

Correspondence

Probabilistic Winner-Take-All Segmentation of Images with Application to Ship Detection

Hossam Osman and Steven D. Blostein

Abstract—A recent neural clustering scheme called “probabilistic winner-take-all (PWTA)” is applied to image segmentation. It is demonstrated that PWTA avoids underutilization of clusters by adapting the form of the cluster-conditional probability density function as clustering proceeds. A modification to PWTA is introduced so as to explicitly utilize the spatial continuity of image regions and thus improve the PWTA segmentation performance. The effectiveness of PWTA is then demonstrated through the segmentation of airborne synthetic aperture radar (SAR) images of ocean surfaces so as to detect ship signatures, where an approach is proposed to find a suitable value for the number of clusters required for this application. Results show that PWTA gives high segmentation quality and significantly outperforms four other segmentation techniques, namely, 1) K -means, 2) maximum likelihood (ML), 3) backpropagation network (BPN), and 4) histogram thresholding.

Index Terms—Artificial neural networks, cluster analysis, image segmentation, SAR imagery, target detection.

I. INTRODUCTION

Image segmentation, for our purposes, is defined as the partitioning of image pixels into different segments with each segment being perceptually homogeneous. Clustering-based methods of image segmentation have demonstrated success [1]–[4]. Among clustering techniques that have been successfully utilized in image segmentation, K -means [5] and maximum likelihood (ML) [5] have received much attention [1]–[4]. Their popularity generally stems from their relative simplicity and low computational cost, especially when compared to simulated annealing-based procedures [6]–[8]. K -means and ML, on the other hand, frequently suffer from the cluster underutilization problem [9]–[12]. This problem occurs when one or more clusters are rarely adapted or are dominated by single data samples and thus they do not contribute to modeling the data. Recently, Osman and Fahmy [12] have developed a novel neural clustering scheme called “probabilistic winner-take-all (PWTA).” PWTA is adaptive in nature and retains, to a large extent, the simplicity and low computational cost of both K -means and ML. A detailed comparison of the PWTA complexity to the complexity of both K -means and ML has been provided in [12].

This paper investigates employing PWTA in image segmentation. It demonstrates that PWTA avoids underutilization of clusters by adapting the form of the cluster-conditional probability density function as clustering proceeds. The paper also introduces a modification to PWTA so as to explicitly utilize the spatial continuity of image regions and thus improve the PWTA segmentation performance. The

modification is to utilize in clustering an estimate of the cluster *a priori* probabilities, determined at each image pixel, and thus generate a tendency to assign neighboring pixels to the same image segment. Experimental results are presented. The application selected is that of segmenting airborne synthetic aperture radar (SAR) imagery of ocean surfaces so as to detect ship signatures. This SAR ship detection application is of significant industrial relevance and has received recent attention [13]–[17]. It is a challenging application, since SAR images are degraded by the presence of speckle. Speckle adversely affects the assumption that feature vectors extracted from one image class would cluster together in feature space. The validity of this assumption is crucial for the success of utilizing any clustering technique in image segmentation. Moreover, for this particular SAR ship detection application, the patterns to be detected, the ships, usually occupy small image regions and thus the occurrence of the cluster underutilization problem would result in poor detection performance [18]. Previous work [13]–[16] tends to rely on heuristic steps and ad hoc parameters and the development of a general robust algorithm for the application of detecting ships in SAR images is still a subject of active research. Here, it should be mentioned that preliminary experimental results for the PWTA segmentation of SAR images that were preprocessed for speckle removal have been reported in [18] and [19].

This paper is organized as follows: Section II justifies and enhances the PWTA applicability to image segmentation. Section III utilizes PWTA in the SAR ship detection application and compares its performance to that of four other segmentation techniques, namely

- 1) K -means;
- 2) ML;
- 3) backpropagation network (BPN);
- 4) histogram thresholding.

II. PROBABILISTIC WINNER-TAKE-ALL SEGMENTATION OF IMAGES

Consider employing PWTA to partition an image into K different segments. The image in question is raster scanned on a pixel-by-pixel basis and at each pixel position an n -dimensional feature vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ is extracted. This yields a set $\mathcal{T}_L = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^L\}$ of L feature vectors, where L is the number of image pixels and the i th feature vector is denoted by \mathbf{x}^i . PWTA is then applied to subgroup the vectors in \mathcal{T}_L into K different clusters. Once clustering is complete, if \mathbf{x}^i is assigned to the j th cluster, C^j , then the image pixel at which \mathbf{x}^i was extracted will be assigned to the j th image segment. Assume that cluster C^j has a known *a priori* probability $P(C^j)$ and has an axes-aligned elliptical Gaussian density with unknown mean vector $\mathbf{m}_j = (m_{j1}, m_{j2}, \dots, m_{jn})^t$ and unknown diagonal covariance matrix Σ_j whose diagonal components constitute the vector $\boldsymbol{\sigma}_j = (\sigma_{j1}, \sigma_{j2}, \dots, \sigma_{jn})^t$. Then, clustering the feature vectors in \mathcal{T}_L is reduced to estimating \mathbf{m}_j and $\boldsymbol{\sigma}_j$ for $j = 1, \dots, K$. PWTA estimates these unknowns by iterating a one-layer neural network through \mathcal{T}_L until insignificant variation in the estimated values is observed. A PWTA network is shown in Fig. 1, where the features extracted at each pixel position are taken as the texture information included within a $\sqrt{n} \times \sqrt{n}$ window centered on the pixel in question. The j th network unit in Fig. 1 corresponds to the j th cluster. It is fully connected to the n components of feature vector \mathbf{x} , and its activation $a_j(\mathbf{x})$ corresponds to the j th cluster-conditional probability density function. When the i th feature vector, \mathbf{x}^i , is presented to the PWTA network, the j th cluster evaluates its density

