DETECTORS AND DESCRIPTORS FOR PHOTOGRAMMETRIC APPLICATIONS

Fabio Remondino

Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland
E-mail: fabio@geod.baug.ethz.ch
http://www.photogrammetry.ethz.ch

Commission III

KEY WORDS: Features Detection, Orientation, Precision

ABSTRACT

This paper reports about interest operators, region detectors and region descriptors for photogrammetric applications. Features are the primary input for many applications like registration, 3D reconstruction, motion tracking, robot navigation, etc. Nowadays many detectors and descriptors algorithms are available, providing corners, edges and regions of interest together with n-dimensional vectors useful in matching procedures. The main algorithms are here described and analyzed, together with their proprieties. Experiments concerning the repeatability, localization accuracy and quantitative analysis are performed and reported. Details on how improve to location accuracy of region detectors are also reported.

1. INTRODUCTION

Many photogrammetric and computer vision tasks rely on features extraction as primary input for further processing and analysis. Features are mainly used for images registration, 3D reconstruction, motion tracking, robot navigation, object detection and recognition, etc. Markerless automated orientation procedures based on image features assume the camera (images) to be in any possible orientation: therefore the features should be invariant under different transformations to be re-detectable and useful in the automated matching procedures.

[Haralick and Shapiro, 1992] report these characteristics for a distinctive matching feature: distinctness (clearly distinguished from the background), invariance (independent from radiometric and geometric distortions), interpretability (the associated interest values should have a meaning and possibly usable for further operations), stability (robustness against image noise) and uniqueness (distinguishable from other points).

We should primarily distinguish between feature detectors and descriptors. *Detectors* are operators which search 2D locations in the images (i.e. a point or a region) geometrically stable under different transformations and containing high information content. The results are generally called 'interest points' or 'corners' or 'affine regions' or 'invariant regions'. *Descriptors* instead analyze the image providing, for certain positions (e.g. an interest point), a 2D vector of pixel information. This information can be used to classify the extracted points or in a matching process.

In photogrammetry, interest points are mainly employed for image orientation or 3D reconstruction applications. In vision applications, regions have been recently also employed, for object detection, recognition and categorization as well as automated wide-baseline image orientation.

In the literature different detectors and descriptors have been presented. The achieved results vary, according to the used images and parameters, therefore assesses of the performances are required. Previous works comparing feature point detectors have been reported in [Schmid et al., 1998; Zuliani et al., 2004; Rodehorst and Koschan, 2006]. [Mikolajczyk et al., 2005] compared affine regions detectors while [Mikolajczyk & Schmid, 2003] reported about local descriptors evaluation.

Usually different measures and criterion are used to assess performance evaluations of interest points or regions detectors: for example, given a ground-truth, the geometrical stability of the detected interest points is compared between different images of a given (planar) scene taken under varying viewing conditions.

Selecting the best procedure to compare the operators is very difficult. In our work, the evaluation is performed calculating the number of correct points detected, their correct localization, the density and analyzing the relative orientation results between stereo-pairs. In all the experiments, the results are checked by visual inspection and statistical evaluations. No comparison of the detection speed is performed as difficult to achieve and as the efficiency of a detector (or descriptor) strongly depends on its implementation.

In the context of this work, we only consider points and regions, excluding edges. An overview and comparison of edge detectors is presented in [Heath et al., 1997; Ziou & Tabbone, 1998].

2. POINT AND REGION DETECTORS

2.1 Point detectors

Many interest point detectors exist in the literature and they are generally divided in contour based methods, signal based methods and methods based on template fitting. Contour based detectors search for maximal curvature or inflexion points along the contour chains. Signal based detectors analyze the image signal and derive a measure which indicates the presence of an interest point. Methods based on template fitting try to fit the image signal to a parametric model of a specific type of interest point (e.g. a corner). The main properties of a point detector are: (1) accuracy, i.e. the ability to detect a pattern at its correct pixel location; (2) stability, i.e. the ability to detect the same feature after that the image undergoes some geometrical transformation (e.g. rotation or scale), or illumination changes; (3) sensitivity, i.e. the ability to detect feature points in low contrast conditions; (4) controllability and speed, i.e. the number of parameters controlling the operator and the time required to identify features.

