BiTS – A Biologically-Inspired Target Screener for Detecting Manmade Objects in Natural Clutter Backgrounds

Mark J. Carlotto, *General Dynamics Advanced Information Systems*

***Abstract*— Motivated by biologically-inspired architectures for video analysis and object recognition, a new single band electro-optical (EO) object detector is described for aerial reconnaissance and surveillance applications. Our bio-inspired target screener (BiTS) uses a bank of Gabor filters to compute a vector of texture features over a range of scales and orientations. The filters are designed to exploit the spatial anisotropy of manmade objects relative to the background. The background, which is assumed to be predominantly natural clutter, is modeled by its global mean and covariance. The Mahalanobis distance measures deviations from the background model on a pixel-by-pixel basis. Possible manmade objects occur at peaks in the distance image. We measured the performance of BiTS on a set of 100 ground-truthed images taken under different operating conditions (resolution, sensor geometry, object spacings, background clutter, etc.) and found its probability of detection (PD) was 12% higher than a RX anomaly detector, with half the number of false alarms at a PD of 80%.**

***Index Terms*— anomaly detection, Gabor filters, texture representation, target screening.**

I. Introduction

The detection of manmade objects in complex natural backgrounds is an important early processing stage in many automatic target recognition (ATR) systems. Although some ATR approaches can be applied to an entire image (e.g., optical correlators), model-based ATRs are usually cued to specific (focus of attention) locations containing possible objects of interest [1]. In synthetic aperture radar, target cues are based on radar cross-section, in thermal infrared, targets may be warmer or colder than the background, and in multispectral or hyperspectral imagery, they are often distinguishable by their spectral response. However, unless strong edges or cast shadows are present, detecting manmade objects in single band (panchromatic) electro-optical (EO) imagery is extremely difficult, and even then, may be confused with trees, rocks, bushes, and other natural features.

Object detection techniques that involve training [2] are in effect coarse ATRs, requiring some explicit characterization of the objects of interest. Our interest here is in techniques that attempt to discover and exploit intrinsic differences between manmade (artificial) objects and natural background features without training. Examples of previous approaches include autoregressive models [3], fractals [4,5] spatial and/or spectral anomaly detection [6,7] and others. These approaches involve detecting possible manmade objects as significant deviations from a statistical model of the background. Where a matched filter can detect instances of a specific object in an image, it cannot find general classes of objects, such as any ground vehicle. On the other hand, anomaly detectors

usually cannot differentiate vehicles from other features having similar size, shape, or spatial frequency content. Such distinctions are usually left to classification and identification stages in an ATR system.

The idea that biology and the study of the brain can provide useful insights into computation and artificial intelligence can be traced back to Weiner's collaboration with McCulloch and Pitts in the 1930's [8]. Thirty years later Hubel and Weisel mapped the primate visual cortex and showed it consists of different kinds of cells. First are simple (S1) cells in layer V1 that respond to localized changes (edges) at particular locations and orientations in the visual field. Outputs from simple cells are combined by complex C1 cells and passed on to the next layer in the cortex (V2) where object recognition is thought to occur. Daugman [9] modeled the spatial frequency response of the S1 cells by Gabor functions. Serre and Riesenhuber [10] developed a set Gabor filters that have been successfully used for object detection [11] and recognition [12].

A model for visual attention proposed by Itti, Koch and Niebur [13] functions much like an object detector. They use multiscale features and center-surround differences that are combined into a saliency map. Objects of interest tend to be located at peaks in the map. Our idea, which is similar, is to use the set of Gabor filters as a basis for anomaly detection – for detecting object regions that are texturally different from the background. The approach, known as BiTS (biologically-inspired target screener), passes the input image through a bank of Gabor filters, which provides a rich vector of texture features over a range of scales and orientations at each pixel location in the image. The background, assumed to be predominantly natural clutter, is modeled by its global mean and covariance. The Mahalanobis distance measures deviations from the background model on a pixel-by-pixel basis. Possible manmade objects whose texture differs from that of the background are detected based on their distance.

