International Journal of Advanced Technology and Engineering Exploration, Vol 10(101)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2022.10100139

Research Article

Feature based image registration using CNN features for satellite images having varying illumination level

Laukikkumar K. Patel[1*] and Manish I. Patel[2]
Research Scholar, Department of Electronics and Communication Engineering, Sankalchand Patel University, Visnagar-384315, India[1]
Assistant Professor, Department of Electronics and Communication Engineering, Institute of Technology, Nirma University, Ahmedabad-382 481, India[2]

Received: 08-September-2022; Revised: 02-April-2023; Accepted: 06-April-2023

Abstract
Many times, meaningful information is derived from the image fusion. In applications like change detection, cancer growth detection, etc., there is a need of alignment of two or more images first. In the image registration process, two images are geometrically aligned, which is an important pre-processing step in the fields such as remote sensing, medical, etc. This paper focuses specifically on satellite images, which are commonly used in applications like change detection, weather forecasting, and growth monitoring. In these applications, image registration is a crucial step, and the accuracy of the registration process is essential. However, there are various challenges to image registration, one of them is illumination change in multi-sensor, multi-spectral satellite images. To address this challenge, paper proposes a feature based approach, where feature detection using speeded up robust feature (SURF) and descriptor from modified visual geometry group (VGG16) convolutional neural network (CNN) structure are used. The descriptors are generated from the initial convolutional layers of the modified VGG16 structure for each key point detected by SURF. The main goal of this approach is to reduce incorrect matches, which in turn improves image registration. Results of this experiment demonstrate 20% to 40% of significant improvement in correct match rate (CMR), which in turn improves image registration by the proposed approach, as compared to the method where only the original SURF is used for feature detection and descriptor generation. Therefore, it is found that, the use of CNN features as descriptor with SURF as feature detector provides improved results in terms of CMR and thus improves image registration compared to the taken method in comparison. This shows that the use of learned feature as descriptor has potential to improve the image registration.

Keywords
Image registration, Speeded up robust feature (SURF), Convolutional neural network (CNN), VGG16, Correct match rate (CMR).

1.Introduction
In the image registration, the sensed image is aligned to the reference image to find the subtle changes in the two images. These images may be of the same scene, but it can be multi-view or multi-sensor or multi-temporal. Various fields in which image registration is required are remote sensing, where registration is required for image mosaic, landscape planning, etc. The medical field, where image registration is required for monitoring of tumor growth/treatment, magnetic resonance image, etc., and computer vision, where image registration is required for automatic change detection, target template matching, etc.

Based on various criteria, image registration methods can be categorized. Brown divided image registration into four classes based on image acquisition [1], while Zitova and Flusser classified image registration techniques into area-based methods and feature-based methods [2]. Area based methods are applied on pixel intensity values which provide distinctive information. Feature based methods use features of the image where distinctive information is provided

*Author for correspondence

by features of the image like corners, edges, lines, blobs etc. In this paper, feature based method is used.

Generally, there are four steps in the majority of image registration methods [2]. Detection of feature, matching of feature, estimation of transform model, image re sampling and transformation. In feature based image registration, different features like edges, corners, etc. are detected and then they are matched to estimate parameters of geometric transformation. It is very important to select an appropriate method for feature detection, feature descriptor generation, and feature matching for better accuracy of image registration because an error in any of the steps of image registration will propagate to the next stage and reduce the accuracy of registration of image.

There are various challenges to image registration, but one of them is illumination change in multi-sensor, multi-spectral satellite images [3], which can reduce correct matches and in the process of image registration, it is important to have more number of correct matches for better estimation of transformation parameters, which lead to a good image registration. So, it is necessary to select appropriate feature detection and description methods for different applications to obtain more correct matches for improvement in the correct match rate (CMR) and image registration, which provided motivation to find different combination of methods for improvement in CMR.

For feature detection and descriptor generation, various methods are available, like scale invariant feature transform (SIFT) [4], speeded up robust feature (SURF) [5], oriented fast and rotated brief (ORB) [6], etc. A comparative study of some methods is found in [7], where five methods of feature detection are compared for rotation, blur, scale change, illumination change, affine transformations. Also in [8], a survey of various methods for detection and description of handcrafted to learning based features is provided. There are some papers which show a comparison of SIFT and SURF and advantages of SURF over SIFT [9, 10]. So here, in this paper, SURF is used for the feature detection process. There are some SURF based approaches available that provided improvised results like, in [11], more matching points extracted by the normalized SURF compared to the original SURF, in [12], marginal improvement is found in approach, where features are detected by smallest univalue segment assimilating nucleus (SUSAN) and described by SURF algorithm, in [13], CMR for multi-modal images is improved by modifying SURF descriptor according to gradient reversal, etc.

In the last decade, machine learning methods had given good results for applications related to image processing. In machine learning, the machine is trained to learn attributes of local features from examples given for training and after the completion of a training, test images are given to the machine to find local features from test images based on examples given in a training. Deep learning is a subfield of machine learning, provided good results in remote sensing whose survey is found in [14–16]. Deep learning based approaches like deep neural network (DNN), convolutional neural network (CNN), etc. can be used for one or more steps of the image registration process. In deep learning, CNN has achieved more attention in different deep learning methods and also provided good results for remote sensing images [17]. There are some approaches, where CNN is used for different steps of image registration like, in [16], some approaches are proposed, where different CNNs are used in image registration process and discussed its effects. In [18], CNN is used with SIFT and it provided improved results for taken cases. As mentioned in [18], lower layers of CNN detect low-level features and higher layers detect high-level features, instead of using high level features used in [18], features from some initial layers are used for feature descriptor generation in our proposed approach here. So, one of the purposes of proposed approach is to modify CNN model for training purpose. Another purpose is to use features from initial different convolutional layers of modified CNN structure. These features are used to generate feature descriptor for each keypoint generated by traditional method SURF to increase CMR and hence to improve image registration, where images are having illumination level change due to which correct matches are reduced.

Hence, due to the advantages of SURF over SIFT [9,10] and the availability of different levels of features from CNN [18], the proposed approach in this paper includes SURF as a feature detector and CNN (visual geometry group (VGG16)) for descriptor generation, to address the issue of illumination level change in reference and sensed satellite images. The results of the proposed approach are compared with approach, where SURF is used for feature detection and feature descriptor generation. Methods for the rest of the steps of image registration like feature matching, transformation parameter estimation and image re sampling and transformation, are same for both approaches. The proposed approach shows improvement in CMR and thus improvement in image registration due to the use of CNN features used in the proposed approach for images having illumination change. So, the main contributions of the paper are

- Modification of the original VGG16 structure.
- Novel approach to generate feature descriptor from different convolutional layers of modified VGG16 structure.
- Improvement in CMR and hence improvement in image registration.

The remaining sections of the paper are organized in five more sections. Section-2 covers brief discussion of SURF and CNN related work and review of different approaches for image registration, section-3 describes proposed CNN modification and the proposed approach, section-4 includes description of datasets used, experimental setup and simulation results, section-5 includes discussion of obtained results and limitation and section-6 consist of conclusion and future work.

2.Literature review

In literature review, initially overview of SURF and CNN is provided followed by review of different approaches related to the image registration.

SURF [5] is rotation and scale invariant that is derived from SIFT, and compared to SIFT, it is fast. In SURF, Hessian matrix based feature detection is there. It uses an integral image and box filter for speeding up calculation and uses Haar wavelet response in orientation assignment. In the first step to extract feature descriptor, square region is constructed around keypoint where window size is kept 20s. This region is divided in 4×4 square sub-regions regularly. Then Haar wavelet responses (dx in horizontal direction and dy in vertical direction) for each sub-region is calculated and weighted with Gaussian. Also, the sum of absolute values of Haar wavelet response |dx|, and |dy| are calculated. Thus, descriptor vector for each sub-region is created which is (Σdx, Σdy, Σ|dx|, Σ|dy|) and then the combined descriptors of all 4×4 sub-regions provide descriptor having length 64 for each interest point.

CNN is a class of artificial neural network in deep learning. CNN has good results for image classification [19], image retrieval [20], etc. CNN consists of various types of layers, like convolutional layer, fully connected layer, pooling layer, etc. In convolutional layer, input is convolved with filters. Pooling layer reduces the size of data means size of feature maps. Generally, two types of pooling are in use, max pooling and average pooling. By fully connected layer, each neuron of one layer is connected to each neuron of the other layer. Currently, various CNNs are available, like AlexNet [21], VGGNet [22], GoogleNet [23], etc. In this paper, from VGGNet, VGG16 architecture is used, which is modified here at fully connected layers and used for feature descriptor generation.

For the image registration process, SIFT and SURF have been widely used in the last decade. These methods use low level information of image for image registration. Because of different characteristics of satellite images like it can be multi-sensor, multi-spectral etc and large size of image, some problems are faced by conventional algorithms of image registration which are used for computer vision and medical applications. SURF also provides incorrect matches. Some traditional SURF based approaches for image registration are reviewed here. As per [11], the proposed normalized SURF descriptor keeps invariance of rotation and scale, and impact of hue in remote sensing images is also reduced. Compared to the original SURF, more matching points can be extracted by the normalized SURF, but the robustness and stability of this normalized SURF require further study. In [12], features are detected by SUSAN and described by SURF algorithm found marginal improvement. Also proposed algorithm provides advantages like reduction in calculation, and increase in speed. In [13], a novel method for multi-modal image registration is proposed to provide higher speed and less time consumption than other multi-modal image registration algorithms taken in paper and results showed robustness to noise, rotation, luminance variations, and blurring. In this method, according to gradient reversal, SURF descriptor is modified and for multi-modal images, it improves the CMR. In [24], three feature detection and matching methods, (1) SIFT, (2) principal components analysis (PCA)-SIFT, and (3) SURF, are summarized, and it is found that SURF is faster among three and finally, they used SURF method for image registration of multi-view satellite image registration. In [3], histogram of oriented gradient (HOG) based descriptor with SURF detector reduces incorrect matches and improved CMR than SURF detector with its original Haar response based descriptor for images having variation in illumination, like multi-sensor and multi-spectral satellite images. Authors used four different types of satellite image datasets and compared CMR between proposed method and SURF with Haar response based descriptor, where it is found that the proposed approach provides improved results. Also, estimation of transformation parameter like translation is done which provides improved result.

