Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 115 (2017) 431–439

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Advances in Computing & Communications, ICACC-2017,
22-24 August 2017, Cochin, India

Bag-of-Spatial Words(BoSW)Framework for Predicting SAR Image Registration in Real Time Applications

B Sirisha*[a], B Sandhya[b], Chandra Sekhar Paidimarry[c], A S Chandrasekhara Sastry[d]

[a] *Department of Electronics and Communication Engineering,K L University,Vaddeswaram,Guntur,A.P,India.*
[b]*Department of ComputerScience and Engineering,MVSR Engineering college,Hyderabad,India.*
[c]*Department of Electronics and Communication Engineering,UCE,Osmania University,Hyderabad, India.*
[d]*Department of Electronics and Communication Engineering,K L University,Vaddeswaram,Guntur,A.P,India.*

Abstract

SAR Image registration is a precursor for several remote sensing applications, which need precise spatial transformation between the real time moving image and fixed off-line image. In such applications, the processing time in finding whether the moving images can be registered with fixed image constitute an overhead. Hence we have approached the problem by trying to predict if the given SAR images can be registered or not even without registering them. The proposed image registration approach incorporates a classifier into the standard pipeline of feature based image registration. The attributes for the classifier model are derived from fusing the spatial parameters of the feature detector to the descriptor vector in bag of visual words framework.

Keywords: Image Registration; Feature Detection; Feature Description; Bag of visual words;Classification.

1. Introduction

Identifying spatial correspondence between two or more images of the same scene is a classical research problem in remote sensing and photogrammetry. Applications such as change detection [1] , image fusion [2],image mosaicing [3], target tracking [4] which directly rely on SAR image registration. There has been wide-ranging research in image registration; the review papers by Brown L.G [5] and Zitova ,Flusser [6] illustrate literature in detail. Zitova categorized image registration as area based image registration(AIR) and feature based image registration(FIR). AIR algorithm operates directly on entire image using gray level pixel intensities.Widely used AIR methods include cross-

* Corresponding author. Tel.: +0-901-027-7400.
E-mail address: sirishavamsi@gmail.com

10.1016/j.procs.2017.09.102

cumulative (CC) residual entropy [7],mutual information [8], normalized cross-correlation (NCC) coefficient .Though the approach is robust, it is limited by the amount of invariance towards deformations[8].In contrast, FIR method instead of using entire image structure uses distinct local structural features like points, lines and regions.These features are less sensitive to reflectance,geometric, photometric inconsistency [9] .This approach has gained momentum in recent years due to advent of many feature extraction algorithms providing invariance to a large range of photometric and geometric deformations[10].

Applications like target detection in remote sensing, require registration of images from sequence of images, where one image is captured offline and saved (I_f-fixed image), and other image is captured real time (I_m-moving image). In such scenarios, instead of computing an inaccurate transformation between images it is important to be able to estimate how well images are registered. Hence we have approached the problem by trying to predict if the two given SAR images can be registered using a specific feature detector and descriptor without registering them. Automatic prediction of the given input SAR image pair can be registered or not, even without registering them is a recent development and the literature is limited in remote sensing. This has been achieved by describing an image using feature detector and descriptor values in the framework of machine learning. We have proposed an approach for SAR image registration where the images are registered only if they have been predicted accurately, by feature level fusion of detector and descriptor parameters. In the first stage features are extracted and described. However before proceeding to the correspondence and transformation estimation, we feed the features to a prediction model which predicts if the images can be registered without actually registering it. Our contribution is twofold:

1 Incorporate knowledge into the standard feature based image registration pipeline by building a prediction model which can be used for real time registration of Terra SAR X band images.
2 Propose bag of spatial words(BoSW) image representation framework ,which fuse feature detectors spatial,scale parameters in addition to descriptor values. BoSW framework is further used in building the prediction model.

The paper is organized as follows: Methodology of the proposed image registration approach is illustrated in Section II, Section III reports the experimental results, discussion and in Section IV the conclusion of the paper is presented.

