A comparison between "old and new" feature extraction and matching techniques in Photogrammetry

Andrea LINGUA, Prof., Politecnico di Torino - DITAG, Italy, andrea.lingua@polito.it
Davide MARENCHINO, PhD, Politecnico di Torino - DITAG, Italy, davide.marenchino@polito.it
Francesco NEX, PhD student, Politecnico di Torino - DITAG, Italy, francesco.nex@polito.it

Abstract: *The development of new photogrammetric systems has changed the user demand. At present, the images acquired by Unmanned Aerial Vehicles (UAV) and Mobile Mapping Technologies (MMT) are far from the normal condition and they also still need to reach reliable results for bad-textured images. The interest point operators and image matching techniques that have traditionally used in Photogrammetry are unable to give good results for these applications. The algorithms that are used in Computer Vision (CV) community could instead assure good results, in terms of number of matched points. For this reason, a comparison analysis between the SIFT operator [1] and traditional photogrammetric feature extraction and matching techniques has been carried out. Many experimental tests on UAV aerial images and MMS terrestrial acquisitions with high geometric distortions (rotation, 3D viewpoint, scale) have been performed, in order to evaluate the reliability of SIFT for automatic homologous point extraction.*

Keywords: *Feature extraction and matching, SIFT, region detectors, tie point extraction, UAV, MMS.*

1. Introduction

Feature extraction and matching techniques play an important role in the digital photogrammetric process. In aerial and close-range photogrammetry, image correspondences are necessary for automatic collimation procedures, such as image orientation, Digital Surface Model (DSM) generation, 3D reconstruction and motion tracking. The techniques developed over the last 25 years allow the automatic extraction of homologous points in normal stereoscopic conditions. On the other hand, these algorithms are unable to give good results if the convergent taking geometry, strong affinity transformations and 3D illumination changes affect the image acquisition. New image acquisition technologies, such as Unmanned Aerial Vehicles (UAV), Mobile mapping Technologies (MMT), Oblique Photogrammetric Cameras, currently provide image acquisitions that are far from the normal condition. Therefore, the development and analysis of other techniques is necessary. The Computer Vision (CV) community has been dealing with these problems since the early 1990s. In CV, features are used in quite varied applications including: wide baseline matching for stereo pairs, model based recognition, texture recognition, robot localization, building panoramas, symmetry detection and object categorization. The techniques developed by the CV community are called region operators. These methods are not sensitive to image scale changes, rotations, affine distortions, or illumination changes. One of the most important operators developed in the CV community is SIFT (Scale Invariant Feature Transform, [1]). SIFT can extract a high number of features that are invariant to the scale, rotation and to the weak affine distortions. Although many publications and researches about feature detectors have been carried out in the CV community, detailed studies about the accuracy of the SIFT operator in new photogrammetric acquisition systems have never been performed. A performance analysis of the SIFT technique has been carried out in this work. The algorithm has been implemented in Matlab code and validated on synthetic images [2]. An accuracy analysis was then performed by means of a comparison with the results supplied by the feature extraction and matching techniques used in Photogrammetry (Forstner operator [3], Cross-Correlation, Least Square Matching [4]). Experimental tests have been carried out on both close-range images and aerial images, with particular attention to the images acquired by the Pelican UAV [5] and the MMS [6] developed by the Politecnico di Torino.

2. Feature extraction and matching (FEM) techniques

In the last few years, the increase in digital cameras has led to the development of automatic techniques for feature extraction and matching. Two main scientific communities have worked in this research field. On one hand, automatic methods have been developed for photogrammetric purposes (tie point extraction, DSM generation, etc.). On the other hand, the automatic extraction of image correspondences has been extensively examined in the CV community for 3D reconstruction and recognition in automatic control and robotic applications.

