Evaluation of Interest Point Detectors

Cordelia Schmid, Roger Mohr and Christian Bauckhage

INRIA Rhône-Alpes

655 av. de l'Europe

38330 Montbonnot, France

Cordelia.Schmid@inrialpes.fr

Keywords

Interest Points, Quantitative Evaluation, Comparison of Detectors, Repeatability, Information Content

Abstract

Many different low-level feature detectors exist and it is widely agreed that the evaluation of detectors is important. In this paper we introduce two evaluation criteria for interest points: repeatability rate and information content. Repeatability rate evaluates the geometric stability under different transformations. Information content measures the distinctiveness of features. Different interest point detectors are compared using these two criteria. We determine which detector gives the best results and show that it satisfies the criteria well.

1 Introduction

Many computer vision tasks rely on low-level features. A wide variety of feature detectors exist, and results can vary enormously depending on the detector used. It is widely agreed that evaluation of feature detectors is important [36]. Existing evaluation methods use ground-truth verification [5], visual inspection [20, 27], localization accuracy [3, 6, 8, 22], theoretical analysis [11, 15, 40] or specific tasks [45, 46].

In this paper we introduce two novel criteria for evaluating interest points : repeatability and information content. Those two criteria directly measure the quality of the feature for tasks like image matching, object recognition and 3D reconstruction. They apply to any type of scene, and they do not rely on any specific feature model or high-level interpretation of the feature. Our criteria are more general than most existing evaluation methods (cf. section 1.1). They are complementary to localization accuracy which is relevant for tasks like camera calibration and 3D reconstruction of specific scene points. This criterion has previously been evaluated for interest point detectors [3, 6, 8, 22].

Repeatability explicitly compares the geometrical stability of the detected interest points between different images of a given scene taken under varying viewing conditions. Previous methods have evaluated detectors for individual images only. An interest point is "repeated", if the 3D scene point detected in the first image is also accurately detected in the second one. The repeatability rate is the percentage of the total observed points that are detected in both images. Note that repeatability and localization are conflicting criteria - smoothing improves repeatability but degrades localization [7].

Information content is a measure of the distinctiveness of an interest point. Distinctiveness is based on the likelihood of a local greyvalue descriptor computed at the point within the population of all observed interest point descriptors. Descriptors characterize the local shape of the image at the interest points. The entropy of these descriptors measures the information content of a set of interest points.

In this paper several detectors are compared using these two evaluation criteria. The best detector satisfies both of these criteria well, which explains its success for tasks such as image matching based on interest points and correlation [49]. In this context at least a subset of the points have to be repeated in order to allow feature correspondence. Furthermore, if image-based measures (e.g. correlation) are used to compare points, interest points should have distinctive patterns.

1.1 Related work on the evaluation of feature detectors

The evaluation of feature detectors has concentrated on edges. Only a few authors have evaluated interest point detectors. In the following we give a few examples of existing edge evaluation methods and a survey of previous work on evaluating interest points. Existing methods can be categorized into methods based on: ground-truth verification, visual inspection, localization accuracy, theoretical analysis and specific tasks.

1.1.1 Ground-truth verification

Methods based on ground-truth verification determine the undetected features and the false positives. Ground-truth is in general created by a human. It relies on his symbolic interpretation of the image and is therefore subjective. Furthermore, human interpretation limits the complexity of the images used for evaluation.

For example, Bowyer et al [5] use human marked ground-truth to evaluate edge detectors. Their evaluation criterion is the number of false positives with respect to the number of unmatched edges which is measured for varying input parameters. They used structured outdoor scenes, such as airports and buildings.

1.1.2 Visual inspection

Methods using visual inspection are even more subjective as they are directly dependent on the human evaluating the results. López et al [27] define a set of visual criteria to evaluate the quality of detection. They visually compare a set of ridge and valley detectors in the context of medical images. Heath et al [20] evaluate detectors using a visual rating score which indicates the perceived quality of the edges for identifying an object. This score is measured by a group of people. Different edge detectors are evaluated on real images of complex scenes.

1.1.3 Localization accuracy

Localization accuracy is the criterion most often used to evaluate interest points. It measures whether an interest point is accurately located at a specific 2D location. This criterion is significant for tasks like camera calibration and the 3D reconstruction of specific scene points. Evaluation requires the knowledge of precise 3D properties, which restricts the evaluation to simple scenes.

Localization accuracy is often measured by verifying that a set of 2D image points is coherent with the known set of corresponding 3D scene points. For example Coehlo et al [8] compare the localization accuracy of interest point detectors using different planar projective invariants for which reference values are computed using scene measurements. The scene contains simple black polygons and is imaged from different viewing angles. A similar evaluation scheme is used by Heyden and Rohr [22]. They extract sets of points from images of polyhedral objects and use projective invariants to compute a manifold of constraints on the points.

Brand and Mohr [6] measure the localization accuracy of a model-based L-corner detector. They use four different criteria : alignment of the extracted points, accuracy of the 3D reconstruction, accuracy of the epipolar geometry and stability of the cross-ratio. Their scene is again very simple : it contains black squares on a white background.

To evaluate the accuracy of edge point detectors, Baker and Nayar [3] propose four global criteria : collinearity, intersection at a single point, parallelism and localization on an ellipse. Each of the criteria corresponds to a particular (very simple) scene and is measured using the extracted edgels. Their experiments are conducted under widely varying image conditions (illumination change and 3D rotation).

As mentioned by Heyden and Rohr, methods based on projective invariants don't require the knowledge of the exact position of the features in the image. This is an advantage of such methods, as the location of features in an image depends both on the intrinsic parameters of the camera and the relative position and orientation of the object with respect to the camera and is therefore difficult to determine. However, the disadvantage of such methods is the lack of comparison to true data which may introduce a systematic bias.

1.1.4 Theoretical analysis

Methods based on a theoretical analysis examine the behavior of the detectors for theoretical feature models. Such methods are limited, as they only apply to very specific features. For example, Deriche and Giraudon [15] study analytically the behavior of three interest point detectors using a L-corner model. Their study allows them to correct the localization bias. Rohr [40] performs a similar analysis for L-corners with aperture angles in the range of 0 and 180 degrees. His analysis evaluates 10 different detectors.

Demigny and Kamlé [11] use three criteria to theoretically evaluate four step edge detectors. Their criteria are based on Canny's criteria which are adapted to the discrete domain using signal processing theory. Canny's criteria are good detection, good localization and low responses multiplicity.

1.1.5 A specific task

A few methods have evaluated detectors through specific tasks. They consider that feature detection is not a final result by itself, but merely an input for further processing. Therefore, the true performance criterion is how well it prepares the input for the next algorithm. This is no doubt true to some extent. However, evaluations based on a specific task and system are hard to generalize and hence rather limited.

Shin et al [45] compare edge detectors using an object recognition algorithm. Test images are cars imaged under different lighting conditions and in front of varying backgrounds. In Shin et al [46] the performance of a edge-based structure from motion algorithm is used for evaluation. Results are given for two simple 3D scenes which are imaged under varying 3D rotations.

1.2 Overview of the paper

Section 2 presents a state of the art on interest point detectors as well as implementation details for the detectors used in our comparison. Section 3 defines the repeatability criterion, explains how to determine it experimentally and presents the results of a comparison under different transformations. Section 4 describes the information content criterion and evaluates results for different detectors. In section 5 we select the detector which gives the best results according to the two criteria, show that the quality of its results is very high and discuss possible extensions.

2 Interest point detectors

By "interest point" we simply mean any point in the image for which the signal changes two-dimensionally. Conventional "corners" such as L-corners, T-junctions and Y-junctions satisfy this, but so do black dots on white backgrounds, the endings of branches and any location with significant 2D texture. We will use the general term "interest point" unless a more specific type of point is referred to. Figure 1 shows an example of general interest points detected on Van Gogh's sower painting.

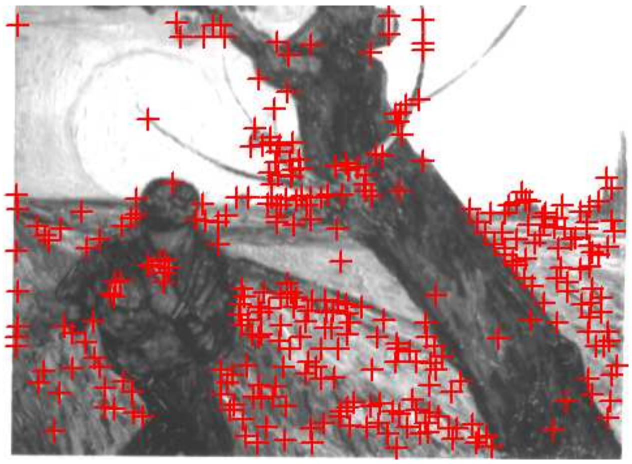

Figure 1: Interest points detected on Van Gogh's sower painting. The detector is an improved version of the Harris detector. There are 317 points detected.

2.1 State of the art

A wide variety of interest point and corner detectors exist in the literature. They can be divided into three categories : contour based, intensity based and parametric model based methods. Contour based methods first extract contours and then search for maximal curvature or inflexion points along the contour chains, or do some polygonal approximation and then search for intersection points. Intensity based methods compute a measure that indicates the presence of an interest point directly from the greyvalues. Parametric model methods fit a parametric intensity model to the signal. They often provide sub-pixel accuracy, but are limited to specific types of interest points, for example to L-corners. In the following we briefly present detection methods for each of the three categories.

2.1.1 Contour based methods

Contour based methods have existed for a long time ; some of the more recent ones are presented. Asada and Brady [1] extract interest points for 2D objects from planar curves. They observe that these curves have special characteristics : the changes in curvature. These changes are classified in several categories : junctions, endings etc. To achieve robust detection, their algorithm is integrated in a multi-scale framework. A similar approach has been developed by Mokhtarian and Mackworth [29]. They use inflexion points of a planar curve.

Medioni and Yasumoto [28] use B-splines to approximate the contours. Interest points are maxima of curvature which are computed from the coefficients of these B-splines.

Horaud et al [23] extract line segments from the image contours. These segments are grouped and intersections of grouped line segments are used as interest points.

Shilat et al [44] first detect ridges and troughs in the images. Interest points are high curvature points along ridges or troughs, or intersection points. They argue that such points are more appropriate for tracking, as they are less likely to lie on the occluding contours of an object.

Mokhtarian and Suomela [30] describe an interest point detector based on two sets of interest points. One set are T-junctions extracted from edge intersections. A second set is obtained using a multi-scale framework : interest points are curvature maxima of contours at a coarse level and are tracked locally up to the finest level. The two sets are compared and close interest points are merged.

The algorithm of Pikaz and Dinstein [37] is based on a decomposition of noisy digital curves into a minimal number of convex and concave sections. The location of each separation point is optimized, yielding the minimal possible distance between the smoothed approximation and the original curve. The detection of the interest points is based on properties of pairs of sections that are determined in an adaptive manner, rather than on properties of single points that are based on a fixed-size neighborhood.

2.1.2 Intensity based methods

Moravec [31] developed one of the first signal based interest point detectors. His detector is based on the auto-correlation function of the signal. It measures the greyvalue differences between a window and windows shifted in several directions. Four discrete shifts in directions parallel to the rows and columns of the image are used. If the minimum of these four differences is superior to a threshold, an interest point is detected.

