

Interest Region Description Using Local Binary Pattern of Gradients

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. Multispectral imaging system maps the contents of a scene to different intensity levels with in spectral images. This imaging process induces spectral variations among the different wavelength band images of the same scene and results in uncorrelated interest region descriptors for cross spectral image matching. This paper presents Local Binary Pattern of Gradients (LBPG) to improve the strength of interest region description under such spectral variations. In LBPG the image gradients are first transformed into binary patterns and then the gradient patterns are used instead of raw gradients for interest region description. We validate the LBPG approach on the spectral images of six different indoor and outdoor scenes. The experimental results confirm better cross spectral image matching performance as compared to SIFT and Center Symmetric Local Binary Patterns.

Keywords: Image matching, multispectral imaging, interest regions, SIFT and local binary patterns.

1 Introduction

Multispectral imaging system captures the scene at multiple discrete wavelength channels [2]. Each channel images the spectral responses of the scene. These responses are useful in understanding the reflectance phenomena of the scene across the spectrum to solve the visual computing problems efficiently [1,4]. In the remote sensing domain, they have been widely studied [4], however, their usage in other applications such as scene recognition [1] and visual surveillance [6] are now growing rapidly. In these applications, several spectral channel images are fused together to integrate the spectral responses. The fusion process usually uses Scale Invariant Feature Transform (SIFT) [7] to geometrically align the images before integration [4,13]. SIFT extracts scale invariant interest points from the spectral images and describes their surrounding information contents through spatial histograms of image gradients [7]. It has been shown in the recent studies [11,13] that the SIFT performance decreases with the increase in the spectral variations among the images in the multispectral domain.

Several modifications have been proposed to address this problem. For instance in [13] similar scale interest points have been suggested for SIFT descriptor matching. It is because the scale parameter specifies the interest region size

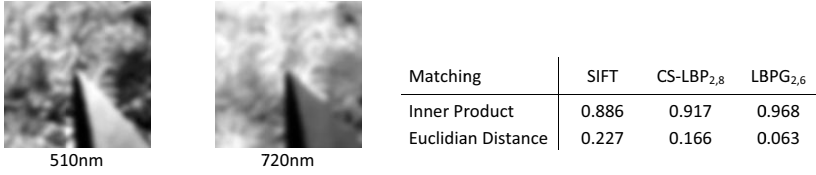


Fig. 1. Interest regions showing same scene contents at 510nm and 720nm spectral channels. SIFT, CS-LBP_{2,8} and LBPG_{2,6} descriptors are constructed for them and then they are matched based on Euclidean distance and Inner product metrics. In the case of perfect match these metrics produce 0 and 1 matching scores respectively.

for description and any variation in scale may lead to a different description. In the orientation restricted SIFT [3,12], the problem of intensity reversal among the spectral images has been addressed by gradient orientation mapping to $(0, \pi)$ radian range prior to description. These modifications improve the performance of SIFT for cross spectral image tasks but still the performance remains low for the images where the spectral variations are high [11].

To elaborate the problem at hand, consider two interest regions in Figure 1. They underlie the same scene contents in 510nm and 720nm spectral channels but the intensity values make the contents visually different. Few pixels have undergone intensity reversal while the other are under illumination variations. We find similar intensity behavior in almost every interest region in this work. This intensity behavior makes the Center Symmetric Local Binary Patterns (CS-LBP) an appropriate description approach for such interest regions due to its gray level invariant texture description nature [5]. It transforms the intensity into binary patterns through illumination invariant process and use them instead of image gradients for interest region description. This description approach has been found robust under illumination variations [5] and our experimental results also confirm its robustness towards spectral variations as compared to SIFT.

Due to such characteristics of CS-LBP, we are using it but in the gradient domain to compute the Local Binary Pattern of Gradients (LBPG). The idea is to transform the gradient magnitude and orientation maps into binary patterns and use such patterns (LBPG) for region description instead of CS-LBP and raw gradients to achieve better image matching under spectral variations. The matching scores of LBPG in Figure 1 show improvement over SIFT and CS-LBP in the cross spectral domain. These scores are computed from descriptor vector matching based on Euclidean distance and Inner product. In the case of perfect match these metrics produce 0 and 1 scores respectively. We use Harris Laplace [9] interest regions in the experiments to evaluate the LBPG approach on the spectral images of indoor and outdoor scenes. The results confirms its better spectral invariant characteristics as compared to SIFT and CS-LBP.