2. Methodology of the Proposed Image Registration Approach

In this section, a robust and effective approach for accurate SAR image registration is presented. The proposed approach can be divided into four main stages, as illustrated in Figure 1. When considering pair of SAR images to be registered, we use the term fixed image I_f and moving image I_m, where the fixed image is the image defining the coordinate system, while the moving image is the one to be transformed to this coordinate system. The approach is intended for registering sequences of images acquired from identical or similar earth observation sensors. Let two images, a fixed image I_f and moving image I_m where $i \in 1, 2, ...n$ are a sequence of input images, to be registered. The main stages of the proposed approach are briefly shown in Figure 1 are described as follows:

2.1. Stage 1- Local Invariant Feature Extraction

In the first stage, a local affine invariant feature detector and descriptor is used to extract features in both fixed image I_f and moving image I_m. A gray scale image is taken as input to the feature detection algorithm, Hessian Affine[11] in this case. It detects affine invariant feature points across the image. Each feature point is characterized by a seven element vector consisting of X-location, Y-location, scale-λ and 4 values of affine transformation matrix $a_{11}, a_{12}, a_{21}, a_{22}$. If N_f is the number of feature points of fixed image I_f, the output of feature detection on fixed image is $N_f * 7$. If N_{mi} is the number of feature points of moving image I_{mi}, the output of feature detection on moving image is $N_{mi} * 7$ The image and its corresponding feature points are fed to a feature descriptor ,SIFT[12] in our case which describes each of the features with a vector values. Hence fixed image I_f is described by $N_f * 128$ values, say set D_f and moving image I_{mi} is described by $N_{mi} * 128$ values, set D_{mi}.

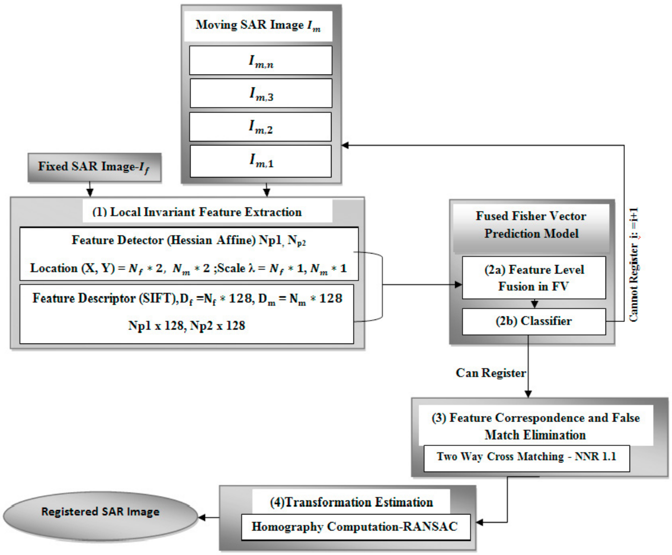

Moving SAR Image I_m
$I_{m,n}$
$I_{m,3}$
$I_{m,2}$
$I_{m,1}$
Fixed SAR Image-I_f
(1) Local Invariant Feature Extraction
Feature Detector (Hessian Affine) Np1, N_{p2}
Location (X, Y) = $N_f * 2$, $N_m * 2$;Scale λ = $N_f * 1$, $N_m * 1$
Feature Descriptor (SIFT),$D_f = N_f * 128$, $D_m = N_m * 128$
Np1 x 128, Np2 x 128
Fused Fisher Vector Prediction Model
(2a) Feature Level Fusion in FV
(2b) Classifier
Cannot Register i: =i+1
Can Register
(3) Feature Correspondence and False Match Elimination
Two Way Cross Matching - NNR 1.1
(4)Transformation Estimation
Homography Computation-RANSAC
Registered SAR Image

Fig. 1. Methodology and stages of the proposed SAR image registration approach

2.2. *Stage 2-Bag-of-Spatial Words(BoSW) Prediction Framework*

The two main phases of stage 2 are feature level fusion in bag of visual words (BoSW) and classification. We have integrated a classifier into the standard pipeline of image registration.

