

A Study of Morphological Pre-Processing Approaches for Track-Before-Detect Dim Target Detection

John Lai, Jason J. Ford, Peter O'Shea, and Rodney Walker
Queensland University of Technology, GPO Box 2434, Brisbane Qld 4001, Australia
j2.lai@student.qut.edu.au, {j2.ford, pj.oshea, ra.walker}@qut.edu.au

Michael Bosse
Autonomous Systems Laboratory, CSIRO ICT Centre QCAT, Brisbane Qld 4069, Australia
Mike.Bosse@csiro.au

Abstract

The *track-before-detect* processing technique has been employed in numerous computer vision based algorithms addressing the dim target detection problem. This processing technique has been shown to be effective under certain conditions; but in particularly noisy or highly cluttered environments, detection performance may be improved by introducing an image pre-processing stage to enhance the raw sensor measurements prior to integration. In this paper, we compare the 'Close-Minus-Open' (CMO) and 'Preserved-Sign' (PS) morphological image pre-processing techniques for suppressing unwanted noise and emphasising target features in the measurement images. This investigation is motivated by the unmanned aerial vehicle "sense-and-avoid" application, where morphology-based filters have demonstrated a degree of success in the detection of small point-like features that may correspond to collision-course aircraft. For completeness, we also briefly examine two well published track-before-detect temporal filtering techniques which may be combined with the morphological pre-processing to detect dim, sub-pixel sized targets. Results from our simulation studies show that the PS approach achieves a higher detection rate than the CMO approach.

1 Introduction

The ability to detect and track targets in naturally lit, high noise environments is becoming increasingly important. Significant challenges arise however in the use of machine vision for target detection and tracking because of the need to contend with not only the inherent noise of imaging sensors, but also with noise introduced by changing and unpredictable ambient conditions. The need to overcome these challenges has significantly driven the development of image filtering and processing techniques.

Over the last three decades, a two-stage processing paradigm has emerged for the simultaneous

detection and tracking of dim, sub-pixel sized targets [Gandhi *et al.*, 2006; Gandhi *et al.*, 2003; Arnold *et al.*, 1993; Barniv, 1985]. These two stages are: 1) an image pre-processing stage that, within each frame, highlights potential targets with attributes of interest; and 2) a subsequent temporal filtering stage that exploits target dynamics across frames. The latter temporal filtering stage is often based on a track-before-detect processing concept where target information is collected and collated over a period of time before the detection decision is made. In this paper, we are primarily interested in the image pre-processing stage of the above two-stage paradigm; but for completeness, we also briefly investigate the performance of two well known track-before-detect temporal filtering techniques which may be combined with the morphological pre-processing to detect slow dim sub-pixel sized targets.

Generally, the goal of the image pre-processing stage is to enhance potential target features whilst suppressing background noise and clutter. There is an abundance of techniques and algorithms available which may be considered for this image processing role. In particular, non-linear spatial techniques such as median subtraction filters [Deshpande *et al.*, 1999] have been widely discussed in the literature. Another non-linear image filtering approach that has received much attention over the last decade has its basis in mathematical morphology [Dougherty and Lotufo, 2003]. Numerous morphology-based filters have been proposed for the detection of small targets in infrared (IR) images [Zhu *et al.*, 2000; JiCheng *et al.*, 1996; Tom *et al.*, 1993]. Specific implementations of the morphological filtering approach include the Hit-or-Miss filter [Schaefer and Casasent, 1995], Close-Minus-Open filter [Casasent and Ye, 1997], and the Top-Hat filter [Braga-Neto *et al.*, 2004]. Although a large proportion of research has focused on IR images, there are recent examples of morphological filters being incorporated into target detection algorithms operating on video images [Carnie *et al.*, 2006; Gandhi *et al.*, 2006; Gandhi *et al.*, 2003]. Moreover, a sign of the increasing popularity of morphological filters for small target detection is evident in the host of studies undertaken into the issue of parameter design [Zeng *et al.*, 2006; Yu *et al.*, 2003]. Finally, there have been efforts made to compare

existing techniques with the morphology-based filters [Gandhi *et al.*, 2006; Warren, 2002; Tom *et al.*, 1993; Barnett *et al.*, 1993], with the median filtering technique often featuring in the comparison studies.

The second focus of this paper concerns the temporal filtering stage that follows the image pre-processing. The temporal filter is designed to extract image features that possess target-like temporal behaviour. For this role, there are two particular filtering approaches that have received much attention in the literature: Viterbi based approaches and Bayesian based approaches.

The Viterbi algorithm has formed the basis of the temporal filtering stage in numerous track-before-detect algorithms [Davey *et al.*, 2008; Gandhi *et al.*, 2006; Tonissen and Evans, 1996; Arnold *et al.*, 1993; Barniv, 1985]. This is in part due to its utility in the context of tracking where, under a number of assumptions, it is able to efficiently determine the optimal target track within a data sequence [Forney, 1973]. Some analysis of the Viterbi algorithm's detection and tracking performance can be found in [Johnston and Krishnamurthy, 2000; Barniv and Kella, 1987; Tonissen and Evans, 1996], and modifications that enhance the algorithm's tracking performance in the presence of non-Gaussian clutter noise have been proposed in [Arnold *et al.*, 1993]. An alternative temporal filter design for track-before-detect algorithms is based on Bayesian filtering [Davey *et al.*, 2008; Bruno, 2004; Bruno and Moura, 2001; Bruno and Moura, 1999]. In [Bruno and Moura, 1999], the typical white Gaussian noise assumptions are relaxed, with consideration given to spatially correlated clutter. Moreover, in [Bruno and Moura, 2001], the modelling of clutter is expanded to encompass a variety of Gaussian and non-Gaussian, correlated and uncorrelated clutter types, and the Bayesian algorithm is extended to accommodate multiple targets that may feature randomly varying amplitudes or intensities.

