives a measure of the least amount of work needed to fill the holes with earth. One unit of work corresponds to transporting a unit of earth by a unit distance. This means that apart from the distributions we want to match (e.g. the top points) a *weight distribution* and a *distance measure* are important.

4.2. Distance measure

To calculate the distance between two points in scale space we cannot use a simple Euclidean measure, because we are not dealing with a regular three dimensional space, but with a scale space of a 2D image. Eberly [8] proposed a metric that depends on a parameter $\rho > 0$, weighing the relative importance of spatial and scale measurements. The metric is defined by:

$$ds^2 = \sum_{i=1}^{n} \frac{dx_i^2}{\sigma^2} + \frac{1}{\rho^2}\frac{d\sigma^2}{\sigma^2} \qquad (2)$$

In which x_i represent the spatial positions of a point in scale space and σ represents the scale of the point.

In order to define distance between two points (x_1, y_1, σ_1) and (x_2, y_2, σ_2), a geodesic connecting these points should be found. By solving the resulting differential equations with correct boundary conditions [9] the distance S between two points in scale space can be expressed as:

$$S = (r_2 - r_1)/\rho \qquad (3)$$

$$r_i = \ln(\sqrt{b_i^2 + 1} - b_i), \quad i = 1, 2 \qquad (4)$$

$$b_i = \frac{\sigma_2^2 - \sigma_1^2 - (-1)^i \rho^2 R^2}{2\sigma_i \rho R}, \quad i = 1, 2 \qquad (5)$$

Where R is the Euclidean distance between (x_1, y_1) and (x_2, y_2). For two points exactly above each other ($x_1 = x_2$ and $y_1 = y_2$), $S = |\ln(\sigma_2/\sigma_1)|/\rho$.

We now have a distance measure that we can use in our EMD algorithm. But we have also introduced a tunable parameter $\rho(> 0)$, for which an optimal value still has to be found experimentally.

4.3. Stability norm

It is very important that the position of the points we want to match is stable under minor noise perturbations. If the points would move a lot under a small perturbation of the image, a matching scheme based on the position of the points would certainly fail.

It is obvious that the position 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. Problems however do arise in areas in the image that consist of almost constant intensity (genericity implies that flat plateaux do not occur in the image). One can imagine that the position of the extrema (and thus the critical paths and top points) is 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 the vicinity to interact with. To account for these instable 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 variation* (TV) norm:

$$TV(\Omega) \stackrel{\text{def}}{=} \frac{\sigma^2 \int_\Omega \|\nabla u(x)\|^2 dV}{\int_\Omega dV} \quad (6)$$

We calculate the TV norm in a circular area with radius $\lambda\sigma$ around a top point at position (x_c, t_c). Note that the size of the circle depends on the scale σ. The integration area of the TV norm Ω is defined by:

$$\Omega : \|x - x_c\|^2 \leq \lambda^2 \sigma^2. \quad (7)$$

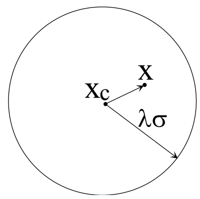


Fig. 3. Integration area around critical point Xc.

Consider a critical point $u(x_c, t_c)$ (For simplicity we shall denote it as u and use the Einstein summation convention). A spatial Taylor series around that point yields:

$$u(x,t) = u + \frac{1}{2} u_{ij} (x - x_c)^i (x - x_x)^j + \mathcal{O}(\|x - x_c\|^3) \quad (8)$$

