SYSTEMATIC ANALYSIS ON TRADITIONAL METHODS OF IMAGE MATCHING

Anuradha Uggi
ee21resch01008@iith.ac.in

Parvez Alam
ai21resch01005@iith.ac.in

Pendota Amulya
ee21mtech12003@iith.ac.in

Rangapuri Deeksha
ee20mtech01001@iith.ac.in

ABSTRACT

Over the past decades various approaches have been proposed for image matching, starting from traditional methods to deep learning models. Deep learning has paved the way for automatic feature learning resulting in superior performance over traditional handcrafted features based methods. Viewpoint, scale, illumination variations, occlusion, blur etc have made the image matching task more challenging. In this analysis paper, we attempt to investigate and understand the robustness of a variety of existing image matching techniques in the literature and analyse the suitability of those techniques for various real time applications.
The related work can be viewed at https://github.com/Anuradha-Uggi/IVP-PROJECT.

Keywords: SURF,SIFT,ORB,BRISK.

1.INTRODUCTION

Image processing can be defined as a process that transforms a degraded image to another image of better quality in order to facilitate its posterior interpretation. Image matching is a part of image processing in which we seek the similarity between two images based on analysing the content,features, structure, relationship, texture and grayscale value. Image matching has applications in the areas of Image Retrieval, Object detection and tracking, Simultaneous localization and Mapping (SLAM), Image Fusion, Image registration, stitching, Structure from the motion, and Visual Homing. So, the performance of image matching algorithms directly affects the accuracy of detection and tracking. Researchers have proposed many image matching algorithms which can be classified mainly into two parts: Traditional feature based and the Deep learning based methods.

1. PROBLEM STATEMENT

In this work we proposed to perform a systematic analysis of different techniques in the literature and try to demonstrate the performances of those techniques in terms of the redundancy that every method offers while matching a disturbed image against the corresponding reference image.

3. LITERATURE REVIEW

Efficient feature learning forms a strong foundation for Image matching. Proposed approaches in the literature can be classified into Traditional and Deep Learning methods. Traditional methods are non-parameterized methods which rely upon the geometry of images.

David Lowe[1] has proposed SIFT which detects key points and corresponding descriptors at different scales using the DoG (Difference of Gaussian) function to perform matching operations between images. SIFT features are observed to be invariant to scale, rotation, illumination variations. The pipeline of SIFT includes :

Scale-space peak selection: The first stage of computation searches over all scales and image locations.

Key point localization: At each candidate location, a detailed model is fit to determine location and scale.

Orientation Assignment: One or more orientations are assigned to each keypoint.

Keypoint descriptor: Robust neighbourhood is extracted around key point. Though SIFT gives us dense features but is highly expensive in computations.

On the other hand Herbert Bay et.al[2] has proposed SURF which uses Integrated images and block filters to generate scale space and the concept of integrated images has been observed to be less expensive compared to SIFT. The inherent difficulty in extracting suitable features from an image lies in balancing two competing goals: high quality description and low computational requirements.

As a novel method for generating key points from an image BRISK has been proposed by Leutenegger et.al[7] in which key points are detected in octave layers of the image pyramid and as well as in layers in-between in order to boost the efficiency of

computation. The proposed BRISK method is demonstrated to achieve comparable quality of matching at much less computation time.

The standard traditional method ORB proposed by Tyagi et.al[8] offers strong invariance to rotation. ORB is basically a fusion of FAST keypoint detector and BRIEF descriptor with many modifications to enhance the performance. First it uses FAST to find keypoints, then applies Harris corner measure to find top N points among them. It also uses pyramid to produce multiscale-features. ORB has a binary descriptor that is similar to BRIEF, with the added advantages of rotation invariance and learned sampling pairs.

Traditional methods are non-trainable which effectively reduces the human efforts in annotating data and best suits for real time applications where time is a major concern.

On the other hand deep learning based approaches are parametric models which learn task-specific features. Again these are classified into strongly supervised, weakly supervised and self-supervised.