The rest of the paper is organized as follows. In Section 2 the SIFT, CS-LBP and LBPG approaches for region description are briefly discussed. Section 3 describes the experimental setup in detail for the experiments of Section 4. Finally, we conclude the paper in Section 5.

2 Region Descriptors

In this section we describe the SIFT, CS-LBP and LBP_G approaches for interest region description. Each approach uses the difference between the pixel intensities to describe the contents of interest regions.

2.1 SIFT Descriptor

The SIFT descriptor is a function of image gradients and their spatial locations. Its feature histogram consists of 8 gradient orientation bins [7]. Every orientation bin holds a value computed from the sum of the product of respective gradient magnitudes with a Gaussian window overlaid over the description region. The window assigns high weights to the gradient samples near the center as compared to the region boundary. A soft binning approach is used to distribute the gradient samples into adjacent bins to compensate the effect of region shift.

2.2 CS-LBP Descriptor

The CS-LBP is a gray level invariant region descriptor [5]. It transforms the pixel intensities into binary patterns [10] and use them for description instead of image gradients. The CS-LBP process [5] is described in (1) where N equally spaced n_i samples at a radial distance of R from the central pixel n_c are used to transform the gray level of n_c into a binary pattern [5]. The $s(z)$ operator is used to binarize the intensity difference between pixel n_i and its center-symmetric neighbour $n_{i+(N/2)}$. The CS-LBP approach is based on a fact that under variable illumination conditions the magnitude of the pixel difference varies but the sign of this difference is normally preserved [5]. The parameter N produces $2^{N/2}$ distinct binary patterns for description whereas the radial distance is often kept small i.e., $R = \{1, 2\}$ to maintain gray level uniformity among the samples.

$$CS-LBP_{R,N} = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i, \quad s(z) = \begin{cases} 1 & z \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

2.3 LBP_G Descriptor

The LBP_G is an extension of CS-LBP but in the gradient domain. In LBP_G the image gradients are first processed through CS-LBP approach and then the modified gradients i.e., LBP_G are used for interest region description in order to achieve better image matching results under spectral variations.

The LBP_G description process is illustrated in Figure 2. It begins with the estimation of gradient magnitudes (ρ) and orientations (ϕ) for an interest region according to $\rho = (G_x^2 + G_y^2)^{1/2}$ and $\phi = \tan^{-1}(G_y/G_x)$ where G_x and G_y represents image gradients along x and y directions respectively. In the next step ρ and ϕ maps are converted into binary patterns via CS-LBP scheme as described in (2) and (3) where m_i is a gradient magnitude sample at a distance of $2R$ from

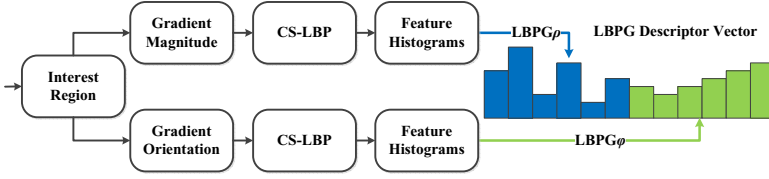


Fig. 2. Illustration of Local Binary Pattern of Gradients (LBP) description process

its center symmetric neighbour $m_{i+(N/2)}$. These two samples are subtracted and their difference is binarized through $s(z)$ operator. This process is repeated for each m_i sample to transform the gradient magnitude sample under study into a binary pattern. Similar description also applies to gradient orientation samples p_i . The parameter N represents the number of samples which produces $2^{N/2}$ binary patterns for description. In the next step the binary patterns of ρ and ϕ are transformed into $LBPGRho$ and $LBPGRho$ descriptors respectively using the SIFT spatial feature histogram scheme [7]. At the end they are concatenated to build an LBP descriptor for the interest region.

$$LBPGRho_{R,N} = \sum_{i=0}^{(N/2)-1} s(m_i - m_{i+(N/2)})2^i \quad (2)$$

$$LBPGRho_{R,N} = \sum_{i=0}^{(N/2)-1} s(p_i - p_{i+(N/2)})2^i, \quad s(z) = \begin{cases} 1 & z \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

3 Experimental Setup

In this section we describe the experimental setup. We discuss the test images, interest regions and evaluation measures for cross spectral image matching.