Photogrammetry

The feature extraction algorithms developed in Photogrammetry are usually called interest operators or point/edge detectors. Interest operators extract salient image features, which are distinctive in their neighbourhood and are reproduced in corresponding images in a similar way [7]. At the same time, interest operators supply one or more characteristics, which can be used during the later image matching. These operators are also called corner detectors, although they do not select just corners, but rather any image location that has large gradients in all directions. Interest point operators are generally divided into Contour Based methods (CB), Intensity Based methods (IB) and Parametric Model Based methods (PMB). IB are the most frequently used operators in photogrammetry: they compute measures that indicate the preference of an interest point directly from the grey values. The Forstner operator [3], and the Harris operator [8] for instance are widely employed in commercial software for automatic tie point extraction.

Edge detectors extract local discontinuities or significant changes in image intensity or texture [9]. They are widely used for segmentation operations (extraction of 3D objects and geometrical discontinuities. The Canny detector [10] is the most frequently used edge detector, due to its high performance and low sensitivity to threshold parameters variation.

The establishment of correspondences between features of two or more images is carried out by means of Image Matching techniques. Different methods (Area Based Matching, Feature Based Matching, and Relational Matching) are developed in Photogrammetry. Nowadays, the Area Based Matching (ABM) methods are the most commonly used. Cross-Correlation (CC) and Least Square Matching (LSM; [4]; [11]) are implemented in commercial software and they provide very good results (sub-pixel accuracy), especially in aerial applications, where the acquisition is near to the normal case.

Computer Vision

The CV community has performed many researches into the feature extraction and matching techniques. The basic goal of these algorithms is the detection of image regions that are invariant under certain transformations [12]. As regards photogrammetric applications, the image acquisition in CV is not necessary forced to the normal condition. Therefore, feature extraction methods, called region detectors, must be invariant under a generic camera movement, such as scale change, rotation, and 3D viewpoint changes. The most investigated geometrical transformation are the scale and the affinity, thus region detectors are classified as scale-invariant (SIFT detector [1]) and affine-invariant detectors (Harris-affine detector and Hessian-affine detector [13]). The word "region" is different from the term point, because region refers to a set of pixels, i.e. any subset of the image, whose boundaries do not have to correspond to changes in image appearance, such as colour or texture [12]. A set of characteristic attributes is computed for each feature, in order to provide a "feature descriptor" (i.e. SIFT descriptor [1]). Descriptors are usually made under certain assumptions regarding the local geometry of the object to be reconstructed and the geometric or radiometric constraints. In some cases, attributes are even provided by region detectors, in other ones, they are computed by the FBM method. Corresponding features from different images are obtained by means of a comparison of the descriptors, which assures good invariance with respect to geometric and radiometric variations.

3. SIFT detector/descriptor

The SIFT operator is one of the most frequently used operators in the region detector/descriptor panorama. It was first thought up by Lowe [1], [14] and it is currently employed in different CV and photogrammetric applications [15], [16].

SIFT extracts image features that are invariant to image scaling and rotation and partially invariant to changes in illumination and 3D camera viewpoints (affine transformation). The features (keypoints) are detected in a Difference of Gaussians (DoG) scale space (Fig. 1-a), which represents the difference of Gaussian convolutions of the original image. A predominant orientation of the radiometric gradients, which assures the invariance to rotations (Fig. 1-b), is assigned to each local maximum of the DoG function. Finally, a "descriptor" is associated to each keypoint. The "descriptor" is a vector of dimension 128 which summarizes the radiometric content of the neighbourhood of the keypoint (Fig. 1-c). The correspondence between two candidate points is found through the evaluation of the minimum distance between the "descriptors". The distances used are the Euclidean or the Mahanalobis ones. A detailed description of the SIFT algorithm can be found in [1].

