A Modified SIFT Descriptor for Image Matching under Spectral Variations

Sajid Saleem and Robert Sablatnig

Computer Vision Lab, Institute of Computer Aided Automation,
Vienna University of Technology, 1040 Vienna, Austria
{ssaleem,sab}@caa.tuwien.ac.at

Abstract. In multispectral imaging multiple discrete wavelength bands are used to image a scene. The imaging process maps the scene contents to different intensity levels and varies the scene appearance from band to band. This induces intensity variations among the spectral images and effects the performance of SIFT for cross spectral image matching. This paper proposes modifications to the SIFT descriptor in order to improve its robustness against spectral variations. The proposed modifications are based on fact, that edges remain well preserved in multispectral imaging and we can achieve better image matching results by boosting the contribution of local edges in the SIFT descriptor construction process. Therefore, we propose a Local Contrast (Δ) and a Differential Excitation (ξ) function for the construction of SIFT descriptors. The experimental results show, that the performance of Δ-SIFT and ξ-SIFT is superior to standard SIFT for image matching under spectral variations.

Keywords: SIFT, spectral images, interest regions, image matching.

1 Introduction

Multispectral imaging decomposes a scene into multi-band images [2]. The decomposition process generates useful information about the scene contents to solve the visual computing problems related to scene recognition [1], remote sensing [4] and visual surveillance [6] efficiently. The Scale Invariant Feature Transform (SIFT) [7] has been widely used for these applications [1,4]. It extracts keypoints from the images and constructs keypoint descriptors using image gradients. Recent studies show, that intensity differences among the spectral images effect the performance of SIFT [10,12].

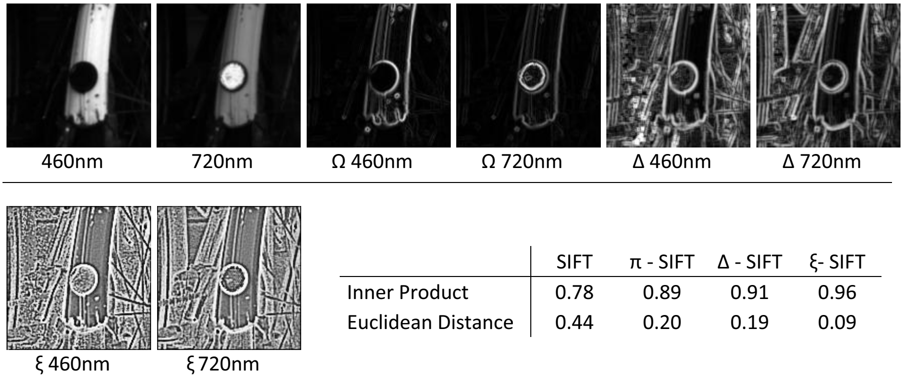
To overcome the effects of such spectral variations Yi et al. propose scale restricted SIFT [13]. They use similar scale SIFT as candidates for descriptor matching. The scale restriction is found efficient in reducing the number of outliers and improves the SIFT performance [13]. In the orientation restricted SIFT (π-SIFT) approach [12] the gradient orientations are mapped to the $[0,\pi]$ range to overcome the intensity reversal problem in the images which in turn improves the SIFT performance against spectral variations. However, it has been found that the performance of SIFT remains low for images where spectral variations are high inspite of such modifications [10].

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 652–661, 2013.

In multispectral images we observe that, edges remain well preserved due to their sharp changes in intensity nature. This motivates us to boost their contributions in the descriptor construction process in order to achieve better robustness for SIFT against spectral variations. Therefore, we propose a Local Contrast (Δ) [11] and a Differential Excitation (ξ) [3] function in this paper to construct modified SIFT descriptors [7]. Each function has a spectral invariant response to local edges and improves the performance of SIFT for spectral images as compared to gradient magnitude (Ω) based SIFT descriptors [7].

The Δ function [11] uses minimum and maximum gray levels to estimate the edge magnitude in a local window, whereas ξ estimates the edge magnitude via the ratio between the local sum of gray level differences to the gray level of the pixel under study [3]. Each function assigns high magnitude weights to edge pixels which in turn cast spectral invariant votes for their corresponding gradient orientation feature histogram bins in the Δ-SIFT and ξ-SIFT descriptors and improves the performance of SIFT for image matching under spectral variations.

