Suppression of subpixel sensor jitter fluctuations using temporal whitening

Steven M. Adler-Golden*, Steven C. Richtsmeier and Robert M. Shroll
Spectral Sciences, Inc. 4 Fourth Avenue, Burlington, MA USA 01803-3304

ABSTRACT

Sensor jitter introduces non-white noise fluctuations in imagery of cluttered scenes. These fluctuations are a major source of interference in the detection of weak time-dependent signals, which may be associated with a subject's appearance, motion, or brightness modulation. Due to the presence of sensor pattern noise and uncertainty in the scene's subpixel spatial structure, standard frame-to-frame registration methods have limited ability to model and remove these fluctuations. A simple temporal whitening approach, applicable to a wide variety of imaging systems, is found to be highly effective for suppressing subpixel jitter effects, leading to dramatic (up to several orders of magnitude) improvement in signal detection ability.

Keywords: jitter, image, detection, signature, anomaly

1. INTRODUCTION

Temporal signatures associated with fluctuating or moving subjects in a scene can be detected in sequences of images from a staring sensor. Such signatures may result from appearance, disappearance or motion of the subject as well as its brightness variation, such as flickering of a flame. A major, and often dominant, source of interference in detecting low-level temporal signatures is image jitter, which may be caused by tracking error, vibration or other motion of the sensor. In addition to blurring the images, jitter-induced fluctuations in the pixel signals obscure the subject's characteristic waveform. One method for mitigating jitter is to perform frame-by-frame image registration. While this is very effective for jitter at the pixel level, sub-pixel registration can be computationally intensive and limited in accuracy by the presence of sensor pattern noise and uncertainty in the sub-pixel background structure. Therefore, there is a need for more efficient and effective means for suppressing sub-pixel jitter effects.

Here we focus on problems in which the signature to be detected is weak and localized and occurs over multiple image frames, in contrast to an impulse event that is observable in two consecutive frames. A treatment of impulse detection in the presence of jitter is presented by Davis *et al.*[1] When the signature is in the same direction over multiple frames (such as an additive light signal), frame-averaging may be applied, resulting in some reduction of the jitter background, but this strategy becomes ineffective as the time-average ("DC") level stabilizes. More general signature waveforms may be detected by averaging "AC" (DC-subtracted quadrature) signals. If the background and subject Power Spectral Densities (PSDs) are dissimilar, some discrimination between the two may be provided by Fourier frequency filtering. However, this scheme is far from optimal, as it fails to distinguish among waveforms with identical frequency content but different phases.

The present investigation was motivated by recognizing that detection of a weak, localized temporal signature in an image sequence in the presence of jitter or other background noise is completely analogous to detection of an anomalous spectral signature in a cluttered hyperspectral image (HSI) "cube". In both cases the data consist of an image sequence in which the subject to be detected maintains its nominal pixel location. In the HSI community, the RX algorithm[2], which is based on data whitening, is considered to be a benchmark method for anomaly detection. The whitening approach is equally applicable to temporal image sequences, and derives from the following premises:

1. Detection is based on a quantity, or metric, that is monotonic in the likelihood that the data vector contains the subject signature relative to the likelihood that it is a pure background vector.

*sag@spectral.com; phone 1 781-273-4770; fax 1 781-270-1161; spectral.com

2. Assuming an unknown signature and additive background, this likelihood ratio is inversely proportional to the multidimensional background probability distribution function (PDF). Therefore the detection threshold (or decision surface in multidimensional space) is an isobar of the PDF.
3. For simplicity, the PDF is represented as a multivariate Gaussian. This reduces to a spherically contoured PDF in a linearly transformed (whitened) coordinate system, in which the data covariance is the unit matrix.
4. Since the spherically contoured PDF is monotonic in the mean-square or root-mean-square amplitude of the whitened data vector, either of these quantities may be taken as the detection metric. The latter is referred to as Mahalanobis distance.

The Gaussian PDF assumption can be relaxed to include more general, elliptically contoured data distributions (e.g., "heavy-tailed"), which also transform to spherically contoured PDFs in the whitened coordinates. This fact may contribute to the success of the RX method with HSI data.

A subspace projection method, which is mathematically related to the current approach, has been used for jitter mitigation since the 1980's.[3,4] Here, an empirically determined number of leading principal component waveforms of the data are identified as jitter and removed. As will be discussed, the whitening and subspace projection approaches may be regarded as employing linear transformations of the data in which different principal component weightings are used.

