Inherent limitation of digital imagery: spatial-phase vacillations and the ambiguity function

R Barry Johnson and Kaveh Heidary

1 Department of Electrical Engineering and Computer Science, Alabama A&M University, 4900 Meridian St., Normal, AL 35762, United States of America
2 Department of Physics, Chemistry, and Mathematics, Alabama A&M University, 4900 Meridian St., Normal, AL 35762, United States of America

E-mail: barry.johnson@aamu.edu

Keywords: digital image processing, image enhancement, image recognition devices, image recognition, algorithms and filters, image reconstruction techniques, pattern recognition, target tracking

Abstract
The spatial relationship of a scene image upon a pixilated focal-plane-array can temporally change due to scene or camera motion, microjitter of camera line-of-sight, etc. It was found that degradation caused by subpixel spatial-phase vacillations (SPV) upon the performance of matched filters was unexpectedly significant. Subpixel spatial-phase vacillations can cause degradations in matched filter performance to detect desired objects and discriminate other objects. SPV appear to be an inherent limitation of digital imagery when processed using matched filter methodology and can negatively impact the performance of systems. Mitigation of this degradation was found to be possible by utilizing one of several matched filter constructions such as a multi-filter enhanced matched filter (EMF) bank and a single-filter EMF. A significant conclusion of this investigation is that, for automatic target recognition applications, improved overall task performance should be realized by the use of an EMF bank. Dramatic reduction in computational resources for image matching multispectral imagery was accomplished by spectrally-collapsing multispectral imagery to form pseudoimages with the spectral information appearing as a texture in a grayscale image. The pattern recognition ambiguity function, which sets the fundamental performance limit of an image processing system, is introduced.

1. Introduction

Prior to the 1970’s, imagery was typically produced using film or by scanning a linear detector array (comprising a single detector up to over 100 detectors) to form an electronic image. With the advent of the CCD detector array in the 1970’s, and subsequently the widespread availability of digital sensors and cameras, and digital image processing, significant investigations of the technology and its applications have occurred. The basic theory of sampled-data imaging systems has been well studied, as has the effect of oversampling, microscanning, microdithering for pixilated imagers [1–12]. For the sake of argument, it is assumed that the fill factor of most focal plane arrays (FPA) is near 100%. An appropriate manner to view microscanned generated imagery is to consider it as spatially-phase enhanced imagery. This is true whether or not the dither is used to (i) fill in gaps in a sparsely-populated FPA (shifts by one or more integer detector-widths) or (ii) provide sampling at sub-pixel intervals. A simple way to think about the generation of the sampled image is to imagine a single detector smoothly moved over the entire image formed by the optical system. The resulting contiguous image is the convolution of the optical image and the detector, i.e. if the image spatial distribution is represented by \(f(x,y)\) and the detector spatial response by \(d(x,y)\), the resultant convolution \(g(x,y)\) is given in the general case by

\[
g(x,y) = f(x,y) * d(x,y) = \int \int f(\alpha,\beta) d(x - \alpha, y - \beta) \, d\alpha \, d\beta
\]
where the integration is over all space. For this technique to be valid, it is required that the content of the image be temporally stable during the time needed to perform the multiple image samples and that the image remains spatially stable with respect to the optical system during this time period. This convolution sheet represents all of the information obtainable from the system and this concept of the convolution sheet will be involved later in the discussion of spatial-phase vacillations (SPV). If now an FPA having N by M pixels is used, the output image comprises N \cdot M values of \( g(x,y) \), i.e. the image generated is a discrete sampling of the convolution sheet. If the FPA is now shifted with respect to the scene, or conversely, by a subpixel amount, a new N \cdot M output image is formed using the new samples of the convolution sheet. Examination of this new image will reveal that it is slightly different from the preceding frame. This movement of the FPA can be considered spatial-phase shifting. By moving the FPA in discrete subpixel amounts and interleaving the data to correspond to the position on the convolution sheet, a notably improved image will be observed. The image geometric spatial-resolution has not been improved, but the image presentation provides more information as a consequence of the spatial phase shifting. Generally, for a well-designed imaging sensor, three to four subpixel shifts in both directions are adequate to achieve quite enhanced imagery; increasing the number of subpixel shifts further will not noticeably improve the observable information content of the displayed image because the optical system is spatial-frequency band-limited. Figure 1 illustrates the preceding concept where a hypothetical sensor generates 400 low-resolution images by methodically forming low-resolution pixels from 20 \times 20 blocks of pixels from the original image (figure 1(a)). The subpixel shift is 1/20th of a low-resolution pixel. A typical low-resolution image is shown in figure 1(b). Taking all 400 low-resolution images and suitably interdigitating them, the resulting spatially-phase enhanced image is shown in figure 1(c). The low-resolution images exhibit aliasing although it is difficult to observe in figure 1(b). The effect of combining the 400 low-resolution images significantly mitigates aliasing since the resulting spatial sampling frequency is significantly higher than the Nyquist frequency of the basic FPA (figure 1(b)).

