# Are MSER features really interesting?

R. Kimmel *Fellow, IEEE,* C. Zhang, A. Bronstein *Member, IEEE,* M. Bronstein *Member, IEEE,*

**Abstract**—Detection and description of affine-invariant features is a cornerstone component in numerous computer vision applications. In this note, we analyze the notion of maximally stable extremal regions (MSER) through the prism of the curvature scale space, and conclude that in its original definition, MSER prefers regular (round) regions. Arguing that interesting features in natural images usually have irregular shapes, we propose alternative definitions of MSER which are free of this bias, yet maintain their invariance properties.

**Index Terms**—MSER, feature detector, affine invariance, stable region, correspondence.

## 1 Introduction

In recent years, feature descriptors extracted through linear scale-space analysis of an image have proven to be a powerful tool in object matching and recognition [1]. One of the most popular descriptor is the *scale-invariant feature transform* (SIFT) introduced by David Lowe [2]. It first locates points of interest in a linear scale-space, and then assigns a descriptor vector constructed as local histograms of image gradient orientations around the point. The descriptor itself is oriented by the dominant gradient direction, which makes it rotation-invariant. SIFT uses linear scale-space in order to search for feature points that appear at multiple resolutions of the image, which makes the method also scale-invariant.

One of the main disadvantages of SIFT is that is is not affine-invariant (see a recent work of [3] on an affine-invariant version of SIFT). An affine-invariant alternative to the SIFT widely used in computer vision applications is the *maximally stable extremal region* (MSER) [4]. This approach extracts stable regions from the image by considering the change in area with respect to the change in intensity of a connected component defined by thresholding the image at a given gray level. The change of area, normalized by the area of the connected component, is used as the stability criterion. The area ratio is invariant to affine transformations and so does the the extracted region after appropriate canonization. [1]Benchmarks comparing the MSER, SIFT, other approaches, and affine-invariant alternatives thereof [9], [10] show that SIFT performs well for planar objects (like a graffiti wall) while the MSER performs better in most scenarios involving less trivial objects.

In this paper, we relate MSER to geometric scale-space analysis and image evolution by the level set curvature flow. We observe that the stability criterion in the original formulation of MSER prefers regular regions, and arguing that interesting features in natural images usually have irregular shapes, propose alternative definitions of MSER which are free of this bias, yet maintain their invariance properties. The rest of this paper is organized as follows. In Section 2, we briefly overview the theory of image representation as level sets and curve evolution. In Section 3 we formulate the MSER algorithm and analyze its preference for round regions. Section 4 is dedicated to shape normalization. Section 5 defines alternatives to the MSER stability criterion. Section 6 shows experimental results. Finally, Section 7 concludes the paper.

## 2 Image as a collection of level sets

Let $X \subset \mathbb{R}^2$ be a domain on which a grayscale image $I : X \to [0,1]$ is defined. Every image can be fully represented as a collection of its level sets. A *level set* of $I$ at some given $t \in [0,1]$ is the set $\{x \in X : I(x) = t\}$. Topologically, a level set may contain zero or more connected components of dimension zero (points) or one (isolines).

Thinking of $t$ as time and observing the evolution of the level sets over time, we will see connected components appear, split, change genus, join and disappear. The study of the changes of topology of the level sets with infinitesimal changes of $t$ belongs to the domain of Morse theory, a branch of differential topology. The *contour* or *component graph* of $I$ is a graph in which

• *R.K., A.B. and M.B. are with the Department of Computer Science, Technion – Israel Institute of Technology, Haifa, 32000, Israel. Email: ron@cs.technion.ac.il, alexbron@ieee.org, bronstein@ieee.org.*

1. See [5], [6] for a closely related approach that also allows for the analysis of contour segments, as well as [7], [8] for an axiomatic framework of differential affine-invariant signatures of planar shapes.

(i) a leaf vertex represents the creation or deletion of a component; (ii) an interior vertex represents the joining/splitting of two or more components; and (iii) an edge formed by two vertices with $t = t_1$ and $t = t_2$ represents a component in the level sets for all values of $t_1 \leq t \leq t_2$. This graph recording the topological events in the level set evolution was shown to be a tree. Each edge of the component tree represents the evolution of a single connected component in some contiguous range of values of $t \in [t_1, t_2]$. We will denote such components by $\partial R_t$ implying the entire sequence $\{R_t\}_{t=t_1}^{t_2}$; $\mathrm{int}(R_t)$ will denote the open set in $X$ enclosed by $\partial R_t$, and $R_t$ will denote the union of the two (the region with its boundary). Components $R_t$ along the edge are *nested* inside each other.