Section 2 describes an anomaly detector that uses Serre and Riesenhuber's S1 filters. In Section 3 we show how the performance of the anomaly detector (probability of detection vs. probability of false alarm) can be improved by using an alternative set of filters that better exploit the spatial anisotropy of manmade objects. A method of phase encoding and multiplexing these alternative real-valued filters as complex-valued composite filters [14] is described in Section 4 that significantly reduces the computational complexity of the algorithm. Test results (Section 5) on a set of ground-truthed images shows BiTS outperforms a traditional RX anomaly detector [6].

II. TEXTURE ANONALY DETECTION

Let $F = f(x,y)$ be a single band (panchromatic) input image that we convolve with a set of *K* Gabor filters:

$$g_k(x,y) = \exp\left(-\frac{u_k^2 + \gamma^2 v_k^2}{2\sigma_k^2}\right)\cos\left(\frac{2\pi}{\lambda_k}u_k\right) \tag{1}$$

with $u_k = x\cos\theta_k + y\sin\theta_k$ and $v_k = -x\sin\theta_k + y\cos\theta_k$. The filter bank $\{G_k\}$ can be implemented in the frequency domain using the fast Fourier transformation (FFT), where precomputed FFTs of the Gabor filters are multiplied by the FFT of the input *G*, and the inverse FFT of each product is computed to produce a set of filtered images, $Z_k = F * G_k$. Thus, for K filters, K+1 FFTs are required. Several different filter banks are explored in this paper.

The stack of K filtered images can be viewed as an image of K-element feature vectors, $\mathbf{z}(x,y)$. We conjecture

that the feature vectors of natural textures are close to each other relative to those of artificial textures produced by manmade objects. If the frequency of occurrence of manmade objects is small, than the background statistics can be modeled by the global means and covariances:

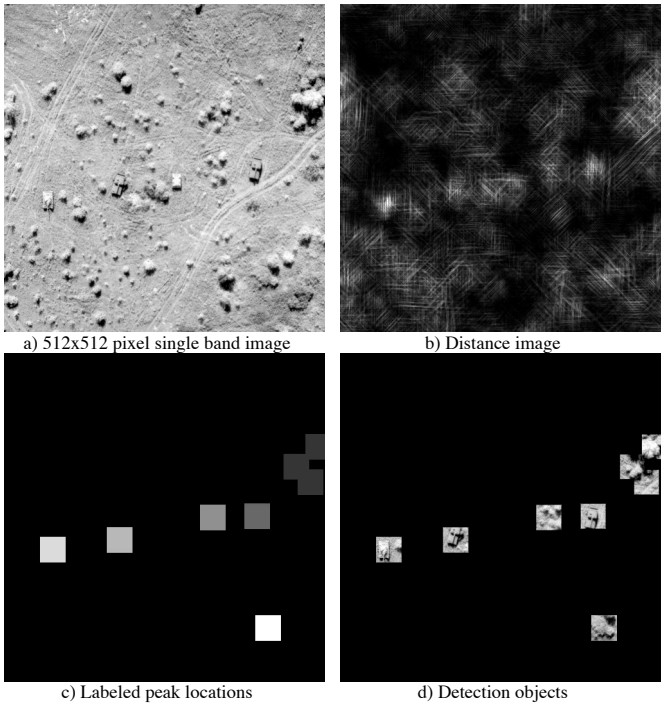
$$\mathbf{m} = \begin{bmatrix} E[Z_1] \\ \vdots \\ E[Z_K] \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} Var[Z_1 Z_1] & \cdots & Var[Z_K Z_1] \\ \vdots & \ddots & \vdots \\ Var[Z_1 Z_K] & \cdots & Var[Z_K Z_K] \end{bmatrix} \tag{2}$$

and the Mahalanobis distance

$$d(x, y) = \left[\mathbf{z}(x, y) - \mathbf{m}\right]^T \mathbf{C}^{-1} \left[\mathbf{z}(x, y) - \mathbf{m}\right] \tag{3}$$

used to measure the deviation of each pixel from the background. If natural textures are close to one another, the distance between background pixels and the background model will be small, while the distance between manmade object pixels and the background model will be large.

