LOCAL SELF-SIMILARITY FREQUENCY DESCRIPTOR FOR MULTISPECTRAL FEATURE MATCHING

Seungryong Kim, Seungchul Ryu, Bumsub Ham, Junhyung Kim, and Kwanghoon Sohn

Digital Image Media Laboratory (DIML),
School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
Email: khsohn@yonsei.ac.kr

ABSTRACT

This paper describes a robust feature descriptor called the local self-similarity frequency (LSSF) for the multispectral RGB-NIR feature matching, which uses the frequency response of the local internal layout of self-similarities. A nonlinear relationship between multispectral image pairs makes conventional descriptors be sensitive to spectral deformation. To alleviate this problem, the LSSF employs a weighted correlation surface reducing the discrepancy between multispectral images. Furthermore, the LSSF provides a rotation invariance exploiting the frequency response of maximal values on log-polar bins based on the fact that a cyclic shift on the log-polar representation leads only a phase shift in a frequency domain. Experimental results show that LSSF outperforms state-of-the-art descriptors in terms of a recognition rate for multispectral RGB-NIR image pairs.

Index Terms— Local self-similarity, multispectral, near-infrared, image registration, descriptor, frequency domain

1. INTRODUCTION

The multispectral imaging, e.g., the combination of color (RGB) and near-infrared (NIR), provides a rich representation from different sensors which cannot be acquired in mono-spectral sensors [1, 2, 3, 4]. Many computer vision tasks have been revisited to address conventional problems via the multispectral imaging, such as a scene categorization [1], an image segmentation [2], a color de-hazing [3], and a dark flash photography [4].

An image registration is one of the fundamental tasks for these applications. Since multispectral image pairs are in a nonlinear relationship due to different sensor specifications, finding the correspondence is a quite challenging problem. A robust feature descriptor is a key factor to address the multispectral correspondence problem. Although conventional descriptors based on the histogram of gradient orientations, such as scale invariant feature transform (SIFT) [5] and speeded-up robust features (SURF) [6], show a satisfactory performance for mono-spectral image pairs (e.g., RGB-RGB), they cannot describe the features for multispectral image pairs properly. Binary descriptors, such as binary robust independent elementary features (BRIEF) [7] and anisotropic binary feature transform (ABFT) [8], provided a low computation complexity while enforcing the robustness. However, these descriptors were not also tailored to the multispectral image matching, thus providing the limited performance.

A number of variants of SIFT were introduced to alleviate this problem. Firmenich et al. [9] proposed the gradient invariant SIFT (GDSIFT) for a RGB-NIR image registration. Yi et al. [10] proposed the gradient orientation modification SIFT (GOM-SIFT). However, since the gradient information are varied across different spectral images, such a variant also cannot performs well similar to the SIFT [11]. Discrete curve evolution (DCE) method proposed in [12] provides a satisfactory result for a planar scene. However, it cannot be perfectly adapted to non-planar scene.

(a) (b) (c) (d) (e) (f)

Fig. 1: (a), (c), (e) NIR images. (b), (d), (f) RGB images. It shows that there exists the gradient variation between multispectral pairs.

Recently, the local self-similarity (LSS) descriptor was introduced in [13] and considered as an alternative descriptor for the multispectral image matching due to its robustness [14, 15]. The LSS describes a statistical co-occurrence of a small image patch in a larger surrounding image region capturing a locally internal geometric layout of self-similarities within an image region. In [14], the LSS was employed to find the correspondence of multispectral images. Torabi et al. [15] applied the LSS as a multispectral similarity metric, and showed the adequacy and strength for a human ROIs registration. However, the direct application of LSS to multispectral images leads to an inaccuracy matching due to a nonlinear relationship across different spectral images. Furthermore, the LSS is not robust to geometrical deformation such as a rotation.