### 2.1 Curvature flow and geometric scale space

In the SIFT method, interesting feature points are located by looking for local maxima of the discrete image Laplacian at different scales obtained by convolving the image with Gaussians of different variances. This procedure is known as linear scale-space analysis. While providing SIFT with scale-invariance qualities, the linear scale-space breaks the geometric relation between images of the same scene captured at different view points, in particular, it is not affine-invariant. Moreover, it is well known that such a scale-space does not necessarily simplify the image structure. This is especially acute when level sets are considered, as linear scale space can disconnect simply connected shapes [11], [12].

Better scale-invariant quantities that are simplified with scale are provided by the curvature scale-space or its affine variations [13], [14], [15], [16], [17], [18]. Yet, involving a non-linear heat flow, the construction of a geometric scale-space may seem to be more demanding computationally. The question we try to answer in this section is whether we can use the structure provided by geometric scale-space without explicitly computing it, a property that was trivially accomplished for the linear scale-space.

In the construction of the curvature scale-space of an image, the image level sets are propagated by their curvature vector. Let $C(s) : [0, L] \to \mathbb{R}^2$ be an arclength-parameterized contour. Then, the curvature flow for the contour is given by

$$C_t(s) = C_{ss},$$

where $C_{ss} = \kappa \vec{n}$ is the curvature vector, normal to the curve at $C(s)$. The whole process can be evaluated simultaneously for all the level sets using the remarkable property proven by Grayson [14] that embedding is preserved along the curvature flow and no self-intersections occur until the contour vanishes at a circular point. The equation governing the image evolution is given by

$$I_t = \mathrm{div}\left(\frac{\nabla I}{\|\nabla I\|}\right)\|\nabla I\|,$$

and can be easily established by the Osher-Sethian level set formulation [16]. Another important property of this flow is that each level set contour vanishes at a time proportional to its area at $t = 0$ [15], [14].

The topology of the curvature scale-space can be again captured entirely by a component tree, identical to the one we defined before.

## 3 Maximally Stable Extremal Regions

Let $R_t$ be the family of connected components representing an edge in the component tree. Matas *et al.* refer to such regions as to *extremal* since either $I|_{\mathrm{int}(R_t)} < I|_{\partial R_t}$ or $I|_{\mathrm{int}(R_t)} > I|_{\partial R_t}$, i.e., all the pixel values in the regions are either strictly darker or strictly brighter than those on the boundary, where the intensity is exactly equal to $t$.

The *stability* of a region $R_t$ is defined as

$$\Psi_1(R_t) \quad = \quad \frac{A(R_t)}{\frac{d}{dt}A(R_t)},$$

where $A(R_t)$ denotes the area of $R_t$. A region is considered stable if its area changes only slightly with the change of the threshold $t$. A region $R_t$ is called *maximally stable* if $\Psi_1(R_t)$ has a local maximum at $t$. Such regions are image features detected by the MSER algorithm.

In [4], Matas *et al.* showed that MSER is affine-covariant. This observation stems directly from the fact that area ratios are preserved under affine transformations, which implies that $\Psi_1(R_t)$ is an affine-invariant property. This, in turn, implies that for an affine transformation $T$ of the domain $X$, the corresponding regions $R$ and $R'$ detected in images $I$ and $I(T^{-1})$, respectively, will be related by $TR = R'$.

### 3.1 Stability and shape factor

Let us now look closer at the stability measure $\Psi_1$. Observing that

$$\frac{dA(R_t)}{dt} \quad = \quad \lim_{\epsilon \to 0} \frac{1}{\epsilon}(A(R_{t+\epsilon}) - A(R_t)) \quad = \quad \int_{\partial R_t} \frac{ds}{\|\nabla I\|},$$

we re-write $\Psi_1$ as

$$\Psi_1(R_t) \quad = \quad \frac{A(R_t)}{\displaystyle\int_{\partial R_t} \frac{ds}{\|\nabla I\|}}.$$