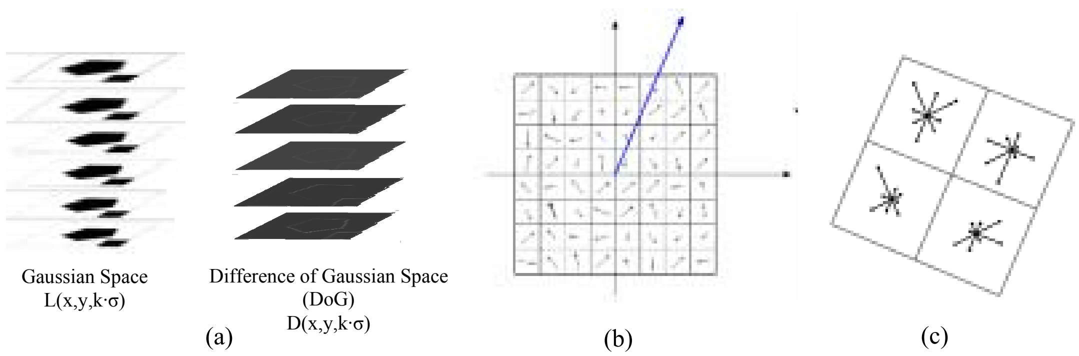


Fig 1. (a). Difference of the Gaussian (DoG) scale space. (b) Predominant orientation of the radiometric gradient. (c) SIFT descriptor.

4. Performance analysis of the SIFT operator in photogrammetric applications

Although SIFT has been developed for CV applications, the high number of features (keypoints) that the region operator can detect and the good performances that can be obtained by the feature matching have led to an increasing interest in the photogrammetric community, especially concerning the close-range field. Nevertheless, only a few researches about the performance of SIFT for photogrammetric applications can be found in literature [17], [12]. Hence, an accurate and reliable assessment of the SIFT performance will be illustrated in the next sections.

The evaluation of the potentiality of the SIFT operator has been carried out through a comparison with the performances of the classical operators used in photogrammetry (Forstner, CC, LSM). Many experimental tests have been performed on both terrestrial and aerial images for this purpose. The tests have concerned images acquired by the “Pelican” UAV and the terrestrial MMT developed by the Geomatic group at the Politecnico di Torino. For this purpose, a complete implementation of the SIFT algorithm, the Forstner operator, the Cross-Correlation and the LSM techniques (ForLSM) has been performed in Matlab code. A robust estimation of the relative orientation between image pairs has been implemented. The Least-Median-Square algorithm (LMS, [18]) has also been employed to detect the gross errors. Geometrical constraints have been introduced for the choice of the points in the relative orientation estimation, in order to assure a good distribution of the homologous points. The congruency of the relative orientation parameters has been checked by a statistical comparison with a set of reference parameters. Therefore, for each test, the relative orientation has been carried out through a set of manual stereoscopic measures, with a well-distributed geometry (according to Von Gruber indications).

The performance of the operators has been evaluated through the analysis of the number of homologous points extracted and the maxima residual parallaxes of the robust relative orientation.

5. UAV applications

First, the algorithms implemented in the Matlab code were validated on synthetic and aerial images [2]. The experimental results have underlined the good performance of the SIFT operator, which assured invariance to rotation and scale changes.

The operators were then applied to images acquired by the Pelican UAV with the RICOH GR (3264x2448 pixel) digital camera. The flight tests, which have been carried out since 2006, highlighted the difficulty of the UAV in the automatic image acquisition. The images are often very far from the normal case, they are rotated and they have 3D viewpoint distortions. They also have drag problems, due to the vibration of the platform.

The SIFT operator was tested on a couple of images of a ploughed field. The images had a very low stereoscopic coverage, and they were badly textured, with a high rate of repetitive patterns. Different tests were performed using as comparison the Forstner operator for the feature extraction and the Cross-Correlation/LSM technique for the automatic collimation and the matching refinement (For-CC-LSM). The total number of matched points (*Mth*); the number of correspondences that are preserved by the LMS estimator (*Mthok*); and the maximum residual parallax (p_{max}) are computed in relation to the rejection thresholds L of the LMS algorithm and the Normalized Cross Correlation (*NCC)* value. As can be seen in Table 1, the maximum number of homologous points extracted is about 100 for ForLSM and more than 450 for SIFT while the p_{max} is less than half a pixel in both cases.

Table 1. The UAV test. The experimental results in relation to the *NCC*, $|D(\hat{x})|$ and *L* parameters.