To illustrate the significance of Δ and ξ functions two regions of interest from 460nm and 720nm wavelength bands are shown in Figure 1. It can be seen that, only sharp edges are visible in the Ω images whereas the edges in the low contrast regions are suppressed. Also the corresponding edge magnitudes are different which lead to less correlated Ω based SIFT descriptors. In the Δ images the edges are equally enhanced irrespectively to image contrast which makes the Δ-SIFT descriptors more spectral invariant as compared to SIFT. In the ξ images the edge magnitudes are more spectral invariant as compared to Ω and Δ functions which increases the correlation between ξ-SIFT descriptors as suggested by the table which summarizes the cross spectral descriptor matching scores. Our experiments on spectral images of three different scenes suggest that, the performance of ξ-SIFT and Δ-SIFT is superior to Ω based SIFT [7] for image matching under spectral variations.



	SIFT	π - SIFT	Δ - SIFT	ξ- SIFT
Inner Product	0.78	0.89	0.91	0.96
Euclidean Distance	0.44	0.20	0.19	0.09

Fig. 1. Two regions of interest from 460nm and 720nm wavelength bands. The edge magnitude images are obtained via gradient magnitude (Ω), local contrast (Δ) and differential excitation (ξ) functions. The table summarizes the cross spectral descriptor matching score for each SIFT type.

The rest of the paper is organized as follows. In Section 2 we discuss the Ω, Δ and ξ functions for edge magnitude estimation. In Section 3 we describe the experimental setup in detail. Section 4 presents the experimental results and finally, we conclude the paper in Section 5.

2 Edge Magnitude Estimation

In this section we briefly describe the Ω, Δ and ξ functions for edge magnitude estimation. We also discuss their significance in the SIFT descriptor construction process.

2.1 Gradient Magnitude (Ω)

The gradient magnitude function uses local gray level changes to estimate the edge magnitudes i.e, $\Omega = (G_x^2 + G_y^2)^{1/2}$ where G_x and G_y are image gradients along x and y directions respectively using $[-1, 0, 1]$ and $[-1, 0, 1]^t$ filter kernels, where t stands for matrix transpose operation. The Ω based descriptor in this paper is referred to as SIFT [7]. In Figure 1 we can see, that gradient magnitude function is sensitive to local gray level changes and produces high magnitude edges in the high contrast regions whereas the edges in the low contrast regions are suppressed due to small changes in the gray levels.

2.2 Local Contrast (Δ)

The Δ function [11] uses minimum $I_{min}(x, y)$ and maximum $I_{max}(x, y)$ gray levels in a local window around each image pixel $I(x, y)$ to estimate the edge magnitude according to (1).

$$\Delta(x, y) = \frac{I_{max}(x, y) - I_{min}(x, y)}{I_{max}(x, y) + I_{min}(x, y) + \epsilon} \tag{1}$$

where (x, y) represents spatial location of the pixel under study and ϵ is an infinitely small positive number added when $I_{max}(x, y)$ is equal to 0. We use a local window of 3×3 size to compute $I_{min}(x, y)$ and $I_{max}(x, y)$ gray levels. The Δ function produces high magnitude edges irrespectively to image contrast as shown in Figure 1 which suggest that, Δ based edge magnitudes are superior to Ω and result in superior performance for Δ-SIFT under spectral variations.

2.3 Differential Excitation (ξ)

Differential excitation is an edge operator [3]. It estimates edge magnitude in a 3×3 region around every image pixel as described in (2) where $d(x, y)$ is a local sum of gray level differences and $I(x, y)$ is the gray level of the pixel under study. The ξ function simulates the local salient variations similar to human perception [3]. Therefore, the edge pixels receive high magnitude weights via ξ and the non edge pixels are suppressed.