We begin with the set of Gabor filters developed by Serre and Riesenhuber [10]. K=64 filters are defined in 16 sizes from 7x7 pixels to 37x37 pixels, each with four orientations, 0°, 45°, 90°, and 135°. Consider a test image containing four military vehicles (Fig. 1a). Bright areas in the Mahalanobis distance image (Fig. 1b) indicate areas that differ from the background texture, which include some of the vehicles as well as small trees, bushes, and vehicle tracks.



a) 512x512 pixel single band image b) Distance image

c) Labeled peak locations d) Detection objects

Fig. 1 Aerial image containing four vehicles. 3 of 4 vehicles detected with ranks 1, 3, and 5 out of 6.

The Mahalanobis distance (3) provides the basis for detection. In theory, candidate object regions can be detected by thresholding the Mahalanobis distance based on a constant false alarm assumption [6], where the threshold

depends on the number of degrees of freedom *K* and the false alarm rate. However, in practice, we find the blurring effect of the filters causes detections to be much larger that the actual size of the objects and so instead use a peak detector with lateral inhibition [15]. Initially a simple detector marks square regions centered at the top *N* peaks whose area is equal to the size of the smallest object we wish to find. The number of detections is a user-specified parameter that depends on the expected density of objects. The top eight detected areas (Fig. 1c) include three of the four vehicles plus 5 false alarms. Fig. 1d shows the actual detected objects.

III. Characteristics of Manmade Objects

The anisotropy (degree of directionality) and rectilinearity (presence of directional components 90º apart) of the power spectrum are useful in discriminating images of natural and artificial features [16]. Methods for computing the anisotropy and rectilinearity from the power spectral density (PSD) are described in Appendix A. Fig. 2 shows four detection chips from Fig. 1 along with their PSD, fractal dimension, fractal model fit (correlation coefficient between log frequency and log power), anisotropy, and rectilinearity. On average vehicles tend to have a higher anisotropy and rectilinearity and lower fractal model fit than natural clutter. Serre and Riesenhuber's filters measure texture in only four directions. Can a different set of filters be used to better distinguish spatially anisotropic targets from clutter?

For image representation, increasing the number of orientations in a Gabor basis improves the accuracy of the representation [17]. Consider an alternative set of filters in four sizes: 7x7, 11x11, 15x15, and 19x19 pixels, each with 16 orientations. Fig. 3a shows the distance image and Fig. 3b the detections. Picking off the first eight detections yields all four vehicles with three false alarms.

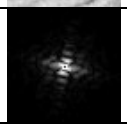
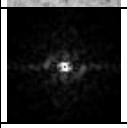
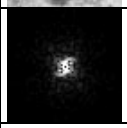
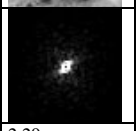
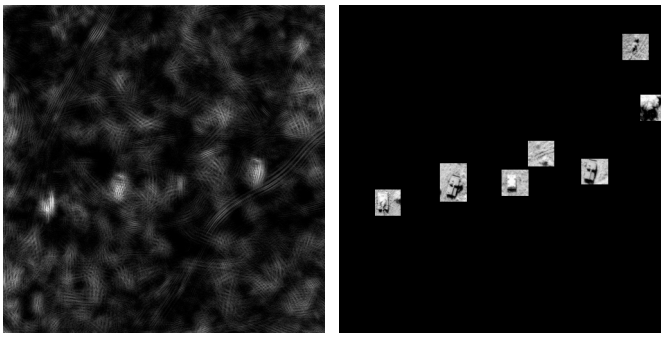
Image chip				
PSD (stretched for display)				
Fractal Dimension	2.27	2.57	2.21	2.29
Fractal Fit (ρ)	0.992	0.984	0.997	0.996
Anisotropy	2.26	3.05	2.57	1.87
Rectlinearity	0.56	0.39	0.01	-0.53

Fig. 2 Comparison of detection chips in terms of features derived from their 2-D power spectral density (PSD).