In this paper, we show how the whitening approach provides dramatic detection improvement in jittering images compared with simple AC signal processing. A simple semi-analytic model for Receiver-Operator Characteristic (ROC) curves is developed based on readily derived image and signature properties. Then the whitening method is validated using data examples containing anomalous signatures, and the detection enhancement associated with whitening is shown to be semi-quantitatively explained by the simple model.

2. SCENE SIMULATIONS

The synthetic jittering image sequence, containing typically between 25 and 200 images, is generated starting from a higher-resolution scene. The first, optional step is to convolve the scene with an instrument optical blur function. Next, the image is sequentially shifted along both x and y directions by a chosen integer number of pixels, determined via Gaussian random sampling. The output frames are then generated by averaging a small number of adjacent shifted images to model the time integration during the frames and resampling the results to a coarser (10x or more) spatial resolution, such that the original image shift increments correspond to 0.1 pixel or less at the output resolution. Finally, a small amount of Gaussian random noise (typically 100 ppm of the peak intensity), representing detector noise, is added to the resampled image. The Gaussian shift-sampling method yields an inverse-frequency-squared dependence of the jitter PSD, similar to what is measured in many real systems.

For evaluation of detection performance using embedded signatures, a signature waveform is generated, and scaled values of this waveform at each time point are added to ten selected pixels in the original high-resolution scene at the start of the simulation process. We have used two different types of waveform, a sine wave and a random wave. The latter signature is generated using the jitter waveform routine, and therefore has the same general PSD behavior as the jitter.

The data simulation and processing have been applied to a variety of input background images and signature waveforms. An example image is the green channel of a 2536 x 1536-pixel photograph of the Boston skyline (courtesy of Tudor Hulubei, http://hulubei.net/tudor). The first frame of the resampled output image (251 x 151 pixels) is shown in Fig. 1 (left). There is a maximum jitter displacement of ½ pixel in both x- and y-directions between any two images in the sequence. With this jitter level the ten signature-containing pixels embedded in the input image yield ten output spots comprising a total of 35 pixels with partial signatures, shown in Fig. 1 at right. To study the effect of different signal-to-noise levels, the embedded signatures in the left-hand-column were scaled to be 50 times brighter than those in the right-hand column.

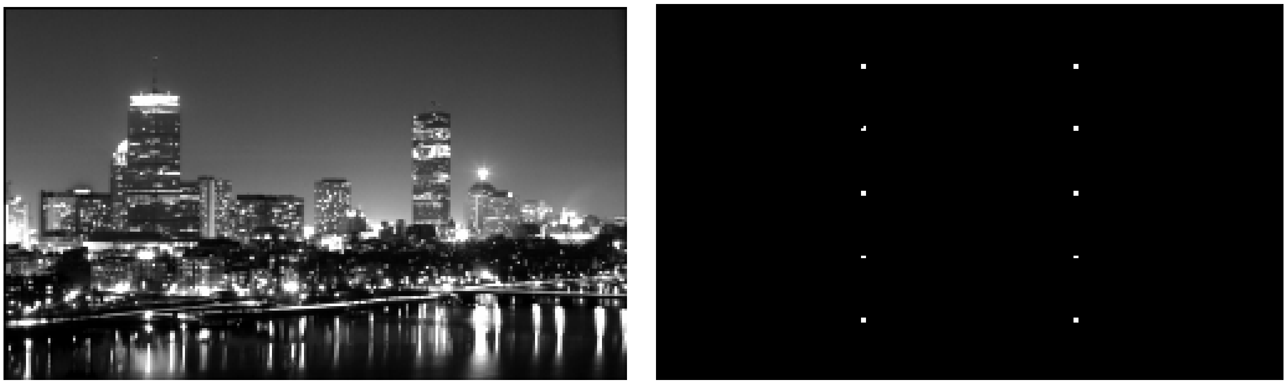

Fig. 1. Output background image (left) of Boston skyline and signature location mask (right). Embedded signatures were 50 times brighter in the left-hand column of spots than in the right-hand column.

3. DETECTION MODEL

This section describes the data processing steps and develops an approximate model for the detection problem that allows efficient construction of ROC curves and affords insight into detection performance. A comparison of the approximate model results with embedded-signature simulations is given in the next section.