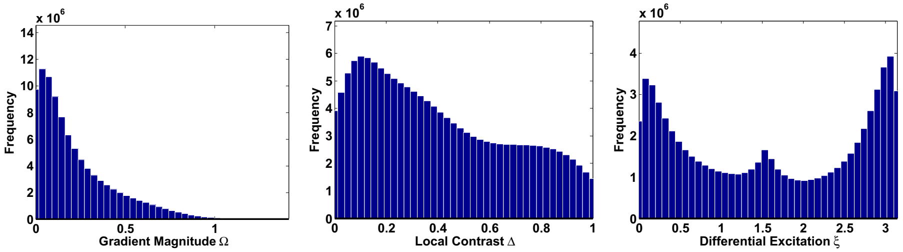
$$\xi(x,y) = \arctan\left(\frac{d(x,y)}{I(x,y)}\right); \quad d(x,y) = -9I(x,y) + \sum_{i=-1}^{i=1}\sum_{j=-1}^{j=1} I(x+i,y+j) \quad (2)$$

The domain range for ξ is $[-\pi/2, \pi/2]$. We map ξ to the $[0,\pi]$ range via $\xi := \pi/2 + \xi$ in order to avoid negative values in the descriptor construction process. The descriptor constructed from ξ in this paper is referred to as ξ-SIFT. It can be seen from Figure 1 that, the ξ function boosts the local edges irrespectively to the wavelength band, which increases the correlation between corresponding ξ-SIFT descriptors and results in superior descriptor matching measures as compared to the SIFT descriptor in the presence of spectral variations.

2.4 Statistics of Edge Magnitudes

The histograms of Ω, Δ and ξ based edge magnitudes are shown in Figure 2. Each histogram is constructed from 12,000 different Harris Laplace (HarLap) regions [5,9] of size 41×41 pixels with gray levels in the normalized [0,1] range. We use HarLap regions in this paper to evaluate the performance of Δ-SIFT and ξ-SIFT. Therefore, the edge magnitude statistics in Figure 2 are useful to understand the distribution and the contribution of each function values in the descriptor construction process.

The histogram of Ω exhibits decaying exponential nature with majority samples in the [0, 0.3] range. This range represent homogenous and low contrast regions where the edges are low in magnitude. The Δ distribution is relatively uniform as compared to Ω because it boosts the edges irrespectively to image contrast. In the ξ case, the majority samples lie close to the function boundaries which represent regions of low gray levels. The edge pixels also belong to such regions because the scene edges diffract the incoming light rays and less of them



(a) (b) (c)

Fig. 2. The histograms of (a) gradient magnitudes (b) local contrast and (c) differential excitations computed from 12,000 different intensity normalized Harris Laplace regions [5,9]

are reflected back towards the camera. The ξ values in the middle represent homogenous regions where the local sum of gray level differences is close to zero in magnitude.

3 Experimental Setup

This section describes the experimental setup in detail. It covers test images, the SIFT descriptor construction process and performance evaluation measures.

3.1 Test Images

We use 460nm and 720nm wavelength band images of three different scenes from Real World Hyperspectral Images (RWHI) dataset[1] as test images. These images are shown in Figure 3. The test images are under different levels of spectral variations which make the cross spectral image matching a challenging problem to solve. Here motivation is to evaluate the performance of SIFT [7] under such intensity variations and then improve its spectral invariant characteristics via Δ and ξ functions.

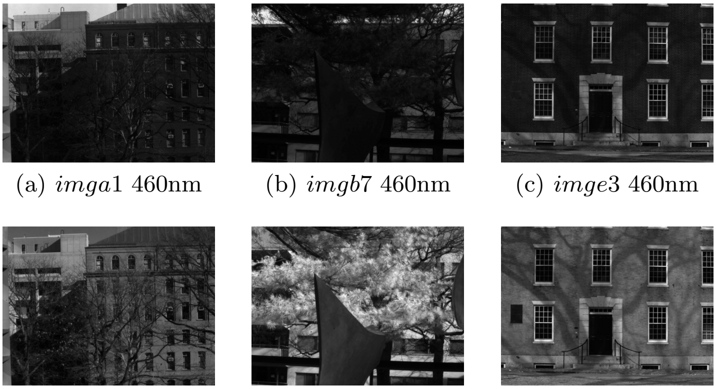
3.2 Interest Regions

We use HarLap regions [9] for cross spectral image matching. The descriptors constructed from HarLap regions are scale invariant. We can also construct rotation invariant descriptors by rotating the regions in the direction of their dominant gradient orientations according to the application requirement. In the experiments we resize each HarLap region to the size of 41×41 pixels and normalize its gray level to the [0, 1] range for descriptor construction [5,8].

