Defence Science Journal, Vol. 58, No. 1, January 2008, pp. 159-167

Object Hierarchy-based Supervised Characterisation of Synthetic Aperture Radar Sensor Images

Ish Rishabh and Subrata Rakshit

Centre for Artificial Intelligence & Robotics, Bangalore–560 093

ABSTRACT

A method of supervised characterisation of synthetic aperture radar (SAR) satellite images has been discussed in which simple object shape features of satellite images have been used to classify and describe the terrain types. This scheme is based on a multilevel approach in which objects of interest are first segmented out from the image and subsequently characterised based on their shape features. Once all objects have been characterised, the entire image can be characterised. Emphasis has been laid on the hierarchical information extraction from the image which enables greater flexibility in characterising the image and is not restricted to mere classification. The paper also describes a method for giving relative importance among features, i.e., to give more weights to those features that are better than others in distinguishing between competing classes. A method of comparing two SAR sensor images based on terrain elements present in the images has also been described here.

Keywords: Object segmentation, object characterisation, image characterisation, terrain-based matching, SAR imagery

NOMENCLATURE

A	Area of the object
P	Perimeter of the object
D	Diameter of the object, defined as the maximum Euclidean distance between any two peripheral points
R_{avg}	Average distance of the peripheral points from the centroid
$R_{\max}$	Radius of the object, defined as maximum Euclidean distance of any peripheral point from the centroid (C) of the object. Need not to be half of D.
th_{bright}	Threshold for bright objects
μ_p	Mean intensity of the image
σ_p	Standard deviation of pixel intensities

1. INTRODUCTION

Classification of images is one of the most extensively researched areas of image processing. It is often required for both military and civil applications. Classification systems working on satellite imagery often use pixel-level properties to classify pixels in an image. This, however, restricts the applicability of such systems to answer queries related to classification only. Since the decision is made at pixel-level only, such an approach cannot characterise the entire image. In the proposed approach, multiple levels of information extraction

Received 25 January 2007, revised 25 May 2007

has taken place, each one arising from the one below it. User-level queries like classification are answered at the final level. This permits enough flexibility to answer the other queries like terrain-based matching of images, detection of changes in scene contents, and so on. The efficacy of this approach has been demonstrated on synthetic aperture radar (SAR) imagery.

Several attempts have been made to classify terrain types present in the SAR images. Statistical and neural network approaches have been used to classify multi-temporal/multiband SAR sensor images using spatial analaysis[1]. Another approach uses an unsupervised scheme using self-organising map (SOM) neural network to classify the SAR images based on texture classification through gray-level co-occurrence probabilities[2]. Many features, that can be used for the SAR classification, have been explored and compiled[3]. The authors have examined the utility of nonstandard features, eg, spatial variability between neighbouring pixels for classification.

The methods cited above build a detector or classifier for a specific task. The proposed scheme is different in that as it retains enough information at the image level to answer not only classification or detection queries but also some other user queries like textual image description, terrain-based image matching, object extraction, and so on. The paper describes the multilevel hierarchical characterisation of images and further explains building up of scene-level description from the objects.

2. HIERARCHICAL IMAGE CHARACTERISATION

As mentioned, a hierarchical approach is adopted in which information extracted from pixels is used to build object-level information, which, in turn, is used to derive terrain-level information, subsequently leading to information at the scene-level.

The purpose of this study is to demonstrate the implementation of such a hierarchical scheme for characterisation of satellite imagery. The information extracted at the lowest-level is used not only to classify objects into various terrain types but also serves as input to higher-level information structures which can then be used to answer the higher-level queries. The implementation has been done on the SAR images. Several simplifying assumptions have been made in this regard. Only the magnitude of SAR images has been used. Phase or polarisation information has not been made use of. It has also been assumed that lakes and rivers give dark signatures in the magnitude image, whereas mountainous and urban objects give bright reflection signatures.

