Saliency Attention and SIFT Keypoints Combination for Automatic Target Recognition on MSTAR dataset

Ayoub Karine, Abdelmalek Toumi, Ali Khenchaf, Mohammed El Hassouni

Saliency Attention and SIFT Keypoints Combination for Automatic Target Recognition on MSTAR dataset

Conference Paper · May 2017

DOI: 10.1109/ATSIP.2017.8075558

4 authors:

Ayoub Karine
ENSTA Bretagne

Abdelmalek Toumi
ENSTA Bretagne

Ali Khenchaf
ENSTA Bretagne

Mohammed El Hassouni
Mohammed V University in Rabat

3rd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2017, May 22-24, 2017, Fez, Morroco.

Saliency Attention and SIFT Keypoints Combination for Automatic Target Recognition on MSTAR dataset

Ayoub Karine[1, 2], Abdelmalek Toumi[1], Ali Khenchaf[1], Mohammed El Hassouni[2, 3]
[1] Lab-STICC UMR CNRS 6285, ENSTA Bretagne, Brest, France
{abdelmalek.toumi, ali.khenchaf}@ensta-bretagne.fr
[2] LRIT URAC 29, Faculty of sciences, Mohammed V University in Rabat, Morocco
{ayoub.karine, mohamed.elhassouni}@gmail.com
[3] DESTEC, FLSHR, Mohammed V University in Rabat, Morocco

***Abstract*—This paper aims to present a novel method for automatic target recognition based on synthetic aperture radar (SAR) images. In order to describe a region of interest (target area), we use a saliency attention model. Then, the produced saliency map is used as a mask on SAR image in order to separate the ground target from the background. After that, we calculate the scale invariant feature transform (SIFT) descriptors of the transformed SAR image. In this way, we maintain only the SIFT keypoints located in the salient region. This strategy leads not only to reduce the dimensionality but also enhances its discriminative power. For recognition step, a matching approach between vector descriptors of unknown image target and all known images stored in training data set is adopted. To validate the proposed approach, MSTAR data set is used. The obtained experimental results show that our approach can effectively describe a SAR image, and obviously improve the recognition rate.**

Keywords*—Automatic target recognition, synthetic aperture radar, SIFT, saliency attention model, matching

I. Introduction

Automatic target recognition (ATR) using Synthetic aperture radar (SAR) images has become an essential research topic for several application fields such as military defense. The ATR-SAR task aims to recognize in automatic way the unknown targets based on its SAR images. For achieving the recognition task, several steps are usually required including data acquisition, feature extraction and classification to build decision making [1, 2]. In the first stage, the SAR images are constructed. It is followed by feature extraction that consists of calculating a signature from each target image. Finally in last stage, these feature vectors are used in classification step to recognize unknown targets. We focus in this paper on the feature extraction and classification steps.

Various ATR approaches based on SAR images have been proposed in recent years. Zhao et al. [3] exploit the raw pixels of SAR images as the input of SVM classifier. In [4], the authors extract the 2-D discrete Fourier transform (DFT) coefficients of the cropped images. The global features descriptor used in the recognition task in [5] is composed by the combination of different feature descriptors which was used in the SVM classifier. Srinivas et al. [6] propose a meta-feature vector by combining three features extraction methods and different classifiers. The obtained features are classified by using SVM or AdaBoost classifiers. Agrawal et al. [7] propose to use SIFT descriptors applied on segmented SAR image. Karine et al. [8] proposed a new statistical approach for SAR recognition. H. Song et al. [9] use a sparse representation-based classification with different feature extraction methods. Recently, some authors [10, 11] lie in the direction to use SAR image pixels directly as an input of deep learning method. In this way, the feature extraction step is not taken into account in target recognition process.