3.3 Descriptor Construction

In the descriptor construction stage every HarLap region is split into 4×4 cells and for each cell a gradient orientation feature histogram is constructed [7]. The gradient samples lying inside the cell region cast votes for their corresponding orientation bins. The votes are computed from the product of gradient magnitudes with a Gaussian window. The window uses high weights for samples near to the region center as compared to the region boundary. A soft binning approach is then used to distribute the votes among the adjacent bins to compensate the effect of region shift. Finally, the feature histograms are concatenated over all cells to form a descriptor vector. The vector is then normalized to unit norm and the elements are limited to 0.2 value [7]. At the end, the descriptor vector is renormalized to unit norm. We use Δ and ξ functions for the construction of Δ-SIFT and ξ-SIFT instead of gradient magnitudes whereas SIFT and π-SIFT descriptors are constructed from gradient magnitudes.

[1] http://vision.seas.harvard.edu/hyperspec



(a) *imga*1 460nm (b) *imgb*7 460nm (c) *imge*3 460nm

(d) *imga*1 720nm (e) *imgb*7 720nm (f) *imge*3 720nm

Fig. 3. Spectral images of three different test scenes for performance evaluation of SIFT descriptors

3.4 Evaluation Criteria

The performance evaluation criterion is based on the number of correct and false matches. We use the descriptor matching strategies of [8] for cross spectral image matching. A match is declared where the distance between the descriptor vectors of two interest regions is below a threshold. The ground truth for this match is established via the overlap error [9]. This error measures how well two regions A and B correspond under a known homography H and it is computed from the ratio of intersections to union of the regions i.e, $\epsilon_s = 1 - (A \cap H^T BH)/(A \cup H^T BH)$ [9]. A match is declared as a correct match if $\epsilon_s < 0.5$, otherwise it is considered as a false match. At the end, recall and 1-precision scores are computed using (3) for performance comparison. The *num_correspondences* term represents the number of matching regions ($\epsilon_s < 0.5$) between the images and the *num_all* is the sum of correct (*num_correct*) and false (*num_false*) matches. The perfect descriptor would give a recall value equal to 1 for any precision score [8]. We also compute the area under the recall versus 1-precision curve (AUC) as a single valued measure for performance comparison.

$$recall = \frac{num\_correct\ matches}{num\_correspondences}; \quad 1\text{-}precision = \frac{num\_false\ matches}{num\_all\ matches} \qquad (3)$$

4 Experimental Results

This section presents the experimental results for image matching between the spectral images of the test scenes shown in Figure 3. The image matching uses HarLap regions with SIFT, π-SIFT, Δ-SIFT and ξ-SIFT based descriptor approaches. We use three different descriptor vector matching strategies for image matching [8] i.e, distance threshold, nearest neighbour and distance ratio.

Each matching strategy computes corresponding descriptor matches between the 460nm and 720nm wavelength band images.

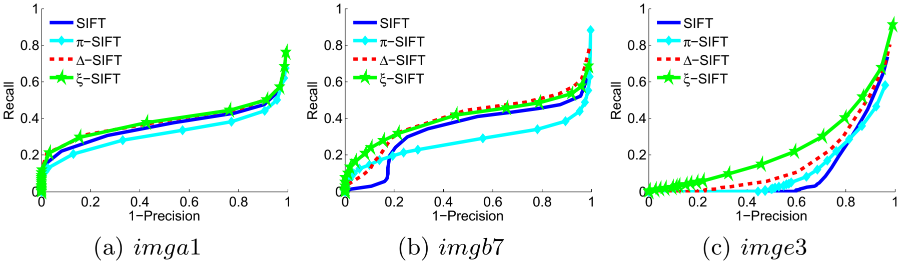
4.1 Discussion on Test Images

The spectral images of *imga*1 scene have variations in illumination as well as spectral variations. However the gray levels of corresponding pixels suggest that illumination variations are dominant. We use normalized HarLap regions (see Section 3.2), therefore, each SIFT approach is illumination invariant and we expect similar performance measures for each of them in this experiment. In the *imgb*7 case, the spectral images appear different due to intensity reversal. This effects the correlation between the corresponding HarLap region descriptors. But there also exists HarLap regions which have only illumination variations, therefore, each SIFT approach also performs better in this experiment. The spectral images of *imge*3 are challenging because most of the corresponding HarLap regions are under spectral variations. This experiment is useful in comparing the evaluation measures for each SIFT type under such spectral variations.

