INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Matching by local invariants

Cordelia SCHMID , Roger MOHR

N 2644

August 1995

PROGRAMME 4

Rapport de recherche

ISSN 0249-6399

Matching by local invariants

Cordelia Schmid , Roger Mohr

Programme 4 — Robotique, image et vision
Projet Movi (Gravir–Imag)

Rapport de recherche n ° 2644 — August 1995 — 37 pages

Abstract: This paper presents a matching method which is based on invariants of the luminance function. This makes matching possible in the case of important geometric transformations between images. The method is robust to occlusion, as local properties of the image signal are used. We propose to use differential measures invariant to rotations in a multi-scale framework. Promising results are obtained which might be used for a variety of applications.

Key-words: matching, differential invariants, multi-scale, interest points

(Résumé : tsvp)

This work was supported by the European Community within the framework of the Human Capital and Mobility program.

This work was done in the context of a common project between CNRS, INRIA, Université Joseph Fourier and Institut National Polytechnique de Grenoble.

Unité de recherche INRIA Rhône-Alpes
46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1 (France)
Téléphone : (33) 76 57 47 77 – Télécopie : (33) 76 57 47 54

Mise en correspondance par invariants locaux

Résumé : Le but de ce papier est de proposer une alternative aux méthodes de mise en correspondance. Cette dernière repose sur le calcul d'invariants de la fonction de luminosité. Ceci permet d'effectuer une mise en correspondance dans le cas d'une transformation importante entre deux images. Pour ce faire, nous utilisons des mesures différentielles invariantes en rotation et une approche multi-échelle. Une mise en œuvre simple de la méthode fournit des résultats de bonne qualité autorisant des applications variées.

Mots-clé : mise en correspondance, invariants différentiels, multi-échelle,point d'intérêt

Table of Contents

1 Introduction

The problem we are going to solve is matching between image points. This is one of the key problems in computer vision; potential applications include stereovision, object identification and localisation as well as target tracking.

In order to see the complexity of the matching we want to solve, the reader is referred to figures 9 to 14.

1.1 Existing matching methods

It is almost impossible to give an overview of such a wide-spread subject as matching. In the following we are only going to discuss the principal matching methods in order to explain what our contribution is.

The matching method which has been used for the longest period of time is signal correlation (i.e. see [Fau 92] which compares different methods, [Zha 89] or [Zab 94] which use a fast binary correlation). To obtain satisfactory results the signal has to be taken under very similar conditions such that the correlation windows can be superimposed by simple translation. If we have to deal with non-trivial image rotation, this method fails. However, if the epipolar geometry is known, it is possible to compensate for the geometric transformations and to obtain subpixel accuracy ([Ack 84], [Moh 94], [Lot 94]).

To restrict matching to image parts containing significant information, many of the methods use contours. It is then possible to get a structure of contours (the structure of a graph [Hor 90]) which captures the global structure of the scene. Global similarity can be used for the matching if optimizing techniques such as relaxation [Lon 86] or combinatorial techniques such as search for maximal cliques [Hor 89] are applied. However, the complexity is very high for these methods, and therefore additional constraints like the epipolar constraint in stereovision have to be added in order to keep run time low.

Local information might be calculated for contour points. Theoretically, any contour point can be characterized by an associated invariant [Wei 91]. Yet, this kind of method is unstable. If a sufficient part of the contour is seen, semi-local methods [Die 94], [Rot 92] allow to calculate invariants for the observed part of the contour. An approximative correspondence could then be calculated; to our knowledge, theses techniques have only been used for limited object recognition problems.

It is evident that all the contour based methods can only be applied if the segmentation process obtains good quality results, as it is the case for simple scenes. However, even for realistic simple scenes they remain sometimes difficult to manage.

1.2 Context and proposed approach

We want to develop a tool for matching which is sufficiently general in that we can consider images taken under very different conditions. This is necessary in case of a camera–hand auto-calibration of a manipulator: this kind of calibration implies different rotations with non parallel axes which leads to very different images of a same scene [Hor 95]. Thus, we have to identify objects and to be insensitive to any rigid displacement in the image (image rotation and translation). As the camera can approach the object, we also want to absorb a change of the scale within some limits, say a factor of 2. It is evident that we have to restrict the scale factor as too drastic changes of scale modify the structure of an object. If the scale factor is too large, a concave curve is transformed into a convex curve and eventually into a single point.

We also have to be able to deal with occlusions and different movements of the objects as well as to operate on complex scenes (i.e. figures 11 to 14).

In absence of global structure we have to use a method which uses local information. The large variety of deformations excludes the use of correlation; the absence of significant contours excludes the use of this type of information.

To obtain perfect matching results in the above cited context is unrealistic. Experiments with robust methods [Sch 90] have shown that it is sufficient for most applications if 1/3 to 2/3 of the initial matches are correct. An example of such an application is the calculation of the epipolar geometry [Zha 94]. We hence fix our goal to obtain at least 2/3 of correct matches.

The imposed geometric constraints have led us to study the signal in the neighborhood of particular points, the extracted interest points: they are local and in these locations the grey level signal contains most information. This makes their identification possible.

A related approach in litterature is to apply correlation to interest points [Zha 94]. To improve such a correlation Deriche [Der 94] adds the following local characteristics : the direction of the gradient, the curvature and a disparity measure which has to be above a fixed threshold. To obtain insensitivity to rotation and scale change, Hu [Hu 94] suggests to apply the correlation in several directions and Ballard [Bal 93] uses steerable filters which are steered in the direction of the gradient.

It is also possible to characterize a signal locally in the frequency space. Gabor filters are often used for stereo-correspondence or for verging an active stereo head [Wes 92, Fle 91, San 88]. These filters allow to obtain local phase information. This information is then used to estimate the disparity between two images. However, this measure is local in frequency and directional in space. It, therefore, only resists to small image rotations and small scale changes. In order to solve this problem, Wu [Wu 94] has proposed to use Gabor filters in several directions and at several scales.

The solutions of Hu [Hu 94] and Wu [Wu 94] suggest a spatial discretization. Thus, they rely on the hypothesis that a linear interpolation between two adjacent directions is possible. This hypothesis is difficult to verify. In addition, their method needs to calculate several values. The solution of Ballard [Bal 93] introduces an additional instability by first calculating the direction of the gradient. To solve these problems we propose to use local measures which are invariant to rotation. We then do not need any discretization and our calculation does not rely on auxiliary unstable parameters. In addition, our method is robust to scale change by using a multi-scale approach.

1.3 Introduction to invariants

In the following a short introduction to invariants is given. It is limited to the minimum required for understanding the work that follows. For more information on this topic in the context of computer vision, the reader is referred to [Mun 92] and to [Gro 92].

Let $I(x, y)$ be an image. A new image I' of the same scene can be taken under a new environment (camera with other parameters, motion of the camera, etc.). Invariants are properties that remain true under a given set of admissible changes. Three kinds of change might be considered:

1. change in the intensity only, i.e. $I'(x, y) = f(I(x, y))$. This occurs when the camera is fixed and only illumination is changing. For this reason f has to be considered as a monotonic function. Simpler approximations of this case might be made assuming that f is only a translation or an affine function, i.e.

$$I'(x, y) = I(x, y) + \mu \quad \text{or} \quad I'(x, y) = \lambda \times I(x, y) + \mu$$

2. change in geometry only, i.e. $I'(x, y) = I(g(x, y))$, where g is a mapping from the 2D plane onto itself. The simplest cases for g are translations. In case of a planar patch, it is well known that g is a homography, i.e. a linear function in homogeneous coordinates. If the change in camera position is not too large, and if the patch under consideration is small enough, then g is close to an affine function [Ull 91, Bas 94].

   If we consider the images of a ground plane taken by an airplane flying horizontally, and with images taken approximatively from the same location, then g is very close to a similarity; with the additional condition of constant altitude, g is close to a rigid motion; moreover if the airplane keeps constant direction, then g is close to a translation.

3. in the real case, both changes (1 and 2) might occur.

Even in only the case 2 above, no geometric invariant exists for a single image of 3D points [Bur 90, Mos 92]. In the case of monotonic change of intensity and rigid motion for g, Kœnderink and his co-workers [Flo 91] have shown that the only invariants are the curvature of the isophote line and the flow-line. Curvature computation is not a stable process and so are the isophotes in images under real noisy condition. For this reason, the applicability of this result is limited.

In what follows here, we assume that intensity might have be translated. If the mean luminance intensity is used prior normalisation has to be done; extending this to the affine change in intensity would be straightforward. We are considering small image patches, and therefore the local planar assumption is a good approximation. Therefore we should consider for g the group of affine transform. However, as it was illustrated in [Gro 94], similarity invariant are good quasi-invariant which are more stable in their computation. For this reason we restrict g to the group of similarity transform; f is restricted to the translational case.

1.4 Report outline

In section 2 our approach is described. It consists of three stages: extraction of interest points, characterization of these points and matching. The characterization is given by a set of values invariant to the group of image displacements and robust to a scale change. Section 2.1 presents the invariants; we rely heavily on the invariance approach developed by Kœnderink [Koe 87] and Romeny [Rom 94b]. Section 2.2 introduces a multi-scale approach robust to scale changes. The two remaining steps of our matching algorithm are given in section 2.3. In section 3 extensive experiments confirm our approach. We validate our theory for the different kinds of modifications and for different types of images.

2 The proposed method

2.1 Invariant signal characterization

We propose a characterization of the luminance signal by using differentials invariant under the group of displacements within the image. This characterization is based on combinations of derivatives invariant to image rotation and translation. The set of invariants used is stacked in a vector denoted by $\vec{\mathcal{V}}_i$. Simple examples of such invariants are gradient magnitude and Laplacian.

