Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Adapting Interest Point Detection to Illumination Conditions

Flore Faille

Institute for Real-Time Computer Systems
Technische Universität München, D-80333 Munich, Germany
Flore.Faille@rcs.ei.tum.de

Abstract. The objective of the presented work is to improve the stability of interest point detection under illumination changes. The use of a global threshold is shown to be insufficient and three methods are proposed to enhance a state–of–the–art algorithm: the Harris corner detector. These methods are based on different principles: a local image normalization as preprocessing, a local threshold adaption, and a local automated threshold selection based on clustering. All methods are compared on several image series created by varying the lighting conditions. For performance evaluation repeatability and false positive rates are used. All methods allow a stability enhancement under complex illumination changes. The algorithm based on threshold adaption performs best.

1 Introduction

Many computer vision tasks are based on the detection of interest points in images. This includes 3D reconstruction [1–3], tracking [4], content–based image retrieval [3, 5], object recognition [3, 6, 7] and mobile robot localization [8, 9]. Using such local features can improve the robustness of algorithms, e.g. to partial occlusions [7, 5]. Furthermore interest point detection is used to reduce the data flow and consequently the computational costs, as all consecutive processing steps such as characterizing the points' neighborhood and matching them are solely applied to the interest points. Therefore the detection should be reliable and should ideally deliver the same points under all possible imaging conditions. Higher level tasks may be able to compensate a partial feature loss using statistical or robust methods (e.g. RANSAC), however the number of misdetections influences the efficiency and accuracy of such algorithms [2]. As a consequence, it is worth spending some effort to improve the detector stability under changing imaging conditions. The repeatability [10] is often used to measure this stability.

In this paper the focus is put on varying lighting conditions, which is still one weak point of interest point detection. Changes of the intensity, of the wavelength composition and above all of the position and orientation of the light source(s) can induce strong variations in the appearance of a scene. In most environments, lighting conditions cannot be controlled. This problem must be overcome – among others – during interest point detection. Non–uniformly lighted

scenes reveal well the difficulty: state–of–the–art detectors tend to extract interest points exclusively in the most illuminated areas, so any change of the light source position may result in poor repeatability as can be seen in Fig. 1.

Different methods are presented in this work to improve a state–of–the–art and popular interest point detector: the Harris corner detector [11]. These methods perform an adaption of the detection to the local lighting conditions and are compared on several test series that involve different types of illumination changes.

The next section explains why the Harris corner detector was chosen and introduces how it works. Section 3 presents the influence of illumination changes on the detection and some solutions used in other image analysis areas. The different methods proposed to improve the interest point detector are given in section 4. Finally experiments and results are described in section 5.

2 Interest Point Detector

As pointed out in the introduction, interest point detectors have important applications and have consequently received a lot of attention. There exist many ways to define interest points. Thus, many detectors have been designed based for example on local grey value extrema [3, 6], on curvature maxima along contours [12, 10] or on the local grey value distribution [13]. The Harris corner detector [11] selects points for which the autocorrelation function significantly drops in two perpendicular directions. That way the interest points can be optimally retrieved after a limited camera motion e.g. in tracking applications [4]. Comparisons between several algorithms [10, 2] have shown that the Harris corner detector reaches the best repeatability rate for moderate changes of the imaging conditions. Recently scale and affine invariant detection of interest points gained importance [6, 3, 12]. The Harris corner detector can be extended to being invariant to such transformations as shown in [14]. Furthermore, it was proved that the interest points extracted with the Harris detector possess high information content [10] and high saliency [15]. These facts are confirmed by means of the different applications in which it was successfully used, e.g. in [1, 2, 4, 5, 7, 9].

