Invariant image descriptors and affine morphological scale-space

Frédéric Sur

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Invariant image descriptors and affine morphological scale-space

Frédéric Sur

N° 6250

Juillet 2007

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--6250--FR+ENG

Invariant image descriptors and affine morphological scale-space

Frédéric Sur

Thème COG — Systèmes cognitifs
Projet Magrit

Rapport de recherche n° 6250 — Juillet 2007 — 17 pages

Abstract: In this research report we propose a novel approach to build interest points and descriptors which are invariant to a subclass of affine transformations. Scale and rotation invariant interest points are usually obtained via the linear scale-space representation, and it has been shown that affine invariance can be achieved by warping the smoothing kernel to match the local image structure. Our approach is instead based on the so-called Affine Morphological Scale-Space, a non-linear filtering which has been proved to be the natural equivalent of the classic linear scale-space when affine invariance is required. Simple local image descriptors are then derived from the extracted interest points. We demonstrate the proposed approach by robust matching experiments.

Key-words: Interest points, local descriptors, scale-space representation, affine morphological scale-space.

Unité de recherche INRIA Lorraine
LORIA, Technopôle de Nancy-Brabois, Campus scientifique,
615, rue du Jardin Botanique, BP 101, 54602 Villers-Lès-Nancy (France)
Téléphone : +33 3 83 59 30 00 — Télécopie : +33 3 83 27 83 19

Descripteurs d'images invariants et scale-space affine morphologique

Résumé : Dans ce rapport de recherche nous proposons une nouvelle approche pour construire des points d'intérêt et les descripteurs associés invariants par une sous-classe de l'ensemble des transformations affines. Les points d'intérêt invariants par changement d'échelle et rotation sont habituellement obtenus par représentation multi-échelle linéaire. De plus, il a été montré que l'invariance affine peut être obtenue en adaptant le noyau de lissage à la structure locale de l'image. Notre approche est, elle, basée sur le scale-space affine morphologique, un filtre non linéaire qui est l'équivalent naturel du scale-space linéaire quand une invariance affine est recherchée. Des descripteurs locaux simples sont ensuite extraits des images autour des points d'intérêt. Nous démontrons la validité de l'approche proposée par des expériences d'appariement robuste.

Mots-clés : Points d'intérêt, descripteurs locaux, représentation multi-échelle, scale-space affine morphologique.

1 Introduction and related work

Interest point detection is one of the very first step of many computer vision problems, such as structure from motion, stereovision, image recognition and comparison, etc. These applications often aim at identifying pairs of interest points between two (or several) views, which correspond to the same real point. Descriptors of interest point vicinities are compared for this purpose. Ideally, these descriptors should capture the invariant information that permits to match the corresponding interest points, whatever the viewpoint or illumination change. For example, if one assumes that interest points lie on locally planar structures, descriptors as well as interest point extraction should be invariant to homographies. However, the larger the class of invariance, the more difficult it is to obtain reliable and reproducible descriptors and interest points with respect to this class. That is why smaller classes are considered, such as scale-rotation or affine invariance. Affine invariance is highly desirable since it corresponds to a first order approximation of the actual planar transformation.

Similarity invariant (i.e. scale and rotation invariant) interest points and descriptors have been the subject of many works and can be quite reliably used since the major breakthrough represented by Lowe's SIFT descriptor [9] and its spin-offs (such as SURF [5] or GLOH [12].) However, a larger class of invariance should enable a higher accuracy in descriptor comparison when viewpoint changes are considered. Affine invariance has been the subject of several recent works which have been thoroughly surveyed and compared in [12] and [13]: see for example [10] and [18] which deal with affine covariant regions, or Mikolajczyk and Schmid's affine-adapted Harris detector [11]. As in this last article, we aim at detecting covariant features (namely characteristic angle and scale) along with interest points. Our work focuses on an intermediate class of transformations between similarity and affine transformations. The main difference with [11] is that we consider a truly affine invariant scale-space instead of adapting the linear scale-space, as explained below.

Most of the time, interest points are detected through linear scale-space of images. Although it is a fully sound approach when scale-rotation invariance is required (as explained in [8]), this calls for an adaptation in the affine invariant case. We give in Section 2 a digest of the scale-space theory and introduce the affine morphological scale space (AMSS) which has been shown to be the only affine invariant scale-space. Our main contribution is described in Section 3: we show how to 1) accurately extract interest points through the AMSS and 2) associate with each of these points an angle and a characteristic scale. This leads to invariance to a subclass of affine mappings. We then derive a local descriptor for each of the interest points and demonstrate the soundness of the proposed algorithm with robust matching experiments (Section 4.) To the best of our knowledge, AMSS has never been used to derive descriptors. We conclude with Section 5 and give some perspectives and possible improvements.