The detector of Beaudet [4] uses the second derivatives of the signal for computing the measure "DET": $DET = I_{xx}I_{yy} - I_{xy}^2$ where $I(x,y)$ is the intensity surface of the image. DET is the determinant of the Hessian matrix and is related to the Gaussian curvature of the signal. This measure is invariant to rotation. Points where this measure is maximal are interest points.

Kitchen and Rosenfeld [24] present an interest point detector which uses the curvature of planar curves. They look for curvature maxima on isophotes of the signal. However, due to image noise an isophote can have an important curvature without corresponding to an interest point, for example on a region with almost uniform greyvalues. Therefore, the curvature is multiplied by the gradient magnitude of the image where non-maximum suppression is applied to the gradient magnitude before multiplication. Their measure is $K = \frac{I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y}{I_x^2 + I_y^2}$.

Dreschler and Nagel [16] first determine locations of local extrema of the determinant of the Hessian "DET". A location of maximum positive DET can be matched with a location of extreme negative DET, if the directions of the principal curvatures which have opposite sign are approximatively aligned. The interest point is located between these two points at the zero crossing of DET. Nagel [32] shows that the Dreschler-Nagel's approach and Kitchen-Rosenfeld's approach are identical.

Several interest point detectors [17, 18, 19, 48] are based on a matrix related to the auto-correlation function. This matrix $\mathtt{A}$ averages derivatives of the signal in a window W around a point (x, y) :

$$\mathtt{A}(x,y) = \begin{bmatrix} \sum\limits_{(x_k,y_k)\in W} (I_x(x_k,y_k))^2 & \sum\limits_{(x_k,y_k)\in W} I_x(x_k,y_k) I_y(x_k,y_k) \\ \sum\limits_{(x_k,y_k)\in W} I_x(x_k,y_k) I_y(x_k,y_k) & \sum\limits_{(x_k,y_k)\in W} (I_y(x_k,y_k))^2 \end{bmatrix} \tag{1}$$

where $I(x, y)$ is the image function and (x_k, y_k) are the points in the window W around (x, y).

This matrix captures the structure of the neighborhood. If this matrix is of rank two, that is both of its eigenvalues are large, an interest point is detected. A matrix of rank one indicates an edge and a matrix of rank zero a homogeneous region. The relation between this matrix and the auto-correlation function is given in appendix A.

Harris [19] improves the approach of Moravec by using the auto-correlation matrix $\mathtt{A}$. The use of discrete directions and discrete shifts is thus avoided. Instead of using a simple sum, a Gaussian is used to weight the derivatives inside the window. Interest points are detected if the auto-correlation matrix $\mathtt{A}$ has two significant eigenvalues.

Förstner and Gülch [18] propose a two step procedure for localizing interest points. First points are detected by searching for optimal windows using the auto-correlation matrix $\mathtt{A}$. This detection yields systematic localization errors, for example in the case of L-corners. A second step based on a differential edge intersection approach improves the localization accuracy.

Förstner [17] uses the auto-correlation matrix $\mathtt{A}$ to classify image pixels into categories - region, contour and interest point. Interest points are further classified into junctions or circular features by analyzing the local gradient field. This analysis is also used to determine the interest point location. Local statistics allow a blind estimate of signal-dependent noise variance for automatic selection of thresholds and image restoration.

Tomasi and Kanade [48] motivate their approach in the context of tracking. A good feature is defined as one that can be tracked well. They show that such a feature is present if the eigenvalues of matrix $\mathtt{A}$ are significant.

Heitger et al [21] develop an approach inspired by experiments on the biological visual system. They extract 1D directional characteristics by convolving the image with orientation-selective Gabor like filters. In order to obtain 2D characteristics, they compute the first and second derivatives of the 1D characteristics.

Cooper and al [9] first measure the contour direction locally and then compute image differences along the contour direction. A knowledge of the noise characteristics is used to determine whether the image differences along the contour direction are sufficient to indicate an interest point. Early jump-out tests allow a fast computation of the image differences.

The detector of Reisfeld et al [38] uses the concept of symmetry. They compute a symmetry map which shows a "symmetry strength" for each pixel. This symmetry is computed locally by looking at the magnitude and the direction of the derivatives of neighboring points. Points with high symmetry are selected as interest points.

Smith and Brady [47] compare the brightness of each pixel in a circular mask to the center pixel to define an area that has a similar brightness to the center. Two dimensional features can be detected from the size, centroid and second moment of this area.

The approach proposed by Laganière [26] is based on a variant of the morphological closing operator which successively applies dilation/erosion with different structuring elements. Two closing operators and four structuring elements are used. The first closing operator is sensitive to vertical/horizontal L-corners and the second to diagonal L-corners.

2.1.3 Parametric model based methods

The parametric model used by Rohr [39] is an analytic junction model convolved with a Gaussian. The parameters of the model are adjusted by a minimization method, such that the template is closest to the observed signal. In the case of a L-corner the parameters of the model are the angle of the L-corner, the angle between the symmetry axis of the L-corner and the x-axis, the greyvalues, the position of the point and the amount of blur. Positions obtained by this method are very precise. However, the quality of the approximation depends on the initial position estimation. Rohr uses an interest point detector which maximizes $\det(\mathtt{A})$ (cf. equation (1)) as well as the intersection of line segments to determine the initial values for the model parameters.

Deriche and Blaszka [14] develop an acceleration of Rohr's method. They substitute an exponential for the Gaussian smoothing function. They also show that to assure convergence the image region has to be quite large. In cluttered images the region is likely to contain several signals, which makes convergence difficult.

Baker et al [2] propose an algorithm that automatically constructs a detector for an arbitrary parametric feature. Each feature is represented as a densely sampled parametric manifold in a low dimensional subspace. A feature is detected, if the projection of the surrounding intensity values in the subspace lies sufficiently close to the feature manifold. Furthermore, during detection the parameters of detected features are recovered using the closest point on the feature manifold.

Parida et al [34] describe a method for general junction detection. A deformable template is used to detect radial partitions. The minimum description length principle determines the optimal number of partitions that best describes the signal.

2.2 Implementation details

This section presents implementation details for the detectors included in our comparison. The detectors are Harris [19], an improved version of Harris, Cottier[10], Horaud [23], Heitger [21] and Förstner [17]. Except in the case of the improved version of Harris, we have used the implementations of the original authors, with the standard default parameter values recommended by the authors for general purpose feature detection. These values are seldom optimal for any given image, but they do well on average on collections of different images. Our goal is to evaluate detectors for such collections.

The standard Harris detector [19] ("Harris") computes the derivatives of the matrix $\mathtt{A}$ (cf. equation (1)) by convolution with the mask [-2 -1 0 1 2]. A Gaussian ($\sigma = 2$) is used to weight the derivatives summed over the window. To avoid the extraction of the eigenvalues of the matrix $\mathtt{A}$, the strength of an interest points is measured by $\det(\mathtt{A}) - \alpha \operatorname{trace}(\mathtt{A})^2$. The second term is used to eliminate contour points with a strong eigenvalue, α is set to 0.06. Non-maximum suppression using a 3x3 mask is then applied to the interest point strength and a threshold is used to select interest points. The threshold is set to 1% of the maximum observed interest point strength.

In the improved version of Harris ("ImpHarris"), derivatives are computed more precisely by replacing the [-2 -1 0 1 2] mask with derivatives of a Gaussian ($\sigma = 1$). A recursive implementation of the Gaussian filters [13] guarantees fast detection.

Cottier [10] applies the Harris detector only to contour points in the image. Derivatives for contour extraction as well as for the Harris detector are computed by convolution with the Canny/Deriche operator [12] ($\alpha = 2$, $\omega = 0.001$). Local maxima detection with hysteresis thresholding is used to extract contours. High and low thresholds are determined from the gradient magnitude (high = average gradient magnitude, low = 0.1 * high). For the Harris detector derivatives are averaged over two different window sizes in order to increase localization accuracy. Points are first detected using a 5x5 window. The exact location is then determined by using a 3x3 window and searching the maximum in the neighborhood of the detected point.

Horaud [23] first extracts contour chains using his implementation of the Canny edge detector. Tangent discontinuities in the chain are located using a worm, and a line fit between the discontinuities is estimated using orthogonal regression. Lines are then grouped and intersections between neighboring lines are used as interest points.

Heitger [21] convolves the image with even and odd symmetrical orientation-selective filters. These Gabor like filters are parameterized by the width of the Gaussian envelope (σ =5), the sweep which increases the relative weight of the negative side-lobes of even filters and the orientation selectivity which defines the sharpness of the orientation tuning. Even and odd filters are computed for 6 orientations. For each orientation an energy map is computed by combining even and odd filter outputs. 2D signal variations are then determined by differentiating each energy map along the respective orientation using "end-stopped operators". Non-maximum suppression (3x3 mask) is applied to the combined end-stopped operator activity and a relative threshold (0.1) is used to select interest points.

The Förstner detector [17] computes the derivatives on the smoothed image ($\sigma = 0.7$). The derivatives are then summed over a Gaussian window ($\sigma = 2$) to obtain the auto-correlation matrix A. The trace of this matrix is used to classify pixels into region or non-region. For homogeneous regions the trace follows approximatively a χ^2-distribution. This allows to determine the classification threshold automatically using a significance level ($\alpha = 0.95$) and the estimated noise variance. Pixels are further classified into contour or interest point using the ratio of the eigenvalues and a fixed threshold (0.3). Interest point locations are then determined by minimizing a function of the local gradient field. The parameter of this function is the size of the Gaussian which is used to compute a weighted sum over the local gradient measures ($\sigma = 4$).

3 Repeatability

3.1 Repeatability criterion

Repeatability signifies that detection is independent of changes in the imaging conditions, i.e. the parameters of the camera, its position relative to the scene, and the illumination conditions. 3D points detected in one image should also be detected at approximately corresponding positions in subsequent ones (cf. figure 2). Given a 3D point X and two projection matrices P_1 and P_i, the projections of X into images I_1 and I_i are $x_1 = P_1 X$ and $x_i = P_i X$. A point x_1 detected in image I_1 is repeated in image I_i if the corresponding point x_i is detected in image I_i. To measure the repeatability, a unique relation between x_1 and x_i has to be established. This is difficult for general 3D scenes, but in the case of a planar scene this relation is defined by a homography [42] :

$$x_i = H_{1i} x_1 \quad \text{where } H_{1i} = P_i P_1^{-1}$$

P_1^{-1} is an abusive notation to represent the back-projection of image I_1. In the case of a planar scene this back-projection exists.