For these reasons, the Harris corner detector was chosen in this work for interest point extraction. The implementation proposed by Schmid et al. [10] is used. The detection of interest points is based on the following matrix $\mathbf{C}$ that represents the local statistics of the first order derivatives around a pixel (x, y):

$$\mathbf{C} = G(\sigma) \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \tag{1}$$

where $G(\sigma)$ is a Gaussian with standard deviation σ and $\otimes$ is the convolution operator. The first derivatives I_x and I_y are estimated by convolving the grey value image $I(x, y)$ with the derivatives of a Gaussian to reduce noise and aliasing effects [10] (here $\sigma_{deriv} = 1.2$ is used). Interest points are pixels for which $\mathbf{C}$ has two big eigenvalues. The "cornerness" function R allows a direct detection:

$$R = \det(\mathbf{C}) - \alpha \operatorname{trace}^2(\mathbf{C}) \quad \text{with} \quad 0.04 \leq \alpha \leq 0.06 \text{ [11]}. \tag{2}$$

In this work, $\alpha = 0.06$ and $\sigma = 3.0$ are used. Interest points are found at local maxima of R above a given threshold T ($T > 0$). Some authors [2, 4] favor the use of the eigenvalues for detection. We observed that the use of the cornerness function R not only allows faster calculation but as well lower sensitivity to noise and aliasing problems.

3 Influence of Illumination Variations on the Detection

Illumination changes represent a problem for the interest point detection as shown in Fig. 1: only 46% of the points are re–detected after the lamps have been turned on. Despite the many algorithms available to overcome these changes for color images (see e.g. [16]), the attempts to consider lighting conditions during interest point detection from grey value images stayed modest so far. Most of the time, a fixed threshold is used [4, 7, 9, 13, 12]. In [10] the threshold is set to 1% of the maximum of R to adapt to the global intensity. It is shown to yield a reasonable stability for planar scenes and moderate lighting changes. Other authors select the N points with the highest cornerness values [3, 2] or the top ε% of the points [1]. In [1] the image is furthermore divided into 4 parts before setting thresholds to achieve a more uniform interest point distribution.

Fig. 1. In the left image the scene is illuminated by sunlight, whereas in the right image the light comes from neon lamps. The detected interest points are indicated by circles of radius 3σ. Here $T = 0.01 \cdot \max(R)$ is used as proposed in [10]. Only 46.0% of the interest points of the left image are re–detected in the right image.

To understand why the use of a global threshold is not sufficient for a stable detection, the image formation model for lambertian surfaces and a distant point light source [3, 16, 17] can be considered:

$$I = \mathbf{e} \cdot \mathbf{n} \int S(\lambda)E(\lambda)F(\lambda)d\lambda, \tag{3}$$

where I is the pixel intensity, λ the wavelength, $\mathbf{e}$ the orientation of the incident light, $\mathbf{n}$ the normal of the viewed surface, $S(\lambda)$ the surface reflectance, $E(\lambda)$ the illumination spectrum and $F(\lambda)$ the camera response function.

According to (3), a variation of the light intensity ($E(\lambda) \rightarrow aE(\lambda)$) results in a multiplicative factor common to all pixels. But if the illumination spectrum changes ($E(\lambda) \rightarrow E'(\lambda)$) or if the light source moves ($\mathbf{e} \rightarrow \mathbf{e}'$), the intensity variations depend on the viewed color and on the underlying 3D structure. Consequently, the transformation between two images I_1 and I_2 of a scene under two different illuminations is often approximated by a linear equation which paramaters may vary from pixel to pixel [3, 17]:

$$I_2(x, y) \approx a(x, y)I_1(x, y) + b(x, y). \tag{4}$$

The parameters $a(x, y)$ and $b(x, y)$ are often assumed to remain constant over small image regions [3, 17]. Using this assumption and the definition of the Harris detector, the effect on the cornerness function can be approximated as follows:

$$R_2(x, y) \approx a^4(x, y)R_1(x, y), \quad \text{where } a(x, y) \text{ varies slowly in space.} \tag{5}$$

This formula is not accurate for local neighborhoods containing 3 or more colors, or situated near depth or surface normal discontinuities. Shadow and light patterns, reflections and specularities cannot be accounted for either.

