2014 22nd International Conference on Pattern Recognition

A Gradient Extension of Center Symmetric Local Binary Patterns for Robust RGB-NIR Image Matching

Sajid Saleem†, Abdul Bais‡ and Robert Sablatnig†
† Computer Vision Lab, Institute of Computer Aided Automation,
Vienna University of Technology, 1040 Vienna, Austria.
Email: {ssaleem,sab}@caa.tuwien.ac.at
‡Faculty of Engineering and Applied Science, University of Regina, Canada.
Email: abdul.bais@uregina.ca

***Abstract*—Scene acquisition using RGB and Near Infra-Red (NIR) filters generates useful visual information about scene contents. But it induces significant intensity and textural changes between RGB and NIR images of the same scene. It becomes a challenging problem to perform interest point based image matching under such intensity and textural changes. To cope with this problem, a novel method for the description of interest points is proposed. The method proposed is based on Center Symmetric-Local Binary Patterns (CS-LBP) which extracts distinct image features from intensity and gradient magnitude maps of the image patches centered at interest points. Those features are then used in the SIFT algorithm to compute robust descriptors against intensity and textural changes. The experimental results show that the method proposed improves the descriptor matching between RGB and NIR images and achieves better image matching results than CS-LBP and SIFT based methods for the description of interest points.**

I. Introduction

Scene acquisition using Red-Green-Blue (RGB) and Near Infra-Red (NIR) filters generates distinct visual information about scene contents [1]. This information is used in scene category recognition [1], remote sensing [2], biometric authentication [3] and visual surveillance [4] to solve the visual computing problems related to these applications efficiently. But such a scene acquisition induces significant intensity changes between RGB and NIR images and makes RGB-NIR image matching a challenging problem. In Figure 1 an image pair is shown in this regards to illustrate the intensity changes between RGB and NIR images. The scene contents in the image pair shown are same but due to high intensity changes the same scene contents appear significantly different. This also induces significant differences between the interest point descriptors and affects the image matching between RGB and NIR images. To cope with these problems a novel method for the description of interest points is proposed.

Several methods in this regards have been also proposed in the literature which are useful in this work in order to compute robust interest point descriptors for RGB-NIR images. Among these methods, the Scale Invariant Feature Transform (SIFT) [5] has been widely used and found more effective under intensity changes [6], [7]. The key to success for SIFT lies in image gradients which are used as features in the description of interest points. The image gradients not only make SIFT robust to intensity changes but also improve its robustness towards scale, rotation and affine changes [5], [8].

To improve the SIFT performance further, several modifications have been also proposed. For instance, Mikolajczyk and Schmid propose Gradient Location-Orientation Histogram (GLOH) [8] and show better descriptor matching results by computing SIFT on a log-polar location grid instead of cartesian location grid [5]. Heikkilä et al. [9] use the same cartesian location grid of SIFT but achieve robustness against intensity changes by replacing the image gradients in the SIFT algorithm with Center-Symmetric Local Binary Patterns (CS-LBP). Similarly, Wang et al. [10] use Local Intensity Order Patterns (LIOP) as image features in the intensity order based location grid and show better descriptor matching results under intensity changes.

Yi et al. [6] use a Gradient Orientation Modification (GOM) and restrict the domain range of gradient orientations between 0 and π radians to compute GOM-SIFT descriptors under intensity changes. They show an improvement of 7.04% in the correct match rate for GOM-SIFT over SIFT. However, Vural et al., [7] show that GOM on one hand improves SIFT robustness towards intensity changes but on the other hand it effects the rotation invariance. To deal with this problem they propose Orientation Restricted (OR)-SIFT [7]. They compute SIFT descriptors and then combine the elements of SIFT descriptors in the opposite orientation directions to obtain OR-SIFT descriptors. They show better image matching results under intensity reversal and intensity changes than SIFT and GOM-SIFT. Saleem and Sablatnig [11] show that intensity changes result in more gradient magnitude changes than gradient orientation changes. Instead of modifying the

Fig. 1. Illustration shows intensity changes between RGB (left) and NIR (right) images when the same scene is acquired using RGB and NIR filters [1].