In [25], SURF descriptor is improved by using the richness of color in the image. The proposed change in SURF descriptor uses different color spaces where SURF-HSV3 uses Hue, Saturation, Value (HSV) color space (3rd plan), SURF-NTSC2 uses National Television System Committee (NTSC) space (2nd plan) to compute descriptor. And apart from transition to other color space, SURF-HISTEQ uses color's histogram equalization before transformation to gray scale from color image to improve SURF descriptor. Depending on the scene's nature, these three different descriptors provide good results than other techniques taken and are robust to rotation, scale, and illumination changes. The proposed approach is used to register remote sensing images. In [26], SURF based approach is used to register unmanned aerial vehicle (UAV) and light detection and ranging (LiDAR) images. Results showed best registration by use of the red band (Band-3) from input data. In the proposed approach, canny and sobel edge detectors are used in pre-processing, which provided increment in features detected by SURF but it was found that it provided more noise and overall ineffectiveness. In [27], performance evaluation of different descriptors in combination with SIFT and SURF as feature detectors is done, where it is found that fast retina keypoint (FREAK) with SURF provide better results than other combinations taken in comparison in terms of structural similarity and visual quality. In [28], optimized SURF is used for image registration, where corners are detected by Shi-Tomasi algorithm and then SURF is used for descriptor generation for detected corner points. In the proposed approach, matching process is carried out by bi-directional matching algorithm. The proposed algorithm provided nearly the same accuracy of registration but time consumption of registration is reduced compared to methods taken comparison. In [29], synthetic aperture radar (SAR) image registration is done where multi-level features from accelerated segment test (FAST) (MFAST) is used for extraction of key points, SURF is used for descriptor generation and improved random sample consensus (RANSAC) is used to remove mismatched feature points to address high mismatch rate issue in image registration of SAR images. The proposed approach provided more than 3 times registration accuracy than SURF algorithm. In [30], image registration of printed circuit board (PCB) is done using improved SURF algorithm to solve the issue of time consumption and low accuracy of matching. The proposed approach uses Shi-Tomasi algorithm instead of using SURF for extracting feature points and descriptor is generated by SURF. After completion of matching, progressive sampling consensus (PROSAC) instead of RANSAC for refining matched feature point pairs. In comparison with traditional SURF, proposed approach reduces time of registration of image by 34%, and efficiency of registration is also improved. In [31], improved SURF based image registration algorithm is proposed in wavelet domain. In this process, by using wavelet transform, reference and sensed images are decomposed in low and high frequency components and low frequency component is used as input image to improved SURF. The proposed approach in this paper provides improvement in registration speed and accuracy of registration compared to the methods taken for comparison. In [32], the proposed method uses SURF features and local cross correlation information for image registration. In this approach, first, features are extracted using SURF for initial rough registration. Then, by using local key area's correlation coefficient, homography matrix is calculated, which is applied to rough registered image for rotation transformation. The proposed approach provides better robustness and higher accuracy of registration compared to the methods taken for comparison.

There are also some approaches that include a deep learning technique for image registration, where different deep learning models are used for various purposes. In [16, 18, 33–37], different models are used for feature descriptor generation purpose. In [38], a deep leaning model is used for semantic segmentation. In [39], a self supervised deep learning network for SAR image registration is proposed. The proposed network is having three parts, one for detection of feature points, another for matching and third one for unstable point removal. In this approach, used SAR images are multi temporal. Authors in [40] also proposed separate CNN model (self supervised) for multi-modal image registration. The model consists of three blocks, first block does extraction of deep features, second block does optical flow field estimation and third block register image. In [41], deep learning regression model is proposed that learns displacement parameters of four corners of block of sensed images. They also developed a dual deep learning network that share weights for fully extracting features of registration pair image. From the above different deep learning based approaches, our focus is on approaches, which are used for feature descriptor generation from different layers of CNNs. Like, in [18], SIFT and CNN are used in the image registration process where SIFT and CNN features are fused and combined features are integrated in position scale orientation (PSO)-SIFT algorithm, which provides better performance than the other methods compared in the paper. In [16], survey of methods based on deep learning for image registration of different types of images of remote sensing and medical field is presented. They focused on feature based approaches and did experimentation using three different approaches for image registration of aerial images having deformations like variation in scale. The first approach uses CNN (AlexNet) for feature descriptor generation for keypoint detected by SIFT, second approach uses CNN (siamese network) for similarity finding and third approach uses CNN (siamese network) for homography estimation. From this experimentation, they found that promising results are provided by second approach among three approaches for conditions taken in experimentation. In [33], registration of multi-source high-resolution remote sensing images is done using SIFT and residual network (ResNet). The proposed approach is used to improve feature matching accuracy. In this approach, ResNet is trained on ImageNet and then, by using sample set, which is constructed from registered remote sensing images having high resolution, it is fine tuned. Then feature descriptor from SIFT and ResNet are combined, but before combining these two descriptors, a normalization operation is applied because of a difference in descriptors of SIFT and ResNet. In this approach, two ResNets are used, ResNet34 and ResNet50 and results are compared with Patch-SIFT and SIFT. The proposed approach provides an increment in tie points and accuracy of registration compared to the methods taken for comparison. In [34], image registration of multi-temporal images is done, where robust feature descriptor from CNN (VGG16) is generated and dynamic inliers selection process is introduced, which provided improvement in feature point registration. In this, multi-temporal satellite image and unmanned aerial vehicle image datasets are used and the approach provides improved results compared to four SIFT based methods. In [35], a twostep image registration method is proposed that is based on local and deep features. In first step, approximate spatial relation is calculated by CNN. Once the approximate transformation matrix is calculated, local feature based method is used to improve matching and their proposed method finds best match from top-10 nearest neighbours in second step. This approach uses deep and local features both, but increases the complexity of algorithm. In [36], a framework based on deep learning is proposed for generation of descriptor for keypoint detected by FAST and also proposed novel loss function for training their model. The proposed method is applied to optical images taken at different time and provided better results than the methods taken for comparison.

From the literature review, it is found that, in traditional methods of image registration for solving the different issues of image registration of various types of images, different methods are used for feature detection or descriptor generation to get the benefits of different methods at a time in image registration. Also, from review of some papers, it is seen that, SURF is having some advantages over SIFT, so SURF based approaches are reviewed here, and found that, SURF with other methods also provides good results in image registration for solving different types of issues. In the last decade, use of deep learning provided good results in image based applications and from literature review it is found that, different deep learning models can be used for various purposes like descriptor generation, semantic segmentation etc. From different purposes to use deep learning models, one of it shows that, CNN features can also be used for feature descriptor generation and can be used separately as descriptor or can be fused with descriptor from traditional methods like SIFT. And this use of CNN features also provided improvised results for image registration. So, in this paper, SURF is used for feature detection purpose and CNN is used for feature descriptor generation to get the advantages of SURF and CNN for improvement in CMR and image registration.

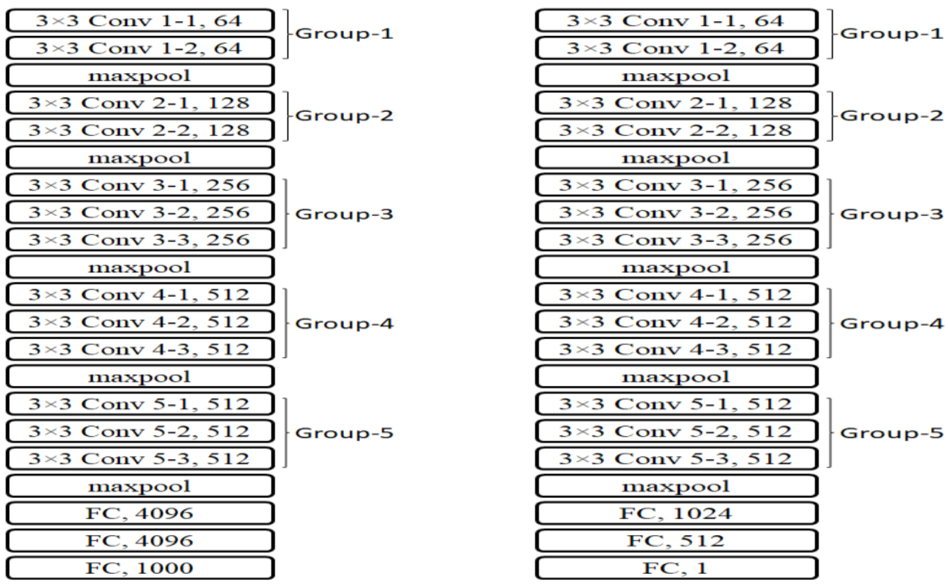
3.Methodology

3.1Proposed CNN modification

We used VGG16 [22] with some modification in our proposed approach from the available different types of CNNs.