However, it shows that a detection with a global threshold can only cope with a change of the light intensity (where $a(x, y) = a$). To achieve a better stability, an image normalization based on the locally estimated contrast[1], mean or standard deviation can be applied as preprocessing before the detection. This method is used to gain robust local characteristics in [7, 17]. In [18] normalization and characteristics extraction are achieved simultaneously by means of a normalized convolution. Several thresholds can as well be used like in [1]: the local illumination dependent adaption is performed on the threshold. An overview and a comparison of locally adaptive thresholding methods are given e.g. in [19] for document image analysis. Algorithms which are general enough to be adapted to our case and require low computation times are based on two approaches: either the threshold is adapted based on local measures like e.g. mean and standard deviation, or a clustering method like the k–means or Otsu's [19] is locally applied to compute the threshold based on the local distribution.

To summarize, three different principles can be followed for a better detection stability under lighting variations. A method is proposed for each principle in the next section: in 4.1 a local image normalization, in 4.2 a local threshold adaption, and in 4.3 a local automated threshold selection based on clustering. Detection examples and a comparison are given in section 5.

4 Adaption of the Detection to Local Lighting Conditions

4.1 Local Image Normalization

The first principle for an illumination dependent adaption consists of a local image normalization followed by a detection with a global threshold. To minimize

[1] difference of the maximal and minimal grey values

the effects of normalization artifacts, the normalized convolution proposed in [18] was chosen. The author compared three normalizations based on contrast, variance and energy. We have chosen the energy normalized convolution:

$$output(x,y) = \frac{\sum_{(i,j)\in W} I(x+i,y+j)kernel(i,j)}{\sqrt{\sum_{(i,j)\in W} I(x+i,y+j)^2}\sqrt{\sum_{(i,j)\in W} kernel(i,j)^2}}, \quad (6)$$

as it shows the best behavior in the presence of noise [18].

For interest point detection, the matrix $\mathbf{C}$ is calculated as in (1) with derivatives I_x and I_y obtained by means of normalized convolution. As the two derivation kernels have the same weighting factors, the division by the kernel energy in (6) is here superfluous and can be suppressed. The window for energy summation should have the same size as the derivation kernels (here $7 \approx 6\sigma_{deriv}$). The Harris detector then remains unchanged: the function R is computed according to (2) and its local maxima above T ($T > 0$) are detected. T is a user–defined threshold and should be set to obtain a proper number of interest points. This method will be called Normalized Harris Corner Detector (N–HCD) in the following.

4.2 Local Threshold Adaption

The second method to improve the Harris detector is based on a locally adaptive threshold. From [19] two ideas were retained: (a) homogeneous areas are filtered out to avoid noise–induced false detections, and (b) the threshold is adapted using a local measure reflecting the lighting parameter $a(x,y)$.

The main noise source in modern cameras is photon noise. As a result, R is affected by a signal dependent, approximately multiplicative noise. Even with a locally adaptive threshold, local maxima created by texture in dark areas cannot be distinguished from the ones caused by noise in bright homogeneous regions. To transform this multiplicative noise to additive noise, the processing will be done on the logarithm of R. Hence, the variance of the noise on $\ln(|R|)$ is independent of the image grey values. This can be verified on image series taken with a constant setup, where noise is the only source for pixel variations. A test on the local standard deviation can then be used to filter noise–induced local maxima (step a). As the local mean $\mu(x,y)$ reflects the lighting parameter $a(x,y)$ and is needed for standard deviation estimation, it is used for threshold adaption (step b). The following scheme for interest point detection is proposed:

1. Calculate $R(x,y)$ and $\ln(|R(x,y)|)$ as indicated in (1) and (2).
2. Compute the local mean $\mu(x,y)$ and standard deviation $\sigma(x,y)$ of $\ln(|R(x,y)|)$
3. (x,y) is an interest point
   - if it is a local maxima of R with $R(x,y) > 0$,
   - if $\sigma(x,y) > T_1$, (step a: filtering of noise–induced maxima)
   - if $\ln(R(x,y)) > \mu(x,y) + T_2$. (step b: locally adaptive thresholding)

This algorithm will be referred to as the Adaptive Threshold Harris Corner Detector (AT–HCD). It is equivalent to a local normalization of R with an additional filtering of noise–induced false detections.

