HAL
open science

Interest Point Detection with Wavelet Maxima Lines

Christophe Damerval, Sylvain Meignen

To cite this version:

Christophe Damerval, Sylvain Meignen. Interest Point Detection with Wavelet Maxima Lines. 2007. inria-00171678

HAL Id: inria-00171678

https://inria.hal.science/inria-00171678v1

Preprint submitted on 12 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interest Point Detection with Wavelet Maxima Lines

Christophe Damerval and Sylvain Meignen

Laboratoire Jean Kuntzmann, LJK, Grenoble, France

Abstract. In this paper, we propose a new approach in image processing for interest point detection. Thanks to the link between linear scale-space and wavelet decompositions, we are able to build a structure in scale-space based on wavelet maxima lines. This framework allows robust detection of objects (or blob-like structures) present in the image, and the computation of their characteristic scale. Eventually the estimation of their shape can be performed.
Keywords: scale-space representation, maxima lines of wavelet decomposition, blob detection, scale selection.

1 Introduction

Interest point detection is an important low-level operation in computer vision, since it is a first step towards more complex tasks such as the determination of local deformations in images, the extraction of scale-invariant interest points or the computation of local descriptors ([4] [6] [8]). So this paper is organized as follows: first, we present the framework of the linear scale-space put forward by Lindeberg in [5] (section 2) ; we also highlight the link with wavelet decompositions, especially with wavelet maxima lines (ML), and explain how scale selection can be performed on the basis of these ML. Then we detail the different steps of our method (section 3), laying the emphasis on ML construction. Besides, we will see how ML can be used to estimate the shape of the detected objects (section 4). Eventually, we will see some applications on geometric images and on natural scenes and also some results on the influence of noise (section 5).

2 Linear Scale Space and Wavelet Maxima Lines

2.1 Linear Scale Space

We first recall the definition of linear scale-space in the one-dimensional case. Linear scale-space representation consists of convolutions with Gaussian kernels at different scales. If we put $g(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ and $g_t(x) = \frac{1}{\sqrt{t}} g(\frac{x}{\sqrt{t}})$, we define as in [5] the linear-scale space by:

$$L(x,t) = g_t * f(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{t}} g(\frac{x-u}{\sqrt{t}}) f(u) du.$$

Scale selection is then carried out through the study of the appropriately normalized derivatives of L. Lindeberg [4] suggests to normalize the derivatives as follows : $\partial^m_{x,\gamma norm} L(x,t) = t^{\frac{\gamma m}{2}} \partial^m_x L(x,t) = \int_{\mathbb{R}} t^{\frac{\gamma m}{2}} \frac{1}{\sqrt{t}} g\left(\frac{u}{\sqrt{t}}\right) f^{(m)}(x-u)du$ $= t^{\frac{\gamma m}{2}} f^{(m)} * g_t(x)$. In a bidimensional context, the study is very similar to that done in one dimension. We now put $g(x_1,x_2) = \frac{1}{2\pi}\exp(-\frac{x_1^2+x_2^2}{2})$ and $g_t(x_1,x_2) = \frac{1}{t} g(\frac{x_1}{\sqrt{t}}, \frac{x_2}{\sqrt{t}})$. The bidimensional linear scale-space definition is identical to its one dimensional counterpart. Let us define $\partial_{x^\alpha} L = L_{x^\alpha} = \partial_{x_1^{\alpha_1} x_2^{\alpha_2}} L = L_{x_1^{\alpha_1} x_2^{\alpha_2}}$, with $\alpha = (\alpha_1, \alpha_2)$, we consider differentiations of the linear scale-space of the form: $\mathcal{D}L = \sum_{i=1}^{I} c_i L_{x^{\alpha^i}}$ where $|\alpha_i| = \alpha_1^i + \alpha_2^i = M$ is independent on i. For such a definition of the linear scale-space, the appropriate normalization of the operator $\mathcal{D}$ is given by [4]: $\mathcal{D}_{x,\gamma norm} L = t^{\frac{M\gamma}{2}} \mathcal{D}L$. The Hessian matrix of L defined by $\mathcal{H} = \begin{pmatrix} L_{x_1x_1} L_{x_1x_2} \\ L_{x_2x_1} L_{x_2x_2} \end{pmatrix}$ is a basic tool for the analysis of the characteristic scale of objects in images [8][4]. A simple normalized feature is usually extracted from this matrix and is: trace $(\mathcal{H}_{\gamma norm}) = t^\gamma \Delta L$. Now, when $\gamma = 1$, it is interesting to draw a parallel with wavelet decompositions. Using two integrations by part and because of the fast decay of the Gaussian function and of its derivatives, we get: trace$(\mathcal{H}_{\gamma norm}) = f * (\Delta g)_t$ so that as Δg is even, this quantity is exactly the L^1-normalized wavelet decomposition using the Δg wavelet at scale $\sqrt{t}$.