In the original VGG16 structure, there are 13 convolutional layers and three fully connected layers. These 13 convolutional layers are segregated into 5 groups, having max-pooling layer after each group. In first and second group, two convolutional layers are there and the rest of the groups consist of three convolutional layers in each group. Once the input image is passed to convolutional layers, filters with small receptive fields of size 3×3 are used. In this structure convolutional stride of 1 pixel is used. In the case of max-pooling layer, 2×2 pixel window is used for max-pooling, and stride is kept to 2 pixels. The number of channels (width of layers) of convolution layers starts from 64 (first layer) and reach up to 512 channels. In three fully connected layers, the first two layers consist of 4096 channels and 1000 channels are there in the third layer.

The original VGG16 [22] structure classifies 1000 categories but here in this approach, the aim is to train model for one dataset at a time, so modification is done at fully connected layers and used modified structure for feature descriptor generation. Each dataset contains one reference image and one sensed image. The modified VGG16 structure shown in *Figure 1* is trained separately for each dataset using 64×64 patches of reference image and augmented patches generated using image data augmentation during training. As original VGG16 structure is modified here, the total parameters for the modified VGG16 structure are 17338177 and parameters at 2nd (Conv 1-2) layer are 36928 and 4th (Conv 2-2) layer are 147584.



(a) (b)

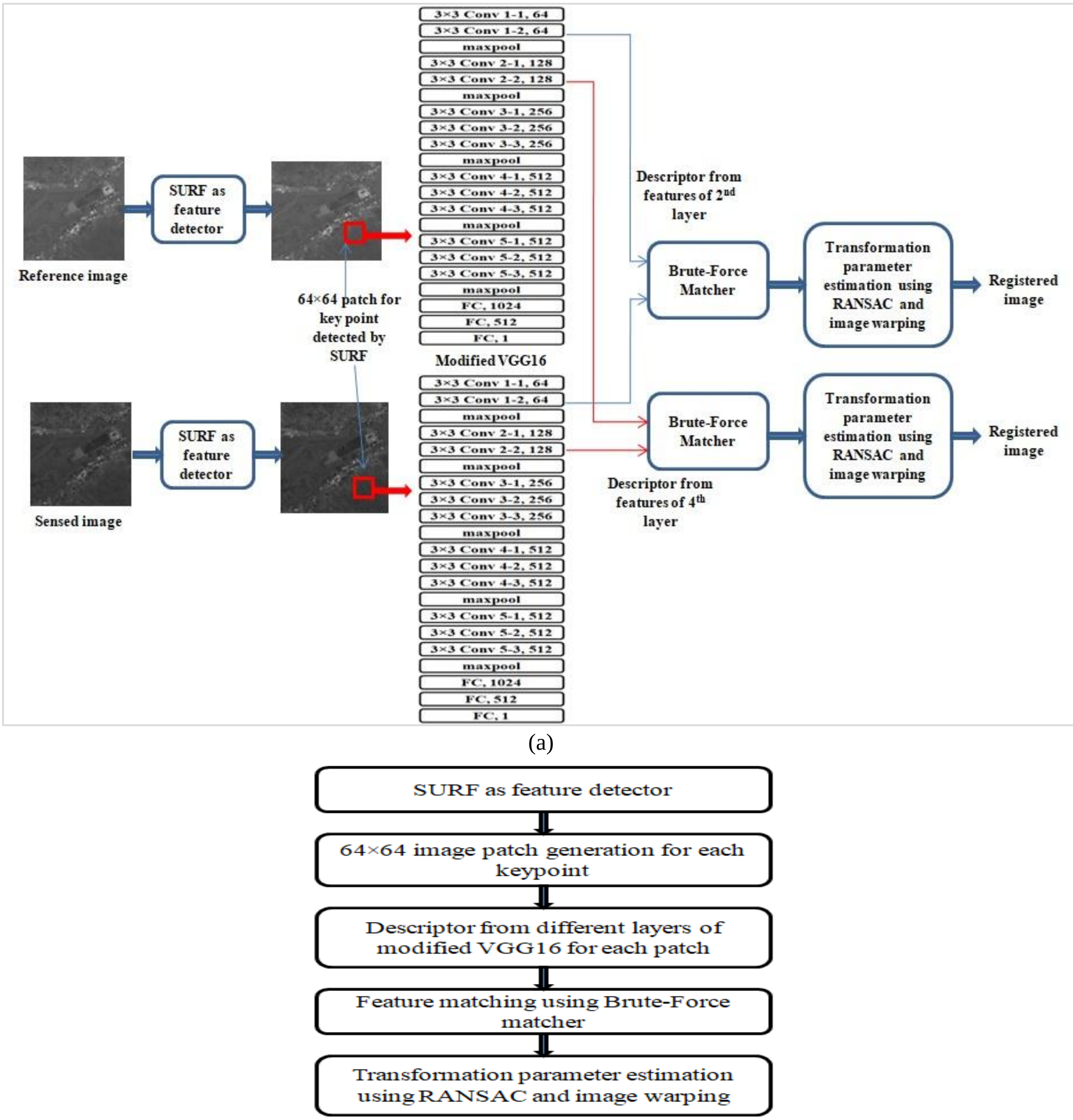
Figure 1 Brief structure of (a) VGG16 [22] and (b) modified VGG16

3.2Proposed method

In the proposed approach, feature detection is done using SURF that provides feature points, which are also known as keypoints. Once features are detected, descriptor is generated from the lower convolutional layers of the modified VGG16 structure. For the generation of descriptor, 64×64 patch is generated for each keypoint. This patch is given as input to the modified VGG16 structure and output is taken from two convolutional layers (2nd layer (Conv 1-2) from Group-1 and 4thlayer (Conv 2-2) from Group-2) of it. So the input to this modified VGG16 structure is 64×64. For the matching process, the Brute-Force matcher [42] is used for matching the generated descriptors. After completion of matching process, total found matches are sorted in ascending order of their Euclidian distance, where matches with low distance come first. In the matching process, descriptors are taken from the same layer for both reference and sensed image (if descriptor from 2nd layer for reference image is taken, then for sensed image 2nd layer descriptor is taken for further process). Once the matching process for selected layer(s) is completed, the CMR is calculated. Framework and steps involved in the proposed approach are shown in *Figure 2.*

For further improvement in CMR, common matched keypoints pairs from two different layers, like from 4th and 2nd convolutional layer (layer 4&2) are searched and if sufficient pairs required for further process in image registration are found then again CMR is calculated for layer 4&2 and improvement compared to single layer is analyzed.

After completion of the matching process, RANSAC [43] is used for outlier removal and transformation parameter estimation. Based on it, sensed image is registered. Comparison of registered image is done on visual bases as only illumination change effect is considered here. In this paper, the proposed approach is compared with the approach, where the original SURF is used for feature detection and descriptor generation both. So the approach, where the original SURF is used for both feature detection and generation of descriptor, is considered as Approach-1 and the proposed approach is considered as Approach-2. The feature matching process, outlier removal, and transformation parameter estimation process for Approach-1 is the same as mentioned in proposed approach (Approach-2).



(b)

Figure 2 (a) Framework of Proposed approach (b) Block diagram of proposed approach with SURF and modified VGG16 (Approach-2)

4.Results

4.1Datasets

In this experiment, four different types of satellite image datasets are taken, where each dataset contains reference image and sensed image. Dataset-1(near bay of Kutch) and Dataset-2 (near Ahmedabad city) are multi-spectral satellite image datasets, taken from linear imaging and self scanning sensor-III (LISS-III) that provides multi-spectral data in four bands, which are used in [3], taken from [44], and the size of it is kept to 300×300. Dataset-3 is used in [3], taken from [45], and it is kept to size 300×259. Dataset-4 is multi-sensor satellite image dataset, where reference image is an aerial photo and sensed image is taken by Indian remote sensing satellite-1C, that is used in [3], taken from [46], and it is kept to size 300×300. Dataset-1, Dataset-2, Dataste-3 and Dataset-4 are shown in first, second, third and fourth row of *Figure 3* respectively. Here, the size of image is reduced to minimize the complexity of calculation, as CNN is used for feature descriptor generation and also in matching process, descriptors of higher size are matched.