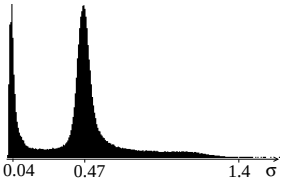

Fig. 2. Histogram of the noise standard deviation.

The threshold T_1 should be adjusted to the noise level on R. It only depends on the camera and on the parameters in (1) and (2). To set T_1 the histogram of the standard deviation of $\ln(|R|)$ on an image series taken with a constant setup can be used, as it represents merely noise effects. As shown in Fig. 2, the histogram contains two peaks: the first one corresponds to textured regions, the second one to homogeneous ones. Higher values occur in areas with a small gradient (e.g. areas where shadows slowly disappear) in which noise effects are amplified. Here T_1 is set to 1.4. The size W of the window used to compute $\mu(x, y)$ and $\sigma(x, y)$ should be chosen such that the local *spatial* standard deviation $\sigma(x, y)$ in homogeneous areas matches the noise standard deviation ($\sigma(x, y)$ is much higher in textured areas). This makes sure that the filtering of homogeneous regions (step a) works properly. For this, several window sizes were tested and the smallest one delivering values consistent with the second peak in Fig. 2 was selected (here $W = 21 \times 21$). W depends only on σ of the Gaussian in (1). The threshold T_2 must be positive and as for the N–HCD should be set by the user to get a proper number of interest points.

4.3 Local Threshold Selection Based on Clustering

The last proposed method involves a clustering algorithm applied on local neighborhoods of the function R [19]. Most methods assume classes with Gaussian distributions. To achieve a better fulfillment of this condition, the logarithm of $|R|$ is used. The k–means algorithm was chosen for clustering as it delivers here the same results as Otsu's method (both methods optimize the a posteriori between–class variance) and is faster (no exhaustive search is performed).

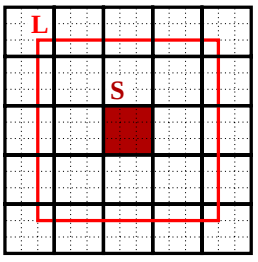

Fig. 3. Discretization for the local thresholding.

To save execution time, the discretization shown in Fig. 3 is adopted: pixels in the large window L are considered for clustering, but the threshold is only applied to the small neighborhood S [19]. The k–means algorithm is initialized with the mean of $\ln(|R|)$ over all pixels. Although this represents the global lighting conditions, it allows a consistent convergence. The pixels inside L are clustered into two classes with the two means μ^{low} and μ^{high}. If a cluster contains no pixel, or if $\mu^{high} - \mu^{low} < T_1$, no significant local maximum is detected in S. Otherwise all local maxima in S such that $\ln(|R_{x,y}|) > \mu^{high} + T_2$ are detected. This method will be referred to as the Local K–Means Harris Corner Detector (LKM–HCD).

The employed window sizes depend on the Gaussian used to calculate R. For L the results of section 4.2 can be used: the repeatability drops when the size of L is smaller than W. In this paper, windows of sizes 25×25 and 5×5 are used. The parameter T_1 represents some kind of required signal–to–noise ratio.

T_2 allows to control the density of detected interest points in textured areas. The influence of both parameters on the repeatability is relatively limited, so that one threshold can be fixed (here $T_1 = 2.5$) and the other can be used to detect an appropriate number of interest points (here T_2). If the application requires the maximum number of interest points, T_1 and T_2 can be set to 0.