Daniel DeTone et.al[3] has proposed SuperPoint, a self-supervised interest point detector and descriptor, trained on large synthetic data consisting of simple geometric shapes like quadrilaterals, triangles, lines etc. It has a fully convolutional neural network that computes SIFT-like 2D interest point locations and its associated descriptors of fixed length in a single forward pass.

Homographic Adaptation technique is used to transform a key point into SuperPoint which is robust to viewpoint and illumination variations. The use of the MagicPoint detector along with the homographic adaptation increases the performance of the detector.

Superpoint produces a larger number of correct matches and outperforms classical detectors like LIFT, SIFT and ORB and performs on par with classical detectors under viewpoint changes.It is found to fail in case of extreme in-plane rotation between images as the training dataset do not consist of such training examples.

Superpoint combined with SuperGlue has resulted in improved performance. Sarlin et.al[4] has proposed SuperGlue built on GNNs (Graph Neural Networks) to extract dense features jointly between a pair of images. It sits in the middle of SLAM architecture to refine superpoint key points by updating nodes in GNNs. Every node in GNN aggregates self and cross attention using powerful transformers.

Rocco et.al[5] has proposed a Neighbourhood consensus network (NCNet) which is composed of global and local filtering stages to effectively downweight non matching feature points. Without need for costly manual annotations of point-point correspondences, a weakly supervised loss function is used to update weights of the network. NCNet takes the neighbourhood around interest points into account to eliminate non matching points. All above defined models are highly effective at learning features but removes a large number of interest points when repetitive, textureless patches are present in the images.

In this work we deeply review a few of the above matching techniques and demonstrate through quantitative performances.

4. RESULTS

As a part of this analysis different image matching techniques SIFT, SURF, ORB and BRISK are applied on RGB and Thermal images Different test images are generated which includes Rotated, Scaled, Affine Transformed, Translated, AWGN and salt-pepper Noisy images. Matching techniques are applied on these test images in order to observe the robustness of each method to each disturbance.

I):RGB IMAGES

Table-1 describes the experimental results obtained on test images from the RGB data set. For every test image key points and descriptors are detected using methods described. While matching a test image with the reference pristine image outliers or unmatched key points are eliminated through Brute Force Matching, where the matching metric(%M) is computed as:

%M=100(#Keypoints-#Outliers/ #keypoints) --*eq (1)*

Where this metric describes the redundant key points detected in the image which are generally not useful to match the image with corresponding pristine image. Table 1 results show how the performance of different methods vary over different kinds of disturbed images.

II): THERMAL IMAGES

The same set of experiments are repeated to find out how the performance of different methods vary when a thermal image is given as input. Traditional methods in general fail at detecting key points on edges, textureless images. Thermal images in general come with less texture in it. By comparing Table 1 results with Table 2 results it can be seen that the number of key points detected are very less in Thermal images compared to the number of keypoints detected on RGB images.

RGB Images

Test Images	SIFT	SURF	ORB	BRISK
Ref Image (Rf_I)	#k_p: 1700	#k_p: 754	#k_p: 474	#k_p: 1225
Rotated Image (R_I)	#k_p: 1473 % Match: 77.93	#k_p: 704 % Match: 62.35	#k_p: 474 % Match: 59.70	#k_p: 2170 % Match: 60.87
Awgn noisy Image (N_I)	#k_p: 1004 % Match: 74.90	#k_p: 790 % Match: 80.88	#k_p: 475 % Match: 50.10	#k_p: 2540 % Match: 99.37
scaled Image (S_I)	#k_p: 1394 % Match: 59.25	#k_p: 603 % Match: 56.55	#k_p: 468 % Match: 42.09	#k_p: 1227 % Match: 52.73
Affine transformed Image (Af_I)	#k_p: 974 % Match: 70.32	#k_p: 595 % Match: 50.58	#k_p: 477 % Match: 36.05	#k_p: 1650 % Match: 40.48
Translated Image (T_I)	#k_p: 1059 % Match: 94.33	#k_p: 516 % Match: 79.26	#k_p: 476 % Match: 48.31	#k_p: 1968 % Match: 62.24