4.2 Distance Threshold Based Matching

In distance threshold (t_d) based matching two interest regions are considered as a match if the distance between their descriptors are less than a threshold [8]. This matching strategy allows several matches for a single query descriptor and several of them may be correct due to $\epsilon_s < 0.5$. The evaluation curves for t_d based matching are shown in Figure 4. The performance measures of each SIFT type are almost similar for *imga*1 due to illumination differences between the spectral images. However, Δ-SIFT and ξ-SIFT perform slightly better than SIFT.

In *imgb*7 the intensity reversal makes the contents of 720nm band image spectrally different from its 460nm band image. This effects the SIFT performance because the corresponding gradient magnitudes are different, however, Δ-SIFT and ξ-SIFT perform relatively better than SIFT especially in the low 1-precision range. In *imge*3 every corresponding HarLap region is under spectral variations



(a) *imga*1 (b) *imgb*7 (c) *imge*3

Fig. 4. Performance comparison of SIFT, π-SIFT, Δ-SIFT and ξ-SIFT using distance threshold (t_d) based matching between the spectral images of Figure 3

which in turn effects the correlation between corresponding descriptors. However, we can see that, the evaluation measures of ξ-SIFT is less effected by such variations as compared to its counterparts. The t_d based image matching suggest, that by boosting the contribution of local edges via Δ and ξ functions we can improve the SIFT robustness against illumination and spectral variations.

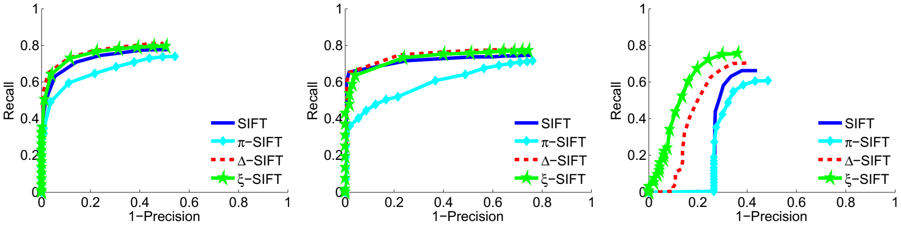
4.3 Nearest Neighbour Based Matching

In nearest neighbour (t_n) based descriptor matching, a nearest neighbor descriptor match is searched in the 720nm band image for each query descriptor of the 460nm band image and a match is declared if distance between the query descriptor and its nearest neighbor is found below a threshold. This matching strategy allows only one match for each query descriptor, which in turn results in better evaluation scores as compared to t_d based matching as shown in Figure 5. This is because the nearest neighbour matching ends up with the correct matches [8]. The t_n based matching suggest that, the performance of ξ-SIFT and Δ-SIFT is superior to SIFT and π-SIFT for image matching under spectral variations.

4.4 Distance Ratio Based Matching

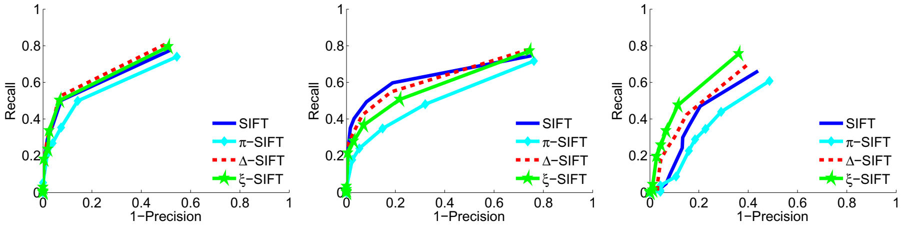
In distance ratio (t_r) based descriptor matching, the distance ratio between the nearest and the second nearest neighbour is computed. If the ratio is below a threshold then a match is declared for the query descriptor. The evaluation curves are shown in Figure 6. The curves suggest, that the SIFT performance is superior in the *imgb*7 case but for the other scenes the ξ-SIFT and Δ-SIFT evaluation measures are superior to SIFT for t_r based image matching.