2.2 Wavelet Decompositions and Maxima Lines

The normalized Laplacian used in our framework can be written as the L^1-normalized bidimensional continuous wavelet transform (CWT) of the image f using the wavelet Δg, $\Delta g(x,y) = \frac{1}{2\pi}\left(-2 + (x_1^2 + x_2^2)\right)\exp(-\frac{x_1^2+x_2^2}{2})$:

$$\forall (x,y) \in \mathbb{R}^2, \forall s \in \mathbb{R}_+^*, \mathcal{W}f(x,y,s) = \frac{1}{s^2}\int_{\mathbb{R}^2} f(u,v)\Delta g(\frac{u-x}{s}, \frac{v-y}{s})dudv$$

Since the wavelet Δg can be expressed analytically in the Fourier domain, the computation of the CWT can be done as follows: first, compute $\widehat{\Delta g}$ times the discrete Fourier transform of the image (using a FFT), and then obtain the CWT by the inverse Fourier transform. Note that the cost of this step is $O(NLog_2N)$ (where N is the size of the data), thus providing an efficient computation. Given a certain scale $s \in \mathbb{R}_+^*$, modulus maxima are defined as local maxima in space of the modulus of the CWT of the image $|\mathcal{W}f(.,.,s)|$. Let us recall that the absolute value of the CWT has the property of low response where the image is smooth, whereas it has a high response where there are singularities (boundaries, edges, corners or isolated peaks). So the point of studying modulus maxima is that they are related to singularities present in the image. Note that at the finest scales, modulus maxima arise from isolated singularities, while at coarser scales, modulus maxima result from several singularities since the translated-dilated wavelet $\Delta g(\frac{\cdot - x}{s}, \frac{\cdot - y}{s})$ covers a wider area. In particular, as we will see, a modulus maxima at a certain scale can be related to the presence of a significant structure.

In our context, the maximum principle (see [9] [1]) ensures that modulus maxima propagate towards finer scales, thus making connected curves in scale-space, called maxima lines (ML). The construction of these ML on the basis of the modulus maxima at each scale is detailed in the next section. So a ML can be viewed as a path in scale-space which does not interrupt when the scale decreases. As the scale s increases, (x_l, y_l) gives the spatial drift of the ML while $|\mathcal{W}f((x_l, y_l, s)|$ gives the evolution of the response. Denoting s_{int} the scale at which the ML is interrupted ($s_{int} \in \overline{\mathbb{R}}_+^*$), one maxima line $L_l (l \in \mathbb{N})$ is denoted as:

$$\begin{aligned} L_l = \{&(x_l(s), y_l(s), s, \mathcal{W}f(x, y, s))_{s \in]0, s_{int}[}, \\ &\forall s \in]0, s_{int}[, (x_l(s), y_l(s)) \text{ local maximum of } |\mathcal{W}f(., ., s)|\} \end{aligned}$$

2.3 Scale Selection

As pointed out by Lindeberg in [5], stable features (robust to noise or to slight deformations) can be obtained by considering the extrema of the image decomposition using an appropriate operator. In particular, when normalized derivatives of the linear scale-space are used, such features can be related to the actual position and size of the objects appearing in the image. Mainly, our approach differs from classical scale space selection to the extent that instead of determining local maxima in the 3D scale space (denoted Max_{3D}) we build a structure (the skeleton of the maxima lines) made of chains of modulus maxima, and take a certain maximum along some of them (the resulting set is noted Max_{ML}). So as to see the link between these two sets which are subsets of the scale-space $S_{3D} = \mathbb{R}^2 \times \mathbb{R}_+^*$, let us define:

$$\begin{aligned} \mathcal{V}(x^*, y^*, s^*) &= \{ \text{ cube of } S_{3D} \text{ centered at } (x^*, y^*, s^*) \text{ of size } \eta \times \eta \times \eta (\eta > 0)\} \\ Max_{3D} &= \{(x^*, y^*, s^*) \in S_{3D}, \text{ for a given } V \in \mathcal{V}(x^*, y^*, s^*), \\ &\quad \forall (x, y, s) \in V, |\mathcal{W}f(x^*, y^*, s^*)| \geq |\mathcal{W}f(x, y, s)|\} \\ Max_{ML} &= \{(x^*, y^*, s^*) \in S_{3D}, \text{ there exists a ML } L_l \text{ containing } (x^*, y^*, s^*) \\ &\quad \text{ such that the response attains a local maximum along } L_l : \\ &\quad \exists V \in \mathcal{V}(x^*, y^*, s^*), \forall (x_l(s), y_l(s), s) \in L_l \cap V, \\ &\quad |Wf(x^*, y^*, s^*)| \geq |\mathcal{W}f(x_l(s), y_l(s), s)|\} \end{aligned}$$