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.150

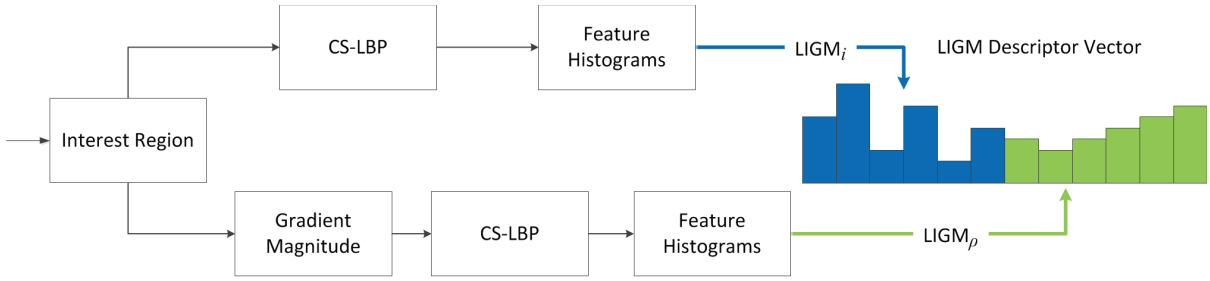


Fig. 2. Illustration shows the description of an interest region based on proposed LIGM method.

gradient orientations, better results are obtained when modification of gradient magnitudes is used. To that end, they propose Local Contrast (LC) magnitudes [12] and Differential Excitation (DE) magnitudes [13] as features instead of gradient magnitudes to compute more robust LC-SIFT [11] and DE-SIFT [11] descriptors under intensity changes.

However, in Local Binary Pattern of Gradients (LBPG) [14] method, it is shown that both components of the image gradients need a modification to cope with intensity changes. For modification, the CS-LBP [9] approach is used which computes LBPG$_\rho$ and LBPG$_\phi$ features from the gradient magnitude (ρ) and gradient orientation (ϕ) maps around the interest points for descriptor construction respectively. It is shown [14] that by using the LBPG features in the SIFT algorithm result in better descriptor matching performance under intensity changes than OR-SIFT and CS-LBP.

Our proposed Local binary patterns of Intensity and Gradient Magnitude (LIGM) method is based on both CS-LBP [9] and LBPG [14] methods. In the method proposed they are used to compute LIGM descriptors in order to achieve better image matching results between RGB-NIR images. To that end, in the method the CS-LBP approach is first applied on the magnitude map of image gradients and LIGM$_\rho$ features are computed according to the LBPG [14] method. Then instead of computing features from the orientation map of image gradients in the method proposed the LIGM$_i$ features are computed by applying the CS-LBP approach on image intensity compared to LBPG [14] method.

The idea is to overcome the instability associated with the LBPG$_\phi$ features [14] which occurs due to the cyclic nature of gradient orientations that makes 0 radian equal to 2π radians near the orientation boundary. This instability results in unstable LBPG$_\phi$ features and effects the description and the descriptor matching of the interest points. However, in the LIGM method proposed such cyclic problems do not occur. To evaluate the performance of LIGM, the RGB-NIR pairs of 52 different scenes are used. The experimental results show better descriptor matching results for LIGM than SIFT, CS-LBP and LBPG methods for image matching between RGB-NIR images.

The rest of the paper is structured as follows. In Section II the LIGM approach is described. In Section III the experimental setup and results are presented. Finally, the paper is concluded in Section IV.

II. Proposed Method

In Figure 2 the proposed LIGM method is illustrated. In the following sections, each building block of the illustration is described.

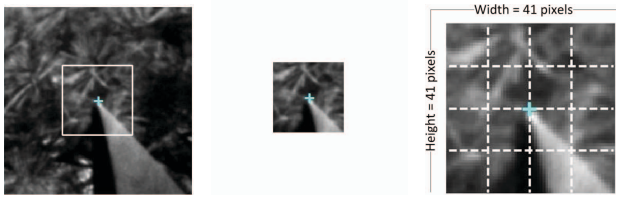
A. Interest Region

Harris Laplace interest points [8] are used in this paper as feature points to compute LIGM descriptors. These feature points are invariant to scale and rotation changes. They are based on Harris function with scale adaption [15]. In the method proposed the scale of each Harris Laplace interest point is used to select an image patch centered at the interest point for descriptor construction. The patch is then intensity normalized and resized to a new region of size 41×41 pixels. This region is referred to as an Harris Laplace (HarLap) interest region [8].