Some comparison between the Viterbi and Bayesian approaches has been made at the theoretical level [Bruno and Moura, 2001], as well as on the practical level via Monte Carlo simulation trials [Davey *et al.*, 2008]. However, conclusions about detection and false-alarm performance are beyond the scope of the theoretical analysis provided in [Bruno and Moura, 2001], while the simulation results of [Davey *et al.*, 2008] are limited to a specific false-alarm rate (the detection tradeoffs for various false-alarm rates that would be useful to a system designer are not available).

The main aim of this paper is to investigate the use of two alternative morphological filtering approaches as the pre-processing stage for track-before-detect algorithms operating on image sequences. This investigation is motivated by the unmanned aerial vehicle "sense-and-avoid" application, where morphology-based filters have demonstrated a degree of success in the detection of small point-like features that may correspond to collision-course aircraft [Carnie *et al.*, 2006]. Here, we compare the 'Close-Minus-Open' (CMO) approach with the less well characterised 'Preserved-Sign' (PS) technique. This comparison is performed in the context of a hidden Markov model (HMM) temporal filtering stage (which has recently been shown to be an effective choice for track-before-detect algorithms [Davey *et al.*, 2008]). As a second objective, we briefly focus our attention on the temporal filtering aspect and consider the detection

performance of a Viterbi-based approach as compared to a Bayesian HMM approach when using morphologically pre-processed image measurements as input.

We assess our detection algorithms via detection, time-to-detection, and false-alarm statistics that provide valuable insight into the tradeoffs in performance from using the two different pre-processing and temporal filtering approaches. The performance of the morphological pre-processing is examined under cross-tracking target scenarios featuring a range of target speeds and signal-to-noise ratios with different noise conditions. In our temporal filtering comparison, we consider not only cross-tracking targets, but also 'emerging' targets that gradually become more distinct over time, such as those that might be expected in an airborne collision avoidance scenario [BASI, 1991]. Our simulation studies show that a PS morphological pre-processing approach in combination with the HMM temporal filter provides the best tradeoff between detection and false-alarm performance.

2 Morphological Image Pre-Processing

In this section, we provide a brief review of the morphological filtering approaches compared in this paper.

2.1 Morphological Operations

We are concerned with greyscale morphological operations designed to be applied to discrete 2D image data quantised to a finite number of intensity or greyscale levels, such as might be expected from the output of an electro-optical sensor. Here, we combine two fundamental morphological operations known as 'dilation' and 'erosion' to create more sophisticated filtering operations for extracting small, point-like features that are present within an image frame.

Let $Y \oplus S$ and $Y \ominus S$ denote the dilation and erosion respectively of a greyscale image Y by a morphological structuring element S (see [Soille, 2003; Dougherty and Lotufo, 2003] for formal definitions of the dilation and erosion operations). The dilation and erosion operations can be combined to form secondary operations that play key roles in morphological image processing; these being 'opening' and its dual 'closing'. Let $Y \circ S$ and $Y \bullet S$ denote the opening and closing respectively of a greyscale image Y by a morphological structuring element S . The opening operation is simply defined as an erosion followed by a dilation

$$Y \circ S = (Y \ominus S) \oplus S, \quad (1)$$

and closing the reverse – a dilation followed by an erosion

$$Y \bullet S = (Y \oplus S) \ominus S. \quad (2)$$

An intuitive understanding of the morphological opening procedure can be gained by visualising the opening operation as the darkening of locally bright regions (which are smaller than the size of the structuring element) to the values of their neighbourhood pixels. In a similar manner, morphological closing may be regarded as the process of brightening locally dark regions (which are smaller than the structuring element) to match the values of neighbouring pixels.

It follows from the above concepts that taking the difference between an image and its morphological opening will produce an output identifying positively contrasting features (pixel groups that are brighter than their neighbours). This corresponds to what is referred to in the literature as a ‘top-hat’ transformation [Gonzalez, 2004]. Let $TH(Y, S) = Y - (Y \circ S)$ denote the top-hat transformation of image Y . Its dual, the ‘bottom-hat’ transformation of image Y , is defined as $BH(Y, S) = (Y \bullet S) - Y$ [Gonzalez *et al.*, 2004] and accordingly will highlight negatively contrasting features (pixel groups that are darker than their neighbours). In each case, only features smaller than the structuring element are preserved, whilst larger features are suppressed. Thus, via carefully chosen structuring elements, the above morphological transformations and operations represent a powerful set of image processing techniques for identifying features of interest based on their geometrical size.

The two morphological filters to be investigated in this paper are based on a combination of the top-hat and bottom-hat transformations. Both filtering approaches aim to differentiate between genuine and non-genuine features of interest based on size via appropriate design of the structuring elements.

2.2 Close-Minus-Open Filtering Approach

The Close-Minus-Open (CMO) filtering approach is given by the sum of the top-hat and bottom-hat transformations

$$\begin{aligned} CMO(Y, S) &= TH(Y, S) + BH(Y, S) \\ &= \{Y - (Y \circ S)\} + \{(Y \bullet S) - Y\} \quad (3) \\ &= (Y \bullet S) - (Y \circ S), \end{aligned}$$

which simplifies down to the difference between the closing and opening of an image. We take advantage of this combination of secondary morphological operations, which has been referred to elsewhere in the literature as a self-complementary top-hat filtering approach [Soille, 2003], to simultaneously preserve both positively and negatively contrasting features with sizes that match potential features of interest (both positive and negative contrasting features result in a non-negative output).