This paper proposes a local self-similarity frequency (LSSF) descriptor for the multispectral RGB-NIR feature matching. The LSSF descriptor enables reducing the discrepancy between RGB-NIR image pair using the weighted correlation surface. Exploiting the frequency component of maximal values on log-polar bins, LSSF descriptor provides the rotation invariance and improved robustness. In addition, non-informative descriptors are eliminated based the entropy of the LSSF descriptor to improve the matching performance. Experimental results show that the LSSF descriptor outperforms state-of-the-art descriptors in terms of a recognition rate for a multispectral (RGB-NIR) feature matching.

The remainder of this paper is organized as follows. Section 2 introduces a multispectral image acquisition model and the challenging problem of multispectral image matching. Section 3 describes the LSSF descriptor for the multispectral feature matching. Experimental results are given in Section 4. Finally, the conclusion is given in Section 5 with suggestions for future works.

2. RGB-NIR IMAGE ACQUISITION MODEL AND CHALLENGING SPECTRAL DEFORMATION PROBLEM

The CCD camera sensors are made of silicon sensitive to radiation from 350 to 1100 nm, and the 400-700 nm range is a visible band

while the 700-1100 nm range belongs to a NIR band [1]. For this reason, by preventing a NIR response with a specific infrared blocking filter, the image captured by the CCD camera can be modeled by the product of an illumination spectral distribution, a surface spectral reflectance, and a sensor sensitivity [16] as follows:

$$I^i(\mathbf{p}) = \int E(\mathbf{p},\lambda)S(\mathbf{p},\lambda)Q_i(\lambda)F^{VIS}(\lambda)d\lambda, \quad (1)$$

where $i \in \{R,G,B\}$. $I^i(\mathbf{p})$ with a position $\mathbf{p} \in \mathbb{N}^2$ and λ denote RGB components and the wavelength of light, respectively. $E(\mathbf{p},\lambda)$ is an illumination spectral distribution. $S(\mathbf{p},\lambda)$ is a surface spectral reflectance, and $Q_i(\lambda)$ is a sensor sensitivity of each color channel. $F^{VIS}(\lambda)$ is a spectrum of a visible pass filter.

By replacing the infrared blocking filter by a piece of glass, a single NIR component is formed as

$$I^{NIR}(\mathbf{p}) = \sum_i \int E(\mathbf{p},\lambda)S(\mathbf{p},\lambda)Q_i(\lambda)F^{NIR}(\lambda)d\lambda, \quad (2)$$

where $F^{NIR}(\lambda)$ is a spectrum of a NIR pass filter. With the Planck's law, Lambertian reflectance, and an assumption of narrow band sensitivity [17, 18], the image acquisition model for RGB and NIR images can be simplified as

$$I^i(\mathbf{p}) = E(\mathbf{p},\lambda_i)S(\mathbf{p},\lambda_i)F^{VIS}(\lambda_i)\nu_i, \quad (3)$$

where λ_i and ν_i is a wavelength and scaling factor for each channel, respectively. In a similar way,

$$\begin{aligned} I^{NIR}(\mathbf{p}) &= \sum_i E(\mathbf{p},\lambda_{NIR})S(\mathbf{p},\lambda_{NIR})F^{NIR}(\lambda_{NIR})\nu_i \\ &= E(\mathbf{p},\lambda_{NIR})S(\mathbf{p},\lambda_{NIR})F^{NIR}(\lambda_{NIR})\nu_{NIR}. \end{aligned} \quad (4)$$

As in (3) and (4), the multispectral images can be modelled as the multiplication of different varying factors, such as $E(\mathbf{p},\lambda_i)$ and $S(\mathbf{p},\lambda_i)$. These factors are nonlinearly transformed across multispectral images with different wavelength λ. Furthermore, these factors are spatially varied. Such a nonlinear relationship between multispectral images induces a distinction of local characteristics, especially gradient information as shown in Fig. 1. It makes the conventional gradient-based descriptors and binary test-based descriptors be sensitive to a spectral deformation. Furthermore, the orientation assignment using gradient information commonly employed in conventional descriptors to provide the rotation invariance is no longer guaranteed in multispectral images.