In Figure 3 the computation of an HarLap region is illustrated. An image patch centered at an Harris Laplace interest point '+' is shown inside a bounding box. The patch is cropped, resized and intensity normalized in the method proposed to obtain an HarLap region which is shown in Figure 3(c).

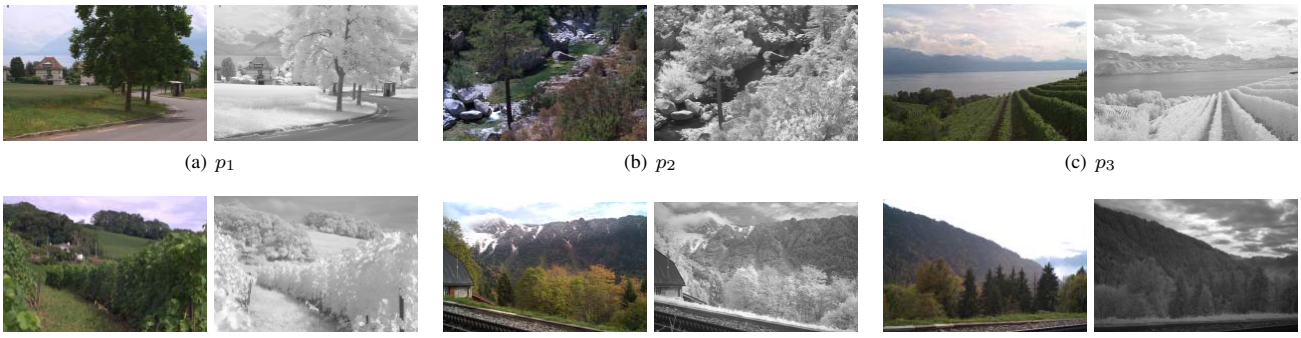
B. CS-LBP

CS-LBP is a gray level invariant texture primitive [9]. It converts the pixel gray levels into binary patterns through a center symmetric pixel subtraction scheme as described in Equation 1 where n_i denotes N number of equally spaced samples at a radius of R from a central pixel n_c. The gray level of each n_i is subtracted from its center-symmetric neighbor $n_{i+(N/2)}$ and the difference is binarized through $s(z)$ function as described in Equation 2. The binarized values are then



(a) (b) (c)

Fig. 3. Illustration shows (a) an image patch centered at a Harris interest point which is depicted with a '+' sign (b) the patch is cropped from the image and (c) it is resized and intensity normalized to obtain a 41×41 pixels HarLap region. In the LIGM method proposed, the HarLap region is then divided into 4×4 location bins [5] for descriptor construction.



(a) p_1 (b) p_2 (c) p_3

(d) p_4 (e) p_5 (f) p_6

Fig. 4. A subset of scenes from evaluated Country Category of the RGB-NIR Scene dataset [1]. For each scene an RGB (left) and an NIR (right) image are shown.

multiplied by 2^i and summed over all $N/2$ center symmetric pairs to obtain a CS-LBP feature for n_c. This process is repeated for each and every image pixel by setting it as a central pixel n_c. The parameter N generates $2^{N/2}$ distinct CS-LBP features and the radius R is usually kept small [9].

$$CS\text{-}LBP_{R,N} = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i \quad (1)$$

$$s(z) = \begin{cases} 1 & \text{z} \geq 0.01 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

In the method proposed, the CS-LBP scheme is used twice. First it is applied on the intensity map of the HarLap region and LIGM$_i$ features are computed. Then the same process is repeated on the gradient magnitude map of the HarLap region and LIGM$_\rho$ features are computed.