2.2.1. *Bag-of-Spatial Words(BoSW) Image Representation*

The attributes to the classifier are generated using BoSW framework. Before exhibiting the BoSW model , a brief review of the standard bag of visual words is presented .Bag of Visual Words[13],[14] (BoW) features are extracted in following steps i) automatically detect feature points, line, regions of interest from images, ii)local descriptors are computed over the detected feature points, iii) quantize the local descriptors into visual words to form visual vocabulary, iv) find the instances/occurrences in the input image pair of each visual word in the visual vocabulary for building bag of visual words feature vector or histogram of visual word frequencies.The setback of such representation is that histogram of descriptors of an image does not carry any spatial information and it has no order. It only stores the frequency of visual word in each image which is very effective in symbolizing the kind of content in the image. However, when two images are being registered, in addition to photometric characteristics, spatial deformations of the images also affect the outcome[15]. Hence we propose to fuse the feature detector parameters like XY locations, scale λ in the bag of visual words model. Figure 2 shows the proposed approach to compute attributes for the BoSW prediction model.Features are detected using Hessian Affine feature detector algorithm, and a patch around each detected point is represented using SIFT descriptor, as a vector for both fixed (source) and moving (target) images. In the standard BoW representation , image descriptors of both the images are together clustered into k clusters,called as visual words. Each image is then represented as a histogram of k bins representing distribution of descriptors in each cluster. In the proposed model location information(XY)ie:spatial coordinates ,scale λ information and descriptors of both images is clustered into k clusters and a k bin histogram is built for each image based on the corresponding detectors and descriptors membership to the cluster.This representation is called Bag of Spatial words BoSW.

2.2.2. *Classifier*

To obtain significant analysis,proposed BoSW is compared with standard BoW,location -XY-BoW,scale-BoW.

Following are given as input to classifier:

Proposed-BoSW:X,Y locations,scale parameters and descriptors belonging to feature points of both images are clustered into k clusters and for each image a histogram of cluster membership of the corresponding descriptors,detectors is generated.Let Wp_f represent fused vector for image I_f and Wp_m for image I_m.

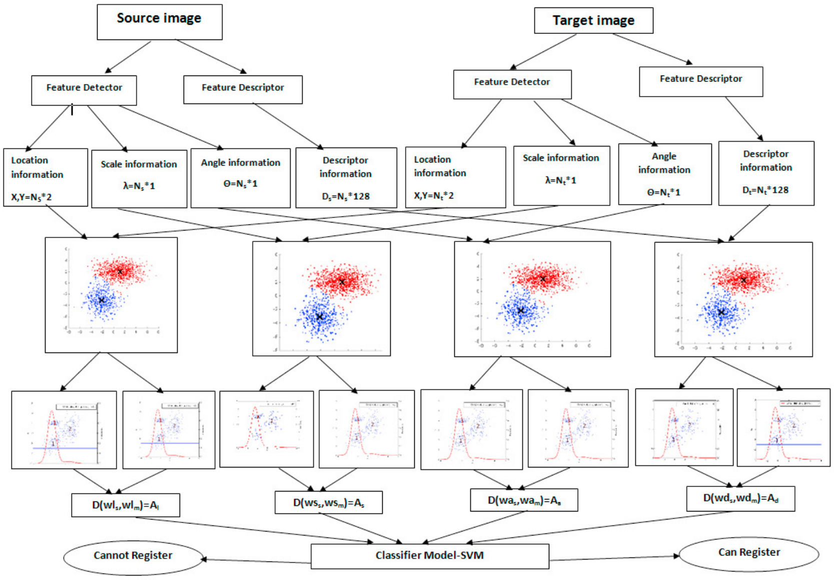

Source image
Target image
Feature Detector
Feature Descriptor
Feature Detector
Feature Descriptor
Location information X,Y=N$_s$*2
Scale information λ=N$_s$*1
Angle information Θ=N$_s$*1
Descriptor information D$_s$=N$_s$*128
Location information X,Y=N$_t$*2
Scale information λ=N$_t$*1
Angle information Θ=N$_t$*1
Descriptor information D$_t$=N$_t$*128
D(wl$_s$,wl$_m$)=A$_l$
D(ws$_s$,ws$_m$)=A$_s$
D(wa$_s$,wa$_m$)=A$_a$
D(wd$_s$,wd$_m$)=A$_d$
Cannot Register
Classifier Model-SVM
Can Register

Fig. 2. Proposed Feature Level Fusion of BoSW Prediction Model for SAR Image Registration