Differential invariants have been theoretically studied by Kœnderink and his co-workers [Koe 87, Sal 92, Flo 93, Rom 94b, Rom 94a]. They have made explicit mathematical results first derived by Hilbert [Hil 93] and they have proposed a stable implementation for the derivatives. In our work invariants up to third order are used. It is, therefore, important to solve the problem of their stability and precision.

2.1.1 Calculation of the derivatives

The calculation of the derivatives is ill-conditioned, as it lacks robustness due to the presence of noise in the image data. Even a small noise can disturb the calculation significantly. To illustrate such a lack of robustness, let us consider the functions $f(x)$ and $\hat{f}(x) = f(x) + \varepsilon \sin(\omega x)$. They are similar for a small ε. However, $f'(x)$ can be very different from $\hat{f}'(x)$ if ω is big. Consequently, a high frequency noise can significantly modify the first derivative and even more higher order derivatives.

Prior to any derivation, it is therefore necessary to smooth the signal. As the differentiation commutes with the convolution, i.e. $\partial_i(g * f) = g * \partial_i f = \partial_i g * f$, smoothing can be obtained by either smoothing the image or by smoothing the operator of derivation. A simple way to stabilize the calculation of the derivatives is therefore to use the derivatives of a smoothing function. A very common choice is to use the Gaussian function [Wit 83, Tor 86, Rom 94a, Flo 93, Lin 94]. This choice also coincides with a definition of scale-space which will be important for our multi-scale approach. The formulation of the rotationally invariant Gaussian $G(\vec{x}, \sigma)$ and its derivatives $G_{i_1 \ldots i_n}(\vec{x}, \sigma)$ for a vector $\vec{x} = (x_1, \ldots, x_D)$ ($D = 2$ in case of an image) is:

$$G(\vec{x}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^D} exp(-\frac{\vec{x}^2}{2\sigma^2}) \quad (1)$$

$$G_{i_1 \ldots i_n}(\vec{x}, \sigma) = \frac{\partial^n}{\partial_{i_1} \ldots \partial_{i_n}} G(\vec{x}, \sigma) \quad (2)$$

$$where \ \ n = 0, 1, 2 \ldots \ \ and \ \ i_k = x_1, \ldots, x_D \ \ for \ \ k = 1, \ldots, n.$$

The σ of the Gaussian function determines the quantity of smoothing. This σ can also be interpreted as a scale factor. In the following, σ will be referred to as *size* of the Gaussian.

Having calculated the derivatives of a function in a point up to Nth order, the Taylor expansion describes this function locally up to this order. Thus, an image in a point is then described by the set of derivatives which is stacked in a vector. This vector has been used by Kœnderink [Koe 87] who named it "local jet".

Definition 2.1 *Local Jet*
Let I be an image and σ a given scale. The "local jet" of order N at a point

$\vec{x}$, denoted by $J^N[I](\vec{x},\sigma)$, is defined by

$$J^N[I](\vec{x},\sigma) = \{L_{i_1\ldots i_n}(\vec{x},\sigma) \mid (\vec{x},\sigma) \in I \times \mathbb{R}^+ \,; n = 0,\ldots,N\}$$

in which $L_{i_1\ldots i_n}(\vec{x},\sigma)$ is the convolution of image I with the Gaussian derivatives $G_{i_1\ldots i_n}(\vec{x},\sigma)$ defined in equation 2 :

$$L_{i_1\ldots i_n}(\vec{x},\sigma) = (G_{i_1\ldots i_n} * I)(\vec{x},\sigma)$$

The local jet takes into account the size σ of the Gaussian used for calculating the derivatives. This makes it possible to define a multi-scale local jet and hence to characterize a function at several scale levels ([Koe 87]). In addition, the calculation is optimal as it can be adjusted to the considered image [Lin 94].

2.1.2 Size normalization of the image

To obtain derivatives invariant to rotation, it is absolutely necessary that the pixel size is quadratic. Otherwise, the calculation of the derivatives is influenced by the relation length/width of a pixel. Thus, we have to normalize the image. This is done by linear interpolation on the columns of the image using a reduction factor equal to the factor "α_v/α_u". This normalization factor "α_v/α_u" corresponds to the ratio between length and width of a pixel. Different experiments have shown that this factor is stable and almost independent of calibration. For the results presented in section 3, the factor used is 1.47.

2.1.3 Differential invariants

Having normalized the image size, it is possible to calculate a local jet for any point of this image. Differential invariants under the group of displacements are then calculated using the local jet. The invariants used in our experiments are stacked in a vector, denoted by $\vec{\mathcal{V}}_i$.

The first part of this vector contains the complete and irreducible set of differential invariants up to 2nd order (cf. equation 3). The formulation of this first part of the vector is given in Einstein notation and in Cartesian coordinates.

$$\vec{\mathcal{V}}_i[0..4] = \begin{bmatrix} L \\ L_iL_i \\ L_iL_{ij}L_j \\ L_{ii} \\ L_{ij}L_{ji} \end{bmatrix} = \begin{bmatrix} L \\ L_xL_x + L_yL_y \\ L_{xx}L_xL_x + 2L_{xy}L_xL_y + L_{yy}L_yL_y \\ L_{xx} + L_{yy} \\ L_{xx}L_{xx} + 2L_{xy}L_{yx} + L_{yy}L_{yy} \end{bmatrix} \qquad (3)$$

The L_i are the elements of the "local jet" defined in definition 2.1 where L represents the luminance function convolved with a Gaussian. It is therefore possible to calculate the L_i and consequently the invariants for different sizes σ of the Gaussian.

The luminance function depends on a two-dimensional vector $\vec{x}$. In Cartesian notation we have $\vec{x} = (x, y)$. The indices x and y represent the derivation with respect to the variables x and y respectively, i.e. $L_{xy} = \frac{\partial^2}{\partial_x \partial_y} L$. In Einstein notation, an index i signifies the summation of the derivations with respect to the set of variables: $L_i = \sum_i L_i = L_x + L_y$ and $L_{ij} = \sum_i \sum_j L_{ij} = L_{xx} + L_{xy} + L_{yx} + L_{yy}$.

The second part of the vector contains a set of complete invariants of third order in Einstein notation (cf. equation 4).

$$\vec{\mathcal{V}}_i[5..8] = \begin{bmatrix} \varepsilon_{ij}(L_{jkl}L_iL_kL_l - L_{jkk}L_iL_lL_l) \\ L_{iij}L_jL_kL_k - L_{ijk}L_iL_jL_k \\ -\varepsilon_{ij}L_{jkl}L_iL_kL_l \\ L_{ijk}L_iL_jL_k \end{bmatrix} \qquad (4)$$

with ε_{ij} the 2D antisymmetric Epsilon tensor defined by $\varepsilon_{12} = -\varepsilon_{21} = 1$ and $\varepsilon_{11} = \varepsilon_{22} = 0$.

2.2 Robustness to a scale change

To obtain robustness to a significant scale change, the invariants have to be adapted to the scale change. A change in image scale can be caused either by a change in the distance of the camera to the object or by a change in the focal length (zoom). Such a scaling of the image is in the following denoted by α in order to distinguish it from the size σ of the Gaussian used for calculating the derivatives.

Experiments have shown that matching which uses invariants resists to a change in scale of 20 %. This observation is coherent with the one made by Fleet [Fle 91] in the context of Gabor filters. Equivalently, correlation based matching can, in case of absence of rotation, be used up to a change in scale of 20 %.

2.2.1 Adapting the invariants to a scale change

For a function f, a scale change α can be described by a simple change of variables, $f(x) = g(u)$ where $g(u) = g(u(x)) = g(\alpha x)$. We then obtain:

$$f(x) = g(u)$$

$$f'(x) = \alpha g'(u)$$
$$f''(x) = \alpha^2 g''(u)$$

This shows that the first derivatives are equal up to a multiplicative factor α. Equivalently, the second derivatives (resp. nth) are equal up to a multiplicative factor α^2 (resp. α^n). Thus $f'(x)^2/f''(x)$ is a theoretical invariant to scale change.

However, in case of a discrete representation of the function, as for an image, the derivatives are calculated on a Gaussian window. If the size of the window is not adapted to the scale change, the differential invariants to scale change are not valid. For two images I_1 and I_2 where I_2 is changed by a scale factor α we have:

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G(\vec{x}, \sigma) d\vec{x} = \int_{-\infty}^{+\infty} I_2(\vec{u}) G(\vec{u}, \sigma\alpha) d\vec{u}$$

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1}(\vec{x}, \sigma) d\vec{x} = \alpha \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1}(\vec{u}, \sigma\alpha) d\vec{u} \tag{5}$$

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1 i_2}(\vec{x},\sigma) d\vec{x} = \alpha^2 \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1 i_2}(\vec{u},\sigma\alpha) d\vec{u}$$

in which the G_{ii} are the derivatives of the Gaussian as defined in equation 2.

Equation 5 shows that the size of the Gaussian, that is the calculation support, has to be adjusted. If the scale change α between two images is known, it is possible to obtain same differential invariants up to a multiplicative factor α. The remaining problem is to either estimate α or to be capable of dealing without this estimation.