2 The scale-space paradigm

Multiscale (or scale-space) analysis is one of the cornerstones of image processing. The basic idea is to associate with an image u_0 (considered as a function from $\mathbb{R}^2$ to $\mathbb{R}$) a family of coarser and smoother representations $(T_t(u_0))_{t>0}$. The parameter t is called *scale* of the analysis. It can be shown that under some natural conditions (mainly linearity of T_t, shift and rotation invariance, and conditions preventing from the creation of new "structures" along the scales) the Gaussian kernels are the unique scale-space filters (see for example [4].) The multiscale representation is thus given by the convolution of u_0 with a Gaussian kernel G_t with variance t

$$T_t(u_0) = G_t \star u_0 \tag{1}$$

In an equivalent manner, $T_t(u_0)(x, y)$ is the solution of the isotropic diffusion equation

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{2}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) \\ u(0, x, y) = u_0(x, y) \end{cases} \tag{2}$$

with u from $\mathbb{R} \times \mathbb{R}^2$ to $\mathbb{R}$.

Note that T_t is invariant to rotation since $T_t(Ru_0) = RT_t(u_0)$ for any rotation R [1]. Besides, the following equation relates a spatial scale change and a scale of analysis change

$$ST_t(u_0) = T_{st}(Su_0) \tag{3}$$

where S is a similarity with zoom factor s. This property yields the normalized derivatives introduced by Lindeberg [8] which allows him to set a local characteristic scale as the extremum of some operator.

It is no more possible to consider a linear scale-space when invariance to affine transformations is desired. Alvarez, Guichard, Lions and Morel [2] and independently in an equivalent formulation Sapiro and Tannenbaum [16] have discovered that imposing affine invariance yields the following partial differential equation (PDE):

$$\begin{cases} \frac{\partial u}{\partial t} = t^{1/3}\left(\frac{\partial^2 u}{\partial x^2}\left(\frac{\partial u}{\partial y}\right)^2 - 2\frac{\partial^2 u}{\partial x \partial y}\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + \frac{\partial^2 u}{\partial y^2}\left(\frac{\partial u}{\partial x}\right)^2\right)^{1/3} \\ u(0, x, y) = u_0(x, y) \end{cases} \tag{4}$$

Let us remark that this PDE is sometimes stated without the "$t^{1/3}$" term. Both PDEs are actually equivalent apart from the rescaling $t \to \frac{3}{4}t^{4/3}$.

Provided u_0 is smooth enough, this PDE admits a unique solution $T_t(u_0)(x, y)$ which is called *affine morphological scale-space* (AMSS.) "Affine" since this solution is invariant to any (special) affine transformation (it satisfies $T_t(Au_0) = AT_t(u_0)$ for all linear mapping A from $\mathbb{R}^2$ to $\mathbb{R}^2$ such that $\det(A) = 1$), and "morphological" since it is invariant to any contrast

[1] Here Ru denotes $u \circ R$ where u is an image and R a planar rotation.

change (that is $T_t(\phi(u_0)) = \phi(T_t(u_0))$ for any non-decreasing real function ϕ) which meets the basic assumption of morphological mathematics "only isophotes matter". Similarly to the linear scale-space (cf equation 3), considering affine transformations that do not preserve area (i.e. with $\det(A) \neq 1$) involves scale change. This is reflected by the following equality

$$AT_t(u_0) = T_{\det(A)^{1/2}t}(Au_0). \tag{5}$$

In this report, we propose to extract invariant keypoints and descriptors by considering image evolution under the AMSS equation. Let us remark that the AMSS is not related at all with what Mikolajczyk and Schmid [11] call *affine Gaussian scale space*, which consists in iteratively warping a Gaussian kernel to match the local image structure around interest points. Strictly speaking, this latter approach is not a scale-space in the sense of the commonly acknowledged axiomatic definition.

3 Interest point detection and invariance

3.1 Interest point localization and characteristic angle evaluation

The proposed interest point detection is based on a result by Alvarez and Morales [3] about the evolution of a corner across the scale of the AMSS. They prove that a corner joining two edges evolves on the bisector line separating these edges in the initial image, and that the evolution speed is constant and only depends on the angle made by the edges. Let us be more specific and consider an image u_0 of a "perfect" corner, that is to say $u_0(x, y) = R(H(\tan(\alpha)x - y)H(y))$ (where R is any rotation, H is the Heaviside step function and α is the angle between the edges of the corner.) Let $T_t(u_0)$ denote the solution of AMSS with initial condition u_0. Then it can be shown by computing the analytic expression of $T_t(u_0)$ (theorem 2 in [3]) that the extremum of curvature of $T_t(u_0)$ (which marks the position of the corner at analysis scale t) lie on the bisector line between the edge of the corners, and that the distance to the initial position is λt, where $\lambda = \tan(\alpha/2)^{-1/2}$.

Alvarez and Morales define an affine invariant corner by tracking these curvature extrema along AMSS evolution. Interpolation allows them to estimate both corner orientation and angle α (through speed λ.) The main drawback of this approach is that curvature computation is known to be numerically unstable and sensitive to noise. As a consequence, curvature extrema are difficult to track in a reliable and accurate manner. However, results can be obtained in the camera calibration context, where corners are detected on a well contrasted black and white calibration target [1].

In their adapted affine Gaussian scale-space, Mikolajczyk and Schmid [11] propose to track points extracted from Harris detector [7]. Although Harris cornerness measure is not affine invariant and neither are its local maxima, we only need here a corner detector that is repeatable enough across the evolution scale so that tracking is possible over a long enough range. For that reason, we also propose to track corners from Harris detector instead of curvature extrema. Actually, Harris corners are experimentally proved much easier to track.