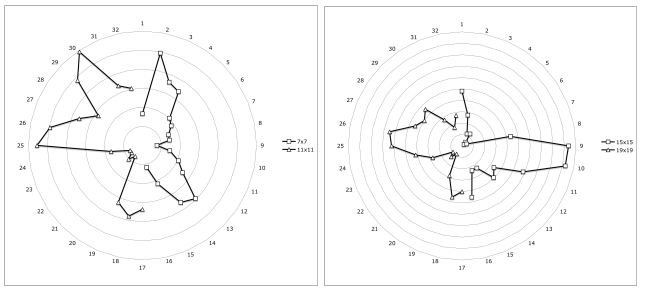
The reason these filters work better than the original set for anomaly detection can best be understood by examining their response around a target and in clutter. Fig. 4a plots the former. Anisotropies 90° apart at the lower spatial frequencies (15x15 and 19x19 windows) are due to the rectangular shape of the vehicle. Those at higher spatial frequencies (7x7 and 11x11) are caused by its internal structure. We contrast this to the response to isotropic clutter (Fig. 4b) where filters respond more or less uniformly in direction. Analysis of plots at other locations suggests that only a subset of the filters around targets tend to respond to produce an anomaly. In general we do not

know what filters will respond and so a large number of filters are required, which increases the computational cost. By analyzing a large number of target and clutter chips it might be possible to find the subset of filters with the best discrimination, although it is likely that any set will not be general, depending on image content and sensor/illumination geometry.

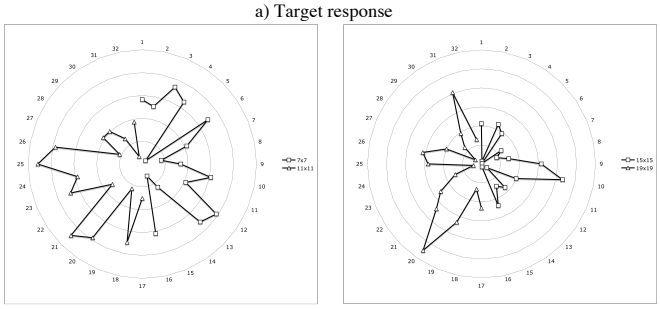


a) Distance image b) Detection objects

Fig. 3 Results using alternative set of Gabor filters. 4 of 4 vehicles detected with ranks 1, 2, 3, and 5 out of 7.



a) Target response



b) Clutter response

Fig.4 Polar plots of filter responses for 7x7 and 11x11 (left) and 15x15 and 19x19 (right). Responses are modulo 180°.

IV. Phase-Multiplexed Gabor Filters

Instead of reducing the number of filters to reduce computational complexity, we combine all of the filters using a complex-valued encoding scheme that runs multiple filters in parallel at different phases. Our combination scheme exploits the tendency of manmade objects to be both spatially anisotropic and non-fractal. Let us make the passband p and directionality q of the Gabor filter explicit:

$$g_{p,q}(x,y) = \exp\left(-\frac{u_q^2 + \gamma^2 v_q^2}{2\sigma_p^2}\right)\cos\left(\frac{2\pi}{\lambda_p}u_q\right) \tag{4}$$

where $u_q = x\cos\theta_q + y\sin\theta_q$ and $v_q = -x\sin\theta_q + y\cos\theta_q$.