Among the different interest point detectors presented in the literature, the most used operators are afterwards shortly described:

- Hessian detector [Beaudet, 1978]: it calculates the corner strength as the determinant of the Hessian matrix ($I_{xx}I_{yy}-I^2_{xy}$). The local maxima of the corner strength denote the corners in the image. The determinant is related to the Gaussian curvature of the signal and this measure is invariant to rotation. An extended version, called Hessian-Laplace [Mikolajczyk & Schmid, 2004] detects points which are invariant to rotation and scale (local maxima of the Laplacian-of-Gaussian).

- Moravec detector [Moravec, 1979]: it computes an un-normalized local autocorrelation function of the image in four directions and takes the lowest result as the measure of interest. Therefore it detects point where there are large intensity variations in every direction. Moravec was the first one to introduce the idea of 'point of interest'.
- Förstner detector [Förstner, W. & Guelch, E., 1987]: it uses also the auto-correlation function to classify the pixels into categories (interest points, edges or region); the detection and localization stages are separated, into the selection of windows, in which features are known to reside, and feature location within selected windows. Further statistics performed locally allow estimating automatically the thresholds for the classification. The algorithm requires a complicate implementation and is generally slower compared to other detectors.
- Harris detector [Harris & Stephens, 1988]: similar to [Moravec, 1979], it computes a matrix related to the auto-correlation function of the image. The squared first derivatives of the image signal are averaged over a window and the eigenvalues of the resulting matrix are the principal curvatures of the auto-correlation function. An interest point is detected if the found two curvatures are high. Harris points are invariant to rotation. Extended versions of the Harris detector have been presented in [Mikolajczyk & Schmid, 2001; Brown et al., 2005] where the detected points are invariant to scale and rotation.
- Tomasi and Kanade detector [Tomasi & Kanade, 1991]: they developed a features tracker based on a previous work of [Lucas & Kanade, 1981]. Defining a good feature 'the one that can be tracked well', a feature is detected if the two eigenvalues of an image patch are smaller that an empirically computed threshold.
- Haralick operator [Haralick & Shapiro, 1992]: it first extracts windows of interest from the image and then computes the precise position of the point of interest inside the selected windows. The windows of interest are computed with a gradient operator and the normal matrix; the point of interest is determined as the weighted centre of gravity of all points inside the window.
- Heitger detector [Heitger et al., 1992]: derived from biological visual system experiments, it uses Gabor filters to derive 1D directional characteristic in different directions. Afterwards the first and second derivatives are computed and combined to get 2D interest locations (called 'keypoints'). It requires a lot of CPU processing.
- Susan detector [Smith & Brady, 1997]: it analyzes different regions separately, using direct local measurements and finding places where individual region boundaries have high curvature. The brightness of each pixel in a circular mask is compared to the central pixel to define an area that has a similar brightness to the centre. Computing the size, centroid and second moment of this area, 2D interest features are detected.

2.2 Region detectors

The detection of image regions invariant under certain transformations has received great interest, in particular in the vision community. The main requirements are that the detected regions should have a shape which is function of the image transformation and automatically adapted to cover always the same object surface. Under a generic camera movement (e.g. translation), the most common transformation is an affinity, but also scale-invariant detectors have been developed. Generally an interest point detector is used to localize the points and afterwards an elliptical invariant region is extracted around each point.

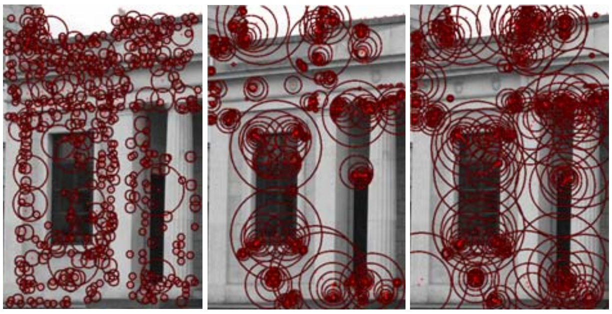

Figure 1: Scale-invariant regions extracted with DoG detector (left) [Lowe, 2004] and affine-invariant regions extracted with Harris-affine (center) and Hessian-affine detector (right) [Mikolajczyk and Schmid, 2002].