The repeatability rate is defined as the number of points repeated between two images with respect to the total number of detected points. To measure the number of repeated points, we have to take into account that the observed scene parts differ in the presence of changed imaging conditions, such as image rotation or scale change. Interest points which can not be observed in

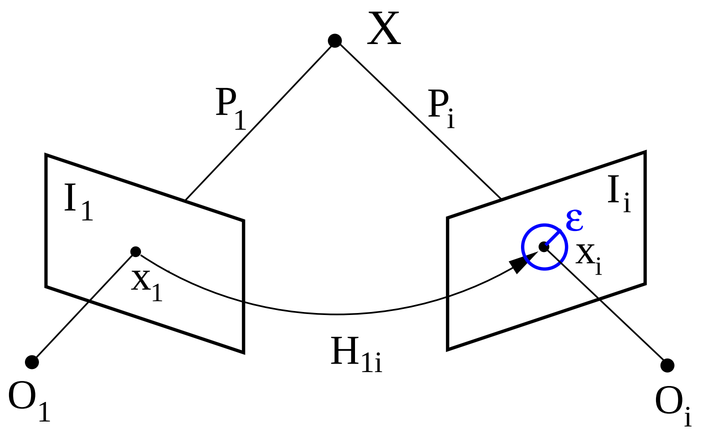

Figure 2: The points x_1 and x_i are the projections of 3D point X into images I_1 and I_i : $x_1 = P_1X$ and $x_i = P_iX$ where P_1 and P_i are the projection matrices. A detected point x_1 is repeated if x_i is detected. It is ϵ-repeated if a point is detected in the ϵ-neighborhood of x_i. In case of planar scenes the points x_1 and x_i are related by the homography H_{1i}.

both images corrupt the repeatability measure. Therefore only points which lie in the common scene part are used to compute the repeatability. This common scene part is determined by the homography. Points $\tilde{x}_1$ and $\tilde{x}_i$ which lie in the common part of images I_1 and I_i are defined by:

$$\{\tilde{x}_1\} = \{x_1 \mid H_{1i}x_1 \in I_i\} \text{ and } \{\tilde{x}_i\} = \{x_i \mid H_{i1}x_i \in I_1\}$$

where $\{x_1\}$ and $\{x_i\}$ are the points detected in images I_1 and I_i respectively. H_{ij} is the homography between images I_i and I_j.

Furthermore, the repeatability measure has to take into account the uncertainty of detection. A repeated point is in general not detected exactly at position x_i, but rather in some neighborhood of x_i. The size of this neighborhood is denoted by ϵ (cf. figure 2) and repeatability within this neighborhood is called *ϵ-repeatability*. The set of point pairs $(\tilde{x}_1, \tilde{x}_i)$ which correspond within an ϵ-neighborhood is defined by:

$$R_i(\epsilon) = \{(\tilde{x}_1, \tilde{x}_i) \mid dist(H_{1i}\tilde{x}_1, \tilde{x}_i) < \epsilon\}$$

The number of detected points may be different for the two images. In the case of a scale change, for example, more interest points are detected on the high resolution image. Only the minimum number of interest points (the number of interest points of the coarse image) can be repeated. The repeatability rate $r_i(\epsilon)$ for image I_i is thus defined by:

$$r_i(\epsilon) = \frac{|R_i(\epsilon)|}{\min(n_1, n_i)}$$

where $n_1 = |\{\tilde{x}_1\}|$ and $n_i = |\{\tilde{x}_i\}|$ are the number of points detected in the common part of images I_1 and I_i respectively. We can easily verify that $0 \leq r_i(\epsilon) \leq 1$.

The repeatability criterion, as defined above, is only valid for planar scenes. Only for such scenes the geometric relation between two images is completely defined. However, the restriction to planar scenes is not a limitation, as additional interest points detected on 3D scenes are due to occlusions and shadows. These points are due to real changes of the observed scene and the repeatability criterion should not take into account such unreliable points [43, 44].

3.2 Experimental conditions

Sequences The repeatability rates of several interest point detectors are compared under different transformations: image rotation, scale change, illumination variation and viewpoint change. We consider both uniform and complex illumination variation. Stability to image noise has also been tested. The scene is always static and we move either the camera or the light source.

We will illustrate results for two planar scenes: "Van Gogh" and "Asterix". The "Van Gogh" scene is the sower painting shown in figure 1. The "Asterix" scene can be seen in figure 3. The two scenes are very different: the "Van Gogh" scene is highly textured whereas the "Asterix" scene is mostly line drawings.

Estimating the homography To ensure an accurate repeatability rate, the computation of the homography has to be precise and independent of the detected points. An independent, subpixel localization is required. We therefore take a second image for each position of the camera, with black dots projected onto the scene (cf. figure 3). The dots are extracted very precisely by fitting a template, and their centers are used to compute the homography. A least median square method makes the computation robust.

While recording the sequence, the scene and the projection mechanism (overhead projector)

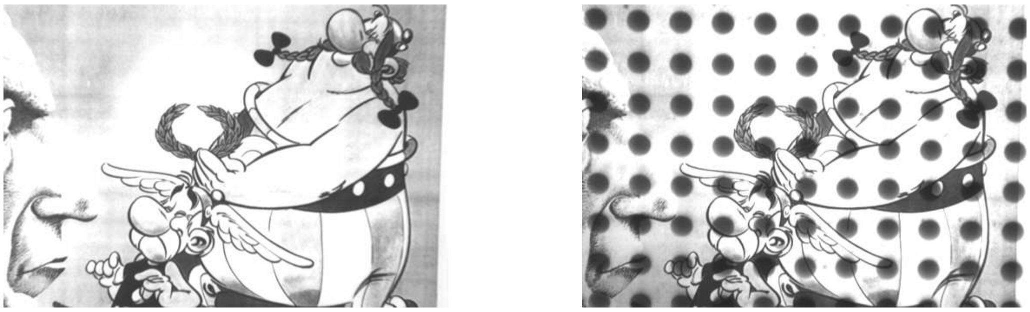

Figure 3: Two images of the "Asterix" scene. On the left the image of the scene and on the right the image with black dots projected.

remain fixed. Only the camera or the light source move. The projection mechanism is displayed in figure 4.

Figure 4: Projection mechanism. An overhead projector casts black dots on the scene.

3.3 Results for repeatability

We first compare the two versions of Harris (section 3.3.1). The one with better results is then included in the comparison of the detectors (section 3.3.2 - 3.3.6). Comparisons are presented for image rotation, scale change, illumination variation, change of viewpoint and camera noise. Results are presented for the "Van Gogh" scene; results for the "Asterix" scene are given in appendix B. For the images used in our experiments, we detect between 200 and 1200 interest points depending on the image and the detector used. The mean distance between a point and its closest neighbor is around 10 pixels. Measuring the repeatability rate with ϵ=1.5 or less, the probability that two points are accidentally within the error distance is very low.

3.3.1 Comparison of the two Harris versions

Figure 5 compares the two different versions of the Harris detector in the presence of image rotation (graph on the left) and scale change (graph on the right). The repeatability of the improved version of Harris (ImpHarris) is better in both cases. The results of the standard version vary with image rotation, the worst results being obtained for an angle of 45°. This is due to the fact that the standard version uses non-isotropic discrete derivative filters. A stable implementation of the derivatives significantly improves the repeatability of the Harris detector. The improved version (ImpHarris) is included in the comparison of the detectors in the following sections.

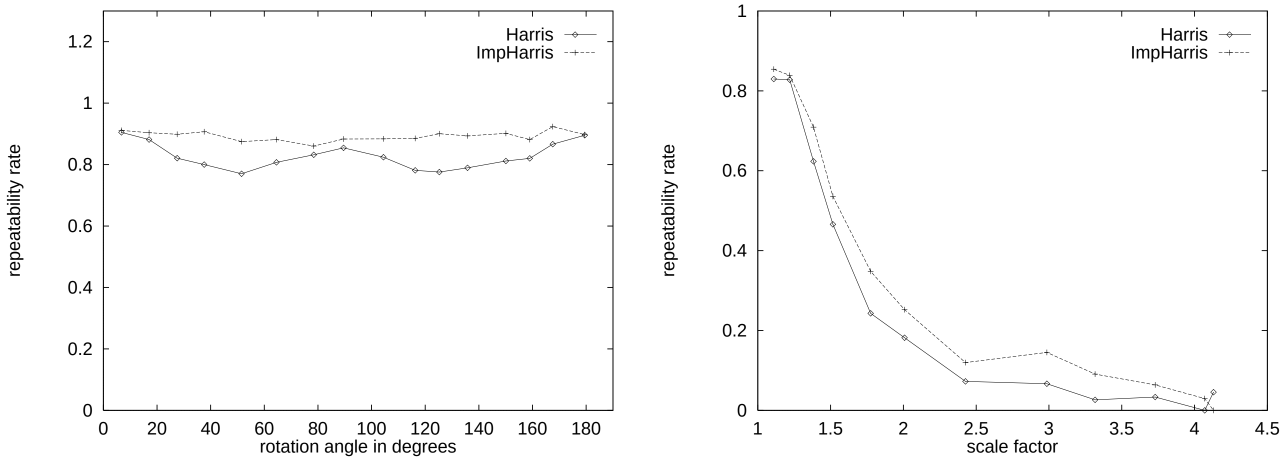

Figure 5: Comparison of Harris and ImpHarris. On the left the repeatability rate for an image rotation and on the right the rate for a scale change. $\epsilon = 1.5$.

3.3.2 Image rotation

In this section, we compare all detectors described in section 2.2 for an image rotation. Image rotations are obtained by rotating the camera around its optical axis using a special mechanism. Figure 6 shows three images of the rotation sequence. The left image is the reference image. The rotation angle for the image in the middle is 38° and for the image on the right 116°. The interest points detected for these three images using the improved version of Harris are displayed in figure 7.

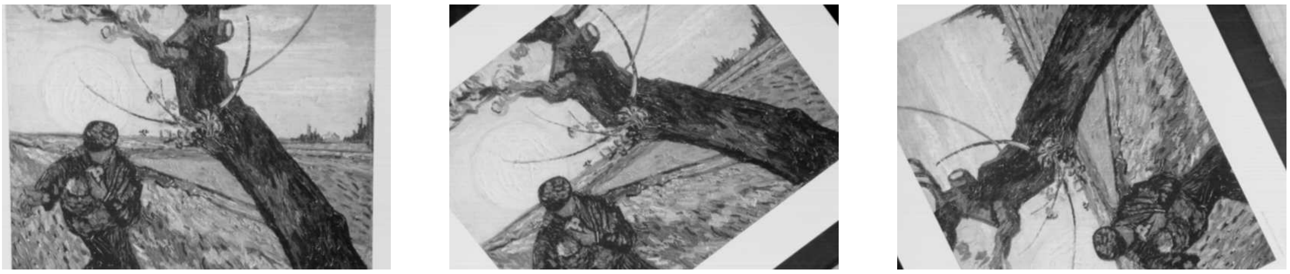

Figure 6: Image rotation sequence. The left image is the reference image. The rotation angle for the image in the middle is 38° and for the image on the right 116°.