2.2.2 Estimating the change in scale

A solution for estimating the scale change α between two images has been proposed by Manmatha [Man 94b, Man 94a] in the general context of local calculation of an affine transformation. His idea is based on the fact that convolution of an image with a Gaussian is equivalent to the convolution of an affine transformed image with the affine transformation of this Gaussian. In the case of simple change in scale this idea is expressed in equation 5. This implies that it is sufficient to continually change the size σ of the Gaussian to obtain the change in α, that is to calculate the value of the convolution of the image with the Gaussian at several sizes σ. The estimation of the scale factor is thus obtained by:

$$\operatorname*{argmin}_{\alpha}\left(\int_{-\infty}^{+\infty} I_1(\vec{x}) G(\vec{x},\sigma) - \int_{-\infty}^{+\infty} I_2(\vec{u}) G(\vec{u},\sigma\alpha)\right)$$

Our experiments have revealed that this method is very unstable: the computed value is very sensitive to noise and to small illumination changes. Satisfactory results can only be obtained if we use accurate point positions, an important support for calculation and if the influence of noise and illumination changes are insignificant. This is incompatible with our constraints (cf. section 1.2).

We suggest here to use the multi-scale approach in another way. Not only the mean but also derivatives or invariants are computed at different scales.

They are used to estimate the scale factor α. Using invariants makes the characterization not only robust to a change in scale, but also invariant to image rotation. The vector $\vec{\mathcal{V}}_i$ is calculated for several sizes of the Gaussian σ_k. An important problem is the discretization in the size σ. Often discretization of half-octave is used, as for example Wu [Wu 94]. We have found that this is not sufficient. Based on the fact that our measurements are robust to a scale change up to 20%, our discretization ensures that the difference between two consecutive sizes is inferior to 20%. As we have allowed a change in scale up to 2, the sizes σ_k used vary between 0.4 (1/2.5) and 2.5.

2.3 Steps of the matching process

We are now going to describe the different stages of our matching algorithm. At first interest points are extracted. It is obvious that the quality of the final results depends on the quality of the extracted points. The detector used should be invariant to rotations, repeatable in position and approximatively invariant to a scale change. Having compared different corner detectors [Bea 78, Kit 82, Mor 79, Har 88, Hei 92, Ros 92], we use the detector of Heitger and Rosenthaler.

With their method an image is convolved with even and odd directional filters in several directions (which can be specified by the user). The filters used are very similar to Gabor filters. The results of the even and odd filters in a same direction are used to calculate the local energy of the image in this direction. This energy corresponds to the 1D characteristics of the image. To obtain 2D characteristics of the image, the first and second derivatives are calculated for each direction. This measurement detects the 2D characteristics, but also gives false responses to 1D characteristics. A method which relies on the systematic nature of the errors makes it possible to eliminate these false responses to the 1D characteristics. The remaining responses are then thresholded.

Having extracted the interest points, these points are characterized by $\mathcal{V}(\vec{x}, \sigma_k)$ as described in section 2.1 and 2.2. To match these points we use the following algorithm : for each vector of a point $\vec{x_1}$ of the first image calculated at scale 1 we look for a vector of a point $\vec{x_2}$ of the second image calculated at a scale σ_{k_2} such that $d(\vec{\mathcal{V}}_i(\vec{x_1}, 1), \vec{\mathcal{V}}_i(\vec{x_2}, \sigma_{k_2}))$ is minimal. The process is then reversed. Hence, we obtain two lists of matched points. The matches kept are the points which choose each other mutually and for which $\sigma_{k_1} = 1/\sigma_{k_2}$. This matching algorithm is simple and can be improved by adding problem dependent constraints like consistency of the matches or by using statistical thresholding.

The distance d used to compare two vectors is the Mahalanobis distance (cf. [Ram 89, Aya 90, Fau 93]). This distance makes it possible to compare two random vectors whose components have different sizes. Additionally, the Mahalanobis distance is a random variable which has a χ^2 distribution. If we use a table of this distribution, it is possible to threshold this distance such that we keep k% of the values. This allows for the rejection of (1-k) % of the values which are not very likely and which correspond to false matches.

3 Experiments

In this section we present experimental results. To validate our theoretical results, we first present separately results for each type of image transformation : in section 3.2 for an image rotation and in section 3.3 for a scale change. Then in section 3.4 results for similarity transformations (rotation and scale) are shown. Section 3.5 displays results for three dimensional displacements.

The quality of the results is measured by the percentage of correct matches in relation to the number of total matches. This evaluation has been made for more than 100 images. On the other hand, the number of total matches is often larger than 100. Thus, it is impossible to evaluate manually the results. We have therefore used a method for automatic evaluation of the results which is presented in section 3.1. Another criterion is the percentage of total matches in relation to the number of points extracted. This percentage varies depending on the image as well as on the type of transformation between the images ; the average percentage of matched points is about 50%.

For our experiments, a vector of characteristics which contains 9 components (cf equations 3 and 4) has been used. This includes third order derivatives. Experiments, which will not be discussed here, have shown that using third order derivatives in addition to second order ones improves the percent of good matches with about 10%.

3.1 Automatic evaluation of the results

To automatically evaluate the results, geometric constraints between two views have been used. In case of a planar scene, any change between two images of this scene can be modeled by a homography [Sem 52]. This homography defines a point to point relation between two images. By applying the homography, it can be determined, whether or not a match is correct : the percentages given in the following figures correspond to the rate of correct matches in relation to the found matches.

For our experiments we have calculated the homography between two images using the found matches. This calculation uses a method of least-square median and is therefore robust to errors up to 50%.

In case of non-planar scenes, the geometric constraint between two views is defined by the fundamental matrix : to a point in the first image corresponds an epipolar line in the second image. A match (a, b) is evaluated as correct if the point b lies on the epipolar line which corresponds to a. This evaluation is not perfect : two points of a false match might lie on corresponding epipolar lines. As the probability that a mismatch satisfies the epipolar constraint is very low, this is still a good estimation of the number of correct matches.

The epipolar geometry between two images has been calculated using the found matches. This calculation uses a method of least-square median (cf. [Zha 94, Bou 95]).

3.2 Invariance to image rotation

To test invariance to image rotation, we have made experiments for two different planar scenes. The first one is a painted folder referred to as "Sanja" in the following. The second one is the painting "the sower" of "Vangogh". "Vangogh" is more difficult, as it contains texture. Texture makes interest point extraction more imprecise.

For each scene we have taken a sequence of 20 images between which we have rotated the camera around the optical axis. Figure 1 shows the results for a pair of images of the sequence "Sanja" for which the rotation angle is 133.0 degrees. The white crosses indicate the correct matches and the black ones the false matches. The rate of correct matches is 88.52%. Figure 2 displays the results for the sequence "Vangogh" for which the rate of correct matches is 74.26%.

