RESEARCH ARTICLE OPEN ACCESS

A Novel Hybrid Coarse to Fine Registration Pipeline for UAV Systems

Alugubilli Harikrishna[1] and Durga Ganga Rao Kola[2]

[1]*PG Student, Dept of ECE. University College of Engineering, JNTU Kakinada, Andhra Pradesh.*

[2]*Assistant Professor, Dept of ECE. University College of Engineering, JNTU Kakinada, Andhra Pradesh.*

ABSTRACT

Multimodal image registration suffers from intensity differences and geometric distortions. In case of visible and infrared images captured from Unmanned Aerial Vehicle (UAV), it is difficult to register due to variations in focal lengths and Field of Views (FOV). This paper introduces a novel hybrid registration pipeline and it integrates the Coarse-to-fine feature matching with Adjacent Self similar Three-Dimensional Convolutional (ASTC) descriptors. In this method, features are extracted via block FAST method and Adjacent self-similarity (ASS) model is applied for creating multidimensional features. The 3-D convolution is used for enrichment of features. For robust feature matching, an ASTC is used. Hybrid Thin Plate Spline (TPS) transformation is applied for non-rigid alignment. The proposed method is tested on VIR-UAV dataset and confirm that it outperforms over traditional methods such as RIFT, LGHD.

Keywords: Adjacent Self similarity 3-D convolution (ASTC), image registration, Thin Plate Spline (TPS), Unmanned Aerial Vehicle (UAV).

Date of Submission: 25-08-2025 Date of acceptance: 03-09-2025

I. Introduction

Unmanned Aerial Vehicles (UAVs) have become important instruments for remote sensing and data collection in numerous applications across various landscapes, due to their ability to integrate multiple sensors for data acquisition [1]. Alignment of two or more images of the same scene taken across time, differing viewpoints, or different sensors is called image registration. Single modal image registration techniques are not suitable to align multimodal images. So Multimodal remote sensing image registration (MRSIR) is very important for the multi-sensor data integration process that contributes to applications such as image fusion applications, change detection applications, and mosaicking applications [2].

Therefore, the evolution from single-mode, such as optical sensors to multi-mode systems, including multispectral, hyperspectral, light detection and ranging (LiDAR), and synthetic aperture radar (SAR), had produced multi-modal remote sensing images (MRSIs), resulting in multiple different spatial, temporal, and spectral resolutions, which yielded different insights about the earth surface. However, having different sensors also brings challenges as different platforms can produce geometric distortions, and nonlinear radiometric differences (NRDs) between the several modalities (e.g., optical-infrared, optical-SAR, or optical-LiDAR) [2].

Area-based methods like Normalized Cross Correlation (NCC) and Mutual Information (MI) can mitigate nonrigid transformation, but these are limited when there's a large geometric distortion and noise [3]. Feature-based methods can improve matching of images using maximum index maps to reduce matching problems for a given distortion in the radiometry, thus improving repeatability. However, they are still susceptible to some variations in scale and viewpoint [4]. Moreover, deep learning models applied to multimodal data sets, have promising applications but these are sensitive to scene complexity and computationally intensive which demonstrates a need for better approaches to be developed [5].

To Address the challenges in Multimodal image registration such as uneven feature distribution, low repeatability of feature points and differences in Field of Views (FOV), a novel hybrid coarse to fine registration pipeline is proposed. The major contribution of this method is as follows:

1) Introduces a coarse to fine registration to avoiding scaling differences.
2) ASTC descriptors for enhanced multimodal matching is used.
3) A hybrid TPS transformation for non-rigid warping is considered.

II. Adjacent Self Similarity 3-D Convolution (ASTC)

ASTC builds upon self-similarity in structure on images by applying an approach using a uniform feature point extractor based on block FAST (Features Accelerated Segment Test) detection strategy and an enhanced descriptor based on adjacent self-similarity (ASS) modeling. The 3-D convolution is adopted for saliency enhancement of the features [6]. This method improves efficiency of matching of feature points. The similarity measure employs a Fast Fourier transform (FFT)-based approach, and outlier rejection uses Fast Sample Consensus (FSC).

The ASTC method mainly include the following three steps:

1) Feature point extraction using Block FAST method.
2) Construction of ASTC feature Descriptor
3) Similarity measure and outlier rejection

For ensuring feature points spread evenly, the image is divided into n x n sub blocks, then perform FAST detection to each block. ASTC feature descriptor is constructed by using Adaptive Histogram Equalization (AHE), Adjacent Self-Similarity (ASS) feature computation and three-dimensional convolution optimization. AHE is used to enhance the contrast of image by measuring local histogram of the image [9]. ASS computation is compute the Sum of Squared Differences (SSD) between mid-patch and offsets patches [10]. The 3-D Gaussian convolution sort out self-similarity values, filter noise and improves the structural prominence.