3. MULTISPECTRAL FEATURE MATCHING WITH THE LSSF DESCRIPTOR

The framework of LSSF descriptor is summarized in Fig. 2. First, LSSF descriptor computes the weighted correlation surface, and localizes the maximal values within log-polar bins on these surfaces. Second, the maximal values are transformed into the frequency domain with an energy normalization to provide rotation invariance and improve the discrimination. In addition, based on the entropy of the LSSF descriptor, non-informative features are refined to improve the matching performance.

3.1. Weighted Correlation Surface

Let us define $I^i(\mathbf{p})$ be the multispectral channels $i \in \{R,G,B,NIR\}$. Denote $\Phi_\mathbf{p}$ and $\Omega_\mathbf{p}$ be a larger surrounding image region and small image patch centered at the pixel $\mathbf{p}$, respectively. To reduce the discrepancy between RGB-NIR image pairs, LSSF descriptor eliminates the illumination spectral distribution $E(\mathbf{p},\lambda)$, filter spectrum $F^i(\lambda)$, and scaling factor ν_i in (3) and (4) with a normalization within local regions. Since a uniform normalization induces an edge-blurring effect, LSSF descriptor exploits a weight function to penalty far positions and different color values. It computes a weighted normalization intensity $\rho^i(\mathbf{p})$ as follows:

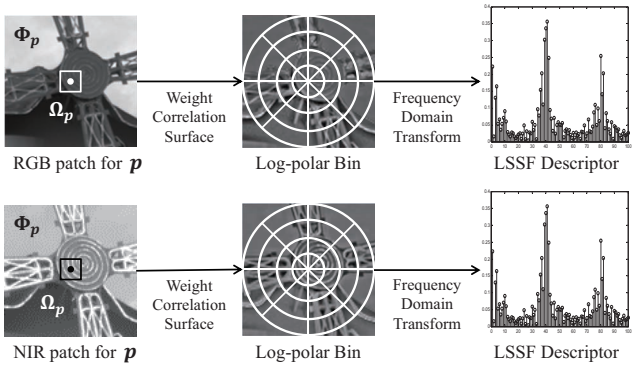


Fig. 2: Frameworks of the LSSF descriptor.

$$\rho^i(\mathbf{p}) = \frac{I^i(\mathbf{p})}{\sum_{\mathbf{q}\in\Omega_\mathbf{p}} \omega(\mathbf{p},\mathbf{q})I^i(\mathbf{q})/Z(\mathbf{p})}, \quad (5)$$

where $Z(\mathbf{p}) = \sum_{\mathbf{q}\in\Omega_\mathbf{p}} \omega(\mathbf{p},\mathbf{q})$ is a normalization factor, and a weight $\omega(\mathbf{p},\mathbf{q})$ between neighboring pixel $\mathbf{q} \in \Omega_\mathbf{p}$ and center pixel $\mathbf{p}$ is computed as the product of two Gaussian functions of geometrical proximity and color similarity as in (6).

$$\omega(\mathbf{p},\mathbf{q}) = \exp\left(-\frac{||\mathbf{p}-\mathbf{q}||^2}{\gamma_g}\right)\exp\left(-\frac{||I^i(\mathbf{p})-I^i(\mathbf{q})||^2}{\gamma_c}\right), \quad (6)$$

where γ_c and γ_g are variables used to normalize the color and spatial distances, respectively.

Under the assumption that $E(\mathbf{p},\lambda)$, $F^i(\lambda)$, and ν_i are nearly constant within local region, e.g., $E(\mathbf{q},\lambda) \simeq E(\mathbf{p},\lambda)$, they are eliminated as follows:

$$\begin{aligned} \rho^i(\mathbf{p}) &= \frac{E(\mathbf{p},\lambda_i)S(\mathbf{p},\lambda_i)F(\lambda_i)v_i}{\sum_{\mathbf{q}\in\Omega_\mathbf{p}} \omega(\mathbf{p},\mathbf{q})E(\mathbf{q},\lambda_i)S(\mathbf{q},\lambda_i)F(\lambda_i)v_i/Z(\mathbf{p})} \\ &= \frac{E(\mathbf{p},\lambda_i)S(\mathbf{p},\lambda_i)F(\lambda_i)v_i}{E(\mathbf{p},\lambda_i)F(\lambda_i)v_i\sum_{\mathbf{q}\in\Omega_\mathbf{p}} \omega(\mathbf{p},\mathbf{q})S(\mathbf{q},\lambda_i)/Z(\mathbf{p})} \\ &= \frac{S(\mathbf{p},\lambda_i)}{\sum_{\mathbf{q}\in\Omega_\mathbf{p}} \omega(\mathbf{p},\mathbf{q})S(\mathbf{q},\lambda_i)/Z(\mathbf{p})}. \end{aligned} \quad (7)$$

Then, in order to describe local internal layout for $\Phi_\mathbf{p}$, the LSSF descriptor builds a weighted correlation surface $\Psi^i(\mathbf{p},\hat{\mathbf{p}})$ for $\hat{\mathbf{p}} \in \Phi_\mathbf{p}$ defined as

$$\Psi^i(\mathbf{p},\hat{\mathbf{p}}) = \exp(-\sum_{\mathbf{q}\in\Omega_\mathbf{p},\hat{\mathbf{q}}\in\Omega_{\hat{\mathbf{p}}}} ||\rho^i(\mathbf{q}) - \rho^i(\hat{\mathbf{q}})||^2), \quad (8)$$

where $\Omega_{\hat{\mathbf{p}}}$ is a same size of shifting window within $\Phi_\mathbf{p}$.

The correlation surface in the LSS is computed by simple sum

of squared difference (SSD) as $\sum_{\mathbf{q}\in\Omega_{\mathbf{p}},\hat{\mathbf{q}}\in\Omega_{\hat{\mathbf{p}}}} ||I^i(\mathbf{q}) - I^i(\hat{\mathbf{q}})||^2$. In contrast, the proposed weighted correlation surface $\Psi(\mathbf{p},\hat{\mathbf{p}})$ is computed from a weighted normalized intensity $\rho^i(\mathbf{p})$. Thus, it provides a more robustness to a spectral deformation by eliminating some wavelength-dependent factors.

3.2. Frequency Description for Weighted Correlation Surface

Since the conventional LSS descriptor is described on a fixed local neighborhood, it is sensitive to a geometrical deformation such as rotation or viewpoint changes. Although an orientation assignment based on the gradient orientation histogram enables to provide the rotation invariance, it cannot be possible for multispectral RGB-NIR image pairs due to the gradient variation.

To alleviate these problems, the LSSF descriptor is built as frequency components for maximal values of a weighted correlation surface. The log-polar bins on $\Phi_{\mathbf{p}}$ centered at $\mathbf{p}$ are partitioned with angle $\varphi \in \{1,...,N_\varphi\}$ and radius $\mu \in \{1,...,N_\mu\}$ as follows:

$$\mathbf{bin}_{\mathbf{p}}(\varphi,\mu) = \{\hat{\mathbf{p}} | \hat{\mathbf{p}} \in \Phi_{\mathbf{p}}, \lceil |\mathbf{p}-\hat{\mathbf{p}}| \rceil = \mu, \lceil \angle(\mathbf{p}-\hat{\mathbf{p}}) \rceil = \varphi\}, \tag{9}$$

where $\lceil\cdot\rceil$ and $\angle$ is the rounding up and angle operator, respectively. Then, the maximal correlation value is selected within each log-polar bin as

$$\mathrm{M}^i_{\mathbf{p}}(k) = \max_{\hat{\mathbf{p}}\in\mathbf{bin}_{\mathbf{p}}(\varphi,\mu)} \Psi^i(\mathbf{p},\hat{\mathbf{p}}), \tag{10}$$