We consider a sequence of n image frames, denoting the pixel waveforms as a set of column vectors $x_p = [x_{p1}, x_{p2}, ... x_{pn}]$. Assuming further that the waveforms all have zero mean (i.e., are DC-subtracted), the AC power in each pixel is the quadrature (squared) signal averaged over frames,

$$X_p = 1/n \, \Sigma_j \, x_{pj}^{\,2} \tag{1}$$

If a subject signal waveform s is added to the background signal, the resulting AC power is given by

$$X_s = 1/n \, \Sigma_j \, (x_{pj}+s)^2 = 1/n \, \Sigma_j \, (x_{pj}^{\,2}+s^2+2x_{pj}s) \tag{2}$$

With large n, x_{pj} and s are likely to be nearly orthogonal, hence the third term in Eq. (2) is small and the AC background and subject quadrature signals can be modeled as additive. This makes it straightforward to construct ROC curves, which are plots of the detection probability PD versus the false alarm probability PFA. If the subject is detected by setting a threshold T on the AC power, then

$$\mathrm{PFA} = 1 - \mathrm{CDF}(T) \tag{3}$$

$$\mathrm{PD} = 1 - \mathrm{CDF}(T - S) \tag{4}$$

where CDF(X) is the cumulative distribution function associated with X.

The whitening approach, illustrated schematically in Fig. 2, recognizes that the n-dimensional PDF associated with the background pixel waveforms x_p contain a high degree of correlation. Here the optimal threshold for detection is an isobar of the PDF rather than a value of the AC power. Presuming that the PDF can be linearly transformed, or whitened, to a spherical Gaussian distribution, the quantity to be thresholded is the frame-averaged AC power (quadrature signal) in the whitened coordinate system, which we denote X'. The standard whitening transform may be written as

$$x' = ED^{-1/2}E^T x \tag{5}$$

Here $D^{-1/2} = diag(d_1^{-1/2}, d_2^{-1/2}, ... \, d_r^{-1/2})$, d_i being the eigenvalues of the covariance matrix, which has rank r, and E^T is the orthogonal matrix whose rows are the corresponding covariance matrix eigenvectors. Using the whitened data x', Eqs. (1) through (4) may be evaluated to compute X'_p, X'_s, and the corresponding ROC curves.

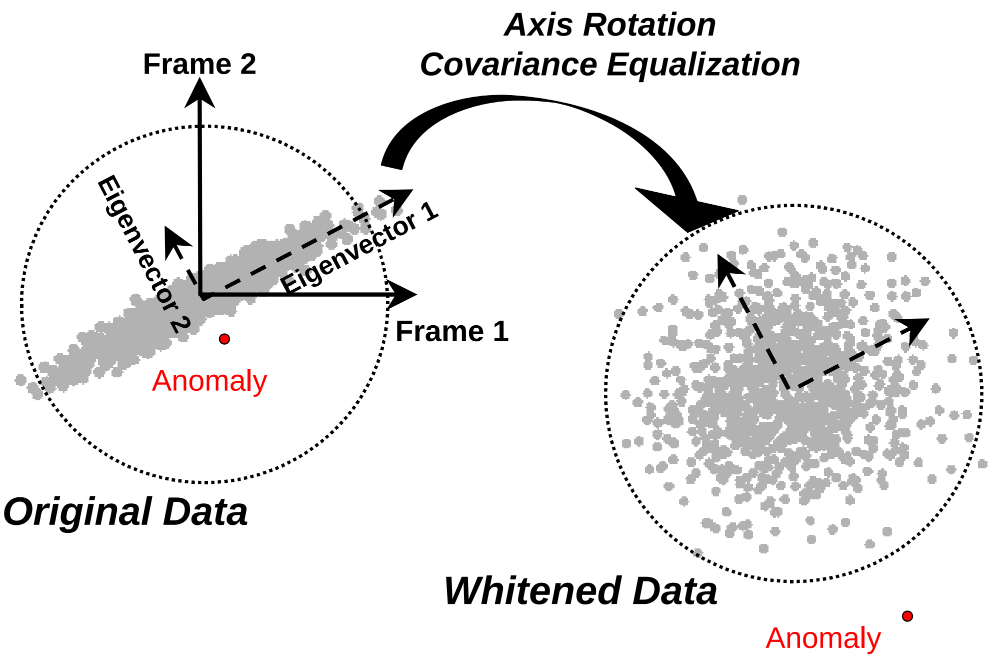

Fig. 2. Schematic comparison of unwhitened and whitened anomaly detection in two data dimensions (frames). Dotted line denotes a threshold on the quadrature signal, represented as radial distance. Superior detection is achieved with the whitened data.