Methods for detecting *scale-invariant regions* were presented in [Lindeberg, 1998; Kadir & Brady, 2001; Jurie & Schmid, 2004; Lowe, 2004; Leibe & Schiele, 2004]. Generally these techniques assume that the scale change is constant in every direction and search for local extrema in the 3D scale-space representation of an image (x, y and scale). In particular, the DoG (Difference of Gaussian) detector [Lowe, 2004] showed high repeatability under different tests: it selects blob-like structures by searching for scale-space maxima of a DoG (FIG). On the other hand, *affine-invariant region* detector can be seen as a generalization of the scale-invariant detector, because with an affinity, the scale can be different in each direction. Therefore shapes are adaptively deformed with respect to affinities, assuming that the object surface is locally planar and that perspective effects are neglected. A comparison of the state of the art of affine region detectors is presented in [Mikolajczyk et al., 2005]. The most common affine region detectors are:

- the Harris-affine detector [Mikolajczyk & Schmid, 2002]: the Harris-Laplace detector is used to determine localization and scale while the second moment matrix of the intensity gradient determines the affine neighbourhood.
- the Hessian-affine detector [Mikolajczyk & Schmid, 2002]: points are detected with the Hessian matrix and the scale-selection based on the Laplacian; the elliptical regions are estimated with the eigenvalues of the second moment matrix of the intensity gradient.
- the MSER (Maximally Stable Extremal Region) detector [Matas et al., 2002]: it extracts regions closed under continuous transformation of the image coordinates and under monotonic transformation of the image intensities.
- the Salient Regions detector [Kadir et al., 2004]: regions are detected measuring the entropy of pixel intensity histograms.
- the EBR (Edge-Based Region) detector [Tuytelaars & Van Gool, 2004]: regions are extracted combining interest points (detected with the Harris operator) and image edges (extracted with a Canny operator).
- the IBR (Intensity extrema-Based Region) detector [Tuytelaars & Van Gool, 2004]: it extracts affine-invariant regions studying the image intensity function and its local extremum.

3. DESCRIPTORS

Once image regions (invariant to a class of transformations) have been extracted, (invariant) descriptors can be computed to characterize the regions. The region descriptors have proved to successfully allow (or simplify) complex operations like wide baseline matching, object recognition, robot localization, etc. Common used descriptors are:

- the SIFT descriptors [Lowe, 2004]: the regions extracted with DoG detector are described with a vector of dimension 128 and the descriptor vector is divided by the square root of the sum of the squared components to get illumination invariance. The descriptor is a 3D histogram of gradient location and orientation. It was demonstrated with different measures that the SIFT descriptors are superior to others [Mikolajczyk & Schmid, 2003]. An extended SIFT descriptor was presented in [Mikolajczyk, K. & Schmid, C., 2005]: it is based on a gradient location and orientation histogram (GLOH) and the size of the descriptor is reduced using PCA (Principal Component Analysis).
- Generalized moment invariant descriptors [Van Gool et al., 1996]: given a region, the central moments M^{a}_{pq} (with order $p+q$ and degree a) are computed and combined to get invariant descriptors. The moments are independent, but for high order and degree, they are sensitive to geometric and photometric distortion. These descriptors are suitable for color images.
- Complex filters descriptors [Schaffalitzky & Zissermann, 2002]: regions are firstly detected with Harris-affine or MSER detector. Then descriptors are computed using a bank of linear filters (similar to derivates of a Gaussian) and deriving the invariant from the filter responses. A similar approach was presented in [Baumberg, 2000].