C. Feature Histograms

LIGM$_i$ and LIGM$_\rho$ features are then sampled by using the grid like feature histogram scheme of SIFT [5] and two descriptors for an HarLap region are computed i.e. LIGM$_i$ and LIGM$_\rho$. To that end, the HarLap region is divided into 4×4 location bins as illustrated in Figure 3(c). For each location bin a feature histogram is computed by using the LIGM$_i$ features of the location bin and then features histograms over all location bins are concatenated to obtain an LIGM$_i$ descriptor for the HarLap region. The same process is then repeated for LIGM$_\rho$ features and a LIGM$_\rho$ descriptor for the HarLap region is obtained as illustrated in Figure 2.

D. LIGM Descriptor

In this paper, $R = 2$ and $N = 6$ are used in Equation 1. These parameters results in LIGM$_{2,6}$ descriptors for the HarLap regions. The parameter N generates $2^{6/2}$=8 number of bins for each feature histogram and the concatenation of such feature histogram results in a vector of length 4×4×8=128 for both LIGM$_i$ and LIGM$_\rho$ descriptors. By concatenating these two descriptors, a LIGM$_{2,6}$ descriptor vector of length 128+128=256 is obtained for the HarLap region. The LIGM$_{2,6}$ descriptor vector is normalized to unit length, the vector elements are thresholded to 0.2 value and the vector is re-normalized to unit length [5].

III. Experimental Results

This section presents experimental setup and results. All results presented are based on image matching between RGB-NIR images. The image matching is performed between pairs of RGB and NIR images of the same scene using LIGM$_{2,6}$, SIFT, GLOH, CS-LBP$_{2,8}$, LIOP, GOM-SIFT, OR-SIFT, LC-SIFT, DE-SIFT and LBPG$_{2,6}$ descriptors.

A. Image Dataset

In this paper the images of Country Category Scenes (CCS) of the RGB-NIR Scene Dataset[1] [1] are used. This category consists of RGB-NIR image pairs of 52 different scenes. A subset of scenes belonging to CCS is shown in Figure 4 where an RGB-NIR image pair for each scene is shown. Each pair shows high intensity changes between RGB and NIR images.

B. Image Matching

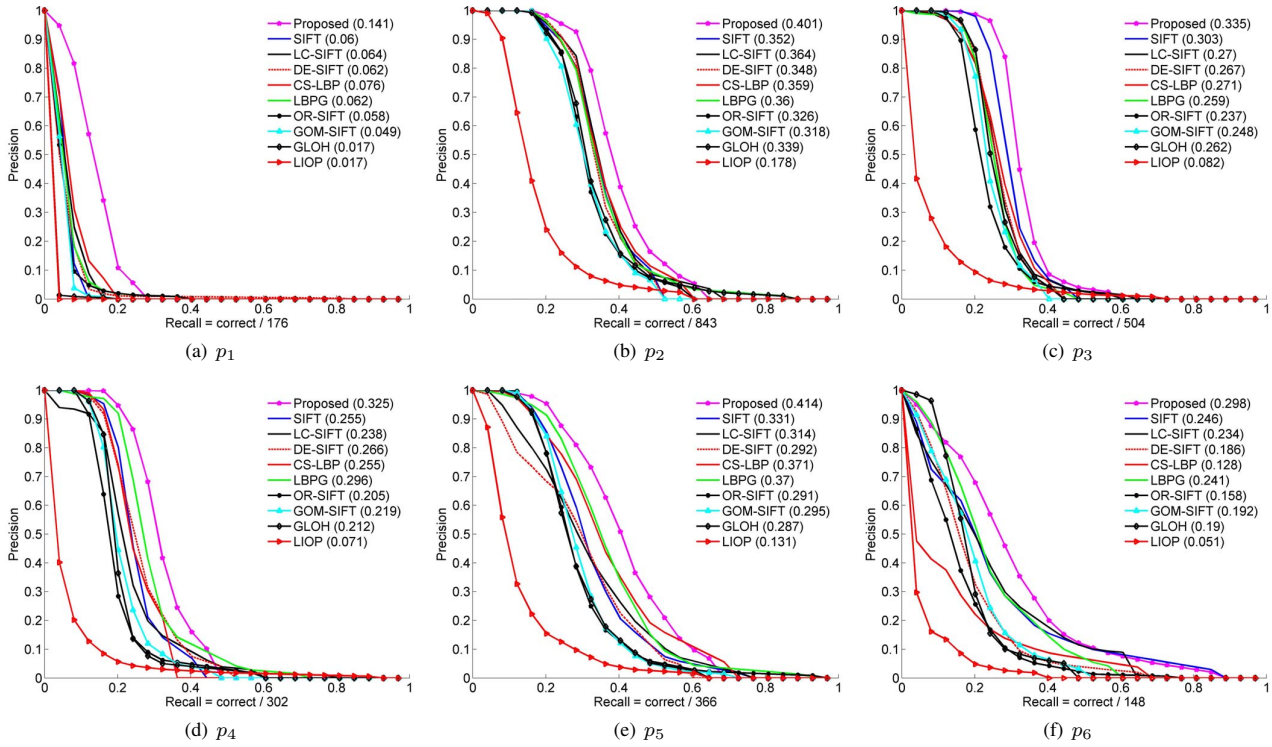
To perform image matching, the gray scale version of the RGB image is used and set as a reference to perform its image matching with the NIR image (target) of the same scene. HarLap regions are computed from reference and target images as described in Section II-A. Then LIGM$_{2,6}$, SIFT, GLOH, CS-LBP$_{2,8}$, LIOP, GOM-SIFT, OR-SIFT, LC-SIFT, DE-SIFT and LBPG$_{2,6}$ descriptors are constructed. The image matching is then carried out by performing descriptor matching between the reference and target images based on three different descriptor matching strategies [8]: (i) distance threshold t_d (ii) nearest neighbor t_n and (iii) distance ratio t_r.

