Improving Harris corner selection strategy

F. Bellavia D. Tegolo C. Valenti

Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Italy
E-mail: domenico.tegolo@unipa.it

Abstract: This study describes a corner selection strategy based on the Harris approach. Corners are usually defined as interest points for which intensity variation in the principal directions is locally maximised, as response from a filter given by the linear combination of the determinant and the trace of the autocorrelation matrix. The Harris corner detector, in its original definition, is only rotationally invariant, but scale-invariant and affine-covariant extensions have been developed. As one of the main drawbacks, corner detector performances are influenced by two user-given parameters: the linear combination coefficient and the response filter threshold. The main idea of the authors' approach is to search only the corners near enhanced edges and, by a z-score normalisation, to avoid the introduction of the linear combination coefficient. Combining these strategies allows a fine and stable corner selection without tuning the method. The new detector has been compared with other state-of-the-art detectors on the standard Oxford data set, achieving good results showing the validity of the approach. Analogous results have been obtained using the local detector evaluation framework on non-planar scenes by Fraundorfer and Bischof.

1 Introduction

Feature detection is a fundamental step in many image analysis applications. For example, three-dimensional reconstruction [1, 2], mosaicing [3] and object detection [4, 5] generally require locating interest points which are usually extracted by feature detectors. These detectors are strictly related to feature descriptors [6], effectively employed in high-level image tasks, which give a compact characterisation of the regions of interest. Cross-correlation among local patches can be applied as feature descriptors, whereas better results can be obtained by using steerable filters [7] or the popular scale-invariant feature transform descriptor (SIFT) [8]. A variety of detectors have been proposed in the literature, sensitive to different kinds of features and invariant to image transformations.

The definition of corner detector has been improperly used in the literature to enclose a more general class of feature detectors which not only detect corners, but rather define a more general kind of interest points. That general definition can be traced back to the work by Moravec [9], which has been improved through the autocorrelation matrix by Harris and Stephens [10], Förstner [11] and Shi and Tomasi [12]. Recently, Kenney *et al.* [13] have investigated some properties of these detectors under a more general setting, extending the image space dimensions. Although the Shi and Tomasi detector seems to be more suitable under this framework, the Harris corner detector is still widely used with good results in many computer vision tasks. Recent implementations introduced by Mikolajczyk and Schmid [14] are scale-invariant (called Harris-Laplace) and affine covariant (Harris-affine). In the latter case, viewpoint changes can be locally approximated by affine transformations [15].

The smallest univalue segment assimilating nucleus [16] and the features from accelerated segment test [17] detectors are based on a circular mask applied on the interest region. Junction detectors are also to be mentioned as a special case of corner detectors. Lindeberg [18] investigated the theoretical behaviour of the junctions at different scales, supported by experimental validation, while Parida *et al.* [19] modelled complex junctions that admit an effective dynamic programming solution.

Blob detectors have been extensively used in recent years, especially after the introduction of the SIFT detector by Lowe [8], which can be led back to an efficient approximation of the difference of Gaussians detector [20]. Blob detectors, based on the Hessian matrix, exist [21, 22], as well as scale-invariant and affine-covariant extensions, implemented in the Hessian-Laplace and Hessian-affine detectors [14], likewise, the Harris-based counterpart. The speeded up robust features [23] and the maximally stable extremal region (MSER) [24] detectors, respectively, based on wavelets and watershed, together with the entropy-based salient region detector [25], intensity-based region and the edge-based region detectors [26] have been successfully applied to computer vision tasks.

Feature detectors have been extensively evaluated in the literature [15, 27, 28] and although some detectors may perform better than others, it appears that results are not unquestionable since features closely depend on the input images. Moreover, it has been observed that, when applied to fully 3D objects, the performances degrade noticeably [27, 28]. The scale-invariant feature operator (SFOP) detector [29], which is a scale-space extension of the detector by Förstner, has been proposed to unify different types of features within the same framework by using the general spiral feature model of Bigün [30]. Dickscheid and

Förstner [31] reported that combining features by different detectors usually provides better results. Indeed, a variety of feature detectors can be put together in a sequential pipeline or in a collaborative fashion to improve the stability of the results, through voting strategies.

The underlying idea of the new Harris-based corner detector proposed here is the removal of a user-defined coefficient, by adapting the corner response function through the so called z-score, and the search for corners near edges by a coarse gradient mask. Given an observation scale, this method lets us obtain fine and stable results, in comparison with state-of-the-art corner detectors as the Harris-affine one.

In Section 2, an introduction to Harris-based detectors is given, together with concepts of the scale-space theory [32], which provides the background for the development of affine scale-invariant detectors. Section 3 introduces our detector, with a description of the algorithm and its implementation details. In Section 4, performances are evaluated in terms of repeatability index and matching score [15], using the Oxford benchmark data set [33] which has been extensively used for detector testing [15, 23], and thus it allows a comparison with other detectors. To validate our detector on non-planar scenes, the evaluation framework by Fraundorfer and Bischof [34] has been considered, too. In Section 5, the conclusion and final remarks are discussed.

2 Harris-based feature detectors

In this section, the standard Harris detector [10] is introduced. Basic concepts of the scale-space theory [32], which allow a generalisation of the Harris detector to affine space [35], are also presented.

2.1 Standard Harris detector

Given an image I, the autocorrelation matrix $\mu(\boldsymbol{p})$ of the point $\boldsymbol{p} \equiv (x, y)$ in the neighbourhood N of $\boldsymbol{p}$ is defined as

$$\mu(\boldsymbol{p}) = \begin{bmatrix} \sum_{\boldsymbol{q}\in N} I_x^2(\boldsymbol{q}) & \sum_{\boldsymbol{q}\in N} I_x(\boldsymbol{q})I_y(\boldsymbol{q}) \\ \sum_{\boldsymbol{q}\in N} I_x(\boldsymbol{q})I_y(\boldsymbol{q}) & \sum_{\boldsymbol{q}\in N} I_y^2(\boldsymbol{q}) \end{bmatrix} \tag{1}$$

where $I_x(\boldsymbol{p})$ and $I_y(\boldsymbol{p})$ are the partial derivatives, with respect to $I(\boldsymbol{p})$. The matrix $\mu(\boldsymbol{p})$ is symmetric, and thus real eigenvalues λ_1 and λ_2 exist, with $\lambda_1 \geq \lambda_2$.

These eigenvalues represent the intensity variation strength along the principal directions, given by the respective eigenvectors, in the neighbourhood N. Image regions can be classified according to the eigenvalues of the autocorrelation matrix as follows [10]:

- *plain regions:* $\lambda_1 \simeq 0$ and $\lambda_2 \simeq 0$. Both eigenvalues are small, and thus the intensity variation is negligible in every direction. Therefore the region can be considered almost plain;
- *edges:* $\lambda_1 \gg \lambda_2$. The difference between the eigenvalues is high, which implies that intensity variation is noticeable only in one direction, orthogonal to the edge;
- *corners:* $\lambda_1 \simeq \lambda_2$, with $\lambda_1, \lambda_2 \gg 0$. Intensity variation is strong along all directions. It is usually achieved by 'small spot' regions or junctions.