Two important utilizations of digital imagery are automatic target recognition (ATR) and target tracking [14–17]. The imagery can be generated by one or more visible monochrome, visible color, near infrared, mid-wave infrared, long-wave infrared, mm-wave, hyperspectral, etc sensors. The predominant approach to accomplish the ATR and target tracking functions is to use an appropriate bank of matched filters. This can become a very onerous task when there are numerous training images and even more so when the images are multispectral or hyperspectral. In order to mitigate a portion of the computational burden, a hyperspace angle map (HAM) method to form spectrally-collapsed pseudoimagery from n-channel multispectral imagery, previously developed by the authors [17], will be discussed in section 2. Considering that the predominance of digital imagery does not include the advantages of microscanning/oversampling, the inconsistency of the autocorrelation of a training image with other supposedly identical images with each having an unknown subpixel shift of the target with respect to the camera FPA caused by microjitter of camera line-of-sight, etc were investigated. The principal objective of this paper is to explore, at the basic research level, the inherent limitation that spatial-phase vacillations impose upon single frames of digital imagery when processed using matched filter methodology. To do so, it was assumed that the typical vagaries of noise, non-uniformity of FPA responsivity, etc and temporal or spatial perturbation of the light field traversing from the scene/object to the FPA are at such a level as to be ignorable. It is further assumed that the image incident on the FPA suffers the same subpixel shift for all pixels in the FPA; however, the amount of subpixel shift for a given image is fundamentally random which means that a plethora of possible images can be generated by a subpixel-scale movement of the image with respect to the FPA. Such movement can be considered as a phase
shift in the spatial frequency domain (see equation (1)). Affects caused by atmospheric turbulence and the like were ignored because they can cause subpixel shifts over the FPA that vary stochastically amongst the pixels and degrade image processing performance beyond that of the SPV uniform phase shifts over the FPA. In this investigation, only a single image of the possible images was processed by the HAM and enhanced matched filter bank (EMFB) as explained in sections 2 and 4. The peak cross-correlations of this image with a large number of other captured images taken from the aforementioned plethora of possible images were determined to demonstrate that there exists a nontrivial variation between these possible images which creates an inherent limitation of the imaging system to be able to discern one image from another, e.g. the ability to detect a subtle spatial difference between say object A and object B that are remarkably similar to each other. The reality and impact of the spatial-phase vacillations (SPV) concept, introduced by the authors, upon the performance of ATR, target tracking, etc systems employing matched filters were a further focus of this study and are discussed in section 4 in which it is shown that spatial-phase vacillations appear to be an inherent limitation of digital imagery when processed using matched filter methodology. In section 5, the concept of Pattern Recognition Ambiguity Function (PRAF) is introduced by the authors. How the SPV and variation in the distance to the object are incorporated into the PRAF is explained and it is argued that the PRAF sets a fundamental limit to the performance capability of an image processing system.

2. Spectrally-collapsed pseudoimage

Although a large matched filter bank can be constructed to account for each spectral band of each object in the training set of images, an alternative effective method to dramatically compress the matched filter bank is to form pseudoimages by spectrally-collapsing the multispectral data cubes. The methodology used to form the pseudoimages is to determine the hyperspace angles between the chromatic vector \( \vec{a} \) of the reference pixel (typically user designated) and the chromatic vectors \( \vec{b} \) of the surrounding pixels. These angles form a reference ‘pseudoimage’ comprising the aforementioned hyperspace angles having values of 0 to \( \pi/2 \). The chromatic vector elements comprise the signal values associated with the multispectral bands of the hyperspace data cube. The spectrally-collapsed pseudoimage comprises the (nonreversible) spectral information appearing as a texture in a grayscale image. This pseudoimage is designated as a Hyperspace Angle Map (HAM) that was introduced in [17]. Example pseudoimages are presented in section 4. The \( n \)-channel hyperspace angle between \( \vec{a} \) and \( \vec{b}_{i,j} \) is given by

\[
\theta_{i,j} = \arccos \left( \frac{\vec{a} \cdot \vec{b}_{i,j}}{|\vec{a}| |\vec{b}_{i,j}|} \right) = \arccos \left( \hat{a} \cdot \hat{b}_{i,j} \right)
\]  

(2)

where \( i \) and \( j \) are the pixel position indices. There are several approaches to comparing the multispectral data cube being generated by the sensors to the training set of hyperspace angle maps to determine the best match for target recognition. The first approach, and followed in this investigation, is to treat it as a matched filter problem. An alternative approach uses a dynamically-adaptive temporal-domain minimum-variance method which is believed to be more tolerant to statistical variations of the multispectral images forming the HAMs. Dynamic tracking of an object can be readily achieved by using HAM difference maps [17]. A temporally-adaptive spatial-spectral filter for the HAMs, based on an extension of work on the enhanced matched filter bank (EMFB), is utilized for this investigation and is presented in the next Section.

3. Enhanced matched filter bank

Given a set of object-class training images, an ordered set of distortion-tolerant filters are synthesized that can classify new renditions of the target robustly. The set of synthesized filters and the respective thresholds constitute the enhanced matched filter bank [18–23]. The performance of the EMFB defined in terms of its detection and false-alarm characteristics approaches that of the corresponding full-set matched filter bank (MFB) which comprises the entire population of training images. It has been shown through examples that an EMFB comprised of a relatively small number of enhanced matched filters (EMF) outperforms, in terms of detection and false-alarm performance, MFBs consisting of a far greater number of matched filters (MF) [24]. The computational complexity of processing a typical test image with an EMF is identical to that of a MF. The substantial reduction of the number of filters makes EMFB a suitable candidate for applications where processing speed is an important design factor.

