OPTIMIZED FEATURE-BASED IMAGE REGISTRATION FOR RGB AND NIR PAIRS

Amir Hossein Farzaneh, Xiaojun Qi

Department of Computer Science, Utah State University, Logan, UT, 84322 USA
farzaneh@aggiemail.usu.edu, xiaojun.qi@usu.edu

ABSTRACT

Image registration is a viable task in the field of computer vision with many applications. Researchers propose various local modules insensitive to illumination changes across cross-spectral image pairs to handle the registration challenges under different spectrum conditions. In this paper, we develop an optimized feature-based approach to register natural cross-spectral image pairs. It works on the phase information to quickly identify and describe reliable keypoints that are insensitive to illumination. It then employs a sequence of outlier removal processes to accurately find the matching feature points and the direct linear transformation to estimate the geometric transformation to align the image pair. We benchmark the proposed method and six state-of-the-art feature-based methods on the dataset provided by École Polytechnique Fédérale De Lausanne (EPFL), which includes 477 pairs of RGB-NIR images. The comprehensive analysis demonstrates that the proposed method achieves up to 13.90% accuracy improvement over the second best registration method.

Index Terms— cross-spectral registration, phase congruency, near infrared, feature-based image registration

1. INTRODUCTION

A cross-spectral image pair is referred to as a pair of two corresponding images captured in different imaging configurations, such as different camera exposures, different camera positions, and different sensors. This makes the images in one pair not perfectly aligned, hence registering them is a challenging task in computer vision applications. When registering two images, the aim is to find a geometric transformation between a pair of corresponding images to compensate for the rotation, translation, and scaling differences. The transformation is then used to spatially align, superimpose or match the images in a pair. With two registered images, it is easier to fuse information or describe differences between them. Cross-spectral image registration has wide applications in remote sensing, object detection, noise reduction, 3D image reconstruction, image fusion, video surveillance, medical image analysis, and image mosaicking.

In this paper, we focus on registering RGB spectrum and near-infrared (NIR) spectrum image pairs. A pair might have differences in translation, scale, or rotation in the viewpoint. Additionally, because different sensors capture different color spectrums, each corresponding pixel between two images has a different range of values, which is regarded as intensity variation in this application. The intensity variation presented in this type of cross-spectral images imposes an additional challenge in the task of registration

Registration methods are categorized into two classes, i.e., similarity measure-based global methods and feature-based local methods. Methods relying on similarity measures are mainly built on global statistical dependencies between images. Mutual Information (MI), which was initially introduced by Maes et al. [1], is a widely used similarity measure capturing the global structure of an image. Therefore, MI is not capable of describing local structures and differentiating local intensity variations. These deficiencies compelled researchers to develop optimized MI-based registration methods to grasp local information [2–4]. Although these proposed approaches inject some form of local representation in a global-based method, they are computationally complex, sensitive to noise, and optimized for medical images. Therefore, they cannot handle natural images with richer details and higher intensity variation.

Locally solving the problem of cross-spectral registration has been tackled mostly by finding matching features extracted with a specific descriptor [5–8]. An end-to-end feature-based method finds a correspondence between the matching keypoints and estimates a transformation from one spectrum to another. It usually consists of three major modules, namely, keypoint extraction, feature extraction, and outlier removal. The disadvantage of feature-based methods is that finding repeatable and robust features between different spectrums and different image content is often a challenging task.

Other local-based methods [9, 10], which do not fall in the category of feature-based approaches, have been also introduced. But they are either sensitive to noise or can only tolerate a very small amount of noise. Deep learning-based approaches have also been explored. For example, Large Deformation Diffeomorphic Metric Mapping (LDDMM) [11] is utilized to develop a 3D Convolutional Neural Network (CNN) architecture called Quicksilver to register two unaligned medical images. Quicksilver is optimized for medical

978-1-5386-1737-3/18/$31.00

images represented in 3D voxels. Additionally, deep learning methods require a large pre-aligned dataset to train a network, which is not always easy to craft.