It can be noted that the product $\lambda_1\lambda_2$ and the sum $\lambda_1 + \lambda_2$ are sensitive to corners and to both edges and corners, respectively. It is also known from elementary algebra that the determinant and the trace of a general diagonalisable matrix agree, respectively, with the product and the sum of its eigenvalues. The above argumentation provides the justification to use the following function to measure the corner response

$$H(\boldsymbol{p}) = \det(\mu(\boldsymbol{p})) - \kappa \operatorname{tr}^2(\mu(\boldsymbol{p})) \tag{2}$$

where $\det(\mu(\boldsymbol{p}))$ and $\operatorname{tr}(\mu(\boldsymbol{p}))$ are, respectively, the determinant and the trace of the autocorrelation matrix (1). The coefficient κ is determined empirically and the values, usually adopted, range in [0.04, 0.16]. A point $\boldsymbol{p}$ is selected as a corner if $H(\boldsymbol{p})$ is a local maximum greater than a given threshold t_1.

2.2 Scale-space theory

In the previous definition, corners are only rotationally invariant, since they involve the eigenvalues of the autocorrelation matrix. They are generalised to be scale and affine invariant by the scale-space theory [32]. The importance of the observation scale for representing a signal at multiple resolutions has been pointed out since the early works on quadtrees [36] and pyramidal representations [37]. A first mathematical formalisation of the concept of scale-space has been given by [38], providing a multiscale representation that comprises a continuous scale parameter and preserves the same spatial sampling. Under these assumptions, it has been proved that different scales can be only simulated by Gaussian convolutions, which guarantee that no details will be at coarser scales [39]. The definitions of integration and differentiation scales have been also introduced [35], strictly related to the concepts of inner and outer scales [32]: the former gives the minimal scale for which an object of interest is completely visible, whereas the latter represents the minimal scale at which desired details begin to appear, that is, the minimal resolution required for the observations.

The autocorrelation matrix for a point $\boldsymbol{p}$ in the uniform scale-space is defined as

$$\mu(\boldsymbol{p}, \sigma_{\mathrm{I}}, \sigma_{\mathrm{D}}) = g_{\sigma_{\mathrm{I}}} \otimes \begin{bmatrix} I_x^2(\boldsymbol{p}, \sigma_{\mathrm{D}}) & I_xI_y(\boldsymbol{p}, \sigma_{\mathrm{D}}) \\ I_xI_y(\boldsymbol{p}, \sigma_{\mathrm{D}}) & I_y^2(\boldsymbol{p}, \sigma_{\mathrm{D}}) \end{bmatrix} \tag{3}$$

where σ_{I} and σ_{D} represent, respectively, the integration and differentiation scales, explicitly introduced to stress the scale-space framework; the operator $\otimes$ represents the usual convolution; g_σ is a uniform Gaussian kernel with radius 3σ, zero mean and standard deviation σ; $I_x(\boldsymbol{p}, \sigma_{\mathrm{D}})$ and $I_y(\boldsymbol{p}, \sigma_{\mathrm{D}})$ are the scale-space derivatives, obtained by the partial derivatives $I_x(\boldsymbol{p})$ and $I_y(\boldsymbol{p})$

$$\begin{aligned} I_x(\boldsymbol{p}, \sigma_{\mathrm{D}}) &= \sigma_{\mathrm{D}} g_{\sigma_{\mathrm{D}}} \otimes I_x(\boldsymbol{p}) \\ I_y(\boldsymbol{p}, \sigma_{\mathrm{D}}) &= \sigma_{\mathrm{D}} g_{\sigma_{\mathrm{D}}} \otimes I_y(\boldsymbol{p}) \end{aligned} \tag{4}$$

where the multiplicative factor σ_{D} is needed to normalise the derivatives after the convolution [32].

It should be noted that according to the scale-space theory the corner response function H (2) depends on three parameters:

the integration and differentiation scales, respectively, σ_I and σ_D, and the linear coefficient κ.

Through non-uniform Gaussian kernels, the affine scale-space [35] further extends the uniform scale-space to take into account affine transformations. The affine transformation inducted when the autocorrelation matrix is used as a covariance matrix, acts by transforming a unit circle into an ellipse with axis directions and lengths given, respectively, by the eigenvectors and eigenvalues of μ [35].

According to the scale-space theory, affine normalisation processes is applied to improve the Harris detector [14, 40]. The key idea is to extract corners at different scales, and then iteratively try to refine both the autocorrelation matrix and the scale. The values of σ_I and σ_D that have been adopted in [5, 14, 40, 41] to represent the scale-space are of the form $\sigma_{I_i} = \xi^i \sigma_{I_0}$ and $\sigma_{D_i} = s\sigma_{I_i}$, where ξ is the scale factor between successive levels, s is a constant factor and the index i increases with coarser scales. The autocorrelation matrix is improved by maximising the ratio λ_2/λ_1, that is, by maximising the corner response to H. The characteristic scale, defined as the scale for which a given function attains an extremum over the scales, is computed to refine the scale. The Laplacian operator has been used as a characteristic scale function with good results by several authors [14, 20, 40]. An efficient implementation of the above approaches is provided by the Harris-affine detector [14].

3 Proposed corner detector

This section starts by giving the motivation and the basic assumptions of our method, followed by a detailed description of the algorithm and its implementation.

3.1 Motivations and basic assumptions

As described in Section 2, the coefficient κ controls the corner response function H (2). In particular, the sensitivity of H is reduced when κ is increased. Moreover, a point is selected as a corner if its response to H is greater than t_1. It should also be noted that both t_1 and κ rely upon local properties of the input image, such as luminosity, noise or intrinsic structures (e.g. textured and non-textured image regions).

Referring to mathematical statistics, the mean can be adopted as a reference point to compare inhomogeneous data, while the standard deviation can be used as a unit of measurement. Therefore a good normalisation choice is given by the z-score [42] function

$$Z(x) = \frac{x - \bar{x}}{\sigma} \tag{5}$$

where $\bar{x}$ and σ are, respectively, the mean value and the standard deviation. If the mean values $\bar{x}_1$ and $\bar{x}_2$ of different quantities are associated with the same conditions, then x_1 and x_2 can be compared on the basis of their z-score normalisations.

These argumentations provide the basic assumptions for our detector. Indeed, both the average values of the determinant and the trace of the autocorrelation matrix of a gradient magnitude image can be associated with flat regions, usually considered as the background. Therefore their z-score normalisations provide a good corner response, with respect to the processed image. Actual corners have to be searched for near edges, or where a strong intensity variation is noticeable. The mean gradient magnitude itself provides a rough separation between flat regions and edges.

Our algorithm starts by computing an edge mask, used to stress the derivatives according to the edges. Next, a function H_z, based on H using the z-score normalisation, is used to measure the corner degree of the points. According to H_z, corners should be points $\boldsymbol{p}$ for which $H_z(\boldsymbol{p}) > 0$. Selected points which attain a local maximum over H_z and lay near the edges, according to the edge mask, are finally selected as corners. All these steps are repeated at different scales.