In this paper, we propose a new ATR system based on a combination of SIFT and saliency attention methods to compute feature descriptors which are used in the classification step. The classification task is achieved by a matching function between an unknown feature vectors and the known features vectors stored in training dataset. Despite that the SIFT method demonstrates best performance in different applications [7, 12, 13], it presents the limitation of the huge dimension of produced descriptor. To overcome this limit, we propose to reduce the SIFT keypoints using an Itti saliency attention model [14]. For doing so, we firstly compute a saliency map of SAR image with the aim to locate the salient region of an image which is the target area in this study. After that, we calculate the SIFT descriptor of the produced image. In this way, we keep only the SIFT the salient region keypoints. To recognize the SAR target, we adopt a matching approach between produced keypoints of unknown and training targets. The flowchart of the proposed approach is illustrated in Figure 1.

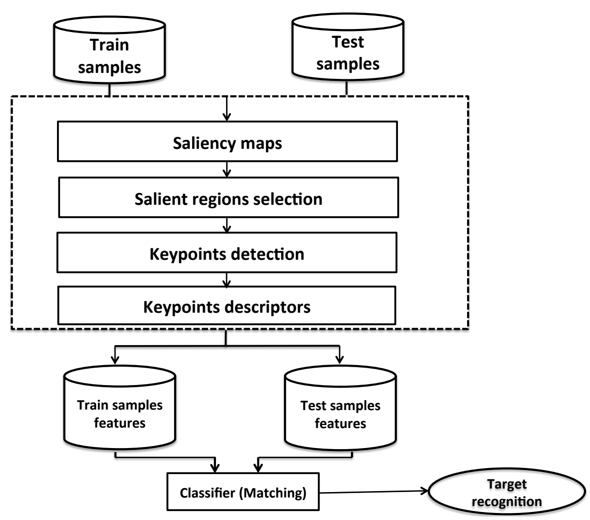

Fig. 1: The general steps of the proposed system.

The organization of this paper is as follows. In section II, we give an overall scheme of different steps of the proposed descriptor. Section III describes the classification process step. The experimental results will be presented in section IV. Finally, the paper is concluded in section V.

II. FEATURE EXTRACTION

In this step, The SIFT descriptors are proposed to describe target in SAR images. However, not all the extracted SIFT keypoints are useful to represent the SAR image as shown in Figure 2(b). For this reason, we aim to compute only the keypoints existing in the salient region in order to describe only the target area image. For doing so, we firstly compute a saliency map using Itti model [14]. This map is used as a mask of SAR image. After that, we compute the SIFT descriptors of the produced SAR image. We display in Figure 2 the application of the mentioned steps on an example of SAR image. By taking a visual comparison between Figure 2(b) and 2(d), it is clear that all salient keypoints are concentrated in the salient region of the SAR image. We give in next subsection a brief review of Itti's model as well as the SIFT method.

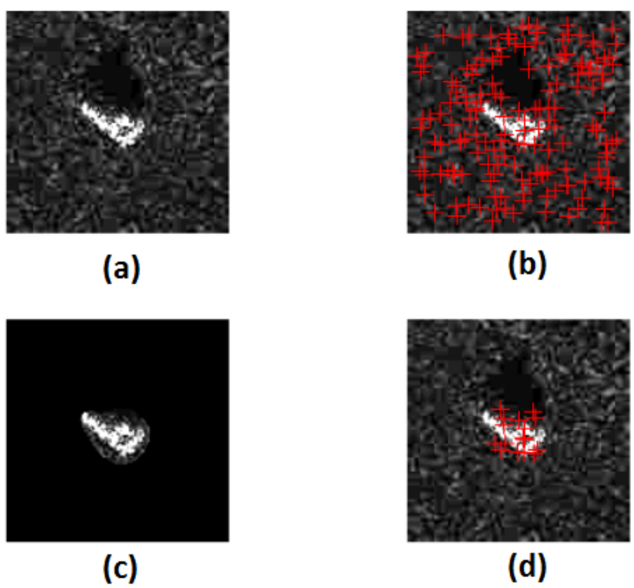

Fig. 2: SIFT and Salient SIFT keypoints distribution. (a)SAR image. (b) SIFT keypoints distribution. (c) Salient region of the SAR image. (d) Salient SIFT keypoints distribution.