The quantity X' is analogous to the RX metric used in hyperspectral imagery, although there may be some differences arising from the choice of data means. The covariance matrix calculation formally requires referencing each data frame to its own mean, so that the mean is a spectrum or waveform. With an image time sequence this computation step may be unimportant, as frame-to-frame variation of the mean with a staring sensor is small and the pixel DC levels are already removed in our treatment. Here the data covariance matrix may be approximated by the correlation matrix. A more important difference relative to typical HSI data analysis is that here we have chosen to let the mean signal level vary with the pixel, thus defining a mean image rather than a mean waveform. As a result, the maximum rank r of the correlation or covariance matrix, which defines the number of significant eigenvectors and eigenvalues, turns out to be one less than the number of frames n.

We have also processed time sequence imagery without DC subtraction, in which case the whitening is performed with the covariance matrix, which is typically found to have rank $r = n$. This is the RX implementation typically used with HSI data. Here the whitened signal power may be written simply as

$$1/n\, X'^T X' = 1/n\, X^T E D^{-1} E^T X = 1/n\, X^T K^{-1} X \qquad (6)$$

where K^{-1} is the covariance matrix inverse.

Some insight into the performance improvement afforded by whitening can be gained by considering the detection from a signal-to-noise standpoint. The whitened to unwhitened AC power ratios for the signal and background are given by, respectively, X'_s/X_s and $r/\Sigma_j d_j$, where j runs from 1 to r. We may therefore define an effective whitening gain factor, or quadrature signal-to-noise improvement, as

$$G^2 = 1/r\, \Sigma_j d_j\, X'_s/X_s \qquad (7)$$

G itself is in linear signal units (i.e., radiance). As shown in the next section, the G value allows one to crudely estimate the effect of whitening on the detection ROC curve. Individual values contain some variation associated with the specific jitter x- and y-waveforms, which depend on the random sampled variables. Typical values with our synthetic imagery are of order 100.

Eq. (7) may be compared with a simple statistical estimate of G^2 that assumes uniform projection of the subject waveform onto the correlation matrix eigenvectors, i.e., $X'_s/X_s = 1/r\, \Sigma_j 1/d_j$:

$$G^2_{\text{stat}} = 1/r^2 \; \Sigma_j d_j \; \Sigma_j 1/d_j \qquad (8)$$

We have found that G estimates from this formula, which contains only background information, are typically within a factor of two of those from Eq. (7). Accordingly, detection performance is found to depend much more on the AC power of the signature, X_s, than on its specific waveform.

Finally, noting that the quantity $E^T x$ in Eq. (5) is the principal component expansion of the data, the data whitening step amounts to principal component extraction, weighting the components via the diagonal $D^{-1/2}$ matrix, and re-assembling the data. Thus, the subspace projection method described by Hulsmann and Barry [1985] and Kirk and Donofrio [1996] amounts to replacing some number of the leading (smallest) elements of $D^{-1/2}$ with zero and setting the remaining elements to a constant. This procedure is completely analogous to subspace projection in HSI applications and has similar advantages and disadvantages versus whitening. Zeroing a matrix element may provide a detection advantage if along that associated dimension the data distribution has an especially heavy tail. On the other hand, subspace projection is less than optimal for a multivariate Gaussian PDF, and furthermore some trial and error is needed to determine the optimum number of elements to set to zero.

4. SIMULATION RESULTS

A jittering 50-frame image sequence was generated using the Boston skyline image with embedded random-wave signatures. The RMS signal-to-background ratio in the Fig. 1 signature mask pixels ranged from 0.002 to 0.05 for the dim signatures and from 0.06 to 4 for the bright signatures. The unwhitened and whitened quadrature outputs X and X' from the data analysis are shown in Fig. 3. Since the values cover a very wide dynamic range, the gray scale was chosen to be linear in the fourth root of the value. The row of bright signature spots is seen to be at a similar level to the background in the unwhitened image, consistent with signal-to-background ratios of order unity. The spots are much brighter than the background in the whitened image. The dim spots are not visible here; however, the topmost dim spot, which is against the sky, is faintly discernable with a more sensitive display.