	FOR-CC-LSM					SIFT				
L	***NCC***	***Mth***	***Mthok***	***%***	p_{max} *(pixel)*	$\mid D(\hat{x})\mid$	***Mth***	***mthok***	***%***	p_{max} *(pixel)*
3	**0.9**	28	16	57.1	0.023	**0.03**	205	30	14.6	noconv
	0.8	182	36	19.7	0.148	**0.01**	745	35	4.7	noconv
5	**0.9**	28	20	71.4	0.067	**0.03**	205	76	37.0	noconv
	0.8	182	79	43.4	noconv	**0.01**	745	87	11.7	0.019
10	**0.9**	28	21	75.0	0.087	**0.03**	205	151	73.7	0.173
	0.8	182	99	54.4	noconv	**0.01**	745	283	38.0	0.082
15	**0.9**	28	23	82.1	0.172	**0.03**	205	177	86.3	noconv
	0.8	182	102	56.0	0.463	**0.01**	745	457	61.3	0.193

Fig. 2 shows the homologous points extracted and matched by these techniques. No features were extracted using ForLSM in the ploughed area, due to the repetitive pattern and the drag problems of the left image. Furthermore, the lack of a good geometric distribution of the matching candidates caused instability in the robust estimation of the relative orientation.
The experimental tests with the SIFT operator instead provided interesting results. The residual parallaxes in fact were lower than half a pixel and, moreover, about 15 points were detected in the ploughed area, where the repetitive patterns and the bad texture made even a manual collimation difficult.

(a) (b) (c) (d)

Fig. 2 (a-b). Homologous points extracted with For-CC-LSM on images acquired by the Pelican UAV over a ploughed field (*NCC*= 0.8 – L= 15). (c-d) Image correspondences extracted with the SIFT operator (|*D(x)*|= 0.01, *L*=15).

These results were obtained changing some of the SIFT detector threshold parameters. The default value of the contrast threshold $|D(\hat{x})|$ of the SIFT detector [1] is too high, therefore many candidates for matching are discarded. In this case, the reduction of the $|D(\hat{x})|$ value from 0.03 to 0.01 allowed us to double the extraction of the image correspondences. This operation is necessary in aerial applications, since, in the original SIFT implementation [13], the threshold values for the detection of the keypoints were set according to the mean texture of close-range images.

However, some residual problems of convergence still remained. They were due to the sensibility of the LMS algorithm to the camera calibration. The RICOH camera has got a retractable lens system, therefore the camera calibration changes for each switch-on/off operation. This problem can be solved by locking the optical system.

These encouraging results led to the application of the SIFT operator for the automatic tie point extraction on images acquired by the Pelican UAV over two archaeological sites in Benevagienna (Piedmont, Italy). In these photogrammetric applications, the camera was calibrated by means of a bundle adjustment self-calibration with the Leica Photogrammetric Suite (LPS) software. This operation allowed the relative orientation parameters to be estimated, and the problems of non-convergence to be reduced. The test on the amphitheatre ruins of Benevagienna (Table 2; Fig. 3) has in fact given good results.

Table 2. The Amphitheatre test. The experimental results of the SIFT operator in relation to the *Th_key* and *L* parameters.

	SIFT				
L	Th_key	Mth	Mthok	%	p_{max} **(pixel)**
5	**0.0005**	1681	75	**4.5**	**0.068**
8	**0.0005**	1681	184	**10.9**	**0.075**
10	**0.0005**	1681	275	**16.3**	**0.088**
15	**0.0005**	1681	599	**44.3**	**0.132**

Almost 600 points were extracted for the high rejection thresholds, with the maximum residual parallaxes lower than half a pixel (Table 2).

These results were obtained by changing another threshold parameter of the SIFT detector. The *Th_key* value, which selects the local maxima in the DoG space was halved (from the default value 0.001 to 0.0005). This setting has allowed matching correspondences to also be estimated in the maize field.

Fig. 3. Photogrammetric survey of the Benevagienna Amphitheatre ruins.