Given $(x^*, y^*, s^*) \in Max_{3D}$, we note that it is a modulus maximum at scale s^*, which belongs to a certain ML, denoted L_l (modulus maxima propagate towards finer scales), so we have $(x^*, y^*, s^*) = (x_l(s^*), y_l(s^*), s^*)$. According to the definition of Max_{3D}, we get : $\forall (x_l(s), y_l(s), s) \in L_l \cap V$, $|\mathcal{W}f(x_l(s^*), y_l(s^*), s^*)| = |\mathcal{W}f(x^*, y^*, s^*)| \geq |\mathcal{W}f(x_l(s), y_l(s), s)|$. So a local maximum in 3D scale-space is also a local maximum along a certain ML. Conversely, let us consider $(x^*, y^*, s^*) \in Max_{ML}$: $\exists V_1 \in \mathcal{V}(x^*, y^*, s^*)$, $\forall (x_l(s), y_l(s), s) \in L_l \cap V_1, |\mathcal{W}f(x_l(s^*), y_l(s^*), s^*)| \geq |\mathcal{W}f(x_l(s), y_l(s), s)|$. We can distinguish between two cases: isolated and non-isolated. First, considering a ML L_l that is isolated in scale-space at scale s^*, there exists $V_0 \in \mathcal{V}(x^*, y^*, s^*)$ such that L_l is the unique ML intersecting V_0 ; taking $V = V_0 \cap V_1$, we obtain that $(x^*, y^*, s^*) \in Max_{3D}$.

The case of non-isolated ML corresponds to the junction of some ML $(L_j)_{j\in J}$ which merge at scale s^*, and in general, we cannot say that $(x^*, y^*, s^*) \in Max_{3D}$. Indeed, let us consider the example of two ML L_1 and L_2 merging at (x^*, y^*, s^*) and assume that the response (i.e., the modulus of the wavelet coefficient) attains a local maximum at s^* while it decreases along L_2: $(x^*, y^*, s^*) \notin Max_{3D}$. In practice, we observe that when the response along L_2 decreases, the latter corresponds either to noise fluctuations – and L_1 is not significant (lower response than noise) – or to a ML which has attained a local maximum at a previous scale – so L_1 does not correspond to a significant structure compared to that associated to L_2.

Eventually, given the ML $(L_j)_{j\in J}$ involved in the junction, and assuming the response attains a local maximum along each of these ML, we can consider the dominating ML, denoted by L_k, defined as the one which gives the maximum response at scales preceding s^*. Also note that the response along L_k is also larger at scales following s^* since these ML merge. Thus there exists $V \in \mathcal{V}(x^*, y^*, s^*)$ such that the response along $L_k \cap V$ is larger than along $(L_j \cap V)_{j\neq k}$ at each considered scale, and therefore higher than any modulus of wavelet coefficient in V. Since (x^*, y^*, s^*) is a local maximum along L_k, it belongs to Max_{3D}. In the following sections, we make the assumption that if a ML attains a maximum at (x^*, y^*, s^*) and merge with some other ML also attaining a maximum at (x^*, y^*, s^*) (there is a junction of ML at this scale-space location), it corresponds to a local maximum at s^* along each of the ML involved in that junction. With this hypothesis, the elements of Max_{ML} of interest are all in Max_{3D} ; a theoretical study should be carried out to settle to what extent this assumption is valid.

3 Local Features Extraction from Wavelet Maxima Lines

3.1 Modulus Maxima at Each Scale

As shown before, the linear scale space representation of a bidimensional signal f can be viewed as a continuous wavelet transform of this signal, using the wavelet Δg and the L^1 normalization. In practice, we use integer scales $s = 1, 2, ..., s_{max}$ (the choice of s_{max} depends on the size of the sought objects). Once the CWT has been computed for all desired scales, we consider its modulus maxima : for each scale s ($1 \leq s \leq s_{max}$), we compute the local maxima of $|Wf(.,.,s)|$ (using a 3x3 neighborhood) and we keep the wavelet coefficient $Wf(x, y, s)$ as a relevant information: its modulus quantifies the importance of the response and its sign indicates the type of singularity. So a modulus maxima is characterized by (x, y, s, v), where (x, y) is its location, s the associated scale and v the value of $Wf(x, y, s)$.