3.1 Test Images

We use the 510nm and 720nm wavelength channel images from the Real World Hyperspectral Images¹ (RWHI) database. These images are shown in Figure 3 and are used for cross spectral image matching in this paper. The test images are under different level of spectral variations to make image matching a challenging task to accomplish for SIFT, CS-LBP and LBP.

3.2 Interest Region Detection and Normalization

We use scale and rotation invariant Harris Laplace (HarLap) interest points [9] in the experiments. The scale value of each interest point specifies the region size for description. In the experiments we resize every region to 41×41 pixels and normalize its pixel intensity to $(0, 1)$ range prior to description [8].

¹ <http://vision.seas.harvard.edu/hyperspec/>

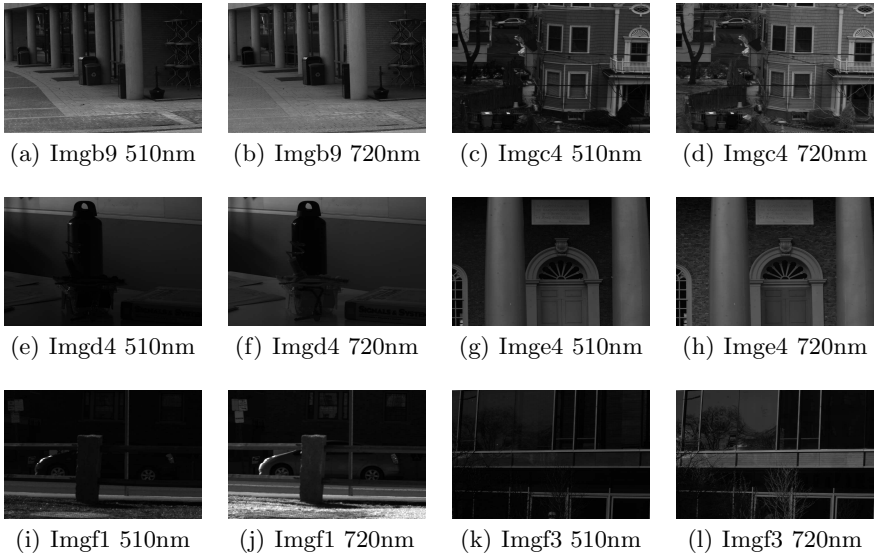


Fig. 3. Selected images from RWHI database of indoor and outdoor scenes

3.3 Region Description

In the region description step, each normalized interest region is split into non overlapping 4×4 cells and for each cell a feature histogram is constructed [7]. The feature histograms are then concatenated over the cells to build a descriptor vector for the interest region. The descriptor is normalized to unit norm and the elements are truncated to 0.2 in order to reduce the impact of large element values in the descriptor matching process. This description scheme is used in the construction of SIFT, CS-LBP and LBPG descriptors in this paper.

In SIFT a feature histogram is consist of 8 gradient orientation bins that results in a $4 \times 4 \times 8 = 128$ dimensional descriptor vector [7]. In the case of CS-LBP_{2,8} [5] the feature histogram is consist of $2^{8/2} = 16$ binary pattern bins which leads to $4 \times 4 \times 16 = 256$ dimensional descriptor. For LBPG we are using $R = 2$ and $N = 6$ parameter values which produce feature histogram of $2^{3/2} = 8$ binary pattern bins and result in $4 \times 4 \times 8 = 128$ dimensional vectors for LBPG _{$\rho_{2,6}$} and LBPG _{$\phi_{2,6}$} descriptors. At the end LBPG _{$\rho_{2,6}$} and LBPG _{$\phi_{2,6}$} are concatenated to produce a 256 dimensional LBPG_{2,6} descriptor vector.