Standard BoW:The vectors in sets D_f and D_m are clustered into k clusters and for each image a histogram of cluster membership of the corresponding descriptors is generated. Let D_f generate a vector WD_f and D_m generate WD_m.
Location XY-BoW: X,Y locations belonging to feature points of both images i.e. N_{p1} and N_{p2} are together clustered into k clusters and a k bin histogram is built for each image based on the corresponding detectors membership to the cluster. Let WL_f represent location vector for image I_f and WL_m for image I_m.
Scale-BoW :The scale parameter belonging to feature points of both the images are clustered and k bin histograms built. Let WS_f and WS_m be the scale vectors for images I_f and I_m respectively.
The distances between Bag of visual words vectors are computed and taken as attributes for the prediction model. The list of similarity measures considered in experimental study are Euclidean, Manhattan,Chi-Square,Bhattacharya distance measures.

Let A_d , A_l ,A_s and A_p represent distances between $\{WD_f, WD_m\}$, $\{WL_f, WL_m\}$, $\{WS_f, WS_m\}$, $\{WP_f$,$WP_m\}$ respectively. The four element vector $\{A_d, A_l, A_s, A_p\}$ is given as the input to a classifier which has been trained and validated. If the classifier predicts that images cannot be registered, subsequent image is taken for registering with fixed image or else it proceeds to the following stage.

2.3. Stage3,4-Feature Correspondence and Transformation Estimation:

Feature descriptors of sets D_f and D_m are matched using a threshold on nearest neighbour ratio of distances. To improve the accuracy of matches, matching is performed twice between D_f and D_m i.e. finding a neighbour of each point in D_f among the points in D_m and vice-versa. Neighbours found in two way matching are only considered as matched.X, Y Locations of matched descriptors are input to RANSAC which estimates the transformation matrix, H in addition to finding inliers among the matches. The moving image I_{mi} can be transformed using H to the coordinate system of fixed image I_f .

3. Experiment Results

Evaluation Dataset: Four Terra SAR X band images, of dimension 10556*9216 of the same scene but captured at different look angles are used for the evaluation.Images of size 850 x 1000 have been cropped from these four images to generate 11 fixed images. An area similar to the fixed image is cropped from the other look angle SAR images,to which transformations are applied to generate moving image.One look angle image is considered as fixed image and another look angle image transformed by applying a known geometric transformation as the moving image.We have created 11 datasets in which the amount of common area (overlap) between the fixed and moving images is varied. In each dataset there are 54 SAR images, hence a total of 594 image pairs is generated as the pairs differ in varying degrees of scale, rotation, noise and overlap. The induced deformations are rotation-35 images (vary by 10 degree each),speckle-10 images (vary by 0.04, 0.08, 0.12, 0.16,0.2, 0.24, 0.28,0.32,0.36,0.4 variance),Look-angle-3 images (vary by 2 degree each) and scale- 6 images (scale factor of 0.5,2,2.5,3,3.5,4).In this section, we report experimental results of proposed BoSW prediction framework using, 594 SAR images that vary in look angle,scale,rotation and speckle noise.Performance and robustness of the framework is assessed by three evaluations:

1 Cluster Count and Distance Measure Analysis.
2 Evaluation of proposed BoSW image representation and prediction framework by
    i Qualitative evaluation of classifier performance is done with six attributes.
    ii Establishing a relation between feature detector distance and deformation between the SAR images.
    iii Establishing a relation between feature detector distance and registration error using regression model.
3 Time Analysis of proposed BoSW and standard BoW.

3.1. *Cluster Count and Distance Measure Analysis*

Figure 3 shows the classification accuracies for four distance measures with cluster count size of 5,75, 150,300 and 500 on 594 SAR image dataset.It is observed that accuracies become stable when the size of the cluster is greater than 200. The accuracy of the classifier reduces if the cluster count k is less than 50, which confirms that a compact clusters has an inadequate discriminative capability. The optimal value of the cluster count can be approximately 100-300. It is also observed from figure 3 that the classification accuracies vary for varied distance measures, indicating that the distance measures have impact on the performance of the BoSW representation. The Bhattacharya distance outperforms other distance measure for all cluster count values.For our experiments we have fixed the value of K=150 and used Bhattacharya distance measure.

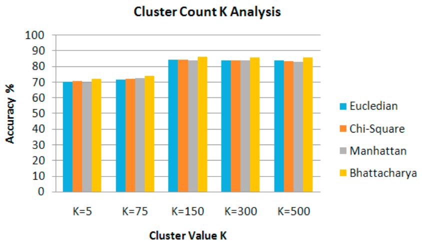


Fig. 3. Classification Accuracies for four distance measures with Cluster Count K= 5,75, 150,300 and 500.