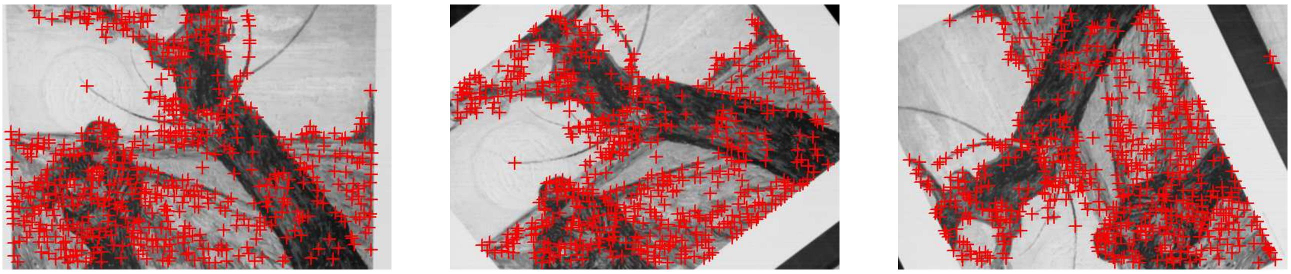

Figure 7: Interest points detected on the images of figure 6 using the improved version of Harris. There are 610, 595 and 586 points detected in the left, middle and right images, respectively. For a localization error ϵ of 1.5 pixels the repeatability rate between the left and middle images is 91% and between the left and right images 89%

The repeatability rate for this rotation sequence is displayed in figure 8. The rotation angles vary between 0° and 180°. The graph on the left displays the repeatability rate for a localization error ϵ of 0.5 pixels. The graph on the right shows the results for an error of 1.5 pixels, that is the detected point lies in the pixel neighborhood of the predicted point.

For both localization errors the improved version of Harris gives the best results; results are not dependent on image rotation. At $\epsilon = 1.5$ the repeatability rate of ImpHarris is almost 100%. Computing Harris only on the image contours (Cottier) makes the results worse. Results of the Heitger detector depend on the rotation angle, as it uses derivatives computed in several fixed directions. Its results are worse for rotation angles between 40° and 140°. The detector of

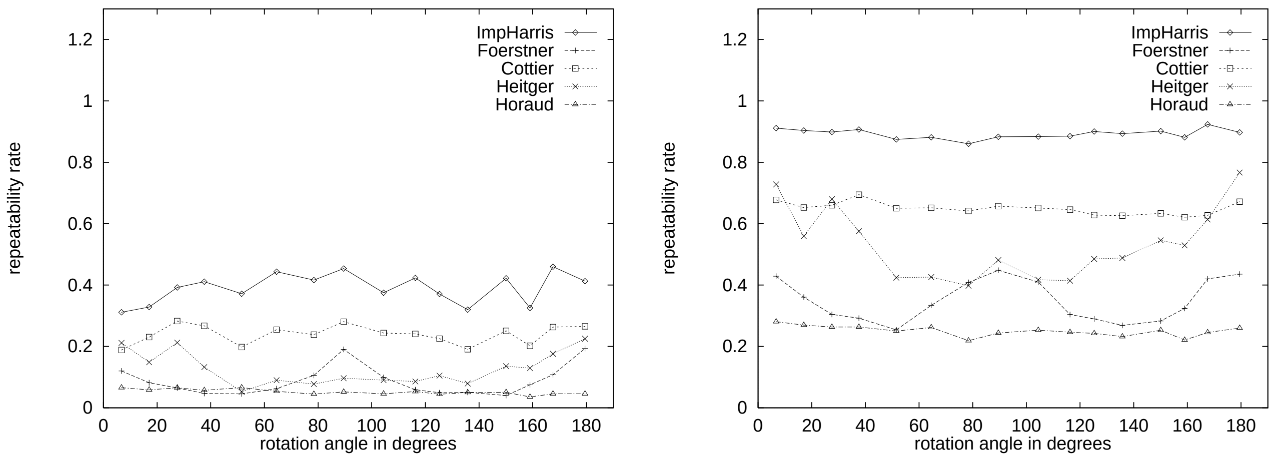

Figure 8: Repeatability rate for image rotation. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

Förstner gives bad results for rotations of 45°, probably owing to the use of anisotropic derivative filters. The worst results are obtained by the method based on the intersection of line segments (Horaud).

Figure 9 shows the repeatability rate as a function of the localization error ϵ for a constant rotation angle of 89 degrees. The localization error varies between 0.5 pixels and 5 pixels. When increasing the localization error, the results improve for all the detectors. However, the improved version of Harris detector is always best and increases most rapidly. For this detector, good results are obtained above a 1 pixel localization error.

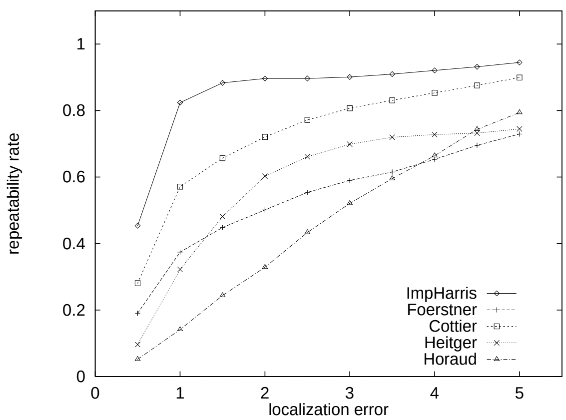

Figure 9: Repeatability rate as a function of the localization error ϵ. The rotation angle is 89°.

3.3.3 Scale change

Scale change is investigated by varying the focal length of the camera. Figure 10 shows three images of the scale change sequence. The left image is the reference image. The scale change for the middle one is 1.5 and for the right one 4.1. The scale factors have been determined by the ratios of the focal lengths. The interest points detected for these three images using the improved version of Harris are displayed in figure 11.

Figure 10: Scale change sequence. The left image is the reference image. The scale change for the middle one is 1.5 and for the right one 4.1.

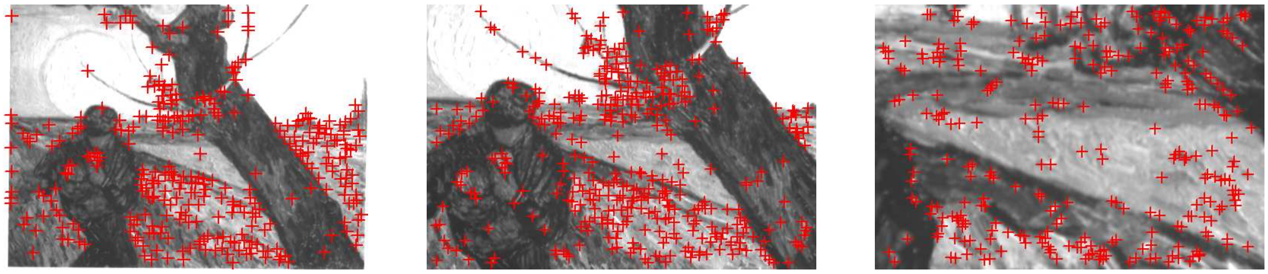

Figure 11: Interest points detected on the images of figure 10 using the improved version of Harris. There are 317, 399 and 300 points detected in the left, middle and right images, respectively. The repeatability rate between the left and middle images is 54% and between the left and right images 1% for a localization error ϵ of 1.5 pixels.

Figure 12 shows the repeatability rate for scale changes. The left graph shows the repeatability rate for an ϵ of 0.5 and the right one for an ϵ of 1.5 pixels. Evidently the detectors are very sensitive to scale changes. For an ϵ of 0.5 (1.5) repeatability is very poor for a scale factor above

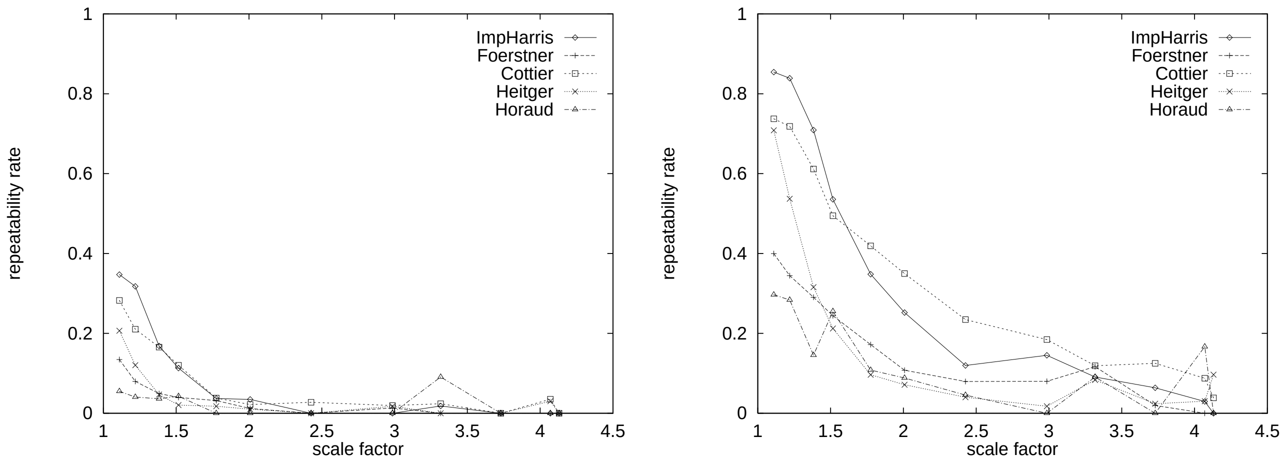

Figure 12: Repeatability rate for scale change. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

1.5 (2). The improved version of Harris and Cottier detectors give the best results. The results of the other detectors are hardly usable. Above a scale factor of about 2.5, the results are mainly due to artifacts. At larger scales many more points are found in the textured regions of the scene, so accidental correspondences are more likely.

Figure 13 shows the repeatability rate as a function of the localization error ϵ for a constant scale change of 1.5. Results improve in the presence of larger localization errors ; the repeatability rates of ImpHarris and Cottier increase more rapidly than those of the other detectors.

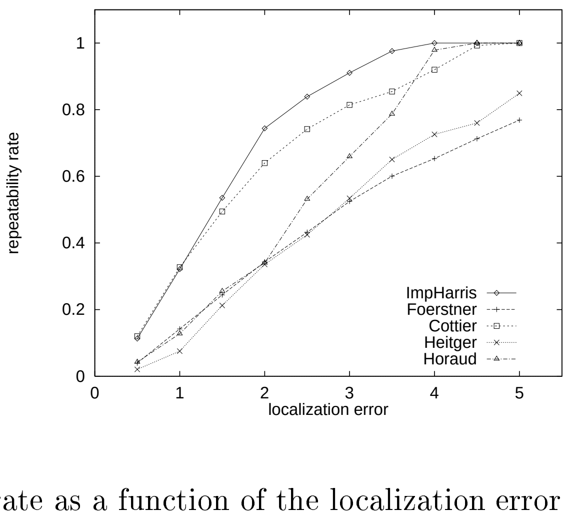

Figure 13: Repeatability rate as a function of the localization error ϵ. The scale change is 1.5.

3.3.4 Variation of illumination

Illumination can vary in many different ways. In the following we consider both a uniform variation of illumination and a more complex variation due to a change of light source position.

Uniform variation of illumination Uniform illumination variation is obtained by changing the camera aperture. The change is quantified by the "relative greyvalue" - the ratio of mean greyvalue of an image to that of the reference image which has medium illumination. Figure 14 shows three images of the sequence, a dark one with a relative greyvalue of 0.6, the reference image and a bright one with a relative greyvalue of 1.7.

Figure 14: Uniform illumination variation: from left to right images with relative greyvalue 0.6, 1 and 1.7.