III. Proposed Method

The block diagram of proposed method is represented in Fig-1. The visible and infrared images are the input images and the scaling differences between the input images are reduced in coarse registration stage by using Affine transformation. Feature maps are generated using Phase Congruency then PC map of coarse registration result I_{ir}' by divided it into n × n blocks. Based on ASTC descriptor, the feature points in both images are optimized. Those feature points are matched by using FFT technique.

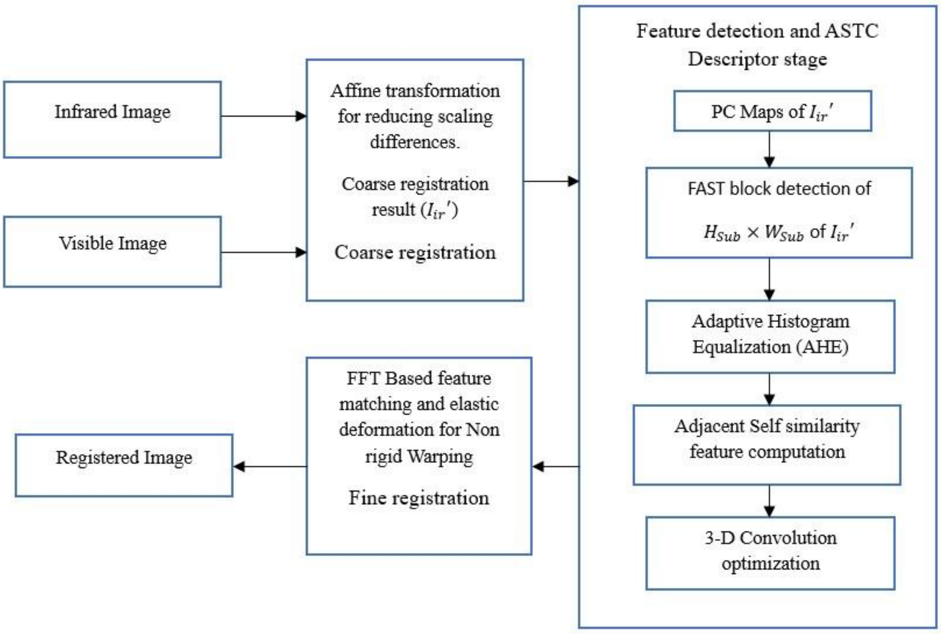
Infrared Image
Visible Image
Affine transformation for reducing scaling differences.
Coarse registration result (I_{ir}')
Coarse registration
Feature detection and ASTC Descriptor stage
PC Maps of I_{ir}'
FAST block detection of $H_{Sub} \times W_{Sub}$ of I_{ir}'
Adaptive Histogram Equalization (AHE)
Adjacent Self similarity feature computation
3-D Convolution optimization
FFT Based feature matching and elastic deformation for Non rigid Warping
Fine registration
Registered Image

Fig-1: Block Diagram of proposed method

3.1 Phase Congruency (PC)

Phase Congruency is a dimensionless quantity that is invariant to changes in image brightness or contrast, it provides an absolute measure of the significance of feature points. The phase congruency function in terms of the Fourier series expansion of a signal at some location x is [7]:

$$PC(x) = \frac{\sum_n W(x)\lfloor A_n(x)(\cos(\triangle\phi(x)) - |\sin(\triangle\phi(x))| - \hat{T}\rfloor}{\sum_n A_n(x) + \varepsilon} \quad (1)$$

$W(x)$ is the frequency spread weight, $\triangle\phi(x)$ is the deviation between $\phi(x)$ and $\bar{\phi}(x)$. Here constant ε is introduced to ignore division by zero. $A_n(x)$ is implies the amplitude of Fourier components [7].

PC feature maps for both images ensure robust feature extraction across modalities with noise compensation.

3.2 Coarse registration

Let visible and infrared images be represented by I_{vis} and I_{ir}. The translation parameters and scaling parameters are included in similarity transformation and this is special affine transformation.