C. Evaluation Criteria

The performance evaluation is based on number of correct and false descriptor matches which are obtained from each descriptor matching strategy in order to compare the performance of LIGM$_{2,6}$ with others. In each descriptor matching strategy, a descriptor match is considered correct or false based on an overlap error (ϵ_s) [8]. This error measures how well two HarLap regions A and B correspond under a known homography H and is computed as ratio of the intersection to the union of the regions [8]:

$$\epsilon_s = 1 - (A \cap H^T BH)/(A \cup H^T BH). \quad (3)$$

[1] http://ivrg.epfl.ch/supplementary_material/cvpr11/index.html



(a) p_1 (b) p_2 (c) p_3

(d) p_4 (e) p_5 (f) p_6

Fig. 5. Results for RGB-NIR image matching for six different scenes which are shown in Figure 4. The AUC value achieved by each descriptor in each image matching is shown inside brackets in the figure legend.

In this paper, $\epsilon_s < 0.5$ [8] is used. Then the following measures based on number of correct and false matches are computed [8]:

$$Precision = \frac{num\_correct\ matches}{num\_all\ matches} \quad (4)$$

$$Recall = \frac{num\_correct\ matches}{num\_correspondences} \quad (5)$$

where $num\_correspondences$ represents the number of corresponding HarLap regions between reference and target images which are obtained by using the $\epsilon_s < 0.5$ criterion. The num_all represents the sum of correct ($num\_correct$) and false matches. In addition to these measures, the Area Under a Precision–Recall Curve (AUC) is also computed and used as a single valued measure for performance comparison.

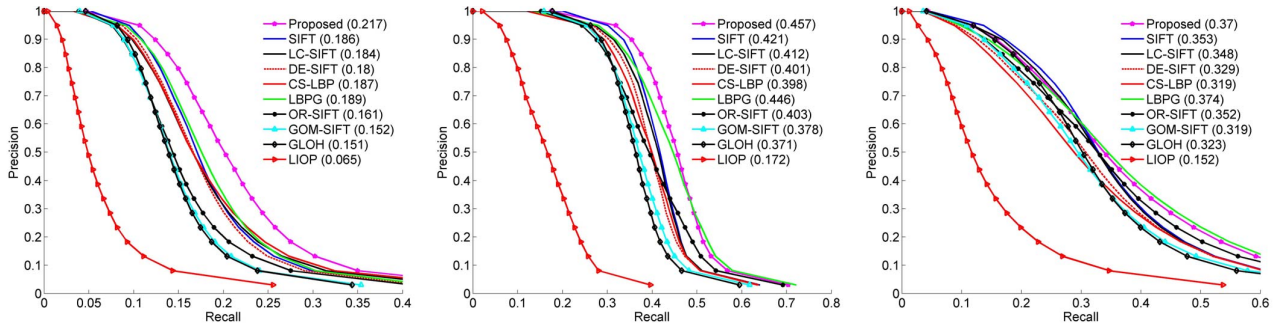
D. Distance threshold based descriptor matching