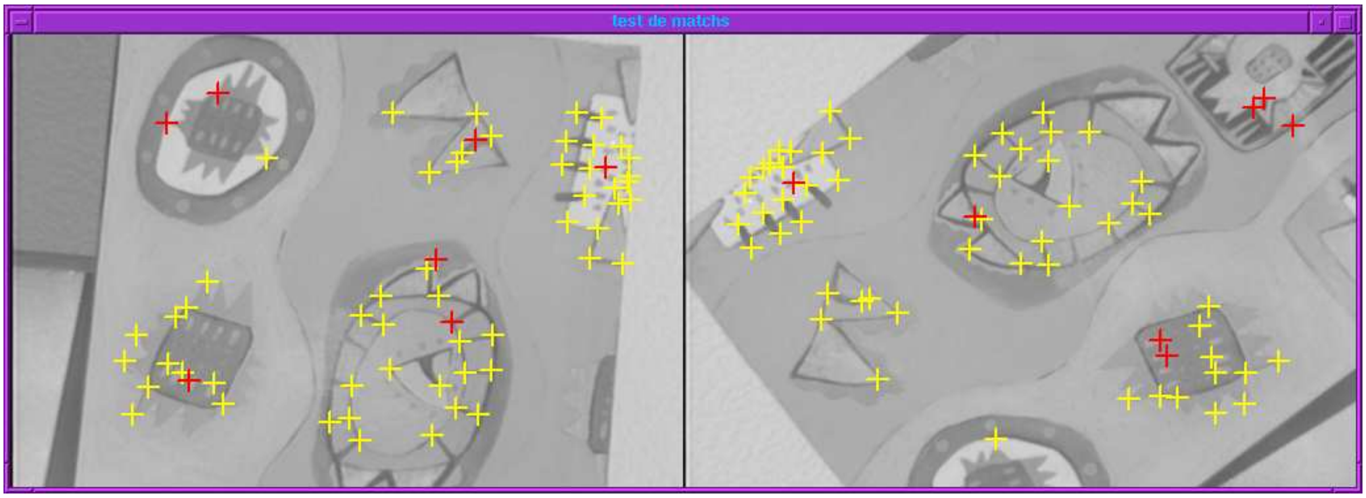

Figure 1: 88.52% correct matches for a rotation of 133.0 degrees

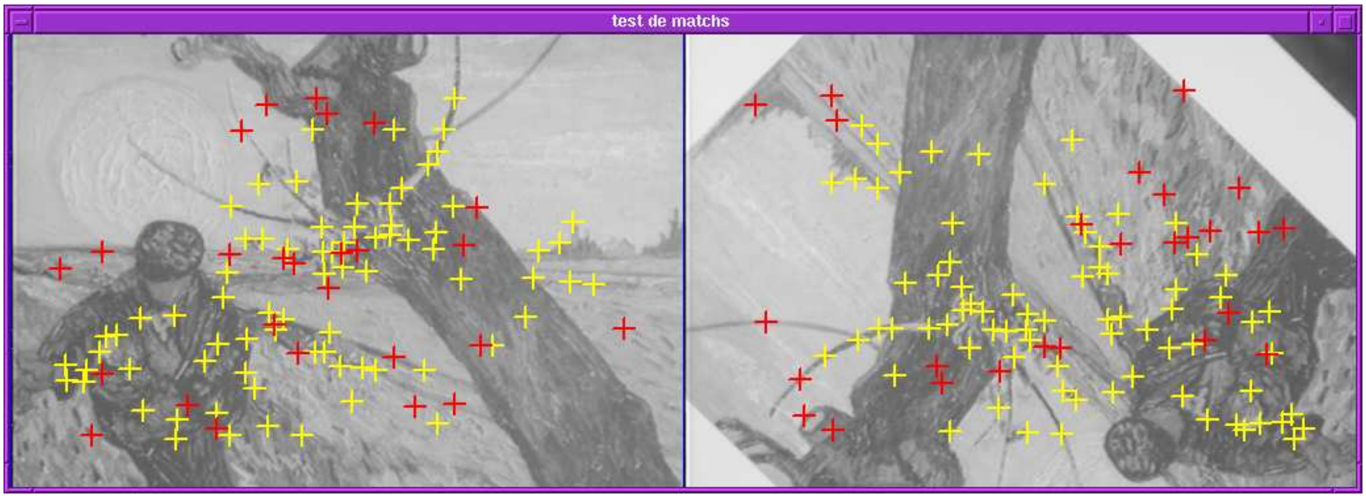

Figure 2: 74.26% correct matches for a rotation of 134.8 degrees

Figure 3 illustrates the percentage of correct matches as a function of the angle for the sequence “Vangogh”. This percentage is calculated for three different sizes σ of the Gaussian. The curves differ according to the size σ used. The bigger this size is, the better the results are. This is mainly due to the fact that bigger sizes σ are less sensitive to imprecision of the position as important smoothing makes the transitions less abrupt. However, sizes bigger than 7 do not improve the results anymore.

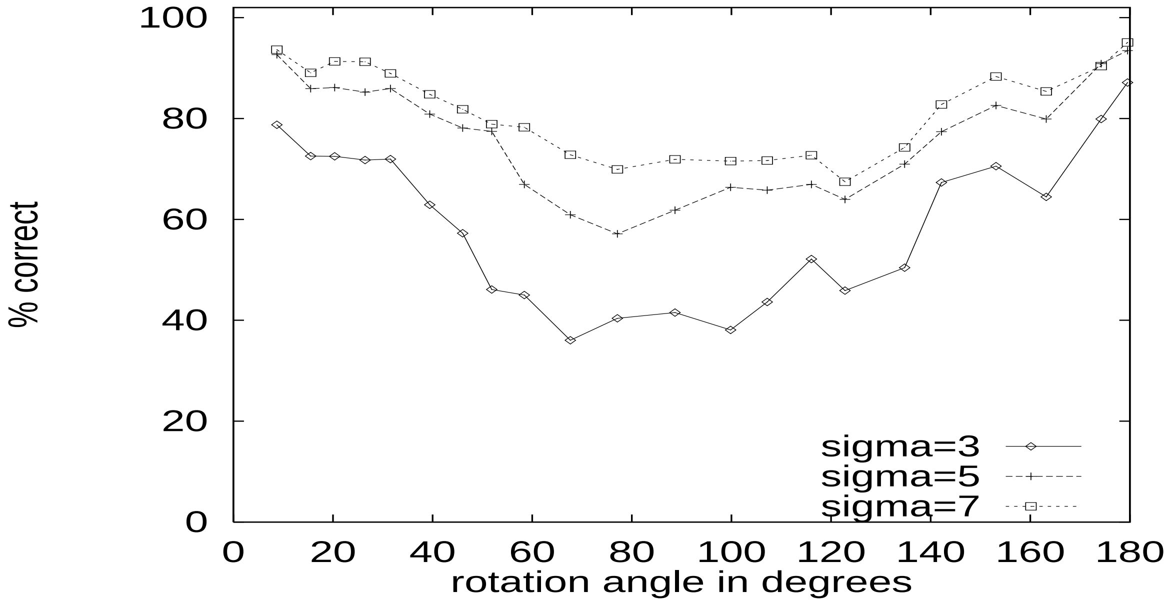

Figure 3: Percentage of correct matches with respect to the rotation angle

Equivalent results are obtained for the sequence “Sanja” (cf. figure 16 in appendix A). The results are slightly better than those obtained for the sequence “Vangogh”.

Figure 3 also shows that the results get worse for an important rotation. The worst results are obtained for a rotation angle of $\pi/2$. This is due to the fact that the results of the detector are not repeatable at different orientations as well as the fact that the images do not cover up for such an angle. In figure 4 the influence of these factors are shown.

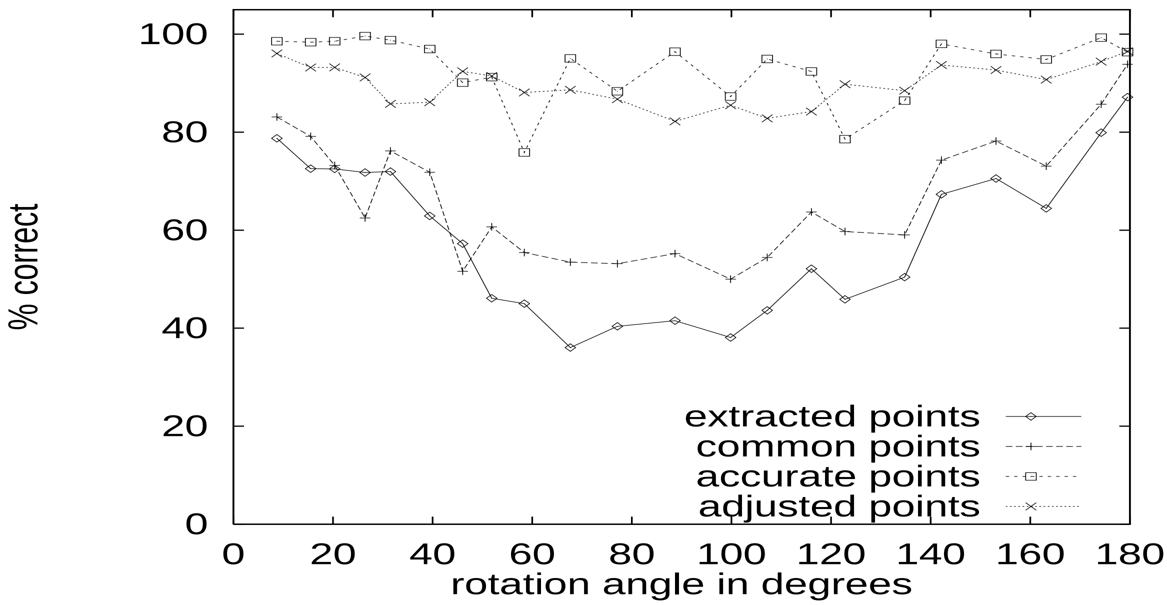

Figure 4: Evaluation of the influence of the detector; percentage of correct matches with respect to the rotation angle, the size σ of the Gaussian is 3

Figure 4 compares the results for the extracted points to the results for the points restricted to the common zone and to the results for accurate points projected by the homography (no subpixel accuracy is used). The results are given for the sequence "Vangogh" and a size σ equal to 3. We can see that the method is mainly sensitive to the accuracy of the points used. The same results are given for sequence "Sanja" in figure 17 in the appendix A.

As our method is sensitive to the position of the points, we can improve our algorithm by adjusting our points; that is by looking for the best match in a neighborhood of a point. The results for adjusting our points is given by the graph named "adjusted points" in figure 4 and in figure 17. To obtain similar results as by using the precise points projected by the homography, a neighborhood of size 1 had to be used for sequence "Sanja" and a neighborhood of size 2 for sequence "Vangogh".

3.3 Robustness to a scale change

For each scene we have taken 40 images for which we have changed the focal length of the lens (zoom). Figure 5 shows the results obtained for a pair of images of the sequence "Sanja" for which the scale factor is 1.5. The white crosses indicate the correct matches, the black ones the false matches. The rate of correct matches is 90.91%. Figure 6 shows the results obtained for a pair of images of the sequence "Vangogh" for a scale factor of 1.5. The rate of correct matches is 70.31%.

The percentage of correct matches depending on the scale factor is given in figure 7. We can see that up to a scale factor of 1.5 we obtain satisfactory results. For a more important scale change we obtain insufficient results. This is due to the imprecision of the detected interest points. Figure 8 shows that the interest point detector is not robust to scale changes, as it was not to orientation.

Similar results for the sequence "Sanja" are shown in figures 18 and 19 in appendix A. We obtain satisfactory results up to a scale factor of 1.8 in comparison to a scale factor of 1.5 for "Vangogh". This is due to the absence of texture in this sequence. Thus, for the sequence "Sanja" the interest point detector works well for more important scale changes.