Table-1: Comparison Table

Thermal Images

Test Images	SIFT	SURF	ORB	BRISK
Ref Image (Rf_I)	#k_p: 614	#k_p: 592	#k_p: 393	#k_p: 381
Rotated Image (R_I)	#k_p: 491 % Match: 80.24	#k_p: 477 % Match: 63.73	#k_p: 416 % Match: 59.70	#k_p: 238 % Match: 60.87
Awgn noisy Image (N_I)	#k_p: 552 % Match: 56.15	#k_p: 1504 % Match: 23.27	#k_p: 445 % Match: 50.10	#k_p: 390 % Match: 99.37
scaled Image (S_I)	#k_p: 116 % Match: 93.97	#k_p: 390 % Match: 58.71	#k_p: 21 % Match: 76.09	#k_p: 5 % Match: 100
Affine transformed Image (Af_I)	#k_p: 526 % Match: 59.14	#k_p: 605 % Match: 48.26	#k_p: 424 % Match: 34.19	#k_p: 478 % Match: 28.23
Translated Image (T_I)	#k_p: 431 % Match: 97.22	#k_p: 412 % Match: 87.37	#k_p: 454 % Match: 38.98	#k_p: 407 % Match: 62.24

Table-2: Comparison Table

Results for RGB Images

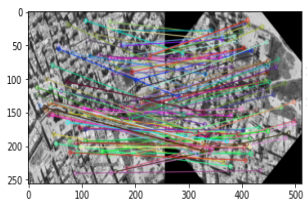
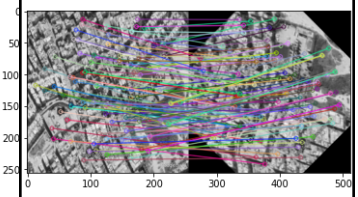
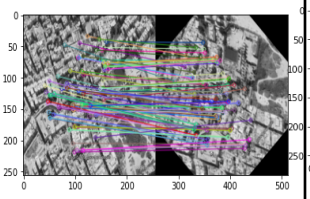