3.4 Evaluation Criteria

In the experiments the performance of SIFT, CS-LBP_{2,8} and LBPG_{2,6} descriptors are evaluated on the basis of cross spectral image matching. The HarLap regions used in this paper are scale invariant, however, if required they can be rotated in the direction of dominant gradient orientations for rotation invariance. The evaluation criteria for each experiment is based on the number of

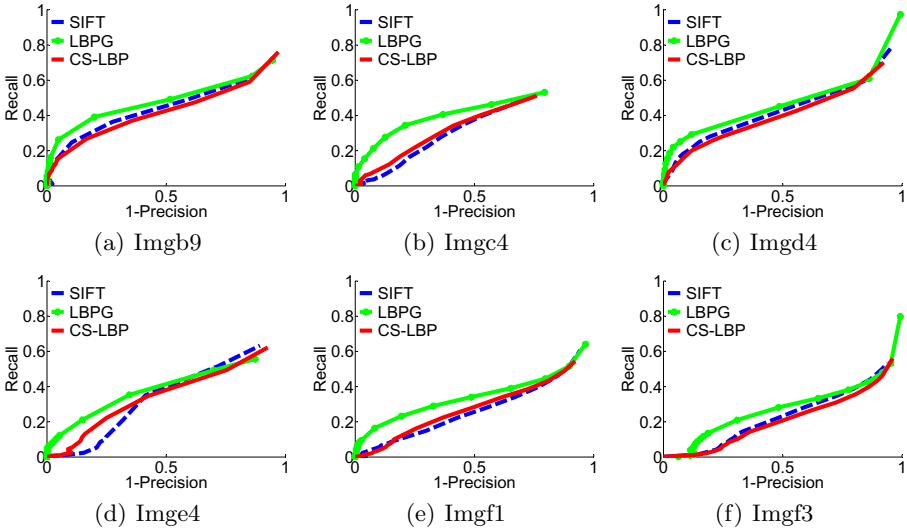


Fig. 4. A performance comparison of SIFT, CS-LBP_{2,8} and LBPG_{2,6} descriptors for image matching between 510nm–720nm channel images of Figure 3. Each approach uses SIFT spatial feature histogram structure for descriptor construction. SIFT descriptor is 128 dimensional whereas CS-LBP_{2,8} and LBPG_{2,6} are 256 dimensional.

correct and false matches. A match is declared correct if the Euclidean distance between the descriptor vectors of two interest regions is below a threshold [8]. The ground truth for a match is established through overlap error [9]. This error tells how well regions A and B correspond under a known homography H [9]. This error is estimated from the ratio of the intersection to the union of the regions i.e., $\epsilon_s = 1 - (A \cap H^T B H) / (A \cup H^T B H)$. A match is assumed to be correct if $\epsilon_s < 0.5$ [9]. This matching strategy allows several matches for a descriptor and several of them may be correct. Therefore, *recall* versus *1-precision* criterion [8] is used for evaluation (4) where $\#correspondences$ stands for the ground truth. The perfect descriptor would give a recall value equal to 1 for any precision score.

$$recall = \frac{\#correct\ matches}{\#correspondences}, \quad 1 - precision = \frac{\#false\ matches}{\#all\ matches} \quad (4)$$

4 Experimental Results

This section presents the experimental results for cross spectral image matching based on the test images of Figure 3. In each experiment, the 510nm and 720nm channel images of a scene is used. The experiment begins with the detection of HarLap interest regions followed by their description using SIFT, CS-LBP_{2,8} and

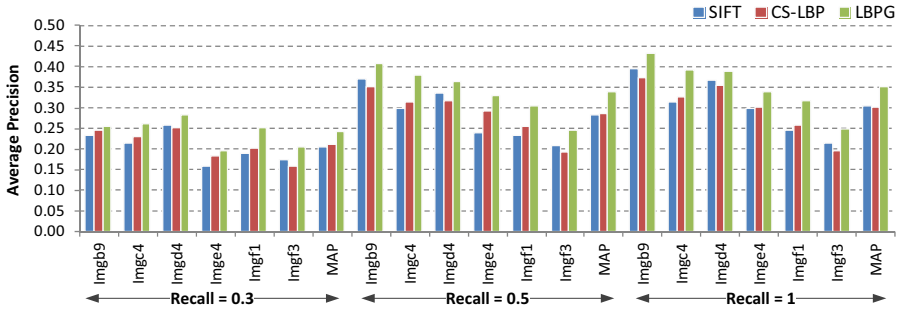


Fig. 5. Comparison of Average Precision measures at three different recall levels

LBP $G_{2,6}$ approaches. The image matching is then carried out by matching every descriptor of 510nm channel image with each descriptor of 720nm channel. At the end Recall versus 1-Precision curves are computed for each descriptor type as shown in Figure 4.