We are therefore led to track corners across several evolution times $t_0, t_1 \dots t_n$ in order to estimate corner location as well as an associated angle. We choose to start interest point extraction at scale t_0 slightly greater than 0 (that is on a smoothed representation of the original image u_0) rather than $t_0 = 0$ (i.e. rather than extracting directly from u_0) in order to get rid of noise and quantization problems. This makes the tracking easier, since it initially concentrates on a smaller number of interest points that are more reliable.

Similarly to the classical scale space, two close curvature extrema end up merging along the AMSS. The straight line displacement hypothesis is of course in this case no more valid. In practice this problem is not very important because in most cases we stop the AMSS evolution before it happens. However, this assumes that the corners to be detected are spaced enough.

The algorithm is illustrated on Fig. 1 and works as follows.

- Extract Harris interest points $(P_i^0)_{i\in\{1\dots n\}}$ at scale t_0. Let C_i be the chained list $[P_i^0]$ made of a single point.
- For each evolution step t_j,
  - extract interest points $(P_k^j)_{k\in\{1\dots m_j\}}$ in $T_{t_j}(u_0)$ (which is the solution of the AMSS PDE at scale t_j);
  - for each chain C_i ($i \in \{1 \dots n\}$), add the nearest point P_k^j to the last point that has been collected in C_i.
- For each chain C_i, project its points on the first principal direction from the Principal Components Analysis of these points. One therefore has a collection (t_i, d_i) where d_i is the signed distance to the barycenter of C_i. Linear least square fitting gives the position at scale $t = 0$, as well as λ (and then α) such as $d = d_0 + \lambda t$.

Note that the algorithm cannot handle interest points that are not detected at scale t_0 but possibly at scale t_i, $i \geq 1$. We have experimentally checked that this situation is quite unusual.

This algorithm is slightly different from the one in [1] and [3]. In these articles, $t_0 = 0$: initial position is therefore known and there is no need for interpolating it. We found that taking $t_0 > 0$ gives better results and a stronger robustness to noise. Using Harris corners instead of extrema of curvatures also makes useless the robust estimation of both λ and the bisector line as in [3]. Let us also remark that [3] does not provide anything but an angle. The aim of the following section is precisely to associate a characteristic scale with it.

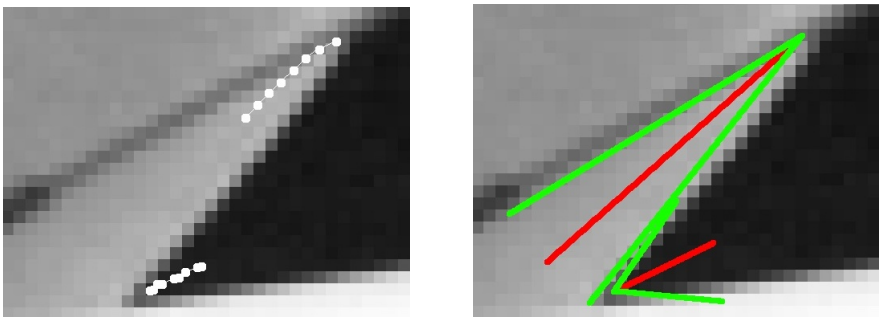

Figure 1: Two examples of corner detection. Left: tracking corner position across AMSS, superimposed over the original image. Corner location moves away at speed λt, where $\lambda = \tan(\alpha/2)^{-1/2}$ and α is the corner angle. Note on this figure that the lower α, the larger the evolution speed λ is. One can see that, due to sub-pixel interpolation, points are not perfectly regularly distributed along the bisector line. Right: the estimated corner (green.) Orientation is measured via the bisector line orientation (red), and angle α via the estimated λ (see algorithm of Section 3.1.) Corner size is set here as in Section 3.3.

3.2 Scale selection

In the linear scale-space setting, Lindeberg [8] has proposed an automatic scale selection technique. It is based on the so-called normalized derivatives which relies on the link between spatial scale and evolution scale (equation 3.) This permits to define characteristic scales as extrema of some operators. Nevertheless, the normalized derivatives concept is no longer valid in the non-linear AMSS framework. It turns out that in this case a similar approach can still be used to estimate a global characteristic scale.

Let u_0 and $\tilde{u}_0$ be two images related by $\tilde{u}_0(x,y) = Au_0(x,y)$, with A a direct affine mapping (i.e. any linear mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ with $\det(A) > 0$.) The problem of interest is to estimate $\det(A)^{1/2}$ (denoted by s in the following) given u_0 and $\tilde{u}_0$.

Let us denote $K(x,y,t) = \frac{\partial}{\partial t} T_t(u_0)(x,y)$ and $\tilde{K}(x,y,t) = \frac{\partial}{\partial t} T_t(\tilde{u}_0)(x,y)$. As stated by equation 5, that is $T_{st}(\tilde{u}_0) = AT_t(u_0)$, one has $s\tilde{K}(x,y,st) = AK(x,y,t)$.