The EMFB comprises a set of filter-threshold pairs. The threshold of each EMF is adjusted in accordance to the user-specified relaxation parameter \( T_n = (1 - 0.01\gamma)\Omega_n \), where \( \Omega_n \) and \( T_n \) denote, respectively, the computed and adjusted threshold values for the \( n \)th EMF, and \( \gamma \) is the relaxation parameter. Increasing the
relaxation parameter results in lowering the threshold values for all EMFs and a more permissive EMFB which in turn results in higher detection and false-alarm rates. The input test image is applied to each EMF. It is declared Target if its peak cross-correlation with respect to one or more EMF is greater than or equal to the respective threshold; otherwise it is classified as non-Target. It is noted that increasing the relaxation parameter results in lowering the threshold values for all EMFs and a more permissive EMFB which in turn results in higher detection and false-alarm rates. Also, as with the basic matched filter, the construction of the EMFB does not consider any data other than that associated with the specific class under consideration.

An EMF is determined using a weighted combination of its constituent training matched filters (TMFs). The domain of effectiveness of an EMF is typically much greater than that of any of the TMFs; however, it has been demonstrated that the EMF provides essentially the same performance as its constituent TMFs while rejecting those not in its effectiveness domain [24].

4. Spatial-phase vacillations

In contrast to the orderly microscanning involved with enhanced spatial sampling frequency and aliasing mitigation, most digital cameras simply have the fundamental FPA to produce the digitally sampled image. By whatever mechanism, the spatial relationship between the FPA and the image formed by the optical system can temporally change by subpixel distances between the frame acquisitions in an unorderly manner. It was observed in the example shown in figure 1 that the spatial frequencies comprising figure 1(a) are significantly higher than the Nyquist frequency associated with figure 1(b) and the shifted low-resolution images are quite different from one another. Although this example has orderly spatial-phase shifts that were used to produce figure 1(c), the impact of spatial-phase perturbations of shifted low-resolution images will be explored later in this section.

Equation (1) shows that the signal generated by each pixel in a FPA is given by \( g(x_i, y_j) \) where the array \([x_i, y_j]\) are the spatial coordinates of the pixel centers. The geometric relationship between the FPA and the image formed by the optical system is illustrated in figure 2 by the solid white grid. Should a spatial-phase shift occur, the target image upon the FPA moves by some subpixel amount as depicted by the solid yellow grid in figure 2. This, in general, results in a change in the observed signal from each pixel of the FPA. The subpixel shift shown in figure 2 retains a common portion of the signals (blue hatched area) from the two positions of the FPA, a portion of the former pixel signal is subtracted (red area), and portions of the signals from two adjacent pixels are added (green area). Consequently, a variety of similar yet observably different images can be created by subpixel spatial-phase shifts. Assuming that the spatial frequency content of the optical image exceeds the Nyquist sampling frequency of the FPA, the resulting subpixel spatial-phase shifted FPA-produced images will contain aliased power from higher spatial frequencies which may result in artifact-laden images. These artifacts, resulting from subpixel spatial-phase shifts, are the fundamental origin of the signal vacillations and why the cross-correlation of a non-perturbed image with its subpixel spatial-phase shifted images can be significantly less than unity. It should also be recognized that the detector transfer function is of the form sinc\((x,y)\) and progressively attenuates its response as the spatial frequency increases [25]. Once the detector cutoff frequency is greater than or equal to the cutoff frequency of the optics, no aliasing or subpixel spatial-phase shift vacillations will occur [26].
A meaningful manner to investigate the impact of subpixel spatial-phase perturbations is through the utilization of matched or correlation filters. To accomplish this examination, it is necessary to have a high-resolution image to serve as 'ground truth.' Such an image can then be transformed into multiple low-resolution images. Figure 3(a) shows RGB photographs of five people, numbered 1 through 5, having both (i) dramatically different spectral and spatial characteristics and (ii) distinct similarities in some cases. From each of Face-1 through Face-5, a corresponding set of one-hundred low-resolution images were formed where each pixel in the low-resolution images comprises 100 pixels (10 pixels by 10 pixels) from the high-resolution images. Figure 3(b) shows a representative low-resolution image for each of the five faces. As explained in section 2, n-channel spectral images can be spectrally-collapsed to form grayscale pseudoimages which are hyperspace angle maps (HAM). The HAM for each image contained in figure 4 is presented in figures 4(a) and (b).

To assess the impact of the fluctuation in these images, the quasiautocorrelation of the reference or base low-resolution HAM is computed with respect to the 100 HAMs resulting from the 1/10th subpixel shifts. The base HAM is given by \( h(x_i, y_j) \) where the \( i \) and \( j \) are integers divisible by 10 in accordance with equation (1) where \( h(x, y) \) is the hyperspace angle map. Figure 5(a) shows the quasiautocorrelation of low-resolution hyperspace angle maps of Face-1 with values ranging from unity to 0.67. The cross-correlation of the reference low-resolution hyperspace angle map of Face-1 with all 100 HAMs of Face-2 is shown in figure 5(b) with the cross-correlation values ranging from 0.31 to 0.35. Although the orientation of these two faces is similar, they are easily differentiated using the matched filter; however, the 0.04 variation in the cross-correlation is due to the subpixel vacillations. If now all 100 of the HAMs of Face-1 are cross-correlated with the 100 HAMs of Face-2, the cross-correlation then ranges from 0.31 to 0.43 which is three times the deviation of just considering the Face-1 reference HAM. In a like manner, if cross-correlations of all Face-1 low-resolution image HAMs with all the Face-3, Face-4, and Face-5 HAMs are examined, the range of values are 0.18–0.26, 0.20–0.30, and 0.16–0.22, respectively. It is evident that all five multispectral faces are easily differentiated from one another by using matched filters on their HAMs, but the most important point to recognize is that subpixel scale vacillations in the image-FPA spatial relationship can cause operationally significant fluctuations in the correlation using the traditional matched filter process. For example, in an ATR task it might be required to differentiate between two vehicles that have very similar characteristics.
where the cross-correlation between their reference images is say 0.85 and system noise is not a factor. If now the SPV of either vehicle causes the minimum autocorrelation to be less than 0.85, then it is evident that the false alarm rate may be non-trivial and unacceptable.