3.2 Algorithm description

Let us compute the image gradient magnitude through the derivatives $I_x(\boldsymbol{p}, \sigma_D)$ and $I_y(\boldsymbol{p}, \sigma_D)$ [(4), see Fig. 1]

$$G(\boldsymbol{p}, \sigma_D) = \sqrt{I_x^2(\boldsymbol{p}, \sigma_D) + I_y^2(\boldsymbol{p}, \sigma_D)} \tag{6}$$

A threshold by the mean value $\overline{G}$, separates flat regions and edges

$$K(\boldsymbol{p}, \sigma_D) = \begin{cases} 0 & \text{if } G(\boldsymbol{p}, \sigma_D) \le \overline{G} \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

By assuming that discontinuities have the same order as the current image resolution, actually given by σ_D, K is convolved by the Gaussian kernel g, thus reducing very crisp edges [see Figs. 2*a*–*b*]

$$M(\boldsymbol{p}, \sigma_D) = g_{\sigma_D} \otimes K(\boldsymbol{p}, \sigma_D) \tag{8}$$

The initial derivatives $I_x(\boldsymbol{p}, \sigma_D)$ and $I_y(\boldsymbol{p}, \sigma_D)$ are enhanced by a pixel-wise multiplication with the edge mask M

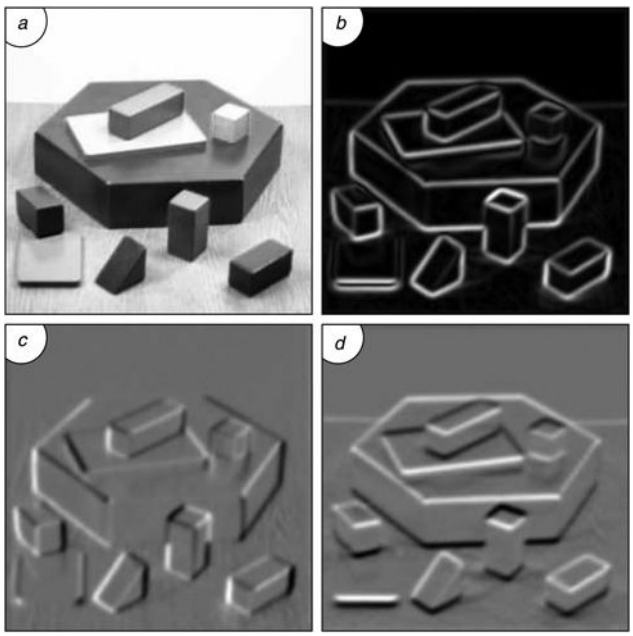

Fig. 1 *Input image and its derivatives*

a Input image I
b Gradient magnitude G
c Horizontal derivative I_x
d Vertical derivative I_y
In the following examples, these parameters were set: $i = 3$, $\sigma_{I_i} = 1.4^i$ and $\sigma_{D_i} = 0.7\sigma_{I_i}$

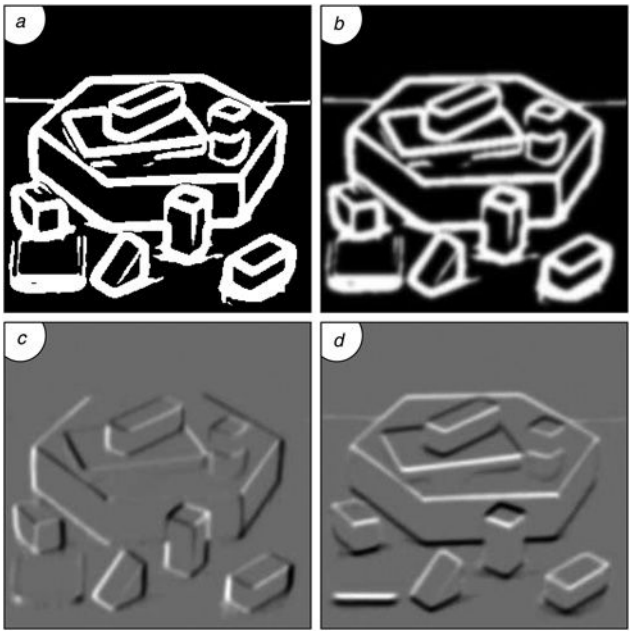

Fig. 2 *Image derivative enhancement*

a Binary edge mask K
b Smoothed mask M
c Enhanced horizontal derivative L_x
d Enhanced horizontal derivative L_y

[see Figs. 2*c*–*d*]

$$L_x(\boldsymbol{p}, \sigma_D) = M(\boldsymbol{p}, \sigma_D) I_x(\boldsymbol{p}, \sigma_D)$$
$$L_y(\boldsymbol{p}, \sigma_D) = M(\boldsymbol{p}, \sigma_D) I_y(\boldsymbol{p}, \sigma_D)$$

The resulting derivatives $L_x(\boldsymbol{p}, \sigma_D)$ and $L_y(\boldsymbol{p}, \sigma_D)$ are used in place of $I_x(\boldsymbol{p}, \sigma_D)$ and $I_y(\boldsymbol{p}, \sigma_D)$ in the autocorrelation matrix $\mu(\boldsymbol{p}, \sigma_I, \sigma_D)$ (3), to suppress most of the noise.

The new corner response function is defined as

$$H_z(\boldsymbol{p}, \sigma_I, \sigma_D) = Z(\det(\mu(\boldsymbol{p}, \sigma_I, \sigma_D))) - Z(\mathrm{tr}^2(\mu(\boldsymbol{p}, \sigma_I, \sigma_D))) \quad (9)$$

where the mean and the standard deviation involved in the z-score (5) are computed on the whole image. We want to point out that this new function allows the comparison of the determinant and the squared trace of the autocorrelation matrix without the usual coefficient κ. We recall that the mean values of both determinant and trace, which are equal to zero after the z-score normalisation, can be associated with flat regions. Fig. 3*a* shows negative and positive values of H_z as dark and bright regions, respectively.

Recalling that the determinant of the autocorrelation matrix is sensitive to corners, while the trace is sensitive to both edges and corners, it follows that:

- $H_z(\boldsymbol{p}, \sigma_I, \sigma_D) \gg 0$, when the corner response is greater than the edge response;
- $H_z(\boldsymbol{p}, \sigma_I, \sigma_D) \ll 0$, when the edge response is greater than the corner response.

