Characterizing Natural Backgrounds for Target Detection

Songnian Rong and Bir Bhanu
College of Engineering
University of California
Riverside, California 92521

Abstract
In order to reduce false alarms in automatic target recognition applications, it is important to develop not only the models for man-made targets but also the models of the natural backgrounds. In this paper, we present a self-organizing map (SOM) based approach to construct and maintain a concise and accurate background model by learning from examples. Features used to characterize the natural backgrounds include Gabor transform based features, and localized statistics of pixel values.

1 Introduction
Automatic Target Detection and Recognition (ATD/R) is a challenging application for the general techniques developed by image processing and image understanding communities. This challenge is mainly due to the lack of control of the environment in a typical ATR mission. As a result, there are many variables that can affect the performance of an ATR system. Because a target could appear on many different backgrounds, it is hard to select a set of fixed target features which can be used to detect targets under so many different environmental conditions. To be reliable, an ATR system must possess the intelligence to adapt its detection criteria based on the background encountered. One way to achieve this is to build up models of the background as well as models of the target and use them together to do the detection and recognition. A background model captures the characteristics (localized statistics as discussed in this paper) of a certain background under different environmental conditions. In the following section, we first discuss our approach of using supervised self-organizing maps to construct models for natural backgrounds. Experimental results are shown in Section 3. Section 4 concludes this paper with a summary of the completed work and future research directions.

2 Learning Background Models via Self-Organizing Maps
2.1 The Background Model Bank
In recent years, two streams of approaches have been developed by ATR researchers to characterize the natural background in infrared images. The first one is using heat transfer equations to model the thermal behaviors of different materials. The second stream of approaches focuses on the image features rather than the thermal-physical meaning behind these images. Two important steps in the image feature-based background modeling approach presented in this paper are: (1) to develop effective features that can characterize a certain natural background against man-made target(s) and other different backgrounds, (2) to develop a suitable representation for the background model so that we can control the potential risk of memory explosion while learning the background models from examples. To facilitate the following discussion, we first introduce two terms, feature cell and feature cell size, which will be used through this paper to refer to regions used in feature computation.

Definition: A Feature Cell is a rectangular region within the image from which an image feature is computed. The Feature Cell Size is a measure of this rectangular region.

Since natural backgrounds can occur in a wide variety, background characterization must rely on multiple features. To efficiently use the available features, we need a proper representation to hold all the information together. One way to attack such a problem is to organize all the features into a high dimensional feature vector (i.e., long feature vector) and classify the backgrounds based on the position of the vector in some high dimensional feature space. The other way is based on short feature vectors. The key ideas behind this later scheme are: we need to understand the physical meaning of each feature and put each feature in a group of features that have closely related physical meanings. We investigate the discriminating power of each feature group independently, and build a background model for each such group. Thus, for a given background we will have a collection of simple (i.e., low dimensional) models. We refer to such a collection of models as a Background Model Bank (BMB), and each model in this bank as a BMB member. Figure 1 shows how this BMB would work once it is constructed. Although many papers in the literature have used statistical distributions in their analysis of natural clutters in IR images, there is no strong evidence that thermal natural clutters possess a certain statistical distribution. Instead of artificially assigning a
distribution model to background models, we construct our BMB from real images through a supervised learning process. Since reliable statistical models can only be obtained through analysis of a large population of samples, space and time complexity of algorithms becomes a major concern when selecting learning schemes. In our approach, each BMB member is represented by a self-organizing feature map (SOM). By controlling the size of the SOM, we can easily control the space and time complexity of the learning process. Figure 2 and Figure 3 show the training process of a BMB member and its validity scope respectively. The validity scope of a BMB member is a lookup table indexed by contextual parameters which remembers the performance of a BMB member under certain conditions. Major contextual parameters include sensor types, range and depression angle, weather conditions, etc.

2.2 Supervised Self-organizing Map Algorithm

There are several established learning schemes that can be used to learn the BMB. Learning from example memorizes all the training examples and uses k-nearest-neighbor method to do classification. A potential risk with Learning from example is the memory explosion while the learner gets more experience. Instance based learning (IBL) focuses on maintaining a good classification performance with the least storage of training examples. However, the class descriptor produced by IBL may not be a complete description of the class, i.e. the distribution computed from the memorized examples may differ from the real distribution of the whole training set. Ritter [2] discovered that self-organizing map is a more general version of vector quantization, which utilizes unsupervised competitive learning to find out the representative set. Because of the topology preserving property of SOM, it best describes the distribution of the original examples with a given number of representatives. This distribution information can be very helpful for classification process because a confidence value can be computed from it.