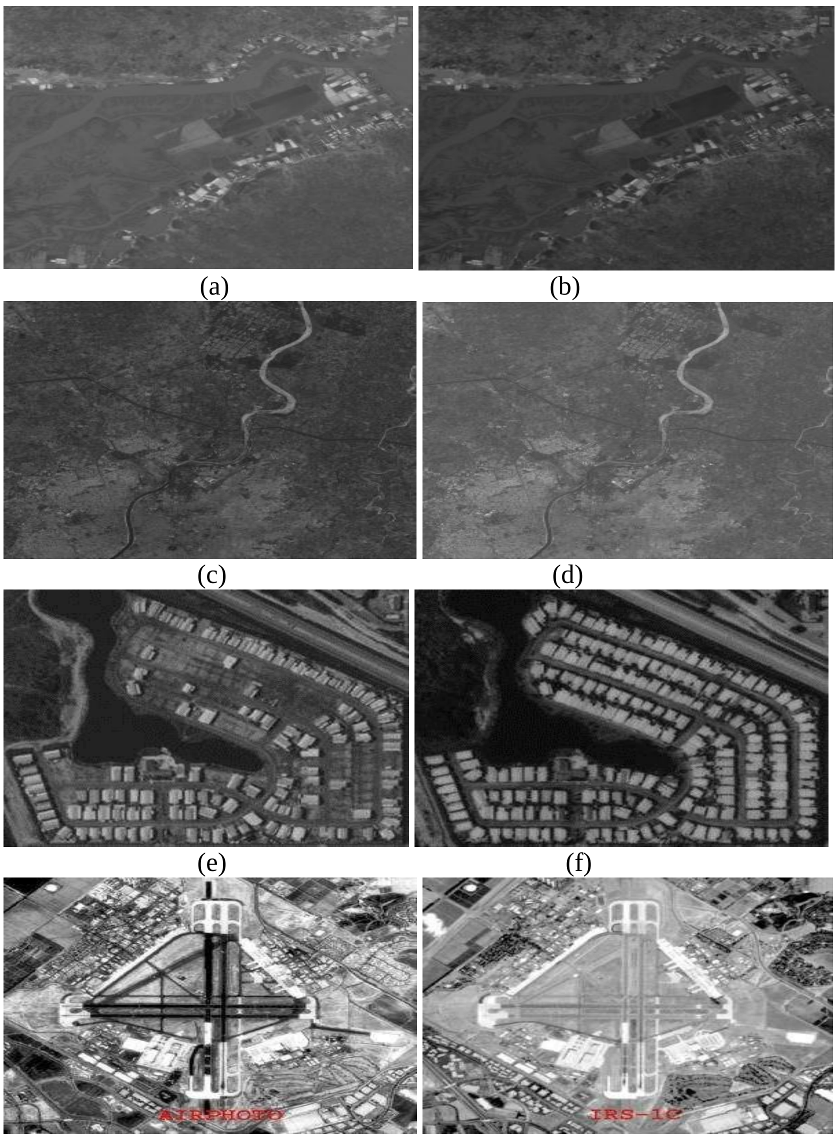


Figure 3 Datasets: (a) reference image- Dataset-1, (b) sensed image- Dataset-1, (c) reference image- Dataset-2, (d) sensed image- Dataset-2, (e) reference image- Dataset-3, (f) sensed image- Dataset-3, (g) reference image- Dataset-4, (h) sensed image- Dataset-4

4.2Experimental setup and simulation results

In this experiment, once the keypoints are detected by SURF, descriptor for each keypoint is generated from 2nd convolutional layer (Conv 1-2), 4th convolutional layer (Conv 2-2) of modified VGG16 structure is taken, where descriptor size of 2nd layer is 262144, and for 4th layer it is 131072. In this experiment, most of the arrays used are of 'integer' type and the data type of values of descriptor used in matching is 'uint8'.

For finding CMR, total matches given by Brute-Force matcher are taken and out of those, correct matches are found. In Brute-Force matcher, L2 norm is used as distance measurement and another parameter crosscheck is kept 'true', which provides only those matches, where two features in both sets match each other. For each keypoint in sensed image, if distance between its mapped point and the corresponding keypoint in reference image in matches is within five pixels, then it is treated as a correct match [13].

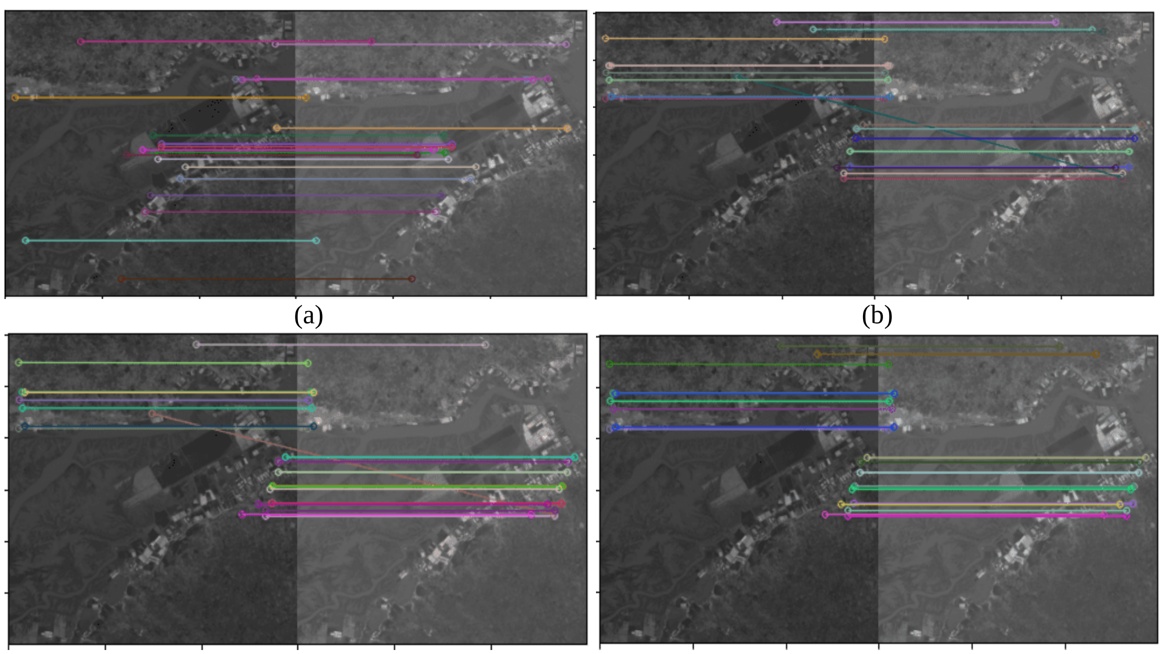
In the experiment, to evaluate the effectiveness of the proposed method quantitatively, CMR is used, which is calculated based on Equation 1.

$$\mathrm{CMR} = \left(\frac{\text{Correct matches}}{\text{Total matches}}\right) \times 100 \qquad (1)$$

In the calculation of CMR, total matches are found from Brute-Force matcher and from the total found matches, correct matches are derived by process mentioned earlier.

Implementation of this experiment is done in pycharm IDE with python 3.7, on intel core i7 processor with 2.80 GHz and 16 GB RAM, and training of modified VGG16 structure is done in google colab with hyperparameters, learning rate=1e-4, epoch-10, batch size (training and validation each) = 5, etc. As training of model is done for single class, so no splitting of patches is applied during training and used same patches for training and testing both.

In *Figure 4*, *Figure 5*, *Figure 6* and *Figure 7*, some matched features are shown using Approach-1 and Approach-2 for different datasets, which shows visualization of true matches and false matches. As the matches are in large number and it is difficult to differentiate in visual representation for the calculation, so some matches are shown for understanding purpose.



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

Figure 4 Matched features of Dataset-1 by (a) Approach-1, (b) Approach-2 (2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)

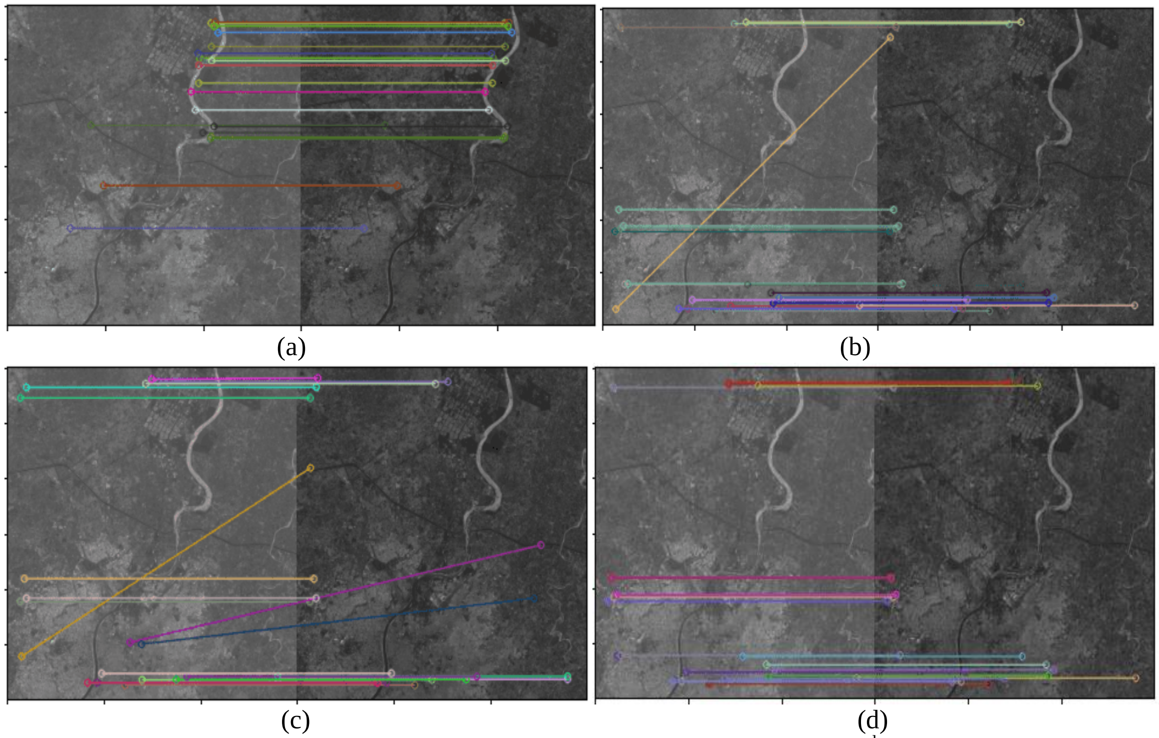


Figure 5 Matched features of Dataset-2 by (a) Approach-1, (b) Approach-2 (2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)