2.3 Preserved-Sign Filtering Approach

An alternative to the CMO technique is the Preserved-Sign (PS) filtering approach, defined as the difference between the top-hat and bottom-hat transformations [Carnie *et al.*, 2006]:

$$\begin{aligned} PS(Y, S) &= TH(Y, S) - BH(Y, S) \\ &= \{Y - (Y \circ S)\} - \{(Y \bullet S) - Y\} \quad (4) \\ &= 2Y - (Y \circ S) - (Y \bullet S). \end{aligned}$$

The above definition shows that the PS filtering approach is in fact a variation on the top-hat contrast enhancement operator described in [Soille, 2003].

The distinguishing feature of the PS filtering approach is that the response to positively contrasting features is non-negative, whereas the response to negatively contrasting features is non-positive. This is in contrast with the CMO approach where any contrasting

feature (positive or negative) is expressed as a non-negative output. Thus, unlike the CMO technique, the PS filtering approach identifies contrasting features and provides additional information regarding the contrast ‘polarity’.

2.4 Proposed Filter Implementation

In this paper, we configure our CMO and PS filtering implementations to serve as powerful tools in the identification of small point-like features within the measurement image. For performance and computational reasons, we exploit a directional decomposition technique [Casasent and Ye, 1997] in our implementation of the CMO and PS morphological filters. In the case of the CMO approach, we take the minimum response from a pair of CMO filters using orthogonal 1D structuring elements. Here, one CMO filter operates exclusively in the vertical direction, while the other operates exclusively in the horizontal direction. Similarly, for the PS approach we implement two separate filters (one operating in the vertical direction and the other in the horizontal direction) using orthogonal 1D structuring elements, but take the minimum *magnitude* response from the pair of PS filters as the final output.

3 Temporal Filtering

In many electro-optical based detection problems, the existence of a target in a 3D volume of space must be determined from observations of a projection of the target space onto a 2D image plane. Here, target detection can be viewed as evaluating the likelihood of 2 alternate hypotheses, where H_1 denotes the hypothesis that there is a single target present in the camera field of view, and H_2 denotes the hypothesis that there is no target present. The temporal filtering approaches implemented in this paper will assume that under hypothesis H_1 , the projected target motion resides on a 2D plane fixed in space that is represented by the set of discrete 2D grid points $\{(i, j) | 1 \leq i \leq N_v, 1 \leq j \leq N_h\}$, with vertical and horizontal resolutions N_v and N_h respectively. Let $N = N_v N_h$ denote the total number of grid points. The measurements are provided by an electro-optical imaging sensor whose field of view is represented by a 2D grid of image pixel locations aligned with the target space and denoted $\{(p, q) | 1 \leq p \leq N_v, 1 \leq q \leq N_h\}$.

3.1 Hidden Markov Model Filtering

We will assume that, when present, the target is located within a particular pixel of the image frame at each time instant. Thus, each pixel (i, j) represents a unique *state* of the HMM in our target detection problem. For notational convenience, we stack the columns of the image frame to form a vector of pixel locations. In this way, each state may be referenced by a single index, in the sense that if the target is at pixel location (i, j) , this corresponds to it being in the state $m = [(j-1)N_v + i]$.

Let X_k denote the state (target location) at time k . Between consecutive image frames the target may move to different pixel locations – that is, the target can transition between the states. The likelihood of state

transitions can be described by the HMM's *transition probabilities* $A^{mn} = P(X_{k+1} = \text{state } m | X_k = \text{state } n)$ for $1 \leq m, n \leq N$, which is the probability of moving from any one pixel position (state) n to any other pixel position (state) m . The transition probabilities can therefore be used to describe the expected mean target motion. For example, in the case of slow moving targets we tend to assign low probabilities to transitions between distant pixels. Moreover, *initial probabilities* $\pi^m = P(X_1 = \text{state } m)$ for $1 \leq m \leq N$ are used to specify the probability that the target is initially located in state m . Finally, to complete the parameterisation of the HMM, there are the *measurement probabilities* $B^m(Y_k) = P(Y_k | X_k = \text{state } m)$ for $1 \leq m \leq N$ that are used to specify the probability of obtaining the observed measurement Y_k , given that the target is actually in pixel location (state) m (see [Elliott *et al.*, 1995] for more details about the parameterisation of HMMs).

Detection Strategy

The HMM filtering approach performs temporal integration of the input measurements by recursively propagating α , an unnormalised probabilistic estimate of the target state X_k , over time. This is achieved via the forward part of the forward-backward procedure describe in [Rabiner, 1989], which can be decomposed into two stages: initialisation and recursion.

For $1 \leq m \leq N$

- 1) Initialisation: Let α_k^m denote the probability $P(Y_1, Y_2, \dots, Y_k, X_k = \text{state } m)$. Then $\alpha_1^m = \pi^m B^m(Y_1)$.
- 2) Recursion: At time $k > 1$, set $\alpha_k^m = \left[\sum_{n=1}^N \alpha_{k-1}^n A^{nm} \right] B^m(Y_k)$.

The forward procedure filtering result is closely related to the two probabilistic measures that we use to facilitate the detection of targets: 1) the probability of measurements up to time k assuming H_1 , given by

$$P(Y_1, Y_2, \dots, Y_k | H_1) = \sum_{m=1}^N \alpha_k^m, \quad (5)$$

and 2) the conditional mean filtered estimate of the target state m given measurements up to time k and assuming H_1 , given by

$$\begin{aligned} \hat{X}_k^m &= E[X_k = \text{state } m | Y_1, Y_2, \dots, Y_k, H_1] \\ &= \frac{\alpha_k^m}{\sum_{n=1}^N \alpha_k^n}, \end{aligned} \quad (6)$$

where $E[\cdot]$ denotes the mathematical conditional expectation operation (see [Billingsley, 1995] for more details). The probability $P(Y_1, Y_2, \dots, Y_k | H_1)$ may be interpreted as an indicator of target presence (following

the probabilistic distance results of [Xie *et al.*, 2005]), and the conditional mean estimate can be regarded an indicator of likely target locations.