3.2 Construction of Wavelet Maxima Lines

In this section, we present a method for the effective construction of ML. In the present context, the maximum principle ensures there is no interruption of the

ML when the scale decreases ([9]), while some ML may be interrupted when the scale increases. When dealing with discrete data and increasing scales, a technical difficulty may arise when some modulus maxima drift spatially by a large amount between two integer scales ; for instance, this is the case when two singularities with opposite signs interfere. In such a case, as the ML still exists at this scale, the corresponding modulus maxima should be chained. An option may consist of refining the scale discretization so that the spatial drift between two successive scales is at most one pixel ; but this would entail an extremely fine discretization in scale, and therefore a prohibitive algorithmic cost. That is why we keep integer scales and propose an alternative rule.

Once modulus maxima are known at each scale, we chain them from scale s to scale $s+1$ (starting from the finest scale $s=1$). At this step, one must give rules for connecting a modulus maxima given by (x, y, s, v) to another at scale $s+1$. We first consider the 3x3 neighborhood of (x, y) at scale $s+1$; if there is one modulus maxima of the same sign in this neighborhood, we make the connection (in case of several modulus maxima, we take the one with the larger modulus). Otherwise, if the scale is sufficiently large ($s > 3$), we allow to connect this modulus maxima to the closest modulus maxima of the same sign, provided there is no modulus maxima of opposite sign between them ; for example, if (x, y, s) and $(x', y,', s+1)$ are positive modulus maxima there should not be any negative modulus maxima at scale $s+1$ in the rectangle defined by the corners (x, y) and (x', y'). These rules ensure that a ML will not cross another ML with opposite sign. So we obtain a set of ML which are chains of modulus maxima $\{(x(s), y(s), s, Wf(x, y, s)), 1 \leq s \leq \min(s_{int}, s_{max})\}$.

3.3 Selection of Relevant Maxima Lines for Interest Point Detection

Among all ML $(L_k)_{1 \leq k \leq N_{ML}}$ (N_{ML} is the number of modulus maxima at the finest scale), only few will be relevant. First, it is proved in [7] that for a white noise, the number of modulus maxima decreases by a factor of 2 when the scale increases by 2 on average. So in order to obtain a set of maxima lines of interest, a basic condition is that these ML should persist until a sufficiently large scale ($s = 5$). We also eliminate the ML on which the modulus of wavelet coefficient is strictly increasing with scale since this will not correspond to a structure with characteristic scale smaller than s_{max}. Besides, as pointed out by Mallat in [7], along the ML due to noise fluctuations, the modulus of wavelet coefficients decreases on average with scale (see Fig. 1 (B')) so these are discarded. In the framework proposed in [1], a rule based on the spatial stability of the ML has been proposed for scale selection. Note that this criterion is valid in the case of an isolated blob (no interference with other singularities). However, in a natural scene, this criterion may lead to evidence only few structures, whereas applications in computer vision require a large number of features so as to lead to satisfactory results. So there is a need for an alternative criterion to select the scales of interest.

3.4 Characteristic Scale Computation – Junction of Maxima Lines

Given a relevant ML, it is possible to follow the evolution of the response along this ML with respect to the scale $s \mapsto |Wf(x(s), y(s), s)|$. We compute $s^* > s_{rel}$ as the scale corresponding to the first local maximum of the modulus of wavelet coefficients along this ML (in practice, the choice of $s_{rel} = 4$ avoids spurious maxima due to noise). So, we obtain a potential interest point (x^*, y^*, s^*, v^*) for each relevant ML, which is the first local maximum that is beyond the noise level. Since we use an appropriate normalization, s^* can be viewed as a characteristic scale of the object, while (x^*, y^*) corresponds to its localization. As we will see in the following section, this can be refined in a more precise estimation of the object shape. In order to group some ML corresponding to the same object, the criterion we propose here is based on the junction of ML : we group the ML for which the values (x^*, y^*, s^*) computed previously are identical (in practice, so as to avoid numerical approximation errors, we group ML whose values of (x^*, y^*, s^*) are very close). So our method selects some points in scale-space (x^*, y^*, s^*) where there is a junction of ML, and for which some ML involved in that junction give a maximum response ; in practice, we note that this is valid for all ML involved in the junction (nevertheless, this point should by studied theoretically).