Manuscript received April 5, 1998; revised February 13, 2000. This work was supported in part by Lockheed Martin Electronic Systems, Montreal, Quebec, Canada, and in part by the Natural Science and Engineering Research Council (NSERC) of Canada under Grant 177119. This paper was recommended by Associate Editor M. S. Obaidat.

The authors are with the Department of Electrical and Computer Engineering, Queen's University, Kingston, Ont. K7L 3N6, Canada.

Publisher Item Identifier S 1083-4419(00)04127-3.

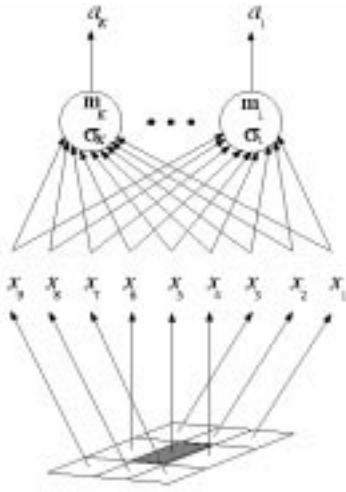


Fig. 1. K -unit PWTA network for image segmentation. Once clustering is complete the shaded, center pixel of the input window is assigned to one of the K segments.

function using [12]

$$a_j(\mathbf{x}^i) = \left(\frac{\mu_j \phi_j^2}{\pi(1 + \mu_j)N_j} \right)^{n/2} \prod_{k=1}^n \frac{1}{\sqrt{\sigma_{jk}}} \times \left(1 + \frac{\mu_j}{(1 + \mu_j)N_j} \frac{(x_k^i - m_{jk})^2}{\sigma_{jk}} \right)^{-N_j/2} \quad (1)$$

where the parameter μ_j reflects the confidence in the value of \mathbf{m}_j as the true unknown value, the parameter N_j reflects the confidence in the value of σ_j as the true unknown value, and ϕ_j is defined by

$$\phi_j \triangleq \frac{\Gamma\{0.5N_j\}}{\Gamma\{0.5(N_j - 1)\}} \quad (2)$$

with $\Gamma\{\cdot\}$ being the Gamma function. After evaluating (1) for all clusters, the j th cluster computes the *a posteriori* probability of \mathbf{x}^i belonging to it using

$$A_j(\mathbf{x}^i) = \frac{a_j(\mathbf{x}^i)P(C^j)}{\sum_{k=1}^K a_k(\mathbf{x}^i)P(C^k)}. \quad (3)$$

Let g denote a randomly generated number that is uniformly distributed between 0 and 1, then \mathbf{x}^i is assigned to cluster C^ℓ if

$$\sum_{j=1}^{\ell-1} A_j(\mathbf{x}^i) \leq g < \sum_{j=1}^{\ell} A_j(\mathbf{x}^i). \quad (4)$$

Assume that at the l th iteration, \mathbf{x}^i is assigned to cluster C^ℓ . Then, the parameters of C^ℓ are adapted using [12]

$$\mathbf{m}_\ell(l) = \frac{\mu_\ell(l-1)\mathbf{m}_\ell(l-1) + \mathbf{x}^i}{1 + \mu_\ell(l-1)} \quad (5)$$

$$\mu_\ell(l) = \mu_\ell(l-1) + 1 \quad (6)$$

$$\sigma_{\ell k}(l) = \frac{N_{\ell k}(l-1)}{N_{\ell k}(l-1) + 1} \sigma_{\ell k}(l-1) + \frac{1}{N_{\ell k}(l-1) + 1} \times \frac{\mu_\ell(l-1)}{1 + \mu_\ell(l-1)} (x_k^i - m_{\ell k}(l-1))^2 \quad (7)$$