Matching procedures can be afterwards applied between couple of images, exploiting the information provided by the descriptors. A typical strategy is the computation of the Euclidean or Mahalanobis distance between the descriptor elements. If the distance is below a predefined threshold, the match is potentially correct. Furthermore, cross-correlation or Least Squares Matching (LSM) [Gruen, 1985] could also be applied to match the regions (see Section 5) while robust estimators can be employed to remove outliers in the estimation of the epipolar geometry.

4. EXPERIMENTAL SETUP AND EVALUATION RESULTS

Five interest point detectors (Förstner, Heitger, Susan, Harris and Hessian) have been firstly compared with different tests, as described in Section 4.1 and Section 4.2 while in Section 4.3 and 4.4 two region detectors/descriptors (Harris-affine and Lowe) are also considered.
In our work, the evaluation is performed calculating the number of correct corners detected (Section 4.1), their correct localization (Section 4.2), the density of detected points/regions (Section 4.3) and analyzing the relative orientation results between stereo-pairs (Section 4.4). The operators used in the comparison have been implemented at the Institute of Geodesy and Photogrammetry (ETH Zurich), except Harris-affine [Mikolajczyk & Schmid, 2002] and [Lowe, 2004] operators, available on the Internet.

4.1 Corner detection under different transformations

A synthetic image containing 160 corners is created and afterwards rotated, distorted and blurred (Figure 2). Corners are firstly detected with the mentioned operators and then compared with the ground-truth (160).
In Table 1 the numbers of detected corners are presented. Förstner and Heitger performed always better than the other detectors in all the analyzed images.

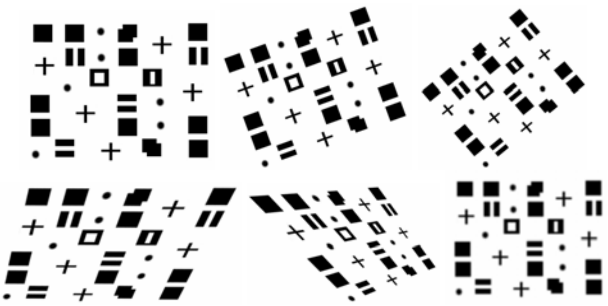

Figure 2: Synthetic images used for the corners detection. The images are numbered left to right from the top-left (1).

	IMAGE 1	IMAGE 2	IMAGE 3	IMAGE 4	IMAGE 5	IMAGE 6 (blur)
Förstner	160/160	159/160	154/160	149/160	145/160	145/160
Heitger	160/160	157/160	158/160	148/160	145/160	148160
Susan	150/160	139/160	118/160	90/160	121/160	141/160
Harris	140/160	139/160	136/160	140/160	121/160	144/160
Hessian	150/160	144/160	142/160	149/160	145/160	140/160

Table 1: Results of the interest point detection on the synthetic images of Figure 1.

4.2 Localization accuracy

The localization accuracy is a widely used criterion to evaluate interest points. It measures whether an interest point is accurately located at a specific location (ground truth). The evaluation requires the knowledge of precise camera and 3D information or simply requires the knowledge of the precise 2D localization of the feature in image space. This criterion is very important in many photogrammetric applications like camera calibration or 3D object reconstruction.
In our experiment, performed on Figure 3 (upper left), the correct corner localizations are achieved with manual measurements. The detected corners obtained from the different operators are afterwards compared with the manual measurements and the differences plotted, as shown in Figure 3. Heitger detector presents only 2 times one-pixel shifts while Harris and Hessian detectors have always a constant shift of one pixel. This might be an implementation problem, but tests performed with other detectors available on the Internet reported the same results.

4.3 Quantitative analysis based on relative orientation between image pairs

Interest points and regions detectors are also used to automatically compute the relative orientation of image pairs. Firstly points (regions) are detected, then matched and finally the coplanarity condition is applied. The correspondences are double-checked, by means of visual inspection and blunder detection (Baarda test and RANSAC estimator), therefore no outliers are present in the data. The extracted points are also well distributed in the images, providing a good input for a relative orientation problem. For each image pair, the same interior orientation parameters are used.