Let us now apply $\Psi_1$ to two equal-area regions, one is a perfect circle, while the other is a more interesting less round shape. Under the simplifying assumption that the change of intensity along the boundaries is the same in both regions, say $\|\nabla I\| = 1$, we have

$$\Psi_1(R_t) \quad = \quad \frac{A(R_t)}{\displaystyle\int_{\partial R_t} ds} \quad = \quad \frac{A(R_t)}{L(\partial R_t)},$$

where $L(\partial R_t)$ is the boundary length of $R_t$. Similar to the *shape factor* $\frac{4\pi A}{L^2}$ which is always smaller or equal to one with equality achieved for the circle, the ratio $\frac{A}{L}$ prefers regular shapes. In fact, $\Psi_1$ is maximized by a large circle, and in general, for two shapes with the same area and same change of intensity along their boundaries, the one with a shorter boundary would be preferred by $\Psi_1$. However, such shapes are not necessarily the most interesting and descriptive features in a natural image, in which interesting features typically have irregular boundaries.

Based on this observation, our goal is to correct the bias of $\Psi_1$ towards round shapes and define an alternative stability measure that prefers less regular and more interesting shapes while still enjoying the affine invariance and stability of $\Psi_1$.

### 3.2 Non-commutativity with blur

Affine covariance of maximally stable regions is the consequence of covariance of the level sets of the image with affine transformations of the coordinates. However, this property holds only if the boundaries of objects in the scene are smooth, which is violated in real-world scenarios. Specifically, for the affine covariance of the level sets to hold, we need the optical point spread function of the camera to be small compared to the natural smoothness of objects in the scene. In other words, we need to assume that the world is blurred to begin with, and that the image formation is primarily a geometric transformation of that blurred image of the world. A more realistic model is to assume that blur occurs *after* the geometric transformation. Figure 1 demonstrates the two cases, where in the upper row smoothing occurs in the imaging phase, while at the bottom row the boundaries are blurred to begin with and the imaging process is modeled as an affine transformation. In other words, real view point transformations constitute (locally) affine transformations followed by blur in the image plane with the point spread function of the camera, and these two constituents do not commute.

As in most practical cases the image formation involves non-negligible blur due to the optical acquisition process, it may happen that the criterion $\Psi_1$ is not truly invariant to view point transformations. In fact, a much better quantity for the stability or edginess of a region would be the weighted gradient magnitude along its boundary. Here weight could be the affine arclength $dv = |\kappa|^{1/3} ds$ for an affine-invariant measure, that explicitly yields

$$\Psi_2(R_t) \quad = \quad \frac{A(R_t)}{\displaystyle\int_{\partial R_t} \frac{\|I_{xx} I_y^2 - 2 I_x I_y I_{xy} + I_{yy} I_x^2\|^{1/3} ds}{\|\nabla I\|}},$$

or any alternative robust filter like the median could represent the significance of the boundary sufficiently well [2].

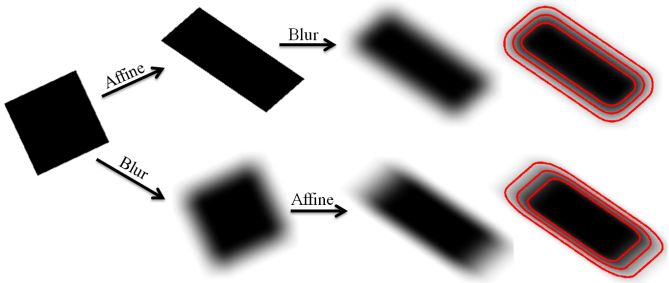


Fig. 1. Top row assumes affine transformation followed by imaging blur. Bottom row, assumes affine transformation of a given blurred object. On the right are three corresponding level sets for both cases.

## 4 Affine-Invariant Normalization

In typical applications, maximally stable regions found by MSER undergo a process of affine-invariant normalization or canonization. Normalization can be though of as a mapping $N : \mathbb{R}^2 \to \mathbb{R}^2$ receiving a region $R$ and returning another region $N(R)$ such that $N(TR) = N(R)$ for any affine transformation $T$. Canonization of a given shape can be viewed as part of a descriptor computation in which the goal is to compensate for arbitrary transformations of the shape due to the acquisition process.

2. Note that the two basic independent affine-invariant second order differential descriptors are $J(I) = I_{xx} I_y^2 - 2 I_x I_y I_{xy} + I_{yy} I_x^2$, and the determinant of the hessian $H(I) = I_{xx} I_{yy} - I_{xy}^2$ [19], while the second order approximation for the affine-invariant curvature of the level sets is given by $\mu = H/J^{2/3}$ [20].