$$N_{\ell k}(l) = N_{\ell k}(l-1) + 1 \quad (8)$$

$$\phi_\ell(l) = \frac{0.5(N_{\ell k}(l-1) - 1)}{\phi_\ell(l-1)}. \quad (9)$$

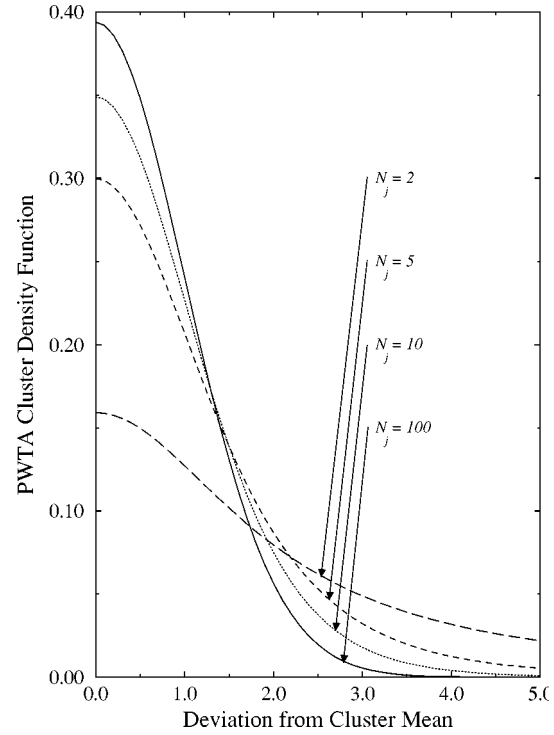


Fig. 2. PWTA cluster density function versus the deviation from the cluster mean for different values of the confidence parameter N_j . The vertical axis is scaled by the standard deviation and the horizontal axis is measured in terms of standard deviation units.

Thus, given input \mathbf{x}^i , the clusters by evaluating their posterior probabilities compete for being assigned this feature vector and only the probabilistically chosen cluster, the winner, is awarded by learning this vector. Once insignificant variation in the estimated values of all means and variances is observed, each feature vector is then assigned to one of the K clusters using

$$\mathbf{x}^i \in C_j, \quad \text{if } A_j(\mathbf{x}^i) > A_k(\mathbf{x}^i), \quad k = 1, \dots, K, \quad k \neq j, i = 1, \dots, L. \quad (10)$$

It should be mentioned that for an asymptotically large number of iterations, it has been proved that the j th cluster density $a_j(\mathbf{x})$ becomes Gaussian with mean \mathbf{m}_j and diagonal variances σ_{jk} , where if the distribution of the input data is identifiable, the estimated values of \mathbf{m}_j and σ_{jk} are identical to the true values [12]. PWTA possesses similar attractive properties in comparison to clustering techniques as K -means [5] and ML [5] that have been successfully employed in image segmentation [1]–[4]. PWTA retains, to a large extent, the simplicity and low computational cost of both K -means and ML. Like ML, it adapts both means and variances of clusters. Therefore, relative to K -means where only means of clusters are adapted, PWTA gives rise to a better modeling of the input distribution. PWTA, on the other hand, avoids underutilization of clusters which is frequently encountered when either K -means or ML is employed [9]–[12]. The problem of underutilization of clusters occurs when one or more clusters are rarely adapted or are dominated by single data samples, and thus they do not contribute to modeling the data. Obviously, a mechanism that is provided by PWTA to avoid this problem is the probabilistic, rather than deterministic, selection of the cluster to be adapted. Thus, for each input presentation, all clusters have a chance to learn, and the cluster with the highest probability of winning is not necessarily chosen as the winner. Originally [12], it has been thought that this is the only mechanism provided by PWTA to refrain from underutilization of clusters. However, in the following subsection, a second mechanism is demonstrated. This will be

followed by introducing a modification to PWTA to enhance its applicability to image segmentation.

A. Avoiding Underutilization of Clusters

Using PWTA, the activation function of the j th network unit, $a_j(\mathbf{x})$, corresponds to the j th cluster-conditional density function and is computed using (1). From (6) and (8), it follows that μ_j is related to N_j through

$$\mu_j = N_j - N_j(0) + \mu_j(0) \quad (11)$$

where $\mu_j(0)$ and $N_j(0)$ are the initial values of μ_j and N_j , respectively. Thus, in view of (1) and (2), $a_j(\mathbf{x})$ can be seen as a function of \mathbf{x} , N_j , \mathbf{m}_j and σ_j , i.e.,

$$a_j(\mathbf{x}) = f(\mathbf{x}, N_j, \mathbf{m}_j, \sigma_j). \quad (12)$$

Since N_j is incremented by 1 each time the j th cluster wins, it is directly proportional to the number of times the j th cluster was the winner. Thus, PWTA introduces a parameter to the cluster density function that is dependent upon the winning count. Fig. 2 plots $a_j(\mathbf{x})$ versus the deviation from the cluster mean for increasing values of N_j assuming \mathbf{x} is one-dimensional. The vertical axis is scaled by the standard deviation $\sqrt{\sigma_j}$ and the horizontal axis is measured in terms of standard deviation units. As shown, as N_j increases, the probability of the j th cluster winning the competition for inputs sufficiently far from its mean decreases. The one dimension assumption is not restrictive since in view of (1), $a_j(\mathbf{x})$ is the outcome of multiplying n one-dimensional densities. Thus, PWTA changes the form of the cluster density function during clustering such that if a cluster learns frequently, the probability of it learning an input feature vector that is sufficiently far from its mean decreases. This gives the other clusters a better chance of learning this input vector. It should be mentioned that this mechanism for avoiding underutilization of clusters can be viewed as an implementation of Grossberg's conscience principle [10]. However, in contrast to a recently-developed clustering technique called "frequency sensitive competitive learning (FSCL)" [11], PWTA implements Grossberg's conscience principle without making any assumptions about the distribution of the input data.