Table 1 summarizes the area under the t_d, t_n and t_r based recall versus 1-precision evaluation curves (AUC%). The AUC% measures suggest that, the performance of ξ-SIFT on average is superior to other SIFT approaches for t_d and t_n based image matching. However, in t_r based matching SIFT performs better than ξ-SIFT. It means that the nearest and the second nearest neighbour SIFT descriptors are less correlated as compared to ξ-SIFT. From AUC measures we



(a) *imga*1 (b) *imgb*7 (c) *imge*3

Fig. 5. Performance comparison of SIFT, π-SIFT, Δ-SIFT and ξ-SIFT using nearest neighbour (t_n) based matching between the spectral images of Figure 3



(a) *imga*1 (b) *imgb*7 (c) *imge*3

Fig. 6. Performance comparison of SIFT, π-SIFT, Δ-SIFT and ξ-SIFT using distance ratio (t_r) based matching between the spectral images of Figure 3

Table 1. Performance comparison of SIFT, π-SIFT, Δ-SIFT and ξ-SIFT using AUC% measures computed from recall versus 1-precision evaluation curves

(a) *imga*1

	SIFT	π-SIFT	Δ-SIFT	ξ-SIFT
t_d	35.3	31.7	38.5	38.7
t_n	37.4	35.1	37.8	37.8
t_r	31.2	29.8	31.3	31.1

(b) *imgb*7

	SIFT	π-SIFT	Δ-SIFT	ξ-SIFT
t_d	34.1	28.1	40.1	40.0
t_n	54.0	44.8	55.5	54.4
t_r	47.2	37.4	45.1	42.2

(c) *imge*3

	SIFT	π-SIFT	Δ-SIFT	ξ-SIFT
t_d	10.2	10.0	15.6	23.7
t_n	15.4	14.4	16.4	19.2
t_r	17.3	16.1	17.8	18.5

(d) mean

	SIFT	π-SIFT	Δ-SIFT	ξ-SIFT
t_d	26.5	23.3	31.4	34.1
t_n	35.6	31.4	36.6	37.1
t_r	31.9	27.8	31.4	30.6

conclude that ξ and Δ functions improve the SIFT robustness against spectral variations and the idea of boosting local edges in the descriptor construction process produces superior results for image matching.

5 Conclusion

This paper proposes modifications to the SIFT descriptor to improve its performance for image matching under spectral variations. The modifications are based on using the Local Contrast (Δ) and Differential Excitation (ξ) functions instead of gradient magnitudes (Ω) for descriptor construction. Each function produces high magnitude responses to edge pixels and boosts their contributions in the SIFT descriptor construction process. This results in better image matching performance as compared to Ω based SIFT. We validate the proposed Δ-SIFT and ξ-SIFT on the spectral images of three different test scenes.

We use three different descriptor vector matching strategies for image matching i.e, distance threshold (t_t), nearest neighbour (t_n) and distance ratio (t_r). Experimental results show that Δ-SIFT and ξ-SIFT perform better than SIFT for t_d and t_n based image matching whereas the SIFT performance is found superior in t_r based image matching.

Acknowledgments. The first author of this paper was supported by the Vienna PhD School of Informatics, Austria.

References

1. Brown, M., Su, S.: Multi-spectral SIFT for scene category recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 177–184 (2011)
2. Chakrabarti, A., Zickler, T.: Statistics of Real-World Hyperspectral Images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 193–200 (2011)
3. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: WLD: a robust local image descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1705–1720 (2010)
4. Hasan, M., Jia, X., Robles-Kelly, A., Zhou, J., Pickering, M.R.: Multi-spectral remote sensing image registration via spatial relationship analysis on SIFT keypoints. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 1011–1014 (2010)
5. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 511–517 (2004)
6. Leykin, A., Hammoud, R.: Pedestrian tracking by fusion of thermal-visible surveillance videos. Machine Vision and Applications 21(4), 587–595 (2010)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
8. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
9. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision 65(1), 43–72 (2005)
10. Saleem, S., Bais, A., Sablatnig, R.: A performance evaluation of SIFT and SURF for multispectral image matching. In: International Conference on Image Analysis and Recognition, pp. 166–173 (2012)
11. Van Herk, M.: A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. Pattern Recognition Letters 13(7), 517–521 (1992)
12. Vural, M., Yardimci, Y., Temizel, A.: Registration of multispectral satellite images with orientation-restricted SIFT. IEEE International Geoscience and Remote Sensing Symposium 3, 243–246 (2009)
13. Yi, Z., Zhiguo, C., Yang, X.: Multi-spectral remote image registration based on SIFT. Electronics Letters 44(2), 107–108 (2008)