Affine Invariant Interesting Descriptors

Ron Kimmel, Cuiping Zhang, Alexander M. Bronstein, and Michael M. Bronstein

August 19, 2009

Abstract

Component trees or level set graphs have proved to be efficient tools for the extraction of stable features in many image analysis and understanding applications. In this paper, we explore the relation between level sets as feature descriptors and curvature scale-space. While most classical image feature detectors are based on linear scale-space, affine-invariant descriptors can be related to the curvature scale-space, in which closed level sets vanish at time proportional to their area and the connectedness is preserved. We provide observations about the affine invariance of objects in an image and design a framework for extracting more informative invariant features.

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 [25]. One of the most popular descriptor is the shift invariant feature transform (SIFT) introduced by David Lowe [18]. It first locates points of interest in a linear scale-space, and then assigns a 128-dimensional 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.

In many cases, this invariance is insufficient, as objects appearing in the scene are three-dimensional, and image formation includes more complicated geometric transformations, which can be approximated by locally-affine transformations. The SIFT descriptor can be made invariant to such transformations by applying local affine normalization for regions around the feature points. However, the
SIFT feature detector which is based on linear scale space is not affine-invariant. Recently, Yu and Morel [26] proposed an affine-invariant version of SIFT, referred to as affine SIFT (ASIFT).

Another affine invariant alternative to the SIFT is the *maximally stable extremal region* (MSER) [19]. 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. See [7, 4] for a closely related approach that also allows for the analysis of contour segments, as well as [2, 3] for an axiomatic framework of differential affine invariant signatures of planar shapes. Benchmarks comparing the MSER, SIFT, other approaches, and affine invariant alternatives thereof [20, 9] 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 explore the relation between object boundaries, level sets, and affine invariant features, linking between seemingly unrelated problems and computational tools. We relate the MSER to geometric scale-space analysis and image evolution by the level set curvature flow, and based on analysis of image formation propose alternatives to the stability criterion used in the MSER. The structure of the paper is as follows. In Section 2, we start with the notion of level sets and some basic concepts in curve and surface geometry and the representation of an image as a level set graph [6], in which each simple level set contour is a node. In Section 3 we provide a brief description of the curvature flow and its resulting geometric scale-space for images. Section 4 presents the MSER in light of its relation to the geometric structure of the image as a composition of level sets and the curvature scale-space. In Section 5, we revisit the definition of MSER and without giving up its invariance and stability properties, provide a new stability criterion that extracts more interesting features out of a given image. We consider a simple image formation model according to which an optical blur occurs after the geometric transformation due to change in viewing direction. Based on this simplified model we propose to reconsider the assumption that the differential affine invariance of the image level sets holds along the boundaries of so-called stable features. Section 6 provides some examples and Section 7 concludes the paper.
2 Image as a collection of level sets

Consider a gray level image $I : \Omega \subset \mathbb{R}^2 \to [0, 1]$ to be a scalar smooth function, where $I = 0$ corresponds to black and $1$ to white. The image can be represented as a collection of its equal height contours or level sets. Let

$$R(x, y, v) = \{I(x, y) < v\} = \begin{cases} 1 & \text{if } I(x, y) < v, \\ 0 & \text{if } I(x, y) \geq v. \end{cases}$$

be the indicator function of level set $v$. The boundaries $\partial R(\cdot, \cdot, v)$ of the binary shapes formed by $R$ at each level $v$ are the level sets of the image function $I(x, y)$. Next, collapse each simple level set contour (i.e. the boundaries of a connected component of $R$) into a point. The embedding relation of level sets induces the connectivity between the points, while each point is characterized by the corresponding area bounded within its corresponding level set. For practical reasons, we restrict our discussion to a finite number of gray levels $v_i = \frac{i}{n}$ for $i \in \{0, 1, \ldots, n\}$. Each level set is represented as a vertex and edges are connecting the level sets $\partial R(x, y, v_i)$ to the level sets $\partial R(x, y, v_{i \pm 1})$. The obtained graph is called the level set graph.