The coarse registration result is I_{ir}' and it includes scaling (X_{scale}) and translation parameters (t_x,t_y) [8].

$$I_{ir}' = \begin{bmatrix} X_{scale} & 0 & t_x \\ 0 & X_{scale} & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

The translation parameters are represented as t_x and t_y

$$t_x = \frac{1}{2}(Col_{vis} - X_{scale}Col_{ir})$$

$$t_y = \frac{1}{2}(Row_{vis} - X_{scale}Row_{ir})$$

where the Row and Col are the height and width of input images respectively.

To maintain evenly distribution of feature points in the PC Maps of infrared and visible images, the feature points are taken out by using a block-FAST detector. The image is split into $H_{Sub} \times W_{Sub}$ small blocks, then FAST feature detection is conducted for each block.

3.3. Hybrid ASTC descriptor:

For template matching, the descriptor construction is playing key role. The ASTC descriptor is the combination of the AHE, ASS computation and it is integrated with CFOG, finally the 3-D gaussian convolution is applied to the ASS-CFOG features.

3.3.1. Adaptive Histogram Equalization (AHE)

AHE improves the local contrast of the image and get more textural information in an image. Let the image be indicated as S, the AHE function of the image is follows as [6]:

$$S' = f_{AHE}(S) \quad (3)$$

where S' denotes the image after applying AHE function, $f_{AHE}(.)$ is the AHE function.

3.3.2 Adjacent Self Similarity (ASS) feature computation

The ASS feature computation contains of Sum of Squared Differences (SSD) constitution of image block concentrated on the ASS feature point of a pixel. The SSD is defined as [6]:

$$P(i,j) = \exp\left(-\frac{SSD(i,j)}{\max(\lambda,\gamma)}\right) \quad (4)$$

where λ is a constant parameter and γ is the maximal variance.

3.3.3 CFOG integration

The Channel Feature of Oriented Gradient (CFOG) generates gradient map g_0 for each ASS map convolving with 2-D gaussian kernel by using the following equation [15]:

$$g_0 = [\partial S / \partial O]^+ \quad (5)$$

where S is the image, O is the orientation of the derivative. $[]^+$ represents that the enclosed quantity equal to itself when its value is positive or otherwise zero.

3.3.4 3-D convolution optimization

The 3-D gaussian kernel enhances the prominence of the self-similarity features and also it smooths the image's self-similarity values. The 3-D gaussian convolution function is described as [6]:

$$G(x,y,z)_{3D} = \begin{bmatrix} S_\sigma(x,y) \otimes G_1) \\ S_{\sigma+1}(x,y) \otimes G_2) \\ S_{\sigma+2}(x,y) \otimes G_3) \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}^T \quad (6)$$

where $\mathrm{G(x,y,z)_{3D}}$ is result of 3-D gaussian convolution function, x and y represent the row and column directions of the 3-D image, σ indicates the standard deviation of the Gaussian kernel function, and ⊗ denotes the matrix multiplication.

3.4. Fine Registration

For feature matching, the Fast Fourier Transform (FFT) based Sum of squared differences is adopted. The SSD between two images is defined as [8]:

$$S_i(v_i) = \sum_n D_1{}^2(x)T_i(x) + \sum_n D_2{}^2(x - v)T_i(x) - 2\sum_n D_2(x-v)D_1(x)T_i(x) \qquad (7)$$

where D_1 and D_2 are the features of input images of per pixel, x represents the position of a pixel in feature of D and *i* is the template window, $T_i(x)$ is the masking function, and v_i is the offset vector that matches D_1 with D_2.

At final stage of image registration, the Thin Plate Spline (TPS) is used which introduce the local warping by interpolating deformation fields based on features. Further, it employs a non-linear deformation between the images. Finally, the non-rigid alignment of Infrared and visible images done by TPS.

IV. Experimental Results

The proposed fine registration method is tested on dataset VIR-UAV. This dataset contains of 17 pairs of visible and infrared images. These images are divided into three groups. The first and third group images pairs are having 1600 ×1200 pixels of resolution. Second group image pairs having resolution of 1920 × 1080 pixels of resolution.

The proposed method compares with two existing algorithms that is Log Gabor Histogram Descriptor (LGHD) and Radiation variation Insensitive Feature Transform (RIFT). The LGHD method is designed to match feature points between images with non-linear intensity differences, such as those between visible (RGB) and infrared (NIR or LWIR) images [13]. These image pairs often differ significantly due to variations in colour, texture, and gradient directions. RIFT integrates PC with a Maximum Index Map (MIM) to create a radiation-insensitive feature matching framework. RIFT employs PC maps derived from log-Gabor filters to detect features [14].