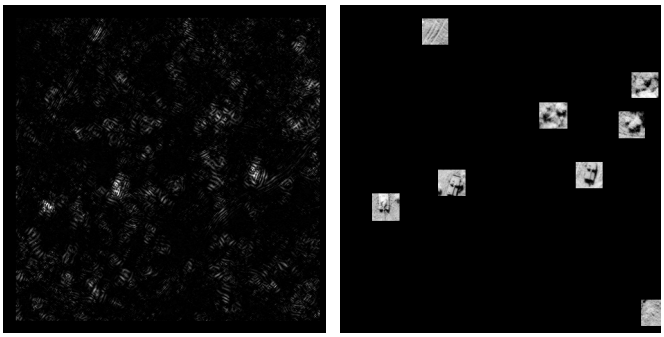
A linear phase coefficient composite filter [14] weights a set of real-valued filters by a linear phase term, representing the sum by a single complex-valued filter. Using this idea we convert the set of 64 filters (4 bands each with 16 orientations) in the previous section into four complex-valued composite filters:

$$\begin{aligned}
h_0(x,y) &= \sum_{q=0}^{Q/4} g_{0,q}(x,y)w(q) + \sum_{q=Q/2}^{3Q/4} g_{1,q}(x,y)w(q) \\
h_1(x,y) &= \sum_{q=Q/4}^{Q/2} g_{0,q}(x,y)w(q) + \sum_{q=3Q/4}^{Q} g_{1,q}(x,y)w(q) \\
h_2(x,y) &= \sum_{q=0}^{Q/4} g_{2,q}(x,y)w(q) + \sum_{q=Q/2}^{3Q/4} g_{3,q}(x,y)w(q) \\
h_3(x,y) &= \sum_{q=Q/4}^{Q/2} g_{2,q}(x,y)w(q) + \sum_{q=3Q/4}^{Q} g_{3,q}(x,y)w(q)
\end{aligned} \tag{5}$$

where the complex weight $w(q) = \exp\left(-j2\pi q / Q\right)$, and Q is the number of directions.

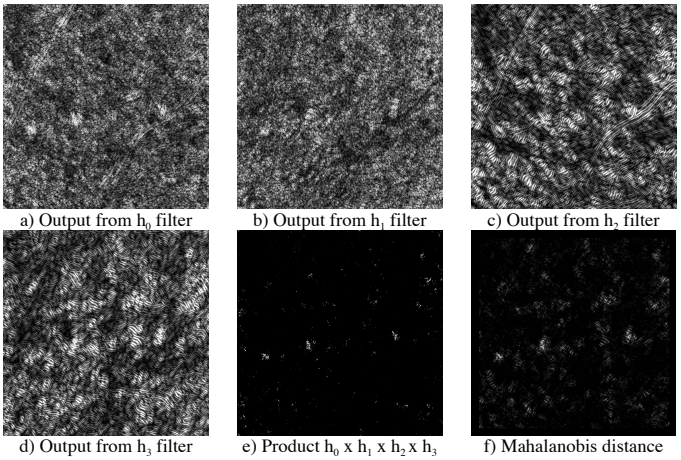
In response to a spatially isotropic fractal texture, Gabor filters at adjacent passbands respond more or less equally in direction resulting in phase cancellation and a small net response. A pattern with an oriented non-fractal texture on the other hand selectively excites only a few of the filters. The LPCCF is designed so the phases add coherently to produce a large net response. Since we do not care which directions respond, we compute the magnitude of the response, discarding the phase, to produce a real-valued 4-vector at each pixel location. Involving only five FFTs per image, the LPCCF implementation reduces the computational complexity by more 10X over the original formulation (1).