2.2.1 Conventional SOM algorithm

In Kohonen’s SOM algorithm, neurons are arranged into an $N \times N$ array. After initialization of the weight vector $w_i$ of each neuron $i$, the algorithm runs inside a loop which contains two operations:

1. given a training feature vector $x$, search is carried out for the winning neuron $c$ which fulfills

$$||x - w_c|| = \min_{i}||x - w_i||,$$

2. update the weight vectors of the winning neuron $c$ and every neuron within a neighborhood of $c$ according to

$$w_i(t + 1) = \begin{cases} w_i(t) + \alpha(t) (x(t) - w_i(t)) & \text{for } i \in N_c \\ w_i(t) & \text{otherwise} \end{cases}$$

Different strategies can be used to control the learning rate $\alpha(t)$ and to adjust the neighborhood $N_c$ as training goes on. Both parameters should decay with time. In the above algorithm, normally the training process terminates when a pre-selected iteration number has been
reached. The selection of this number is mainly based on experiments. To make the learning process autonomous, i.e. without the need for humans intervention, a metric reflecting the SOM’s ordering is needed so that the algorithm can determine how well the SOM has been trained, and thus determine whether it is time to terminate the learning process.

2.2.2 Disorder Index
Since a properly trained SOM asymptotically converges to the distribution of training examples, the variation of the weight vectors with respect to a fixed number of training iterations will decrease asymptotically. So a measure based on this variation can be used as an index for the ordering of SOM. Let \( d_{mx}(t) \) be the mean square distance between the training vectors and the weight vectors at discrete training time \( t \), we have

\[
d_{mx}(t) = \frac{1}{|S_T|} \sum_{x \in S_T} \left( \frac{1}{|N_c|} \sum_{i \in N_c} ||x - w_i(t)||^2 \right)
\]

where \( S_T \) denotes the training set and \( N_c \) is the set of neurons within the neighborhood of the winning neuron. The Disorder Index (DOI) can then be defined as

\[
DOI = d_{mx}(t + k) - d_{mx}(t)
\]

where \( k \) determines the length of the interval when DOI is evaluated. Recently a more sophisticated metric has been proposed for measuring the disorder of a SOM [4].

2.2.3 Near-miss Injection
When DOI is below a pre-selected threshold the SOM is in a well ordered state, and a conventional SOM algorithm can terminate its learning process at this time. In our approach, at this time the learning process will go into the second stage — refining those ambiguous regions in the SOM by using the near-miss injection algorithm and negative examples. By ambiguous regions we mean regions where features of different classes (e.g. background and man-made target) overlap. The near-miss injection algorithm runs inside a loop which contains two steps:

1. given a negative training vector \( y \), search is made for the “hitting” neuron \( h \) using equation 1.
2. update the weight vectors according to

\[
w_i(t + 1) = \begin{cases} 
    \frac{y(t + 1)}{||y - w_i(t)|| + 1} & \text{for } i \in N_h \\
    \frac{y(t + 1)}{||y - w_i(t)||} & \text{otherwise}
\end{cases}
\]

\[
u = \begin{cases} 
    \frac{y(t) - w_i(t)}{||y - w_i(t)||} & \text{if } ||y - w_i(t)|| \neq 0 \\
    \frac{y(t) - w_i(t)}{||y - w_i(t)||} & \text{if } ||y - w_i(t)|| = 0
\end{cases}
\]

\[
\bar{w}_i = \frac{1}{8} \sum_j w_j, \text{ neuron } j \in 8\text{-neighbor of neuron } i
\]

2.3 Image Features for Characterizing Natural Backgrounds
2.3.1 Gabor transform coefficient features
Recently Gabor decomposition has been used to analyze both synthetic and natural textures [3]. Gabor transform can decompose an input image into base functions which are localized both in spatial and frequency domain. This property is particularly desirable in ATR applications. To compute the discrete Gabor transform of an image, we implemented an algorithm developed by Yao [5] which makes the otherwise complex computation more efficient. Given an image matrix \( I_m \), the Gabor transform of the image can be computed through a matrix multiplication operation

\[
A = G^{-1} I_m (G^T)^{-1}
\]

where matrix \( G \) and its transpose can be pre-calculated based on the selected spatial resolution. Each element of matrix \( A \) is complex and represented by its real part \( a_{rs}^r \) and imaginary part \( a_{rs}^i \). Here \( m \) and \( n \) are indexes in spatial domain, while \( r \) and \( s \) are indexes in frequency domain. The amplitude of the transform is given by

\[
A_{m,n} = \left( (a_{rs}^r)^2 + (a_{rs}^i)^2 \right)^{1/2}
\]