Using the Fig. 1 signature mask, ROC curves for detection were constructed (Fig. 4). There is enormous improvement in detection performance for the brighter signatures using the whitening procedure. Detection of the brighter signatures without whitening is roughly similar to detection of the dimmer signatures with whitening, indicating that the effective whitening gain factor G is comparable to the 50:1 signature brightness ratio. This is consistent with Eq. (7), which yields a G value of 57. Whitening also improves detection of the dimmer signatures, but much less dramatically, since large numbers of false detections become unavoidable with any data transformation as the signatures approach the center of the data distribution.

Fig. 3. Output detection images for random-wave signatures: unwhitened AC signal X at left, whitened signal X' at right.

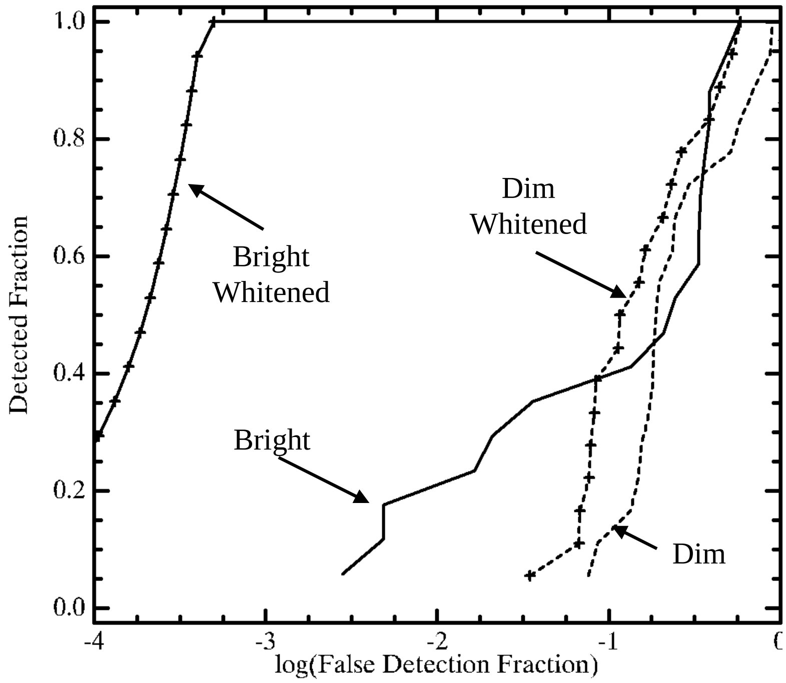

Fig. 4. ROC curves for detection of dim (dashed line) and bright (solid line) signatures with and without whitening.

5. REAL DATA EXAMPLE

An application to real data is shown in Fig. 5, in which a hand-held digital camera was used to record a short video clip of an indoor scene. A modem box sits on the floor, and the small light dot on the box is a blinking LED. The video clip was downsampled to 320 x 240 pixels to reduce the jitter in pixel units, and a 1-s (30-frame) segment of the red channel was selected.

Results from our analysis are shown in Fig. 6. The unwhitened AC image at left shows large contributions at high-contrast edges due to the camera motion. In the whitened image at right these features are suppressed, and the LED emerges as a strong, localized temporal anomaly. The statistical G factor for this image is around 30. It is important to note that here the camera motion includes a rotational component, which is missing in our simulated imagery.

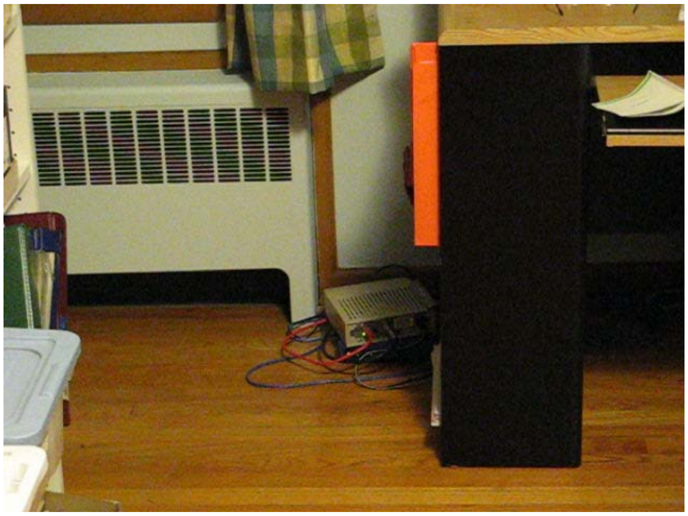

Fig. 5. Indoor scene with modem.