Figure 15 displays the results for a uniform illumination variation. Even for a relative greyvalue of 1 there is not 100% repeatability due to image noise (two images of a relative greyvalue of 1 have been taken, one reference image and one test image). In both graphs, the repeatability decreases smoothly in proportion to the relative greyvalue. The improved version of Harris and Heitger obtain better results than the other detectors.

Complex variation of illumination A non-uniform illumination variation is obtained by moving the light source in an arc from approximately $-45°$ to $45°$. Figure 16 shows three images of the sequence. The light source is furthest right for the left image (image # 0). This image is the reference image for our evaluation. For the image in the middle the light source is in front of the scene (image # 6). Part of this image is saturated. The light source is furthest left for the right image (image # 11).

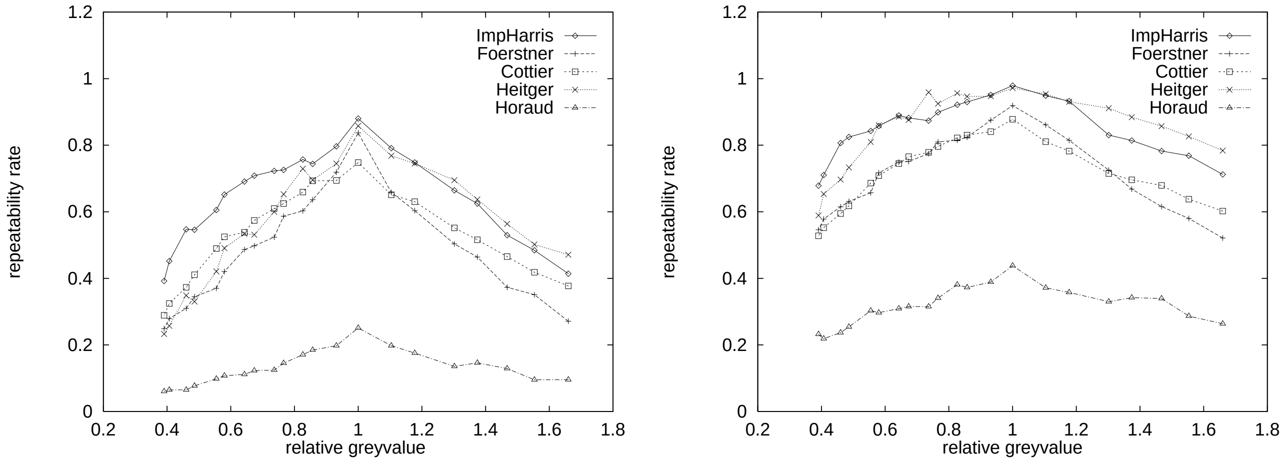

Figure 15: Repeatability rate for uniform illumination variation. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

Figure 16: Complex illumination variation. The reference image is on the left (image # 0), the light source is furthest right for this image. For the image in the middle the light source is in front of the scene (image # 6). For the one on the right the light source is furthest left (image # 11).

Figure 17 displays the repeatability results. The improved version of Harris obtains better results than the other detectors. For an ϵ of 0.5, results slightly decrease as the light position changes. For an ϵ of 1.5 results are not modified by a complex illumination variation. Illumination direction has little effect on the results as the interest point measures are computed locally.

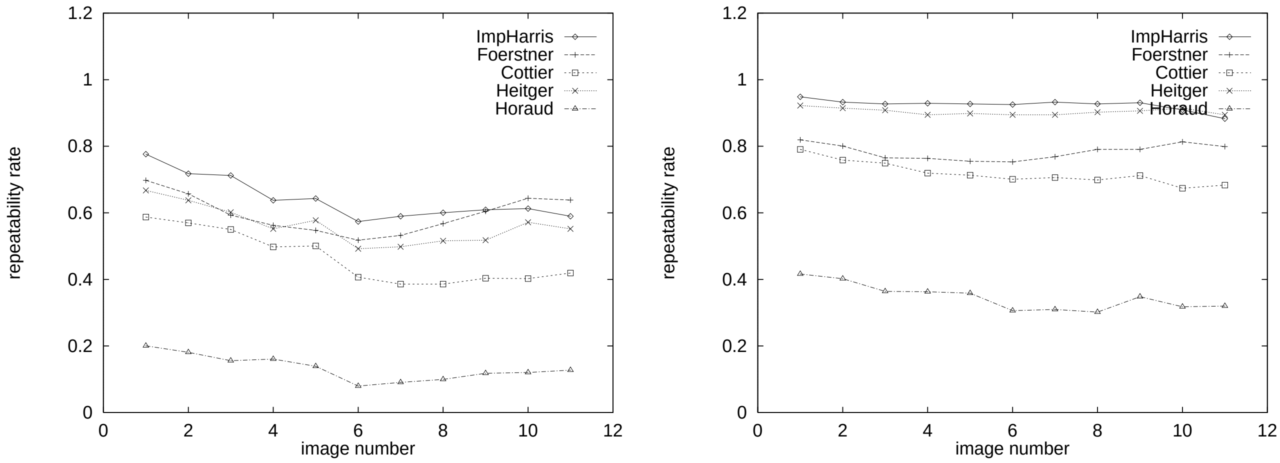

Figure 17: Repeatability rate for complex illumination variation. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.3.5 Viewpoint change

To measure repeatability in the presence of viewpoint changes, the position of the camera is moved in an arc around the scene. The angle varies from approximately $-50°$ to $50°$. The different viewpoints are approximately regularly spaced. Figure 18 shows three images of the sequence. The left image is taken at the rightmost position of the camera (image # 0). For the image in the middle the camera is in front of the painting (image # 7). This image is the reference image for our evaluation. For the right image the camera is at its leftmost position (image # 15).

Figure 19 displays the results for a viewpoint change. The improved version of Harris gives results superior to those of the other detectors. The results degrade rapidly for $\epsilon = 0.5$, but significantly more slowly for $\epsilon = 1.5$. For this ϵ the repeatability of ImpHarris is always above 60% except for image 0. The ImpHarris detector shows a good repeatability in the presence of perspective deformations.

Figure 18: Viewpoint change sequence. The left image is taken at the rightmost position of the camera (image # 0). For the middle image the camera is in front of the painting (image # 7). This image is the reference image for our evaluation. For the right image the camera is at its leftmost position (image # 15).

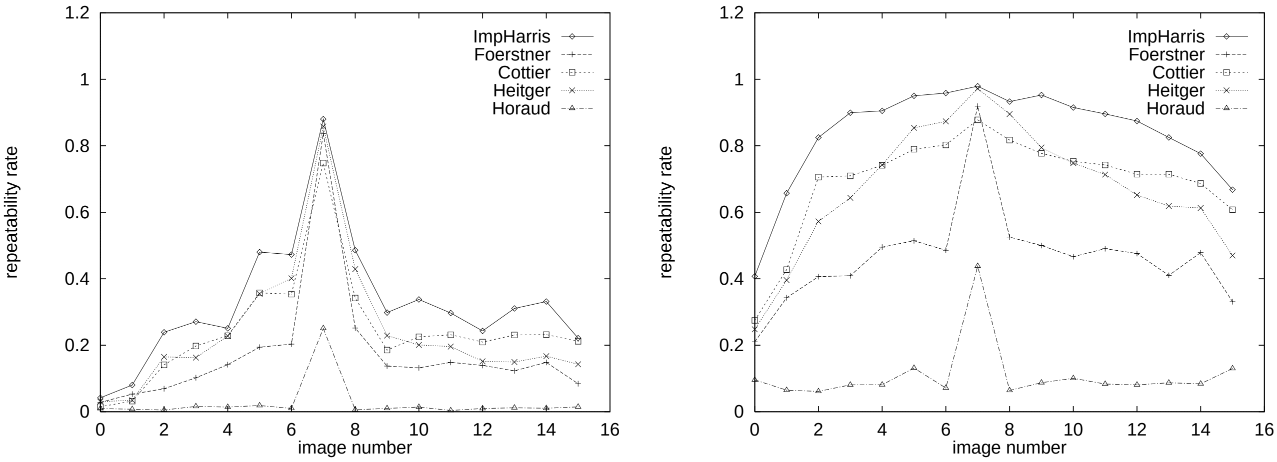

Figure 19: Repeatability rate for the viewpoint change. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.3.6 Camera noise

To study repeatability in the presence of image noise, a static scene has been recorded several times. The results of this experiment are displayed in figure 20. We can see that all detectors give good results except the Horaud one. The improved version of Harris gives the best results, followed closely by Heitger. For $\epsilon = 1.5$ these two detectors obtain a rate of nearly 100%.

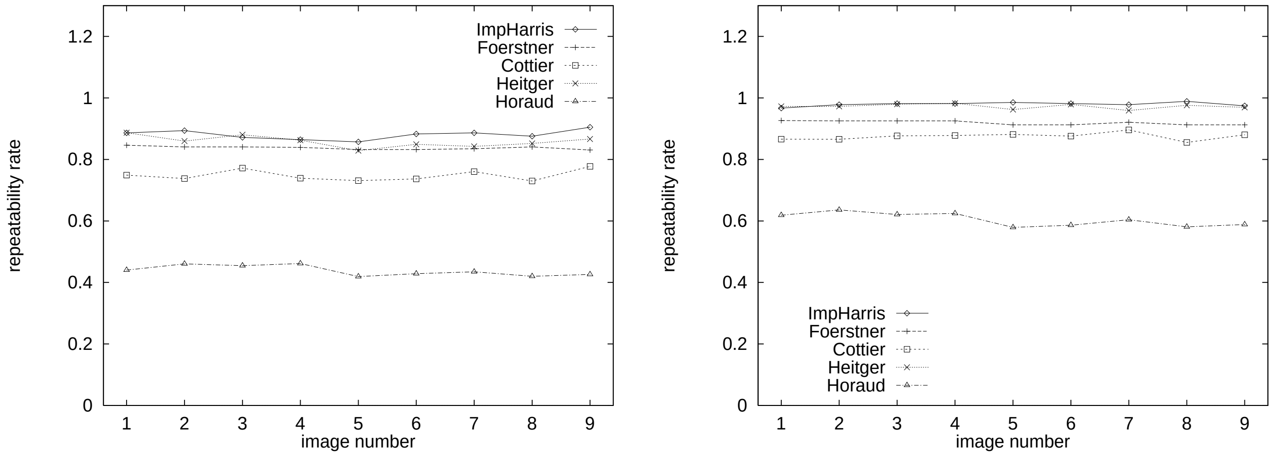

Figure 20: Repeatability rate for the camera noise. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.4 Conclusion for repeatability

Repeatability of various detectors has been evaluated in the presence of different imaging conditions : image rotation, scale change, variation of illumination, viewpoint change and noise of the imaging system. Two different scenes have been used : "Van Gogh" and "Asterix". The results for these two sequences are very similar ; the “Asterix” sequence (cf. appendix B) confirms the results presented above.

Results of the previous section show that a stable implementation of the derivatives improves the results of the standard Harris detector. The improved version of Harris (ImpHarris) gives significantly better results than the other detectors in the presence of image rotation. This is due to the rotation invariance of its image measures. The detector of Heitger combines computations in several directions and is not invariant to rotations. This is confirmed by Perona [35] who has noticed that the computation in several directions is less stable to an image rotation. ImpHarris and Cottier give the best results in the presence of scale changes. Moreover, the computation of these detectors is based on Gaussian filters and can easily be adapted to scale changes. In the case of illumination variations and camera noise, ImpHarris and Heitger obtain the best results. For a viewpoint change ImpHarris shows results which are superior to the other detectors.