The actual details of the lowest-level classification depend on the nature of images and the entities of interest for a given application. These would necessarily have to be optimised by each user for each application. The selection of entities and their expected signatures given below are, therefore, only indicative. The focus of this paper is the construction of a hierarchical characterisation scheme that supports multiple user-level queries.

2.1 Lowest (pixel)-level Description

Figure 1 shows a typical SAR image. It is worth mentioning that pixel intensities can be grouped into three categories: Dark, bright, and mean intensity levels.

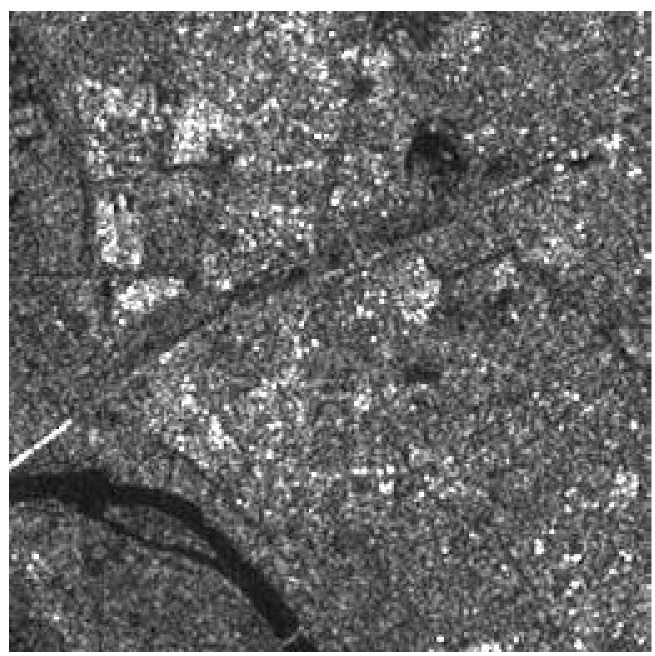


Figure 1. SAR image of an urban region.

The bright spots in Figure 1 occur due to strong reflection of light from right-angled corner reflectors due to vertical structures. In other words, man-made structures like buildings, give bright signatures. Similarly, mountainous terrains give similar bright reflections. On the other hand, rivers, lakes, and other water bodies give darker signatures. A congregation of such pixels forms an object of interest.

Segmentation of these objects is done on the basis of intensity. Thresholds for bright and dark objects are determined separately as

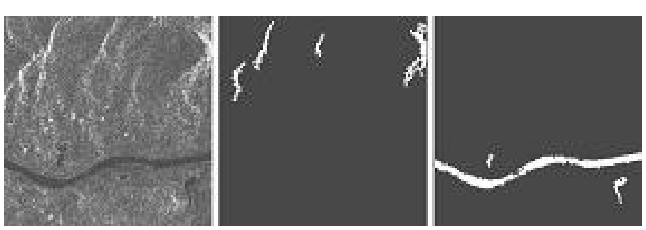
$$th_{bright} = \mu_p + \lambda_b \sigma_p \quad (1)$$

$$th_{dark} = \mu_p + \lambda_d \sigma_p \quad (2)$$

where th_{bright} is the threshold for bright objects, th_{dark} is the threshold for dark objects, μ_p is the mean intensity of the image, and σ_p is the standard deviation of pixel intensities.

To describe an object, only those pixels are considered that adjoin each other (in 8-connected sense) to form objects of reasonable size (in this case, 20 or more). To keep the algorithm independent of the resolution of the image, all thresholds based on sizes are multiplied with a scaling factor (γ), which takes care of resolution change. In this work, γ is 1.0. λ_b and λ_d usually lie between 0.5 and 1.5. These are obtained iteratively till a minimum (pre-decided to be 10) number of objects are segmented out for analysis.

Figure 2 shows a SAR image out of which bright and dark objects have been segmented out.



(a) (b) (c)

Figure 2. Segmentation into bright and dark objects: (a) original image, (b) bright objects, and (c) dark objects.