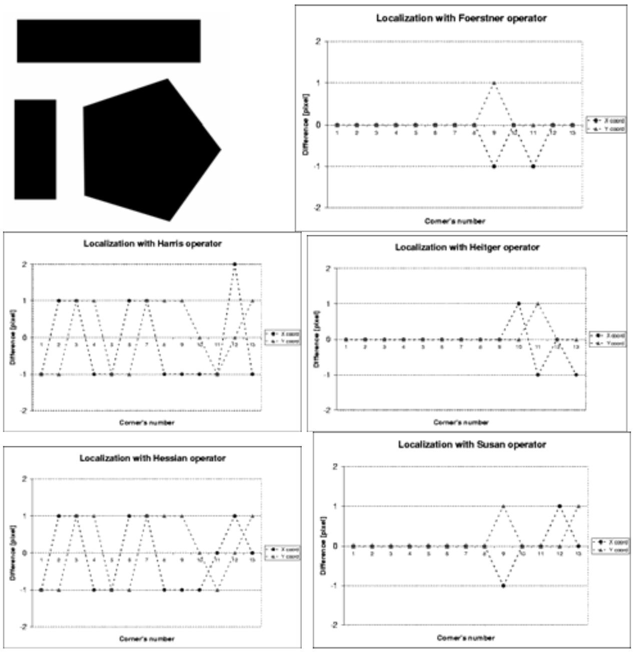

Figure 3: Synthetic image used to evaluate the localization accuracy of the point detectors (upper left). Results of the localization analysis expressed as differences between manual measurements (reference) and automatically detected points.


Figure 4: Two stereo-pairs used for the automated relative orientation computation. Church (1024x768 pixel), Hotel (720x576 pixel).

		CHURCH	HOTEL
Förstner	matched	145	89
	sigma$_0$	0.0183	0.0201
Heitger	matched	133	106
	sigma$_0$	0.0217	0.0207
Susan	matched	127	122
	sigma$_0$	0.0174	0.0217
Harris	matched	184	85
	sigma$_0$	0.0256	0.0425
Hessian	matched	93	91
	sigma$_0$	0.0259	0.0290
Lowe	matched	269	135
	sigma$_0$	0.0341	0.0471
Harris-Affine	matched	139	94
	sigma$_0$	0.0321	0.0402

Table 2: Results of the relative orientation between stereo-pairs in terms of matched points and sigma naught [mm] of the adjustment.

In Table 2 the results of the experiments are reported. To notice the fact that with region detectors (Lowe and Harris-affine operators), the number of matched correspondences is maybe higher but the accuracy of the relative orientation is almost two time worst than with an interest points detector.

5. ACCURACY IMPROVEMENT OF DETECTOR AND DESCRIPTOR LOCATIONS

As shown in section 4.4, region detectors and descriptors provide worst accuracy compared to corners in orientation procedures. The reason might be explained as follow (Figure 5): regions are localized with their centroid and generally matched using the extracted descriptor feature vectors. But, due to perspective effects between the images, the centre of the regions might be slightly shifted, leading to lower accuracy in the relative orientation.

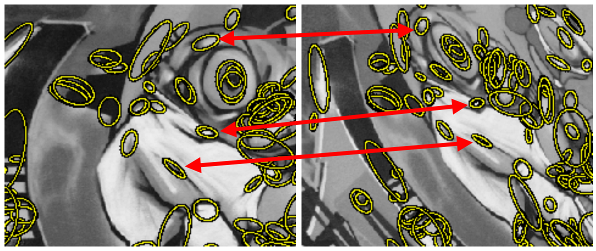
Figure 5: Affine regions detected with Harris detector [Mikolajczyk et al., 2004] with homologues regions. Due to perspective effects, the centre of the regions might be slightly shifted (red arrows).