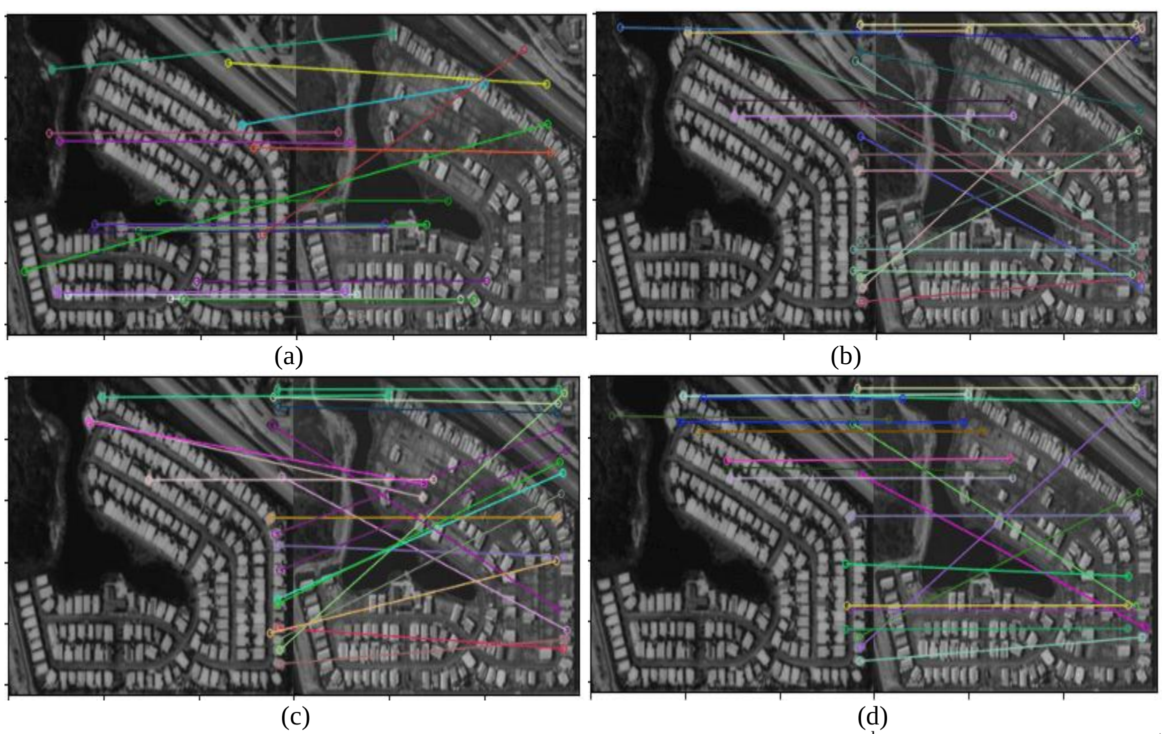
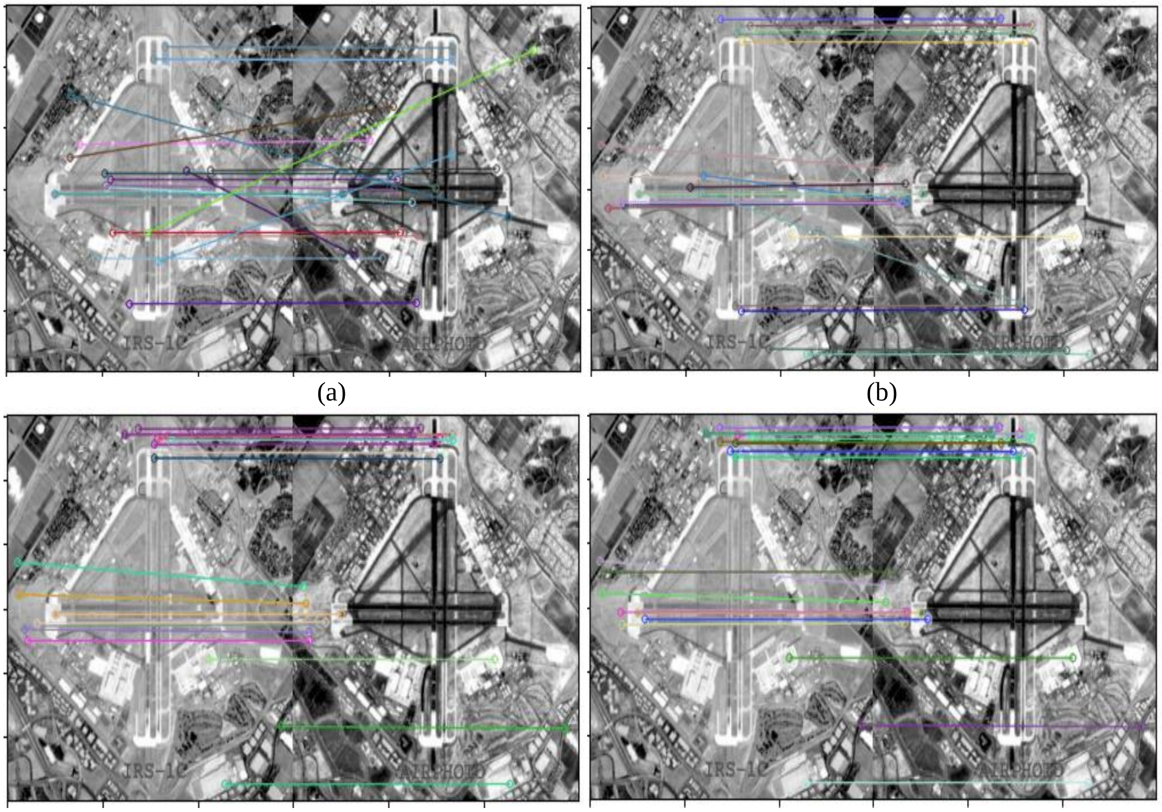


Figure 6 Matched features of Dataset-3 by (a) Approach-1, (b) Approach-2 (2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)



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

Figure 7 Matched features of Dataset-4 by (a) Approach-1, (b) Approach-2 (2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)

After completion of the matching process, generated total matches and correct matches for all datasets are provided in *Table 1* and the results in terms of CMR for all datasets are shown in *Table 2*.

From the simulation results, it is seen that the CMR increases for the proposed approach compared to Approach-1 in major cases used in this experiment. In the proposed approach, for descriptor from 2nd convolutional layer, results are good compared to Approach-1, which can be seen in *Figure 8*, but descriptor from 4th layer provides degraded results for one dataset but for the rest of three datasets, it provides good results compared to Approach-1, which is shown in *Figure 9*. CMR is also further increased by finding common matched keypoints pairs from two layers compared to a single layer, which can be seen in *Figure 10* and *Figure11*. As the correct match increases, it can help in better estimation of transformation parameters and which further improves image registration. Here, four different types of satellite images are taken and in major cases, CMR is increasing for the proposed approach and its comparison with Approach-1 can be seen in *Figure 8, Figure 9, Figure 10,* and *Figure 11*.

Table 1 Total matches and correct matches for Dataset-1 to Dataset-4

Dataset		**Dataset -1**		**Dataset -2**		**Dataset-3**		**Dataset-4**	
Approach		**Total matches**	**Correct matches**	**Total matches**	**Correct matches**	**Total matches**	**Correct matches**	**Total matches**	**Correct matches**
Approach-1		210	183	296	240	375	41	512	47
Approach-2 (Proposed Approach)	2nd layer	181	162	36	30	216	115	375	181
	4th layer	214	200	67	51	280	152	458	259
	layer 4&2	164	162	30	30	146	103	198	157

Table 2 CMR for Dataset-1 to Dataset-4

Approach	CMR in %			
	Dataset-1	**Dataset-2**	**Dataset-3**	**Dataset-4**
Approach-1	87.14	81.08	10.93	9.18
Approach-2 (2nd layer)	89.5	83.33	53.24	48.27
Approach-2 (4th layer)	93.46	76.12	54.29	56.55
Approach-2 (layer 4&2)	98.78	100	70.55	79.29

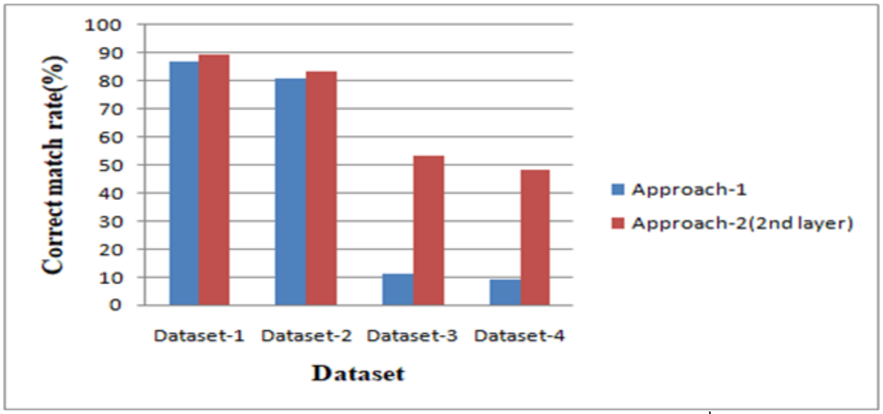


Figure 8 CMR for Datasets with Approach-1 and Approach-2 (2nd layer)