In the interest of computational efficiency, we choose in this paper to evaluate the conditional mean estimate directly from the following expression [Elliott *et al.*, 1995]:

$$\hat{X}_k = N_k B_k(Y_k) A \hat{X}_{k-1}, \quad (7)$$

where N_k is a scalar normalisation factor; $B_k(Y_k)$ is a $N \times N$ matrix where the main diagonal is occupied by the values of $B^m(Y_k)$ for $1 \leq m \leq N$ and all other elements are zero; A is a $N \times N$ matrix with elements A^{mn} ; and \hat{X}_k is a $N \times 1$ vector consisting of elements \hat{X}_k^m for $1 \leq m \leq N$ that are equivalent to those given in (6). Moreover, we note the following relationship between the normalisation factor N_k and the probability of measurements up to time k assuming H_1 :

$$P(Y_1, Y_2, \dots, Y_k | H_1) = \prod_{l=1}^k 1/N_l. \quad (8)$$

HMM Filter Implementation

In this paper, we define the transition probabilities so that only a self-transition or a transition to any one of the 8-connected neighbours is possible from time k to $k+1$. This leads to a sparse A matrix that is efficient to implement in practice. Figure 1 provides a visual representation of the possible state transitions in the HMM filter.

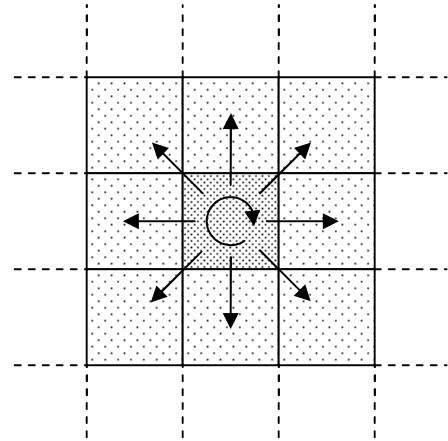


Figure 1. Possible state transitions of the HMM filter. The solid squares define the 8-connected neighbourhood of the state (depicted here as the central cell), with the arrows indicating the transition possibilities from time k to $k+1$. Thus, the state may only undergo a self-transition (dark cell) or a transition to any one of its 8-connected neighbours (light cells).

Furthermore, we note that our implementation of the HMM filter exploits the following probabilistic relationship between target location X_k and the pre-processed measurements Y_k :

$$B^m(Y_k) = \frac{P(Y_k^m | X_k = \text{state } m)}{P(Y_k^m | X_k \neq \text{state } m)}, \quad (9)$$

for $1 \leq m \leq N$. We highlight the computational advantage that (9) affords, given that $P(Y_k^m | X_k = \text{state } m)$ and $P(Y_k^m | X_k \neq \text{state } m)$ can each be determined on a single-pixel basis (rather than requiring the probability of a whole image).

Remark: Strictly speaking, the right hand side of (9) is proportional to $B^m(Y_k)$ (the applicable scaling factor may be absorbed by N_k in the implementation of (7)), and only holds under the following assumptions: 1) the statistical properties of pixel values within an image are spatially independent, and 2) individual pixels do not allow the opportunity of perfect detection, in the sense that $P(Y_k^m | X_k \neq \text{state } m) > 0$ whenever $P(Y_k^m | X_k = \text{state } m) > 0$. Admittedly, the presence of extended (multi-pixel) targets or spatially correlated noise would violate the above assumptions.

3.2 Viterbi-Based Filtering

We implement a Viterbi-based temporal filtering approach that is based on the dynamic programming algorithms of [Carnie *et al.*, 2006] and [Gandhi *et al.*, 2006], where pre-processed image frames are integrated over time along possible target trajectories in order to improve the signal-to-noise ratio. The output is an image, with large pixel values indicating likely target locations.

The motion of the target is modelled in terms of discrete velocity cells, where each cell (u, v) encompasses a range of possible target velocities. Assuming constant target velocity (i.e. no transitions between velocity cells) and a velocity cell resolution of one pixel/frame, it can be shown that if the target is at any particular pixel (i, j) at time k , there exists a neighbourhood of four connected pixels within which the target will be located at time $k+1$ [Tonissen and Evans, 1996]. We denote this neighbourhood of four connected pixels by Q and refer to them collectively as the *forward state transitions*. By symmetry, if the target is at any particular pixel (i, j) at time k , there exists a neighbourhood of four connected pixels within which the target was located at time $k-1$. We denote this neighbourhood of four connected pixels by \bar{Q} and refer to them collectively as the *backward state transitions*. Hence, for each velocity cell (u, v) there exists a corresponding and unique set of forward state transitions Q_{uv} and backward state transitions \bar{Q}_{uv} . This model for target dynamics can be modified to allow for transitions to more than four pixels at a time by adjusting the resolution of the velocity cells. However, previous analysis has shown that better performance is achieved for smaller numbers of possible transitions [Tonissen and Evans, 1996], with four transitions considered to be a reasonable choice for slow, non-maneuvring targets [Gandhi *et al.*, 2006].

Detection Strategy

The Viterbi-based algorithm performs temporal integration of the input measurements by recursively

generating a set of intermediate images a for each velocity cell (u, v) that is considered. This process can be divided into two stages: initialisation and recursion.