B. Exploiting Spatial Continuity of Images

K -means, ML, and even PWTA do not make explicit use of the spatial continuity of image regions. Usage of spatial continuity is crucial for accurate segmentation, particularly when the underlying images have a "salt-and-pepper" appearance as those speckled images produced in a wide variety of important applications including SAR and laser radar. PWTA, as described above, assumes that the j th cluster *a priori* probability, $P(C^j)$, has a known value that is independent of feature vector \mathbf{x} . A variation to PWTA to make it explicitly utilize the spatial continuity of image regions is to replace $P(C^j)$ by one that is \mathbf{x} -dependent and that has to be estimated as clustering proceeds. Thus, let the j th cluster *a priori* probability, now, be denoted by $P(C^j|\mathbf{x})$ and rewrite (3) as

$$A_j(\mathbf{x}^i) = \frac{a_j(\mathbf{x}^i)P(C^j|\mathbf{x}^i)}{\sum_{k=1}^K a_k(\mathbf{x}^i)P(C^k|\mathbf{x}^i)}. \quad (13)$$

Then, for given \mathbf{x}^i , before computing the *a posteriori* probabilities using (13), a window of size $p \times p$ is centered on the pixel where the vector \mathbf{x}^i was extracted and $P(C^j|\mathbf{x}^i)$ is estimated as $p_j/(p \times p)$, where p_j is the number of feature vectors that were extracted at pixels within this window and that are currently assigned to the j th cluster.

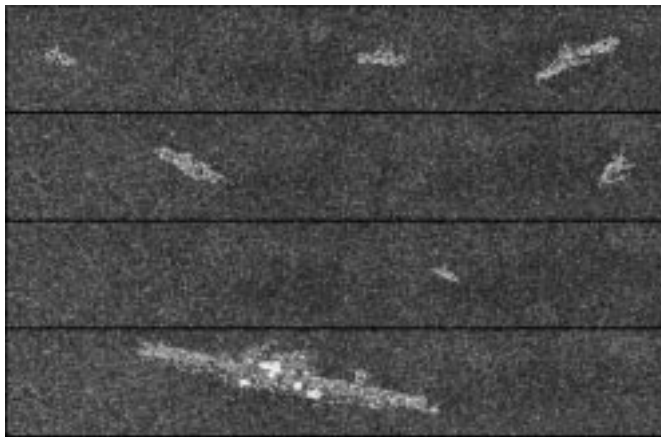


Fig. 3. Synthetic and real airborne SAR images representing ship targets in open-ocean scenes.

The locally-estimated *a priori* probabilities, $P(C^j|\mathbf{x}^i)$, are then substituted into (13) to compute the *a posteriori* probabilities of all clusters. This way, a tendency to assign neighboring pixels to the same image segment is generated. Two remarks are in order here. First, initially, when not all feature vectors are yet assigned to clusters, all $P(C^j|\mathbf{x})$ are set to $1/K$. Second, a tradeoff exists in choosing the size of the window used for estimating $P(C^j|\mathbf{x})$; the smaller the window size the better the utilization of the spatial continuity of image regions, but the less accurate the estimate of the *a priori* probabilities. For the application presented in this paper, experimental observations indicated that window sizes in the vicinity of 10 times the number of *a priori* probabilities provided a good tradeoff and yielded results of similar quality. In the next section, the performance of PWTA is compared to that of other techniques in the application of segmenting SAR images of ocean surfaces so as to detect ship signatures.

III. SHIP DETECTION APPLICATION

We employ the PWTA as modified in Section II-B to partition the airborne SAR images shown in Fig. 3 into two or more segments so as to detect ship signatures. Each of these images has 512×84 pixels. Ship detection is performed by segmenting each image and then determining the ship segment, the segment where most of ship pixels are assigned. The bottom image in Fig. 3 is a real SAR image. The top three are synthetic images that were produced by immersing simulated SAR images of different ship signatures [20] in real SAR open-ocean images. Synthetic images were used to allow a quantitative measure of performance. They were generated using SAR parameters identical to those used for the real ocean background and they serve as representative prototypes for different ship-pixels-to-ocean-pixels ratios. Before employing PWTA, the number of clusters, K , must be selected. In the literature of cluster analysis, this problem is referred to as the cluster validation problem [4]. The method we propose here for solving this problem for the particular application at hand is described in the next subsection.