One could consider level set graphs as the intersection of the epigraph and hypograph of an image,

$$\text{epi } I = \{(x, y, v) : (x, y) \in \mathbb{R}^2, v \in \mathbb{R}, v \geq I(x, y)\},$$
$$\text{hyp } I = \{(x, y, v) : (x, y) \in \mathbb{R}^2, v \in \mathbb{R}, v \leq I(x, y)\}.$$

Note that the level sets are given by the intersection sets of the two, while the structure of the level set graph is the union of the two corresponding topology graphs [5]. Compact representations of these structures as trees, also known as component trees, are in use in watersheds in mathematical morphology, and the construction of shape descriptors in the MSER, see for example [8].

3 Curvature flow and geometric scale-space

In the SIFT method, 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 [23, 12].
Better scale-invariant quantities that are simplified with scale are provided by the curvature scale-space or its affine variations \([14, 11, 10, 22, 24, 1]\). 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 the image, the image level sets are propagated by their curvature vector. Let \( C(s) : [0, L] \rightarrow \mathbb{R}^2 \) 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 [11] 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

\[
I_t = \text{div} \left( \frac{\nabla I}{|\nabla I|} \right) |\nabla I|,
\]

and can be easily established by the Osher-Sethian level set formulation [22]. Another important property of this flow is that each level set contour vanishes at a time proportional to its area at \( t = 0 \) [10, 11]. We notice that the level set graph (with area at each vertex) is a compact way for representing the structure of the curvature scale-space for the whole image.

4 MSER

Using our terminology, stable regions extracted by the MSER are defined as vertices of the component tree for which the change of area while traversing the component tree, is relatively small. Specifically, the authors in [19] normalize the change in area by the area of the connected component, and search for locally minimal ratio along the same branch of the component tree. A connected component in the MSER is defined by the threshold set \( R^v = \{ I(x, y) < v \} \) which is somewhat different from the simply connected level set contours that define the vertices of the level set graph. For that reason, and in order for the stable regions to be invariant to contrast inversion, the process is usually applied to the image and then to its negative. The results are referred to as MSER+ and MSER-, respectively.
The level set graph we propose to explore is the intersection of the two component trees constructed for the MSER+ and the MSER-. The proposed level set graph was also referred to as level lines in a topology map and the related descriptors as level line invariant descriptors (LLD) by Cao et al. in [4]. In [4], the analysis of level lines of interest is based on the integral of the gradient magnitude $|\nabla I|$ along the level set. It is in fact closely related to MSER as the differential change of area used in the MSER is nothing but the integral of $1/|\nabla I|$ along the boundary. The local description allows Cao et al. to analyze open contours rather than limiting the analysis to simple closed contours (or shapes as in MSER). Bitangents, an affine invariant structure, were used in order to partition a boundary of a shape into segments that could be canonized with respect to an affine transformation more than two decades ago (see, for example, Lamdan et al. [17]). The link between boundaries of shapes and level sets lead us to adopt these tools and geometric structures for the extraction of interesting features and computation of invariant descriptors.

Let $R_v$ represent one binary shape (connected component) defined by the threshold set $\{(x, y) : I(x, y) < v\}$. Let $A(R_v)$ define the area of that shape, and $dA(R_v)/dI$ be the change in area of the shape as a function of the change in intensity. Then, the stability measure

$$ \Psi_1(R_v) = \frac{dA(R_v)/dI}{A(R_v)} $$

is used in order to extract stable regions (the smaller $\Psi_1$, the better). This is an affine invariant quantity as it involves a ratio of areas.

As will be shown in the following, the stability criterion $\Psi_1$ tends to prefer round shapes, which are not always interesting and discriminative features in the image. Furthermore, $\Psi_1$ may fail to be invariant under a realistic image formation model.