In [6], Cao et al. argue that normalization of a planar shape that compensates for affine transformations and is based on second-order moments can be unstable. The authors propose alternatives based on the detection of flat intervals along the boundary. The next steps applied by Cao et al. involve center of mass estimation for the two regions created by a line parallel to the flat boundary line that goes through the center of mass. Parallel lines, area ratio, and center of mass are indeed robust measures preserved by affine transformations. On the other hand, a definition of flatness that is based on Euclidean distance and angles is not invariant to affine transformations. Moreover, if we limit our discussion to the analysis of simple closed contours there is a simple alternative for the first step propose in [6].

Experimenting with second order moments based normalization [21] we did not experience the instabilities reported by Cao et al. In fact, the moments based normalization proved to be equally stable as the centers of mass based alternative as can be seen in Figure 2. The method we propose in this section could be used to either initialize the Cao et al. canonization method or as compensation for the rotation ambiguity in moments based normalization.

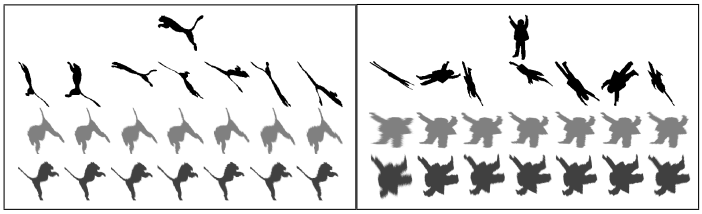

Fig. 2. The original silhouettes of the Puma logo and a boy appear at the top, and their random affine transformations sampled to low resolution $64 \times 64$ patches at the second row in black. The normalized shapes with second-order moments appears in dark gray (bottom row) while the alternative method proposed by Cao et al. is presented in light gray (third row).

Let us assume that the contours we would like to normalize are interesting and therefor non-convex. In fact, convex contours could be classified by the simplest regular polygons that approximate the shape. A rough affine-invariant canonical approximation for convex shapes could be triangles, squares, and circles that represent the rest of the regular polygons. Relying on area ratios and centers of mass, and based on [6], we define a robust affine-invariant method for mapping a given contour into its canonical normalized shape. The steps of the method are as follows:

1) Compute the convex hull of the shape.
2) Find the largest area bounded between the convex hull and the given shape, and use the bitangent line which is part of the convex hull touching the largest area for the next steps (see Figure 3).
3) Next we follow the rest of the steps in [6] using the computed bitangent as the reference axis, see Figure 4.

The reference axis could also be used for compensating for rotation ambiguity in the case of moments based normalization [21]. Using moments based normalization, first the normalization is performed, and then the above rotation cancelation using the convex hull and maximal bounded area is applied.

There are other options to account for rotations, like radial Fourier transform over the shape and consideration of the phase as a rotation angle. Yet, the best computational complexity for the convex hull of a closed contour is $O(n \log h)$ where $h$ defines the number of points in the convex hull ($n > h$), see [22], while the Fourier transform is slightly more costly and requires $O(n \log n)$ operations.

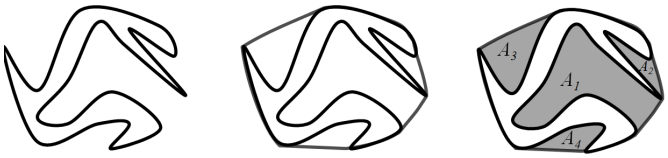


Fig. 3. Left to right: The shape's boundary contour, its convex hull, and the areas formed between the convex hull and the shape. The largest area, $A_1$ in this case, defines the bitangent that is used for normalization (canonization) of the shape or for fixing its orientation.

## 5 Interesting Features

In order to better treat interesting non-convex shapes, we propose a stability measure as an alternative to $\Psi_1$. Unlike the standard MSER where $\Psi_1$ is computed on the components from the component tree, we propose to first normalize each component. We then compute the inverse of the standard Euclidean shape factor

$$\Psi_3(R_t) = \frac{L^2(N(R_t))}{A(N(R_t))},$$

where the operator $N$ means that the measure is applied to the normalized region. Such a function prefers shapes with irregular boundaries, while being affine-invariant.