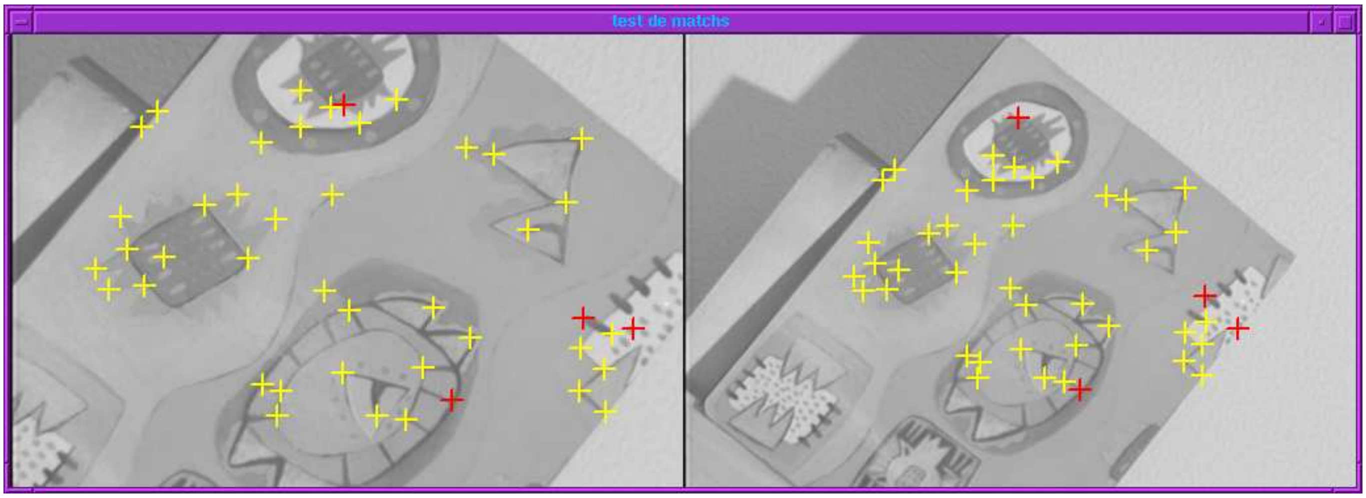

Figure 5: 90.91% correct matches for a scale factor of 1.5

Figure 6: 70.31% correct matches for a scale factor of 1.5

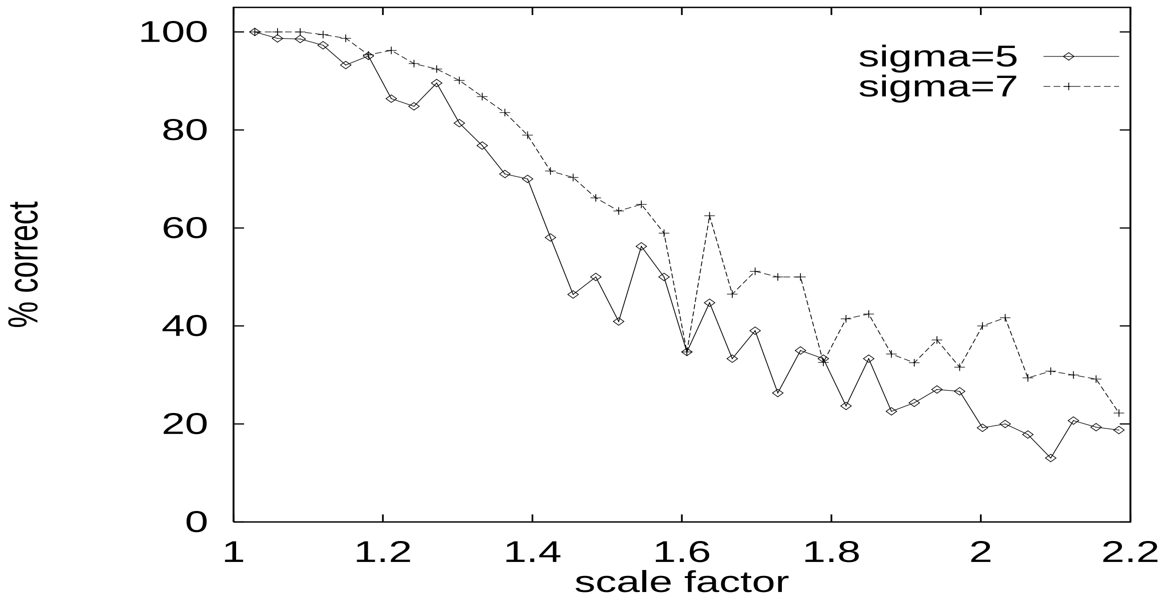

Figure 7: Percentage of correct matches with respect to the scale factor

3.4 Combining rotation and scale change

Our method allows to combine rotation and scale change, since it is invariant to image rotation and robust to scale change at the same time. Figure 9 displays the result for the sequence "Sanja" in case of a rotation of 108.8 degrees and a scale factor of 1.5. As before, white crosses indicate correct matches and black ones false matches. The percentage of correct matches is 82.35%. Figure 10 shows a result for the sequence "Vangogh" in case of a rotation of 99.8 degrees and a scale factor of 1.5. The percentage of correct matches is 68.42%. We can see that the proposed method works well for a similarity transformation.

3.5 Three dimensional scenes

We present results for three dimensional scenes. For the evaluation of our results we have used the algorithm described in 3.1. Figure 11 shows the result obtained on a complex outdoor scene. The transformation between the two images is not very important, although there are many repetitive patterns in

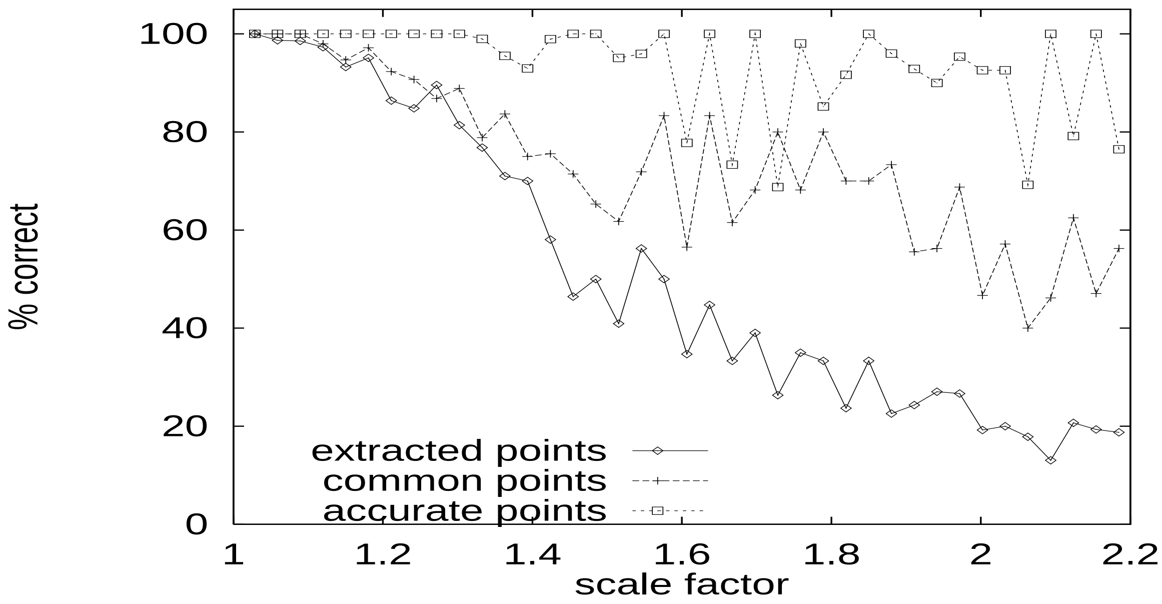

Figure 8: Evaluation of the influence of the detector; percentage of correct matches with respect to the scale factor, the size σ of the Gaussian is 5

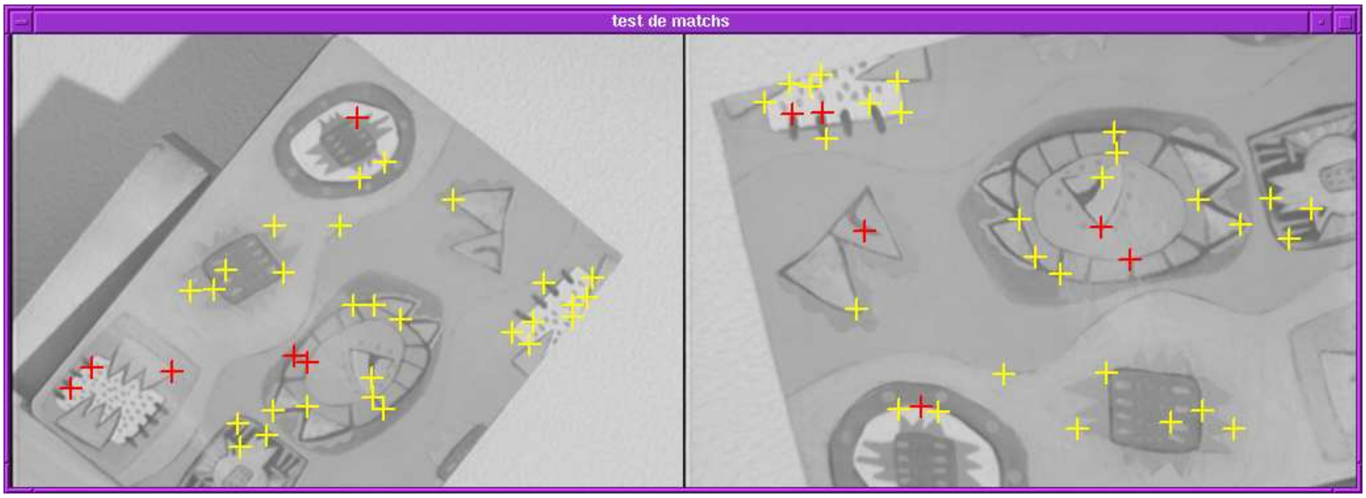

Figure 9: 82.35% correct matches for a rotation of 108.8 degrees and a scale factor of 1.5