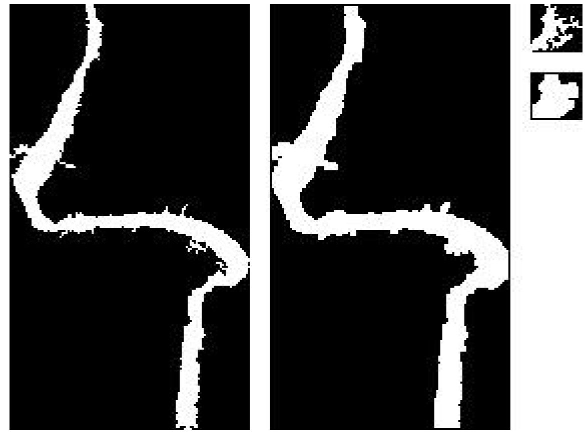


Figure 3. Dilation of extracted objects.

2.2 Object Enumeration and Description

Once these putative objects have been extracted out, they are preprocessed before feature extraction. Each of the objects is dilated so as to fill the irregularities of its boundary. This is done because such indentations unnecessarily increase the perimeter, which is used as part of the features used to characterise its shape. Figure 3 shows objects that have been dilated after extraction. It is further mentioned that dilation reduces unwanted indentations while retaining the overall shape of the objects.

In this study, focus has been put on distinguishing between the following terrains:

- Mountainous terrains
- Urban terrains
- Rivers
- Lakes(includes dams).

As is evident from Figs 1 and 2, bright, long objects characterise mountainous terrains, whereas bright small-oval objects mark urban features. Similarly, rivers and lakes occur as dark objects, which can further be distinguished by their shape features. Shape features have been used to characterise objects. As SAR images do not give reliable edge extracts, edge-based features have not been considered. Features need not be limited to object shapes only. Other features can also be used for this purpose. The main purpose here is to demonstrate buildup of hierarchical information structure from pixel to scene-level. Objects of interest are either long and eccentric or roughly oval. Hence, features that can distinguish between the eccentric and oval shapes have been used. Shape of each object has been described using the following features:

(a) *Roundness*[4]: This is defined as the ratio between the area of a circle whose circumference equals *P*, to the area of the object, *A*.

$$Rnd = \frac{P^2}{4\pi A} \tag{3}$$

(b) *Ovalness*: This is defined as the ratio between the area of the circle whose diameter equals *D*, to the area of the object, *A*.

$$Ovl = \frac{\pi D^2}{4A} \tag{4}$$

(c) *Ratio of Areas*: It is the ratio between the area of a circle whose radius is $R_{\max}$ to the area of the object, *A*.

$$ROA = \frac{\pi R_{\max}^2}{A} \tag{5}$$

(d) *Elliptical Eccentricity*: It is the eccentricity of an ellipse with semi-major axis equal to $R_{\max}$ and semi-minor axis equal to R_{avg}.

$$ee = \sqrt{1 - \frac{R_{avg}^2}{R_{\max}^2}} \tag{6}$$

(e) *Eccentricity*[4]: It is defined as the ratio of the minimum moment of inertia ($I_{\min}$) of the shape, to the maximum moment of inertia ($I_{\max}$).

$$ecc = \frac{I_{\min}}{I_{\max}} \tag{7}$$

These features are grouped to form a feature vector, *f*.

To summarise, the following steps are performed:

- Threshold image (twice) to obtain dark and bright objects separately.
- For each object, extract five features to form a feature vector, *f*.

It is re-emphasised here that only intensity and shape information have been used to obtain the feature vector, *f*. Texture features have not been made use of.

2.3 From Objects to Terrain Elements

Object with shape feature vector (*f*) is classified into one of the terrain types based on its distance from the cluster means of the different terrain types. However, before the classifier for objects is ready, it needs to be trained. This is done using a training set through which cluster mean for each terrain type is determined.