This paper proposes a fast, reliable, and robust image registration method to align the RGB and NIR image pair under different illumination conditions. The contributions of the proposed method are as follows: 1) Employing the Phase Congruency (PC) method to extract the keypoints that are invariant to intensity changes 2) Incorporating the intermediate results from keypoint extraction, namely, the Log-Gabor filter responses, in the feature description step to represent each keypoint using the histogram of oriented Log-Gabor filters; 3) Designing a sequence of outlier removal processes to accurately match corresponding keypoints between the RGB and NIR image pair, which perform well regardless of whether non-rigid or rigid correspondences are present in the data; and 4) Utilizing the Direct Linear Transformation (DLT), a projective transformation, to estimate the geometric transformation for registering all the RGB points in the NIR domain. Finally, we conduct an extensive study of the image registration results on a widely used public dataset for registration applications. The dataset is provided by Computer Vision Lab at École Polytechnique Fédérale De Lausanne (EPFL). The proposed image registration method is evaluated on the registration results on EPFL dataset in terms of the Root Mean Square Error (RMSE). It outperforms other state-of-the-art methods in terms of accuracy and has comparable run-time performance compared to the second most accurate method. To the best of our knowledge, there is not a fully comprehensive study of the registration task in the literature, whereas the major focus is on the evaluation of the keypoint extraction and keypoint description. This is the first attempt to evaluate an end-to-end registration system from the perspective of the performance of key modules in the system and their impact on the whole system.

The remainder of the paper is organized as follows. Section 2 presents the proposed method. In Section 3, the evaluation method and the experimental results on the EPFL dataset are presented. A thorough study is pursued to develop a comprehensive guideline for future research. Finally, conclusions are discussed in Section 4.

2. PROPOSED METHOD

In our application, we intend to register the RGB image onto the NIR image. Hence, the NIR and RGB images will be referred to as the reference image and the moving image, respectively. All the pixels in the moving image (e.g., RGB image) are transformed to the reference image plane via the geometric transformation found in the process.

Our method consists of five components including keypoint extraction, keypoint feature description, keypoint feature matching, transformation estimation, and image registration. The aims of these five components are as follows:

- Keypoint extraction: Extracting distinct reliable and repeatable points in both reference and moving images uing PC.
- Keypoint feature description: Representing the keypoints using Log-Gabor filter responses stored in the PC results. This compact but rich feature vector captures local information and is insensitive to intensity variation.
- Keypoint feature matching: Finding the corresponding matching keypoints between the RGB and NIR images using an exhaustive matching method. The outliers are removed using the Vector Field Consensus (VFC) algorithm.
- Transformation estimation: Finding a geometric relationship between the matching keypoints in the form of a transformation matrix using the Direct Linear Transformation (DLT) algorithm.
- Image registration: Aligning or superimposing the registered RGB image onto the NIR image using the found geometric transformation.

In the following subsections, we will explain each component in detail.

2.1. Keypoint Extraction

A keypoint is a well-defined spatial location representing what stands out in an image based on local information around the selected locations such as the corners. Therefore, unlike global measures, keypoints have to be insensitive to image rotation, translation, scale change, occlusions, and background clutter. Classic keypoint extraction methods such as Harris [12] and FAST [13] extract local information based on statistical measures of gradient. Gradient-based keypoint detectors degrade the performance of cross-spectral image registration when a large intensity variation exists between images. On the other hand, Kovesi suggests employing the Phase Congruency (PC) operator [6] to extract features using local energy and local phase. This operator uses the principal moments of the PC information and a Local Energy Model (LEM) [14] to extract features in an arbitrary image. In the LEM model, features are described as points that are in the most coherence state in the phase domain. Since the extracted information is highly localized with filter responses invariant to intensity changes, PC results in a keypoint extraction module, which is robust to varying illuminations usually existing in image pairs captured in cross-spectral applications. PC at point x is computed as the ratio of weighted and noise compensated local energy summed over all the orientations to the total sum of filter response amplitudes over all orientations and amplitudes. That is:

$$PC(x) = \frac{\sum_n W(x) \lfloor A_n(x)(\cos(\phi_n(x) - \overline{\phi_n(x)}) - |\sin(\phi_n(x) - \overline{\phi_n(x)}|) - T \rfloor}{\sum_n A_n(x) + \epsilon} \quad (1)$$

where the term $W(x)$ is the frequency spread weighting factor, $A_n(x)$ is the amplitude of the nth Fourier component at x, and $\phi_n(x)$ is the phase of the nth component at x. The symbol $\lfloor \rfloor$ returns the enclosed value as it is if the value is positive; otherwise, the symbol returns 0. ϵ (e.g., 0.0001) prevents the PC value from becoming unstable as the term $\sum_n A_n(x)$ becomes very small. This formulation of PC not only provides better localization but also compensates the noise with an empirically determined optimal value T.

In practice, local frequency information is captured using a bank of Log-Gabor filters at different scales and orientations. We take advantage of 4 scales and 8 orientations, which capture enough scale and orientation information, to generate a bank of 32 Log-Gabor filters. The input image is then convolved with these filters and their responses (local energies) are saved to describe features in the next section. PC then proceeds with moment analysis to extract second moments. At this stage, we exploit the minimum moment as a corner information map to extract the corners in both RGB and NIR images. The 1200 corners with the most strength (i.e., the largest 1200 values) in the map are chosen as the keypoints among the candidate corners for each image in a pair and are passed to the next module.

2.2. Keypoint Feature Description

We use a descriptor insensitive to illumination changes to represent features at corner points due to the non-linear intensity variation between the cross-spectral RGB-NIR image pair. We choose the Log-Gabor Histogram Descriptor (LGHD) [7], which is a distribution-based descriptor relying on high frequency components to make it a more robust candidate for our desired application. LGHD uses the Log-Gabor filters in different scales and orientations to build a histogram of oriented filters in a patch of size $S \times S$ around each extracted keypoint. Each patch is divided into 16 smaller sub-regions and then the histogram is calculated.

Since LGHD itself uses PC, we combine the keypoint extraction and feature description module into a single module as our contribution. In other words, the local energies, i.e., the Log-Gabor filter bank responses saved in the keypoint extraction module, are passed to the LGHD. Similar to the keypoint extraction step, we use 4 scales and 8 orientations to construct a feature vector of size $4 \times 8 \times 16$ (512). We empirically use patches of size 50 (e.g., $S = 50$) around each keypoint to compute the histogram. Larger patches would allow us to consider possibly more informative descriptors, but at the same time they would be more susceptible to occlusions and slower to compute. At the end of this step, the RGB image has a set of feature vectors denoted as f_{RGB} to represent the characteristics of each keypoint and the NIR image has another set of feature vectors denoted as f_{NIR} to represent the characteristics of each keypoint.

2.3. Keypoint Feature Matching

As discussed in the previous section, each keypoint is represented by a feature vector of 512 values. An exhaustive matching method is used to compute the pairwise distance between the feature descriptors of the keypoints in each RGB and NIR pair. Two keypoints match if their sum of absolute differences in their feature descriptor in all 512 dimensions is less than a certain threshold. This exhaustive matching method ensures that all potentially matching keypoints are uniquely identified and saved in a set of putative matching points. This, however, leaves us with outliers, which need to be removed to make our transformation estimation more accurate. Maintaining a robust set of corresponding points from a putative set of matching points is an essential step in the registration task prior to transformation estimation. Classic Sample Consensus (SAC) algorithms such as Random Sample Consensus (RANSAC) or M-estimator Sample Consensus (MSAC) are highly sensitive to the proportion of outliers. Moreover, they cannot handle non-rigid (non-parametric) correspondences. For our task, we adopt the idea of Vector Field Consensus (VFC) algorithm [15] to represent the matching points by motion field samples and take advantage of the Expectation Maximization (EM) algorithm [16] to detect inliers and remove outliers. If the observed 2D sets of matching points are $P_m = (x_m, y_m)^T$ and $P_r = (x_r, y_r)^T$ with P_m representing the set of keypoints in the moving image and P_r representing the set of keypoints in the reference image, the motion field vector for each pair of matching points is:

$$v = (s_n, t_n), \quad s_n = P_m, \;\; t_n = P_r - P_m \tag{2}$$

where s_n is the vector's starting point and t_n is the vector's terminal point. Next, we define the motion field set as:

$$S = \{(s_n, t_n) : n \in \mathbb{N}\} \tag{3}$$

The goal is to fit a mapping field function f so that

$$t_n = f(s_n) \tag{4}$$

The robust estimation of f is obtained when there are no outliers present in the data. By assuming a Gaussian noise with zero mean, an arbitrary uniform standard deviation for the inliers, and a uniform distribution for the outliers, VFC employs the EM algorithm to estimate a set of parameters including f, inlier, and outlier distribution parameters. The EM algorithm estimates a posterior probability for each vector by updating the distribution parameters until convergence (i.e., reaching the desired minimum energy). The final solution enforces closeness of f to the inliers and maintains smoothness on the vector field of f. Vectors with the posterior probability lower than a certain threshold (e.g., 0.75) are considered to be outliers.

2.4. Transformation Estimation

Given a reference image r and a moving image m, the goal of image registration is to find a transformation function, $H : \mathbb{R}^d \rightarrow \mathbb{R}^d$, which maps all the pixels in the moving image to their corresponding pixels in the reference image. Here, d denotes the dimension of the data, which in our case is 2D (i.e., the x and y coordinates of matching keypoints). At this step, we aim to find the optimal projective transformation matrix to map all the matching points in the moving image to their corresponding matching points in the reference image. Specifically, we utilize the DLT algorithm [17] to estimate the projective transformation H. We also constrain DLT to require at least 8 matching points instead of 4 matching points to increase its robustness to estimate H. Hence, the task of registration will be tagged as *failed* if the keypoint feature matching method cannot identify at least 8 robust matching points. The 2D inlier set of matching keypoints in the moving RGB image is denoted as $P_m = (x_m, y_m, 1)^T$, where $m = 1, 2, ..., N$, and N is the number of matching points with $N \geq 8$. Similarly, the 2D inlier set of matching keypoints in the reference NIR image is denoted as $P_r = (x_r, y_r, 1)^T$, where $r = 1, 2, ..., N$. The transformation equation is denoted as $P_r = HP_m$. DLT uses Singular Value Decomposition (SVD) to calculate the 9-value vector consisting of the entries of the matrix H:

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \tag{5}$$

Using all pairs of the matching points between the RGB and NIR images, DLT estimates all 9 elements of the H matrix, which is further applied to all the pixels in the RGB image to register the RGB image onto the NIR image.

3. EXPERIMENTAL RESULTS

In this section, we discuss our experiments and the results to further evaluate the proposed image registration method and the state-of-the-art image registration methods. These image registration methods have been tested on the EPFL dataset [18], which includes 477 pairs of RGB-NIR images. The image pairs are categorized into 9 different types of scenes: country, field, forest, indoor, mountain, old building, street, urban, and water. Each category contains at least 50 image pairs. Image pairs in the EPFL dataset are already aligned. This will eliminate the need for manual labeling and facilitate the evaluation. Sample pairs from this dataset are shown in Fig. 1.