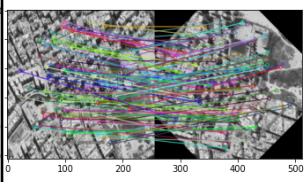
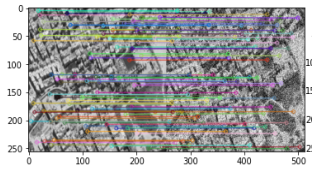
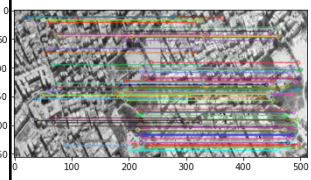
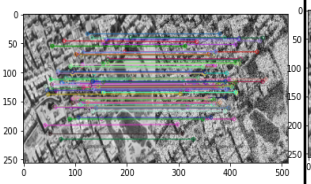
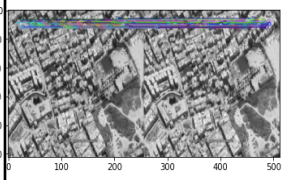
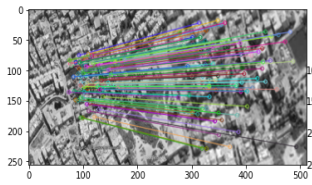
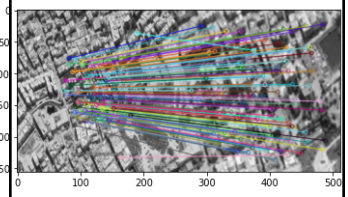
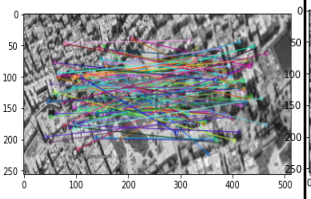
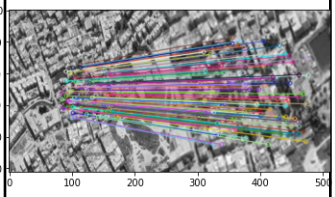
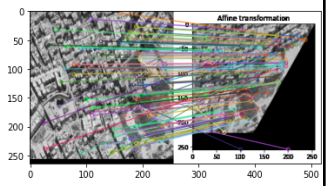
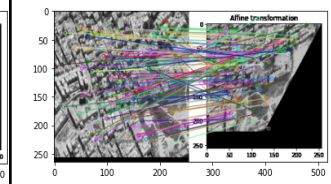
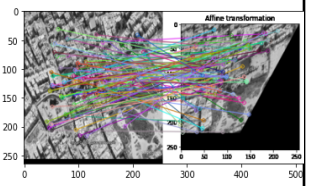
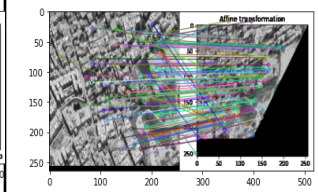
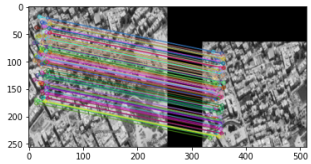
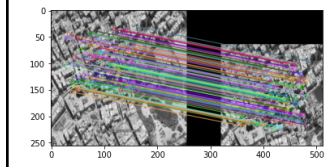
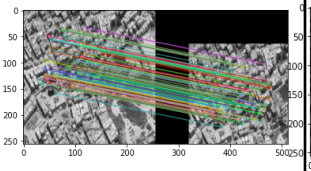
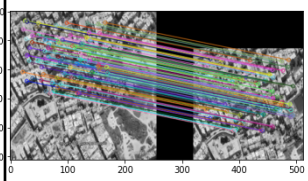
Test Images	SIFT	SURF	ORB	BRISK
Rotated Image (R_I)				
Awgn noisy Image (N_I)				
scaled Image (S_I)				
Affine transformed Image (Af_I)				
Translated Image (T_I)				
Column1	Column2	column3	Column4	Column5