(a) *Training Set*: For each of the terrain types, a large number (50 to 100) of objects belonging to that type are extracted. Then,

- Determine the features mentioned above for each of these extracted objects.
- Find the mean ($\mu_{i,c}$) and standard deviation ($\sigma_{i,c}$) for all features for all the objects belonging to class (*c*) [terrain type]. Here, *i* represents the i^{th} feature.

Table 1 shows the means of features for objects pertaining to specific terrain types

Table 1. Means of features for each terrain type

Feature	Mnt	Urban	River	Lakes
Rnd	10.7290	2.913	16.7720	4.359
Ovl	10.5280	2.325	14.1230	2.763
ROA	11.8120	2.587	16.5290	3.356
ee	0.8728	0.801	0.8679	0.818
ecc	0.0378	0.426	0.0777	0.379

Table 2 shows the standard deviations of features for objects pertaining to specific terrain types.

Table 2. Standard deviations of features for each terrain type

Feature	Mnt	Urban	River	Lakes
Rnd	4.770	1.501	6.518	2.137
Ovl	4.065	0.773	5.082	1.046
ROA	4.649	0.978	6.352	1.416
ee	0.017	0.053	0.021	0.060
ecc	0.057	0.219	0.107	0.221

Standard deviation of the i^{th} feature for all objects together, irrespective of the class to which they belong, is also determined. These are denoted as μ_i and σ_i, respectively. Table 3 shows these values.

Table 3. Standard deviations of features irrespective of terrain types

Ftr.	*Rnd*	*Ovl*	*ROA*	*ee*	*ecc*
	5.621	5.284	6.075	0.054	0.245

(b) *Classification of Objects*: Feature means and standard deviations for each class are represented as vectors for that class by μ_c and σ_c, respectively. Also, σ_f represents the vector formed by the standard deviations in Table 3 above. This vector represents the overall standard deviation of features.

Any object vector (*f*) can be classified into one of the classes by evaluating its distance from each of the mean vectors of the class.

It is mentioned here that the classification of objects should be done separately for objects extracted by the two thresholding steps. In other words, objects are either classified into mountain or urban, or into rivers or lakes. Hence the task is reduced to a 2-class classification problem.

(c) *Distance*: If *f* is the feature vector of the object to be classified, then its Euclidean distance ($D_{Euclidean}$) in feature space from the mean vector (μ_c) of class (*c*) is defined as

$$D_{Euclidean} = \sqrt{\sum_{i=1}^{F} \left(f_i - \mu_{i,c}\right)^2} \tag{8}$$

Here, F is the number of features.

However, the feature axes are usually not isotropic, as is evident from Table 3. Some of the features have higher dynamic ranges (and hence higher standard deviations) than the others. The axes are, therefore, made isotropic by dividing by the Euclidean distance of that feature by the standard deviation of the feature. This distance is known as Mahalanobis distance and is determined as

$$D_{Mahalanobis} = \sqrt{\sum_{i=1}^{F} \left[\frac{f_i - \mu_{i,c}}{\sigma_i}\right]^2} \tag{9}$$

It is found that some features give better discrimination between competing classes than the others and hence should be given more importance or weight than other features. Therefore, the distance equation is further modified to incorporate relative weights between features as

$$D_{weighted} = \sqrt{\sum_{i=1}^{F} w_i \times \left[\frac{f_i - \mu_{i,c}}{\sigma_i}\right]^2} \tag{10}$$

such that

$$\sum_{i=1}^{F} w_i = 1 \tag{11}$$

(d) *Feature Weights*: Weights should be chosen on the basis of the ability of a feature to discriminate between competing classes. Different features give different separation for different classes. Accordingly, the weights should not only be feature-specific, but also class-specific.