A. Selecting Number of Clusters

Since each pixel in the images of Fig. 3 is either that of a ship or an ocean background, it may be expected that two clusters should be employed, one for each image class. However, due to the presence of speckle and underwater structures, the number of clusters needed to model the feature vectors extracted from the images of Fig. 3 may not necessarily be 2. Speckle adversely affects the assumption that feature vectors extracted from one image class would cluster together in feature space. In this application, the number of clusters should be selected not

to best fit the underlying data distribution as typically done [4], but to yield an image segment that has as many ship pixels and as few ocean pixels as possible. Consequently, the selection of K should be directly linked to the approach used for the determination of the ship segment. The approach we propose here for determining this segment is based on the conjecture that the similarity among clusters corresponding to ocean segments is much higher than the similarity between those clusters and the cluster corresponding to the ship segment. In other words, it is conjectured that among all clusters, the cluster corresponding to ship pixels should be the most distinguishable. This conjecture is not restricted to this particular application. For any application where an image segment of interest is immersed into an image background, it may be conjectured that the similarity among clusters corresponding to the image background is much higher than the similarity between those clusters and the cluster corresponding to the image segment of interest. The similarities between cluster C^j , $j = 1, \dots, K$, and each and every other cluster are computed using the Bhattacharyya distance [21] as follows:

$$B_{jk} = \frac{1}{8} (\mathbf{m}_j - \mathbf{m}_k)^t \left(\frac{\Sigma_j + \Sigma_k}{2} \right)^{-1} (\mathbf{m}_j - \mathbf{m}_k) + \frac{1}{2} \ln \left\{ \left| \frac{\Sigma_j + \Sigma_k}{2} \right| / \sqrt{|\Sigma_j| |\Sigma_k|} \right\} \quad k = 1, \dots, K, k \neq j. \quad (14)$$

The Bhattacharyya distance is selected as the distribution similarity measure since relative to many other measures, it offers a reasonable compromise between simplicity and accuracy [21]. Thus, after computing the average Bhattacharyya distance for each cluster using

$$B_j = \frac{1}{K-1} \sum_{k=1, k \neq j}^K B_{jk}, \quad j = 1, \dots, K \quad (15)$$

the ship segment is selected as the segment corresponding to cluster C_s , where

$$B_s = \max_{j=1, \dots, K} B_j. \quad (16)$$

Based on this approach for determining the ship segment, in our experiments, K was varied between 2 and 9 in increments of 1 and the value selected for performance evaluation was the one that yielded the highest value for B_s . In the next subsection, the implementation details and results for the PWTA utilization in the SAR ship detection application are given.

B. PWTA Detection of Ship Pixels

For each image in Fig. 3, the set \mathcal{T}_L was generated by raster scanning the image using a 3×3 window, and then row-ordering the gray levels of the window pixels, at each window position, into a 9-dimensional feature vector. Thus, as in Fig. 1, the features extracted at each image pixel were the texture information included within a 3×3 window. Experimental observations indicated that for good performance, a small-size window was needed so as to match the small sizes of the collected ship signatures. A PWTA network as that of Fig. 1 was initialized. The means of all units were set to K data samples randomly selected from \mathcal{T}_L . All data were assumed to have a covariance matrix given by $\delta \mathbf{I}$, where \mathbf{I} is the identity matrix. The variance δ was estimated from the data and, then, divided by K to yield an initial variance for all units. The parameters μ_j and N_j were initialized to 1 and 2, respectively, where from (2), N_j must be greater than 1. Also, using (2), $\phi_j(0) = \Gamma\{1.0\}/\Gamma\{0.5\} = 0.5$. After initialization, the PWTA network was iterated through \mathcal{T}_L . For each input vector, the cluster density functions were evaluated using (1), the *a priori* probabilities were estimated at the corresponding pixel position



Fig. 4. PWTA segmentation of the top image of Fig. 3 into five segments. Each of the five segments is identified by a different gray level. Notice that most of the ship pixels are assigned to one segment.

TABLE I
PWTA DETECTION RESULTS FOR THE TOP IMAGE OF FIG. 3 FOR DIFFERENT NUMBER OF CLUSTERS. ABD: AVERAGE BHATTACHARYYA DISTANCE OF THE CLUSTER CORRESPONDING TO THE DETECTED SHIP PIXELS, MSP: MISCLASSIFIED SHIP PIXELS, MOP: MISCLASSIFIED OCEAN PIXELS, TMP: TOTAL MISCLASSIFIED PIXELS, TSP: TOTAL SHIP PIXELS. AS REPORTED, FIVE CLUSTERS YIELDED BOTH THE BEST DETECTION RESULTS AND THE HIGHEST ABD