A candidate corner $\boldsymbol{p}$ is selected if it attains a local maximum for H_z greater than zero, within a circular window with the same radius of the Gaussian kernel g_{σ_D} and if $\boldsymbol{p}$ lies near an edge [see Fig. 3*b*], that is when $M(\boldsymbol{p}, \sigma_D) > t_2$, for a proper threshold value t_2 [see Figs. 3*c*–*d*]. Opposed to the

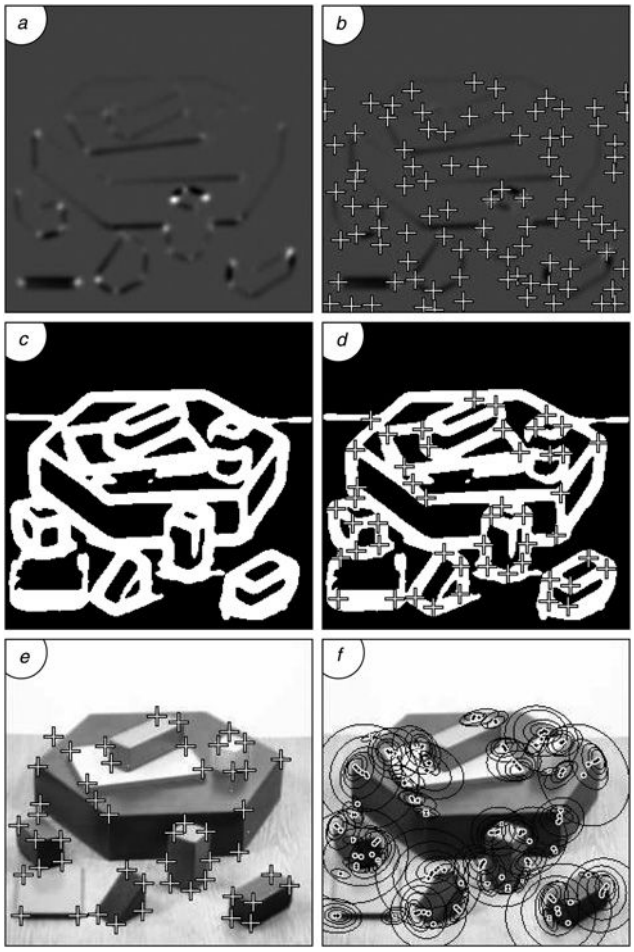

Fig. 3 *Corner selection*

a H_z function
b Local maxima greater than zero
c Binarised mask $M > 0.31$
d Extracted corners for $i = 3$
e Corners superimposed on the input image
f Ellipse representation for $i = 3, \ldots, 8$

Harris-affine detector, no affine refinements nor characteristic scale computation are carried out.

The following general notes must be considered for the choice of the threshold value t_2. Firstly, it must be noted that the values of M range in [0, 1] (8), since they have been obtained by Gaussian convolution of the binary mask K (7). The differentiation scale σ_D can be assumed to be the unit length of a pixel at the current observation scale, and thus its size is σ_D. As in Fig. 4, let us consider a

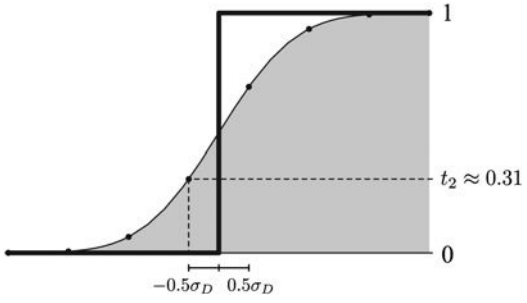

Fig. 4 *Ideal step edge (bold line), centred in 0, is convolved with* g_{σ_D} *(solid line)*

Differentiation scale σ_D gives the current resolution, and thus the size of a pixel is σ_D. The separation border in $-0.5\sigma_D$ is the value $t_2 \simeq 0.31$

long-enough ideal step edge $f(x)$ in $x = 0$, that is the pixel on the ramp lies in $[-0.5\sigma_D, 0.5\sigma_D]$. After the convolution of f with g_{σ_D}, the initial value of the pixel can be restored with the cut-off at $x = -0.5\sigma_D$, which corresponds to $t_2 \simeq 0.31$

$$\begin{aligned} t_2 &= g_{\sigma_D} \otimes f(-0.5\sigma_D) \\ &= \int_{-\infty}^{\infty} f(x) g_{\sigma_D}(x + 0.5\sigma_D) \mathrm{d}x \\ &= \int_{0}^{\infty} g_{\sigma_D}(x + 0.5\sigma_D) \mathrm{d}x \\ &= \int_{0}^{0.5\sigma_D} g_{\sigma_D}(x) \mathrm{d}x \\ &= \Psi_{0,\sigma_D}(0.5\sigma_D) = \Psi_{0,1}(0.5) \simeq 0.31 \end{aligned}$$

where $\Psi_{\eta,\sigma}(x)$ is the normal cumulative distribution with mean η and standard deviation σ (see Fig. 4). Here, the assumption that the pixel is centred in $x = 0$ was made so that the left border of the pixel is in $x = -0.5\sigma_D$. A threshold $t_2 = 0.31$ guarantees a general lower bound which does not depend on the position of the step edge, neither on σ_D. Actually, this value includes the case $t_2 = 0.5$, which corresponds to the pixel border in $x = 0$.

The whole method is repeated for different scales σ_I, and therefore a corner $\boldsymbol{p}$ has to be identified by the triplet $\boldsymbol{c} = (\boldsymbol{p}, \mu, \sigma_I)$. Each corner delimits an elliptic region S_c, centred in $\boldsymbol{p}$, with axis directions and lengths given, respectively, by the eigenvectors and eigenvalues of μ [(3), see Fig. 3*f*].

No image pyramid approach [43] has been used, although it speeds up the process, because it decreases the point location accuracy and requires the affine normalisation process described in Section 2.2.

3.3 Implementation details

The observation scale depends on the level of details required by the particular image processing task and should be provided by the user. The scales have been set according to $\sigma_{I_i} = \xi^i \sigma_{I_0}$ and $\sigma_{D_i} = s\sigma_{I_i}$, with $\xi = 1.4$, $\sigma_{I_0} = 1$ and $s = 0.7$, as reported in [5, 14, 41]. Coarser scales are obtained by increasing the index i.

Table 1 shows the average computational time required to process a single scale in the case of i from 1 to 11 and the respective cumulative time, together with the percentage of extracted corners with respect to $i = 1$. The tests have been performed on the Oxford image data set [33]. The algorithm has been implemented on a Linux system with kernel version 2.6.27, running on Intel® Core™2 Duo E8500 CPUs. When the scale is increased, the computational time almost doubles and the number of corners extracted halves.

Table 1 Computational times and extracted corner percentage, per scale

Scale index, to coarser scales	Corners, %	Time (min:sec)		Cum.time (min:sec)	
		Mean	Max	Mean	Max
1	100.0	0:02	0:03	0:03	0:06
2	49.3	0:02	0:03	0:05	0:09
3	**31.5**	**0:02**	**0:03**	**0:08**	**0:12**
4	**17.1**	**0:03**	**0:04**	**0:10**	**0:15**
5	**9.7**	**0:03**	**0:04**	**0:13**	**0:19**
6	**5.3**	**0:04**	**0:06**	**0:17**	**0:24**
7	**2.9**	**0:05**	**0:07**	**0:22**	**0:31**
8	**1.6**	**0:07**	**0:08**	**0:28**	**0:38**
9	0.8	0:09	0:12	0:37	0:50
10	0.4	0:13	0:16	0:50	1:07
11	0.3	0:19	0:24	1:09	1:30

Values represented in bold are best scales

To further refine the corner selection, points with $\lambda_2/\lambda_1 < 0.25$ have been discarded. Only 1.9% of the initially extracted corners have been removed, underlining the consistency of the new method.

The floating-point computation of the maximum value of H_z [see (9)] within a circular window of radius $3g_{\sigma_D}$ is particularly time consuming in the case of high values of the index i. To speed up the algorithm, we have implemented a heap structure [44] and introduced auxiliary indexes to update the elements on the border of the moving window, as illustrated in Fig. 5.