Note that at the top point the first order derivatives are all zero. If we now calculate the gradient near the top point by using the result of the Taylor series we get:

$$u_i(x,t) = u_{ij}(x - x_c)^j \quad (9)$$

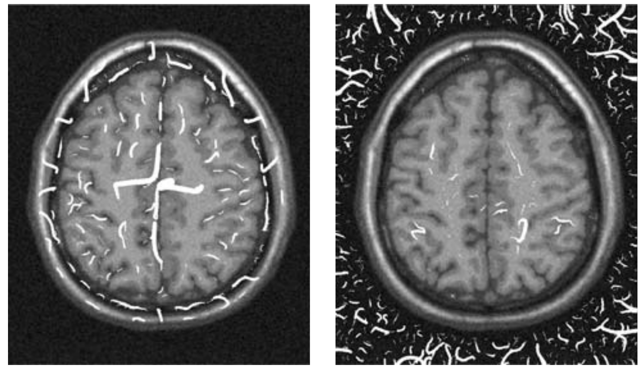
The equation of the TV norm (6) with substitution of the equations (7,9) yields:

$$TV(\lambda) = \frac{\sigma^2 \int \|\nabla u\|^2 d\Omega}{\int_{\|x\|^2 \leq \lambda^2\sigma^2} d\Omega} = \frac{\pi}{4} \lambda^4 \sigma^4 \text{Tr}(H^2) + \text{(h.o.t.)} \quad (10)$$

From this we may define the ”differential TV-norm” by the following limiting procedure:

$$tv \stackrel{\text{def}}{=} \lim_{\lambda \to 0} \frac{4}{\pi} \frac{1}{\lambda^4} TV(\lambda) = \sigma^4 \text{Tr}(H^2) \quad (11)$$

Equation (11) resembles Koenderink’s *deviation from flatness*. It enables us to calculate a stability measure for a top point by using only its second order derivatives.
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. 4.



(a) Stable paths (b) Unstable paths

Fig. 4. Spatial projection of critical paths of a MR brain scan image. The paths are filtered by thresholding the stability norm of their top points. Apparently most instabilities occur in flat regions, as expected.

5. RESULTS

We have tested our matching algorithm with a small faces database from Olivetty Research Laboratory. The test database consists of 200 pictures from 20 different people. Of each person the database contains 10 different images. The differences range from different poses, facial expressions, hair style, presence or absence of glasses, to artificial distortions like skewing parts of the images. These images all have a resolution of 112x92 pixels and contain an average of 80 top points per picture (above the finest scale that we consider).
The experiments we have conducted are not meant to compare our algorithm to other algorithms for face matching, but are merely done to verify the feasibility of using top points for image matching.

Our experiment is a small image retrieval task, given a query image, we want to find ten images that resemble the query image the most. In the best case this would give us all the images of one single person. If the matched image is a picture of the same person as in the query image, we say that the match is correct. Some results of our experiments are depicted in Table 1.
We have compared the Eberly distance described in section 4.2, to the Euclidean distance measure. Experimentally we have found that the value $\rho = 4$ gives us the best results for the Eberly distance measure. The improvement of the amount of correctly matched images using the Eberly distance measure over a normal spatial distance measure is significant as can be seen in Table 1a. and b.
In another experiment we have used the stability measure described in section 4.3, to filter out unstable top points. This reduces the number of top points per picture to an average of 70. The filtering improves the matching results as shown in Table 1c.
Finally we have taken into account the second order derivatives at the top points, so we no longer match solely on the positions of a top point but now also take into account their second order derivatives. We have added this experiment to demonstrate that it is possible to use more features of the top points than only their positions. The incorporation of the second order derivatives in the matching scheme further increase the results as can be seen in Table 1d. A part of these matching results are shown in Fig. 5.

One observes that our algorithm can still match the correct faces even though there are some major differences in the images. Faces are matched with and without glasses and with a lot of dif-

	2	3	4	5	6	7	8	9	10
a	93%	78%	68%	62%	58%	54%	49%	45%	43%
b	93%	82%	80%	73%	70%	68%	63%	59%	56%
c	95%	88%	83%	76%	73%	66%	62%	59%	58%
d	100%	97%	96%	95%	92%	88%	86%	84%	81%

Table 1. The accuracy of the matching, the first column gives the percentage of the second image being matched correctly. The second column gives the percentage of the second and third image being matched correctly, etc. The first image is identical to the query image and is always matched correctly. **a.** Using the Euclidean distance. **b.** Using the Eberly distance. **c.** As b. including the stability norm. **d.** As c. including second order derivatives in the distance measure.