Number of Clusters	ABD	100 × MSP / TSP	100 × MOP / TSP	100 × TMP / TSP
2	1.04	0.5	49.5	50.0
3	1.85	1.4	16.9	18.3
4	2.25	1.9	14.0	15.9
5	2.60	2.7	11.3	14.0
6	2.55	3.3	11.2	14.5
7	2.37	4.0	11.0	15.0
8	2.29	4.5	10.9	15.4
9	2.31	4.7	10.9	15.6

as described in Section II-B, and then the *a posteriori* probabilities of all clusters were computed using (13). The winning cluster was then probabilistically chosen and its parameters were updated using [5]–[9]. After one epoch, where the presentation of all vectors in \mathcal{T}_L marks the end of an epoch, insignificant variation in the means and variances of all clusters was observed. Each feature vector was then assigned to a cluster using (10). If feature vector \mathbf{x}^i was assigned to cluster C^j , the corresponding image pixel was assigned to the j th image segment. Finally, the ship segment was determined as described in Section III-A. Fig. 4 shows, as an example, the PWTA segmentation of the top image of Fig. 3 into five segments. Table I reports the PWTA detection performance for the same image for different numbers of clusters. The average Bhattacharyya distance for the cluster that corresponded to the detected ship pixels is also reported in this table. As shown, there is a positive correlation between the distance value and detection performance; the higher the value, the better the detection rate. It should be emphasized that misclassification implies ship pixels detected as ocean pixels or vice versa. The PWTA detection of ship pixels for all images of Fig. 3 is shown in Fig. 5. Visual inspection of Figs. 3 and 5 indicates detection performance of good similar quality for both synthetic and real images. The detection results when PWTA as in [12] assumed *a priori* probabilities that were equal and independent of the input feature space are shown in Fig. 6. To save space, Fig. 6 shows only the detection performance for two of the images of Fig. 3. These two images were selected because they serve as representative prototypes for different ship-pixels-to-ocean-pixels ratios. Clearly, making explicit usage of the spatial continuity of ship patterns and ocean background, through the utilization of locally-estimated *a priori* probabilities, greatly enhanced the detection performance. In the next subsection, the PWTA performance is compared to that of four other segmentation techniques.

C. Comparative Analysis of Performance

K -means [5], ML [5], BPN [22], and histogram thresholding were also used to segment the top three images in Fig. 3 so as to detect ship pixels. As previously mentioned, these three images are synthetic and thus they allow a quantitative comparison of performance. K -means

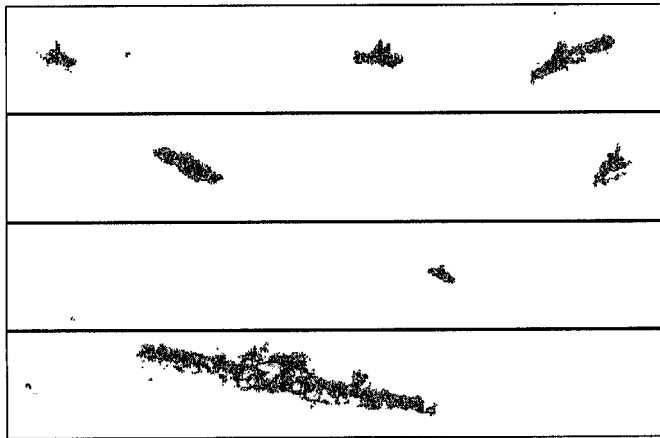


Fig. 5. PWTA detection of ship pixels for the images of Fig. 3 using all steps of Section III-A and Section III-B, including finding the number of clusters, K , that yields the highest average Bhattacharyya distance for the cluster that corresponds to the detected ship pixels.

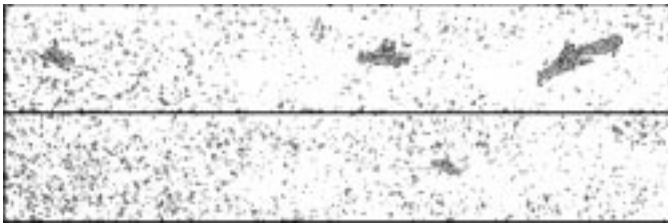


Fig. 6. Detection of ship pixels for the first and third images of Fig. 3 using the original PWTA in [12].

was implemented using a winner-take-all (WTA) network [9] that had the same network layout shown in Fig. 1. ML assumed axes-aligned elliptical Gaussian clusters and was implemented using a MLSC network [23] that also had the same layout shown in Fig. 1. When the number of clusters was greater than four, both K -means and ML frequently encountered the problem of underutilization of clusters, identifiable by the appearance of one-vector clusters, and thus yielded poor segmentation performance. Both K -means and ML took advantage of the spatial continuity of image regions through the same approach introduced for PWTA in Section II-B and their performances were greatly enhanced. BPN was a three-layer network that classified each image pixel as being either a ship pixel or an ocean pixel. It had nine input units and two output units. The input units corresponded to the features extracted within a 3×3 window as done above. The number of hidden units was varied between 2 and 16 in increments of 2. For each number of hidden units, BPN was iteratively trained using a training set of size 2000. The number of hidden units that yielded the best classification accuracy on a validation set of the same size was selected for performance evaluation. The two image classes were represented equally in both the training and validation sets. It is worth mentioning that as in case of PWTA, all of the above techniques needed only one epoch to converge. As for histogram thresholding, it was applied as in [24] after de-speckling the SAR images using the local-statistics noise filtering algorithm in [25]. A direct application of histogram thresholding would certainly fail, since due to the presence of speckle the images in Fig. 3 have unimodal histograms. Fig. 7 shows, as an example, the pixels of the top image of Fig. 3 that were detected as ship pixels in case of using each of the above techniques, whereas Table II quantitatively compares the performance of these techniques to that of PWTA. Inspecting Table II demonstrates the following:

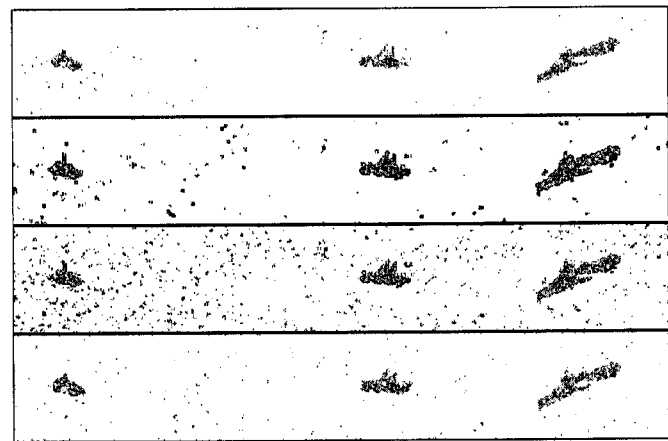


Fig. 7. Using K -means, ML, BPN, and thresholding, from top to bottom, in the detection of the ship pixels of the top image of Fig. 3.

- PWTA significantly outperformed all other techniques when the overall detection accuracy is considered, and it gave an average value of 87.53%. PWTA still performed relatively well when the percentages of misclassified ship pixels and misclassified ocean pixels were examined separately.
- K -means recorded the highest percentage of misclassified ship pixels. This is owing to the presence of speckle and the fact that K -means only adapts a first-order statistic of a cluster [5]. Since K -means minimizes the sum of the squared Euclidean distances between each and every feature vector and the vector's nearest cluster mean, it favored the image class with the larger number of feature vectors and, hence, the number of misclassified ship pixels was larger than the number of misclassified ocean pixels.
- The only supervised segmentation technique utilized, BPN, gave the worst detection performance. This is due to not taking advantage of the continuity of image regions. Since the two image classes were represented equally in the used training and validation sets and since the number of ocean pixels was much larger than the number of ship pixels, the number of misclassified ocean pixels was larger than the number of misclassified ship pixels.
- ML gave the next worst detection performance. This may be attributed to the fact that ML adapts all clusters for each input presentation [5]. Experimental observations demonstrated that adapting all clusters seems to lower cluster discrimination. Low cluster discrimination implies that many of the feature vectors that correspond to one image class were clustered with those corresponding to another. Since the number of ocean pixels was much larger than the number of ship pixels, the number of misclassified ocean pixels was larger than the number of misclassified ship pixels. It is worth remembering that for a given input, PWTA only adapts the winning cluster. Experimental observations indeed demonstrated that PWTA gave, for the cluster corresponded to the detected ship segment, an average Bhattacharyya distance that is significantly higher than that given using ML.
- Histogram thresholding after de-speckling the SAR images yielded a relatively large number of misclassified ship as well as ocean pixels. This is attributed to the fact that it is a one-dimensional clustering approach that does not utilize texture information and only extracts one feature at each pixel position.
- Except for PWTA, all techniques recorded very poor detection performance when applied to Image 3, the third image of Fig. 3, which is characterized by low ship-pixels-to-ocean-pixels ratio. For Image 1 and Image 2, ML and BPN still performed poorly, and the best three performances were obtained using PWTA,

TABLE II
DETECTION RESULTS FOR THE TOP THREE IMAGES OF FIG. 3. MSP: MISCLASSIFIED SHIP PIXELS, MOP: MISCLASSIFIED OCEAN PIXELS, TMP: TOTAL MISCLASSIFIED PIXELS, TSP: TOTAL SHIP PIXELS

Approach	Image 1			Image 2			Image 3		
	$\frac{100 \times \text{MSP}}{\text{TSP}}$	$\frac{100 \times \text{MOP}}{\text{TSP}}$	$\frac{100 \times \text{TMP}}{\text{TSP}}$	$\frac{100 \times \text{MSP}}{\text{TSP}}$	$\frac{100 \times \text{MOP}}{\text{TSP}}$	$\frac{100 \times \text{TMP}}{\text{TSP}}$	$\frac{100 \times \text{MSP}}{\text{TSP}}$	$\frac{100 \times \text{MOP}}{\text{TSP}}$	$\frac{100 \times \text{TMP}}{\text{TSP}}$
PWTA	2.7	11.3	14.0	2.0	9.7	11.7	2.3	9.5	11.8
<i>K</i> -Means	11.3	6.5	17.8	10.8	10.5	21.3	39.0	1648.5	1687.5
ML	2.8	39.7	42.5	2.5	69.5	72.0	3.1	1196.6	1199.7
BPN	1.9	83.3	85.2	1.5	99.1	100.6	2.7	604.9	607.6
Thresholding	5.3	17.2	22.5	4.6	19.2	23.8	1.1	8314.8	8315.9

K-means, and then thresholding. For each of Image 1 and Image 2, let TSP denote the total number of ship pixels and let TMP denote the total number of misclassified pixels. Also, let ε denote the average TMP/TSP calculated over the performances of PWTA, *K*-means, and thresholding. Finally, let γ denote the binomial standard deviation above and below this average evaluated as $\sqrt{\varepsilon(1-\varepsilon)/\text{TSP}}$. Then, for Image 1 and Image 2, $\varepsilon \pm \gamma$ was given by 0.181 ± 0.0111 and 0.189 ± 0.0134 , respectively. Thus, it follows that the improvement obtained using PWTA is statistically significant.