3.2. *Classifier Performance*

The effectiveness of the BoSW image representation is tested and analyzed with three classification algorithms,viz., Naive Bayes, SVM and J48 using 540 training images with tenfold cross validation .The performance of the BoSW model is evaluated and compared with six attributes in terms of accuracy, precision, recall, F measure and ROC area on 540 SAR images across deformations. The six attributes are 1) standard BoW, 2) location XY BoW, 3) Scale Bow, 4) fusing Desc, XY attributes (concatenation of Location XY, Descriptor vector) 5)fused Desc,Scale, 6) fused Desc,location ,scale vector . Table 1 shows the performance when a classifier trained on 10 datasets (D1 to D10-540 SAR images) is tested on dataset 11(D11-54 SAR images). It can be observed that proposed BoSW accuracy is high compared to other attributes.It is evident from 1 that, accuracy of the prediction framework increases only when the spatial information extracted from feature detector is considered along with descriptor information.It is noted that fature detector location parameter XY plays a significant role in improving the classification accuracy irrespective to the type of classifier.

Table 1. Prediction Model Results On 540 Training Images (D1 TO D10) and 54 (D11) Test Images Using 10 Fold Cross Validation.

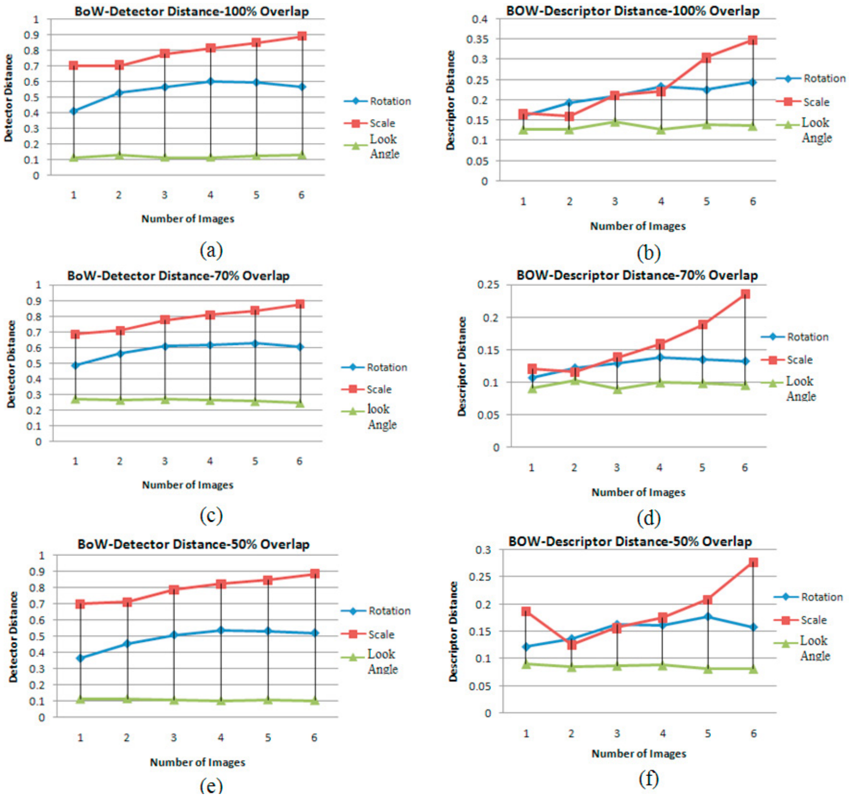
	Attributes	Accuracy	TP Rate	FP Rate	Precision	Recall	Fmeasure	ROC Area
J48	Standard-BoW(Desc)	72.48	0.725	0.326	0.755	0.725	0.707	0.748
	Location XY-BoW	86.52	0.865	0.145	0.866	0.865	0.865	0.9
	Scale-BoW	62.04	0.62	0.418	0.617	0.62	0.608	0.61
	Fused Desc,XY -BOW	86.14	0.861	0.152	0.863	0.861	0.861	0.88
	Fused Desc,Scale-BoW	76.85	0.769	0.262	0.777	0.769	0.763	0.777
	Fused Desc,XY,ScaleBoW	**86.52**	**0.865**	**0.151**	**0.868**	**0.865**	**0.864**	**0.89**
	Proposed BoSW	**89.29**	**0.88**	**0.132**	**0.88**	**0.891**	**0.89**	**0.91**
SVM	Standard-BoW(Desc)	72.29	0.723	0.329	0.753	0.723	0.704	0.697
	Location XY-BoW	86.52	0.865	0.145	0.866	0.865	0.865	0.86
	Scale-BoW	59.96	0.6	0.447	0.595	0.6	0.581	0.576
	Fused Desc,XY -BOW	86.14	0.861	0.146	0.861	0.861	0.861	0.858
	Fused Desc,Scale-BoW	72.67	0.727	0.326	0.76	0.727	0.708	0.7
	Fused Desc,XY,ScaleBoW	**86.14**	**0.861**	**0.146**	**0.861**	**0.861**	**0.861**	**0.858**
	Proposed BoSW	**89.91**	**0.89**	**0.121**	**0.92**	**0.91**	**0.89**	**0.91**
Nave Bayes	Standard-BoW(Desc)	72.48	0.725	0.326	0.755	0.725	0.707	0.752
	Location XY-BoW	86.52	0.865	0.145	0.866	0.865	0.865	0.899
	Scale-BoW	62.04	0.62	0.418	0.617	0.62	0.608	0.609
	Fused Desc,XY -BOW	87.47	0.875	0.132	0.875	0.875	0.875	0.918
	Fused Desc,Scale-BoW	73.43	0.734	0.285	0.734	0.734	0.732	0.775
	Fused Desc,XY,ScaleBoW	**86.14**	**0.861**	**0.147**	**0.861**	**0.861**	**0.861**	**0.909**
	Proposed BoSW	**89.01**	**0.87**	**0.143**	**0.86**	**0.889**	**0.88**	**0.89**
	54 Test Images (D11)							
SVM	Standard-BoW(Desc)	60.66	0.567	0.387	0.626	0.667	0.536	0.652
	Fused Desc,XY,ScaleBoW	**83.33**	**0.833**	**0.169**	**0.835**	**0.833**	**0.84**	**0.832**
	Proposed BoSW	**89.9**	**0.89**	**0.129**	**0.89**	**0.894**	**0.893**	**0.901**