Applying this to the test image in Fig. 1 produces a result (Fig. 5) that is better than the S1 filters but worse than the alternative filter set (Fig. 3). In the visual attention model proposed by Itti, Koch and Niebur [13], feature maps are computed from center-surround differences across scales (passbands) in a Gaussian pyramid. Peaks of activity in the feature maps are 'promoted' using a heuristic that is roughly equivalent to multiplying the feature maps. In BiTS, LPCCFs assign opposing phases to filters to form a complex sum of differences of Gabor filters oriented in the same direction but at different passbands. If we multiply the individual LPCCF outputs (Fig. 6a-d) the product (Fig. 6e) is similar to the Mahalanobis distance (Fig. 6f). The visual attention model is thus like an anomaly detector where salience is related to the distance of a pixel from the background model.



a) Distance image b) Detection objects

Fig. 5 Phase-multiplexed Gabor filter anomaly detection results. 3 of 4 vehicles detected with ranks 1, 2, and 3 out of 8.



a) Output from h_0 filter b) Output from h_1 filter c) Output from h_2 filter

d) Output from h_3 filter e) Product h_0 x h_1 x h_2 x h_3 f) Mahalanobis distance

Fig. 6 Breakdown and comparison of phase-coded Gabor filter outputs

V. RESULTS

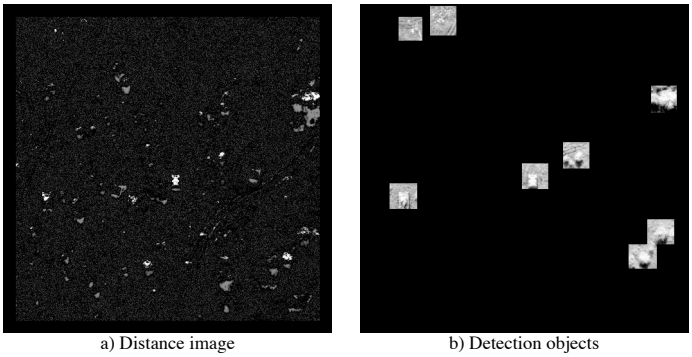
We use the RX algorithm [6] as a baseline for comparison. The single pixel RX detection statistic is

$$d(x,y) = \frac{\left[f(x,y) - m_L(x,y)\right]^2}{\sigma_L^2(x,y)} \tag{6}$$

where $m_L(x,y)$ and $\sigma_L^2(x,y)$ are the local mean and variance within a sliding window. Fig. 7 shows the results for our test image. Where RX is effective in detecting objects that are brighter or darker than the immediate background, it is less effective in detecting textural differences as a review of Figs. 1,3, 5, and 7 reveals.

A set of 100 ground-truthed images containing a variety of vehicles imaged at different resolutions, sensor angles, illumination conditions, and in different backgrounds was used for extended testing. Fig. 8a compares the original set of S1 filters with our alternative filter bank and finds a small improvement in performance, consistent with the earlier example (compare Figs. 1 and 3). Although one would expect the phase multiplexed LPCCF performance to be worse than the non-multiplexed alternative filter bank based on our earlier example, it is roughly comparable over the larger test set (Fig. 8b). Thus with no loss of detection performance the LPCCF implementation is more than 10X faster. Fig. 8c shows BiTS performance against the single pixel RX algorithm. The probability of

detection (PD) was 12% higher than the RX detector, with half the number of false alarms at a PD of 80%.



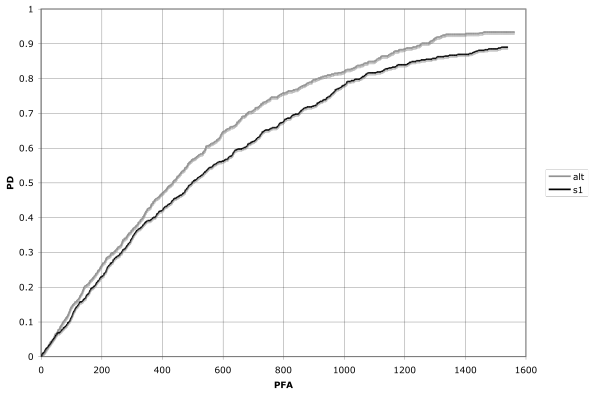
a) Distance image b) Detection objects

Fig. 7 Detection results for RX algorithm. 2 of 4 vehicles detected with ranks 2 and 6 out of 8.