IV. CONCLUSION

Properties desirable for image segmentation have been examined for a recently-developed neural clustering scheme called "probabilistic winner-take-all (PWTA)." One new property that has been demonstrated is that as clustering proceeds PWTA changes the form of the cluster density function such that if a cluster gets adapted frequently, the probability of it being adapted for an input feature vector that is sufficiently far from its mean decreases. This gives the other clusters a better chance of learning this vector. A modification to PWTA to enhance its applicability to image segmentation has also been described. This is the explicit utilization of the spatial continuity of image regions through the inclusion of *a priori* probabilities that depend on the input feature space and that are estimated at each image pixel. Experimental results have been presented. PWTA has been utilized in the segmentation of airborne SAR images of ocean surfaces so as to detect ship signatures, where an approach has been proposed to find a suitable value for the number of clusters required for this application. PWTA performance has also been compared to that of *K*-means, maximum likelihood, backpropagation network and histogram thresholding. It has been demonstrated that PWTA outperformed all techniques and gave high segmentation quality with an average ship detection accuracy of 87.5%. The results presented in this paper are typical and other experimentations led to similar conclusions.

REFERENCES

- [1] G. B. Coleman and H. C. Andrews, "Image segmentation by clustering," *Proc. IEEE*, vol. 67, pp. 773–785, May 1979.
- [2] C. W. Therrien, T. F. Quatieri, and D. E. Dudgeon, "Statistical model-based algorithms for image analysis," *Proc. IEEE*, vol. 74, pp. 532–551, 1986.
- [3] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 403–419, 1994.
- [4] D. A. Langan, J. W. Modestino, and J. Zhang, "Cluster validation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 7, pp. 180–195, Feb. 1998.
- [5] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [6] P. A. Kelly, H. Derin, and K. D. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 36, pp. 1628–1641, Oct. 1988.
- [7] E. Rignot and R. Chellappa, "Segmentation of polarimetric synthetic aperture radar data," *IEEE Trans. Image Processing*, vol. 1, no. 3, pp. 281–300, 1992.
- [8] Y.-F. Wong and E. C. Posner, "A new clustering algorithm applicable to multispectral and polarimetric SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 634–644, May 1993.
- [9] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cogn. Sci.*, vol. 9, pp. 75–112, 1985.
- [10] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance," *Cogn. Sci.*, vol. 11, pp. 23–63, 1987.
- [11] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, "Competitive learning algorithms for vector quantization," *Neural Networks*, vol. 3, pp. 277–290, 1990.
- [12] H. Osman and M. M. Fahmy, "Probabilistic winner-take-all learning algorithm for radial-basis-function neural classifiers," *Neural Comput.*, vol. 6, no. 5, pp. 925–941, 1994.
- [13] M. Rey, J. K. E. Tunaley, J. T. Folinsbee, P. A. Jahans, J. A. Dixon, and M. R. Vant, "Application of radon transform techniques to wake detection in Seasat-A SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 553–560, July 1990.
- [14] M. Rey, J. K. E. Tunaley, and T. Sibbald, "Use of the Dempster-Shafer algorithms for the detection of SAR ship wakes," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 1114–1118, Sept. 1993.
- [15] G. Benelli, A. Garzelli, and A. Mecocci, "Complete processing system that uses fuzzy logic for ship detection in SAR images," *Proc. Inst. Elect. Eng., Radar, Sonar, Navig.*, vol. 141, no. 4, pp. 181–186, 1994.
- [16] K. Eldhuset, "An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 1010–1019, July 1996.
- [17] H. Osman and S. D. Blostein, "New cost function for backpropagation neural networks with applications to SAR imagery classification," in *Proc. SPIE Conf. Automatic Target Recognition IX*, vol. 3718, 1999, pp. 110–117.
- [18] —, "SAR imagery segmentation using probabilistic winner-take-all clustering," in *Proc. SPIE Conf. Algorithms for Synthetic Aperture Radar Imagery III*, vol. 2757, FL, USA, 1996, pp. 217–226.
- [19] —, "SAR image processing using probabilistic winner-take-all learning and artificial neural networks," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. II, Lausanne, Switzerland, 1996, pp. 613–616.
- [20] Technology Service Corporation, *RIG: The Radar Imagery Simulator*, 1996, Version 3.0.
- [21] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1990.
- [22] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
- [23] S. Nowlan, "Maximum likelihood competitive learning," in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1990, pp. 574–582.
- [24] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, 1979.
- [25] J.-S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Comput. Graph. Image Process.*, vol. 17, pp. 24–32, 1981.