The whole absolute time of our detector is quite high, compared to other feature detectors [8, 15, 23], but it is still suitable for offline tasks. In any case, it should be noted that no approximations, which usually improve the computational performances, have been made. Observing the small amount of features corresponding to more time-consuming coarser scales, a reasonable compromise between the number of extracted corners and the computational performances (see Table 1, columns 'corners' and 'time') can be provided with scale indexes $i < 9$.

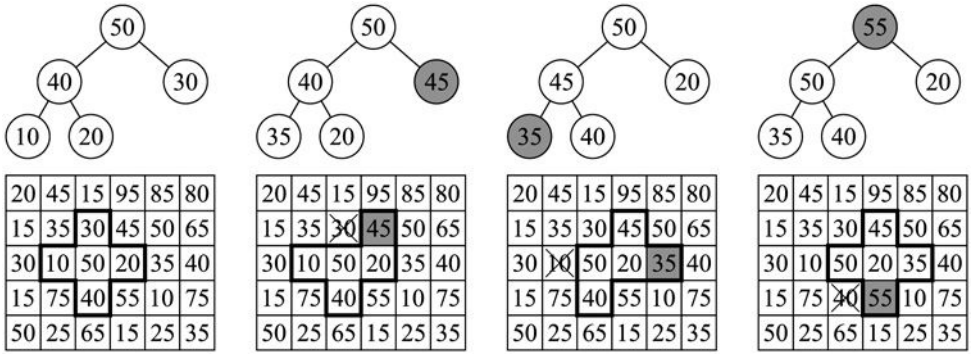

Fig. 5 *Example of heap to find the maximum value within the window (in bold)*

When moving this window on the image, one pixel at a time is replaced and usually just part of the heap has to be updated

4 Experimental results and evaluation

In this section, our algorithm is compared with other state-of-the-art detectors, according to the so-called repeatability index and matching score [15].

4.1 Experimental set-up

Feature detectors have been extensively compared in several papers [15, 27, 28], which show the difficulty in designing an efficient test procedure. Indeed, a ground truth should be known to estimate the correct positions and corresponding features should be matched across different views of the same scene, owing to various image transformations. In planar scenes, the evaluation of correspondences is less difficult than in actual three-dimensional objects, although this latter case provides more general results.

Given two images A and B, let us consider the set of corners $\{\boldsymbol{a}_i\} \in A$ and $\{\boldsymbol{b}_j\} \in B$ which are also present, respectively, in B and A through the maps $T_{A\to B}$ and $T_{B\to A}$, that is $T_{A\to B}(\boldsymbol{a}_i) \in B$ and $T_{B\to A}(\boldsymbol{b}_j) \in A$. The overlap error

$$\varepsilon_{\boldsymbol{a}_i,\boldsymbol{b}_j} = 1 - \frac{S_{\boldsymbol{a}_i} \cap T_{B\to A}(S_{\boldsymbol{b}_j})}{S_{\boldsymbol{a}_i} \cup T_{B\to A}(S_{\boldsymbol{b}_j})}$$

measures how much the elliptic regions $S_{\boldsymbol{a}_i}$ and $S_{\boldsymbol{b}_j}$ overlap each other.

We define the set $Q = \{(\boldsymbol{a}_i, \boldsymbol{b}_j)\}$, so that the corners $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ are considered once and the value $\varepsilon_{\boldsymbol{a}_i,\boldsymbol{b}_j}$ is minimised. The repeatability index measures the stability of the selected corners

$$r_{A,B}(\varepsilon) = \frac{|\{(\boldsymbol{a}_i, \boldsymbol{b}_j) \in Q : \varepsilon_{\boldsymbol{a}_i,\boldsymbol{b}_j} < \varepsilon\}|}{\min\{|\{\boldsymbol{a}_i\}|, |\{\boldsymbol{b}_j\}|\}} \tag{10}$$

where $|\cdot|$ indicates the usual set cardinality. The value of the repeatability index can increase by accidental overlaps when incrementing the number of corners or the size of their

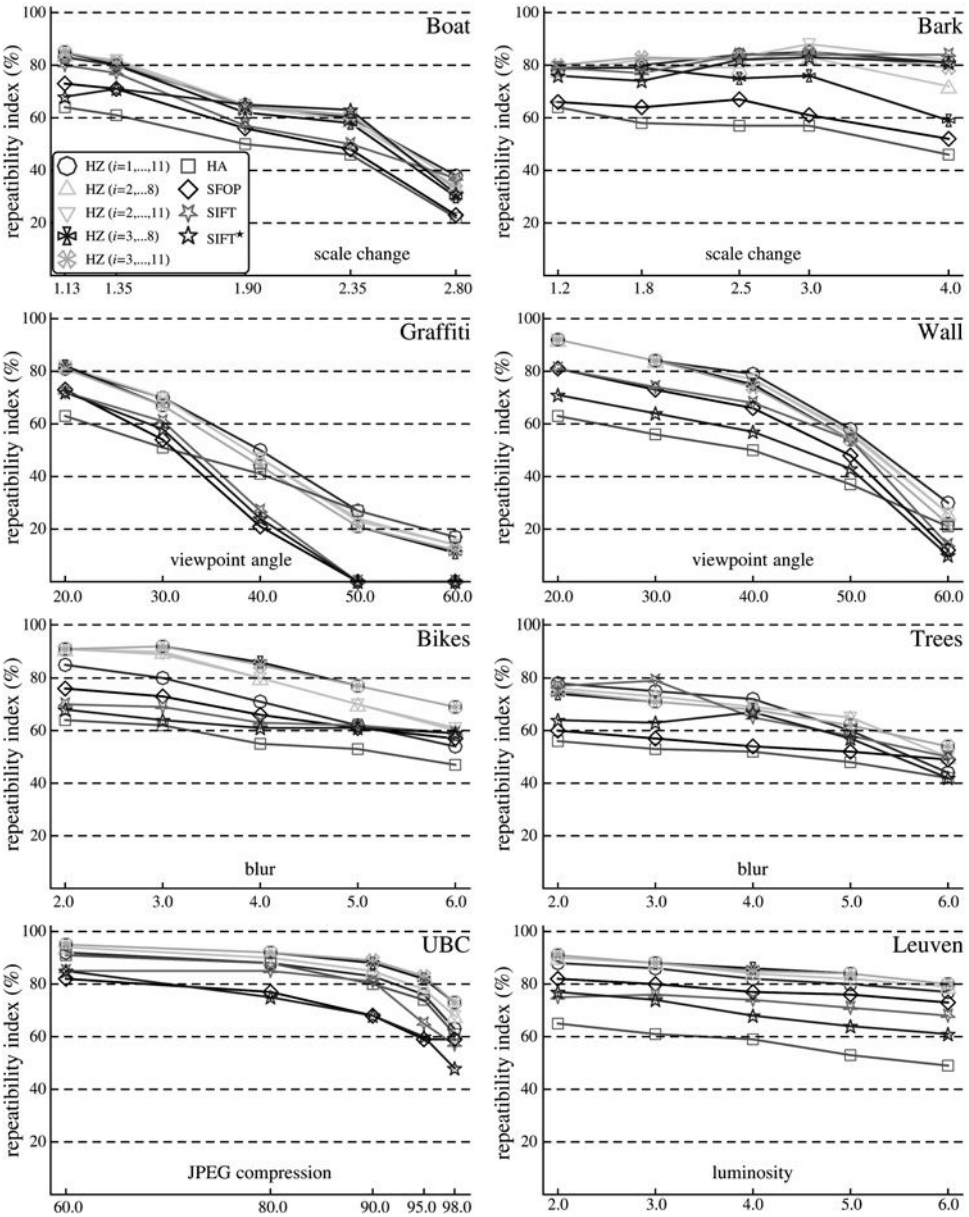

Fig. 6 *Repeatability index obtained for different sequences and transformations (both indicated within the plots) of the Oxford data set*
Labels HA and HZ refer to the Harris-affine algorithm and to our method, respectively