The spectral images of Imgb9 scene are under spectral and illumination differences (see Figure 3). The pixel intensity in some part of 510nm image has undergone reversal in 720nm image. This intensity change adversely affects the descriptor matching score of the regions under such variations. Nevertheless, these regions are less in number as compared to the regions which are under illumination variations. Therefore the evaluation scores are high as shown in Figure 4(a) for each descriptor type. The performance of LBP $G_{2,6}$ is slightly better than SIFT and CS-LBP $_{2,8}$ which suggest its better performance under such spectral and illumination variations.

In Imgc4 case, the spectral images are now under spectral variations. Most of the low intensity pixels in 510nm image are changed into high intensity pixels in 720nm image. This intensity change induces spectral variations among the corresponding interest regions. The low evaluation scores for SIFT and CS-LBP $_{2,8}$ in Figure 4(b) show the affect of such variations on their evaluation scores. However the evaluation measures of LBP $G_{2,6}$ are relatively better which suggest that using processed gradients i.e., LBP G for description instead of raw gradients makes the region description robust towards spectral variations.

In the Imgd4 and Imge4 spectral images the illumination differences are dominant. Under such variations the gradient magnitudes are normally more affected as compared to gradient orientations [5]. The transformation of gradient maps into illumination invariant binary patterns will robustly overcome these illumination variations. The evaluation scores in Figure 4(c) and 4(d) confirm this fact where LBP $G_{2,6}$ performance is found superior to SIFT and CS-LBP $_{2,8}$. These results suggest that LBP $G_{2,6}$ has better illumination invariant characteristics as compared to SIFT and CS-LBP $_{2,8}$.

The next cross spectral image matchings are relatively challenging due to high levels of spectral variations in the Imgf1 and Imgf3 spectral images. The intensity change in these images has altered the appearance and the textural

Table 1. Effect of Gaussian blur on the Average Precision measures of LBPG_{2,6}

σ_s	Descriptor	Imgb9	Imgc4	Imgd4	Imge4	Imgf1	Imgf3	MAP
3	SIFT	0.387	0.303	0.357	0.303	0.240	0.208	0.300
	CS-LBP _{2,8}	0.369	0.319	0.342	0.296	0.255	0.194	0.296
	LBPG _{2,6}	<u>0.425</u>	<u>0.383</u>	<u>0.384</u>	<u>0.334</u>	<u>0.307</u>	<u>0.243</u>	<u>0.346</u>
5	SIFT	0.366	0.272	<u>0.340</u>	0.292	0.228	0.191	0.282
	CS-LBP _{2,8}	0.341	0.291	0.325	0.287	0.241	0.172	0.276
	LBPG _{2,6}	<u>0.383</u>	<u>0.331</u>	0.337	<u>0.325</u>	<u>0.265</u>	<u>0.192</u>	<u>0.306</u>

characteristics of the images under matching. The effect of such variations is evident from the evaluation curves in Figures 4(e) and 4(f). The evaluation scores of LBPG_{2,6} is also dropped in contrast to the previous experiments, nevertheless, still remain significantly better than SIFT and CS-LBP_{2,8}.

The Average Precision (AP) measures of these experiments are shown in Figure 5. They are computed at three different recall levels. For each level Mean Average Precision (MAP) is computed to average the AP over all the test images. The low recall level shows the AP measures for image matching using low Euclidian distance thresholds between the descriptor vectors. AP is a single valued evaluation measure and it is computed from the average value of the precision p as a function of recall r over the interval 0 to 1 as $AP = \int_0^r p(r)dr$. The AP measures also confirms the better performance of LBPG_{2,6} description under spectral variations as compared to SIFT and CS-LBP_{2,8}.

4.1 Effect of Blur

The narrow wavelets range light sources used for the acquisition of a scene at multiple wavelengths channels often produces blurred and low contrast images [2]. In this section we evaluate the LBPG_{2,6} performance under different level of Gaussian blur in the presence of spectral variations. The AP measures for this evaluation are summarized in Table 1. These measures suggest decreases in AP scores with the increase in the standard deviation σ_s of the blur. It is because the image contrast becomes low due to increase in blur level and consequently increases the correlation among the region descriptors. However, the AP measures of LBPG_{2,6} is relatively better than SIFT and CS-LBP_{2,8} in almost every cross spectral image matching under blur.