5 Experimental Results

To illustrate the detection results, all three methods were applied to the left image of Fig. 1. Their variable threshold was chosen to obtain the same number of interest points as with the global threshold used for Fig. 1.[2] As expected, the points detected with the proposed algorithms are not only located in the most illuminated areas (cf. Fig. 4). Furthermore, the AT–HCD and LKM–HCD allow a relatively even distribution of interest points in the image, whereas the N–HCD tends to detect a few bunches of points. In their current implementation, the N–HCD, AT–HCD and LKM–HCD require respectively about 1.11, 1.30 and 2.08 times as much execution time as a detection with a given threshold.

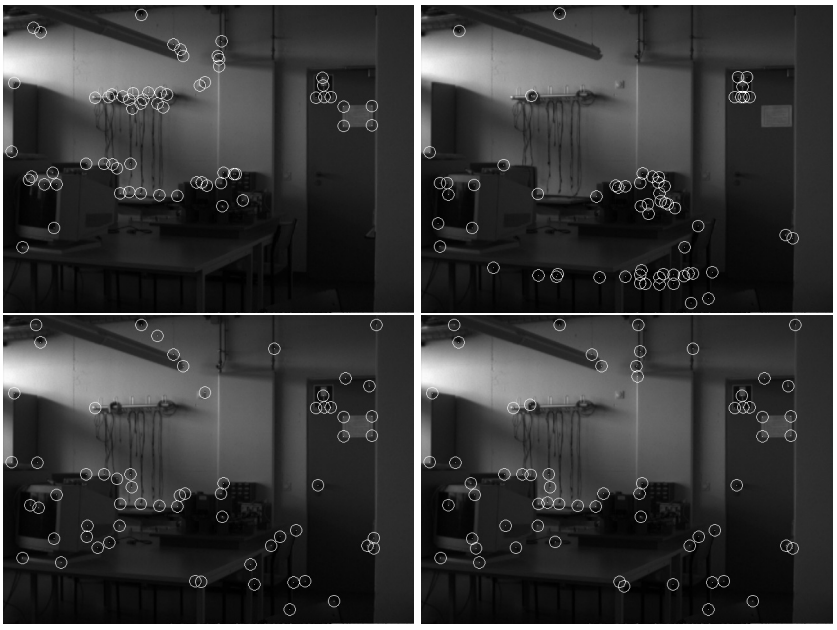

Fig. 4. Detection example for all algorithms. Upper left: global threshold, $T = 0.01 \cdot \max(R)$. Upper right: local image normalization (N–HCD). Lower left: local threshold adaption (AT–HCD). Lower right: local k–means algorithm (LKM–HCD).

[2] 63 interest points are detected. We obtain the following parameters: $T = 0.34$ for the N–HCD, $T_2 = 2.67$ for the AT–HCD, $T_2 = 1.82$ for the LKM–HCD.

To evaluate their stability, the detectors were compared on image series involving different perturbations. The same parameters were used as for the previous experiment. For comparison the results for a Global Threshold $T = 0.01 \cdot \max(R)$ (GT–HCD) and for the selection of the Best N (here 63) Points (BNP–HCD) are given. The performances are measured with the repeatability rate [10], which is here the number of re–detections divided by the number of points in the reference image. A point is considered re–detected if it is within the 8–neighborhood of a reference point. Additionally, the false positive rate is estimated as the number of non–redetected points divided by the number of points in the current image. The consideration of both rates is necessary as the number of detections is not constant. For all series, the reference image is the image in the middle. The results are given in Fig. 5. The table presents the stability for small perturbations due to camera noise (image series taken with a

Small perturbation image series

Methods	Mean repeatability rates over following series:			Mean false positive rates over following series:		
	noise	color	neon	noise	color	neon
GT−HCD	0.9773	0.9326	0.8306	0.0080	0.0943	0.0000
BNP−HCD	0.9841	0.9451	0.9844	0.0231	0.0939	0.0389
N−HCD	0.9077	0.8649	0.9913	0.0062	0.0215	0.0083
AT−HCD	0.9724	0.9503	0.9511	0.0347	0.0792	0.0270
LKM−HCD	0.9533	0.9416	0.9512	0.0276	0.0465	0.0101

In the graphs, repeatability and false positive rates are represented by solid respectively dashed curves.