A. Visual attention model

Recently, several works give a special attention to the visual attention modeling for several applications in image processing [15]. The idea behind this research field is to detect image regions that attract the observe. These images regions are called salient region.

Among the first saliency models proposed in the literature, we find the Itti saliency model [14]. This model integrates three feature channels: intensity, color and orientation. In this work, we test our method on grayscale SAR images, consequently, we don't take into account the color information. Using the intensity information, this saliency model generates a Gaussian pyramid $I(\sigma)$ where $\sigma \in [0..8]$ is the scale. After that, the oriented Gabor pyramids $O(\sigma, \theta)$, where $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ is the orientation angles. Based on intensity and orientation channels, the center-surround difference ($\ominus$) between a center c and a surround s is calculated. As a result, 30 feature maps (FM) are generated :

- 6 FM for intensity:
$$I(c,s) = \mid I(c) \ominus I(s) \mid \tag{1}$$
where $c \in \{2,3,4\}$
- 24 FM for orientation:
$$O(c,s,\theta) = \mid O(c,\theta) \ominus O(s,\theta) \mid \tag{2}$$
where $s = c + \delta, \delta \in \{3,4\}$

The feature maps found in the previous step is normalized ($N(.)$) and combined using the accross-scale addition ($\oplus$). As a result, we obtain two conspicuity maps : $\bar{I}$ for intensity and $\bar{O}$ for orientation:

$$\begin{aligned} \bar{I} &= \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(I(c,s)). \\ \bar{O} &= \sum_{\theta \in \{0^\circ,45^\circ,90^\circ,135^\circ\}} N\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(O(c,s,\theta)) \right). \end{aligned} \tag{3}$$

The final saliency map is simply obtained using the summation and the normalization of the two conspicuity maps as follows:

$$S = \frac{1}{2} \left(N(\bar{I}) + N(\bar{O}) \right). \tag{4}$$

We display in Figure 3 an example of SAR image and the corresponding saliency map. We can see that the region of interest which is target area is located in the saliency map.

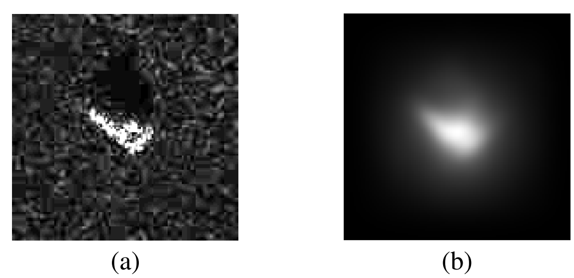

Fig. 3: (a) An example of SAR image. (b) Corresponding saliency map.

B. SIFT keypoint detection and description

The Scale Invariant Feature Transform (SIFT) method aims to extract a local feature of images. It is proposed by [16]. This algorithm is used in several image processing applications, e.g., face recognition, biometric and others. Generally, the SIFT method use four steps:

1) Scale-space extrema detection: To transform an image to a scale space, the convolution of the image $I(x, y)$ with a variable-scale Gaussian $G(x, y, \sigma)$ is performed:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \tag{5}$$

where σ represents standard deviation of the Gaussian distribution. After that, the difference-of-Gaussian (DOG) is calculated as follows:

$$DOG(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma). \tag{6}$$

Where k control the difference between two nearby scales.
A pixel of the DOG is considered as a keypoint if it is the local maxima or minima compared with 26 pixels neighbors as illustrated in Figure 4.

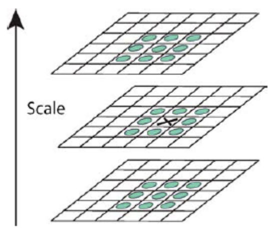

Fig. 4: Comparison of the keypoints with 26 pixel values [16].