where $k = \varphi \cdot N_\mu + \mu \in \{1,...,N = N_\varphi \times N_\mu\}$ with an angle-oriented indexing. Inspired by [19, 20], the LSSF descriptor encodes a frequency response of maximal correlation values as a descriptor, based on the fact that a cyclic shift in a log-polar domain causes a phase shift in the frequency domain. Thus, the rotation variation on log-polar coordinates do not induce the magnitude shift in frequency domain, which will be described in Section 3.2.1. The frequency response of the $\mathrm{M}^i_{\mathbf{p}}(k)$ can be computed by discrete fourier transform (DFT) [20] as follows:

$$\Im^i_{\mathbf{p}}(n) = \sum_{k=0}^{N-1} \mathrm{M}^i_{\mathbf{p}}(k) \exp(-2\pi jkn/N). \tag{11}$$

In order to reduce a bias deformation on the correlation surface, e.g., $\mathrm{M}^i_{\mathbf{p}}(k) \leftarrow \mathrm{M}^i_{\mathbf{p}}(k)' + \beta$, the LSSF descriptor eliminates the zero frequency response to eliminate a DC component. Furthermore, in order to reduce a scaling deformation on the correlation surface, e.g., $\mathrm{M}^i_{\mathbf{p}}(k)' \leftarrow \alpha \cdot \mathrm{M}^i_{\mathbf{p}}(k)$, the LSSF descriptor is normalized by the magnitude of a frequency domain with an energy. Finally, the LSSF descriptor is defined as follows:

$$LSSF_{\mathbf{p}}(n) = \frac{\Im^i_{\mathbf{p}}(n)\overline{\Im^i_{\mathbf{p}}(n)}}{\sqrt{\sum_{n=1}^{N} \Im^i_{\mathbf{p}}(n)\overline{\Im^i_{\mathbf{p}}(n)}}}. \tag{12}$$

where $\bar{\cdot}$ is a conjugate operator.

3.2.1. *Rotation invariance property of the LSSF descriptor*

The rotation variation on log-polar bins provides the cyclic shift into the angle as follows:

$$\mathrm{M}^i_{\mathbf{p}}(k') = \max_{\hat{\mathbf{p}}\in\mathbf{bin}_{\mathbf{p}}(\varphi-\phi,\mu)} \Psi^i(\mathbf{p},\hat{\mathbf{p}}), \tag{13}$$

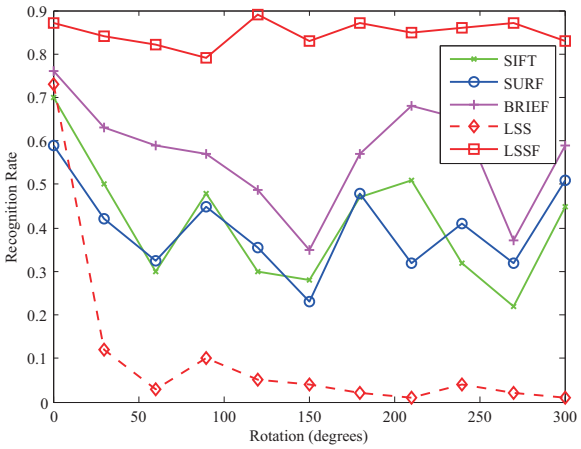


Fig. 3: Rotation invariance of the LSSF descriptor.

where $k' = (\varphi - \phi) \cdot N_\mu + \mu$. Then

$$\Im^i_{\mathbf{p}}(n') = \Im^i_{\mathbf{p}}(n) \exp(-j2\pi r\phi N_\mu/N). \tag{14}$$

The magnitude of the this function as in the LSSF descriptor can be derived as

$$\begin{aligned} &\Im^i_{\mathbf{p}}(n')\overline{\Im^i_{\mathbf{p}}(n')} \\ &= \Im^i_{\mathbf{p}}(n) \exp(-j2\pi n\phi N_\mu/N)\overline{\Im^i_{\mathbf{p}}(n)} \exp(j2\pi n\phi N_\mu/N) \\ &= \Im^i_{\mathbf{p}}(n)\overline{\Im^i_{\mathbf{p}}(n)}. \end{aligned} \tag{15}$$

It shows that the magnitude of the frequency function is invariant to cyclic shifts, thus providing a rotation invariance. Fig. 3 shows a recognition rate for descriptors. It shows the robustness and rotation invariance properties of the LSSF descriptor compared with state-of-the-art descriptors and the LSS descriptor.