Affine invariant regions are generally drawn as ellipses, using the parameters derived from the eigenvalues of the second moment matrix of the intensity gradient [Lindeberg, T., 1998; Mikolajczyk, K. and Schmid, C., 2002]. The location accuracy of the region centers can be improved using a LSM algorithm. The use of cross-correlation would fail in case of big rotations around the optical axis and big scale changes, both typical situations in wide baseline images. The ellipse parameters of the regions (major and minor axis and inclination) can be used to derive the approximations for the affine parameters transformation of the LSM. Indeed LSM can cope with different image scale (up to 30%) and significant camera rotation (up to 20 degrees), if good and weighted approximations are used to constraint the estimation in the least squares adjustment.
An example is shown in Figure 6. Given a detected affine region and its ellipse parameters in the template and search image, LSM is computed without and with initial approximations (provided by the region detector), leading to wrong convergence and correct matching results.

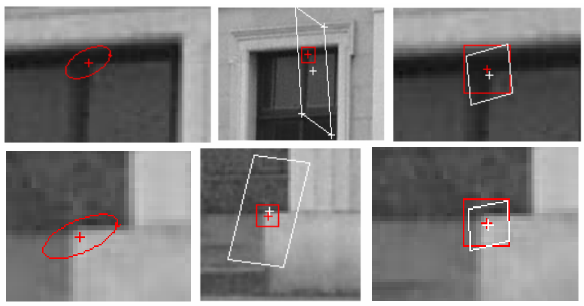
Figure 6: Detected affine region (left). Wrong LSM results with strongly deformed image patch in the search image, initialized with the centroid of the region (centre). LSM result (right) obtained using the approximations derived by the region detector algorithm.

For the church example of Section 4.3, all the extracted Lowe points (regions) were re-located, as previously described, by means of LSM algorithm. The final precision of the relative orientation decreased to 0.0259 mm.

6. CONCLUSIONS

An evaluation and comparison of interest point and region detectors and descriptors has been presented. As the selection of comparison criteria is quite difficult, we tried to used measures and procedures which are typical in photogrammetric applications. Moreover, we showed how to improve to location accuracy of region detectors using a classical least squares measurement algorithm.

From all our tests and results, [Förstner & Guelch, 1987] and [Heitger et al., 1992] operators showed better results than the others examined algorithms. Compared to other evaluation papers, we performed a quantitative analysis of the analyzed point detectors, based on the relative orientation. On the other hand, region detectors and descriptors, as they detect an area and not a single point, reported worst accuracy in the relative orientation problem. In fact they might detect the same region, but the centroid of the region (i.e. the point used to solve for the image orientation) might be shifted due to perspective effects. Nevertheless, they generally provide for affinity invariant parameters, which can be used as approximations for a least squares matching measurement algorithm, which would not converge without good approximations due to the large camera rotations or scale change. Therefore regions could also be good image features for precise and automated orientation procedures, in particular with images acquired under a wide baseline.

As final remark, we should mention that each operator has its own set of parameters which are generally used fix and constant for the entire image. An adaptive parameter selection could help in the optimization of the point selection and distribution.

REFERENCES

Baumberg, A., 2002: Reliable feature matching across widely separated views. Proc. of CVPR, pp. 774-781

Beaudet, P., 1978: Rotationally invariant image operators. Proc. 4th Int. Joint Conference on Pattern Recognition, pp. 579-583

Brown, M. Szeliski, R. and Winder, S., 2005: Multi-image matching using multi-scale oriented patches. IEEE CVPR'2005 Proceedings, Vol. I, pp. 510-517

Förstner, W. and Guelch, E., 1987: A fast operator for detection and precise location of distinct points, corners and center of circular features. ISPRS Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281-305

Gruen, A., 1985: Adaptive least square correlation: a powerful image matching technique. South African Journal of PRS and Cartography, Vol. 14(3), pp. 175-187

Haralik, R.M., 1985: Second directional derivative zero-crossing detector using the cubic facet model. Proceedings of 4th Scandinavian Conference on Image Analysis, pp.17–30

Harris, C. and Stephens, M., 1988: A combined edge and corner detector. Proc. of Alvey Vision Conference, pp. 147-151

Haralick, R.M. and Shapiro, L.G., 1992: Computer and Robot Vision, Addison-Wesley, 630 pp.