We have compared the proposed approach with different combinations of keypoint extractors and keypoint descriptors, which are promised to deliver good results under illumination varied applications. Two powerful keypoint extractors, namely, SIFT [19] and PC, have been chosen for our benchmark. We describe the features at each keypoint using the four most commonly used cross-spectral descriptors such as LGHD, SIFT, Eight Local Directional Patterns (ELDP) [20], and Phase Congruency and Edge Histogram Descriptor (PCEHD) [21]. We exclude keypoint extractors such as FAST because of its poor performance in locating enough robust and reliable keypoints on the EPFL dataset. We extract the SIFT descriptors around each keypoint identified by SIFT since they can be easily extracted from the SIFT keypoint extraction process. However, extracting the SIFT descriptors for other keypoint extractors is difficult and we could not find any reliable online source code to do this. We use the following naming conventions {keypoint extractors + cross-spectral descriptors} to build the benchmark for six state-of-the-art image registration methods including {SIFT+LGHD}, {SIFT+SIFT}, {SIFT+ELDP}, {SIFT+PCEHD}, {PC+ELDP}, {PC+PCEHD}, the proposed method (i.e., an efficient version of {PC+LGHD}), and its variant methods, which remove outliers using different SAC algorithms. It should be noted that the six state-of-the-art image registration methods use the same sequence of outlier removal processes as proposed in the paper (i.e., the exhaustive matching method followed by VFC) to find the reliable matching keypoints. The benchmark is executed on a 3.4GHz Core i7 machine with 16GBs of RAM.

Since the images are in different spectrums, we cannot use the intensity of registered points to evaluate the registration performance. Instead, we use the RMSE to evaluate the accuracy of the estimated transformation H. Since the image pairs in the EPFL dataset is pre-aligned, we use H to register the inlier keypoints extracted from the RGB image to be aligned with their matching points in the NIR image. We then compute RMSE based on the number of pixels that the registered points shift away from their original locations in the NIR image. Specifically, if P_i is the set of matching points in the reference image and the set of their corresponding matching points in the moving image after employing the transformation is denoted by $P_j = HP_i$, RMSE is calculated by:

$$RMSE = \sqrt{||P_i - P_j||^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} ||P_i - HP_i||^2} \tag{6}$$

where N is the number of matching points and i is the index of a pair of matching points in both the reference image and the moving image. Smaller values of RMSE represent a more accurate transformation from RGB to NIR. In literature, an RMSE of below 5 pixels is usually considered to be a fair error [5].

Table 1 summarizes the RMSE and runtime performance of the six state-of-the-art registration methods, the proposed method, and its three variant methods on the EPFL dataset for each category, respectively. Overall, the proposed method outperforms the other methods with an average of 2.29 pixels in RMSE for all the images. This renders as a 13.90%

Fig. 1. Sample RGB-NIR image pairs from the EPFL dataset (top row: RGB images; bottom row: NIR images).

Table 1. Performance summary (i.e., the RMSE in terms of pixels and runtime in terms of seconds) of the proposed image registration method, its three variant methods, and six compared feature-based image registration methods on the EPFL dataset, where RMSE value is followed by the runtime value shown in parenthesis.