Now, let us see some examples of ML which simultaneously merge in scale-space and for which the response attains a maximum along them. First, let us take a one-dimensional example : we perform a 1D-CWT using the Sombrero wavelet on the signal represented on the Fig. 1 (A) ; the skeleton of the ML is shown on the Fig. 1 (A'). Note that the junction of ML evidences the main structures in the signal and that the left structure gives rise to ML which merge at a certain scale s but which are not spatially stable between scales s and $s+2$. Furthermore, if we denote L_1 the ML drifting from $t = 448$ to $t = 512$ and L_2 the ML drifting from $t = 576$ to $t = 512$, we note that L_1 and L_2 merge at a certain scale s_j (on the basis of the skeleton of the ML). Then we retain the first maximum (beyond the noise level) of the response along $L1$ and L_2, attained at the scale s^*, and observe that $s^* = s_j$ (see Fig. 1 (A")). So these two ML of opposite signs have similar behaviors: spatially, each drifts one towards the other, merge at s^*, and both give a maximum response at the scale of junction s^*. In this regard, note that at the finest scales, each modulus maxima is associated with a step singularity whereas at the scale s_j where ML join, the associated modulus maxima corresponds to a crenel singularity (made of two steps), thus leading to a higher response.

Secondly, in a two-dimensional context, let us take a disk to which a Gaussian white noise has been added ($SNR = -1dB$). Our method allows to follow the spatial drift of the ML from the boundaries to the center of the object (see Fig. 1 (B)). Along a ML corresponding to noise fluctuations (see Fig. 1 (B')), the response decreases rapidly as the scale increases. Along a ML associated to a significant structure (see Fig. 1 (B"), considering scales sufficiently large (beyond the noise level), we note that the first maximum corresponds to the junction of all the ML associated to the object. In this case, the object gives

rise to different ML which merge progressively as the scale increases. Note also that when some ML merge at a scale s_1, it may not correspond to a maximum of the response (the response may also increase after s_1) ; in particular, the first junction of ML tend to increase the response when it involves some ML whose wavelet coefficients are of the same sign (at the scale of junction).

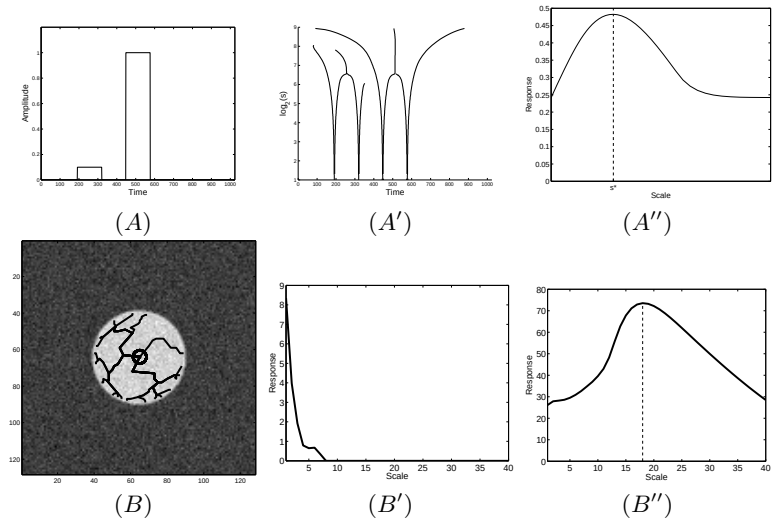


Fig. 1. (A): 1d example of two blobs ; (A'): skeleton of ML ; (A''): response along a relevant ML ; (B): noisy disk and spatial drift (the junction is at $(x^*, y^*, s^*) = (64, 64, 18)$); (B'): evolution of the response along a ML due to noise (focusing on finer scales); (B''): evolution of the response along a ML of interest.

3.5 Global Complexity

If we note N the size of the data and assuming $s_{max} << N$, the computation of the CWT takes $O(Nlog_2N)$ operations, thanks to the fast Fourier Transform, and the computation of the local modulus maxima is done with $O(N)$ operations. If we note N_1 the number of modulus maxima at the finest scale $s = 1$ ($N_1 < N$) the construction of the maxima lines is achieved with $O(N_1)$ operations (this step is in fact time-consuming due to the complexity of the construction of each ML). The following steps (selection of the ML, computation of the characteristic scale, shape estimation) require $O(N_1)$ operations. Finally, the global computation cost is $O(Nlog_2N)$.

4 Interest Point Determination and Shape Estimation

The proposed method based on maxima lines evidences a set of different objects present in the image, more or less complex, called blob-like structures. Thanks to the junction criterion, a blob-like structure can be represented by its associated maxima lines $(L_i)_{i\in I}$. For each L_i, we can follow the spatial drift $(x_i(s), y_i(s))$ and the evolution of the modulus of the wavelet coefficient $|Wf(x_i(s), y_i(s), s)|$ along L_i with respect to the scale s. According to the junction criterion, we get $\forall i \in I, (x_i(s_i^*), y_i(s_i^*), s_i^*) = (x^*, y^*, s^*)$. The drift of the ML $(L_i)_{i\in I}$ from $s = 1$ to $s = s^*$ gives some geometric information on the structure, which allows an estimation of its shape on the basis of the different origins of the ML $(x_i^0, y_i^0)_{i\in I}$ and their common peak (x^*, y^*, s^*).