For all (u, v) , $1 \leq i \leq N_v$, and $1 \leq j \leq N_h$

- 1) Initialisation: Let $a_k^{ij}(u, v)$ denote the ij th pixel of the intermediate image frame at time k for velocity cell (u, v) . Then $a_1^{ij}(u, v) = 0$.
- 2) Recursion: At time $k > 1$, set $a_k^{ij}(u, v) = [(1 - \beta)y_k^{ij}] + \beta \cdot \max_{(i', j') \in Q_{uv}} [a_{k-1}^{i'j'}(u, v)]$, where y_k^{ij} is the ij th pixel of the pre-processed image at time k , and β represents a memory factor that can vary between zero and one.

At anytime T when a detection decision is required, we take the maximum output across corresponding pixels of the intermediate image frames belonging to each velocity cell:

$$a_{\max}^{ij} = \max_{(u, v)} [a_T^{ij}(u, v)], \quad (10)$$

for $1 \leq i \leq N_v$ and $1 \leq j \leq N_h$. This final image a_{\max} that consolidates target information from across all velocity cells then serves as the basis for declaring detections.

Viterbi-Based Filter Implementation

The Viterbi-based filter implemented here mirrors those seen in [Carnie *et al.*, 2006] and [Gandhi *et al.*, 2006], where four velocity cells are used to detect targets that move with constant velocity in any direction, but are limited to a maximum speed of 1 pixel per frame. Nonmaximal suppression is applied to the output of the filter to reduce an undesirable ‘‘dilation’’ effect where pixels in the neighbourhood of the target also attain significantly large values [Gandhi *et al.*, 2006].

4 Performance Characterisation

In our main study, we compare the performance of the two alternative morphological pre-processing techniques

- A Close-Minus-Open filter, and
- A Preserved-Sign filter

in the context of the track-before-detect problem by applying them to a large number of image sequences containing cross-tracking targets having a variety of intensity and speed attributes. This pre-processed data is then sent to a HMM temporal filtering stage. The HMM filtering stage is individually optimised for each particular pre-processing technique to ensure that neither approach is unfairly disadvantaged.

As a second study, we compare the HMM temporal filtering approach from above with a Viterbi-based approach by applying them to a large number of morphologically pre-filtered image sequences containing two different target types. In particular, we consider cross-tracking (constant size and intensity) and looming targets. The cross-tracking target case allows us to compare results with those of existing studies [Gandhi *et al.*, 2006]; whereas the looming target case allows us to examine

performance in a new and important detection scenario.

In the following subsections, we describe our metrics for quantifying performance, our procedure for generating image sequences, specific filtering parameters, as well as the presentation of results. These aspects comprise the simulation framework of our comparison studies.

4.1 Performance Metrics

In the detection of cross-tracking targets, we are interested in detection versus false-alarm statistics, whereas in the detection of looming targets we consider time-to-detection versus false-alarm statistics. If a target is present, the track-before-detect algorithm is considered to have achieved a *detection* if the algorithm correctly identifies the target's presence and locates it to within two pixels of the true position. We define the *detection rate* as the number of detections divided by the maximum number of possible detections. A *false-alarm* event occurs if the track-before-detect algorithm incorrectly declares the presence of a target. We define the *false-alarm rate* as the number of false-alarms divided by the total number of possible false-alarm events. The *time-to-detection* is defined as the frame when detection is first achieved.

For the HMM filtering approach, let η_k , our test statistic for declaring the presence of a target at time k , be given by

$$\eta_k = \frac{1}{k} \log \left(\prod_{l=1}^k \frac{1}{N_l} \right) \quad (11)$$

(in practice, a statistically equivalent recursive moving average would instead be implemented). When η_k exceeds a predefined threshold, the HMM track-before-detect algorithm considers a target to be present and located at state

$$\gamma_k = \arg \max_m (\hat{X}_k^m) \quad (12)$$

at time k . Our definition of η_k and γ_k is motivated by the detection strategy discussed in Section 3.1.

For the Viterbi-based filtering approach, our test statistic λ_k for declaring the presence of a target at time k is given by

$$\lambda_k = \max_{ij} (a_k^{ij}) \quad (13)$$

When λ_k exceeds a predefined threshold, the Viterbi-based algorithm considers a target to be present and located at state

$$\zeta_k = \arg \max_{ij} (a_k^{ij}) \quad (14)$$

at time k . The definition of λ_k and ζ_k follow from the interpretation of the Viterbi-based algorithm's output as an image, where the pixel values correspond to target signal strength. Finally, we note that detection and false-alarm statistics are based on the final temporal filtering output after the last image sequence frame is processed.

4.2 Image Sequence Generation

As much as possible, we attempt to include image sequences in our simulation studies that are sourced from authentic data; however our access to this type of data is very limited. Thus, in order to carry out the large number of trials our simulation studies require, we make use of synthetically generated image sequences. The image frames that comprise our synthetic image sequences is formed by adding a noise component and a target signature (if required) to a uniform background image set at an arbitrary greyscale level of 128.

Consider any finite sized target with a non-zero velocity. As this target traverses across the image, the physical extent of the target is likely to overlap multiple pixels at any time. We model the target signature in each pixel as being proportional to the amount the target overlaps the pixel. For example, if the target occupies half the area of a particular pixel, then the target signature in that pixel is assigned half the value of the target intensity.

The basic target model is that of a cross-tracking target that has fixed target size and intensity for the duration of the image sequence. Conversely, our looming target model represents the scenario of a fixed size object approaching the imaging sensor from a very distant location at constant speed. Based on simple point light source and thin lens optics assumptions, we find that the looming target scenario can be constructed by maintaining a fixed target intensity and increasing the target size exponentially at a rate that is consistent with the closing speed and target distance that is being simulated.