3.3. Feature Refinements with the LSSF descriptor

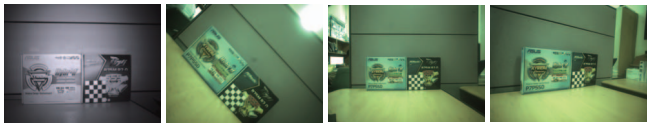
In order to improve the matching performance, the non-informative features are eliminated from each set of descriptors. The non-informative feature contain a high self-similarity in their surrounding regions, e.g. homogeneous region, and does not contain any self-similarity, e.g. salient region [13]. These features can be eliminated by using the entropy of the LSSF descriptor. The non-informative features represent the low entropy of LSSF descriptors since LSSF descriptor encodes the frequency property of local internal layouts. The entropy of the LSSF descriptor is defined as

$$H_{LSSF}(\mathbf{p}) = -\sum_{n=1}^{N} P(LSSF_{\mathbf{p}}(n)) \times \log_2 P(LSSF_{\mathbf{p}}(n)), \tag{16}$$

where $P(\cdot)$ is the probability mass function. $H_{LSSF}(\mathbf{p})$ encodes how much informative the descriptor of the pixel $\mathbf{p}$ has. A feature whose descriptor entropy is below a specific threshold τ is eliminated to improve the matching performance.

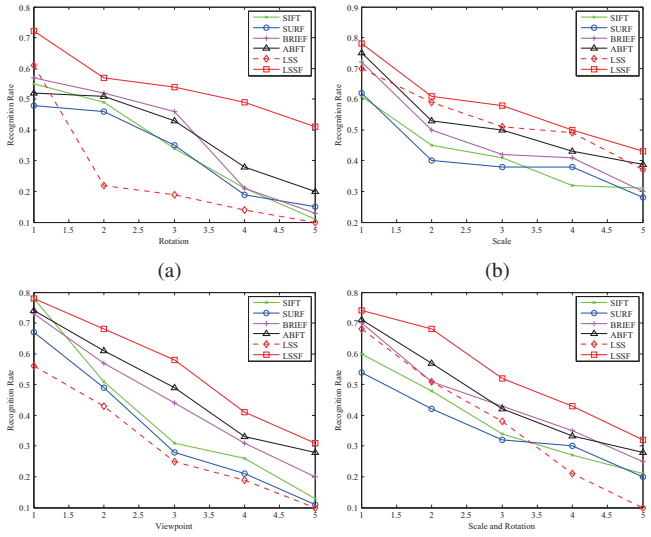
4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the performance of the LSS descriptor for multispectral (RGB-NIR) image pairs was evaluatedin feature matching, and compared with the state-of-the-art descriptors: SIFT [5], SURF [6], BRIEF [7], ABFT [8], and LSS [13]. The parameters of the



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

Fig. 4: An example of the DIML RGB-NIR database [21]. (a) The NIR image. (b) The RGB image with a rotation deformation. (c) The RGB image with a scale deformation. (d) The RGB image with a viewpoint deformation.



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

Fig. 5: The recognition rates for DIML RGB-NIR database. (a) Scale deformation. (b) Rotation deformation. (c) Viewpoint deformation. (d) Rotation with Scale deformation.