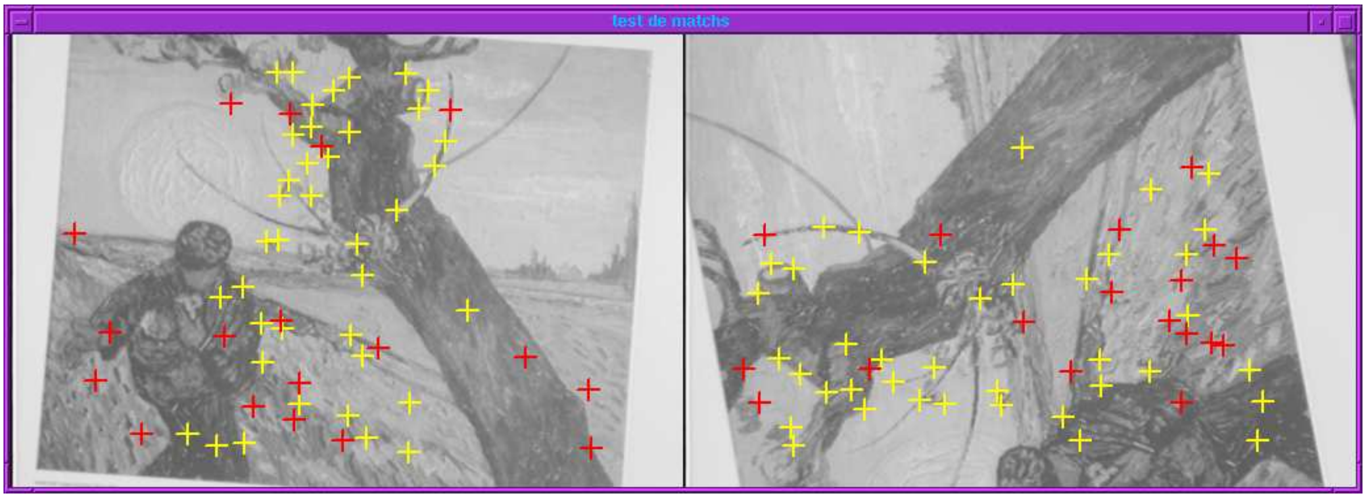

Figure 10: 68.42% correct matches for a rotation of 99.8 degrees and a scale factor of 1.5

the image. We obtain a score of 84.34 %. This can be considered satisfactory for such a scene. We notice that some false matches are due to repetitive patterns such as the ones on the window.

Figure 12 gives the result obtained on a complex scene. The transformation between the two images consists of a scene rotation, camera rotation and a scale change. We obtain 80.0% correct matches. Again false matches are primarily due to repetitive patterns. Figure 13 shows the displacement vectors for the correct matches in figure 12. Classical correlation based algorithms completely fail on such a scene.

Figure 14 gives results obtained on a complex indoor scene. The transformation between the two images consists again of a scene rotation, camera rotation and a scale change. We obtain 84.9% correct matches. Figure 15 shows the displacement vectors for the correct matches in figure 14. We can see that there is one match which has been falsely evaluated as good match. This occurs since we use epipolar geometry for evaluation (cf. section 3.1).

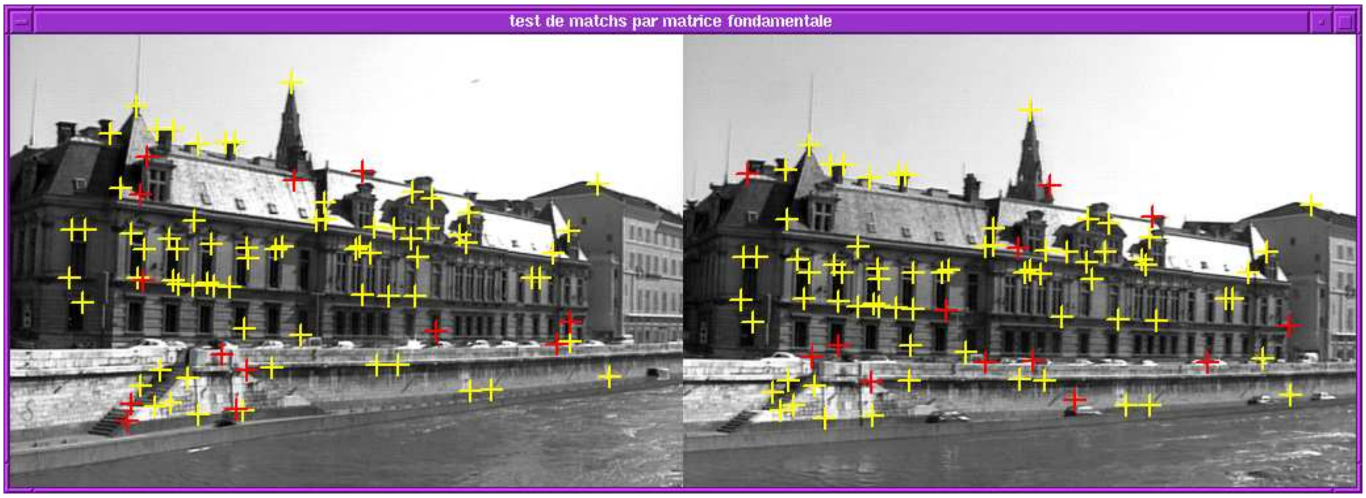

Figure 11: 84.34 % correct matches

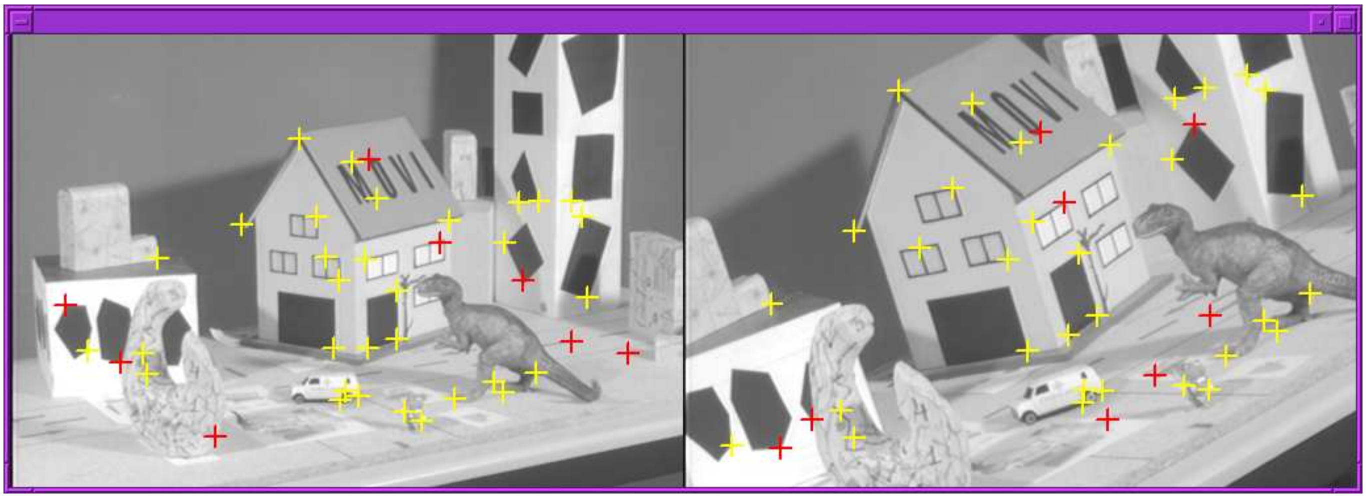

Figure 12: 80.0% correct matches

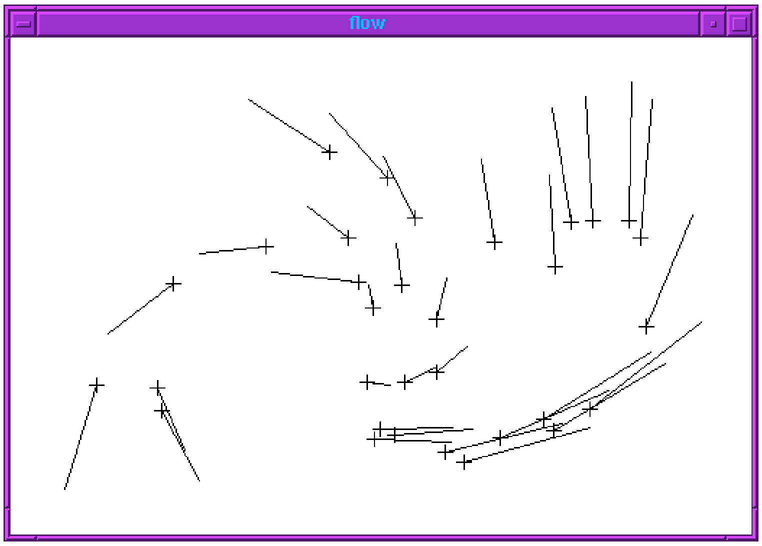

Figure 13: Displacement vectors for the correct matches in figure 12

Figure 14: 84.9% correct matches

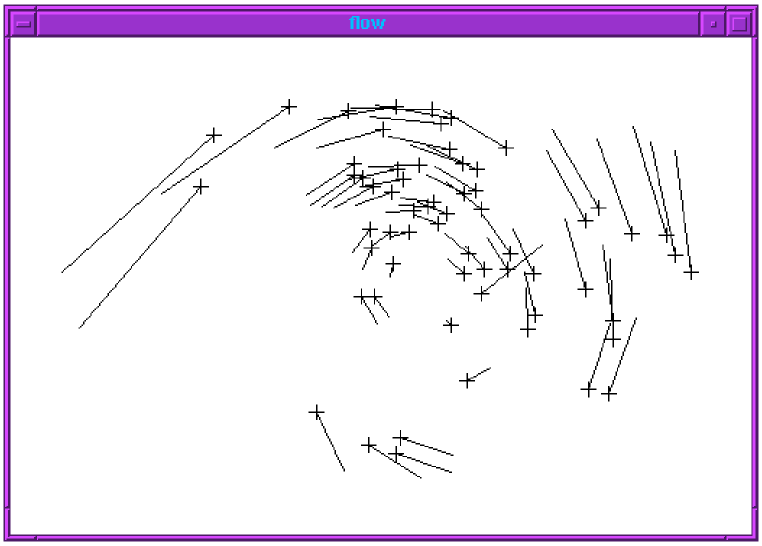

Figure 15: Displacement vectors for the correct matches in figure 14

4 Conclusion and Discussion

Differential invariants can be used to characterize points in the image. If prior works have described these invariants, it remained to determine how to calculate them such that they could practically be applied. In particular, the formulas which are invariant for different scales are not applicable, as these calculations are not supposed to be influenced by the scale factor ; for this reason they were not used here.