In distance threshold (t_d) based descriptor matching strategy, a correct match is declared if the Euclidian distance between two descriptors is below a distance threshold and their HarLap regions fulfill the $\epsilon_s < 0.5$ criterion [8]. By changing the distance thresholds a Precision-Recall curve is computed for each descriptor. Please note, t_d is a one to many descriptor matching strategy which allows a descriptor of the reference image to find several matches in the target image and several of them may be correct if they fulfill the t_d based descriptor matching criteria.

In Figure 5 the results for t_d based RGB-NIR image matching are shown. The results show Precision-Recall curves which are generated separately for six different scenes of Figure 4. To understand the results consider the image matching results for p_1 scene which is shown in Figure 5(a). The results are based on 176 number of corresponding HarLap regions which are computed by applying the $\epsilon < 0.5$ criterion on HarLap regions of RGB-NIR images of p_1.

For a point (Recall, Precision)=(0.2, 0.1) on the proposed LIGM$_{2,6}$ descriptor performance curve in Figure 5(a), the number of correct and false matches can be computed as $0.2 \times 176 = 36$ and $36/0.1 - 36 = 324$ respectively. Similarly, the number of correct and false matches can also be computed for other descriptors at any point on their performance curves in Figure 5(a) in order to compare their performance measures with LIGM$_{2,6}$. The results show that LIGM$_{2,6}$ proposed achieves an AUC of 0.141 (shown inside brackets in the figure legend) and outperforms others for matching between RGB-NIR images of p_1. Similarly, other image matching results which are shown in Figure 5 also show better image matching results for LIGM$_{2,6}$ compared to other descriptors.

In Figure 6(a), the average Precision-Recall curves for t_d based image matching are shown when all RGB-NIR image pairs of CCS are considered. The results are based on 443 number of corresponding HarLap regions which represents an average value and is computed by using $\epsilon < 0.5$ for all image pairs of CCS to perform RGB-NIR image matching. The results show better Precision-Recall curve for LIGM$_{2,6}$ proposed compared to others. These t_d based average Precision-Recall measures show that the LIGM features which are used as image features for the construction of LIGM$_{2,6}$ are more robust to intensity changes than image gradients, CS-



(a) t_d (b) t_n (c) t_r

Fig. 6. Average results for RGB-NIR image matching when all RGB-NIR image pairs of the CCS are considered for matching.

TABLE I. AVERAGE AUC(%) VALUES WHEN ALL RGB-NIR IMAGE PAIRS OF THE CCS ARE CONSIDERED FOR MATCHING.

Matching		SIFT	LC-SIFT	DE-SIFT	CS-LBP	LBPG	OR-SIFT	GOM-SIFT	GLOH	LIOP	Proposed
Distance Threshold	t_d	18.6	18.4	18.0	18.7	18.9	16.1	15.2	15.1	06.5	**21.7**
Nearest Neighbor	t_n	42.1	41.2	40.1	39.8	44.6	40.3	37.8	37.1	17.2	**45.7**
Distance Ratio	t_r	35.3	34.8	32.9	31.9	**37.4**	35.2	31.9	32.3	15.2	37.0

LBP and LBPG features and result in better RGB-NIR image matching compared to other descriptors. In Table I, the average AUC (%) values are summarized for t_d based RGB-NIR image matching when all RGB-NIR image pairs of CCS are considered. The results show best average AUC of 21.7% for $\text{LIGM}_{2,6}$ proposed which is 3.1%, 3% and 2.8% better than SIFT, $\text{CS-LBP}_{2,8}$ and $\text{LBPG}_{2,6}$ respectively. The results show the lowest AUC value of 6.5% for LIOP in the t_d based matching. It suggests that the interest point description based on intensity distribution is less effective due to high intensity changes between RGB-NIR images.