2) Keypoint localization: The founded keypoints are filtered in order to pick the best ones and reject the unstable ones.

3) Orientation assignment: A gradient orientation histogram with 36 bins is calculated by considering a neighbor region around a keypoint. The orientation of the keypoint correspond to the peak of this histogram.

4) Keypoint descriptor: This step aims to compute a feature vector (descriptor) for each keypoint. It is done by considering a neighboring region around a pixel. The size of this region is 16×16 pixels. This region is splited to 16 sub-regions. Each sub-region has the size of 4×4 pixels. For each sub-region, a weighted histogram of 8 bins is calculated. Consequently, the size of the final descriptor is $128 = 4 \times 4 \times 8$ values.

III. Feature matching

We adopt in this work the nearest neighbor rule (NNR) strategy [16] to measure the distance between SAR images. This technique exploits the matching of keypoints. We assume that the built feature for a given SAR image is described as follows:

$$\mathrm{FV} = [\mathrm{KP}_1, \mathrm{KP}_2, ..., \mathrm{KP}_\mathrm{n}] \tag{7}$$

where n is the number of keypoints and KP_i is the descriptor of the keypoint of index $i = 1, ..., n$, the size of each keypoint KP_i is 128 values.
The distance between a feature vector of test sample FV^t with a feature vector of train sample FV^tr can be formulated as:

$$\mathrm{Dist}(\mathrm{FV}^\mathrm{t}, \mathrm{FV}^\mathrm{tr}) = \frac{1}{\mathrm{n_t}} \sum_{i=1}^{\mathrm{n_t}} \min\left(\mathrm{dist}(\mathrm{KP}_i^t, \mathrm{KP}_j^\mathrm{tr})_{j=1,\ldots,\mathrm{ntr}}\right) \tag{8}$$

where "dist" is a distance between two keypoints descriptors and $\mathrm{n_{tr}}$ and $\mathrm{n_t}$ (generally $\mathrm{nt} \neq \mathrm{ntr}$) are the number of keypoints in training and test sets respectively. Finally, the target type (class) of an unknown SAR image is the class of image that gives the small distance.

IV. Experimental results

To evaluate the proposed approach, we use three classes of the Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset[1] [17] which is widely used to assess the performance of SAR-ATR algorithms. This database include three different categories ground military targets : BMP2, BTR70 and T72 (T72 is a tank and the other two vehicles are armored personnel carriers). These SAR images are collected using an X-band SAR sensor at two different depression angles (15° and 17°). Figure 5 depicts an example of SAR images of three types of targets and their corresponding optical images. Each SAR image in the database has the size of 128×128 pixels. This database is divided into training and test sets which use targets at the 17° and 15° depression angles respectively. The number of images in each target type (class) is summarized in Table I. The total number of SAR images in training set is 689 whereas it is 1365 in test set.

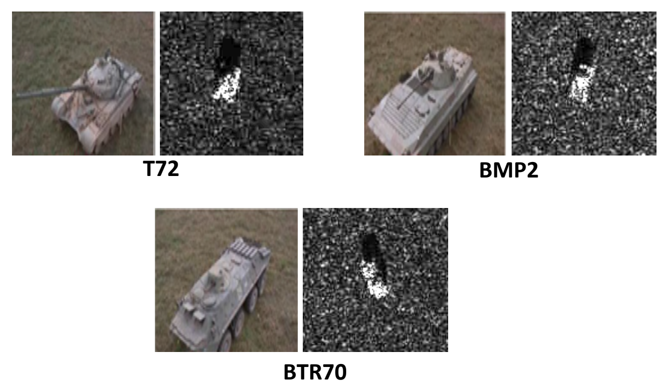

Fig. 5: Example of SAR images in MSTAR database.

TABLE I: Composition of training and test sets of MSTAR database.