The SIFT operator was applied to a couple of images acquired by the Pelican UAV over the archaeological ruins of the theatre of Benevagienna. In this case, the instability of the UAV during the flight operations did not assure the image acquisition in normal conditions. The two images have a non negligible scale difference, and a weak 3D viewpoint distortion, due to the instability of the UAV, especially in the roll angle. Nevertheless, the SIFT operator provided very good results in the automatic operation of the tie point extraction. About 300 image pairs were extracted and matched (Fig. 4), with a maximum residual parallax of about half a pixel.

Fig. 4. Photogrammetric survey of the Theatre ruins of Benevagienna. Image correspondences automatically matched by SIFT (298 points, *Th_key*=0.0005, *L*=15).

6. Terrestrial applications

The performance analysis of the SIFT operator was performed on terrestrial images with high geometrical distortions. In a first step, the SIFT and For-CC-LSM techniques were tested on a stereoscopic pair acquired in the Misericordia church in Turin (Italy) with a Canon EOS 5D camera (Fig. 5).

The Canon EOS 5D is a professional digital camera with a fixed optical system, which allows stability of the internal orientation parameters. The self-calibration of the camera was performed by the Geomatics research group at the Politecnico di Torino with the *Calibra* software. The acquisition was performed in good illumination conditions, nevertheless the images have an geometrical distortion due to a viewpoint change of about 25°. In this application, the evaluation of the matching methods was performed by fixing to *NCC* and *Th_eu* thresholds to 0.9 and 2 respoectively. The statistical analysis of the estimated angular parameters was then carried out. The experimental results are shown in Table 3.

Table 3. The Misericordia test. The experimental results of the examined operators, in relation to the *NCC*, *Th_eu* and *L* parameters

	FOR-CC-LSM					**SIFT**				
L	***NCC***	***Mth***	***Mthok***	***%***	p_{max} *(pixel)*	***Th_eu***	***Mth***	***Mthok***	***%***	p_{max} *(pixel)*
5	**0.9**	256	98	38.2	5.24	**2**	1500	95	**6.4**	**0.032**
8	**0.9**	256	124	48.4	14.2	**2**	1500	206	**13.8**	**0.091**
10	**0.9**	256	131	51.1	9.41	**2**	1500	303	**20.2**	**0.143**
15	**0.9**	256	144	56.2	45.7	**2**	1500	643	**42.9**	**0.349**

It is possible to notice that SIFT offered excellent performances in term of number of pairs extracted (643), which was almost five times higher than the number of pairs extracted with ForLSM (144). Furthermore, the rate of matched points with respect to the aerial case, was quite stable for SIFT, while it was rather reduced for ForLSM (Fig. 6). Finally, the maximum parallaxes maintained sub-pixel values in SIFT for high rejection thresholds, while were drastically increased in ForLSM. Therefore, it is probable that the preserved data in the robust estimation are affected by outliers. The statistical check of the relative orientation parameters confirmed this assumption. In fact, for rejection thresholds higher than 5, the angular parameters do not satisfy the statistical equality with a reference set.

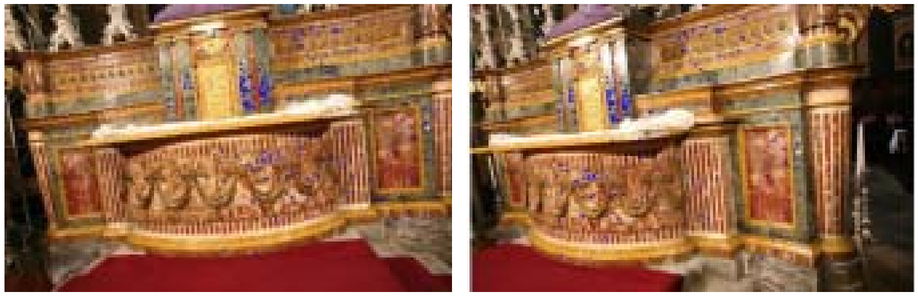
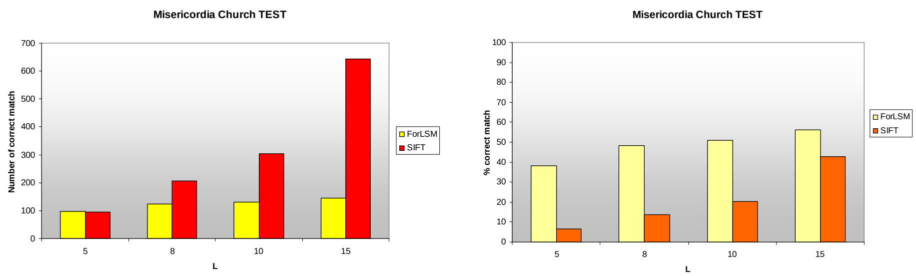