In particular, if K and $\tilde{K}$ are measured at the location of the same corner in $T_{st}(u_0)$ and $T_t(\tilde{u}_0)$ (denoted respectively by K_c and $\tilde{K}_c$), for every t the following equality holds

$$s\tilde{K}_c(st) = K_c(t). \tag{6}$$

A few remarks are needed about the operator K_c and its soundness to detect a characteristic scale. First of all, according to equation 4, $K_c(t)$ can be written as $K_c(t) = t^{1/3}\kappa(t)^{1/3}$, where κ is the curvature of the local level line multiplied by the gradient norm to the power of three. In the linear scale-space setting, operator $t\kappa(t)$ has been used by Lindeberg for

junction detection with automatic scale setting [8]. Furthermore, using the t-derivative of the scale-space representation to detect a characteristic scale is not new in the linear scale-space, since it is the basis of Harris-Laplace detector [11]. In this case the t-derivative is indeed proportional to Δu.

Equation 6 makes it possible to derive a local characteristic scale from the operator K_c, as well as in the scale-adapted Harris detector [8] or Harris-Laplace detector [11]. Contrary to these settings, one can experimentally check that K_c does not attain some extremum over scale. Another choice has to be made. By integrating equation 6, one obtains:

$$\int_0^{st} \tilde{K}_c(u)du = \int_0^t K_c(u)du. \tag{7}$$

We therefore decide to set s such that $\int_0^s |K_c(t)|dt = C$ where C is a positive constant, fixed once and for all. Since the sign of K_c is the sign of the curvature at the corner location (since $K_c = t^{1/3}\text{curv}(u)^{1/3}|\nabla u|$), it remains the same along the scale and is not of interest. Fig. 2 shows an example of scale selection.

One can remark that it would also be possible to fix s by noticing that $sK_c(0) = \tilde{K}_c(0)$ or that equation 7 directly springs from $T_{st}(\tilde{u}_0) = T_t(u_0)$, but the proposed integration is much numerically stable since it relies on several samples. Moreover, K_c does not seem to correspond to a physical measurement that would be related to some geometrical or textural property of the interest point neighborhood. Although it would be interesting to discover such a feature, the aim here is to derive a characteristic scale whose repeatability through reasonable viewpoint change is good enough. This is experimentally checked below.

3.3 Deriving an invariant local descriptor

We build here an invariant local descriptor from each of the detected corner, based on the associated angle α (Section 3.1) and on the characteristic scale s (Section 3.2.) Let us consider such a corner. From what precedes one can associate it with an isosceles triangle which vertices are the interest point and the two points at a distance $D.s$ on the corner edges. We take here $D = 10$. Such a coefficient is rather arbitrary but is unavoidable and also intervenes for example in SIFT descriptors. We then normalize the neighborhood of the corner in $T_s(u_0)$ (where s is the detected characteristic scale) via the linear transformation which maps the isosceles triangle to a given unit right triangle. Interpolation of grey-level values is of course required. Then, we collect the grey values of 17x17 points, regularly distributed over a grid centered on the interest point and directed by the edges of the right triangle. These grey values are 1) weighted by a Gaussian distribution around the interest point to limit edge effects, and 2) normed to achieve invariance to linear contrast changes.

This descriptor is invariant to any affine transformation that maps the unit right triangle to an isosceles triangle[2]. In that sense, the proposed descriptor is invariant to a subclass of

[2]Let us recall that every affine transformation is uniquely defined by the image of three fixed points.

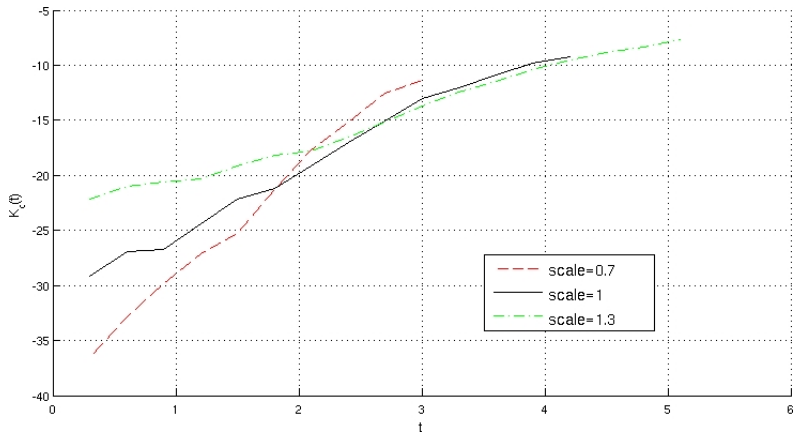

Figure 2: We consider here the same image at three different sizes (zoom factor with respect to the original image: $0.7, 1, 1.3$.) The three curves show the value of $K_c(t)$ against t for the same corner in the three images. The characteristic scale s such that $\int_0^s |K(u)|du = 50$ is 2.574 for image 1, 2.180 for image 0.7, and 3.276 for image 1.3. The ratio between these values is 0.85 (0.7 expected) and 1.27 (1.3 expected.)

the affine transformations, between the similitude transformations (that map the unit right triangle to any isosceles right triangle) and the affine transformations (that map the unit right triangle to any triangle).