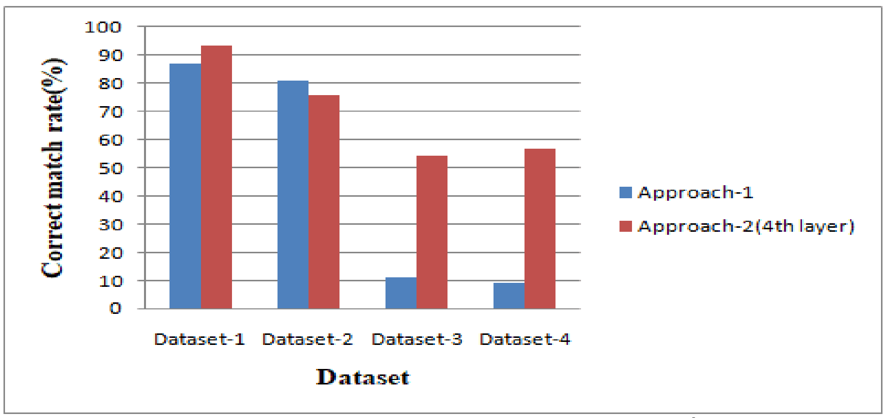


Figure 9 CMR for Datasets with Approach-1 and Approach-2(4th layer)

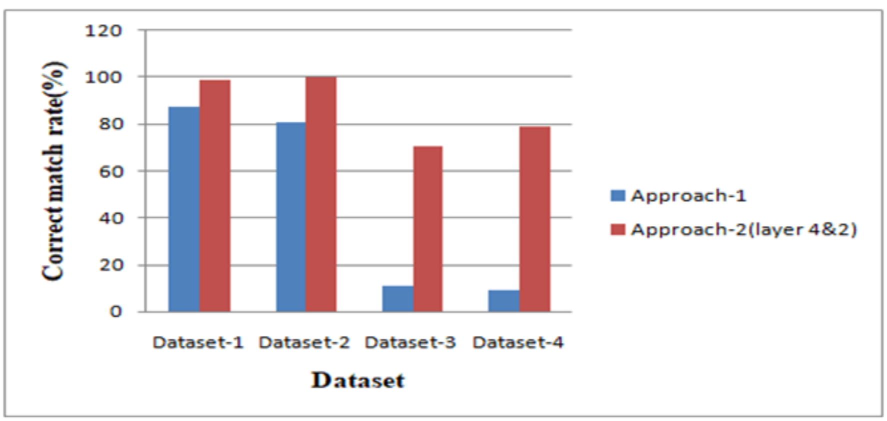


Figure 10 CMR for Datasets with Approach-1 and Approach-2 (layer 4&2)

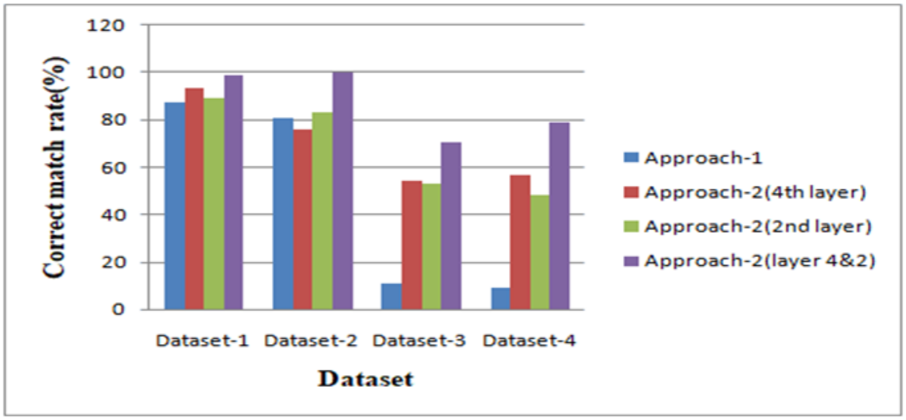
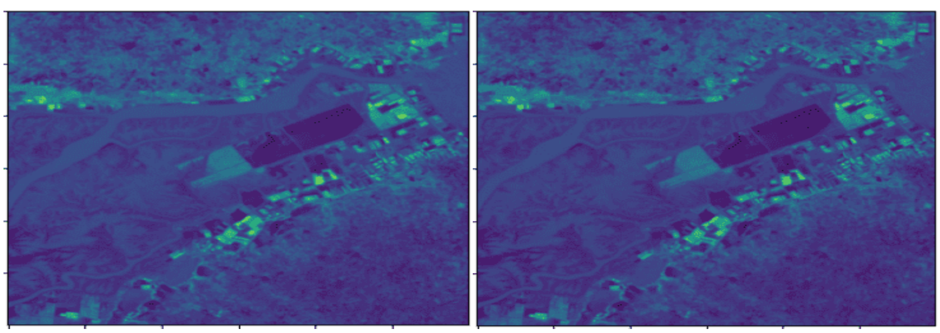
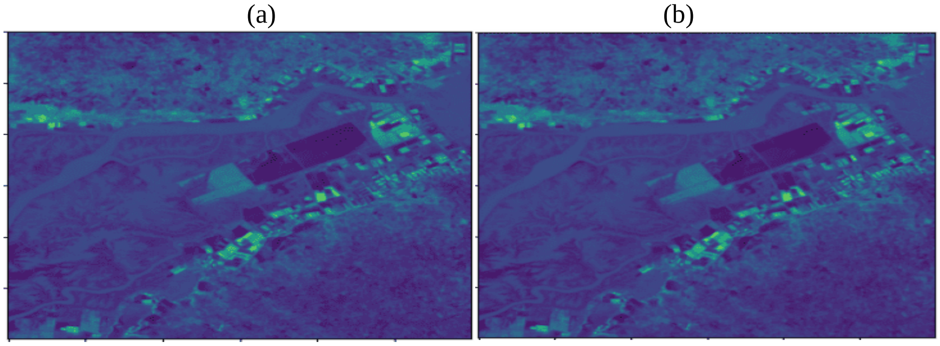


Figure 11 CMR for datasets with Approach-1 and Approach-2

To observe the effect of improved CMR, sensed images in taken datasets are registered. After feature matching, RANSAC is used for outlier removal and transformation parameter estimation. Here, only illumination change effect in reference and sensed image is considered, so comparison is done based on visualization of registered image. In the case of registered images, for Dataset-1 and Dataset-2, visual results of registered images using Approach-1 and Approach-2 seem good, which are shown in *Figure 12* and *Figure 13* respectively. For Dataset-3 and Dataset-4, visual results of Approach-2 are better than Approach-1, which can be seen from *Figure 14* and *Figure 15* respectively. As shown in *Figure 14(a)*, result of registered image by using Approach-1 provides degraded results but results shown in *Figure 14(b)*, *Figure 14(c)*, and *Figure 14(d)* of registered images by Approach-2 provide better results than Approach-1. In *Figure 15(a)*, result of registered image by using Approach-1 provides degraded results but results shown *in Figure 15(b)*, *Figure 15(c)*, and *Figure 15(d)* of registered images by Approach-2 provide better results than Approach-1 which shows the impact of improvement of CMR.

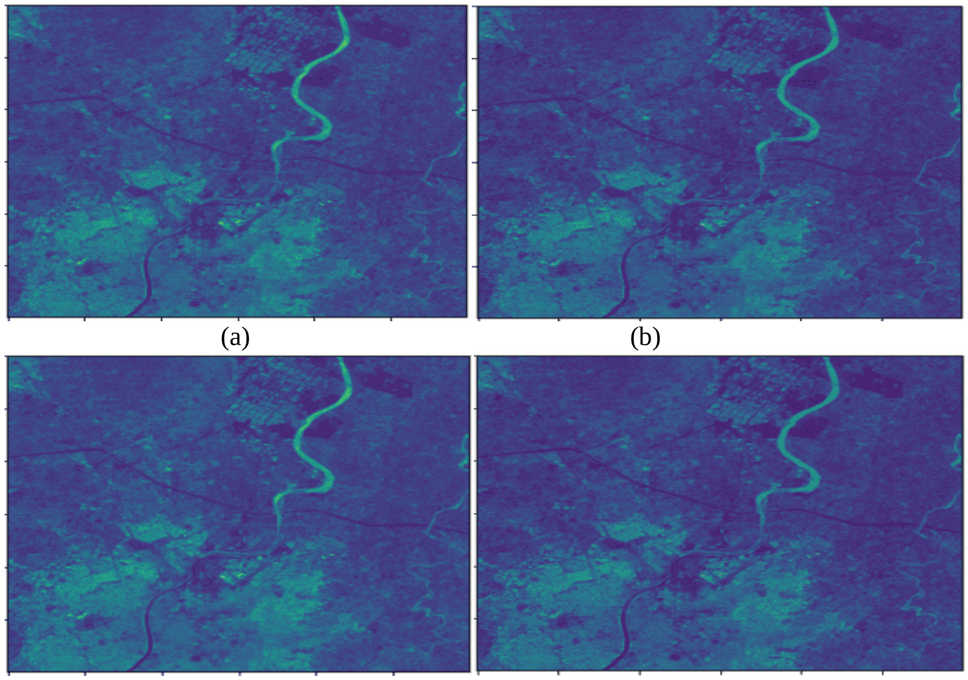


(a) (b)