Training targets	# Training set	Test targets	# Test set
BMP2(snc21)	233	BMP2(snc21)	196
		BMP2(snc9563)	195
		BMP2 (snc9566)	196
BTR70(c71)	233	BTR70(c71)	196
T72(sn132)	232	T72(sn132)	196
		T72(sn812)	195
		T72(sn7)	191

For each SAR image in database, we locate the salient region using the Itti saliency model. From the produced SAR image, we calculate the SIFT descriptor with 4 octaves and 5 levels per octave. As a result, we present a SAR image by a reduced number of SIFT keypoints. Finally, to recognize the

[1]https://www.sdms.afrl.af.mil/index.php?collection=mstar&page=targets

SAR image, we use a matching approach between keypoints of test and training sets using several distances. We compare in Figure 6 different distances function which are: Chi-square, Euclidean and Manhattan. According to Figure 6, it is obvious that the Euclidean distance performs worst than the other two distances. In addition, the Chi-square distance works a little better than the Manhattan distance in different classes and also in average recognition rate (ALL). For the next of experiments, we choose the Chi-square as a distance of choice.

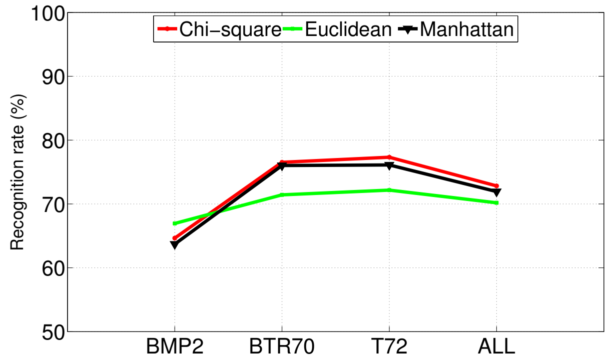

Fig. 6: Recognition performance comparison of different distances functions.

We show in Table II the comparison between the proposed method and two other methods of feature extraction and description which are SIFT [16] and Sal-SIFT [18] in terms of average number of keypoints and matching times. The Sal-SIFT refers to the algorithm proposed in our previous work [18]. It consists on computing firstly the SIFT keypoints for the input radar image and after that we filter the produced keypoints using a saliency model. For SIFT method, the average number of keypoints is 240 for each image training set and 246 for each image in test set. Whereas, it is 12 for for each image in training set and 11 for for each image in test set using the proposed method. Therefore, it is reduced by 95 % and 97 % for training and test sets respectively. Consequently, the runtime for matching the keypoints is significantly reduced by 97 %. For instance, the SIFT descriptor requires 27 times more runtime than the proposed method to match all test keypoints with the training ones. Comparing the both algorithms (Sal-SIFT and proposed one) that combine SIFT and saliency model, the proposed approach has the faster time for matching. We mention here that all algorithms are executed in Matlab 2016 environment with 3.10 GHz Intel processor and 8 Gb of memory.

TABLE II: Comparison between SIFT and proposed method in terms of average number of keypoints and matching time.

	SIFT [16]		Sal-SIFT [18]		Proposed	
	Train	Test	Train	Test	Train	Test
Average number of keypoints	240	246	14	25	12	11
Matching time (s)	8898		388		320	
Matching time per image(s)	6.51		0.28		**0.23**	

We display in Figure 7 a comparison between SIFT and Sal-SIFT methods and the proposed one in terms of recognition rate. It is obvious that the proposed method clearly outperforms the both other methods for each class and also for average recognition rate (ALL). From these results, it is obvious that with a reduced number of keypoints, we achieve a better recognition rate comparing to the SIFT ans Sal-SIFT descriptors on MSTAR dataset. This is because our method exploits only the useful keypoints concentrated in the salient area and remove those located in the background (outliers). For all above performance comparisons, it can be summarized that the proposed method achieves a high trade-off between recognition rate and runtime.