In this framework, knowing the characteristic scale actually just allows us to fix the determinant of a general affine mapping. A full affine invariance (i.e. dealing with general triangles instead of isosceles triangles) would require to set two different scales in each direction. This is not the purpose of this report.

Let us remark that matching the proposed descriptor (based on grey values at fixed pixel locations) is more demanding than matching some descriptors based on statistics of the gradient orientation such as in [9] or [12]. However, this is fully satisfactory for the feasibility assessment of Section 4.

3.4 A few remarks about multiple junctions

The proposed framework consists in modeling interest points as perfect corners. This is however a strong bias. Interest points should instead be seen as multiple junctions such as T-junctions [6] that actually correspond to multiple junctions of level sets. However, as pointed out in [3], a multiple junction evolves through the AMSS as a collection of isolated corners (see illustration on Fig. 3.) This makes the proposed interest point detector still valid in the context of non-decreasing contrast change.

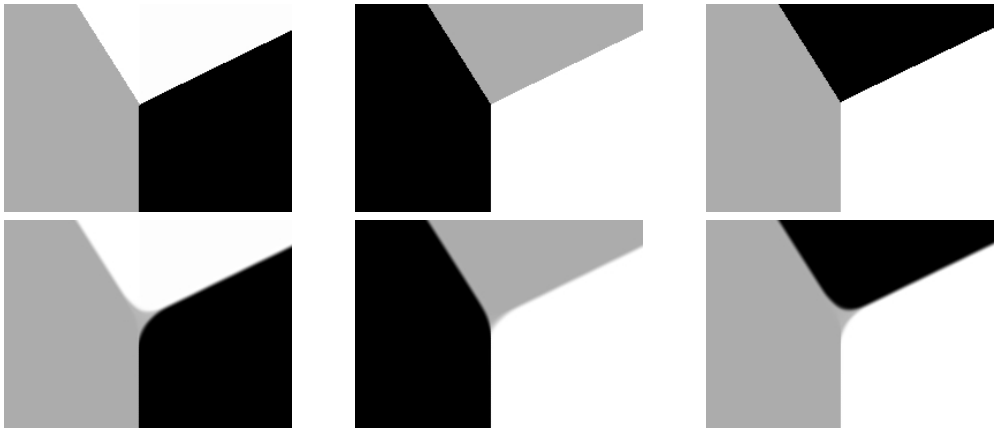

Figure 3: Junction evolution through AMSS. Top: three examples of grey level distribution over a triple point. Bottom: corresponding evolution. As noted in [3], this evolution depends on the ranking of the grey levels. Corners corresponding to level sets evolve in the same way as in Section 3.1. Since we are interested in identifying repeated angles in different views, the approach is still valid provided the possible contrast change is non-decreasing (and thus preserving the local grey level ranking of multiple junctions.)

4 Experimental assessment

In this section, we first discuss some parameters and algorithmic choice. Then we present experiments. Considering two views of the same scene, we match the above descriptors robustly to the underlying image transformation. We consider here planar scenes from different viewpoint; the considered transformation is therefore an homography.

4.1 Parameter setting and algorithmic choices

Subpixel accuracy in the corner position is achieved by quadratic interpolation of Harris cornerness local maxima.

The corner extraction stage (Section 3.1) needs to fix the evolution times $t_0, \dots t_n$ at which corners are extracted. Since corners evolve at speed λt with $\lambda = \tan(\alpha/2)^{-1/2}$, on the one hand this step must be large enough so that estimating λ is still possible for corners with large α, and on the other hand this must be small enough so that small angle corners are still trackable. We set $\lambda = 0.4$ and the step between t_i's to 0.3 so that corners with $\alpha = \frac{3\pi}{4}$ move by 0.2 pixel at each step, and corners with angle of $\alpha = \frac{\pi}{4}$ move by 0.5 pixel.

In order to track corners in different images across almost the same number of evolution steps before their vanishing, the maximal evolution time has to vary linearly as a function

of the size of the image (here $n_r \times n_c$). It is set in our algorithm to $t_n = 4.0 * \sqrt{\frac{n_r * n_c}{640*480}}$ so that in a 640x480 image a corner with $\alpha = \pi/2$ moves by 4 pixels at the final step.

The constant C to fix the local scale (Section 3.2) is set to $C = 50$, which experimentally almost always gives a scale between the first and last steps t_0 and t_n. In order to give an approximation of the scale finer than the discretized scales, K_c is linearly interpolated.

The AMSS and Harris implementations come from the MegaWave free software[3]. Note that this AMSS implementation is not fully consistent with the contrast invariance assumption since it relies on a finite difference scheme. A full contrast invariance would necessitate to decompose the image into its level sets, as explained in [14]. This requires much computation time.