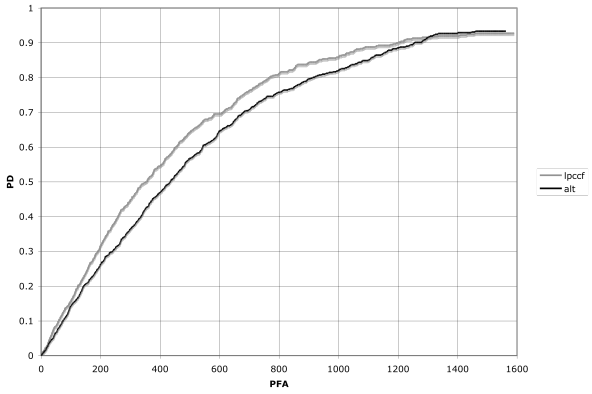
VI. CONCLUSION

The idea of modeling and detecting objects as deviations from the background can be applied to spectral, spatial, textural, and other domains [18]. In this paper we have applied a cluster-based anomaly detection approach to the problem of detecting manmade objects in single band panchromatic EO imagery. The detector uses a bank of Gabor filters to compute a vector of texture features over a range of scales and orientations. The filters exploit the spatial anisotropy of manmade objects relative to the background, which is assumed to be predominantly natural clutter. The Mahalanobis distance measures deviations from the background model on a pixel-by-pixel basis. Possible manmade objects occur at peaks in the distance image.

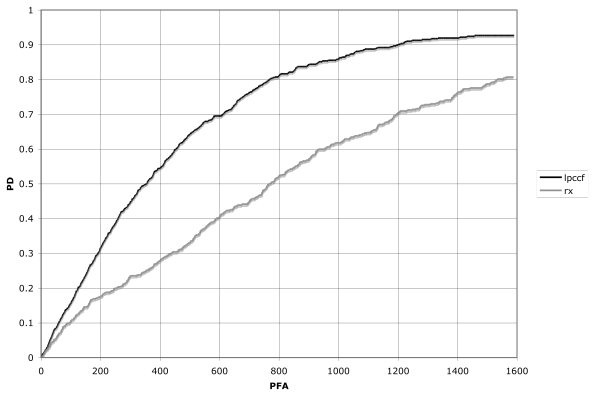
For spectral data, the performance of the anomaly detector depends on the number of clusters [19]. In this paper we have applied BiTS to relatively small images where the background can be modeled reasonably well by a single texture (cluster). Future work will investigate the applicability of the technique on larger images containing multiple background types, including manmade clutter.



a) Original (S1) and alternative filter banks



b) Standard (non multiplexed) and phase multiplexed (LPCCF) implementations of alternative filter bank



c) BiTS (LPCCF) and RX algorithms

Fig. 8 Extended test results on ground-truthed data set

APPENDIX – MEASURING ANISOTROPY AND RECTLINEARITY

Let $S(p,q)$ be the discrete 2-D power spectral density (PSD) of an image chip centered on detection, and $S(\theta)$ be the 1-D angular distribution

$$S(\theta) = \sum_{r} s(p = r\cos\theta, q = r\sin\theta) \tag{A-1}$$

If μ_A and σ_A are the mean and standard deviation of $S(\theta)$ over the angular range $0 \le \theta < 2\pi$, we can define the anisotropy of the 2-D PSD to be

$$A = \frac{\max_{2\pi} S(\theta) - \mu_A}{\sigma_A} \tag{A-2}$$

Images that have directional structure have larger values of A than those that do not. If the direction with the most power