	Category									
METHOD	COUNTRY	FIELD	FOREST	INDOOR	MOUNTAIN	OLD BUILDING	STREET	URBAN	WATER	**Average**
SIFT+LGHD	**2.48** (20.33)	**4.75** (19.99)	0.94 (23.44)	0.68 (13.12)	5.69 (19.34)	3.13 (16.95)	1.78 (17.76)	0.5 (15.68)	**4.03** (18.06)	2.66 (18.3)
SIFT+SIFT	5.79 (4.32)	5.58 (4.29)	1.00 (5.63)	0.94 (2.67)	7.72 (4.72)	3.54 (4.00)	1.99 (3.82)	0.54 (3.56)	8.26 (3.65)	3.93 (4.07)
SIFT+ELDP	3.74 (8.31)	5.50 (8.12)	0.90 (9.84)	1.02 (5.27)	5.69 (8.11)	3.12 (7.16)	1.62 (7.26)	0.54 (6.63)	11.96 (7.03)	3.79 (7.53)
SIFT+PCEHD	3.84 (11.87)	5.66 (11.81)	**0.88** (13.60)	0.87 (8.33)	3.92 (11.95)	2.65 (10.34)	1.96 (10.48)	0.48 (9.28)	8.46 (10.21)	3.19 (10.87)
PC+ELDP	3.64 (5.17)	4.84 (5.26)	failed (5.23)	0.68 (5.12)	3.63 (5.13)	1.18 (4.82)	2.28 (5.20)	0.37 (4.92)	15.06 (5.09)	failed (5.10)
PC+PCEHD	4.89 (5.55)	11.40 (5.68)	failed (5.66)	**0.61** (5.55)	**2.48** (5.49)	1.25 (5.23)	2.81 (5.64)	**0.36** (5.35)	9.57 (5.51)	failed (5.52)
our method										
+VFC	2.92 (9.05)	5.12 (8.87)	1.13 (8.92)	0.64 (8.68)	2.61 (8.73)	**1.16** (8.82)	**1.44** (9.78)	0.37 (8.54)	5.37 (8.69)	**2.29** (8.90)
+RANSAC	7.52 (9.14)	13.53 (9.20)	2.62 (9.24)	1.91 (8.99)	6.42 (8.98)	3.45 (8.70)	3.51 (9.18)	1.26 (8.79)	13.72 (9.04)	5.99 (9.03)
+MSAC	6.40 (8.97)	20.81 (9.01)	3.32 (9.07)	1.20 (8.82)	5.67 (8.81)	3.97 (8.48)	3.64 (8.99)	0.92 (8.58)	11.10 (8.87)	6.34 (8.84)
+MLESAC	7.83 (8.95)	17.14 (9.01)	4.65 (9.07)	2.69 (8.80)	10.68 (8.90)	4.32 (8.78)	3.26 (9.30)	1.99 (8.89)	22.97 (9.24)	8.39 (8.99)
Average	3.89 (9.16)	6.08 (9.12)	2.38 (9.97)	0.77 (7.53)	4.53 (9.01)	2.29 (8.32)	1.99 (8.74)	0.45 (8.02)	8.97 (8.53)	

accuracy improvement compared to the second best method {SIFT+LGHD}. Additionally, the proposed method delivers an RMSE of below 5 pixels across all categories except the water category, while the second best method {SIFT+LGHD} delivers an RMSE of below 5 pixels across all categories except for the mountain category. Compared to {SIFT+LGHD}, the proposed method significantly improves the registration performance for images in mountain and old building categories, slightly improves the registration performance for images in indoor, street, and urban categories, and achieves comparable registration performance for images in the other categories. It also achieves the best accuracy in terms of RMSE in street and old building categories among all the compared methods. This suggests that the proposed method performs the best in scenes with a lot of variety and corners, which are the features commonly seen in buildings and vehicles. The images in the water category seem to be the most challenging to register. This is mainly due to homogeneous texture of water, which makes it hard for the keypoint extractors to find distinctive keypoints.

Table 1 shows that the proposed method is more than two times faster than the second best registration method {SIFT+LGHD}. Other methods such as {SIFT+SIFT}, {SIFT+ELDP}, {PC+ELDP}, and {PC+PCEHD} are faster than our method. However, they do not deliver good accuracy across the categories. For example, the fastest registration method {SIFT+SIFT} achieves the second worst accuracy in RMSE of 3.93 pixels. The second fast registration method {PC+ELDP} fails to register all images. This makes the proposed method the best candidate from the perspectives of both accuracy and speed.

Table 1 also lists the RMSE performance of the proposed method and its three variant methods, which use RANSAC, MSAC, and Maximum Likelihood Estimation Sample Consensus (MLESAC) to remove the outliers. It is clear that VFC achieves better performance than the other three SAC methods. With an average RMSE of 2.29 pixels, VFC improves the second best method (i.e., +RANSAC) by 61.76%. In all categories except for the challenging water category, VFC achieves RMSEs of smaller than 5 pixels. The overall average RMSEs of the three variant methods are all over 5 pixels.

VFC has comparable runtime of 8.90 seconds in average compared to the classic SAC methods. From the perspectives of registration error and speed, the proposed method with VFC as the outlier remover is the best candidate.