4.2 Robust matching experiment

We present now proof-of-concept matching experiments. If descriptors from two images are matched independently of the underlying transformation, it would give too many retrievals (and therefore barely legible results). We thus perform a robust matching that only keeps inliers with respect to the estimated actual transformation. The robust matching procedure simply consists in 1) determine point-to-point (here corner-to-corner) correspondences between two planar views, and then 2) build a group of inliers through a robust statistical estimation of the underlying homography. Point-to-point correspondences are built by keeping for each corner in the first image the nearest neighbor in the second image (in the sense of the Euclidean distance between descriptors) provided its distance is below a given (large) threshold. We use MSAC [17] in the robust estimation step.

Fig. 4 and 5 shows robust correspondence in two different situations: scale and rotation change (Fig. 4) and viewpoint change (Fig. 5.) Fig. 6 stands as an element of comparison with Mikolajczyk and Schmid's algorithm. The images are taken from the database used in the series of articles by Mikolajczyk and coauthors[4]. For the sake of legibility, retrieved corners (in red) are superimposed over lightened version of the original images. Many matches are retrieved, which proves the potentiality of the approach.

4.3 A quantitative comparison

We finish with a quantitative comparison of the proposed interest point detector and descriptor with two popular and state-of-the-art algorithms, namely Harris-Affine detector with GLOH descriptor, and Hessian-Laplace detector with SIFT descriptor. The first one is an affine-invariant algorithm, while the latter is similitude-invariant (see [11] and [12].) The proposed algorithm is therefore in between, since it is invariant to an intermediate class of transformations. Its repeatability to viewpoint change is tested in what follows. Theoretically, it should perform better than the similitude-invariant algorithm, and not as good as the affine-invariant algorithm.

[3] `http://www.cmla.ens-cachan.fr/Cmla/Megawave/index.html`

[4] Available at: `http://www.robots.ox.ac.uk/~vgg/research/affine/index.html`

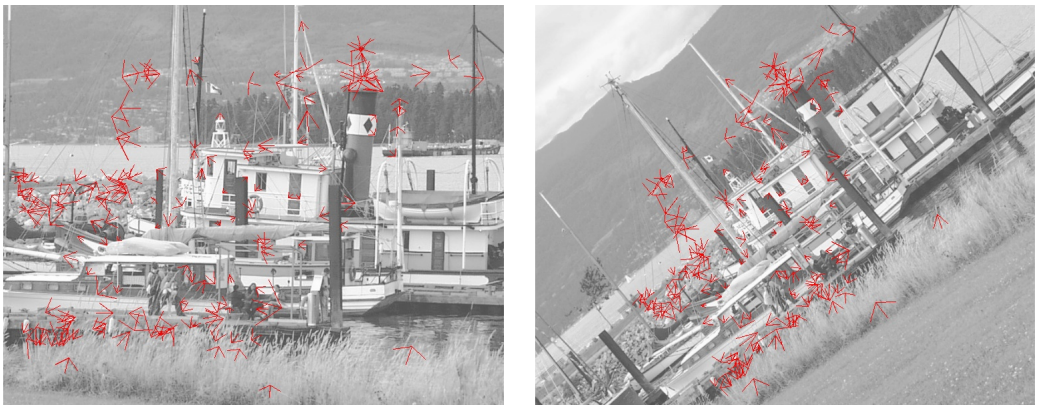

Figure 4: Robust matching experiment, scale and rotation transformation. Here are shown the 177 pairs of inliers to the robustly estimated homography between left and right images. These pairs are retrieved from the 2189 (resp. 1728) corners in the left (resp. right) image. Numerous corners are initially detected because of the large amount of clutter.

The procedure consists in robustly matching five different views of the same planar scene with a reference view (Fig. 7.) Then the proportion of descriptors from the reference view that are retrieved is computed for each of the five views. The robust matching algorithm is the same as in Section 4.2. Thresholds ask for a careful examination in order to get comparable figures. In the MSAC step, the threshold for considering a pair (x, x') of interest points as an inlier to the estimated homography H is set to two pixels, that is to say $d(Hx, x') < 2$. Besides, setting the threshold for building point-to-point correspondences is touchy. If it is too low, then almost nothing is matched, and if it is too high, then too many wrong matches prevent the MSAC step to correctly estimate the underlying transformation. We have therefore perform several experiments with different values of this threshold, and kept the larger proportion of retrievals.

Fig. 8 shows the figures. As expected, the proposed algorithm has a better retrieval rate than the similitude-invariant algorithm, and a slightly lower rate than the affine-invariant algorithm. As a matter of fact, our algorithm gives more accurate localization of the interest points, while the two others are a little less accurate. A more precise estimation of the underlying transformation is therefore enabled, which is required by the two pixels threshold. In compensation, it is less stable in case of a strong viewpoint change (see the last image in the series of Fig. 7, an accurate localization of the interest points is very difficult to obtain with such a small angle between the wall and the camera axis.) While every experimental assessment methodology is questionable, what is important to note here is that the proposed descriptor is clearly competitive with well established ones.