The matching we experimented was only based on local features and is therefore insensitive to occlusion. It is a simple brute force matching involving comparison of a vector of invariants and using the Mahalanobis distance. Direct application to this matching lead to a number of correct matches almost over 50%, even in case of large scaling and rotation.

However, location of interest points is often very poor, and this explains the bad results obtained under such strong changes in the image : in case of a real rotation the point detection can result in errors up to 2 pixels ; in case of a significant scale change the point detection results in even more important errors. If interest points are located with the accuracy of 0.5 pixel, then the results increase up to 90% of correct matches even under the strong conditions we were considering here.

A possible extension of our work is to include affine deformations. The problem is then to determine affine invariants and to calculate them on a window support for which we allow affine variations.

References

[Ack 84] F. Ackermann. Digital image correlation: performance and potential application in photogrammetry. *Photogrammetric Record*, 64(11): 429–439, October 1984.

[Aya 90] N. Ayache. *Stereovision and sensor fusion*. MIT-Press, 1990.

[Bal 93] D. H. Ballard and L. E. Wixson. Object recognition using steerable filters at multiple scales. In *IEEE Workshop on Qualitative Vision, CVPR93*, 1993.

[Bas 94] R. Basri and E. Rivlin. Localization and homing using combinations of model views. Technical report cs94-12, Weizmann Institut of Sciene, June 1994.

[Bea 78] P.R. Beaudet. Rotationally invariant image operators. In *International Conference on Pattern Recognition*, pages 579–583, 1978.

[Bou 95] B. Boufama and R. Mohr. Epipole and fundamental matrix estimation using the virtual parallax property. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, June 1995. to appear.

[Bur 90] J.B. Burns, R. Weiss, and E.M. Riseman. View variation of point set and line segment features. In *Proceedings of* DARPA *Image Understanding Workshop, Pittsburgh, Pennsylvania, USA*, pages 650–659, 1990.

[Der 94] R. Deriche, Z. Zhang, Q.-T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In *European Conference on Computer Vision*, pages 567–576, 1994.

[Die 94] M. Van Diest, L. Van Gool, T. Moons, and E. Pauwels. Projective invariants for planar contour recognition. In J.O. Eklund, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 527–534. Springer-Verlag, 1994.

[Fau 92] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang. Quantitative and qualitative comparisons of some area and feature-based stereo algorithms. In W. Forstner, R. Haralick, and B. Radig, editors, *Robust Computer Vision*, pages 1–26, February 1992. Comparisons of area and feature-based stereo algorithms.

[Fau 93] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint.* Artificial intelligence. M.I.T. Press, Cambridge, MA, 1993.

[Fle 91] D.J. Fleet, A.D. Jepson, and M.R.M. Jenkin. Phase-base disparity measurement. *Computer Vision Graphics and Image Processing*, 53(2): 198–210, 1991.

[Flo 91] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever. General intensity transformations and second order invariants. In *Proceedings of the 7th Scandinavian Conference on Image Analysis, Aalborg, Denmark*, pages 338–345, August 1991.

[Flo 93] Luc Florack. *The Syntactical Structur of Scalare Images.* PhD thesis, Universiteit Utrecht, November 1993.

[Gro 92] P. Gros and L. Quan. Projective Invariants for Vision. Technical Report RT 90 IMAG - 15 LIFIA, LIFIA–IRIMAG, Grenoble, France, December 1992.

[Gro 94] P. Gros. Using quasi-invariants for automatic model building and object recognition: an overview. In *Proceedings of the NSF-ARPA Workshop on Object Representations in Computer Vision, New York, USA*, December 1994.

[Har 88] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference*, pages 147–151, 1988.

[Hei 92] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kuebler. Simulation of neural contour mechanism: from simple to end-stopped cells. *Vision Research*, 32(5): 963–981, 1992.

[Hil 93] D. Hilbert. Ueber die vollen invariantensystemen. *Math. Annalen*, 42: 313–373, 1893.

[Hor 89] R. Horaud and Th. Skordas. Stereo correspondence through feature grouping and maximal cliques. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(11): 1168–1180, 1989.

[Hor 90] R. Horaud, T. Skordas, and F. Veillon. Finding geometric and relationnal structures in an image. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, Lecture Notes in Computer Science, pages 374–384. Springer-Verlag, April 1990.

[Hor 95] R. Horaud, R. Mohr, F. Dornaika, and B. Boufama. The advantage of mounting a camera onto a robot arm. In *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xian, China*, pages 206–213, April 1995.

[Hu 94] X. Hu and N. Ahuja. Feature extraction and matching as signal detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(6): 1343–1379, 1994.

[Kit 82] L. Kitchen and A. Rosenfeld. Gray-level corner detection. *Pattern Recognition Letters*, 1: 95–102, 1982.

[Koe 87] J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55: 367–375, 1987.

[Lin 94] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[Lon 86] P. Long and G. Giraudon. Stereo matching based on contextual line-region primitives. In *Proceedings of the 8th International Conference on Pattern Recognition, Paris, France*, 1986.

[Lot 94] J.L. Lotti and G. Giraudon. Adaptive window algorithm for aerial image stereo. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages 701–703, October 1994.

[Man 94a] R. Manmatha. A framework for recovering affine transforms using points,lines or image brightness. In *Computer Vision and Pattern Recognition*, pages 141–146, 1994.

[Man 94b] R. Manmatha. Measuring the affine transform using gaussian filters. In *European Conference on Computer Vision*, pages 159–164, 1994.

[Moh 94] R. Mohr, P. Brand, and P. Remagnino. Correlation techniques in adaptive template matching with uncalibrated cameras. In *Vision Geometry III, SPIE's international symposium on photonic sensors & control for commercial applications*, volume 2356, pages 252–253, October 1994.

[Mor 79] H. Moravec. Visual mapping by a robot rover. In *IJCAI79*, pages 598–600, 1979.

[Mos 92] Y. Moses and S. Ullman. Limitations of non model–based recognition. In *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 820–828, May 1992.

[Mun 92] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision.* MIT Press, Cambridge, Massachusetts, USA, 1992.

[Ram 89] F. Ramparany. *Perception multisensorielle de la structure géométrique d'une scène.* Thèse de doctorat, Institut National Polytechnique de Grenoble, France, February 1989.

[Rom 94a] B.M. ter Haar Romeny. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.

[Rom 94b] B.M. ter Haar Romeny, L.M.J. Florack, A.H. Salden, and M.A. Viergever. Higher order differential structure of images. *Image and Vision Computing*, 12(6): 317–325, 1994.

[Ros 92] L. Rosenthaler, F. Heitger, O. Kuebler, and R. von der Heydt. Detection of general edges and keypoints. In *European Conference on Computer Vision*, pages 78–86, 1992.

[Rot 92] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mundy. Canonical frames for planar object recognition. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 757–772. Springer-Verlag, May 1992.

[Sal 92] A.H. Salden, B.M. ter Haar Romeny, L.M.J. Florack, M.A. Viergever, and J.J. Koenderink. A complete and irreducible set of local orthogonally invariant features of 2-dimensional images. In *11th International Conference on Pattern Recognition*, pages 180–184, 1992.

[San 88] T. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 2(59): 405–418, 1988.

[Sch 90] B. G. Schunck. Robust computational vision. In *Proceedings of the International Workshop on Robust Computer Vision, Seattle, Washington, USA*, pages 1–19, October 1990.

[Sem 52] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry.* Oxford Science Publication, 1952.

[Tor 86] V. Torre and T.A. Poggio. On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2): 147–163, 1986.

[Ull 91] S. Ullman and R. Basri. Recognition by linear combinations of models. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(10): 992–1006, 1991.

[Wei 91] I. Weiss. Noise-resistant invariant of curves. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Reykjavik, Iceland*, pages 319–344, 1991.

[Wes 92] C.J. Westelius, Knutsson H., and Wiklund J. Robust vergence control using scale-space phase information. Technical Report LiTH-ISY-I-1363, Linöpings tekniska högskola, Department of Electrical Engineering, May 1992. Sweden.

[Wit 83] A.P. Witkin. Scale-space filtering. In *International Joint conference on Artifical Intelligence*, pages 1019–1023, 1983.

[Wu 94] X. Wu and B. Bhanu. Target recognition using multi-scale gabor filters. In *ARPA Image Understanding Workshop, Monterey, CA*, November 1994.

[Zab 94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance, May 1994.

[Zha 89] J. Zhao. *Extraction d'information tri-dimensionnelle par stéréovision.* Thèse de doctorat, Université Paul Sabatier, Toulouse, July 1989.

[Zha 94] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. Rapport de recherche 2273, INRIA, May 1994.

A Some more results

In this appendix we give results for the sequence "Sanja". The equivalent figures have been shown in section 3 for the sequence "Vangogh".