Is there a method to diminish the impact of SPV upon the differentiation between two vehicles that have very similar characteristics? The answer is yes with at least three possible approaches; however, each approach has a cost associated with it in terms of computation and training-set data acquisition. These approaches are a bank of matched filters, an enhanced matched filter, and the ‘Enhanced Matched Filter—Lite’ (EMF–L) which will be explained presently. Consider first the bank of matched filters with the idea of selecting the minimum number of matched filters necessary to achieve a specified minimum quasiautocorrelation value, say 0.95 for example. To identify the set of filters requires that images are available for a plethora of subpixel spatial-phase shifts. Figure 6(a) shows the peak cross-correlation of the Face-1 reference image (1,1) with respect to all 100 Face-1 spatial-phase shifted low-resolution images. The two abscissae specify the location of the spatial-phase shifted images \((i, j)\). It is evident that only those 16 images that lie above the dashed line will have a cross-correlation \(\geq 0.95\). Using the (5,6) Face-1 image as the reference image, 52 images that lie above the dashed line will have a cross-correlation \(\geq 0.95\); however, there is one common image. In selecting the minimum number of matched filters, it should be recognized that some of the filters will have one or more spatial-phase shifted low-resolution images in common which should be expected. In addition, examination of figure 6 illustrates that if the cross-correlation threshold is reduced to say 0.90, then using the (5,6) Face-1 image as the reference image results in only nine of the spatial-phase shifted low-resolution images not being detected. Most likely a few more matched filters will be needed to have full detection at this threshold setting. Nevertheless, it is necessary to have a dense sampling of the subpixel image space in order to be certain the decision threshold will always be exceeded. As the threshold is lowered, in general the essential sampling density of the subpixel image space will become less.
Figure 7. Peak cross-correlation of the Face-1 EMF with respect to one-hundred Face-1 spatial-phase shifted low-resolution images.

An enhanced matched filter bank (EMFB) comprising a single filter was generated using all 100 of the spatially-phased low-resolution images (SPLRI) for Face-1. Figure 7 presents the peak cross-correlation of the Face-1 EMF with respect to all one-hundred Face-1 spatial-phase shifted low-resolution images. The maximum and minimum peak cross-correlation is near unity and 0.87, respectively. The cyclic nature of the peak cross-correlation shown in figure 7 is an artifact of the ordering selection of the abscissa positions \((i, j)\) of the spatial-phase shifted images. It is noted that if the EMFB generating algorithm had been allowed to produce two or three matched filters, the minimum peak cross-correlation observed from the EMFB would be greater than the 0.87 observed for the single-filter enhanced matched filter bank.

Now consider the situation where a matched filter is constructed from a set of randomly selected spatially-phased low-resolution images (SPLRI), which is denoted as an Enhanced Matched Filter–Lite (EMF–L), since it is constructed by simply summing the selected SPLRI and dividing by the number of selected images. The computational effort required to make an EMF–L is significantly less than that needed to construct the EMFB. Indeed, the EMF–L and the EMFB provide the same results if the EMF–L and the single-filter EMFB use all 100 of the SPLRI. However, if the EMFB generator is allowed to create two or more matched filters to form its bank, then the performance will improve and the generating algorithm likely will not use all 100 SPLRI in the construction of the filters. Figure 8(a) shows the peak cross-correlation of the EMF–L constructed from 10 randomly selected SPLRI for Face-1 with all 100 of the Face-1 SPLRI. The maximum peak cross-correlation is 0.99 rather than unity and the minimum peak cross-correlation has increased from 0.67 (see figure 5(a)) to 0.81 which is still less than 0.87 for the single-filter EMFB (see figure 7). This demonstratively illustrates that the minimum peak cross-correlation performance of either the EMFB or EMF–L, when considering spatial-phase vacillations, is superior to the more traditional construction of the matched filter from a single image.

Increasing the minimum peak intra-face cross-correlation for Face-n is beneficial in differentiating Face-n from the other faces and other images particularly as the signal-to-noise in the images degrades. The intention is to maximize the difference between the minimum peak intra-face cross-correlation of Face-n and the maximum peak inter-face cross-correlation with the other faces. For example, figure 8(b) shows the peak cross-correlation of the EMF–L constructed from 10 randomly selected SPLRI for Face-1 with all 100 of the Face-2 SPLRI. The range of maximum peak cross-correlation is 0.37–0.40. In a like manner, the range of maximum peak cross-correlation the same EMF–L with all 100 of the Face-3 SPLRI is 0.18–0.25 (see figure 8(c)). A similar significant difference between the Face-1 EMF–L minimum peak cross-correlation and the maximum peak cross-correlation for Face-4 and Face-5 is observed. It is further noted that the cross-correlation values vary just by few percent when the ten randomly-selected images for the EMF–L are changed. Figure 9 shows the peak cross-correlation for two Face-1 EMF–L, each constructed from different sets of ten randomly-selected SPLRI, with respect to one-hundred Face-1 spatial-phase shifted low-resolution images. EMF–L–1 represents the first filter construction and EMF–L–89 represents the 89th filter construction of the two-hundred random filter constructions. The minimum peak cross-correlation for EMF–L–1 is 0.82 and is 0.75 for EMF–L–89.