4.2 Effect of Noise

In CS-LBP gray level uniformity among the samples is considered important for binary patterns [10]. The presence of noise normally affects this intensity uniformity. In this section we evaluate the performance of LBPG_{2,6} against Additive White Gaussian Noise of variance σ_n^2 with spectral variations. The AP measures in Table 2 suggest that noise adversely effects the LBPG_{2,6} performance

Table 2. Effect of Additive White Gaussian Noise on the Average Precision. The pixel intensity range is normalized to (0,1) prior to noise addition of variance σ_n^2 .

σ_n^2	Descriptor	Imgb9	Imgc4	Imgd4	Imge4	Imgf1	Imgf3	MAP
0.006	SIFT	<u>0.379</u>	0.310	<u>0.355</u>	<u>0.291</u>	0.236	0.209	<u>0.297</u>
	CS-LBP _{2,8}	0.352	0.322	0.335	0.280	0.246	0.186	0.287
	LBPG _{2,6}	0.339	<u>0.342</u>	0.293	0.281	<u>0.257</u>	<u>0.217</u>	0.288
0.010	SIFT	<u>0.373</u>	0.307	<u>0.350</u>	<u>0.283</u>	0.235	<u>0.204</u>	<u>0.292</u>
	CS-LBP _{2,8}	0.342	0.321	0.321	0.274	<u>0.241</u>	0.182	0.280
	LBPG _{2,6}	0.300	<u>0.325</u>	0.262	0.270	0.233	0.195	0.264

Table 3. Effect of radius and number of samples on Average Precision of LBPG_{R,N}

Descriptor	Imgb9	Imgc4	Imgd4	Imge4	Imgf1	Imgf3	MAP
LBPG _{1,6}	0.425	0.381	0.381	0.325	0.300	0.239	0.342
LBPG _{2,6}	<u>0.432</u>	<u>0.393</u>	0.388	<u>0.339</u>	<u>0.318</u>	<u>0.250</u>	<u>0.353</u>
LBPG _{3,6}	0.402	0.369	0.379	0.327	0.301	0.232	0.335
LBPG _{2,4}	0.375	0.322	0.352	0.305	0.248	0.193	0.299
LBPG _{2,8}	0.416	0.381	<u>0.390</u>	0.326	0.307	0.235	0.343

as compared to CS-LBP_{2,8}. It is because the CS-LBP is used twice in LBPG description. The SIFT performance is found best under noise due to the distribution of gradient samples in the adjacent bins through soft binning approach.

4.3 Effect of Radius and Number of Samples

The LBPG_{R,N} description depends upon two parameters i.e, radius R and samples N . These values can be varied according to the application requirement. R is normally kept small to maintain intensity uniformity among the samples for CS-LBP scheme [5]. The number of samples N generates $2^{N/2}$ binary patterns for region description. In LBPG_{R,N} the number of feature histogram bins are equal to the number of binary patterns. Therefore increasing N will increase the descriptor length that will in turn increase the computational complexity of descriptor matching process. AP measures of several LBPG_{R,N} descriptors are given in Table 3. These measures suggest that LBPG_{2,6} is best for the test images of Figure 3.

5 Conclusion

In this paper we present a novel interest region descriptor for image matching under spectral variations. The descriptor is constructed from the Local Binary Pattern of Gradients (LBPG) rather than raw image gradients because they vary

across the spectrum due to variation in pixel intensity. We compute the LBPG by transforming the gradient magnitude and orientation maps into binary patterns via illumination invariant CS-LBP coding scheme. The LBPG are then used in the SIFT spatial histogram structure to construct the region descriptors. We use the Harris Laplace interest regions to evaluate LBPG approach on the spectral images of six different indoor and outdoor scenes with different levels of illumination, spectral, blur and noise variations. The experimental results suggest that LBPG performance is superior to SIFT and CS-LBP under blur, illumination and spectral variations. However, in the case of noise the SIFT performance is found better than LBPG.

Acknowledgement. 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. Firmenichy, D., Brown, M., Susstrunk, S.: Multispectral interest points for rgb-nir image registration. In: IEEE International Conference on Image Processing, pp. 181–184 (2011)
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. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. *Pattern Recognition* 42(3), 425–436 (2009)
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. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 971–987 (2002)
11. 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)
12. Vural, M., Yardimci, Y., Temizel, A.: Registration of Multispectral Satellite Images with Orientation-Restricted SIFT. In: Geoscience and Remote Sensing Symposium, vol. 3, pp. 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)