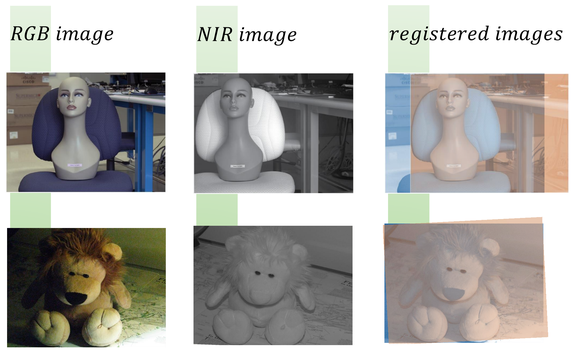
OPTIMIZED FEATURE-BASED IMAGE REGISTRATION FOR RGB AND NIR PAIRS

Amir Hossein Farzaneh, Xiaojun Qi

UtahStateUniversity
Computer SCIENCE

IMAGE REGISTRATION

A cross-spectral image pair is two images of the same scene captured under different imaging configurations. The goal is to automatically align a pair spatially so the corresponding pixels have the same positions. The focus is to register RGB-NIR images in pairs.



- **APPLICATIONS:**
  - remote sensing, object detection, noise reduction, 3D image reconstruction, image fusion, image mosaicking, etc.
- **CHALLENGES:**
  - different translations, rotations, and scales
  - intensity variation

DATASET

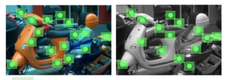
EPFL RGB-NIR dataset containing 477 image pairs in 9 different scene categories: country, field, forest, indoor, mountain, old building, street, urban, and water. This benchmark dataset is available for evaluating the performance of cross-spectral image registration techniques.

THE PROPOSED METHOD

We propose a local feature-based method to tackle the challenges of cross-spectral image registration. Extracting local features in the image phase domain ensures invariance to translation, rotation, scale, and pixel intensity. Below is the summary of the proposed method pipeline:

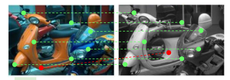
1 KEYPOINT EXTRACTION: Phase Congruency (PC) extracts local frequency information using Log-Gabor filters. Minimum moments are considered as keypoints.

$$PC(x) = \frac{|E(x)|}{\sum_n A_n(x)}$$



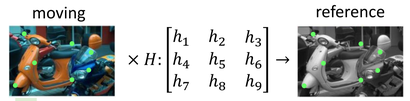
2 KEYPOINT FEATURE DESCRIPTION: The Log-Gabor Histogram Descriptor (LGHD) extracted from the Log-Gabor filtered images resulted from step 1 describes keypoints on both images using local patches around them.

$$f_{image} \in R^{512}$$



3 KEYPOINTS FEATURE MATCHING: An exhaustive matching method finds corresponding feature vectors between two images. Vector Field Consensus (VFC) removes outliers to obtain a reliable set of matching points.

$$s_n = P_m, \quad t_n = P_r - P_m$$
$$t_n = f(s_n)$$



4 TRANSFORMATION ESTIMATION: Direct Linear Transformation (DLT) algorithm estimates a projective transformation matrix H to map all the pixels in the RGB image onto the NIR image.

$$P_r = HP_m$$

RESULTS

Our method is compared against different combinations of keypoint extractors and keypoint feature descriptors designed for cross-spectral applications:

- SIFT + LGHD
- SIFT + SIFT
- SIFT + ELDP
- SIFT + PCEHD
- PC + ELDP
- PC + PCEHD

The Root Mean Squared Error (RMSE) is used to evaluate the accuracy of the estimated transformation matrix H.

Performance summary of the proposed image registration method, its three variant methods, and six compared feature-based image registration methods on the EPFL dataset: *RMSE* (*runtime*)

METHOD	Category: 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)	

- Second best method runs slower.
- Not all methods are capable to register all images
- Our proposed method offers the least RMSE error averaged across all categories.
- Comparable run-time with different outlier rejection methods.
- *Water* is the most challenging scene category.

IEEE International Conference on Multimedia and Expo (ICME) 2018