Fig1: Image Matching on RGB data: Column1 includes all disturbed test images generated, Column2: SIFT key points mapping, Column3: SURF key points mapping, Column4: ORB key points mapping, Column5: BRISK key points mapping.

Results for Thermal Images

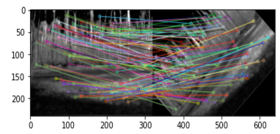
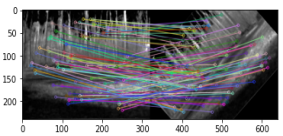
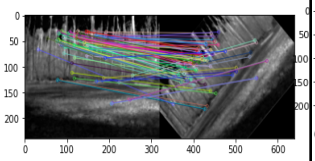
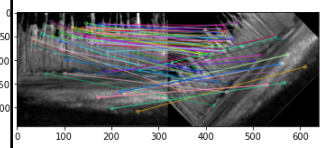
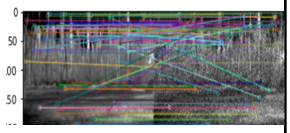
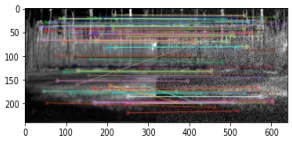
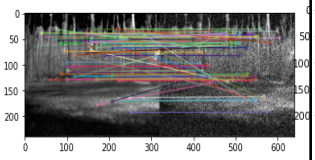
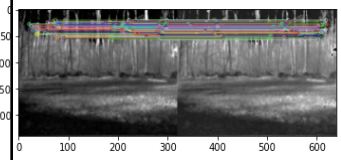
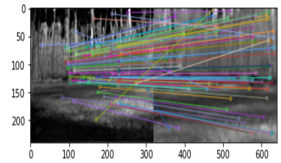
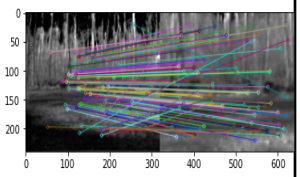
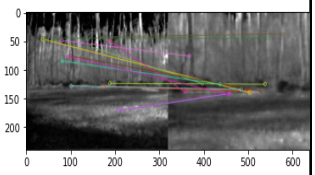
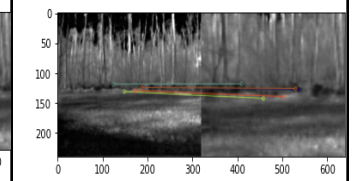
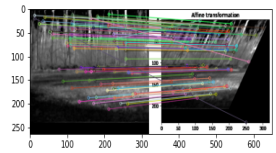
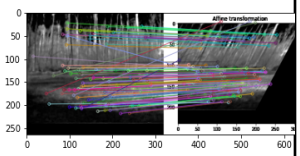
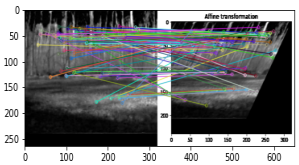
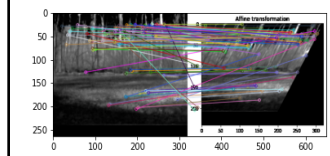
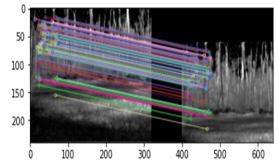
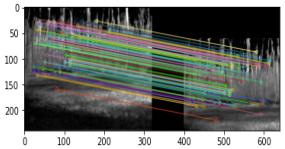
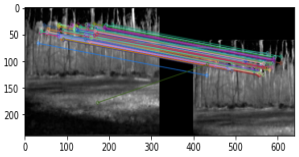
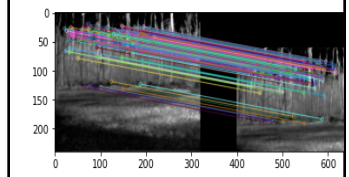
Test Images	SIFT	SURF	ORB	BRISK
Rotated Image (R_I)				
Awgn noisy Image (N_I)				
scaled Image (S_I)				
Affine transformed Image (Af_I)				
Translated Image (T_I)				
Column1	Column2	Column3	Column4	Column5

Fig2: Image Matching on RGB data:Column1 includes all disturbed test images generated, Column2: SIFT key points mapping, Column3: SURF key points mapping, Column4: ORB key points mapping, Column5: BRISK key points mapping.

5. CONCLUSION

This analysis describes the performance variation of different methods over different images especially on different modes of images RGB and thermal. Here, as we see in real time applications due to natural effects, test images might be very different and sometimes distorted. In such cases the method should still be able to match those images. This paper demonstrates sensitivity of Traditional methods to 2d-motions and noisy images.

6. REFERENCES

[1] David G Lowe, “Distinctive ImageFeatures from Scale-Invariant Keypoints”,International Journal of Computer Vision,vol.50, No. 2, 2004, pp.91-110.

[2] SURF: Speeded Up Robust Features Herbert Bay1, Tinne Tuytelaars2, and Luc Van Gool12 .

[3] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich, “Superpoint: Self-supervised interest point detection and description” .

[4] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich “SuperGlue: Learning Feature Matching With Graph Neural Networks”.

[5] Dubey, Shiv Ram. "A decade survey of content based image retrieval using deep learning." IEEE Transactions on Circuits and Systems for Video Technology (2021).

[6] Ma, Jiayi, et al. "Image matching from handcrafted to deep features: A survey." International Journal of Computer Vision 129.1 (2021): 23-79.

[7]Leutenegger, Stefan, Margarita Chli, and Roland Y. Siegwart. "BRISK: Binary robust invariant scalable keypoints." *2011 International conference on computer vision*. Ieee, 2011.

[8]Tyagi, Deepanshu. "Introduction to ORB (Oriented FAST and Rotated BRIEF)." (2019).