Figure 4 shows how weights should be chosen. If there are *N* competing classes with means and standard deviations as ($\mu_{i,c}$, $\sigma_{i,c}$), for i^{th} feature, then for class (*c*), one has:

$$w'_{i,c} = \frac{\left|\left(\mu_{i,c} - \mu_{i,nearest}\right)\right|}{\max\left(\sigma_{i,c}, \sigma_{i,nearest}\right)} \tag{12}$$

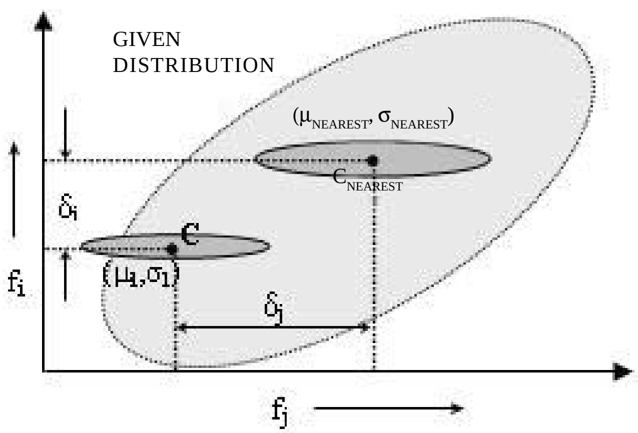


Figure 4. Determination of feature weights.

$$w_{i,c} = \frac{w'_{i,c}}{\sum_{i}^{F} w'_{i,c}} \qquad (13)$$

As is evident from Fig. 4, even though $\delta_i<\delta_j$, feature *i* should get more weight because the two clusters are better separated (less or no overlap) along feature *i* than feature *j*, as is evident from the lower value of the standard deviation of *i* for a class than for *j*.

Table 4 gives the feature weights-specific to the classes (terrain types).

Table 4. Normalised feature weights for each terrain type

Feature	Mnt	Urban	River	Lakes
Rnd	0.1869	0.1869	0.2265	0.2265
Ovl	0.2301	0.2301	0.2658	0.2658
ROA	0.2263	0.2263	0.2466	0.2466
ee	0.1545	0.1545	0.0989	0.0989
ecc	0.2021	0.2021	0.1621	0.1621

Since it is a 2-class classification problem, the weights of class mountain are equal to that of class urban. Similarly, weights of class river are the same as those of class lakes. In general, these weights may not be the same because class *i* may have class *j* as the nearest neighbour but class *j* may not necessarily have class *i* as the nearest neighbour.

3. SCENE DESCRIPTION BASED ON TERRAIN ELEMENTS

This work focuses on four terrain types. To obtain a quantitative description of a terrain, a terrain descriptor is determined. This demonstrates how higher-level information structures can be built using lower-level information. in this case, there are four terrain descriptors for an image, one for each terrain type.

3.1 Terrain Descriptors

A set $S = \{TD_i\}$ is computed for the entire image, where TD_i is the terrain descriptor for class *i*. Here, $i = \{1, 2, 3, 4\}$ where 1 denotes class mountain, 2 denotes class urban, 3 denotes class river and 4 denotes lakes class.

To determine TD_i for class *i*, all objects that have been classified into class *i* are considered. The following attributes are derived for TD_i.

- Number of objects classified into class *i*.
- Area coverage of terrain is the sum of the areas of the objects belonging to that terrain.
- *Number of Prominent Objects*: Prominence is defined on the basis of the distance of the object from terrain's mean vector. A threshold is chosen for prominence through supervised learning.
- *Terrain Vector*: This is determined by weighing the distance of each object, from terrain's mean vector, by the area of the object and dividing the sum of these area-weighted distances by the total area coverage of the terrain (*i*). This feature is used to determine whether the terrain type is prominent in the image or not. As is evident, this quantity is always positive.

S gives a quantitative description of the terrain types present in the image. Comparing terrain vectors can do terrain-type comparisons.

3.2 Image Analysis based on Terrain Descriptors

Image-level queries can be handled once the set *S* has been computed. The following analysis of the image can now be done:

(a) *Image Classification*: Image gets classified into one or more terrain types. Unlike an object, an image can get classified into more than one terrain types. This is valid because there can be more than one terrain types present in the image.