3.3. Establishing Relation between Feature Detector distance and Deformation

Six images with scale, rotation, and look angle deformation are used from datasets (100%,70% and 50% common area)to demonstrate robust characteristics of feature detector . Figure 4 shows a plot of Location XY-BoW detector distances and descriptor distances between 6 pairs of SAR images for deformation wise.It is observed from Figure 4 (a), (c) and (e) that the detector distance between a pair of SAR images in the case of scale deformity falls in the range (0.7-0.9) and rotation in the range (0.4-0.6) and look angle in the range (0.08-0.15).The range of the BoW detector distance for each deformation does not overlap with other deformation range which is not true in case of BoW descriptor distance shown in Figure 4(b), (d) and (f). It is also observed that BoW detector distance values are consistent, irrespective of the common area between the images unlike descriptor distance. This robust characteristics of feature detector along with descriptor plays a significant role to improve the accuracy of proposed BoSW prediction framework.

3.4. Establishing Relation between Feature Detector distance and Registration Error

If the classifier predicts that images can be registered, the features extracted are matched and transformed using computed transformation matrix H estimated by RANSAC.Registration error is the euclidean distance between matched points of source Image which are transformed with computed transformation matrix H and target image matched points. Experiments have been carried out to predict if feature detector parameters can help in estimating the



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 4. BoW Detector and Descriptor distances across deformations like look angle, scale and rotation in Datasets D1, D8 and D10

error of registration between the images.Regression is performed using two models linear regression model (LRM) and linear regression tree(LRT) with seven BoW detector,descriptor and scale attribute distances of 594 image pairs. Log of registration error is used as dependent variable. LRM create a linear model based on the input attributes and registration error.LRT returns regression tree based on the input attributes and registration error.The binary tree obtained where each branching node is split based on values of registration error.For both approaches regression is assessed using mean square error (MSE) between the actual and the predicted registration error values.Figure 5 shows the mean square error for tenfold cross validation. It can be observed that proposed MSE of BoSW is low compared to other attributes.It can be noted that MSE of LRT,LRM is less when detector parameters are fused with the descriptor vector. Hence the spatial and scale information extracted from Bow feature detector distance plays a vital role in predicting and estimating the error of registration between the images, when it is combined with feature descriptor distance.