Basically, we can see (x^*, y^*, s^*) as an interest point, the associated region of interest being a circle centered at (x^*, y^*) and whose radius equals s^*. Alternatively, a geometrically based rule can be applied: for each ML L_i, we compute the distance between the peak localization (x^*, y^*) (i.e., the center of the structure) and the origin of the ML (x_i^0, y_i^0) (i.e., the border of the structure), defined as $\mathcal{S}_i^* = \sqrt{(x^* - x_i^0)^2 + (y^* - y_i^0)^2}$; then we define the interest point as $(X^*, Y^*, S^*) = (x^*, y^*, \text{median}\{S_i^*\})$. This new characteristic scale gives a robust estimation of the size of the structure (see [1]). Let us note that the convex hull of the points $\{(x_i^0, y_i^0)_i \in I\}$ is fully determined by $\{(x_i^0, y_i^0)_i \in K\}$, where K is a subset of I, and that the polygon defined by this convex hull gives an approximate shape of the structure. A smoother shape can also be obtained by considering $(x_i^0, y_i^0)_{i\in K}$ as the knots of a cubic spline curves ; nevertheless, one should be careful, since very closed points may lead to distorted or inappropriate shapes. Besides, note that this approach assumes that the structure is associated to a sufficiently large number of ML ; otherwise, in the case of very few ML, the shape estimation could be irrelevant.

A particular case is when we are looking for some ellipse-shaped regions, which is linked to the detection of affine-invariant interest points. Lindeberg put forward a method for the estimation of local deformations, based on a fixed-point problem, which involves the second moment matrix defined in [4]. We propose here to estimate directly this shape. The estimation of this ellipse can be written as an optimization problem. Let $\alpha = (\alpha_i)_{1\leq p\leq 6} \in \mathbb{R}^6$, we define $F_\alpha : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by $F_\alpha(x, y) = \alpha_1 x^2 + \alpha_2 y^2 + \alpha_3 xy + \alpha_4 x + \alpha_5 y + \alpha_6$, and then solve the least-square problem : $\min_\alpha \sum_{k\in K} F_\alpha(x_k^0, y_k^0)^2$ (note that only the points that define the convex hull are taken into account). Then, given an estimation $\hat{\alpha}$ of α the equation of the sought ellipse is then $F_{\hat{\alpha}}(x) = 0$. A simplification can be obtained by setting the center of the ellipse, for instance as $(x_c, y_c) = (x^*, y^*)$. The equation of this ellipse in the plane can be written as $Z^T AZ = 1$, where Z is the vector of centered coordinates and A a semi-definite positive matrix. We then compute the matrix A and thus obtain the ellipse centered at (x_c, y_c) which fits at best (in the least square sense) the points $(x_k^0, y_k^0)_{k\in K}$. This choice is motivated by the reduction of the number of parameters and the fact that (x^*, y^*) should be close to the center of the ellipse (see results in the following section).

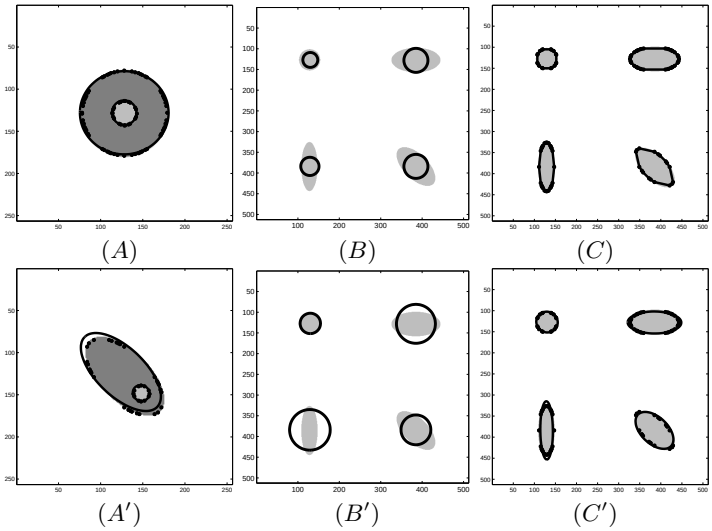


Fig. 2. Basic examples of shape estimation. (A) and (A'): two nested blobs, and shape estimation by the ellipse centered at (x^*, y^*) ; (B): interest points based on the peak (x^*, y^*, s^*); (B'): interest points (x^*, y^*, S^*) (S^* obtained by the rule of the median); (C): convex hull of the origins of the ML (x_i^0, y_i^0) ; (C'): shape estimation by ellipses based on $(x_k^0, y_k^0)_{k \in K}$ (K: convex hull).