Fig. 5. The Misericordia test. The image correspondences automatically matched with SIFT (303 points, *Th_eu*=2, *L*=10).



(a) (b)

Fig. 6. The Misericordia test. (a) The total number of correct correspondences found with ForLSM and SIFT for different rejection threshold values. (b) Rate of the correct correspondences.

In the Misericordia test, the threshold parameters of the SIFT detector and descriptor were set according to Lowe's implementation [1]. The good results of the matching confirm the suitability of the standard parameters for close-range applications.

The SIFT technique was tested on a set of images acquired by the low-cost MMS which was developed by the Geomatic research group at the Politecnico di Torino [6]. The system is equipped with three Logitech QuickCam Pro 9000 webcams, which can acquire image sequences and stereopairs. The webcams have got a high resolution sensor of 2 Mpixel (1600x1200 pixels) and a focal length of 3.7 mm with an autofocus system. They were calibrated with the *Calibra* software. In this application, the automatic tie point extraction between the image sequences was a critical problem, due to the geometry of the acquisition, the remarkable illumination changes and the movement of the vehicle.

Fig. 7 shows the performance of the SIFT operator on a stereopair acquired by the MMS along a railway overpass. The operator extracted more than 300 homologous points (rejection threshold *L*=10), according to the threshold parameters set in the first UAV test. The images were pre-processed using the Wallis filter [19] in order to improve the local dynamic range. Then, a contrast threshold $|D(\hat{x})|$ equal to 0.01 was adopted. Nevertheless, some problems still remain. First, the low stability of the camera calibration did not allow all the outliers in the LMS approach to be eliminated, therefore a self-calibration of the camera just before the survey is necessary. Moreover, some points which did not belong to the terrestrial scene, such as clouds, or objects in movement were detected.

Fig. 7. Homologous points automatically extracted on a stereopair acquired by the MMS.

7. Conclusions

The aforementioned tests have experimentally proved that SIFT provides very robust and accurate results, which can be compared to the ForLSM ones. The number of points extracted and matched is very high (up to 600 on medium-format cameras) and the accuracy is good enough for quick photogrammetric applications. Moreover, SIFT also assures good performances on image pairs with high geometric and photometric distortions, in which the classical image matching techniques usually fail. In these tests, SIFT has given very good results on the aerial images acquired by the mini-UAV or MMS, which have a bad radiometric content, with a repetitive pattern or a lack of texture. On the other hand, SIFT requires the a priori setting of thresholds parameters. This operation is computationally expensive and it can condition the results. Some empirically estimated thresholds for aerial applications have been proposed in this paper, with particular attention to the images acquired by the Pelican mini-UAV over non-urban areas. Nevertheless, in order to improve the performances of the method, a self-adaptive approach is recommended. Further studies are in underway for this purpose.

The performances of the feature extractor and matching operators also depend on the criterion that is used for the blunder detection. In this case, LMS estimation of the relative orientation makes the selection sensitive to the rejection threshold L, and the interior parameters of the camera. Higher rejection thresholds favour the robustness of SIFT, which can extract a high number of pairs. The high number of extracted pairs will be helpful in the DSM production phase, in which a good approximated DSM could reduce the computational costs and allow removal of a classical coarse-to-fine DSM generation approach [20].