Table 1 presents the minimum inter-face peak cross-correlations between Face-1 and the other four faces for the several processes. This is a significantly more exhaustive study of the minimum peak cross-correlations between the different faces. The first row in table 1, denoted i, contains the minimum and maximum values of the 100 000-minimum peak cross-correlations between each of the one-hundred Face-1
Figure 8. Peak cross-correlation range of values for the EMF–L constructed from 10 randomly selected SPLRI for Face-1 with all 100 of the (a) Face-1 SPLRI, (b) Face-2 SPLRI, and (c) Face-3 SPLRI. The range of values shown in the parentheses are for the peak cross-correlation of the single SPLRI (1,1) with all 100 of the Face-1, Face-2, and Face-3 SPLRI.

Figure 9. Peak cross-correlation of two typical Face-1 EMF–L with respect to one-hundred Face-1 spatial-phase shifted low-resolution images.

SPLRI and all one-hundred SPLRI for each Face-2, Face-3, Face-4, and Face-5. The single-filter EMF computed for Face-1 was used to compute the minimum peak cross-correlations, presented in the table 1 row denoted ii, for all one-hundred SPLRI for each Face-2, Face-3, Face-4, and Face-5. Next, a single-filter EMF was constructed for each Face-2, Face-3, Face-4, and Face-5 and is denoted in table 1, row iii, as Face-n EMF. The values shown on the row are the minimum and maximum values of the minimum peak cross-correlations between each of the one-hundred Face-1 SPLRI and the respective Face-n EMF. Row iv in table 1 provides the results when two-hundred randomly selected EMF–L for Face-1 are cross-correlated with all one-hundred SPLRI for each Face-2, Face-3, Face-4, and Face-5. It is noted that number of possible EMF-L for Face-1 is huge. And lastly, the final row in table 1 tabulates the minimum peak cross-correlation values of one-hundred Face-1 SPLRI and the two-hundred each EMF–L constructed for Face-2, Face-3, Face-4, and Face-5. Examination of the data in the table indicates that the maximum inter-face peak cross-correlation never exceeds 0.5 and is mostly below 0.3. Since the minimum peak cross-correlation of Face-1 with any of the matching processes discussed exceeds 0.7, it is evident that Face-1 is easily discriminated from the other four faces. Examination of the inter-face peak cross-correlations between Face-n and the other four faces resulted in quite similar discrimination performance.
Table 1. Minimum inter-face peak cross-correlations between Face-1 and the other four faces.

| Process          | Face-2 min max | Face-3 min max | Face-4 min max | Face-5 min max |
|------------------|----------------|----------------|----------------|----------------|
| i. SPLRI         | 0.3123 0.4709  | 0.1867 0.2926  | 0.2114 0.3142  | 0.1671 0.2393  |
| ii. Face-1 EMF   | 0.4029 0.4465  | 0.2284 0.2872  | 0.2641 0.2952  | 0.1972 0.2378  |
| iii. Face-n EMF  | 0.3416 0.4590  | 0.2367 0.2806  | 0.2280 0.3081  | 0.1943 0.2373  |
| iv. Face-1 EMF–L | 0.3727 0.4615  | 0.2166 0.2923  | 0.2465 0.3107  | 0.1835 0.2394  |
| v. Face-n EMF–L  | 0.3309 0.4684  | 0.2208 0.2888  | 0.2232 0.3126  | 0.1860 0.2398  |

Figure 10. Effect of number of SPLRIs used in the construction of the EMF–L on the minimum peak-cross-correlation of the Face-1 EMF–L with respect to all one-hundred Face-1 SPLRIs.

An interesting question to consider is how the minimum peak-cross-correlation of the Face-1 EMF–L, with respect to all one-hundred Face-1 SPLRI, varies as the number of randomly selected SPLRI increases. Figure 10 presents the variation in terms of the worst case, the median, and mean of the minimum peak-cross-correlation. When a single SPLRI is taken, the worst-case minimum peak-cross-correlation is 0.67, for 10 SPLRI it is 0.745, and progressively rises to 0.87 when all 100 SPLRI are used to construct the EMF–L. As mentioned previously, the EMF–L using all 100 SPLRI yields the same performance as obtained from the EMFB. In this particular case, there is marginal advantage in using more than about a third of the SPLRI to construct the EMF–L if worst-case minimum peak-cross-correlation is the criterion or to use more than a few SPLRI if mean or median minimum peak-cross-correlation is the criterion. The EMF–L is likely to be well suited for practical ‘people tracking in a crowd’ applications since the target-image—FPA spatial phase relationship will in general be vacillating due to varying target motion alone.\(^{17}\)

Now assume that the geometric relationship between target image and the FPA has a slight dynamic spatial instability. Spatial phase errors will occur that will appear as SPV. If the sensor system invokes microscanning to spatially-enhanced images as described in section 1, how much image degradation is caused by subpixel microscanned positions being slightly perturbed? The results of numerous simulations exploring this situation suggest that relatively trivial degradation to spatially-enhanced images occurs even for up to about 30% perturbation of the expected microscanned positions. This observation is reasonably consistent with the finding that impressive performance of the EMF–L can be achieved by using only a small number of randomly located subpixel samples.