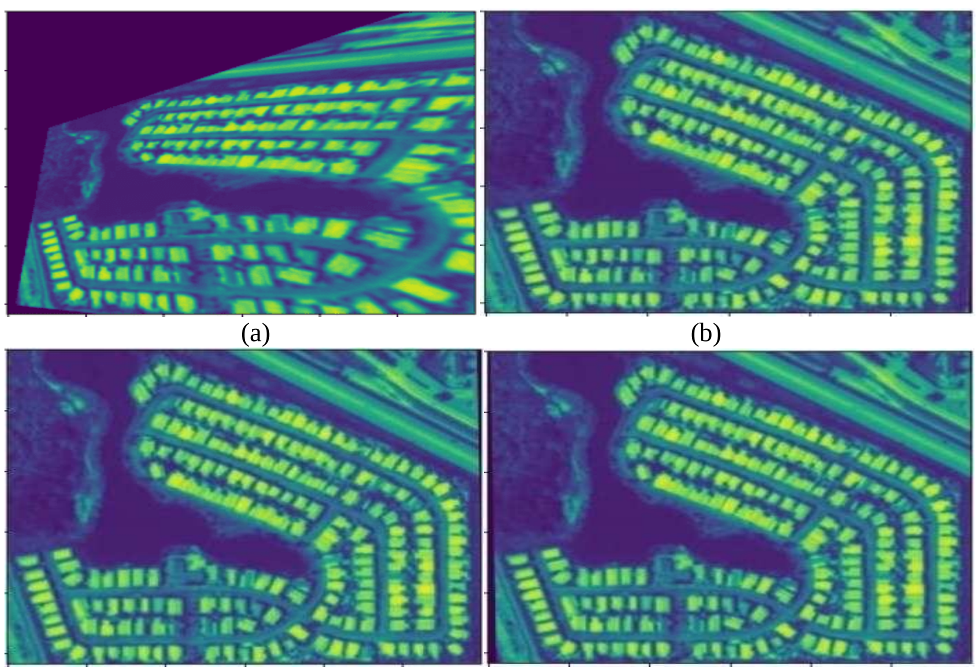
(c) (d)

Figure 12 Registered images for Dataset-1, (a) Approach-1, (b) Approach-2(2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)



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

Figure 13 Registered images for Dataset-2, (a) Approach-1, (b) Approach-2 (2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)



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

Figure 14 Registered images for Dataset-3, (a) Approach-1, (b) Approach-2 (2nd layer), (c) Approach-2 (4th layer), (d) Approach-2 (layer 4&2)

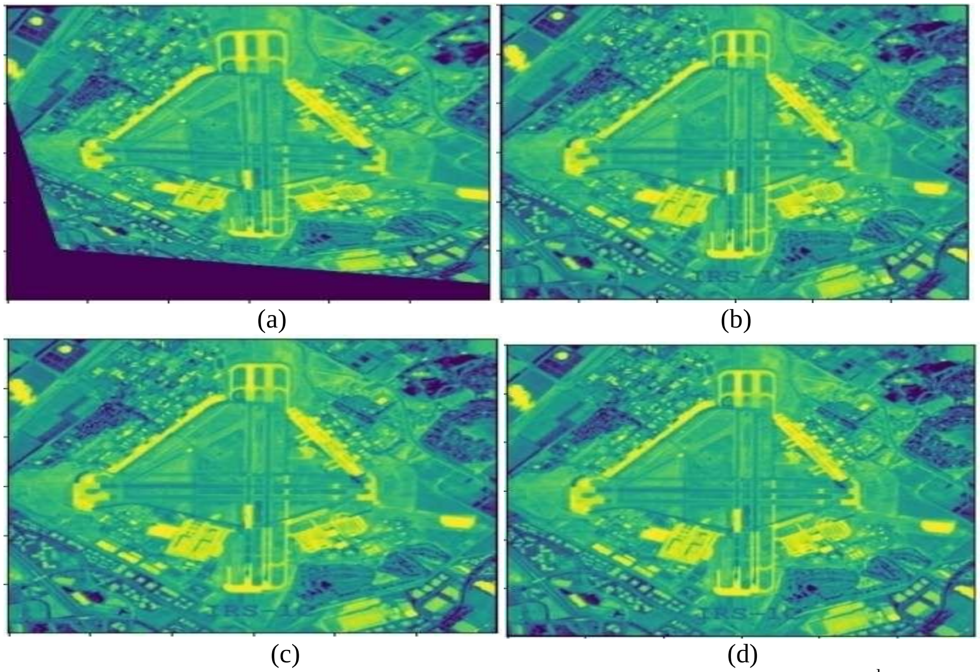


Figure 15 Registered images for Dataset-4, (a) Approach-1, (b) Approach-2 (2^{nd} layer), (c) Approach-2 (4^{th} layer), (d) Approach-2 (layer 4&2)

5.Discussion

From the study of different review and research papers of the related field, it is found that there are various traditional methods used in the image registration process like SIFT, SURF, ORB, etc., which had provided good results but instead of using the same method for feature detection and descriptor generation, the combination of different methods provide even better results due to the use of advantages of different methods on the single platform based on the requirement to solve different issues.

Also, the use of deep learning increased in the last decade for different applications and had provided good results, where one of the applications is image registration. In a few reviewed papers, it is found that the use of convolutional features with traditional methods provide improved results, where different approaches to use convolutional features are proposed. Here, in our paper, convolutional features from some initial convolutional layers (2nd layer and 4th layer) of modified VGG16 layers are used to generate descriptor to get lower level features for preparation of descriptor for each key point detected by SURF. This combination of SURF as feature detector and CNN as feature descriptor generation, provides improved CMR and hence improved image registration compared to SURF used for both feature detection and descriptor generation. So the use of CNN features shows improvement in CMR and thus in image registration. Further improvement is also found in CMR, when common matched keypoints pairs from two different layers, like from 4th and 2nd convolutional layers (layer 4&2) are searched and used for further process of image registration.

As descriptors are generated from different layers (2nd layer and 4th layer), the effect of them on improvement in correct matches, CMR and image registration is analyzed and further improvement by finding common matched points between two layers is also analyzed here. After completion of the matching process, total matches and correct matches provided in *Table 1* are analyzed and it is found that the total matches in case of Approach-1 are more than Approach-2 in major cases. But in case of correct matches, Approach-2 with 2nd layer, 4th layer and layer 4&2 provide more correct matches than Approach-1 for Dataset-3 and Dataset-4. For Dataset-2, Approach-2, 2nd layer, 4th layer and layer 4&2 provide less correct matches than Approach-1. But for Approach-2, as total matches found are less which in turn reduces correct matches but it increases CMR in major cases. For Dataset-1, Approach-2 with 4th layer provide more correct matches and Approach-2 with 2nd layer and layer 4&2 provide less correct matches as total matches found are less which in turn reduces correct matches but it increases CMR.

The results in terms of CMR for all datasets are shown in *Table 2*. For Dataset-1, Approach-1 provides CMR of 87.14 and in case of the proposed approach, it reaches to 89.5 and 93.46 for 2nd layer and 4th layer respectively, which is further improved to 98.78 by finding the common matched keypoints pairs (layer 4&2). For Dataset-2, Approach-1 gives CMR of 81.08 and Approach-2 with 2nd layer provides improved CMR of 83.33 and further improved to 100 for layer 4&2. For Dataset-3, Approach-1 provides CMR of 10.93 and Approach-2 gives 53.24, 54.29 for 2nd layer and 4th layer respectively, which is further improved to 70.55 for layer 4&2. For Dataset-4, Approach-1 provides CMR of 9.18 and Approach-2 with 2nd layer provides CMR of 48.27 and Approach-2 with 4th layer provides CMR of 56.55 which is further improved to 79.29 for layer 4&2.

As datasets are changing, there is a variation in comparative results of two approaches which can be seen from *Figure 8* to *Figure 11*, which can be due to differences of image contents. So, in case of Approach-2, average improvement (taking average of % improvement achieved compared to Approach-1) in CMR for all four datasets taken in this experiment is around 21% for 2nd layer, around 23% for 4th layer and around 40% for layer 4&2. So it is concluded that Approach-2 provides improvement of around average 20% to 40% in CMR for taken datasets.

In this experiment, all the datasets are having change in illumination level between reference and sensed images. For Dataset-1 (multi-spectral images) and Dataset-2 (multi-spectral images), there is a good improvement in CMR in major cases compared to Approach-1, but for Dataset-3 and Dataset-4 (multi-sensor images), there is much more improvement in CMR compared to Approach-1 and improvement of CMR is also reflecting in registered images, where Approach-2 provide better registered images than Approach-1, which shows improvement in CMR also affects the improvement in image registration.

Limitation

In this paper, the proposed approach is applied to the satellite images having varying illumination level, where a modified VGG16 model is trained with available satellite images in our experiment. Here, modified VGG16 structure is trained for each dataset separately so, for different datasets, structure need to be trained again and it becomes time consuming. Also, the proposed approach may not be generalized to all the fields because it may not work well for computer vision or medical field images as during training of CNN only satellite images are considered. Our proposed approach with CNN is compared with the traditional SURF method, so computational complexity and time for completing registration process is more than the traditional SURF method due to involvement of CNN.

A complete list of abbreviations is shown in *Appendix I.*

6.Conclusion and future work

In the proposed approach, feature descriptor is generated from lower convolutional layers (2nd layer and 4th layer) of modified VGG16 structure for features detected by SURF, instead of using descriptor from SURF, which is used in Approach-1. This proposed approach provides improvement in CMR compared to SURF used for feature detection and descriptor generation for images having illumination level change like multi-sensor and multi-spectral satellite images. Further improvement in CMR is also noted by finding common matched pairs in two different layers in Approach-2. Compared to Approach-1, proposed approach in this paper gives good improvement in CMR in the range of 20% to 40%, which results into improvement in image registration. In future work, rotation and translation parameter estimation related issues in image registration will be addressed for different datasets.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contributions statement

Laukikkumar K. Patel: Methodology, simulations, interpretation of results, original draft writing, review and editing. **Manish I. Patel:** Interpretation of results, review and editing.