E. Nearest neighbor based descriptor matching

In nearest neighbor (t_n) based descriptor matching strategy, a nearest neighbor for each descriptor of the reference image is computed in the target image. Then the match is considered correct if $\epsilon_s < 0.5$ and the Euclidean distance between descriptor and its nearest neighbor is below a distance threshold [8]. By changing this distance threshold Precision-Recall curves are computed. The t_n based matching computes only a single match for each descriptor of the reference image in the target image compared to the t_d based matching.

The average Precision-Recall curves for t_n based matching are shown in Figure 6(b) when all RGB-NIR image pairs of CCS are considered. The results show better Precision-Recall measures for $\text{LIGM}_{2,6}$ proposed compared to others. The average $AUC(\%)$ values which are shown in Table I show 3.6% and 5.9% better results for $\text{LIGM}_{2,6}$ compared to SIFT and $\text{CS-LBP}_{2,8}$. However, the results show a slight performance improvement for $\text{LIGM}_{2,6}$ over $\text{LBPG}_{2,8}$ for t_n based matching. This shows that the nearest neighbors of $\text{LIGM}_{2,6}$ in RGB-NIR image matching are more robust under intensity changes compared to others.

F. Distance ratio based descriptor matching

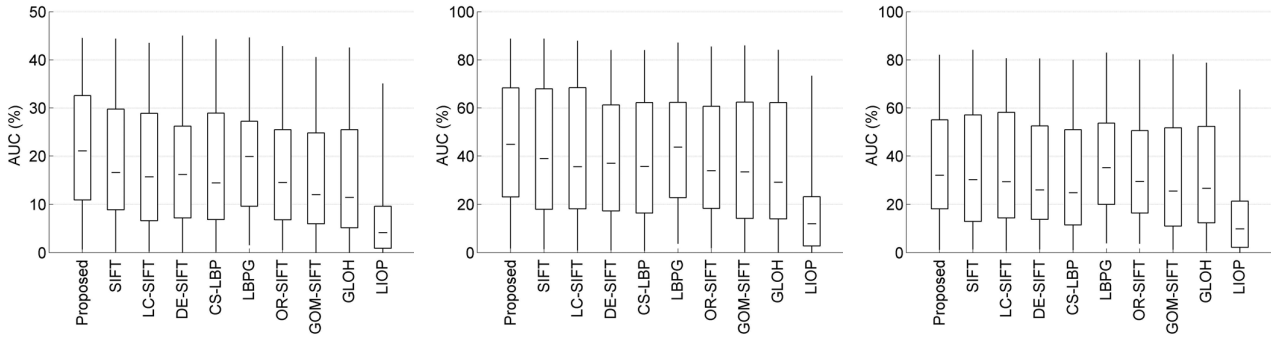
In distance ratio (t_r) based descriptor matching, the distance ratio between nearest and second nearest neighbor is computed. If this ratio is found below a threshold and the $\epsilon_s < 0.5$ criterion is satisfied then the match is declared correct [8]. By changing the distance threshold Precision-Recall curves are computed.

The average results for t_r based RGB-NIR image matching are shown in Figure 6(c). The results show that in t_r based matching SIFT performs better than $\text{LIGM}_{2,6}$ proposed between 0 and 0.3 recall. Afterwards, the performance of $\text{LBPG}_{2,6}$ becomes superior to SIFT and $\text{LIGM}_{2,6}$. The average AUC(%) values in Table I show that $\text{LIGM}_{2,6}$ proposed achieves 1.7% and 5.1% better AUC than SIFT and $\text{CS-LBP}_{2,8}$ in the t_r based matching respectively but compared to $\text{LBPG}_{2,6}$ it shows 0.4% inferior results.