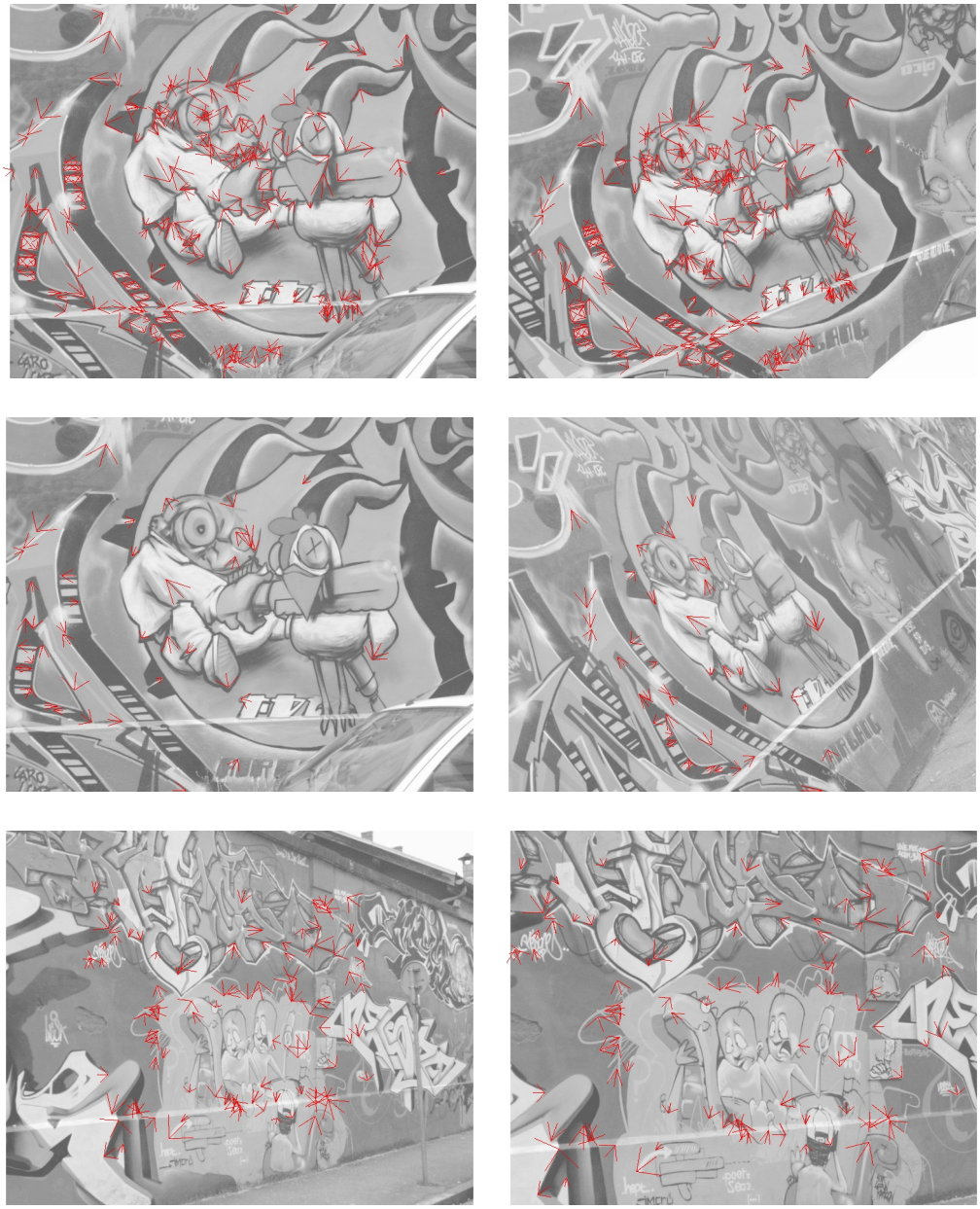

Figure 5: Three robust matching experiments, with different viewpoint and scale changes. Corners from the left image are robustly matched to corners from the right image. Corners matched coherently with the estimated homography are shown in red. Experiment 1: 294 retrieved pairs (601 corners were extracted on the left, 740 on the right; not shown.) Experiment 2: 50 pairs of inliers (left: 601 corners extracted; right: 831.) Experiment 3: 104 pairs of inliers (left: 1240 corners extracted; right: 1154.)

Figure 6: A comparison with Harris-Affine interest point detector, with GLOH descriptor [11] (K. Mikolajczyk's implementation.) The same robust matching algorithm as in Fig. 4 and 5 is used: threshold of step 1) is set so that (almost) no false detection can be seen, leading to 196 inliers to the estimated homography. 4093 descriptors were extracted from the left image, and 4376 from the right one. Although the algorithm is affine-invariant, this gives almost the same proportion of retrievals as in Fig. 5.