LSSF descriptor were fixed such that size of Φ: 40×40 (scale-normalized), size of Ω: 5×5 (scale-normalized), $\gamma_g = 160$, $\gamma_c = 32$, and $N = 256$. For a visible image, the weighted correlation surfaces were constructed by averaging R, G, B channels such that $\Psi^{RGB}(\mathbf{p}, \hat{\mathbf{p}}) = \frac{1}{3}\sum_{i\in\{R,G,B\}} \Psi^{i}(\mathbf{p}, \hat{\mathbf{p}})$. The parameters of other methods follow to the original works. The experiments were conducted on Intel(R) Core(TM) i7-3770 CPU at 3.40 GHz. In order to focusing the performance of a feature descriptor, the feature detector was fixed to ABFT feature detector [8] providing the same keypoints. The recognition rate defined as the ratio of correct matched descriptors in [7] was employed to evaluate the performance of descriptors.

4.1. Performance evaluation for DIML Database

As shown in Fig. 4, the DIML database [21] consists of 4 image sequences of RGB-NIR image pairs, and each sequence has the different deformation conditions such as a scale, a rotation, a viewpoint, and a scale with rotation change. Note that the deformation degree was indexed from 1 to 5. Thus, it enables to evaluate the robustness of feature descriptors for the geometrical deformation. Fig. 5 shows the recognition rate of descriptors for the DIML database. LSSF descriptor outperforms other descriptors for all deformations. The SIFT or SURF descriptors show limited matching performance since these descriptors are computed based on the gradient orientation histogram. The BRIEF and ABFT show the relatively high recognition rate since they are based on the pixel intensity test instead of the gradients. However, they also provide poor results on regions which have an indistinct order of the pixels. The LSS descriptor is sensitive to image deformation such as a rotation or a viewpoint. In contrast, LSSF descriptor show the outstanding performance compared with other descriptors in most cases. Since the LSSF descriptor encodes frequency information for the local neighborhood, it provides the rotation invariance as shown in Fig. 5 (a), (d). Furthermore, the weighted correlation surface in the LSSF descriptor provides the robustness to the discrepancy between multispectral RGB-NIR image pairs.

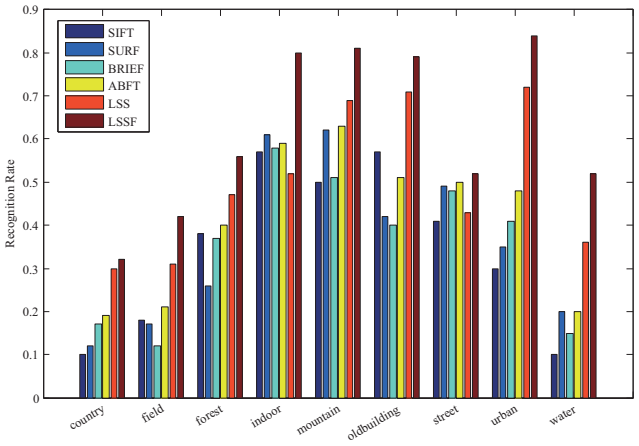


Fig. 6: The recognition rates for EPFL RGB-NIR database.

4.2. Performance evaluation for EPFL Database

The EPFL database consists of 9 scene categories and each category has 99 multispectral image pairs [1]. These databases enable to evaluate the discriminant of descriptors with varying the type of multispectral scenes, such as country, field, indoor, mountain, oldbuilding, street, urban, and water. The recognition rate for EPFL database was evaluated as shown in Fig. 6. Similar to the results of the DIML RGB-NIR databases, the SIFT and SURF descriptors based on the gradient information show some limitations. In addition, the BRIEF and ABFT descriptors also provide the limited performance due to nonlinear transformations between RGB-NIR image pairs. In contrast, the LSS and LSSF descriptor show the robustness compared to other descriptors, since they were designed to encode the geometrical layout instead of visual appearance. Compared to the LSS descriptor, the LSSF descriptor provides higher performance on the multispectral RGB-NIR image pairs.