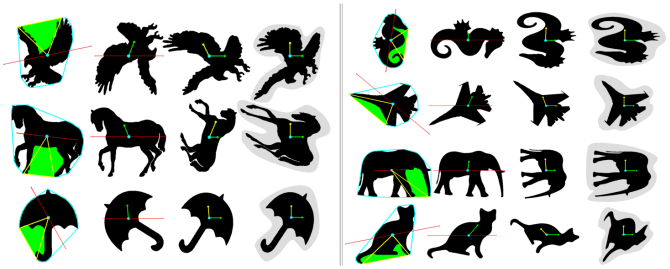

Fig. 4. Normalization steps of a given shape, left to right: Convex hull and maximal bounded area detection, rotation of the parallel to the bitangent through the center of mass, alignment of the center of mass of the upper half of the shape with the $x$-axis, and finally shear of the center of mass of the (new) upper part so that the line connecting it to the center of mass aligns with the $y$ axis. The resulting normalized shape is at the right of each sequence.

Finally, an affine-invariant stability measure for interesting shapes could combine the above measure with $\Psi_1$, like

$$\Psi_4(R_t) = \Psi_1(R_t)\Psi_3(R_t) = \frac{A(R_t)L^2(N(R_t))}{A(N(R_t))\frac{d}{dt}A(R_t)}.$$

Since the computational complexity of region normalization is proportional to the length of the boundary, reversing MSER selection and normalization is not more computationally expensive than first computing $\Psi_1$ and then doing normalization of the remaining MSERs.

## 6 Experimental Results

The goal of our first experiment is the validation of the affine-invariant level set normalization. We applied the modified canonization based on convex hull, maximal bounded area and centers of mass to random affine transformations of two silhouettes collected from the web. Figure 5 demonstrates the fact that various transformations of the same object all lead to a similar canonical shape.

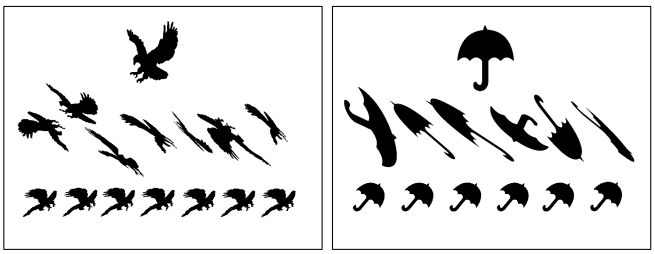

Fig. 5. In each frame a silhouette appears at the top, its random affine transformations in the middle row and their corresponding normalized shapes at the bottom.

The second experiment demonstrates the improved feature matching using a modified MSER, in which the average gradient along the contour is used as an estimation for stability. Figure 6 shows feature matching in an object taken from two video frames of a movie. The MSER regions are normalized and matched based on their canonized shapes, and for each pair the first three matches are considered. The final selection is of features that are supported by consistent neighboring features that are determined by the first ten nearest neighbors. The improvement in performances shows up in the correspondence of features in the two frames as can be seen in Figure 7.

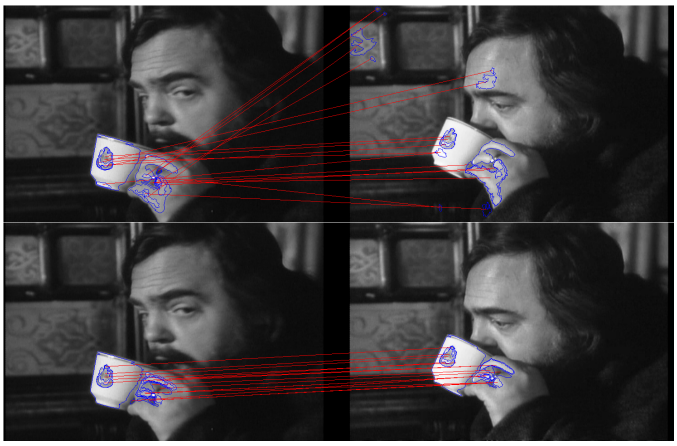

Fig. 6. The top frame demonstrates matching with the classical MSER, while the bottom frame shows the result of a modified stability criteria.