Fig. 5. Matching results.

ferent poses and expressions. Even artificial distortions do not prevent the images from being matched to the correct person, as shown in the 7th row of Fig. 5.
Our algorithm cannot yet cope with extensive zooming, as this has not yet been implemented in the algorithm. This explains some of the inferior results as shown in the last row of Fig. 5.

6. CONCLUSIONS

In this article we have shown that top points can be used for image matching. To match the top points of different images we have successfully used the earth movers distance algorithm.
For this method to work in scale space we needed a distance measure which defines the distance between points in scale space. The measure proposed by Eberly has been used with success and performs better than the often naively used Euclidean distance measure.
To get rid of unstable top points that deteriorate the matching results we have derived a stability measure, based on the total variation norm, that describes the stability of a top point. By only using the second order derivatives at the top point position this stability norm enabled us to filter out any unstable top points.
The results have shown that the EMD, the distance measure and the stability norm contribute to a working matching scheme. It has also been shown that more information than solely the top point positions can be taken into account. In a small demonstrative experiment we have added the second order information of the top point to our matching scheme. This increased the matching results even further.
On the whole we can confirm that using top points for matching images, appears promising. As we have not yet explored every detail of top point matching we expect that with some further improvements, this method may well compete with existing image matching algorithms and thus provide a complete new matching paradigm.

Acknowledgement: This work is part of the DSSCV project supported by the IST Programme of the European Union (IST-2001-35443).

7. REFERENCES

[1] M. Nielsen and M. Lillholm, "What do features tell about images?," In Kerckhove [10], pp. 39–50.

[2] J. Damon, "Local Morse theory for solutions to the heat equation and Gaussian blurring," *Journal of Differential Equations*, vol. 115, no. 2, pp. 368–401, January 1995.

[3] L. Florack and A. Kuijper, "The topological structure of scale-space images," *Journal of Mathematical Imaging and Vision*, vol. 12, no. 1, pp. 65–79, February 2000.

[4] M. Loog, J. J. Duistermaat, and L. M. J. Florack, "On the behavior of spatial critical points under Gaussian blurring. a folklore theorem and scale-space constraints," In Kerckhove [10], pp. 183–192.

[5] F.M.W. Kanters, L.M.J. Florack, B. Platel, and B.M. ter Haar Romeny, "Image reconstruction from multiscale critical points," in *Proceedings of the 4th international conference on Scale Space Methods in Computer Vision (Isle of Skye, UK, June 2003)*, pp. 464–478.

[6] F.M.W. Kanters, B. Platel, L.M.J. Florack, and B.M. ter Haar Romeny, "Content based image retrieval using multiscale top points," in *Proceedings of the 4th international conference on Scale Space Methods in Computer Vision (Isle of Skye, UK, June 2003)*, pp. 33–43.

[7] Y. Rubner, C. Tomasi, and L.J. Guibas, "A metric for distributions with applications to image databases," in *Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India*, 1998, pp. 59–66.

[8] D. Eberly, "A differential geometric approach to anisotropic diffusion," in *Geometry-Driven Diffusion in Computer Vision*, B. M. ter Haar Romeny, Ed., vol. 1 of *Computational Imaging and Vision Series*, pp. 371–392. Kluwer Academic Publishers, Dordrecht, 1994.

[9] J. J. Wijk and W. A. A. Nuij, "A model for smooth image viewing," Accepted for publication, 2004.

[10] M. Kerckhove, Ed., *Scale-Space and Morphology in Computer Vision: Proceedings of the Third International Conference, Scale-Space 2001, Vancouver, Canada*, vol. 2106 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, July 2001.