In conclusion, the SIFT operator is a powerful technique for the detection of homologous points in geometric conditions that are far from the normal case. The tests have shown that SIFT provides good results in prohibitive image acquisition conditions, in which classical image matching techniques (ForLSM) usually fail. The SIFT algorithm therefore plays an important role in automatic point extraction and approximate DSM generation of the photogrammetric process, especially in new applications, such for mini-UAVs and MMS.

8. References

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

2. *Lingua A., Marenchino D., Nex F., 2008. L'operatore SIFT per l'orientamento di immagini acquisite con prese non-normali. Proceeding of XII Conferenza Nazionale ASITA, L'Aquila, Italy.*
3. *Forstner W., Gulch E., 1987. A fast operator for detection and precise location of distinct points, corners and circular features. Proc. Inter. Conference on fast Processing of photogrammetric data.*
4. *Ackermann, F., 1984. Digital image correlation: Performance and potential application in photogrammetry. Photogrammetric record 11(64):429-439.*
5. *Bendea H., Boccardo P., Dequal S., Guglieri G., Marenchino D., Tonolo F.G., 2007. Tecniche di navigazione assistita per l'utilizzo fotogrammetrico di UAV (Unmanned Aerial Vehicle). Proceeding of XI Conferenza Nazionale ASITA, Torino, Italy.*
6. *Bendea H., Cina A., De Agostino M., Lingua A., Piras M., 2008. Realizzazione di un GIS stradale con un veicolo rilevatore basso costo. Proceeding of XII Conferenza Nazionale ASITA, L'Aquila, Italy.*
7. *Rodehorst V., Koschan A., 2006. Comparison and evaluation of feature point detectors. Proc. 5th International Symposium Turkish-German Joint Geodetic Days "Geodesy and Geoinformation in the Service of our Daily Life", Technical University of Berlin, ISBN 3-9809030-4-4, Berlin, Germany.*
8. *Harris C., Stephens M., 1988. A combined corner and edge detector. Alvey Vision Conference, pp. 147-151.*
9. *Zhang L., 2005. Automatic Digital Surface Model (DSM) generation from Linear Array Images. Phd. dissertation, ETH Zurich, Switzerland.*
10. *Canny J.F., 1986. A computational approach to Edge Detection. PAMI, vol. 8, No. 6, pp. 679-698.*
11. *Gruen A., 1985. Adaptive Least Square Correlation: a powerful image matching technique. South Africa Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14, no. 3, pp.175-187.*
12. *Remondino F., 2006. Detectors and descriptors for photogrammetric applications. Photogrammetric and computer vision ISPRS symposium, Bonn, Germany.*
13. *Mikolajczyk K., Tuytelaars T., Schmid C., Zisserman A., Matas J., Schaffalitzky F., Kadir T., Van Gool L., 2006. A comparison of affine region detectors. International Journal of Computer Vision. Vol 65, issue 1-2, pp. 43-72.*
14. *Lowe D.G., 1999. Object recognition from local scale-invariant features. International conference in Computer Vision, Curfu, Greece, pp. 1150-1157.*
15. *Heinrichs, M.; Hellwich, O.; Rodehorst, V., 2008. Robust spatio-temporal feature tracking. In Proceedings of XXI ISPRS Congress, Beijing, China.*
16. *Forstner, W.; Steffen R., 2008. On visual real time mapping for Unmanned Aerial Vehicles. In Proceedings of XXI ISPRS Congress, Beijing, China.*
17. *Mikolajczyk K., Schmid C., 2005. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10):1615–1630.*
18. *Rousseuw P. J., Leroy A. M., 1987. Robust regression and outlier detection. John Wiley & Sons, Inc., ISBN 0-471-85233-3.*
19. *Wallis, R., 1976. An approach to the space variant restoration and enhancement of images. In Proceedings of Symosium on Current Mathematical Problems in Image Science, November 1976; Naval Postgraduate School, Monterey CA, USA, 329-340.*
20. *Baltsavias E., 1991. Multiphoto Geometrically Constrained Matching. Phd. dissertation, ETH Zurich, Switzerland, ISBN 3-906513-01-7*