4.1 Performance Metric

The performance of the proposed method is measured by using Root Mean Square Error (RMSE) [8]. These are defined as follows:

$$\mathrm{RMSE} = \sqrt{1/L \sum_{i=1}^{L}(f_i - \Gamma(m_i))^2} \qquad (8)$$

where (f_i, m_i) is i^{th} checkpoint, L is the number of the selected check points. $\Gamma(.)$ represents the transformation model obtained by different methods.

The table-1 shows the RMSE errors of the registered images of RIFT, LGHD and proposed method. Root Mean Square Error (RMSE) is used to quantify the accuracy of registered image. The RMSE error of RIFT and LGHD methods were reported as 7.98 and 8.78 respectively. The RMSE error of proposed method is 6.12. That means the accuracy of proposed method is superior than the accuracy of RIFT, LGHD methods of registered images.

Table-1 Root Mean Square Error (RMSE) of registered images

	RIFT	LGHD	Proposed method
I	7.9883	8.7859	6.1255
II	6.7255	4.1871	3.8161
III	6.1472	6.9792	5.269
IV	7.6891	8.3395	6.9597
MEAN	7.1375	7.0729	5.5424

4.2 Visual evalution of Proposed method

Fig-2 shows the image registration results of a proposed registration method. The visible and infrared images are shown in Fig-2(a), (b) respectively. The proposed method's registration results are shown in Fig-2(c) and it is observed that the geometric differences and scaling differences are reduced.