In all cases the results of the improved version of the Harris detector are better or equivalent to those of the other detectors. For this detector, interest points are largely independent of the imaging conditions; points are geometrically stable.

4 Information content

4.1 Information content criterion

Information content is a measure of the distinctiveness of an interest point. Distinctiveness is based on the likelihood of a local greyvalue descriptor computed at the point within the population of all observed interest point descriptors. Given one or several images, a descriptor is computed for each of the detected interest points. Information content measures the distribution of these descriptors. If all descriptors lie close together, they don't convey any information, that is the information content is low. Matching for example fails, as any point can be matched to any other. On the other hand if the descriptors are spread out, information content is high and matching is likely to succeed.

Information content of the descriptors is measured using entropy. The more spread out the descriptors are, the higher is the entropy. Section 4.2 presents a short introduction to entropy and shows that entropy measures the average information content. In section 4.3 we introduce the descriptors used for our evaluation, which characterize local greyvalue patterns. Section 4.4 describes how to partition the set of descriptors. Partitioning of the descriptors is necessary to compute the entropy, as will be explained in section 4.2. The information content criterion of different detectors is compared in section 4.5.

4.2 Entropy

Entropy measures the randomness of a variable. The more random a variable is the bigger the entropy. In the following we are not going to deal with continuous variables, but with partitions [33]. The entropy of a partition $\mathcal{A} = \{A_i\}$ is :

$$H(\mathcal{A}) = -\sum_i p_i \log(p_i) \tag{2}$$

where p_i is the probability of $\mathcal{A}_i$.

Note that the size of the partition influences the results. If $\mathcal{B}$ is a new partition formed by subdivisions of the sets of $\mathcal{A}$ then $H(\mathcal{B}) \geq H(\mathcal{A})$.

Entropy measures average information content. In information theory the information content I of a message i is defined as

$$I_i = log(1/p_i) = -log(p_i)$$

The information content of a message is inversely related to its probability. If $p_i = 1$ the event always occurs and no information is attributed to it: $I = 0$. The average information content per message of a set of messages is then defined by $-\sum_i p_i \log(p_i)$ which is its entropy.

In the case of interest points we would like to know how much average information content an interest point "transmits", as measured by its descriptor. The more distinctive the descriptors are, the larger is the average information content.

4.3 Descriptors characterizing local shape

To measure the distribution of local greyvalue patterns at interest points, we have to define a measure which describes such patterns. Collecting unordered pixel values at an interest point does not represent the shape of the signal around the pixel. Collecting ordered pixel values (e.g. from left to right and from top to bottom) respects the shape but is not invariant to rotation. We have therefore chosen to use local rotation invariants.

The rotation invariants used are combinations of greyvalue derivatives. Greyvalue derivatives are computed stably by convolution with Gaussian derivatives. This set of derivatives is called the "local jet" [25]. Note that derivatives up to Nth order describe the intensity function locally up to that order. The "local jet" of order N at a point $\mathbf{x} = (x, y)$ for image I and scale σ is defined by:

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

where $L_{i_1 \ldots i_n}(\mathbf{x}, \sigma)$ is the convolution of image I with the Gaussian derivatives $G_{i_1 \ldots i_n}(\mathbf{x}, \sigma)$ and $i_k \in \{x, y\}$.

To obtain invariance under the group $SO(2)$ (2D image rotations), Koenderink [25] and Romeny [41] compute differential invariants from the local jet. In our work invariants up to second order are used:

$$\vec{\mathcal{V}}[0..3] = \begin{bmatrix} L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2 L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2 L_{xy} L_{xy} + L_{yy} L_{yy} \end{bmatrix} \tag{3}$$

The average luminance does not characterize the shape and is therefore not included. Note that the first component of $\vec{\mathcal{V}}$ is the square of the gradient magnitude and the third is the Laplacian.

4.4 Partitioning a set of descriptors

The computation of entropy requires the partitioning of the descriptors $\vec{\mathcal{V}}$. Partitioning is dependent on the distance measure between descriptors. The distance between two descriptors $\vec{\mathcal{V}}_1$ and $\vec{\mathcal{V}}_2$ is given by the Mahalanobis distance:

$$d_M(\vec{\mathcal{V}}_2, \vec{\mathcal{V}}_1) = \sqrt{(\vec{\mathcal{V}}_2 - \vec{\mathcal{V}}_1)^T \Lambda^{-1} (\vec{\mathcal{V}}_2 - \vec{\mathcal{V}}_1)}$$

The covariance matrix Λ takes into account the variability of the descriptors $\vec{\mathcal{V}}$, i.e. their uncertainty due to noise. This matrix Λ is symmetric positive definite. Its inverse can be decomposed into $\Lambda^{-1} = P^T D P$ where D is diagonal and P an orthogonal matrix representing a change of reference frame. We can then define the square root of Λ^{-1} as $\Lambda^{-1/2} = D^{1/2} P$ where $D^{1/2}$ is a diagonal matrix whose coefficients are the square roots of the coefficients of D. The Mahalanobis distance can then be rewritten as:

$$d_M(\vec{\mathcal{V}}_2, \vec{\mathcal{V}}_1) = \|D^{1/2}P(\vec{\mathcal{V}}_2 - \vec{\mathcal{V}}_1)\|$$

The distance d_M is the norm of the difference of the normalized vectors:

$$\vec{\mathcal{V}}_{norm} = D^{1/2}P\vec{\mathcal{V}} \tag{4}$$

Normalization allows us to use equally sized cells in all dimensions. This is important since the entropy is directly dependent on the partition used. The probability of each cell of this partition is used to compute the entropy of a set of vectors $\vec{\mathcal{V}}$.

4.5 Results for information content

In this section, we compute the information content of the detectors which are included in our comparison. To obtain a statistically significant measure, a large number of points has to be considered. We use a set of 1000 images of different types: aerial images, images of paintings and images of toy objects. The information content of a detector is computed as follows:

1. Extract interest points for the set of images.
2. Compute descriptors (cf. equation (3)) for all extracted interest points (σ=3).
3. Normalize each descriptor (cf. equation (4)). The covariance matrix takes into account the variability of the descriptors.
4. Partition the set of normalized descriptors. The cell size is the same in all dimensions, it is set to 20.
5. Determine the probability of each cell and compute the entropy with equation (2).

The results are presented in table 1. It shows that the improved version of Harris produces the highest entropy, and hence the most distinctive points. The results obtained for Heitger are almost as good. The two detectors based on line extraction obtain worse results. This can be explained by their limitation to contour lines which reduces the distinctiveness of their greyvalue descriptors and thus their entropy.

Random points are included in our comparison : for each image we compute the mean number m of interest points extracted by the different detectors. We then select m random points over the image using a spatially uniform distribution. Entropy is computed as specified above using this random point detector. The result for this detector ("random") is given in table 1. Unsurprisingly, the results obtained for all of the interest point detectors are significantly better than those for random points. The probability to produce a collision is $e^{-(3.3-6.05)} \approx 15.6$ times higher for Random than for Harris.

detector	information content
ImpHarris	6.049526
Heitger	5.940877
Horaud	5.433776
Cottier	4.846409
Förstner	4.523368
Random	3.300863

Table 1: The information content for different detectors.

5 Conclusion

In this paper we have introduced two novel evaluation criteria : repeatability and information content. These two criteria present several advantages over existing ones. First of all, they are significant for a large number of computer vision tasks. Repeatability compares interest points detected on images taken under varying viewing conditions and is therefore significant for any interest point based algorithm which uses two or more images of a given scene. Examples are image matching, geometric hashing, computation of the epipolar geometry etc.

Information content is relevant for algorithms which use greyvalue information. Examples are image matching based on correlation and object recognition based on local feature vectors. Furthermore, repeatability as well as information content are independent of human intervention and apply to real scenes.

The two criteria have been used to evaluate and compare several interest point detectors. Repeatability was evaluated under various different imaging conditions. In all cases the improved version of Harris is better than or equivalent to those of the other detectors. Except for large scale changes, its points are geometrically stable under all tested image variations. The results for information content again show that the improved version of Harris obtains the best results, although the Heitger detector is a close second. All of the detectors have significantly higher information content than randomly selected points, so they do manage to select "interesting" points.

The criteria defined in this paper allow the quantitative evaluation of new interest point detectors. One possible extension is to adapt these criteria to other low-level features. Another extension would be to design an improved interest point detector with respect to the two evaluation criteria. Concerning repeatability, we have seen that detectors show rapid degradation in the presence of scale change. To solve this problem, the detectors could be included in a multi-scale framework. Another solution might be to estimate the scale at which the best results are obtained. Concerning information content, we think that studying which kinds of greyvalue descriptors occur frequently and which ones are rare will help us to design a detector with even higher information content.

Acknowledgments

We are very grateful to Yves Dufournaud, Radu Horaud and Andrew Zisserman for discussions.

A Derivation of the auto-correlation matrix

The local auto-correlation function measures the local changes of the signal. This measure is obtained by correlating a patch with its neighbouring patches, that is with patches shifted by a small amount in different directions. In the case of an interest point, the auto-correlation function is high for all shift directions.

Given a shift $(\Delta x, \Delta y)$ and a point (x, y), the auto-correlation function is defined as:

$$f(x,y) = \sum_{(x_k,y_k)\in W} (I(x_k,y_k) - I(x_k+\Delta x, y_k+\Delta y))^2 \tag{5}$$

where (x_k, y_k) are the points in the window W centered on (x, y) and I the image function.

If we want to use this function to detect interest points we have to integrate over all shift directions. Integration over discrete shift directions can be avoided by using the auto-correlation matrix. This matrix is derived using a first-order approximation based on the Taylor expansion:

$$I(x_k+\Delta x, y_k+\Delta y) \approx I(x_k,y_k) + \begin{pmatrix} I_x(x_k,y_k) & I_y(x_k,y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \tag{6}$$

Substituting the above approximation (6) into equation (5), we obtain:

$$\begin{aligned} f(x,y) &= \sum_{(x_k,y_k)\in W} \left(\begin{pmatrix} I_x(x_k,y_k) & I_y(x_k,y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2 \\ &= \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \begin{bmatrix} \sum_{(x_k,y_k)\in W} (I_x(x_k,y_k))^2 & \sum_{(x_k,y_k)\in W} I_x(x_k,y_k) I_y(x_k,y_k) \\ \sum_{(x_k,y_k)\in W} I_x(x_k,y_k) I_y(x_k,y_k) & \sum_{(x_k,y_k)\in W} (I_y(x_k,y_k))^2 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \\ &= \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \mathtt{A}(x,y) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \end{aligned} \tag{7}$$

The above equation (7) shows that the auto-correlation function can be approximated by the matrix $\mathtt{A}(x, y)$. This matrix $\mathtt{A}$ captures the structure of the local neighborhood.

B Repeatability results for the "Asterix" scene

In this appendix the repeatability results for the "Asterix" scene are presented. Experimental conditions are the same as described in section 3.