G. Discussion

In Figure 7, the AUC values are plotted when all RGB-NIR image pairs of the CCS are considered for t_d, t_n and t_r based matching. In each plot the median AUC value is shown as '-' whereas the edges of the box represent the 25th and 75th percentile AUC values and the vertical lines are extended to the most extreme cases. From the results shown it can be noted that the $\text{LIGM}_{2,6}$ proposed achieves better results for RGB-NIR image matching than other descriptors. The descriptors based on orientation modification like GOM-SIFT and OR-SIFT do no show better performance compared to $\text{LIGM}_{2,6}$ for RGB-NIR image matching. Also the gradient magnitude modification based descriptors like LC-SIFT and DE-SIFT show lower performance measures compared to $\text{LIGM}_{2,6}$. The results also show the effects of intensity changes on the performance measures of GLOH and CS-LBP.

LIOP which uses intensity based approach for the description of interest points shows the least performance compared to intensity difference (image gradient) based descriptors. $\text{LBPG}_{2,6}$ comparatively shows better results which suggests that modification of both components of image gradients i.e. magnitude and orientation generates more robust image



(a) t_d (b) t_n (c) t_r

Fig. 7. AUC based results when all RGB-NIR image pairs of CCS are considered for image matching based on distance threshold t_d, nearest neighbor t_n and distance ratio t_r based descriptor matching strategies. In each plot the median AUC value is depicted with '-', the edges of the box represent the 25th and 75th percentile values and the vertical lines are extended to the most extreme cases.

features which results in better RGB-NIR image matching compared to others. In this paper this matching performance is further improved by using the proposed LIGM method which combines the CS-LBP features with its gradient magnitude extensions in order to compute robust interest point descriptors to achieve better RGB-NIR image matching results.

IV. Conclusion

A novel method for the description of interest points is proposed. The method proposed is based on Center Symmetric Local Binary Patterns (CS-LBP) to cope with high intensity changes between RGB-NIR images. Such intensity changes effect the performance of SIFT and CS-LBP and result in low image matching results. To deal with these problems LIGM features are proposed which are used in the SIFT algorithm instead of image gradients to compute robust LIGM descriptors. To evaluate the performance of LIGM, image matching was performed on RGB-NIR image pairs of 52 different scenes based on three different descriptor matching strategies: distance threshold, nearest neighbor, and distance ratio. The experimental results showed better image matching results for LIGM in each descriptor matching strategy compared to SIFT and CS-LBP.

Acknowledgment

The financial support provided by the Vienna PhD School of Informatics, Austria is gratefully acknowledged.

References

[1] M. Brown and S. Susstrunk, "Multi-spectral SIFT for scene category recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 177–184, 2011.

[2] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2196–2212, 2012.

[3] Z. Guo, D. Zhang, L. Zhang, and W. Liu, "Feature band selection for online multispectral palmprint recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1094–1099, 2012.

[4] A. Leykin and R. Hammoud, "Pedestrian tracking by fusion of thermal-visible surveillance videos," *Machine Vision and Applications*, vol. 21, no. 4, pp. 587–595, 2010.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[6] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electronics Letters*, vol. 44, no. 2, pp. 107–108, 2008.

[7] M. Vural, Y. Yardimci, and A. Temizel, "Registration of multispectral satellite images with orientation-restricted SIFT," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 243–246, 2009.

[8] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[9] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, 2009.

[10] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," *International Conference on Computer Vision*, pp. 603–610, 2011.

[11] S. Saleem and R. Sablatnig, "A modified SIFT descriptor for image matching under spectral variations," *International Conference on Image Analysis and Processing*, pp. 652–661, 2013.

[12] B. Su, S. Lu, and C. L. Tan, "Binarization of historical document images using the local maximum and minimum," *IAPR International Workshop on Document Analysis Systems*, pp. 159–166, 2010.

[13] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: a robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.

[14] S. Saleem and R. Sablatnig, "Interest region description using local binary pattern of gradients," *Scandinavian Conference on Image Analysis*, pp. 468–477, 2013.

[15] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," *International Conference on Computer Vision*, vol. 1, pp. 525–531, 2001.