5. Pattern recognition ambiguity function

It is evident that the spatial-phase vacillations discussed in the prior sections have shown unexpected large variations in the output of the matched filter as the image moves about spatially over the area of a pixel. This implies that a specific matched filter designed for a specific object can potentially yield ambiguous results should this matched filter process signals from other generally similar, yet distinctly different, objects that yield outputs equal to or greater than the minimum SPV output of this matched filter. By this point, it should be clear that SPV is a fundamental limitation to the detection and classification of digital imagery. But how can knowledge of the SPV be utilized to determine the ultimate performance limitation of an image processing system? In the remainder of this section, the Pattern Recognition Ambiguity Function (PRAF) is introduced, how the SPV is incorporated into it is explained, and that the PRAF sets a fundamental limit to the performance capability of an image processing system.
Changing the range (distance) to an object, where \( R_0 \) is the reference range, has the effect of scaling the image size as the input to the matched filter. Such a change in scale will result in a decrease in the output of this matched filter which is also subject to additional vagaries due to the SPV. Consequently, a change in image size with concomitant SPV can also yield ambiguous results. For simplicity, the object is considered to always be in focus.

The term *ambiguity function* is often associated with radar systems to represent the response of a matched filter designed for a specific signal and its doppler-frequency-shifted signals. Dr Merrill Skolnik stated that one should not be distracted by trying to understand why the *ambiguity function* is described by the ambiguous use of the term ‘ambiguity’ [27]. For pattern recognition, the ‘doppler’ axis is replaced by range and the ‘range’ axis is replaced by the spatial phase vacillations (SPV). The ordinate is therefore in effect the filter mismatch. By setting a mismatched threshold, the probability of detection of the target can be determined as a function of SPV and range. By plotting the occurrences of non-target filter outputs that do not exceed the threshold, a measure of the FAR can be determined as a function of range.

In the preceding sections, the SPV behavior was shown to be dependent upon the selection of the spatial-phase coordinates within the pixel. The SPV behavior can be represented by a two-dimensional sheet referenced to the aforementioned selected coordinates. A performance metric can be defined as the volume under the sheet where the filter output is normalized such that when the SPV is nil, the metric equals unity, i.e. the pixel is assumed square and the area is normalized, and the peak quasicorrelation is unity. Let \( V(x,y,R) \) designate the volume under the sheet where \( x \) and \( y \) are the coordinates within the pixel and \( R \) is the range.

Consider now the construction of a three-dimensional ambiguity plot. Let the abscissae represent (i) the location of the matched filter’s spatial-phase coordinates within the pixel and (ii) the relative range given by \( R-R_0 \). The ordinate value is unity minus the volume under the SPV sheet, which is also impacted by the relative range. By defining the ordinate (\( z \)-axis) in this manner, the ‘no ambiguity case’ shows as zero on the plot and non-zero (positive) values are a measure of the potential ambiguity.

The pattern recognition ambiguity function plot can be used to locate the best filter spatial-phase coordinates within the pixel for say \( R_0 \) by selecting the abscissa coordinate having the lowest value. Now consider the question ‘What is the best filter spatial-phase coordinates within the pixel that provides the best overall performance (lowest ambiguity) over a choice of ranges?’ The assumptions are that the matched filter is established for \( R_0 \) and the user can select the evaluation ranges. For any SPV abscissa coordinate, the area under the curve in the range axis is a measure of the potential ambiguity. More specifically, this ambiguity can be ‘selected-range’ normalized in the following manner.

\[
A_{\text{range}}(x,y,R_1,R_2) = 1 - \frac{\int_{R_1}^{R_2} \int V(x,y,R)dR}{\int_{R_1}^{R_2} dR}
\]

where \( R_1 \) and \( R_2 \) are the extents of the range under consideration. The value of \( A_{\text{range}}(x,y,R_1,R_2) \) can vary from zero (no ambiguity) to unity (completely ambiguous). Clearly, the selection of the \((x,y)\) having the lowest value of \( A_{\text{range}} \) is desired.

Another possibly useful plot is to use the same abscissa for \((x,y)\), the delta range \((R_2 - R_0)\) along the other abscissa, and \( A \) along the ordinate. Examination of this plot should provide guidance of the useful ranges for a given ambiguity criterion.

A variety of PRAF profiles were computed for many different images of diverse classes of objects, and an example low-resolution image of a vehicle is shown in figure 11 and the corresponding PRAF profile presented in figure 12 is representative of the overall effect with a range variation of \( \pm 10\% \). The axes represent range variation, spatial phase vacillations, and correlation mismatch. Notice that the ambiguity goes to zero when the range change is zero and the spatial phase vacillation is either zero or unity. The ambiguity mismatch establishes the correlation limitation for a given correlation mismatch for high SNR and low background clutter. The pattern recognition ambiguity function should be used as a metric in the determination of data training sets for the EMFB.

Now consider the ambiguity between classes of objects when processed by a specific matched filter designed for a particular object, say \( O_{\text{ref}} \), and at range \( R_0 \).

\[
\delta A_{\text{ref}}(x,y,R_0) = \begin{cases} 
 A_{\text{ref}}(x,y,R_0) - A_{\text{ref}}(x,y,R_0) & \text{for } 0 > A_{\text{ref}}(x,y,R_0) - A_{\text{ref}}(x,y,R_0) \\
 0 & \text{otherwise}
\end{cases}
\]
where \( i \) indicates the \( i \)th object and \( iCA_{\text{ref}} \) is the cross-ambiguity between the \( \text{ref} \) and \( i \)th objects [28]. When the value of \( iCA_{\text{ref}} \) is non-zero, there exists a possibility that the \( i \)th object can be classified as the \( \text{ref} \) object. A plot can also be created showing the delta range in a like manner to prior plot.