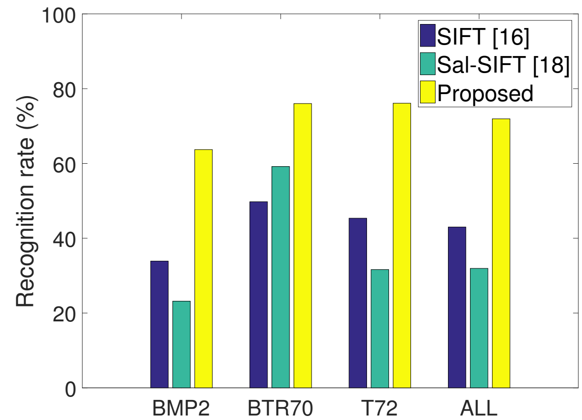

Fig. 7: Recognition performance comparison between SIFT without filtering and the proposed method.

V. CONCLUSION

In this paper, a novel method for automatic target recognition in SAR images is proposed. It consists on matching the SIFT keypoints located in the region of interest (salient region). For this end, we use an Itti's model to select the salient region from a SAR image. This salient region is used as a mask to compute only the SIFT keypoints located in it. In this way, the huge dimension of SIFT descriptors is reduced to distinctive ones. Thanks to the disregard of the keypoints located in image background regions, our method achieves good results in MSTAR database. The next step in the further development of this algorithm will be to take into account also the target shadow information in the recognition process. To achieve this, we will extract the local features of target and shadow areas separately and fuse them in the sparse classification framework. This may improve the recognition performance. This future direction must be tested also on noisy SAR images.

REFERENCES

[1] A. Toumi, A. Khenchaf, and B. Hoeltzener, "A retrieval system from inverse synthetic aperture radar images: Application to radar target recognition," *Information Sciences*, vol. 196, pp. 73–96, 2012.

[2] A. Karine, A. Toumi, A. Khenchaf, and M. E. Hassouni, "Visual salient sift keypoints descriptors for automatic

target recognition," in *IEEE European Workshop on Visual Information Processing (EUVIP), Marseille, France*, 2016.

[3] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, Apr 2001.

[4] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, pp. 112–125, January 2007.

[5] C. Tison, N. Pourthie, and J. C. Souyris, "Target recognition in SAR images with support vector machines (SVM)," in *IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain*, July 2007, pp. 456–459.

[6] U. Srinivas, V. Monga, and R. G. Raj, "Meta-classifiers for exploiting feature dependencies in automatic target recognition," in *IEEE RadarCon (RADAR), Kensas, USA*, May 2011, pp. 147–151.

[7] A. Agrawal, P. Mangalraj, and M. A. Bisherwal, "Target detection in SAR images using SIFT," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Abu Dhabi, UAE*, Dec 2015, pp. 90–94.

[8] A. Karine, A. Toumi, A. Khenchaf, and M. E. Hassouni, "A non-gaussian statistical modeling of sift and dt-cwt for radar target recognition," in *IEEE/ACS International Conference of Computer Systems and Applications (AICCSA), Marrakech, Morocco*, 2016.

[9] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[10] A. Housseini, A. Toumi, and A. Khenchaf, "Deep learning for target recognition from SAR images," in *7th Seminar on Detection systems: Architectures and Technologies, Algiers, Algeria*, February 2017.

[11] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, Aug 2016.

[12] C. Geng and X. Jiang, "Face recognition using SIFT features," in *IEEE International Conference on Image Processing (ICIP), Cairo, Egypt*, Nov 2009, pp. 3313–3316.

[13] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A sift-like algorithm for sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453–466, Jan 2015.

[14] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[15] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, Jan 2013.

[16] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[17] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: a tutorial," in *Aerospace/Defense Sensing and Controls*. International Society for Optics and Photonics, 1996, pp. 228–242.

[18] A. Karine, N.-E. Lasmar, A. Baussard, and M. E. hasosuni, "Sonar image segmentation based on statistical modeling of wavelet subbands," in *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), Marrakech, Morocco*, November 2015.