$$\theta_A = \arg\max_{2\pi} S(\theta) \tag{A-3}$$

define

$$\Omega : \theta_A + \frac{\pi}{2} - \delta \le \theta \le \theta_A + \frac{\pi}{2} + \delta \tag{A-4}$$

to be the angular range to search for rectilinear features. For perfectly rectangular features $\delta = 0$. Providing a tolerance ($\delta > 0$) is useful in cases when the images are oblique and rectangular features appear as parallelograms (due to foreshortening). If μ_R and σ_R are the mean and standard deviation, and θ_R is the direction of the peak response within the interval Ω,

$$R = \frac{S(\theta_R) - \mu_R}{\sigma_R} \Bigg/ \frac{S(\theta_A) - \mu_R}{\sigma_R} \tag{A-5}$$

is a measure of the rectilinearity, whose maximum value is one when the second peak is as large as the first. (Measuring the rectilinearity over the interval Ω instead of from zero to 2π reduces the correlation between *A* and *R*.)

REFERENCES

[1] R. Hummel, "Model-based ATR using synthetic aperture radar, "The Record of the IEEE 2000 International Radar Conference, pp. 856-861.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 11-13 Dec. 2001, Vol. 1, pp. 511-518.

[3] C.W. Therrien, T.F. Quatieri and D.D. Dudgeon, "Statistical model-based algorithms for image analysis," Proc. IEEE, Vol. 74, No. 4, April 1986.

[4] M. C. Stein, "Fractal image models and object detection," Proc. SPIE, Vol. 845, pp 293-300, 1987.

[5] T. Peli, "Multiscale fractal theory and object characterization," J. Opt. Society Am. A, Vol. 7, No. 6, June 1990.

[6] I. Reed and X. Yu, "Adaptive multi-band CFAR detection of an optical pattern with unknown spectral distribution," IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. 38, pp 293-305, March 1990.

[7] M.J. Carlotto, "A Cluster-based Approach for Detecting Manmade Objects and Changes in Imagery," IEEE Trans. Geoscience and Remote Sensing, Vol. 43, No. 2, Feb. 2005.

[8] F. Conway and J. Siegelman, Dark Hero of the Information Age, Basic Books, NEW York, 2005.

[9] J.G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," J. Optical Society of America A, 2:1160–1169, 1985.

[10] T. Serre and M. Riesenhuber, "Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex," Technical Report CBCL Paper 239 / AI Memo 2004-017, Massachusetts Institute of Technology, Cambridge, MA, July 2004.

[11] S. Bileschi and L. Wolf. "A Unified System for Object Detection, Texture Recognition, and Context Analysis Based on the Standard Model Feature Set," British Machine Vision Conference (BMVC), 2005.

[12] T. Serre, L. Wolf and T. Poggio, "A new biologically motivated framework for robust object recognition," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2005.

[13] L. Itti, C. Koch and E. Niebur, "Model of Saliency-based Visual Attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11 Nov. 1998.

[14] L. Hassebrook, B.V.K. Vijaya Kumar and L. Hostetler, "Linear phase coefficient composite filters for optical pattern recognition," SPIE Proc. Optical Pattern Recognition, Vol. 1053, pp 218-226 1989.

[15] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," Human Neurobiology, Vol. 4, pp 219-227, 1985.

[16] M.J. Carlotto, "Detecting Patterns of a Technological Intelligence in Remotely-Sensed Imagery," J. British Interplanetary Soc., Vol. 60, pp 28-39, 2007.

[17] T. S. Lee, "Image Representation Using 2D Gabor Wavelets," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.18 No.10, pp 959-971, October 1996.

[18] Mark Carlotto, "Feature-based anomaly detection," SPIE's 2007 Signal Processing, Sensor Fusion, and Target Recognition XV, Orlando FL.

[19] Mark Carlotto, "Automatic clustering based on an information-theoretic approach with application to spectral anomaly detection," SPIE's 2006 Signal Processing, Sensor Fusion, and Target Recognition XV, Orlando FL.