3.5. *Time Analysis of proposed BoSW Prediction Framework*

This section reports processing time taken for each stage of proposed BoSW approach when executed on 2.19Ghz/3MB cache $IntelCore^{TM}$ i7,8GB RAMx64 bit . The analysis of BoSW Prediction framework is done by comparing the processing time taken for a SAR image when registered and not registered using standard FIR pipeline and proposed BoSW.The Figure 6 shows plot of time taken for each of the four stages of SAR image registration.It is observed from the figure that in BoSW framework when images are registered all the four stages are executed and when they are not only stage 1 and 2 are executed. However in standard feature based image registration pipeline, stages 1,3 and 4 are always executed,so in our scenario out of 594 SAR images 220 images are registered and 374 are not registered.It is clear that we save 22,440 time units using machine learning approach.

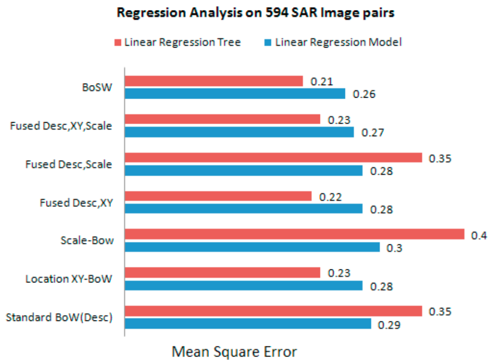


Fig. 5. Regression Analysis on 594 SAR Images for tenfold cross validation.

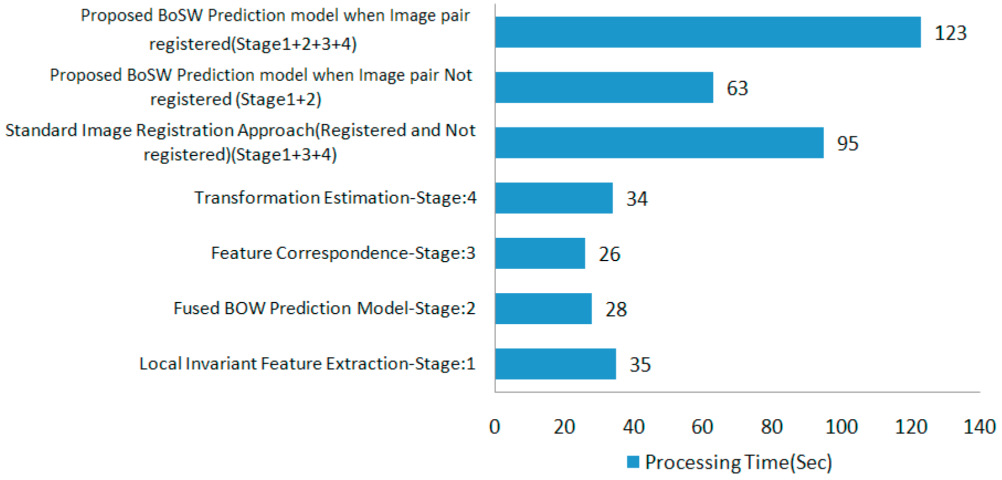


Fig. 6. Time analysis of Proposed BoSW Prediction framework against Standard Image registration framework for both registered or not registered cases.

3.6. Discussion

We have established that the standard BoW model falls short of its expected performance when used for image registration. The goal of registration is to find a spatial transformation between the images. For a specific kind of images,the possibility to register images correctly depends on the kind and amount of deformation between the images. It is known that the descriptor vectors do not have any spatial information encoded in them. Hence spatial information present in key point locations has been exploited by using it in BoW model. There exists a relation between the distances of BoW histograms of XY locations and the kind of deformation observed. This has motivated us to use them for estimating the registration parameters by two approaches : classification and regression. As discussed in the introduction, after feature extraction stage, accuracy of registration depends on the correspondence and the inlier estimation. Both of these are affected by the characteristic differences between the images. BoW vector built using descriptor and detector parameters closely represents these differences in both radiometric and spatial domains. Hence the difference between such vectors has been shown to be effective in the estimation of the registration error in regression. We have also demonstrated that incorporating machine learning approach in standard pipeline of SAR image registration is beneficial in time critical applications. This kind of approach saves the effort in estimating the homography matrix for images which cannot be registered.