(b) *Image Description*: To describe the image, describe each of the terrain types present in the image. For each terrain type, describe its area coverage, number of objects belonging to it and the number of prominent objects.

(c) *Terrain Extraction*: To extract a terrain, all the objects belonging to that terrain type are extracted.

(d) *Terrain Matching*: It often happens that there is a need to match two images and determine how closely they relate to each other in terms of their scene contents. Terrain descriptors facilitate this task.

Let images I_1 and I_2 have terrain descriptor sets S_1 and S_2, where $S_1 = \{TD1_i\}$ and $S_2 = \{TD2_i\}$ and i denotes the terrain type. Treating TD as a vector, for each terrain type i, the cosine of the angle between the $TD1_i$ and $TD2_i$ is found. This gives a value between 0 and 1, which indicates how closely the terrain type present in one image matches with the corresponding terrain type present in the other image. Thus, for the two images, 4 cosine terms are obtained, one each for each terrain type. A weighted average of these cosines is then taken. The weights to be used are the coverage of each terrain in the two images, normalised by the sum total of the weights.

The value obtained lies between 0 and 1. For images depicting similar terrain types, this value is close to 1 and for those images that depict dissimilar terrain types, this value is closer to 0. Hence, a measure of similarity between the two images can be determined.

Advantages of using a hierarchical approach for image analysis are as follows:

- *Flexibility*: As shown here, separate high-level functions like classification, description, extraction and matching can be based upon a single hierarchical framework. In future, if some other kind of query comes, the only modification needed is at the user query-level. The entire code of pixel-level, object-level and scene-level need not be modified.

- *Domain Knowledge*: At the scene-level, domain knowledge can be incorporated specifically. For example, a terrain type can be defined as prominent or not, depending upon its coverage or its terrain vector, and so on. If prominence is defined as terrain coverage, then urban terrain may never appear as prominent as compared to that of mountains because urban objects are considerably smaller than mountainous objects. Domain knowledge can be used here to give weight to urban objects while comparing them to mountainous ones.

4. RESULTS

In this study, only the magnitude of the SAR imagery has been used for characterisation of terrain. Thus, this method can be used with sensors that do not have phase information and is, therefore, generic.

Results have been compiled under the following headings:

- *Image Classification*: The proposed implementation gives a high accuracy of classification. However, there are misclassification errors in which an object of one-terrain type is misclassified into some other-terrain type. These errors are typically low, with the highest being 10 per cent for objects of river class.

- *Image Description*: Figure 5 shows that a textual description of image contents can be generated. This is especially useful for military applications where textual reports need to be generated from surveillance images.

Urban: Total 46 object(s) found out of which 28 are prominent. Total coverage is 2.67 per cent. Average object size is 38.07 pixels. Overall prominence of the terrain is 0.46.

River: Total 1 object found out of which 1 is prominent. Total coverage is 4.03 per cent. Average object size is 2643.00 pixels. Overall prominence of the terrain is 0.44.

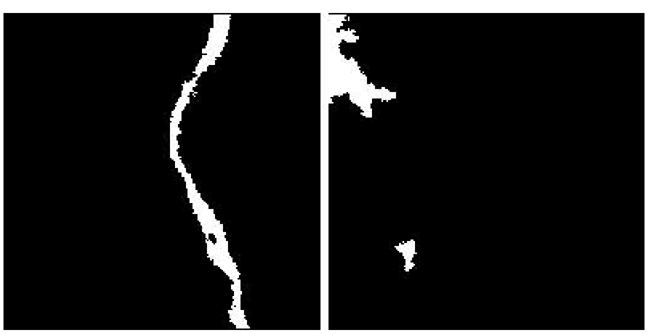
Lakes: Total 2 object(s) found out of which 2 are prominent. Total coverage is 2.92 per cent. Average object size is 959.50 pixels. Overall prominence of the terrain is 0.34.