5 Results

For a given image, we carry out an interest point detection by following the different steps of our method. Once all ML are constructed, we select them on the different criteria explained previously ; in this regard, note that most of the rejected ML correspond to noise (see Tab. 1). Given the set of remaining ML, the objects present in the image are then evidenced on the basis of the junction criterion. Eventually, a slight thresholding on the modulus wavelet coefficient should be performed, since low-response singularities may lead to non-significant structures. Note that in the presence of noise, singularities at the finest scale disappear as the noise level increases. So the construction of ML leads to a much larger number of ML : most of them are due to noise, and only some of them correspond to the sought objects. Besides, for a given object, there are less relevant ML in the noisy case than in the original one (some of these are destructed by the noise). Globally, in presence of noise, the selection step leads to retain less than 1% of all ML on a geometric image, and about 20% on a natural scene example and so our approach still applies in a noisy context. Moreover, we observe that the junction criterion presented here leads to a satisfying number of ML for each detected object.

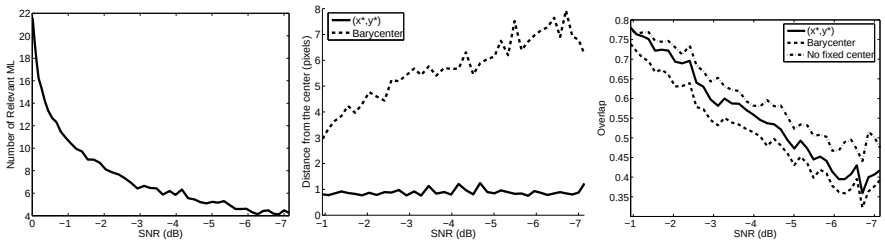
Now, let us consider geometric images containing some objects known *a priori*. First, considering two nested blobs (see Fig. 2(A)), we observe that $c^* = (x^*, y^*)$ corresponds exactly to the actual center c. In the case of a main

Table 1. Numerical examples of our method applied to a geometric image (a disk centered at $(256, 256)$) and a natural scene (classical image *Barbara*). In the noisy case, a Gaussian white noise ($SNR = -1dB$) was added to the original image.

Image 512x512	Disk	Noisy Disk	Barbara	Noisy Barbara
ML constructed	166	9769	4907	6592
ML due to noise	0	9694	3150	5253
Relevant ML	166	75	1736	1324
Objects	1	1	136	129

blob (denoted by 1) containing another blob (denoted by 2), we observe that if $c_1 \neq c_2$, the location of c_1^* does not correspond to its actual center c_1 because of the influence of the smaller blob 2 (see Fig. 2(A')). Nevertheless, we note they remain close ($||c_1^* - c_1|| = 6.4$ pixels for an image of size 256x256). Secondly, we apply our method on an image containing different ellipses. We observe that the characteristic scale $\mathcal{S}^*$ (based on an Euclidean distance – see section 4) approximately equals $2s^*$ (see Fig. 2 (B) and (B')) and the convex hull and the fitted ellipse computed on the basis of the origins of the ML corresponding to an object are very similar (see Fig. 2 (C) and (C')).

Thirdly, we consider an image of size 128x128 representing a disk with added Gaussian white noise, and we apply our method to the image for different noise levels. (for each noise level, we performed 100 simulations of images of size 128x128). On the basis of the ML evidenced by our method, we are able to compute a potential center c_p by choosing either the coordinates of the interest point $c^* = (x^*, y^*)$, either the barycenter of the origins of the ML $c_b = (x_b, y_b) = \sum_{k \in K} w_k (x_k^0, y_k^0)$ (where K is the convex hull and $w_k = \frac{1}{|K|}$) or alternatively the center c_e given by the direct estimation of the ellipse. Then it is possible to compute the Euclidean distance between the actual center c of the object and the potential center c_p ($c_p = c^*, c_b$ or c_e) for each noise level. As the noise level increases, c^* appears as more stable than the barycenter c_b (see Fig. 3 (B)). This is mainly due to the fact that the noise destroys too many relevant ML for a correct estimation of the barycenter (see Fig. 3 (A)). Furthermore, denoting A the area covered by the known object, and A_e the area of a fitted ellipse, we are able to compute an overlap value, defined as $\frac{A_e}{A} \in [0, 1]$. Results are shown on Fig. 3 (C), for the different methods of ellipse fitting explained before. We observe that the best results are obtained when the center of the ellipse is not fixed ; this shape estimation method being efficient provided the points $(x_k^0, y_k^0)_{k \in K}$ allow a correct estimation of all the parameters of this ellipse (which is the case for a geometric image). Nevertheless, note that the objects detected in a natural scene image may be associated to few ML. So a more efficient approach could be obtained by combining both (x^*, y^*) – stable feature – and $(x_k^0, y_k^0)_{k \in K}$ – which gives a correct estimation of the ellipse in our geometric image example.