B.1 Comparison of the two Harris versions

Figure 21 compares the two different versions of the Harris detector in the presence of image rotation (graph on the left) and scale change (graph on the right). The repeatability of the improved version of Harris version is better in both cases. Results are comparable to those obtained for the "Van Gogh" scene (cf. section 3.3.1).

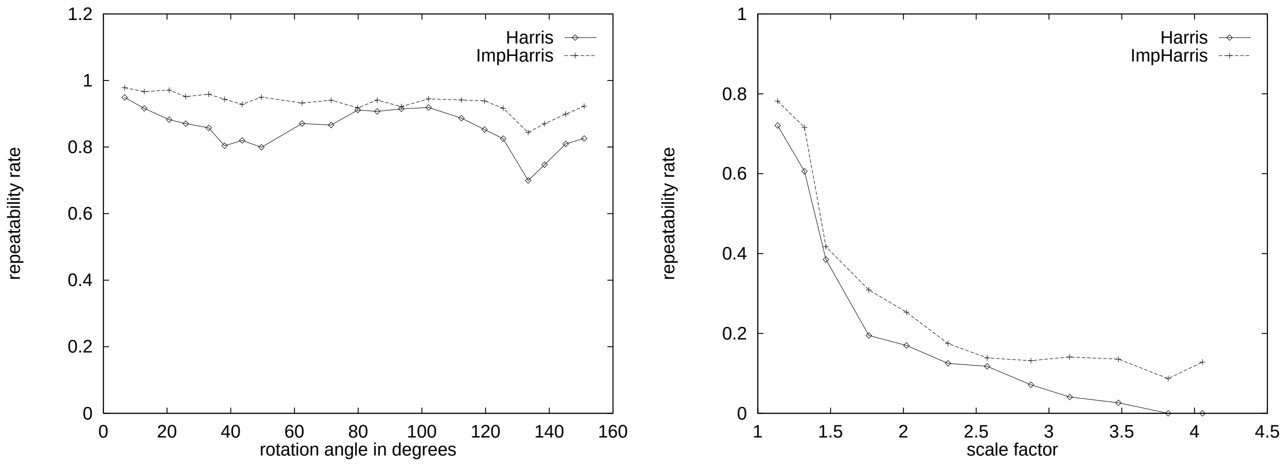

Figure 21: Comparison of Harris and ImpHarris. On the left repeatability rate for an image rotation and on the right the rate for a scale change. $\epsilon = 1.5$.

B.2 Image rotation

Figure 22 shows two images of the rotation sequence. The repeatability rate for the rotation sequence is displayed in figure 23. The improved version of Harris gives the best results as in the case of the "Van Gogh" scene (cf. section 3.3.2). Figure 24 shows the repeatability rate as a function of the localization error ϵ for a constant rotation angle of 93°.

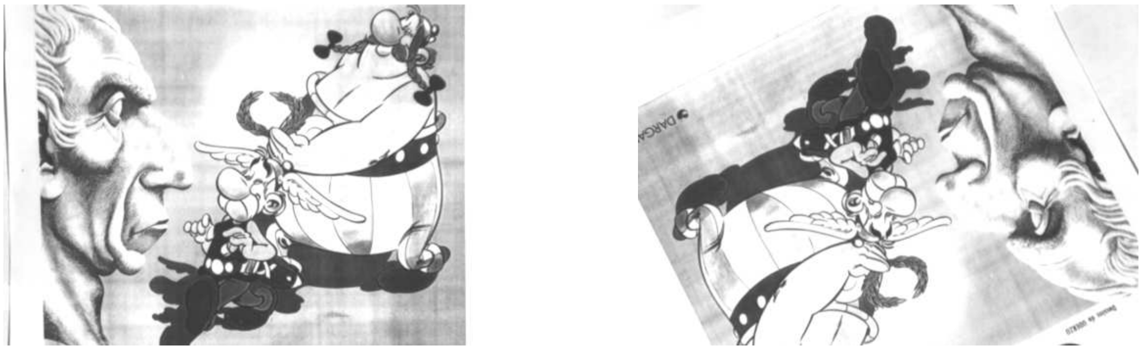

Figure 22: Image rotation sequence. On the left the reference image for the rotation sequence. On the right an image with a rotation angle of 154°.

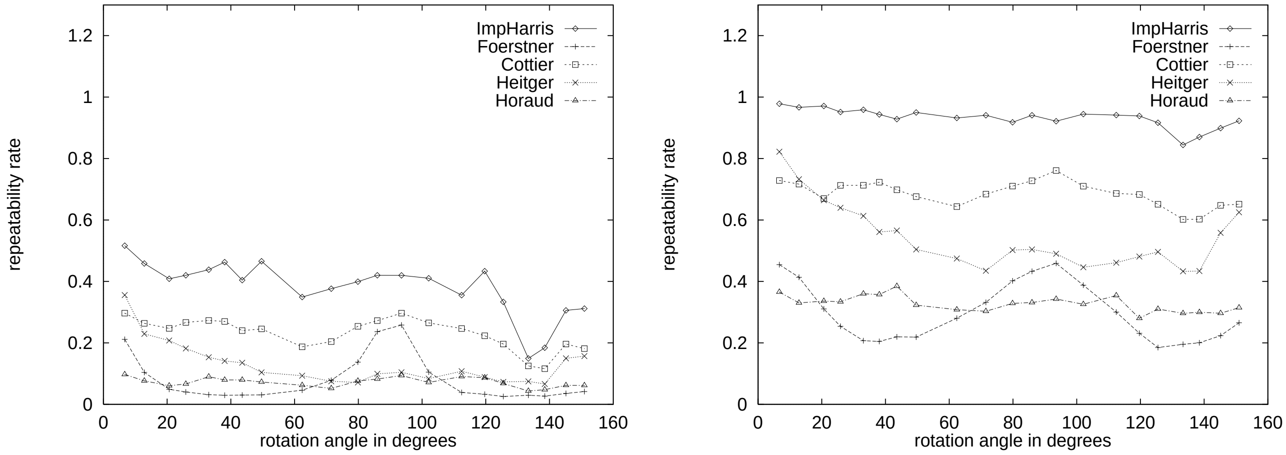

Figure 23: Repeatability rate for the sequence image rotation. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

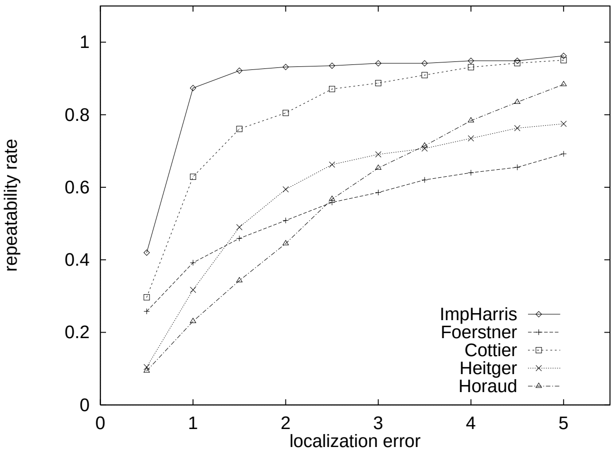

Figure 24: Repeatability rate as a function of the localization error ϵ. The rotation angle is 93°.

B.3 Scale change

Figure 25 shows two images of the scale change sequence. The scale factor between the two images is 4.1. The repeatability rate for the scale change sequence is displayed in figure 26. The improved version of Harris and the Cottier detector give the best results as in the case of the "Van Gogh" scene (cf. section 3.3.3). Figure 27 shows the repeatability rate as a function of the localization error ϵ for a constant scale change of 1.5.

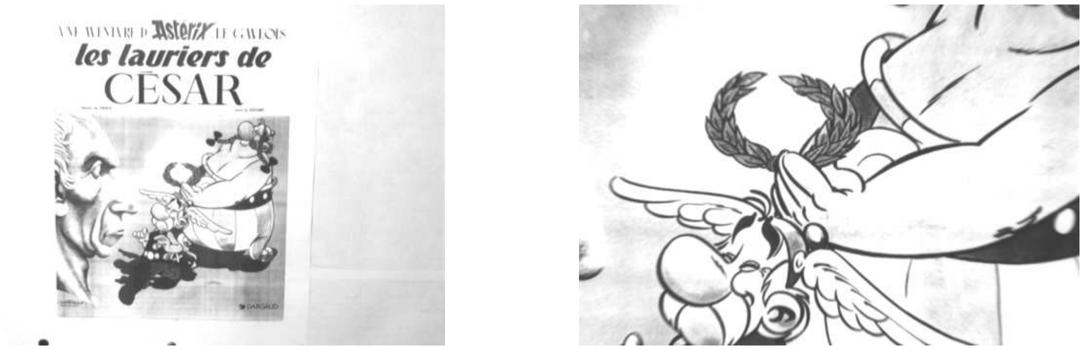

Figure 25: Scale change sequence. On the left the reference image for the scale change sequence. On the right an image with a scale change of a factor 4.1.

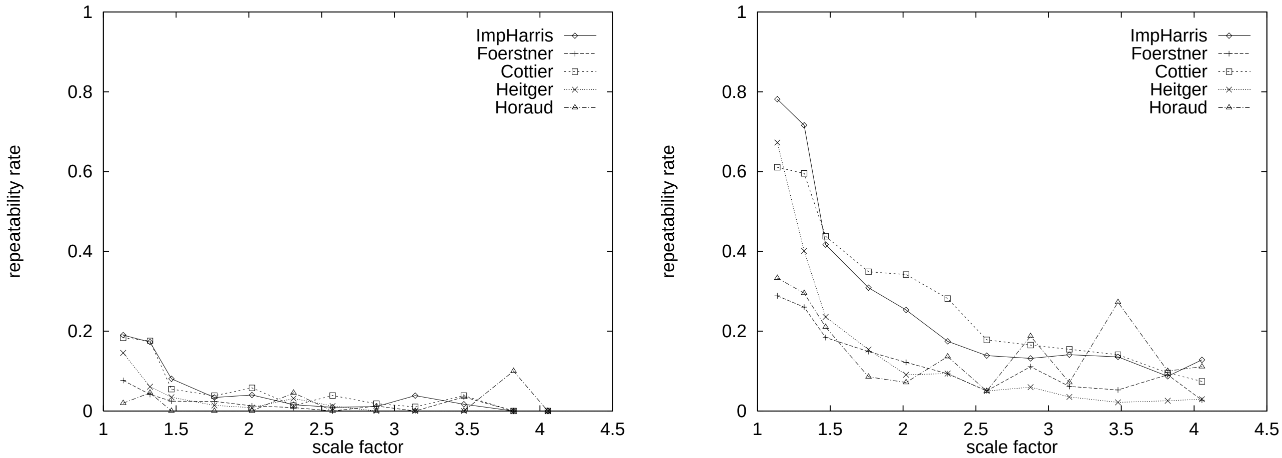

Figure 26: Repeatability rate for the sequence scale change. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

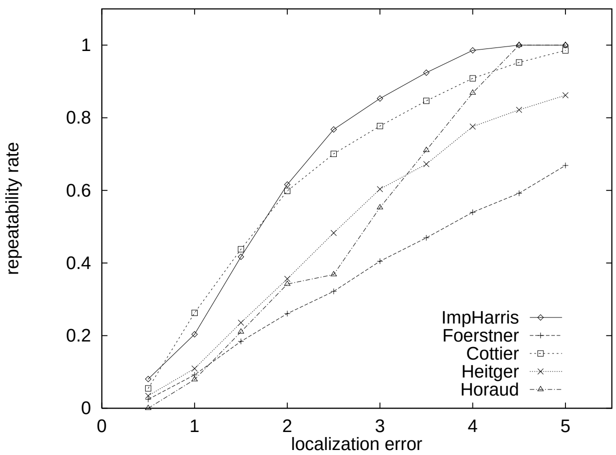

Figure 27: Repeatability rate as a function of the localization error. The scale change is 1.5.