Heath, M., Sarkar, S., Sanocki, T. and Bowyer, K.W., 1997: A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms. IEEE Transactions on PAMI, Vol. 19(12), pp. 1338-1359

Heitger, F., Rosenthalter, L., von der Heydt, R., Peterhans, E. And Kuebler, O., 1992: Simulation of neural contour mechanism: from simple to end-stopped cells. Vision Research, Vol. 32(5), pp. 963-981

Kadir, T. and Brady, M., 2001. Scale, saliency and image description. International Journal of Computer Vision, Vol. 45(2), pp. 83–105

Kadir, T., Zissermann, A. and Brady, M., 2004: An affine invariant salient region detector. Proc. of 8th ECCV

Ji, Q. and Haralick, R.M., 1997: Corner detection with covariance propagation, Proc IEEE Conf Computer Vision and Pattern Recognition, pages 362—367

Jurie, F. and Schmid, C., 2004: Scale-Invariant shape features for recognition of object categories. Proc. of CVPR, vol. 02, pp. 90-96

Leibe, B. and Schiele, B., 2004: Scale-Invariant Object Categorization using a Scale-Adaptive Mean-Shift Search. DAGM Pattern Recognition Symposium, Springer LNCS, Vol. 3175, pp. 145-153

Lindeberg, T., 1998: Feature detection with automatic scale selection. International Journal of Computer Vision, Vol. 30(2), pp. 79-116

Lowe, D., 2004: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, Vol. 60(2), pp. 91-110

Lucas, D. and Kanade, T., 1981: An Iterative Image Registration Technique with an Application to Stereo Vision. Int. Joint Conference on Artificial Intelligence, pp. 674-679

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2002: Robust wide baseline stereo from maximally stable extremal regions. Proc. of British Machine Vision Conference, pp. 384-393

Mikolajczyk, K. and Schmid, C., 2001: Indexing based on scale invariant interest points. Proc. of 8th ICCV, pp. 525-531

Mikolajczyk, K. and Schmid, C., 2002: An affine invariant interest point detector. Proc. of 7th ECCV, pp. 128-142

Mikolajczyk, K. and Schmid, C., 2003: A performance evaluation of local descriptors. Proc. of CVPR

Mikolajczyk, K. and Schmid, C., 2004: Scale and Affine Invariant Interest Point Detectors. Int. Journal Computer Vision, Vol. 60(1), pp. 63-86

Mikolajczyk, K. and Schmid, C., 2005: A performance evaluation of local descriptors. Accepted for PAMI

Mikolajczyk, K. Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L. 2005: A comparison of affine region detectors. Int. Journal of Computer Vision

Moravec, H.P., 1979: Visual mapping by a robot rover. Proc. 6th International Joint Conference on Artificial Intelligence, pp. 598-600

Rodehorst, V. and Koschan, A., 2006: Comparison and Evaluation of Feature Point Detectors. Proc. of the 5th International Symposium Turkish-German Joint Geodetic Days

Schaffalitzky, F. and Zisserman, A., 2002: Multi-view matching for unordered image sets. Proc. of ECCV

Smith, S.M. and Brady, J.M., 1997: SUSAN – a new approach to low level image processing. Int. Journal Computer Vision, Vol. 23(1), pp. 45-78

Schmid, C., Mohr, R. and Bauckhage, C., 1998: Comparing and Evaluating Interest Points. Proc. of ICCV'98

Tuytelaars, T. and Van Gool, L., 2004: Matching widely separated views based on affine invariant regions. International Journal of Computer Vision, Vol. 59(1), pp. 61-85

Tomasi, C. and Kanade, T, 1991: Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132

Van Gool, L., Moons, T. and Ungureanu, D., 1996: Affine / photometric invariants for planar intensity pattern. Proc. of 4th ECCV, pp. 642-651

Ziou, D. and Tabbone, S., 1998: Edge Detection Techniques - An Overview. Journal of Pattern Recognition and Image Analysis. Vol. 8, pp. 537-559

Zuliani, M., Kenney, C, and Manjunath, B.S., 2004: A mathematical comparison of point detectors. Proc. of CVPR 'Image and Video Registration Workshop'