Figure 5. Description of the SAR image of Fig. 6.

- *Terrain Extraction*: Figure 7 shows river and lakes extracted out of the image shown in Fig. 6. Figures 8 and 9 show similar terrain, extraction for mountainous and urban terrains respectively.

- *Terrain Matching*: To show the accuracy of terrain matching, one of the images has been kept common between two different figures.

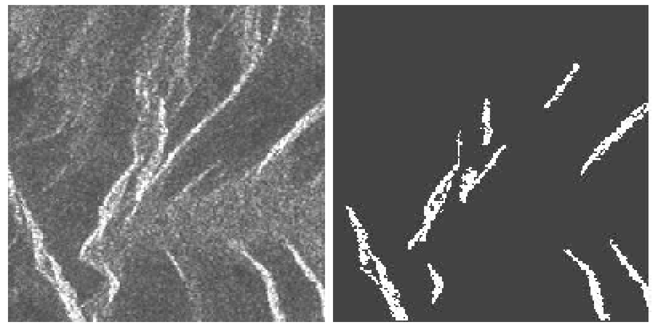
Figure 6. Urban area image with river and lakes.



(a) (b)

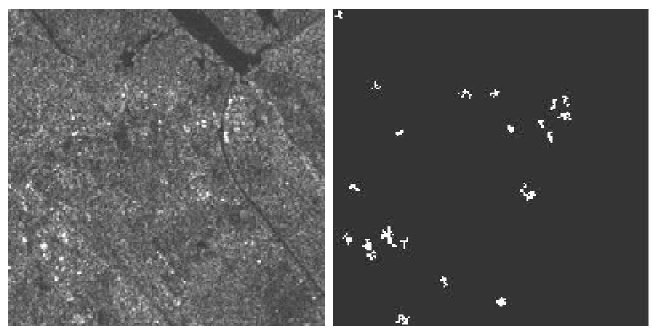
Figure 7. Terrain extraction from Fig. 6: (a) river extract and (b) lakes' extracts.

Figure 10 shows two very similar terrains and their corresponding degree of closeness. Figure 11 also has similar images but the image on the right doesn't have a river. Hence, the two of them aren't as similar as Fig. 10. A similar situation with mountains is shown in Figs 12 and 13.



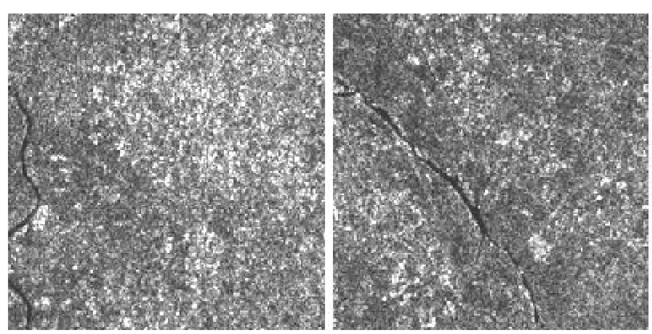
(a) (b)

Figure 8. Terrain extraction: (a) original mountainous image and (b) mountain extracts.



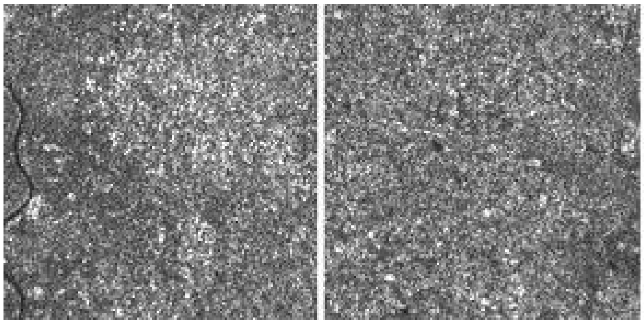
(a) (b)

Figure 9. Terrain extraction: (a) original urban image and (b) urban extracts.