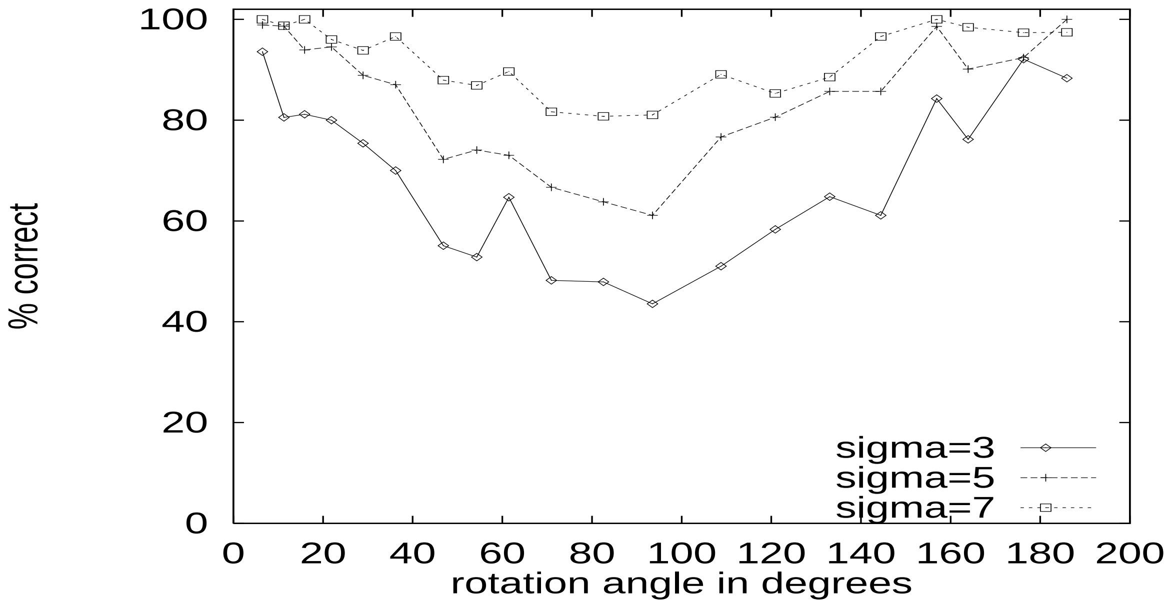

Figure 16: Percentage of correct matches with respect to the rotation angle

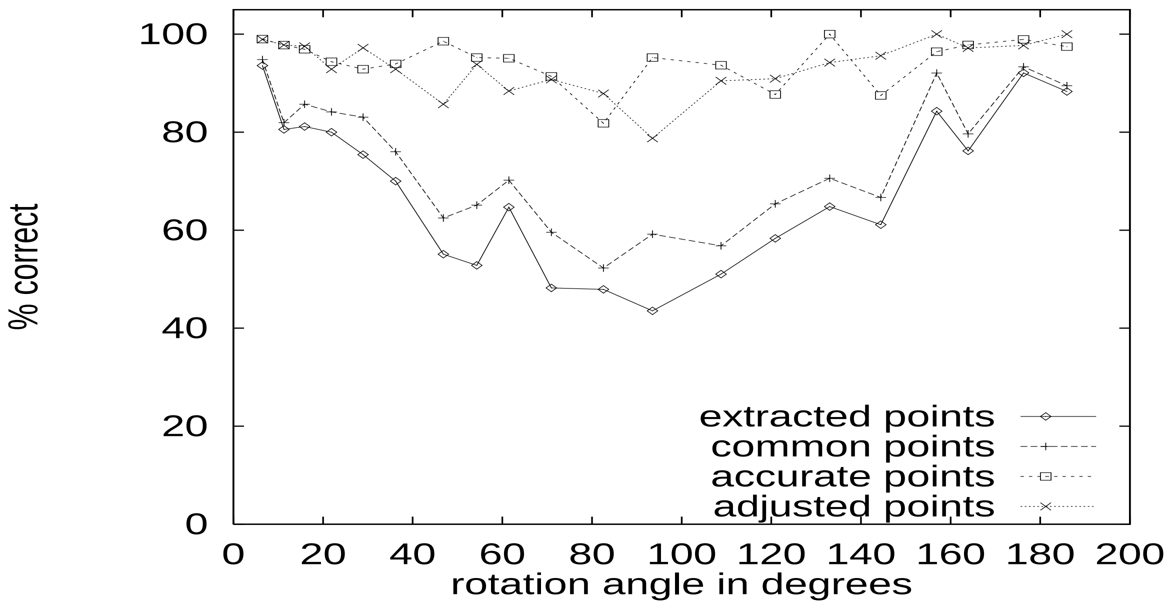

Figure 17: Evaluation of the influence of the detector; percentage of correct matches with respect to the rotation angle, the size σ of the Gaussian is 3

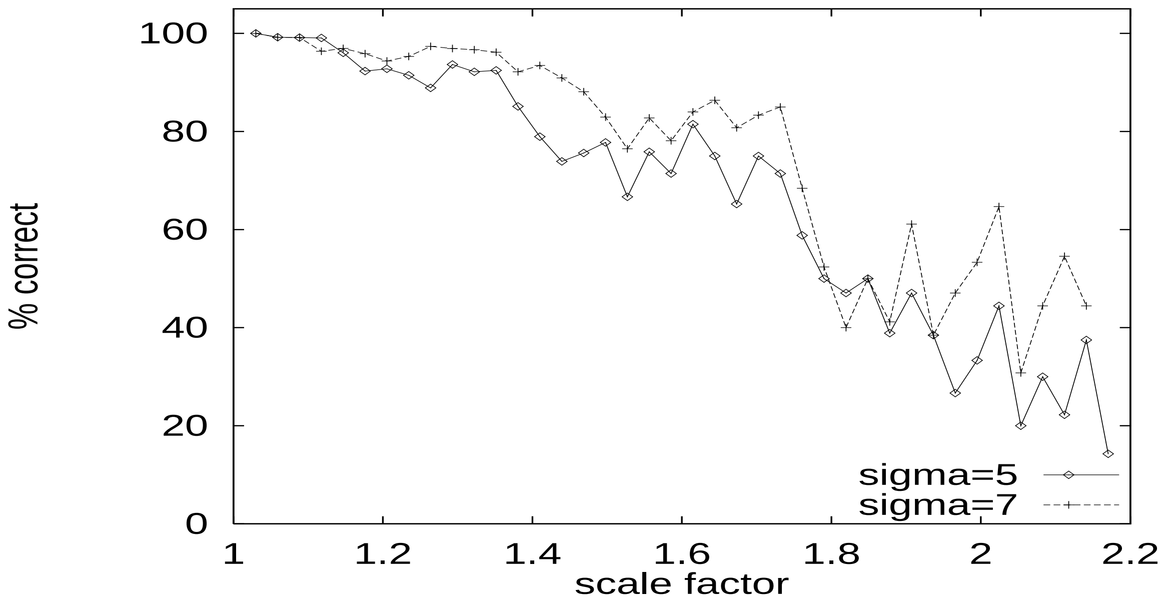

Figure 18: Percentage of correct matches with respect to the scale factor

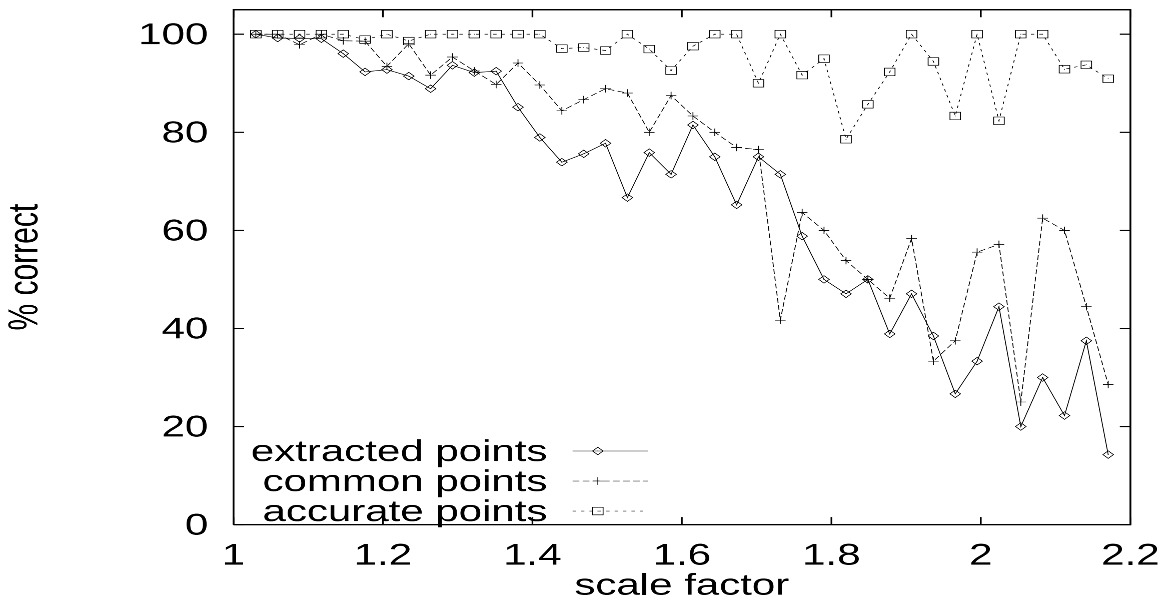

Figure 19: Evaluation of the influence of the detector; percentage of correct matches with respect to the scale factor, the size σ of the Gaussian is 5

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399