The initial ambiguity plot \( A_{\text{ref}}(x, y, R_0) \) can be useful in estimating the number of spatially-phase shifted images that are needed to construct an EMF-L. Examination of the ambiguity plot can indicate regions where the ambiguity is low and the relative displacement from other low regions. Ideally, the images selected for the EMF-L should be taken from these areas. By determining the number of entries along the spatial-phase vacillation axis that exceeds a threshold value of \( A_{\text{ref}} \), one can gain insight into the potential performance of an EMF-L and the number of randomly taken images needed. The percentage of the number of entries that do NOT exceed the threshold is basically the probability that random samples will be acceptable for the EMF-L in achieving anticipated performance.

If one integrates \( A_{\text{ref}} \) along the spatial-phase vacillation axis, the resulting value gives a measure of the potential ambiguity. If now \( \alpha \) represents the spatial-phase vacillation axis, then the probability \( P \) of obtaining an ambiguous result is given by

\[
P(R_0) = 1 - \frac{\int_{\alpha_1}^{\alpha_2} A(\alpha, R_0) d\alpha}{\int_{\alpha_1}^{\alpha_2} d\alpha}.
\]

(5)

If now one selects two or more images along the spatial-phase axis to construct an EMF-L, then the full set of images can be processed by this EMF-L to generate another plot showing the ambiguity. In general, \( P_{\text{EMF-L}} \) should be less than for \( P \). It is anticipated that there should be a subset selection of \( \alpha \) values that can form an EMF-L which yields the minimum ambiguity probability.
6. Multiple-frames spatial-phase vacillations

In the preceding discussion, the SPV and PRAF were used to investigate single images. Actual systems often acquire a number of frames of the same image; however, the image formed on the FPA can move slightly between frames for a variety of reasons such as scene motion, camera motion, microjitter of camera line-of-sight, etc. It has been demonstrated that SPV and PRAF impact the PD and FAR of matched filters. It is anticipated that having multiple frames with subpixel movement between them can be used to improve the PD and FAR performance of the matched filters. The degree of this improvement is yet to be determined as is the appropriate method to affect such improvement.

A related hyperspectral image-analysis application exists where an anomalous object, having spatial extent smaller than a sensor detector footprint, is entirely confined to a single image pixel. When multiple frames are acquired, SPV likely will impact system performance. The technique developed for a single frame is an unsupervised learning algorithm that examines each image pixel in the context of its immediate neighborhood without any a priori knowledge about the spatial and spectral characteristics of the expected background or potential anomalies [29]. A powerful procedure suitable for real-time applications has been developed, by the authors, for the single-image frame case where the anomalous source is located within a pixel. Matched filters are not suitable for addressing this problem since the anomaly has no observable spatial extent or features. When SPV occur as a consequence of acquiring multiple frames, it is possible that the anomalous source can move slightly (less than a pixel footprint) and 'contaminate' two, three, or four pixels adjacent to one another. The amount of contamination of these pixels likely will change between frames and appear similar to modulation of these pixels as well as changes in the spectral content of the contaminated pixels. The questions to be investigated in this case include (i) the impact upon detection of an anomalous source, (ii) its spatial location, (iii) how the PD and FAR change, and (iv) development of a method to mitigate the PD and FAR degradation caused by SPV.

7. Conclusions

This investigation has illustrated that spatial-phase vacillations (SPV) can cause remarkable fluctuations in image cross-correlation values and is the ultimate limiting factor in image matching. Considering that the predominance of digital imagery does not include the advantages of microscanning, many systems suffer SPV-induced degradations of the autocorrelation of a training image with other supposedly identical images resulting from perhaps subpixel shift of the target with respect to the sensor FPA, microjitter of sensor line-of-sight, etc. The impact of subpixel spatial-phase vacillations upon the performance of matched filters has been shown to be unexpectedly significant. The minimum peak-cross-correlation in the presence of SPV can be much lower than anticipated from the straightforward autocorrelation. Even in the absence of signal-to-noise and signal-to-background performance limitations, SPV can cause degradations in the ability of the matched filter to (i) detect the target object and (ii) to discriminate other objects. Spatial-phase vacillations appear to be an inherent limitation of digital imagery when processed using matched filter methodology and can negatively impact the performance of ATR, target tracking, etc systems. Mitigation of this degradation was found to be possible by utilizing one of several matched filter constructions. Superior performance can be obtained by using a multi-filter EMF, although in general a single-filter EMF provides nearly as good performance. Noteworthy performance enhanced over the traditional matched filter was observed using the 'Lite' enhanced matched filter using a relatively few randomly-located subpixel shifted images for its construction. A significant conclusion of this investigation is that, for an ATR application, improved overall task performance should be realized by the use of the EMFB or the EMF–L. In contrast to the typical method of collecting a training set of images for construction of a common matched filter bank, only a modest number of images need to be taken of the object for each view angle, range, and environmental condition assuming that some existing or induced microjitter is present.

Image-matching multispectral-imagery using a complex bank of matched filters can require a large amount of computational resources. One approach to dramatically reducing this requirement has been developed where the multispectral imagery is spectrally-collapsed to form pseudoimages with the spectral information appearing as a texture in a grayscale image designated as a Hyperspace Angle Map (HAM). Both the static examples presented in this paper and those using real-time multispectral imagery [29] have yielded impressive results. Artificial intelligence (AI) and like applications are increasingly incorporating deep-learning methods that require evaluation of massive amounts of data. Utilization of the methods presented in this paper to enhance filter performance, with dramatic reduction in computational time and resources, is anticipated to have major impact on the development of such AI applications.
The use of subpixel microscanning to enhance imagery and mitigate intrinsic aliasing has been investigated for over three decades and has demonstrated very useful improvements in system performance. In this study, the effect of spatial perturbations of the microsamples was investigated. It was observed that perturbations of even 30% had minimal impact upon the resulting enhanced imagery.