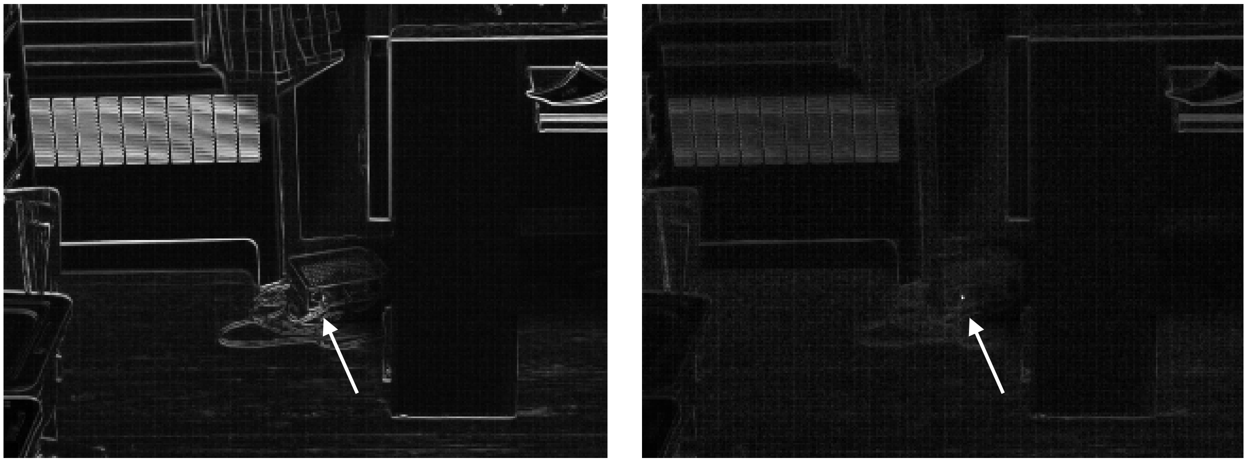

Fig. 6. Outputs from the indoor scene analysis: unwhitened AC signal X at left, whitened signal X' at right. Modem LED is indicated by the arrows.

6. CONCLUSIONS

A data-whitening method, analogous to the RX anomaly detection method used with hyperspectral imagery, has been found to be highly effective for discriminating against jitter in image sequences to enable detection of weak and often sub-pixel time-dependent signatures. In evaluations using both synthetic and real imagery, the typical performance improvement versus simple AC signal processing corresponded to an effective one to two orders of magnitude enhancement in signature brightness. Potential applications of the method are varied, and include detection of moving vehicles and personnel as well as time-varying sources such as warning lights and fires.

It should be noted that data whitening is advantageous for detection of known as well as unknown signatures. Data whitening may be regarded as the basis of the matched filter detector, which can be written as a projection of whitened data onto the whitened signature. A widely used detection technique for hyperspectral imagery, matched filtering is similarly applicable to image sequences in which an independent measurement of the signature waveform is available.

REFERENCES

1. J.C. Davis, J.P. Helferty and J.J. Lisowski, "Target detection in non-Gaussian clutter noise," *Proc. IEEE Aerospace Conf.*, Vol. 5, pp. 2073-2084, March 8-15, 2003.
2. I.S. Reed, and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing*, 38, pp. 1760-1770, 1990.
3. J.D. Hulsmann, and Barry, P.E., "An Eigenvector Procedure for Eliminating Line-Of-Sight Jitter Induced Noise from Staring Mosaic Sensors," presented at *Nineteenth Asilomar Conference on Circuits, Systems and Computers*," Naval Postgraduate School, Monterey CA, 6-8 November 1985.
4. J.A. Kirk, and M. Donofrio, "Principal component background suppression," *Proc. IEEE Aerospace Applic. Conf.*, Vol. 3, pp. 105-119, 3-10 February 3-10 1996.