elliptic regions S_c. Moreover, the repeatability index does not consider the distinctiveness of the corners, which is crucial in the matching phase.

Extracted corners can be matched using different strategies, which require to select a feature descriptor. SIFT-based descriptors [6, 8] usually provide better results. Corners are then matched if their distance in the descriptor space is lower than a defined threshold or according to a nearest-neighbour approach [8]. For a given strategy, let F be the set of matches between $\{\boldsymbol{a}_i\}$ and $\{\boldsymbol{b}_j\}$, so that only a single match is allowed for each pair $S_{\boldsymbol{a}_i}$ and $S_{\boldsymbol{b}_j}$. The matching score is defined as the ratio between the number of correct matches in F and the minimum number of corners

$$m_{A,B}(\varepsilon) = \frac{|\{(\boldsymbol{a}_i, \boldsymbol{b}_j) \in F : \varepsilon_{\boldsymbol{a}_i,\boldsymbol{b}_j} < \varepsilon\}|}{\min\{|\{\boldsymbol{a}_i\}|, |\{\boldsymbol{b}_j\}|\}} \tag{11}$$

Besides the repeatability index and the matching score, which can be seen as complementary indexes, other less common approaches exist based on geometric distances or ROC curves [27], which proved that the accuracy of all detectors and descriptors degrade noticeably in the general case of three-dimensional scenes.

We considered the Oxford database [33] which includes eight greylevel sequences of six structured or textured images with various dimensions (width $= 765, \ldots, 1000$ and height $= 512, \ldots, 700$ pixels), where a structured image contains homogeneous regions with distinctive edges and a textured image has repeated patterns of different forms. Different degrees of transformations, including blur, viewpoint change, scale and rotation, JPEG compression and luminosity variation have been gradually applied to the sequences. Although this data set does not contain actual three-dimensional scenes, most of the detectors have been tested on it, thus allowing a comparison with these detectors.

The evaluation framework by Fraundorfer and Bischof [34] has been adopted to take into account non-planar scenes. This benchmark includes two sequences, subject to viewpoint change, with 800 × 600 and 896 × 1024 pixels. The ground

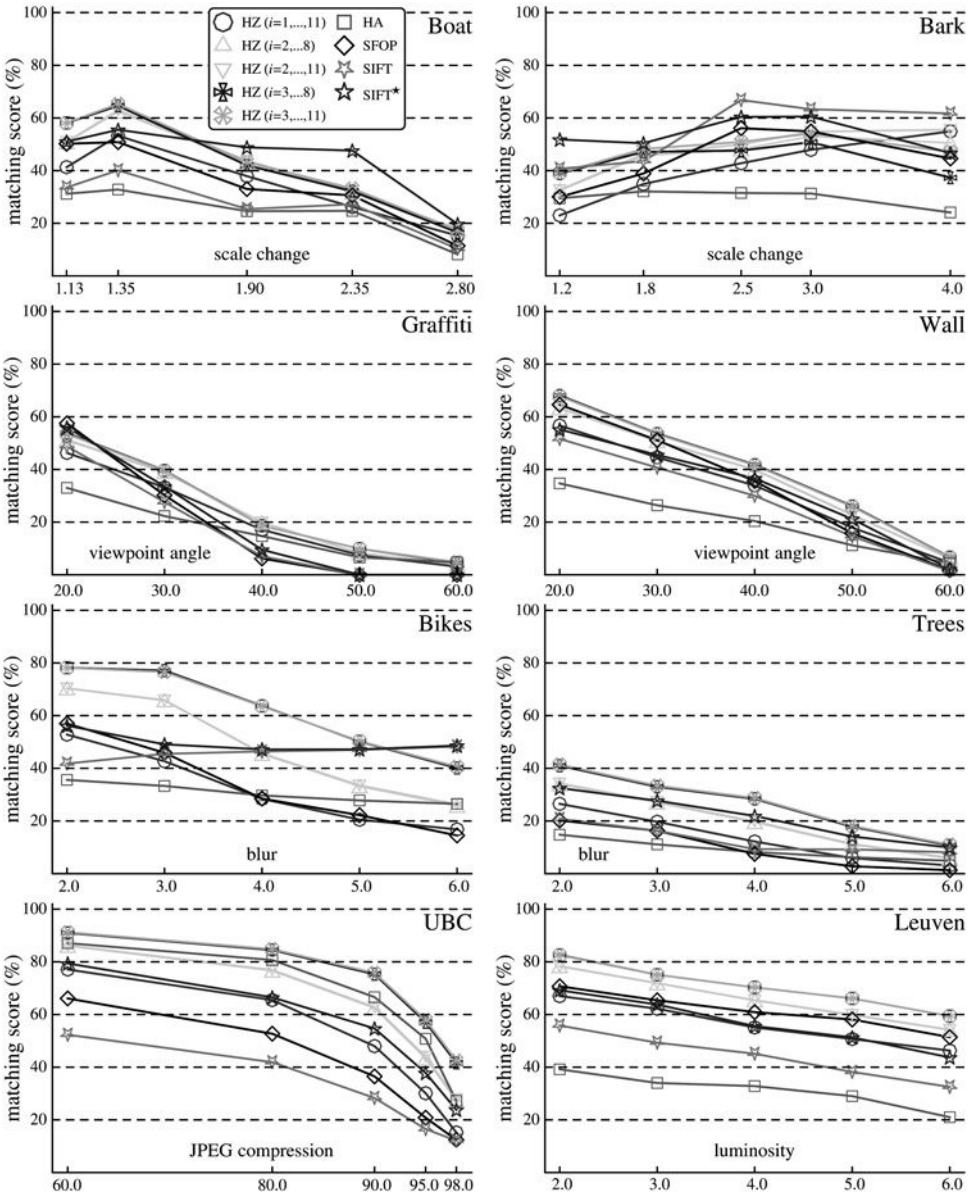

Fig. 7 *Matching score obtained for different sequences and transformations (both indicated within the plots) of the Oxford data set*
Labels HA and HZ refer to the Harris-affine algorithm and to our method, respectively

truth of the corresponding three-dimensional structures were obtained by dense matching and trifocal tensor.