Our synthetic image frames will incorporate either one of two types of noise: 1) Zero-mean Uncorrelated Gaussian Noise, or 2) Spatially Correlated Noise. We model the spatially correlated noise as a Gauss-Markov random field (GMRF) (see [Moura and Balram, 1993; Moura and Balram, 1992] for more details) because it is believed to be representative of an important component of the noise present in electro-optical sensors [Bruno and Moura, 1999]. For computational reasons, a first-order, spatially homogenous field with free or Dirichlet boundary conditions is assumed, parameterised by vertical and horizontal field potentials or interactions.

Signal-to-noise ratio (SNR) is used to provide an indication of how distinct a target is (i.e. how well it stands out from the background, and in turn the ease with which it may be detected). Here, we define SNR as $20 \log_{10}(I/\sigma)$ dB, where I is the maximum target signature value and σ the noise standard deviation. The upper bound for the SNR, the peak SNR (PSNR), is achieved when I is equal to the target intensity. In our simulation studies, the tendency of the SNR to vary with time leads us to quote the PSNR instead as a means of signifying how distinct the target is. We note that the PSNR is only indicative of the target's distinctiveness and provides an over bound on the average SNR for cross-tracking targets only (for looming targets, the SNR grows with time and hence the PSNR is not defined).

4.3 Filtering Parameters

Morphological Pre-Processing

Both the CMO and PS approaches employ the same pair of structuring elements given by $s_v = [1, 1, 1, 1, 1]^T$ and $s_h = [1, 1, 1, 1, 1]$ to filter in the horizontal and vertical

directions, respectively.

Temporal Filtering

For the Viterbi-based filter, we let $\beta = 0.75$, which has been demonstrated in [Carnie *et al.*, 2006] to be a reasonable choice for the memory factor.

For the HMM filter, we assign a probability of $7/15$ to self-transitions and a probability of $1/15$ for transitions to adjacent pixels. We highlight that performance was not overly sensitive to these parameter choices (the particular values selected here were found to give reasonable performance). Furthermore, it is assumed that the target may be initially located anywhere within the image frame with equal probability, and hence we set $\pi^m = 1/N$ for $1 \leq m \leq N$.

To construct our measurement probability matrix $B_k(Y_k)$, we are required to estimate the probabilities $P(Y_k^m | X_k \neq \text{state } m)$ and $P(Y_k^m | X_k = \text{state } m)$. The former describes our prior knowledge about the distribution of pixel values in the absence of a target (i.e. the noise and clutter distribution), while the latter captures our prior knowledge about the distribution of values at pixels containing a target. We estimate the required probabilities for $B_k(Y_k)$ directly from data. The probability $P(Y_k^m | X_k \neq \text{state } m)$ is estimated as the average frequency that each pixel value resulted from a non-target location. Using a similar procedure, $P(Y_k^m | X_k = \text{state } m)$ is estimated as the average frequency that each pixel value measurement resulted from a target location.

Finally, for convenience, we choose in both temporal filtering approaches to process positively and negatively contrasting targets separately when dealing with the PS filtering output.

4.4 Performance Tradeoff Curves

In this paper, our preferred method of presenting the simulation study results is via graphical performance curves that concisely illustrate the tradeoffs between two suitable performance metrics (such as detection vs. false-alarm or time-to-detection vs. false-alarm) independent of threshold values. For this approach it is essential that the performance metrics be generated from a fixed set of threshold values evaluated against the test statistics.

5 Performance Comparison Studies

In our first study, we evaluate the performance of the two morphological filtering techniques by allowing their corresponding track-before-detect algorithms featuring a common HMM temporal stage to process a large number of image sequences containing cross-tracking targets with various target attributes and noise properties. In the second study, we briefly compare the HMM and Viterbi-based temporal filtering approaches using morphologically pre-processed image sequences containing cross-tracking and looming targets.

5.1 Study 1

The CMO and PS pre-processing techniques are compared in two simulation scenarios using synthetically generated image sequences. In scenario 1, we examine the

performance of the pre-processing techniques in uncorrelated Gaussian noise, whereas in scenario 2 spatially correlated noise is considered.

Scenario 1

The aim of this scenario is to investigate the performance of the morphological filters in uncorrelated Gaussian noise (with zero-mean; standard deviation of 1) for a selection of target speeds and PSNRs. Specifically, we consider a 1×1 pixel cross-tracking target travelling at speeds of 0.1, 0.2, and 0.3 pixels/frame and with intensities corresponding to PSNRs of approximately 8, 9.5, and 11 dB. Image frames of size $N_v = 111$ and $N_h = 147$ comprise image sequences that are 151 frames in length. For each combination of target speed and PSNR, detection counts are gathered from 10^5 separate image sequences each containing only a single target (all targets converged towards the centre of the image frame, but began at initial locations evenly distributed about the centre). The same number of image sequences (without targets) is used in the calculation of false-alarm counts.

The simulation results show the PS filtering approach to be superior to the CMO approach for all combinations of PSNRs and target speeds. For example, consider the detection vs. false-alarm tradeoff curves illustrated in Figure 2, which is for a target speed of 0.1 pixels/frame at 8 dB PSNR. Here, the detection rates associated with the PS approach are significantly higher than those for the CMO approach.

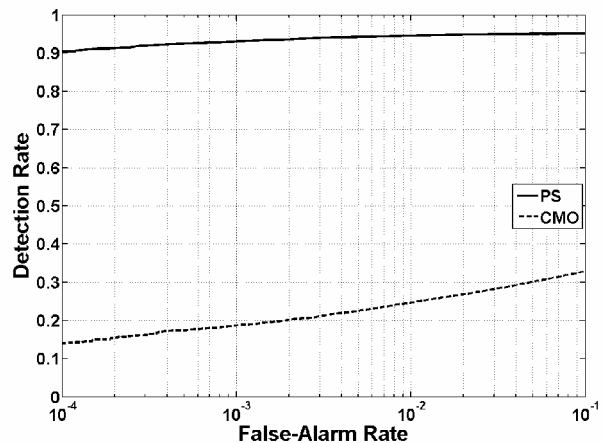


Figure 2. Detection vs. false-alarm performance for CMO and PS morphological filtering (uncorrelated Gaussian noise; target speed 0.1 pixels/frame; PSNR 8 dB)

A comparison of the detection rates at a false-alarm rate of 10^{-3} is provided in Table 1.