The pattern recognition ambiguity function has been shown to answer the question 'What is the best filter spatial-phase coordinates within the pixel that provides the best overall performance (lowest ambiguity) over a choice of ranges?' The assumptions are that the matched filter is established for $R_0$ and the user can select the evaluation ranges. For any abscissa coordinate, the area under the curve in the range axis is a measure of the ambiguity potential. It is anticipated that there should be an EMF-L formulatable which yields the minimum ambiguity probability.

Acknowledgments

U.S. Army Research Laboratory (W911NF-15-1-0531), Alabama A&M University, and Boeing SVS, Inc, each provided, in part, the funding for this investigation. The assistance of Dr Jean Michel Taguenang in generating certain data utilized in this investigation is gratefully acknowledged.

ORCID iD

R Barry Johnson ☑ https://orcid.org/0000-0003-2089-2053

References

[1] Huck F O, Halyo N and Park S K 1980 Aliasing and blurring in 2-D sampled imagery Appl. Opt. 19 2175–81
[2] Watson E A, Muse R A and Blommel F P 1992 Aliasing and blurring in microscanned imagery Proc. SPIE 1689 242–50
[3] Chevrete P and Fortin J 1998 Method and system for fast microscanning US Patent 5,774,179
[4] Hadar O and Boreman G D 1999 Oversampling requirements for pixelated-imager systems Opt. Eng. 38 782–5
[5] Schuler J M and Sibbren D A 1999 Increasing spatial resolution through temporal super-sampling of digital video Opt. Eng. 38 801–5
[6] Park S K and Rahman Z 1999 Fidelity analysis of sampled imaging systems Opt. Eng. 38 786–800
[7] Krapels K, Driggers R G, Vollmerhause R and Halford C 2001 Performance comparison of rectangular (four-point) and diagonal (two-point) dither in undersampled infrared focal plane array imagers Appl. Opt. 40 71–84
[8] Zomet A and Peleg S 2001 Super-resolution from multiple images having arbitrary mutual motion Super-Resolution Imaging ed S Chaudhuri (Norwell: Kluwer Academic)
[9] Sills T G and Williams O M 2004 Aliasing and scintillation reduction in real-time graphics Opt. Eng. 43 1908–15
[10] Wilse J M and Miller J L 2005 Imagery improvements in staring infrared images by employing subpixel micoscan Opt. Eng. 44 05640
[11] Ashley P R and Pittman W C 2008 Image analysis and enhancement system US Patent 7,381,937
[12] Mendlovic D 2013 Toward a super imaging system Appl. Opt. 52 581–566
[13] Image 26 2009 (https://translate.google.com/translate?hl=en&sl=ru&u=https://trinixy.ru/32656-ochen-krasivye-fotografii-37-foto.html&prev=search)
[14] Schmieder D E and Weathersby M R 1983 Detection performance in clutter with variable resolution Aerosp. Electr. Syst. IEE Trans. AES-19 623–30
[15] Sims S R F 2000 Putting ATR performance on an equal basis – the measurement of knowledge base distortion and relevant clutter Proc. SPIE 4040 55–60
[16] Landeau S and Dagobert T 2006 Image database generation using image metric constraints: an application within the CALADIOM project Proc. SPIE 6234 623410
[17] Heidary K and Johnson R B 2013 Autonomous tracking of designated persons in crowded scenes Proc. SPIE 8833 83308
[18] Caulfield H J and Maloney W T 1969 Improved discrimination in optical character recognition Appl. Opt. 8 2354–6
[19] Caulfield H J and Haines R 1980 Generalized matched filtering Appl. Opt. 19 181–3
[20] Caulfield H J and Weinberg M H 1982 Computer recognition of 2D patterns using generalized matched filters Appl. Opt. 21 699–1704
[21] Heidary K and Caulfield H J 2005 Application of supergeneralized matched filters to target classification Appl. Opt. 44 47–54
[22] Heidary K and Caulfield H J 2009 Nonlinear Fourier correlation Proc. SPIE 7340 73400A
[23] Heidary K 2013 Distortion tolerant correlation filter design Appl. Opt. 52 2570–6
[24] Johnson R B and Heidary K 2010 A unified approach for database analysis and application to ATR performance metrics Proc. SPIE 7696 76960Z
[25] Driggers R G, Cox P and Edwards T 1999 Introduction to Infrared and Electro-Optical Systems (Norwood: Artech House) pp 218–23
[26] Holst G C 1995 Electro-Optical Imaging System Performance (Winter Park: JCD Publishing) pp 82–90
[27] Skolnik M 1980 Introduction to Radar Systems 2nd edn (New York: McGraw-Hill)
[28] Vandenberg D 2012 Mathematical survey and application of the cross-ambiguity function MSc Thesis Indiana University, South Bend (https://pdfs.semanticscholar.org/2f94/303f12f19b46ef60eb6d33273e37c10b5.pdf)
[29] Heidary K and Johnson R B 2017 Knowing the unknowable unknowns: subpixel anomalous source detection in hyperspectral imagery IJOER 3 52–65