We tested our algorithm according to the repeatability index (10) and the matching score (11), using the SIFT descriptor and the default experimental settings reported in [15, 28]. The SIFT and the recent SFOP detectors have been also included in the comparison, using the code freely distributed by the authors with its default parameters. To fairly compare the same number of points we detected for indexes $i = 3, \ldots, 8$, we have considered a subset of points, called SIFT*, obtained by the SIFT method with scales $\sigma_I > 1.68$.

4.2 Results

We report in Figs. 6–8 the results, according to the repeatability index and the matching score, referred not only to our algorithm, but obtained also through the Harris-affine detector [15], the SFOP and the SIFT methodologies. Tests are reported in [15, 28], also in the case of further detectors, and show that our approach is better than the Harris-affine methodology and comparable with the Hessian-affine and MSER detectors. We have experimentally verified that our algorithm is comparable also to SFOP and SIFT approaches, although the latter returns a slightly better matching score in the case of the bark and boat sequences, which correspond to scale and rotation changes.

The discrimination between features and noise at finer scales is a difficult task, indeed the feature stability increases with coarser scales. This conclusion can be also deduced from Fig. 9 which shows the average repeatability index with respect to the overlap error.

Furthermore, for high-scale indexes, results are almost unchanged, because only a negligible number of corners is extracted and their computation can be avoided, thus reducing the running time. In order to compare different detectors, both repeatability index and matching score require about the same number of extracted points. Actually, our algorithm with $i \geq 3$, the Harris-affine procedure and the SFOP and SIFT* detectors returned a comparable number of points, but, in the case of the SIFT method, indexes $i \geq 1$ have been considered (see Table 2).

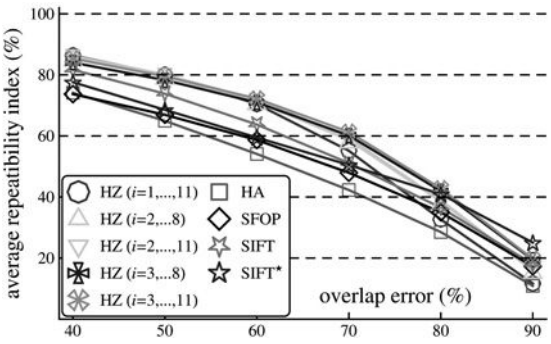

Fig. 9 *Average repeatability index against overlap error, computed on the Oxford data set*

Labels HA and HZ refer to the Harris-affine algorithm and to our method, respectively

Further discussions should be carried out for the Graffiti sequence of Fig. 6, where the new algorithm behaves like the Harris-affine detector for high-viewpoint angles. Projective transformations are assumed piecewise locally affine, in the case of small projective distortions. Moreover, since corners rely on contour stability unlike blob-like

Table 2 Median of extracted point percentages of the proposed algorithm, for various ranges, with respect to other detectors

	Scale index range				
	1–11	2–8	2–11	3–8	3–11
HA, %	378	186	189	106	107
SFOP, %	336	177	179	97	100
SIFT*, %	366	188	190	103	105
SIFT, %	99	55	56	34	35

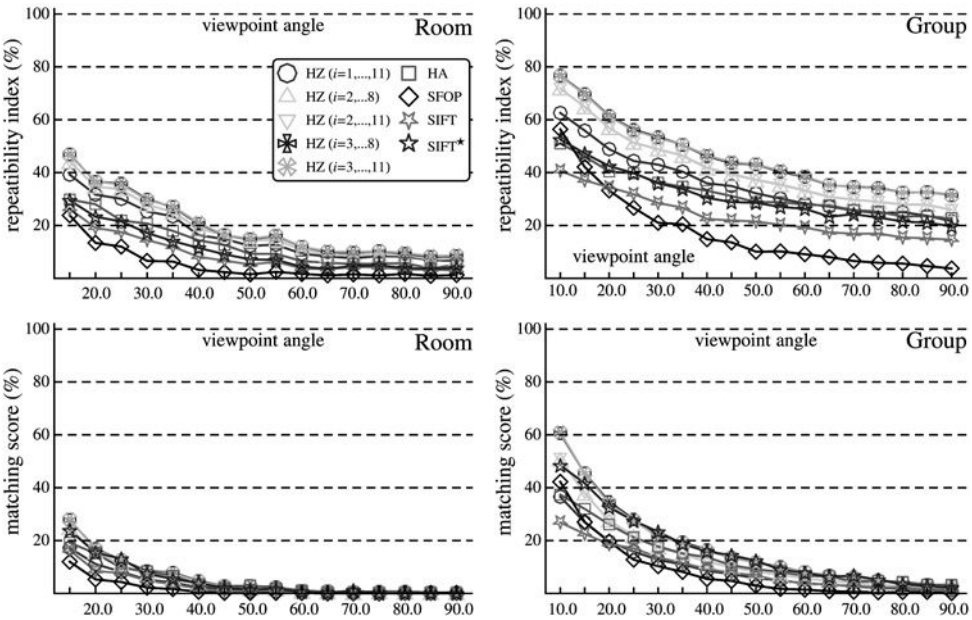

Fig. 8 *Repeatability index and matching score in the case of viewpoint angle transformation on both non-planar sequences of the data set by Fraundorfer and Bischof*

Labels HA and HZ refer to the Harris-affine algorithm and to our method, respectively

features, which are based on homogeneous flat regions, corner detectors are more sensitive to relevant viewpoint changes.

In the case of non-planar scenes, the proposed algorithm returns good results, comparable with those obtained by the MSER detector, thus underlining the stability and validity of our approach.

5 Conclusion and final remarks

As reported in Section 4.2, tests show the validity of the proposed methodology that returns good results when compared with other recent feature detectors.

In terms of the repeatability index and matching score, our algorithm provides better results, under all kinds of transformations applied in the Oxford data set, with respect to the standard Harris-affine detector, and a comparable performance, with respect to the Hessian-affine and MSER approaches. These considerations still hold for the non-planar sequences of the evaluation framework provided by Fraundorfer and Bischof.

The z-score function normalises the determinant and the trace parts of the Harris function and, therefore does not demand the user to set the value of the linear combination coefficient, which depends on the particular image to analyse. The mask to limit the search for corners close to the edges returns a stable output, which is no longer affected by the threshold value that relies on the content of the whole image.

It can be noted that corners are complementary to blob-like features, and thus they provide additional information that can be combined to obtain better results. We have experimentally proved that a complete pyramidal structure is not required and that just a limited number of scales are sufficient to locate almost all corners present in the image.

Although our method is not fast, it is still appropriate for offline tasks which require high accuracy. For example, it can applied in different computer vision problems, such as aerial image segmentation, retinal feature extraction and underwater mosaicing. A possible application regards the combination of the proposed method with other feature detectors, usually through a voting strategy, in order to refine the final result.

6 Acknowledgments

This work makes use of results produced by the PI2S2 Project managed by the Consorzio COMETA, a project co-funded by the Italian Ministry of University and Research (MIUR) within the Programma Operativo Nazionale 'Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione' (PON 2000–2006). More information is available at http://www.pi2s2.it and http://www.consorzio-cometa.it.