PSNR	Target Speed	Detection Rate	
		CMO	PS
8	0.1	0.19	0.93
	0.2	0.07	0.70
	0.3	0.01	0.26
9.5	0.1	0.74	0.99
	0.2	0.35	0.96
	0.3	0.11	0.82
11	0.1	0.98	1.00
	0.2	0.87	0.99
	0.3	0.50	0.97

Table 1. Detection rates for CMO and PS morphological filtering under Gaussian noise at a false-alarm rate of 10^{-3} .

Scenario 2

The simulation parameters of the previous scenario are applied here, with the exception that spatially correlated noise is used instead of uncorrelated Gaussian noise. The spatially correlated noise is modelled by a GMRF having a horizontal and vertical interaction factor of 0.12 and driven by a zero mean Gaussian signal with a standard deviation of 1.

The simulation results show the PS filtering approach to be superior to the CMO approach for all combinations of PSNRs and target speeds. An example of the detection versus false-alarm tradeoff is illustrated in Figure 3, while in Table 2 a comparison of the detection rates at a false-alarm rate of 10^{-3} is provided. Comparing Figures 2 and 3 as well as Tables 1 and 2 reveals little difference in the performance trends of the morphological filtering approaches under the two noise types, with detection perhaps slightly better in correlated noise. However, we anticipate that different results will be obtained with correlated noise parameterised by a higher interaction factor.

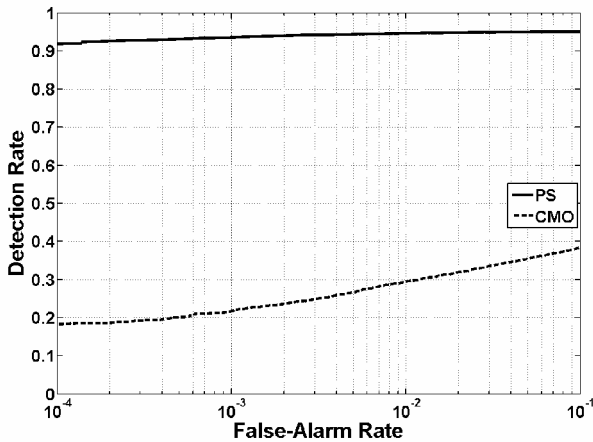


Figure 3. Detection vs. false-alarm performance for CMO and PS morphological filtering (GMRF noise; target speed 0.1 pixels/frame; PSNR 8 dB)

PSNR	Target Speed	Detection Rate	
		CMO	PS
8	0.1	0.22	0.94
	0.2	0.09	0.72
	0.3	0.02	0.27
9.5	0.1	0.79	0.99
	0.2	0.42	0.96
	0.3	0.13	0.83
11	0.1	0.98	1.00
	0.2	0.90	0.99
	0.3	0.57	0.97

Table 2. Detection rates for CMO and PS morphological filtering under GMRF noise at a false-alarm rate of 10^{-3} .

5.2 Study 2

The HMM and Viterbi-based temporal filtering approaches are compared in three simulation scenarios. Scenario 1 simulates a cross-tracking target in spatially correlated noise, whereas scenarios 2 and 3 investigate detection performance for a looming target under spatially correlated noise and real image noise conditions, respectively. We choose to pre-process the image measurements using the PS morphological filtering

approach based on its superior performance results from our first study.

Scenario 1

The performance of the HMM and Viterbi-based temporal filtering approaches in spatially correlated noise is compared in this scenario. The spatially correlated noise is modelled by a GMRF having a horizontal and vertical interaction factor of 0.12 and driven by a zero mean Gaussian signal with a standard deviation of 1. Image frames of size $N_v = 111$ and $N_h = 147$ comprise image sequences that are 151 frames in length. Detection counts are gathered from 5×10^4 separate image sequences each containing only a single cross-tracking target (of 1×1 pixel size, travelling at 0.2 pixels/frame, and with a PSNR of approximately 9.5 dB). The initial locations of the targets are uniformly distributed about the centre of the image. For the calculation of false-alarm counts, the same number of image sequences but without targets is used.

Figure 4 illustrates the detection vs. false-alarm curves for the two temporal filtering approaches. This figure shows the HMM approach demonstrating dramatically better performance than the Viterbi-based approach. We note that this unexpectedly large performance difference is in contrast to recently published results [Davey *et al.*, 2008], which show the HMM (Bayesian) and Viterbi-based (Dynamic Programming) techniques to be quite competitive when morphological pre-processing is not used. We acknowledge that the morphological pre-processing may be more suited to the HMM approach, and that the use of “backtracking” may improve the performance of the Viterbi-based approach implemented here.