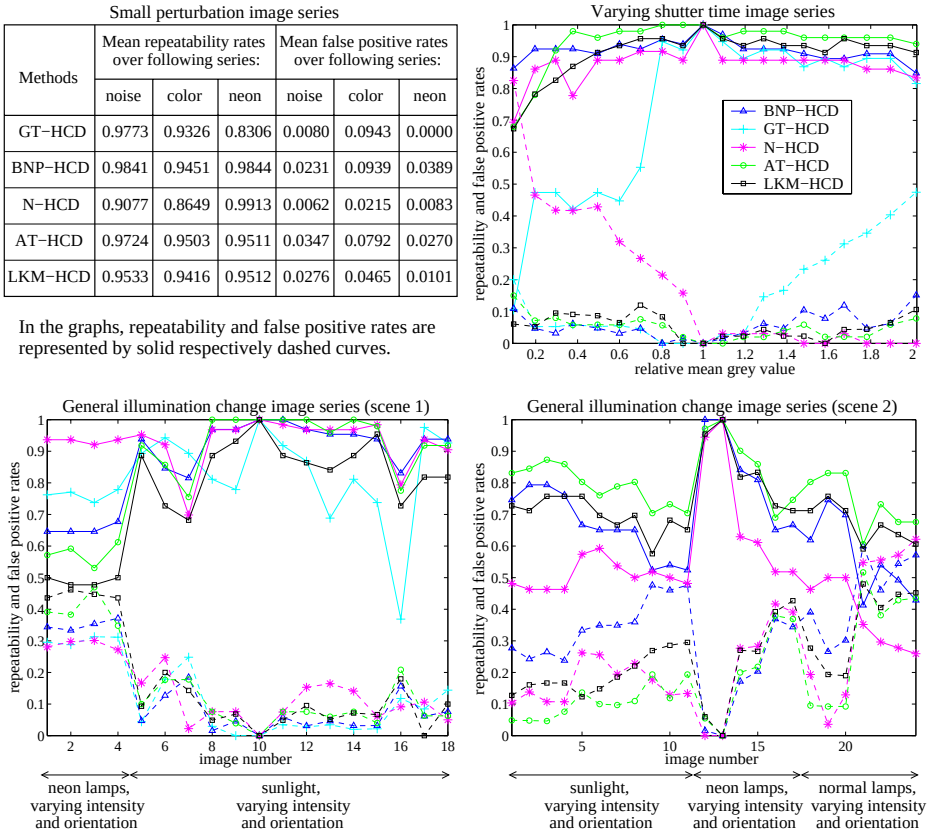

Fig. 5. The table (upper left) gives the mean repeatability and false positive rates under small perturbations. The graphs show the repeatability and false positive rates for image series created by varying the shutter time (upper right) or by changing the type, number, position and orientation of light sources (lower left and right).

constant setup), to global color changes (obtained by varying the white balance parameters of the camera) and to neon lamp flickering. The first graph shows the results for a global intensity change (the shutter time was changed). For the two other graphs, the type, number, position and orientation of the light source(s) were varied (see the graph annotations for more details). In the second graph (scene 2), the results obtained by the GT–HCD have been omitted for better visibility.

All detectors provide good stability for small perturbations. However the N–HCD is more sensitive to noise than all others, as shown in the table and by the many false positives for dark images in the varying shutter time graph. The GT–HCD obtains the worst results: Even global lighting changes may result in many false positives or low repeatability. Consequently, if a global threshold is favored e.g. for efficiency reasons, the best choice is the BNP–HCD as it reaches on the whole a reasonable stability. It provides a constant number of detections, which also can be a drawback depending on image content. The three proposed methods perform similarly to GT–HCD and BNP–HCD for small perturbations or global lighting variations and are more stable under general lighting changes (cf. lower graphs in Fig. 5). The AT–HCD and the LKM–HCD achieve the best results. Their main advantage is an even distribution of interest points in the image, as illustrated by the general illumination change graphs. In the first series (scene 1), the few textured areas in which the N–HCD detected most points were not much affected, so that this one could reach a very good repeatability. In the second graph (scene 2), the N–HCD's results are very poor due to specularities. The AT–HCD and LKM–HCD reach moderate to good stability on both series. Both have the additional advantage that the number of detections varies less with the image content than for N–HCD and GT–HCD. The AT–HCD has slightly better results and requires less computation time.