4. CONCLUSIONS

In this paper, we propose an optimized feature-based approach to quickly, reliably, and robustly register cross-spectral image pairs under different illumination conditions. Our major contributions include:

- Employing the PC method, which performs well under various illuminations, to identify reliable and robust keypoints that are invariant to intensity changes.
- Incorporating the Log-Gabor filter responses obtained from the keypoint extraction step to represent the characteristics around each keypoint using the histogram of the filter responses.
- Designing a sequence of outlier removal processes (i.e., exhaustive matching method followed by VFC) to accurately find the reliable matching keypoints.
- Employing DLT to estimate the geometric transformation to align the image pair.
- Proposing the RMSE measure to evaluate the registration performance.

We evaluate the proposed method, its three variant methods incorporating different outlier removal algorithms, and six common feature-based approaches in the cross-spectral registration field on a public dataset EPFL. The proposed method achieves a 13.90% improvement in accuracy compared to the second best method {SIFT+LGHD}. VFC is the best candidate for outlier removal. Overall, our method outperforms other state-of-the-art feature-based methods that are developed for cross-spectral imagery from the perspectives of both accuracy and speed.

5. REFERENCES

[1] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[2] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE Trans. Med. Imaging*, vol. 25, no. 5, pp. 626–639, 2006.

[3] H. Rivaz, Z. Karimaghaloo, and D. L. Collins, "Self-similarity weighted mutual information: a new nonrigid image registration metric," *Med. Image. Anal.*, vol. 18, no. 2, pp. 343–358, 2014.

[4] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 19–29, 2010.

[5] C. Zhao, H. Zhao, J. Lv, S. Sun, and B. Li, "Multimodal image matching based on multimodality robust line segment descriptor," *Neurocomputing*, vol. 177, pp. 290–303, 2016.

[6] P. Kovesi, "Phase congruency detects corners and edges," in *The Australian Patt. Recog. Soc. Conf.: DICTA*, 2003.

[7] C. Aguilera, A. D. Sappa, and R. Toledo, "LGHD: A feature descriptor for matching across non-linear intensity variations," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*. IEEE, 2015, pp. 178–181.

[8] S. Kim, D. Min, B. Ham, S. Ryu, M. N. Do, and K. Sohn, "DASC: Dense adaptive self-correlation descriptor for multi-modal and multi-spectral correspondence," in *Proc. IEEE Conf. on Comput. Vis. and Patt. Recog.*, 2015, pp. 2103–2112.

[9] C. Wachinger and N. Navab, "Entropy and Laplacian images: Structural representations for multi-modal registration," *Med. Image. Anal.*, vol. 16, no. 1, pp. 1–17, 2012.

[10] M. Chen, A. Carass, A. Jog, J. Lee, S. Roy, and J. L. Prince, "Cross contrast multi-channel image registration using image synthesis for MR brain images," *Med. Image. Anal.*, vol. 36, pp. 2–14, 2017.

[11] X. Yang, R. Kwitt, and M. Niethammer, "Quicksilver: Fast predictive image registration-a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.

[12] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey Vis. Conf.* Manchester, UK, 1988, vol. 15, pp. 10–5244.

[13] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, 2010.

[14] M. C. Morrone, J. Ross, D. C. Burr, and R. Owens, "Mach bands are phase dependent," *Nature*, vol. 324, no. 6094, pp. 250–253, 1986.

[15] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Proc.*, vol. 23, no. 4, pp. 1706–1721, 2014.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.: Ser. B*, pp. 1–38, 1977.

[17] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[18] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *Proc. IEEE Conf. Comput. Vis. Patt. Recog.* IEEE, 2011, pp. 177–184.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[20] M. R. Faraji and X. Qi, "Face recognition under illumination variations based on eight local directional patterns," *IET Biom.*, vol. 4, no. 1, pp. 10–17, 2015.

[21] T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera, and R. Toledo, "Multispectral stereo odometry," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1210–1224, 2015.