(A) (B) (C)

Fig. 3. Influence of noise on object shape estimation ; as the level of noise increases, we represent (A) : Evolution of the number of ML (B) : Distance to the actual center, (C) : Overlap between the object and the fitted ellipse using different approaches.

On natural scenes (see Fig. 4), we observe that some detected regions are consistent with the visual content of the image. However, this is not the case for all regions ; in particular, we note that in some cases, the origins of the ML are almost collinear: this corresponds to some boundaries (edges, corners) which give rise to some ML that join at a certain scale-space location (x^*, y^*, s^*). So (x^*, y^*) is a stable feature, but it may be difficult to interpret it as the center of an object, whereas the origins of the ML $(x_i^0, y_i^0)_{i \in I}$ can be related to some boundaries. Therefore future work should put forward some rules so as to set up which geometry should be associated to these detected points.

6 Conclusion

In this paper we have presented a new method, which exploits the particular structure of the linear scale-space. Starting from the singularities at the finest scale, we build maxima lines in scale space. Their analysis leads to evidence different objects present in the image. We then obtain an interest point detector by taking a certain maximum along these maxima lines. Eventually, an object shape estimation can be performed, in addition to the computation of the localization in scale-space.

Future work will be carried out in different directions. First, a theoretical study should establish to what extent a junction of ML corresponds to a local maxima of the response along these ML. Then, we will deal with affine transformations, and lay the emphasis on repeatability tests (see [8]). Eventually we will see how other operators than the Laplacian may be used, in particular the rules that should be applied to construct empirically some extremum paths, and the conditions that ensure their continuity from a theoretic point of view.

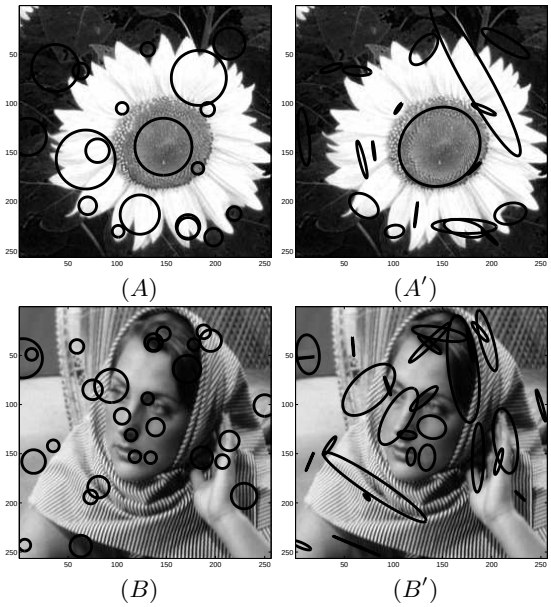


Fig. 4. Natural Scenes Examples (image of sunflower and classical image *Barbara*); (A) and (B): Interest point detection based on (x^*, y^*, s^*) ; (A) and (A'): Shape estimation by performed on ellipses centered at (x^*, y^*) and estimated by $(x_k^0, y_k^0)_{k \in K}$.

References

1. C. Damerval, S. Meignen, *Blob Detection with Wavelet Maxima Lines*, IEEE Signal Processing Letters (2006)
2. C. Harris, M. Stephens, *A Combined Corner and Edge Detector*, in Alvey Vision Conference, pp. 147-151, 1988.
3. F. John, *Partial Differential Equations*, Springer-Verlag, New-York, 1975.
4. T. Lindeberg, J. Garding, *Shape-Adapted Smoothing in Estimation of 3-D Shape Cues from Affine Deformations of Local 2-D Brightness Structure*, Image and Vision Computing, vol. 15, pp. 415-434, 1997.
5. T. Lindeberg, *Feature Detection with Automatic Scale Selection*, International Journal of Computer Vision, vol. 30, pp. 79-116, 1998.
6. D. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision (2004).
7. S. Mallat, W.L. Hwang, *Singularity Detection and Processing with Wavelets* IEEE Transactions on Information Theory, vol. 38, no. 2, 1992.
8. K. Mikolajczyk, C. Schmid, *Scale and Affine Invariant Interest Point Detectors*, International Journal of Computer Vision, vol. 60, pp. 63-86, 2004.
9. A. L. Yuille, T. A. Poggio, *Scaling Theorems for zero-crossing*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, pp. 15-25, 1986.