4. Conclusion

Registering SAR images varying in look angle is a challenging task as small variation in look angle changes the photometric and geometric characteristics of SAR image to a large extent. The proposed approach can predetermine if the images can be accurately registered . The advantage of this approach is that, the parameters needed are computed from the features extracted as part of registration pipeline using BoW framework. It is demonstrated that the detector parameters when fused in the descriptor BoW framework improves the accuracy of prediction. The fused BoW image representation is extensively tested on a large dataset of SAR images to find the exact error of registration. When trying to register two images blindly, i.e. without knowing the kind of deformation or common overlap between them, the fusion of feature detector and descriptor parameters are used in the backdrop of BoW model that is effective in predicting the outcome of registration. The proposed registration approach has been demonstrated to be effective in terms of time as compared to the standard image registration approach.

References

[1] Yifang Ban,Peng Gong,Chandra Giri.(2015)"Global land cover mapping using Earth observation satellite data: Recent progresses and challenges."*Journal of Photogrammetry and Remote Sensing*103:1–6.

[2] Junyi Tao , Stefan Auer.(2016)"Simulation-Based Building Change Detection From Multiangle SAR Images and Digital Surface Models."*IEEE Journal of Applied Earth Observations and Remote Sensing* 9(8):3777–3791.

[3] J. E. Vera , S. F. Mora , J. A. Torres , J. Avendano.(2016)"Analysis of images SAR to flood prevention implementing fusion methods." in (STSIVA),booktitle *XXI Symposium on Signal Processing, Images and Artificial Vision*7743334:1-5

[4] Z. Liu, J. An, and Y. Jing.(2012)("A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration."*IEEE Trans. Geosci. Remote Sens.,* 50(2):514–527

[5] L. G. Brown.(1992)"A survey of image registration techniques."*ACM Computing Surveys (CSUR)* 24(4):325–376.

[6] Zitov, Barbara and Flusser, Jan.(2003)"Image registration methods: a survey."*Image Vision Comput.* ,21 (11):977–1000

[7] M. Hasan, M. R. Pickering, and X. P. Jia.(2012)"Robust automatic registration of multimodal satellite images using CCRE with partial volume interpolation."*IEEE Trans. Geosci. Remote Sens.,* 50(10).

[8] M. A. Siddique, M. S. Sarfraz, D. Bornemann, and O. Hellwich.(2012)"Automatic registration of SAR and optical images based on mutual information assisted Monte Carlo."*IEEE IGARSS, Munich, Germany* :1813–1816.

[9] M. G. Gong, S. M. Zhao, L. C. Jiao, D. Y. Tian, and S. Wang.(2014)"A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information."*EEE Trans. Geosci. Remote Sens.,* 52(7):4328–4338.

[10] B. Fan, C. L. Huo, C. H. Pan, and Q. Q. Kong.(2013)"Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT."*IEEE Geosci. Remote Sens. Lett.,* 10(4):657–661.

[11] K. Mikolajczyk ,C. Schmid.(2004)"Scale and affine invariant interest point detectors." *International Journal of Computer Vision* ,60(1):63–86.

[12] David G. Lowe.(2004)"Distinctive image features from scale-invariant keypoints."*International Journal of Computer Vision*60:91–110.

[13] A.Bosch,X.Munoz, and R.Marti.(2007)"Which is the best way to organize/ classify images by content?."*Image and Vision Computing* ,25(6):778–791.

[14] J. C. Niebles, H. Wang, and L. Fei-Fei.(2008)"Unsupervised learning of human action categories using spatial-temporal words."*International Journal Computer Vision* 79(3):299–318.

[15] Ivan Gonzalez Daz , Murat Birinci,Fernando Diaz-de-MariaEdward J.Delp."Neighborhood Matching For Image Retrieval."*IEEE Transactions on Multimedia*19(3):544–558.