B.4 Uniform variation of illumination

Figure 28 shows two images of the uniform variation of illumination sequence, a dark one with a relative greyvalue of 0.6 and a bright one with a relative greyvalue of 1.5. The repeatability rate for a uniform illumination variation is displayed in figure 29. ImpHarris and Heitger give the best results as in the case of the “Van Gogh” scene (cf. section 3.3.4).

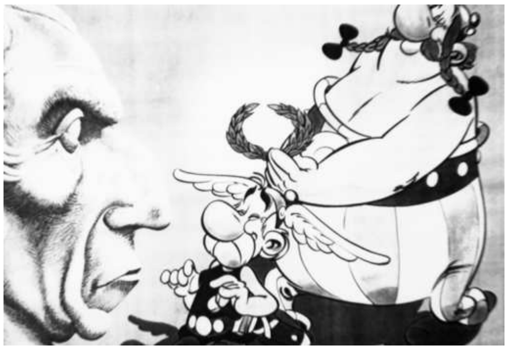

Figure 28: Uniform variation of illumination sequence. On the left an image with a relative greyvalue of 0.6. On the right an image with a relative greyvalue of 1.5.

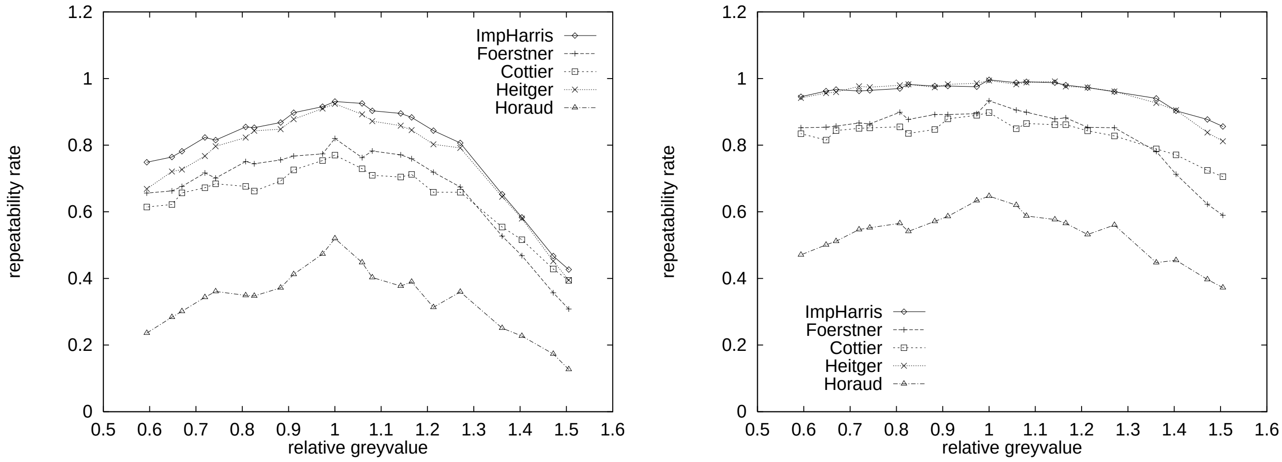

Figure 29: Repeatability rate for the sequence uniform variation of illumination. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

B.5 Camera noise

The repeatability rate for camera noise is displayed in figure 30. ImpHarris and Heitger give the best results.

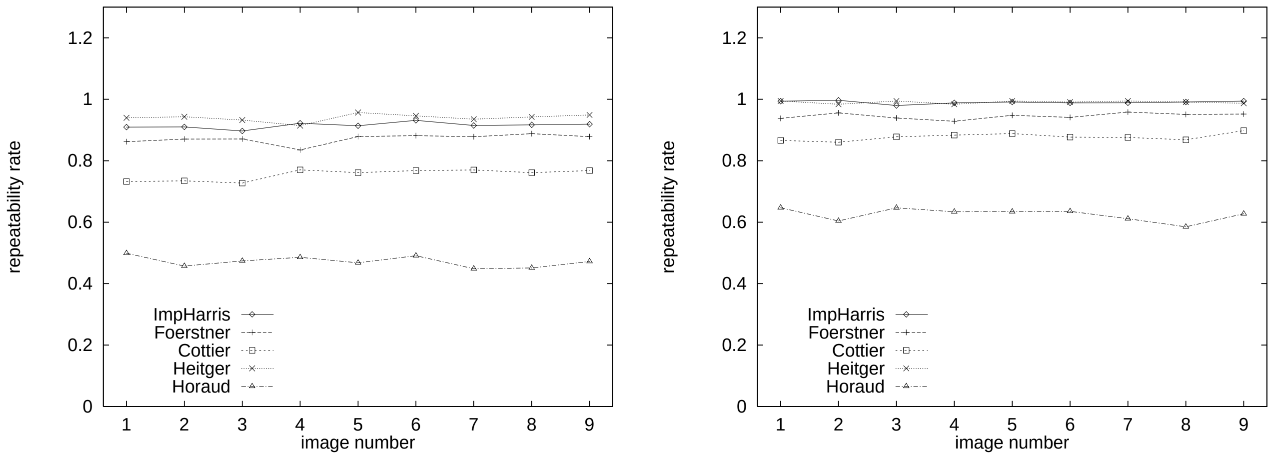

Figure 30: Repeatability rate for camera noise sequence. $\epsilon = 0.5$ for the graph on the left and $\epsilon = 1.5$ for the graph on the right.

References

[1] H. Asada and M. Brady. The curvature primal sketch. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(1):2–14, 1986.

[2] S. Baker, S. Nayar, and H. Murase. Parametric feature detection. *International Journal of Computer Vision*, 27(1):27–50, 1998.

[3] S. Baker and S. K. Nayar. Global measures of coherence for edge detector evaluation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 373–379, 1999.

[4] P.R. Beaudet. Rotationally invariant image operators. In *Proceedings of the 4th International Joint Conference on Pattern Recognition, Tokyo*, pages 579–583, 1978.

[5] K. Bowyer, C. Kranenburg, and S. Dougherty. Edge detector evaluation using empirical ROC curves. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 354–359, 1999.

[6] P. Brand and R. Mohr. Accuracy in image measure. In S.F. El-Hakim, editor, *Proceedings of the* SPIE *Conference on Videometrics III, Boston, Massachusetts, USA*, volume 2350, pages 218–228, 1994.

[7] J. Canny. A computational approach to edge detection. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[8] C. Coelho, A. Heller, J.L. Mundy, D. Forsyth, and A. Zisserman. An experimental evaluation of projective invariants. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Reykjavik, Iceland*, pages 273–293, 1991.

[9] J. Cooper, S. Venkatesh, and L. Kitchen. Early jump-out corner detectors. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 15(8):823–833, 1993.

[10] J.C. Cottier. Extraction et appariements robustes des points d'intérêt de deux images non étalonnées. Technical report, LIFIA–IMAG–INRIA Rhône-Alpes, 1994.

[11] D. Demigny and T. Kamlé. A discrete expression of Canny's criteria for step edge detector performances evaluation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1199–1211, 1997.

[12] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.

[13] R. Deriche. Recursively implementing the Gaussian and its derivatives. Technical report, INRIA, 1993.

[14] R. Deriche and T. Blaszka. Recovering and characterizing image features using an efficient model based approach. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 530–535, 1993.

[15] R. Deriche and G. Giraudon. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2):101–124, 1993.

[16] L. Dreschler and H.H. Nagel. Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene. *Computer Graphics and Image Processing*, 20:199–228, 1982.

[17] W. Förstner. A framework for low level feature extraction. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 383–394, 1994.

[18] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland*, pages 281–305, 1987.

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

[20] M. D. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer. A robust visual method for assessing the relative performance of edge-detection algorithms. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1338–1359, 1997.

[21] 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.

[22] A. Heyden and K. Rohr. Evaluation of corner extraction schemes using invariance methods. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 895–899, 1996.

[23] R. Horaud, T. Skordas, and F. Veillon. Finding geometric and relational structures in an image. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 374–384, 1990.

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

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

[26] R. Laganière. Morphological corner detection. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 280–285, 1998.

[27] A. M. López, F. Lumbreras, J. Serrat, and J. J. Villanueva. Evaluation of methods for ridge and valley detection. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 21(4):327–335, 1999.

[28] G. Medioni and Y. Yasumoto. Corner detection and curve representation using cubic B-splines. *Computer Vision, Graphics and Image Processing*, volume 39, pages 267–278, 1987.

[29] F. Mokhtarian and A. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(1):34–43, 1986.

[30] F. Mokhtarian and R. Suomela. Robust image corner detection through curvature scale space. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1376–1381, 1998.

[31] H.P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, USA*, page 584, 1977.

[32] H.H. Nagel. Displacement vectors derived from second order intensity variations in image sequences. *Computer Vision, Graphics and Image Processing*, 21:85–117, 1983.

[33] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw Hill, 1991.

[34] L. Parida, D. Geiger, and R. Hummel. Junctions : Detection, classification, and reconstruction. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 20(7):687–698, 1998.

[35] P. Perona. Deformable kernels for early vision. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(5):488–499, 1995.

[36] P. J. Phillips and K. W. Bowyer. Introduction to the special section on empirical evaluation of computer vision algorithms. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 21(4):289–290, 1999.

[37] A. Pikaz and I. Dinstein. Using simple decomposition for smoothing and feature point detection of noisy digital curves. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 16(8):808–813, 1994.

[38] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators : The generalized symmetry transform. *International Journal of Computer Vision*, 14:119–130, 1995.

[39] K. Rohr. Recognizing corners by fitting parametric models. *International Journal of Computer Vision*, 9(3):213–230, 1992.

[40] K. Rohr. Localization properties of direct corner detectors. *Journal of Mathematical Imaging and Vision*, 4(2):139–150, 1994.

[41] B.M 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.

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

[43] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 593–600, 1994.

[44] E. Shilat, M. Werman, and Y. Gdalyahu. Ridge's corner detection and correspondence. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 976–981, 1997.

[45] M. C. Shin, D. Goldgof, and K. Bowyer. Comparison of edge detectors using an object recognition task. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 360–365, 1999.

[46] M. C. Shin, D. Goldgof, and K. W. Bowyer. An objective comparison methodology of edge detection algorithms using a structure from motion task. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, USA*, pages 190–195, 1998.

[47] S. M. Smith and J. M. Brady. SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.

[48] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report CMU-CS-91-132, Carnegie Mellon University, 1991.

[49] 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. *Artificial Intelligence*, 78:87–119, 1995.