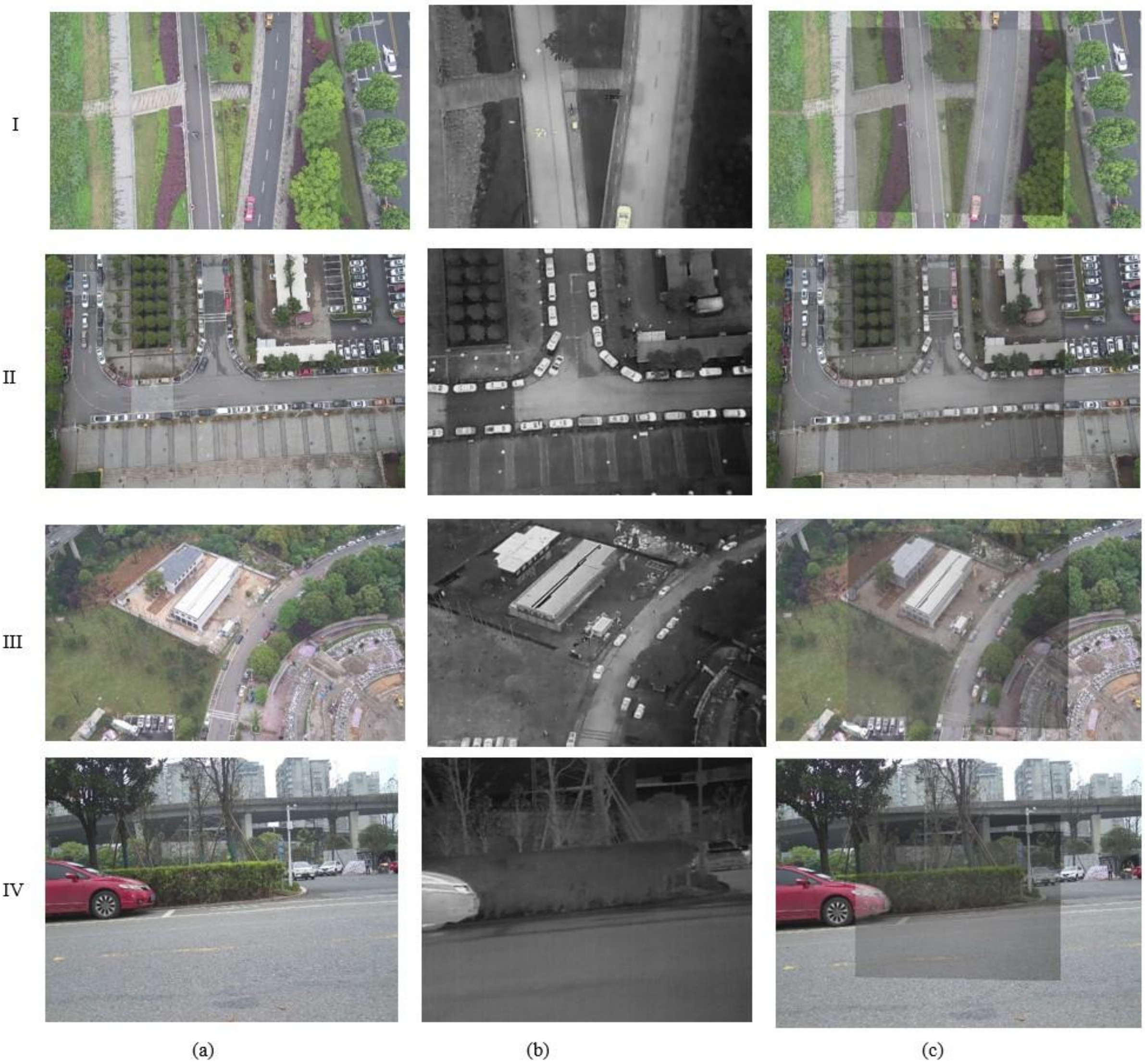

Fig-2: Visual evaluation on the VIR-UAV dataset (a) Visible images. (b) infrared images. (c) Registered images of Proposed method.

V. Conclusion

In this work, a novel hybrid coarse to fine registration method is presented for registering visible and infrared images. In the coarse registration, the scaling differences are reduced by using Affine transformation. The proposed method uses hybrid ASTC descriptor which is a combination of the AHE, ASS feature computation and integrated with CFOG and 3-D convolution optimization. Finally in the fine registration, the feature points are matched based on the FFT technique. The proposed registration method overcomes the scaling differences, intensity difference, and illumination differences. The performance of the registered results is measured by Root Mean Square Error (RMSE). The RMSE error is reduced from 8.78 to 6.12. It is also observed that the proposed method is superior than the RIFT, LGHD.

References

[1]. S. Jung and W. Kim, "Development of an Unmanned Aerial System for Maritime Environmental Observation," in *IEEE Access, vol. 9*, pp. 132746-132765, 2021.

[2]. B. Zhu, L. Zhou, S. Pu, J. Fan and Y. Ye, "Advances and Challenges in Multimodal Remote Sensing Image Registration," in *IEEE Journal on Miniaturization for Air and Space Systems, vol. 4, no. 2,* pp. 165-174, June 2023.

[3]. Y. Byun, J. Choi and Y. Han, "An Area-Based Image Fusion Scheme for the Integration of SAR and Optical Satellite Imagery," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 6, no. 5*, pp. 2212-2220, Oct. 2013.

[4]. J. Li, Q. Hu and M. Ai, "RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform," in *IEEE Transactions on Image Processing, vol. 29*, pp. 3296-3310, 2020.

[5]. B. Debaque et al., "Thermal and Visible Image Registration Using Deep Homography," 2022 25th *International Conference on Information Fusion (FUSION)*, Linköping, Sweden, 2022.

[6]. W. Yang *et al*., "Adjacent Self-Similarity 3-D Convolution for Multimodal Image Registration," in *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1-5, 2024.

[7]. Kovesi, P. (1999). Image Features from Phase Congruency. *Videre: Journal of Computer Vision Research*, *1*(3), 1-26.

[8]. Y. Mo, X. Kang, S. Zhang, P. Duan and S. Li, "A Robust Infrared and Visible Image Registration Method for Dual-Sensor UAV System," in *IEEE Transactions on Geoscience and Remote Sensing, vol. 61,* pp. 1-13, 2023.

[9]. A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process.-Syst. Signal, Image, Video Technol., vol. 38, no. 1*, pp. 35–44, Aug. 2004.

[10]. X. Xiong, G. Jin, Q. Xu, H. Zhang, L. Wang, and K. Wu, "Robust registration algorithm for optical and SAR images based on adjacent self-similarity feature," *IEEE Trans. Geosci. Remote Sens., vol. 60*, 2022, Art. no. 3197357.

[11]. Lisa Gottesfeld Brown, "A Survey of Image Registration Techniques" *ACM Computing Surveys, Vol 24*, *No. 4*, December 1992.

[12]. Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins "*Digital Image Processing Using MATLAB".*

[13]. C. A. Aguilera, A. D. Sappa and R. Toledo, "LGHD: A feature descriptor for matching across non-linear intensity variations," 2015 *IEEE International Conference on Image Processing (ICIP)*, Quebec City, QC, Canada, 2015.

[14]. J. Li, Q. Hu and M. Ai, "RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform," in *IEEE Transactions on Image Processing, vol. 29*, pp. 3296-3310, 2020.

[15]. Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11*, pp. 9059–9070, Nov. 2019.