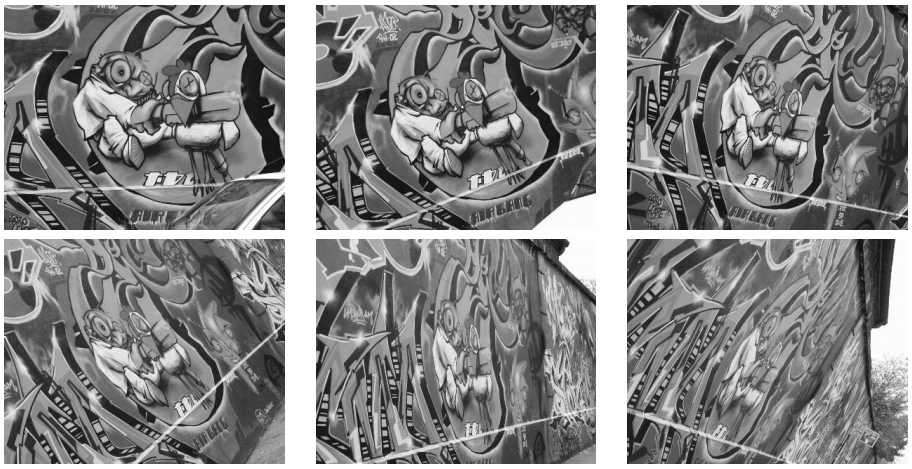

Figure 7: Testing the repeatability to viewpoint change. Each image is a view of the same planar scene, with stronger and stronger viewpoint change. Descriptors are extracted, then the reference image (far left) is robustly matched to the five other images.

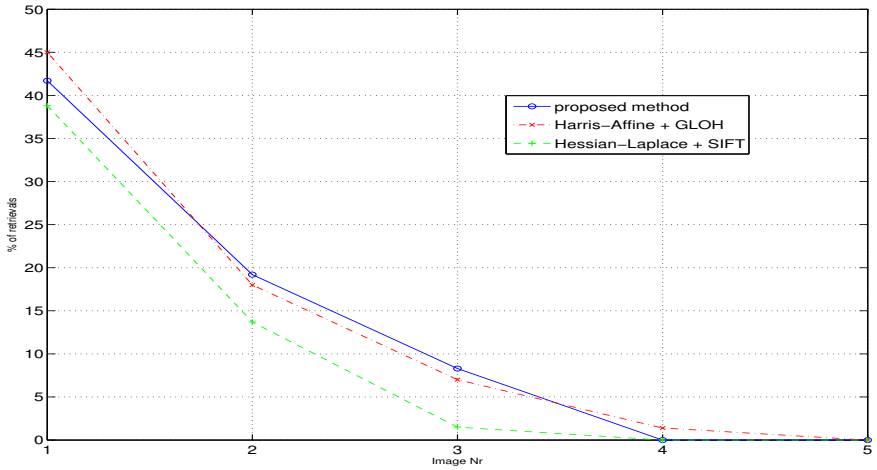

Figure 8: Rate of retrieved descriptors from the reference image that are robustly matched to descriptors from the five other images. The three competing methods are Harris-Affine / GLOH descriptor (affine invariance), Harris-Laplace / SIFT descriptor (similitude invariance), and the proposed AMSS-based descriptor (intermediate invariance.) For all of these methods, the recognition rate decreases with the intensity of the viewpoint change. As expected, the larger the class of invariance, the better the robustness to viewpoint change. None of the methods works with a very strong viewpoint change (image Nr 5.) See text for discussion.

5 Conclusion and perspectives

In this research report we have presented a new method to extract interest points and descriptors, which are invariant to a subclass of affine mappings. The novelty of the method comes from the Affine Morphological Scale Space, which should replace the linear scale-space when affine (and no more scale / rotation) invariance is required. Experiments demonstrate that this approach is valid and competitive.

In the near future, several operators should be investigated to set the scale (Section 3.2), such as affine gradient [15] which is in particular invariant to contrast changes. The next step is also to achieve a full affine invariance.

Acknowledgements. Thanks are due to Marie-Odile Berger for her helpful advices, and to Julie Delon and Yann Gousseau for valuable discussion.

References

[1] L. Alvarez, C. Cuenca, and L. Mazorra. Morphological corner detection. Application to camera calibration. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications*, pages 21–26, Rhodes, Greece, 2001.

[2] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing: Multiscale analysis and P.D.E. *Archive for Rational Mechanics and Analysis*, 16(9):200–257, 1993.

[3] L. Alvarez and F. Morales. Affine morphological multiscale analysis of corners and multiple junctions. *International Journal of Computer Vision*, 25(2):95–107, 1997.

[4] J. Babaud, A.P. Witkin, M. Baudin, and R.O. Duda. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):26–33, 1986.

[5] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In Springer, editor, *Proceedings of the European Conference on Computer Vision*, volume LNCS 3951, pages 404–417, Graz, Austria, 2006.

[6] V. Caselles, B. Coll, and J.-M. Morel. Junction detection and filtering: a morphological approach. In *Proceedings of the International Conference on Image Processing*, volume 1, pages 493–496, Lausanne, Switzerland, 1996.

[7] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, pages 147–151, Manchester, United Kingdom, 1988.

[8] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[9] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[10] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 384–393, 2002.

[11] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[12] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2006.

[14] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An a contrario decision method for shape element recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006.

[15] P. Olver, G. Sapiro, and A. Tannenbaum. Affine invariant detection: edge maps, anisotropic diffusion, and active contours. *Acta Applicandae Mathematicae*, 59(1):45–77, 1999.

[16] G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International Journal of Computer Vision*, 11(1):25–44, 1993.

[17] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.

[18] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.

Unité de recherche INRIA Lorraine
LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399