Three feature groups are constructed from the amplitude. The first group consists of three features — \( MA0 \), \( MA1 \) and \( MA2 \), which are the zero, first, and second level mean amplitudes of a feature cell. If \( Am(i) \) denotes the sorted Gabor amplitude of a \( M \times N \) feature cell, \( i = 0, 1, \ldots, MN - 1 \), then

\[
MA0 = \frac{1}{MN} \sum_{i=0}^{T} \hat{A}m(i)
\]

\[
MA1 = \frac{2}{MN} \sum_{i=0}^{T/2} \hat{A}m(i)
\]

\[
MA2 = \frac{4}{MN} \sum_{i=0}^{T/4} \hat{A}m(i)
\]

where \( T = MN - 1 \). The second group consists of two features — the first order moments of \( Am \) with respect to \( w_x \) and \( w_y \) axis.

\[
Mom_x = \sum_{r,s=1}^{M,N} Am(r,s) \cdot r
\]

\[
Mom_y = \sum_{r,s=1}^{M,N} Am(r,s) \cdot s
\]

The third feature group is based on the analysis done by Field [1], which characterizes natural backgrounds by the changing ratio between logarithmic Gabor amplitude and logarithmic spatial frequency.

2.3.2 Amplitude features
The most basic image features are measures of image amplitude in terms of intensity. The mean and standard deviation of the amplitude (denoted as \( M_n \), and \( S_d \)) are computed from feature cells of size \( M \times N \)

\[
M_n = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} Im(i,j)
\]

\[
S_d = \left( \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (Im(i,j) - M_n)^2 \right)^{1/2}
\]
where $I_m(i,j)$ ($i = 0, 1, \ldots, M - 1, j = 0, 1, \ldots, N - 1$) is the intensity matrix of the feature cell.

3 Experimental Results

In this section, we show experimental results of using Gabor transform-based and amplitude features to characterize natural backgrounds. Samples of FLIR and low resolution visible images used in our experiments are shown in Figure 4 where a man-made target is present on top of a natural terrain. To build up the background models, we manually "cut out" the pure background regions from these training images and use them as positive training examples. Figure 5 shows the values of different feature groups computed from IR images (size 200×200) with a 50×50 feature cell size. All the features show a certain degree of discriminating power between man-made targets and natural backgrounds, but the separation of these two classes becomes not 100% distinct when examples exhibiting variety of environmental conditions are used. This calls for the cooperation of multiple feature groups and a measure to resolve the feature overlapping problem in each feature group. Figure 6 shows the result of using our supervised SOM algorithm to construct a BMB member based on the mean and standard deviation feature group.

4 Conclusion

In this paper, we presented an approach that uses self-organizing maps to construct statistical models of natural backgrounds using visible and infrared images. By alternately using positive examples and near-misses in the training phase, the SOM can learn the distribution of feature vectors and refine its boundaries to deal with the feature overlapping problem. How beneficial the background models can be to ATD process depends on two factors, (1) the effectiveness of the selected features, and (2) better understanding of each feature's validity scope with respect to contextual parameters. Our future work will concentrate on: (1) developing new feature groups based on the analysis of local statistics of geometric elements of natural backgrounds. (2) investigating and identifying the most important contextual parameters for each feature group using reinforcement learning methods. (3) extending our approach to other sensors.

5 Acknowledgement

This work was supported by ARPA/AFOSR grant F49620-93-1-0624 and ARPA grant MDA972-93-1-0010. The contents of the information do not necessarily reflect the position or the policy of the U.S. Government.

References


1994 Image Understanding Workshop

Proceedings
Hyatt Regency Hotel
13 - 16 November
Monterey, California
Volume I

Image Understanding Research Program
University of California at Riverside
Host of 1994 Image Understanding Workshop

Sponsored by:
Advanced Research Projects Agency
Software and Intelligent Systems Technology Office
This document contains copies of reports prepared for the ARPA Image Understanding Workshop. Included are Principal Investigator reports and technical results from both the basic and strategic computing programs within ARPA/SISTO sponsored projects and certain technical reports from selected scientists from other organizations.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the United States Government.