6 Summary and Outlook

This paper represents a first step to reach a higher stability of interest point detection under varying illumination. The current algorithms use a global threshold, which was shown to be insufficient. Three methods based on different principles are proposed to adapt the Harris corner detector to the local lighting conditions. The N–HCD relies on the calculation of locally normalized image derivatives. The AT–HCD performs a local threshold adaption and additionally filters noise–induced false detections. The LKM–HCD consists of a local k–means algorithm to select the threshold automatically. These methods were compared on image series involving different lighting variations. They enable a better stability under complex illumination changes than detectors using a global threshold. The AT–HCD performs best and furthermore allows a relatively even distribution of interest points in the image. Specular highlights, saturation and light or shadow patterns still represent a problem and decrease the detector stability, as they are not accounted for. This will be subject of further research, e.g. by considering color information. The use of CMOS cameras could also help

to reduce the saturation effects. The N–HCD can be enhanced by filtering noise–induced false detections as in the AT–HCD and by constraining a more uniform distribution of the interest points. Future work should as well consider testing under simultaneous illumination changes and camera motion.

References

1. Molton, N., Brady, M.: Practical structure and motion from stereo when motion is unconstrained. International Journal of Computer Vision **39** (2001) 5–23
2. Vincent, E., Laganière, R.: Matching feature points in stereo pairs: A comparative study of some matching strategies. Machine Graphics & Vision **10** (2001) 237–259
3. Tuytelaars, T.: Local, Invariant Features for Registration and Recognition. PhD thesis, Katholieke Universiteit Leuven (2000)
4. Shi, J., Tomasi, C.: Good features to track. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (1994) 593–600
5. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 530–535
6. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proc. of the International Conference on Computer Vision ICCV, Greece (1999) 1150–1157
7. Carneiro, G., Jepson, A.D.: Local phase–based features. In: Proc. of the European Conference on Computer Vision (ECCV), Copenhagen, Denmark (2002) 282–296
8. Se, S., Lowe, D., Little, J.: Vision-based mobile robot localization and mapping using scale–invariant features. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA), Seoul, Korea (2001) 2051–2058
9. Knapek, M., Swain-Oropeza, R., Kriegman, D.: Selecting promising landmarks. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco, USA (2000)
10. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. International Journal of Computer Vision **37** (2000) 151–172
11. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of the 4th Alvey Vision Conference. (1988)
12. Mokhtarian, F., Suomela, R.: Robust image corner detection through curvature scale space. IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998) 1376–1381
13. Smith, S.M., Brady, J.M.: SUSAN – a new approach to low level image processing. International Journal of Computer Vision **23** (1997) 45–78
14. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Proc. of the European Conference on Computer Vision (ECCV), Denmark (2002) 128–142
15. Hall, D., Leibe, B., Schiele, B.: Saliency of interest points under scale changes. In: Proc. of the British Machine Vision Conference (BMVC), Cardiff, Wales (2002)
16. Gevers, T., Smeulders, A.W.: Color–based object recognition. Pattern Recognition **32** (1999) 453–464
17. Montesinos, P., Gouet, V., Deriche, R., Pelé, D.: Matching color uncalibrated images using differential invariants. Image and Vision Computing **18** (2000) 659–672
18. Schiele, B.: Object Recognition Using Multidimensional Receptive Field Histograms. PhD thesis, Institut National Polytechnique de Grenoble (1997)
19. Trier, O.D., Jain, A.K.: Goal–directed evaluation of binarization methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995) 1191–1201