This work has also been partially supported by a grant 'Fondo per il Potenziamento della Ricerca del Dipartimento di Matematica e Applicazioni dell'Università degli Studi di Palermo'.

This work has been produced within the ASIC Project 'Applicazioni e Servizi Innovativi di Comunicazione, Voce, Video e Dati', through grant ASIC-FB #DM29098(GPS).

7 References

1 Schaffalitzky, F., Zisserman, A.: 'Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"'. European Conf. on Computer Vision, 2002, pp. 414–431

2 Hartley, R., Zisserman, A.: 'Multiple view geometry in computer vision' (Cambridge University Press, 2004, 2nd edn.)

3 Brown, M., Lowe, D.G.: 'Recognizing panoramas'. Int. Conf. on Computer Vision, 2003, pp. 1218–1225

4 Mikolajczyk, K., Leibe, B.: 'Local features for object class recognition'. Int. Conf. on Computer Vision, 2005, pp. 1792–1799

5 Lowe, D.G.: 'Object recognition from local scale-invariant features'. Int. Conf. on Computer Vision, 1999, pp. 1150–1157

6 Mikolajczyk, K., Schmid, C.: 'A performance evaluation of local descriptors', *Pattern Anal. Mach. Intell.*, 2005, **27**, (10), pp. 1615–1630

7 Freeman, W., Adelson, E.: 'The design and use of steerable filters', *Pattern Anal. Mach. Intell.*, 1991, **13**, (9), pp. 891–906

8 Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

9 Moravec, H.: 'Obstacle avoidance and navigation in the real world by a seeing robot rover' (Carnegie-Mellon University, Robotics Institute, 1980)

10 Harris, C., Stephens, M.: 'A combined corner and edge detector'. Alvey Vision Conf., 1988, pp. 147–151

11 Förstner, W.: 'A feature-based correspondence algorithm for image matching', *Int. Arch. Photogram. Remote Sens.*, 1986, **26**, pp. 150–166

12 Shi, J., Tomasi, C.: 'Good features to track'. IEEE Conf. on Computer Vision and Pattern Recognition, 1994, pp. 593–600

13 Kenney, C.S., Zuliani, M., Manjunath, B.S.: 'An axiomatic approach to corner detection'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005, pp. 191–197

14 Mikolajczyk, K., Schmid, C.: 'Scale and affine invariant interest point detectors', *Int. J. Comput. Vis.*, 2004, **60**, (1), pp. 63–86

15 Mikolajczyk, K., Tuytelaars, T., Schmid, C., *et al.*: 'A comparison of affine region detectors', *Int. J. Comput. Vis.*, 2005, **65**, (1/2), pp. 43–72

16 Smith, S.M., Brady, J.M.: 'Susan – a new approach to low level image processing', *Int. J. Comput. Vis.*, 1997, **23**, (1), pp. 45–78

17 Trajkovic, M., Hedley, M.: 'Fast corner detection', *Image Vis. Comput.*, 1998, **16**, (2), pp. 75–87

18 Lindeberg, T.: 'Junction detection with automatic selection of detection scales and localization scales'. First Int. Conf. on Image Processing, 1994, pp. 924–928

19 Parida, L., Geiger, D., Hummel, R.A.: 'Kona – a multi-junction detector using minimum description length principle'. First Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, 1997, pp. 51–65

20 Lindeberg, T.: 'Feature detection with automatic scale selection', *Int. J. Comput. Vis.*, 1998, **30**, (2), pp. 79–116

21 Beaudet, P.R.: 'Rotationally invariant image operators'. Int. Joint Conf. on Pattern Recognition, 1978, pp. 578–583

22 Kitchen, L., Rosenfeld, A.: 'Gray-level corner detection', *Pattern Recognit. Lett.*, 1982, **1**, pp. 95–102

23 Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: 'Surf – speeded up robust features', *Comput. Vis. Image Underst.*, 2008, **110**, (3), pp. 346–359

24 Matas, J., Chum, O., Urban, M., Pajdla, T.: 'Robust wide baseline stereo from maximally stable extremal regions'. British Machine Vision Conf., 2002, pp. 384–393

25 Kadir, T., Zisserman, A., Brady, M.: 'An affine invariant salient region detector'. European Conf. on Computer Vision, 2004, pp. 345–457

26 Tuytelaars, T., Gool, L.V.: 'Matching widely separated views based on affine invariant regions', *Int. J. Comput. Vis.*, 2004, **59**, (1), pp. 61–85

27 Moreels, P., Perona, P.: 'Evaluation of features detectors and descriptors based on 3d objects'. Int. Conf. on Computer Vision, 2005, pp. 800–807

28 Fraundorfer, F., Bischof, H.: 'A novel performance evaluation method of local detectors on non-planar scenes'. Workshop on Empirical Evaluation Methods in Computer Vision, 2005

29 Förstner, W., Dickscheid, T., Schindler, F.: 'Detecting interpretable and accurate scale-invariant keypoints'. Int. Conf. on Computer Vision, Kyoto, Japan, 2009

30 Bigün, J.: 'A structure feature for some image processing applications based on spiral functions', *Comput. Vis. Graph. Image Process.*, 1990, **51**, (2), pp. 166–194

31 Dickscheid, T., Förstner, W.: 'Evaluating the suitability of feature detectors for automatic image orientation systems'. Int. Conf. on Computer Vision Systems, Liege, Belgium, 2009

32 Lindeberg, T.: 'Scale-space theory in computer vision' (Kluwer Academic Publishers, 1994)

33 Mikolajczyk, K., Tuytelaars, T., Schmid, C., *et al.*: Affine Covariant Features. Available at http://www.robots.ox.ac.uk/~vgg/research/affine

34 Fraundorfer, F., Bischof, H.: Local detector evaluation. Available at http://icg.tugraz.ac.at/Members/fraunfri/local-detector-evaluation

35 Mikolajczyk, K.: 'Detection of local features invariant to affines transformations' (Institut National Polytechnique de Grenoble, 2002)

36 Klinger, A.: 'Patterns and search statistics'. Optimizing Methods in Statistics, 1971, pp. 303–339

37 Crowley, J.: 'A representation for visual information' (Carnegie-Mellon University, Robotics Institute, 1981)
38 Witkin, A.P.: 'Scale space filtering'. Int. Joint Conf. on Artificial Intelligence, 1983, pp. 1019–1023
39 Koenderink, J.J.: 'The structure of images', *Biol. Cybern.*, 1984, **5**, (50), pp. 363–370
40 Baumberg, A.: 'Reliable feature matching across widely separated views'. IEEE Conf. on Computer Vision and Pattern Recognition, 2000, pp. 774–781
41 Lindeberg, T.: 'Feature detection with automatic scale selection', *Int. J. Comput. Vis.*, 1998, **2**, (30), pp. 79–116
42 DeGroot, M.H., Schervish, M.J.: 'Probability and statistics' (Addison Wesley, 2001, 3rd edn.)
43 Crowley, J.L., Riff, O., Piater, J.: 'Fast computation of characteristic scale using a half octave pyramid'. Int. Workshop on Cognitive Computing, 2002
44 Sedgewick, R.: 'Algorithms in C' (Addison-Wesley, 1997, 3rd edn.)