References

[1] Brown LG. A survey of image registration techniques. ACM Computing Surveys. 1992; 24(4):325-76.
[2] Zitova B, Flusser J. Image registration methods: a survey. Image and vision computing. 2003; 21(11):977-1000.
[3] Patel MI, Thakar VK, Shah SK. Image registration of satellite images with varying illumination level using

HOG descriptor based SURF. Procedia Computer Science. 2016; 93:382-8.
[4] Lowe DG. Distinctive image features from scale-invariant key points. International Journal of Computer Vision. 2004; 60:91-110.
[5] Bay H, Tuytelaars T, Van GL. Surf: speeded up robust features. Lecture Notes in Computer Science. 2006; 3951:404-17.
[6] Rublee E, Rabaud V, Konolige K, Bradski G. ORB: an efficient alternative to SIFT or SURF. In 2011 international conference on computer vision 2011 (pp. 2564-71). IEEE.
[7] El-gayar MM, Soliman H. A comparative study of image low level feature extraction algorithms. Egyptian Informatics Journal. 2013; 14(2):175-81.
[8] Joshi K, Patel MI. Recent advances in local feature detector and descriptor: a literature survey. International Journal of Multimedia Information Retrieval. 2020; 9(4):231-47.
[9] Sreeja G, Saraniya O. A comparative study on image registration techniques for SAR images. In 5th international conference on advanced computing & communication systems 2019 (pp. 947-53). IEEE.
[10] Mistry D, Banerjee A. Comparison of feature detection and matching approaches: SIFT and SURF. GRD Journals-Global Research and Development Journal for Engineering. 2017; 2(4):7-13.
[11] Kai W, Bo C, Lu M, Song X. Multi-source remote sensing image registration based on normalized SURF algorithm. In international conference on computer science and electronics engineering 2012 (pp. 373-7). IEEE.
[12] Huiqing Z, Lin G. Image registration research based on SUSAN-SURF algorithm. In the 26th Chinese control and decision conference 2014 (pp. 5292-6). IEEE.
[13] Zhao D, Yang Y, Ji Z, Hu X. Rapid multimodality registration based on MM-SURF. Neurocomputing. 2014; 131:87-97.
[14] Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine. 2017; 5(4):8-36.
[15] Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing. 2019; 152:166-77.
[16] Kuppala K, Banda S, Barige TR. An overview of deep learning methods for image registration with focus on feature-based approaches. International Journal of Image and Data Fusion. 2020; 11(2):113-35.
[17] Liu X, Han F, Ghazali KH, Mohamed II, Zhao Y. A review of convolutional neural networks in remote sensing image. In proceedings of 8th international conference on software and computer applications 2019 (pp. 263-7).
[18] Ye F, Su Y, Xiao H, Zhao X, Min W. Remote sensing image registration using convolutional neural network features. IEEE Geoscience and Remote Sensing Letters. 2018; 15(2):232-6.
[19] Marmanis D, Datcu M, Esch T, Stilla U. Deep learning earth observation classification using imagenet pretrained networks. IEEE Geoscience and Remote Sensing Letters. 2015; 13(1):105-9.
[20] Zhou W, Newsam S, Li C, Shao Z. Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. Remote Sensing. 2017; 9(5):1-20.
[21] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017; 60(6):84-90.
[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ICLR 2014 (pp.1-14).
[23] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
[24] Shah U, Mistry D, Banerjee A. Image registration of multiview satellite images using best features points detection and matching methods from SURF, SIFT and PCA-SIFT. Journal of Emerging Technologies and Innovative Research. 2014; 1(1):1-13.
[25] Bouchiha R, Besbes K. Automatic remote-sensing image registration using SURF. International Journal of Computer Theory and Engineering. 2013; 5(1):406-10.
[26] Al-najjar H. UAV and lidar image registration: a SURF-based approach for ground control points selection. ISPRS Geospatial Week. 2019:413-18.
[27] Aarathi MR, Raju J. Influence of different descriptors to enhance image registration techniques using FREAK: case study. In 1st international conference on innovations in information and communication technology 2019 (pp. 1-6). IEEE.
[28] Sheng Z, Peihua L, Yuli L, Mingsi Q, Changgang J, Meng Z. Image registration method based on optimized SURF algorithm. American Journal of Optics and Photonics. 2019; 7(4):63-9.
[29] Liao F, Chen Y, Chen Y, Lu Y. SAR image registration based on optimized Ransac algorithm with mixed feature extraction. In IGARSS international geoscience and remote sensing symposium 2020 (pp. 1153-6). IEEE.
[30] Li X, He H, Huang C, Shi Y. PCB Image registration based on improved SURF algorithm. In international conference on image processing, computer vision and machine learning 2022 (pp. 76-9). IEEE.
[31] Wu Y, Wang Z. Remote sensing image registration algorithm based on improved surf in wavelet domain. Journal of Tianjin University Science and Technology. 2017; 50(10): 1084-92.
[32] Zhongjun W, Yanfeng C. Image registration algorithm using SURF feature and local cross-correlation information. Infrared and Laser Engineering. 2022; 51(6):1-6.

[33] Zhao X, Li H, Wang P, Jing L. An image registration method using deep residual network features for multisource high-resolution remote sensing images. Remote Sensing. 2021; 13(17):1-25.
[34] Yang Z, Dan T, Yang Y. Multi-temporal remote sensing image registration using deep convolutional features. IEEE Access. 2018; 6:38544-55.
[35] Ma W, Zhang J, Wu Y, Jiao L, Zhu H, Zhao W. A novel two-step registration method for remote sensing images based on deep and local features. IEEE Transactions on Geoscience and Remote Sensing. 2019; 57(7):4834-43.
[36] Li C, You Y, Cao J, Zhou W. MoatNet: registration for multi-temporal optical remote sensing images using deep convolutional features. In international geoscience and remote sensing symposium IGARSS 2021 (pp. 2154-7). IEEE.
[37] Ying C, Guoqing L, Hengshi C. Multi-temporal remote sensing image registration based on multi-layer feature fusion of deep residual network. In international conference on intelligent informatics and biomedical sciences 2019 (pp. 363-7). IEEE.
[38] Zeng L, Du Y, Lin H, Wang J, Yin J, Yang J. A novel region-based image registration method for multisource remote sensing images via CNN. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2020; 14:1821-31.
[39] Zou B, Li H, Zhang L. Self-supervised SAR image registration with SAR-superpoint and transformation aggregation. IEEE Transactions on Geoscience and Remote Sensing. 2022; 61:1-15.
[40] Ibañez D, Fernandez-beltran R, Pla F. FloU-Net: an optical flow network for multimodal self-supervised image registration. IEEE Geoscience and Remote Sensing Letters. 2023; 20:1-5.
[41] Li L, Han L, Ding M, Liu Z, Cao H. Remote sensing image registration based on deep learning regression model. IEEE Geoscience and Remote Sensing Letters. 2020; 19:1-5.
[42] https://docs.opencv.org/4.x/dc/dc3/tutorial_py_matcher.html. Accessed 24 August 2022.
[43] Zuliani M. Computational methods for automatic image registration. University of California, Santa Barbara; 2007:1-24.
[44] http://bhuvan.nrsc.gov.in. Accessed 20 August 2022.
[45] Gong M, Zhao S, Jiao L, Tian D, Wang S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. IEEE Transactions on Geoscience and Remote Sensing. 2013; 52(7):4328-38.
[46] http://www.landinfo.com/GalSatResComp.htm, Accessed 20 August 2022.

Laukikkumar K. Patel is a Ph.D. scholar in Sankalchand Patel University Visnagar, India. He is also working as a lecturer in the Electronics and communication department at Government Polytechnic Palanpur, India. He obtained his BE degree in Electronics and Communication Engineering in 2009 from Sardar Patel University, Vallabh Vidyanagar and M.Tech degree in Electronics and Communication (CSE) in 2012 from Charotar University of Science and Technology. His research interests include Image processing, Machine Learning and Deep Learning.
Email: laukiksky@gmail.com

Dr. Manish I. Patel is working as an Assistant Professor in Electronics and Communication Engineering Department at Institute of Technology, Nirma University, Ahmedabad. He has more than 18 years of teaching experience. He obtained his M.tech degree in Electronics and Communication (VLSI Design) in 2010 from Nirma University, Ahmedabad. Dr Manish obtained his PhD in 2017 from Gujarat Technological University, Ahmedabad. His areas of interest include signal and image processing, machine learning and VLSI design. He has published more than fifteen papers in journals and conference proceedings. He has guided PG students. He is a life member of ISTE and IETE, India.
Email: manish.i.patel@nirmauni.ac.in

Appendix I

S. No.	Abbreviation	Description
1	CNN	Convolutional Neural Network
2	CMR	Correct Match Rate
3	DNN	Deep Neural Network
4	FAST	Features from Accelerated Segment Test
5	FREAK	Fast Retina Keypoint
6	HOG	Histogram of Oriented Gradient
7	HSV	Hue, Saturation, Value
8	LiDAR	Light Detection and Ranging
9	LISS-III	Linear Imaging and Self Scanning sensor-III
10	NTSC	National Television System Committee
11	ORB	Oriented FAST and Rotated BRIEF
12	PCA	Principal Components Analysis
13	PCB	Printed Circuit Board
14	PROSAC	Progressive Sampling Consensus
15	PSO	Position Scale Orientation
16	RANSAC	Random Sample Consensus
17	ResNet	Residual Network
18	SAR	Synthetic Aperture Radar
19	SIFT	Scale Invariant Feature Transform
20	SURF	Speeded Up Robust Feature
21	SUSAN	Smallest Univalue Segment Assimilating Nucleus
22	UAV	Unmanned Aerial Vehicle
23	VGG	Visual Geometry Group