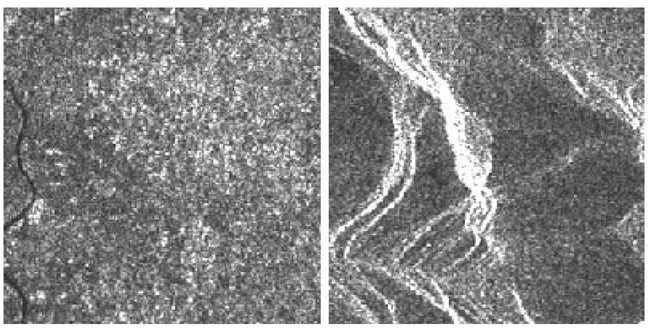
(a) (b)

Figure 10. Terrain matching: (a) urban image with river and (b) another urban image with river. Measure of similarity = 0.99.



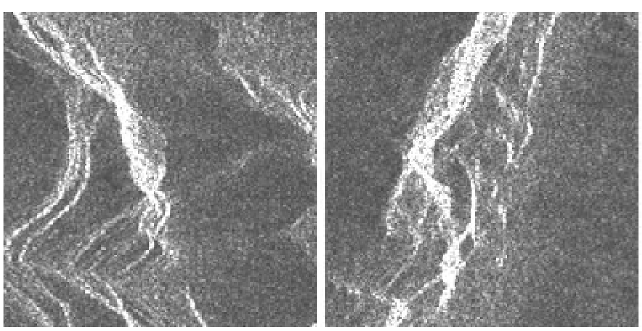
(a) (b)

Figure 11. Terrain matching: (a) urban image with river and (b) another urban image without river. Measure of similarity = 0.69.



(a) (b)

Figure 12. Terrain matching: (a) urban image with river and (b) mountainous terrain. Measure of similarity = 0.29.



(a) (b)

Figure 13. Terrain matching: (a) mountainous terrain and (b) another mountainous terrain. Measure of similarity = 1.

5. CONCLUSION AND FUTURE WORK

Ability to characterise a SAR satellite image using multilevel hierarchical characterisation has been demonstrated in the proposed framework. Only simple shape-based features have been used which are easy and computationally inexpensive to extract. As results show, shape-based features are sufficient to describe, classify, and match terrains in images. Future work can comprise a similar characterisation for higher resolution images. Texture measures can also be used along with shape-based features to obtain a better quantitative description of objects, especially for higher-resolution imagery.

ACKNOWLEDGEMENTS

The authors wish to thank Director, Centre for Artificial Intelligence & Robotics (CAIR), Bangalore, for his support and encouragement, and their colleagues of Computer Vision Group for their help. They also thank Shri M.R. Keshorey, Dr G. Athithan and Dr N. Ramamurthy for their valuable feedback comments on the paper.

REFERENCES

1. Pellizzeri, T.M.; Gamba, P.; Lombardo, P. & Dell'Acqua, F. Multitemporal/multi-band SAR classification of urban areas using spatial analysis: Statistical versus neural kernel-based approach. *IEEE Trans. Geosci. Remote Sensing*, 2003, **41**(10), 2338-353.

2. Ersahin, K.; Scheuchi, B. & Cumming, I. Incorporating texture information into polarimetric radar classification using neural networks. *In* IEEE International Geoscience and Remote Sensing Symposium, IGARSS'04, 20-24 September 2004. pp. 563.

3. Pierce, L. An exploration of features for SAR classification. *In* IEEE International Geoscience and Remote Sensing Symposium, IGARSS'97, 3-8 August 1997. pp. 2053.

4. Jain, A.K. Image analysis and computer vision. *In* Fundamentals of digital image processing, edited by Thomas Kailath. Prentice Hall, New Jersey, 1989. pp. 390-94.

5. Duda, R.O.; Hart, P.E. & Stork, D.G. Bayesian decision theory. *In* Pattern classification. John Wiley & Sons, Inc., Singapore, 2001.