## 5 Interesting stable regions

Let us look closely at the stability measure $\Psi_1(\cdot)$. Specifically, apply $\Psi_1$ to two shapes for which the area is the same, yet, one is a circle while the other is a more interesting less rounded shape. In general, we would like our descriptor to prefer more interesting shapes. Next, assume that the change of intensity along the boundaries is the same, say $|\nabla I| = 1$, for both shapes. Then, we have that

$$ \frac{dA}{dI} = \lim_{\delta I \to 0} \frac{\int_0^L \int_0^{\delta I} \frac{1}{|\nabla I|} dIds}{\delta I} $$
\[
\lim_{\delta I \to 0} \int_0^L \left( \int_0^{\delta I} dI \right) ds
= \lim_{\delta I \to 0} \int_0^L ds = L.
\]

That is, the change in area is proportional to the length of the shape we are exploring, while the stability measure under our over simplified assumptions is proportional to
\[
\Psi_1 \propto \frac{L}{A}.
\]

Similar to the isoperimetric inequality \((4\pi A \leq L^2\) with equality achieved for the circle), the ratio \(\frac{L}{A}\) prefers more regular shapes. \(\Psi_1\) is minimal for a circle, and in general, for two shapes with the same area and same change of intensity along their boundaries, the one with shorter perimeter would be preferred by \(\Psi_1\). However, such shapes are not necessarily the most interesting and descriptive features in a natural image. Typically, interesting features have irregular boundaries.

Based on this observation, we would like to change the bias of \(\Psi_1\) towards rounder shapes and define a measure that prefers less regular and more interesting shapes while still enjoying the affine invariance and stability of \(\Psi_1\). For that goal, we explore alternative affine invariant measures that could be used in order to extract more interesting shapes. One example is the convexity of the shape, expressed as the ratio of the area \(A\) of the shape and the area \(A_{CH}\) of its corresponding convex hull, \(\Psi_2 = \frac{A}{A_{CH}}\). The multiplication of the edge indication measure \(\Psi_1\) and the convexity measure \(\Psi_2\), yields a new measure in which the shape’s area cancels out, \(\Psi_1 \Psi_2 = \frac{dA/dI}{A_{CH}}\). Still, among shapes with the same convex hull and the same image gradient magnitude along the boundary, the one with the shorter boundary would be preferred. Note that the convex hull can be computed in almost linear time for simple polygons, and the ratio of the area of the convex hull and that of the shape is affine invariant.

### 5.1 Affine invariance and stable regions

Affine invariance, by which stable regions are extracted in the MSER procedure, assumes invariance of the level sets of the image to affine transformations of the coordinates. This property holds only if the boundaries of objects in the scene are smooth. Specifically, for the affine invariance assumption to hold between level sets, 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.

As in most practical cases the image formation involves blur due to optical acquisition process, it may happen that the criterion $\Psi_1$ is not invariant to geometric transformations such as change of the camera view point. 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 integration over $|I_{xx}I_y^2 - 2I_xI_yI_{xy} + I_{yy}I_x^2|^{1/3}$, or any alternative robust filter like the median could represent the significance of the boundary sufficiently well.\footnote{Note that the two basic independent affine invariant second order differential descriptors are $J(I) = I_{xx}I_y^2 - 2I_xI_yI_{xy} + I_{yy}I_x^2$, and the determinant of the hessian $H(I) = I_{xx}I_{yy} - I_{xy}^2$ \cite{21}, while the second order approximation for the affine invariant curvature of the level sets is given by $\mu = H/J^{2/3}$ \cite{15}.}

Figure 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.

5.2 Affine invariant shape normalization

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. In [4], Cao et al. argue that normalization (canonization) 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 [4].

Experimenting with second order moments based normalization [13] 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.

![Figure 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 × 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).](image)

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 [4], 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:

8
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 [4] 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 [13]. 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 [16], while the Fourier transform is slightly more costly and requires $O(n \log n)$ operations.

![Figure 3](image)

Figure 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.

In order to better treat interesting non-convex shapes we can adopt a measure that is often used in the Euclidean world in order to classify how far a shape is from being a circle. We apply the isoperimetric ratio ($\frac{\text{area}}{\text{length}^2} \leq \frac{4\pi}{4\pi^2}$) to the normalized canonical shape. We can therefore use the measure

$$\Psi_3 = \frac{A_N}{L_N^2},$$

where the subscript $N$ indicates that the measure is taken with respect to the normalized affine invariant canonical shape.
Figure 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 measures, like $\Psi_4 = \Psi_1 \Psi_3 = \frac{A_N}{L_N^2 A^r}$.

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.

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.
Figure 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.

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