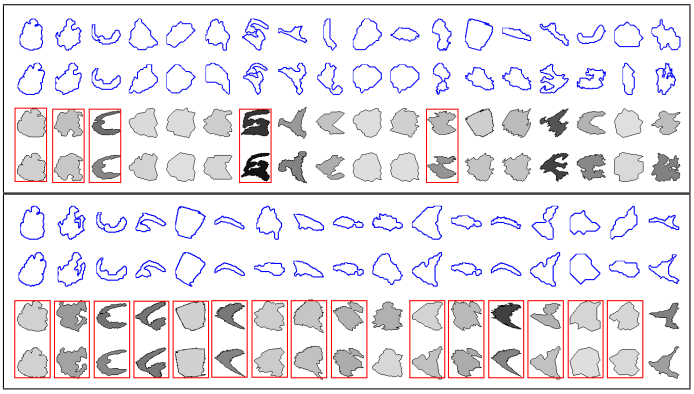

Fig. 7. The top frame demonstrates the matching pairs extracted with the classical MSER. First row: regions found in the first frame. Second row: the matching regions in the second frame. Third row: normalized regions (first frame). Bottom row: Matched normalized regions in the second frame. The order (left to right) is according to the matching score, while the gray level of the canonical shapes corresponds to the isometrimetric ratio. Correct matches appear in a red box. Bottom frame repeats the experiment with the modified stability criterion.

## 7 Conclusions

We stress again the amazing fact that while being only Euclidean-invariant, the curvature scale-space structure is captured by the level set graph which is affine- (and projective-) invariant. This property explains the usefulness of the image level sets and their local density in generating interesting features. The relation between the level set graph, curvature flow, and invariant stable and interesting features provides a theoretical bridge that could be used for various image and shape analysis applications. Finally, we revisited the assumptions of the MSER and redefined some of the criteria that help us extract more informative shape descriptors.

## References


[1] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. of ICCV*, 2003.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[3] G. Yu and J. Morel, "A fully affine invariant image comparison method," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, 2009, pp. 1597–1600.

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. BMVC*, 2002, pp. 384–393.

[5] A. Desolneux, L. Moisan, and J. Morel, "Edge detection by helmholtz principle," *JMIV*, vol. 14, pp. 271–284, 2001.

[6] F. Cao, J. Lisani, J. Morel, P. Musé, and F. Sur, *A theory of shape identification*, ser. Lecture Notes in Mathematics. Springer, 2008, vol. 1948.

[7] A. M. Bruckstein, R. J. Holt, A. Netravali, and T. J. Richardson, "Invariant signatures for planar shape recognition under partial occlusion," *CVGIP: Image Understanding*, vol. 58, no. 1, pp. 49–65, 1993.

[8] A. M. Bruckstein and D. Shaked, "Skew symmetry detection via invariant signatures," *Patrn. Rec.*, vol. 31, no. 2, pp. 181–192, 1998.

[9] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. van Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1/2, pp. 43–72, 2005.

[10] P. Forssèn and D. Lowe, "Shape descriptors for maximally stable extremal regions," in *IEEE ICCV*, Rio de Janeiro, Brazil, 2007.

[11] in *Geometric–Driven Diffusion in Computer Vision*, B. M. ter Haar Romeny, Ed. The Netherlands: Kluwer Academic Publishers, 1994.

[12] F. Guichard and J. Morel, "Image analysis and P.D.E.s," *IPAM GBM Tutorial*, 2001.

[13] B. B. Kimia, "Toward a computational theory of shape," Ph.D. Dissertation, Department of Electrical Engineering, McGill Univ., Montreal, 1990.

[14] M. A. Grayson, "The heat equation shrinks embedded plane curves to round points," *J. Diff. Geom.*, vol. 26, 1987.

[15] M. Gage and R. S. Hamilton, "The heat equation shrinking convex plane curves," *J. Diff. Geom.*, vol. 23, 1986.

[16] S. J. Osher and J. A. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. of Comp. Phys.*, vol. 79, pp. 12–49, 1988.

[17] G. Sapiro, "Topics in shape evolution," D.Sc. Thesis, Technion - Israel Institute of Technology, 1993.

[18] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, "Axioms and fundamental equations of image processing," *Arch. Rat. Mech.*, vol. 123, pp. 199–257, 1993.

[19] P. J. Olver, *Equivalence, Invariants, and Symetry*. Cambridge Univ. Press, 1995.

[20] R. Kimmel, "Affine differential signatures for gray level images of planar shapes," in *Proc. of ICPR*, Vienna, 1996.

[21] M. Hu, "Visual pattern recognition by moment invariants," *IRE T. on Inf. Theory*, vol. 8, pp. 179–189, 1962.

[22] D. G. Kirkpatrick and R. Seidel, "The ultimate planar convex hull algorithm," *SIAM J. on Computing*, vol. 15, no. 1, pp. 287–299, 1986.