5. CONCLUSION

The robust feature descriptor called LSSF has been proposed for multispectral RGB-NIR feature matching. The LSSF computes the weighted correlation surface to reduce the discrepancy between RGB-NIR image pair based on the multispectral image acquisition models, and localizes the maximal values within each log-polar bin. It exploits the frequency response of the maximal values to provide rotation invariance and improved discrimination. In addition, non-informative features was eliminated based on the entropy of LSSFs. Experimental results showed that the LSSF outperforms state-of-the-art descriptors for multispectral RGB-NIR feature matching.

We will apply the LSSF descriptor to address other cross-domain image matching, such as sketch-photo, paint-photo, and multimodal image pairs.

6. REFERENCES

[1] M. Brown and S. Susstrunk, "Multispectral SIFT for Scene Category Recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2011.

[2] N. Salamati, D. Larlus, G. Csurka, S. Susstrunk, "Semantic Image Segmentation Using Visible and Near-Infrared Channels," in *Proc. European Conf. Computer Vision*, 2012.

[3] L. Schaul, C. Fredembach, S. Susstrunk, "Color Image Dehazing Using the Near-Infrared," in *Proc. IEEE Int'l Conf. on Image Processing*, 2009

[4] D. Krishnan and R. Fergus, "Dark Flash Photography," *ACM Trans. Graph.*, vol. 28, no. 3, 2009.

[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91-110, 2004.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008.

[7] M. Calonder et al., "BRIEF : Computing a local binary descriptor very fast," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281-1298, 2011.

[8] S. Kim, H. Yoo, S. Ryu, B. Ham, and K. Sohn, "ABFT : Anisotropic Binary Feature Transform Based on Structure Tensor Space," in *Proc. IEEE Int'l Conf. on Image Processing*, 2013.

[9] D. Firmenich, M. Brown, and S. Susstrunk, "Multispectral Interest Points for RGB-NIR Image Registration," in *Proc. IEEE Int'l Conf. on Image Processing*, 2011.

[10] Z. Yi, C. Zhiguo, and X. Yang, "Multispectral Remote Image Registration based on SIFT," *Electronics Letters*, vol. 44, no. 2, pp. 107-108, 2008.

[11] S. Saleem, A. Bais, R. Sablatnig, "A Performance Evaluation of SIFT and SURF for Multispectral Image Matching," in *Proc. Int'l Conf. on Image Analysis and Reconition*, 2012.

[12] S. Sonn, G. A. Bilodeau, P. Galinier, "Fast and Accurate Registration of Visible and Infrared Videos," in *IEEE Conf. Computer Vision and Pattern Recognition Workshop*, 2013.

[13] E. Shechtman and M. Irani, "Matching Local Self-Similarities across Images and Videos," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2007.

[14] C. Bodensteiner, W. Huebner, K. Juengling, J. Mueller, and M. Arens, "Local Multi-Modal Image Matching Based on Self-Similarity," in *Proc. IEEE Int'l Conf. on Image Processing*, 2010.

[15] A. Torabi and G. A. Bilodeau, "Local Self-Similarity-based Registration of Human ROIs in pairs of Stereo Thermal-Visible Videos," *Pattern Recognition*, vol. 46, no. 2, pp. 578-589, 2013.

[16] S. D. Hordley, "Scene Illuminant Estimation: past, present, and future," *Color Res. Appl.*, vol. 31, no. 4, pp. 303-314, 2006.

[17] G. D. Finlayson and Gerald Schaefer, "Solving for Colour Constancy using a Constrained Dichromatic Reflection Model," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 127-144, 2001.

[18] G. D. Finlayson, S. D. Hordley, C. Lum, and M. S. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 1209-1221, 2001.

[19] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation Invariant Image and Video Description with Local Binary Pattern Features," *IEEE Trans. on Image Processing*, vol. 21, no. 4, pp. 1465-1477, 2010.

[20] R. Maani, S. Kalra, Y. H. Yang, "Rotation Invariant Local Frequency Descriptors for Texture Classification," *IEEE Trans. on Image Processing*, vol. 22, no. 6, pp. 2409-2419, 2013.

[21] http://diml.yonsei.ac.kr/∼srkim/LSSF/.