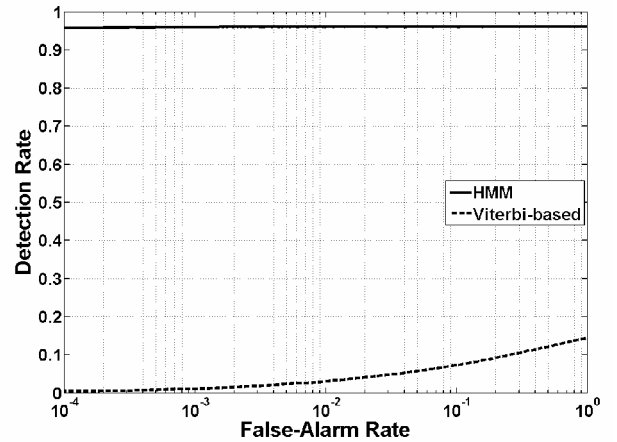


Figure 4. Detection vs. false-alarm performance for HMM and Viterbi-based temporal filtering (GMRF noise; target speed 0.2 pixels/frame; PSNR 9.5 dB)

Scenario 2

In this scenario, we focus on filtering performance for looming targets as opposed to cross-tracking targets. Simulation parameters from the previous scenario are carried over to this scenario, with the exception that image sequences are lengthened to 201 frames for looming targets, and these targets being set to move at 0.02 pixels/frame. The first 20 frames are reserved for algorithm initialisation, leaving the remaining 181 frames for characterising time-to-detection performance. Each image sequence begins with a sub-pixel sized target of

low intensity (approximately 0 dB) during the initialisation frames, which is then allowed to gradually become larger and more distinct such that it models the scenario of a 1 square metre profile object approaching at approximately 103 metres per second from a distance of around 5 kilometres (the time interval between frames is 0.2 second).

Figure 5 shows the time-to-detection vs. false-alarm curves for the two temporal filtering approaches. This figure illustrates that the HMM approach has better looming target detection performance than the Viterbi-based approach (i.e. earlier detection for a specified false-alarm rate). For example, at a false-alarm rate of 10^{-3} , the HMM algorithm detects about 32 frames earlier than the Viterbi-based algorithm. For illustrative purposes, we present in Figure 6 samples of the typical image pre-processing and temporal filtering outputs.

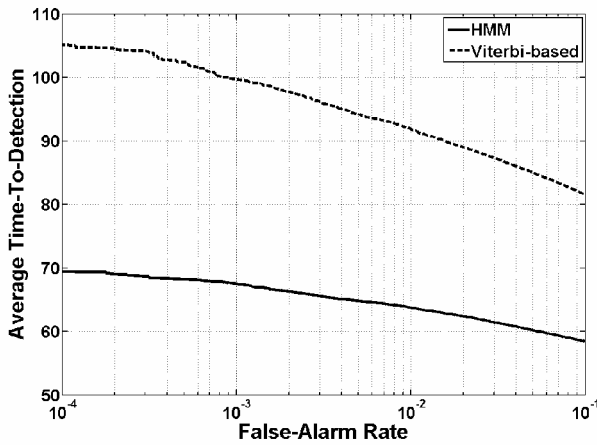


Figure 5. Time-to-detection vs. false-alarm performance for HMM and Viterbi-based temporal filtering (GMRF noise; looming target)

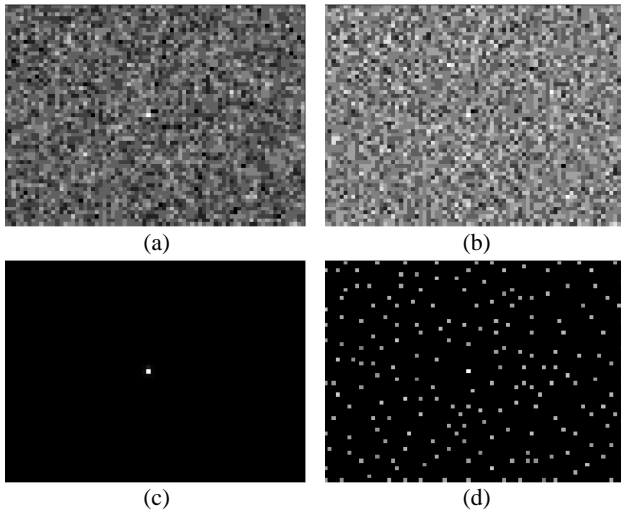


Figure 6. (a) Typical synthetic image frame with correlated noise (GMRF). Target is emerging at centre of image. (b) Image frame after PS morphological pre-processing. (c) and (d) HMM and Viterbi-based temporal filtering output, respectively, after processing 115 frames.

Scenario 3

For scenario 3, the same simulation parameters as the previous scenario are used, except that the synthetic noise

is replaced by noisy image backgrounds obtained from real data sequences. Due to the limited availability of real data, we are unable to provide a comprehensive comparison study of the two algorithms. Processing the real data sets did however provide some useful insight into the existence of non-target semi-persistent features that can delay target detection. These artefacts appear to impact the HMM approach more than the Viterbi-based technique. A possible method of mitigating the detection delays is by introducing extra processing layers that recognise persistent non-target features and reinitialises the filter accordingly. These issues are beyond the scope of this paper and are the subject of ongoing research.

6 Conclusion

In this paper, we compared the Close-Minus-Open (CMO) and Preserved-Sign (PS) morphological image pre-processing techniques in the context of track-before-detect dim target detection. Furthermore, we briefly examined how a HMM temporal filtering approach compares with a competing Viterbi-based algorithm when morphologically processed image data is used. The results from our simulation studies show that the PS pre-processing technique in combination with a HMM temporal filtering approach offers the best tradeoffs for performance in terms of detection, time-to-detection, and false-alarm statistics.

However, it is likely that further refinement of the algorithms is necessary in order to attain a level of performance that would be acceptable in a practical collision avoidance system for UAVs. Our ongoing work includes improving the detection vs. false-alarm tradeoffs through the development of a theoretical basis for the design of HMM filtering parameters (such as the measurement and transition probabilities), as well as addressing persistent non-target features present in noisy real data backgrounds.

Acknowledgements

The authors would like to thank R. Carnie for providing the real image sequences used in this paper. Computational resources and services used in this work were provided by the HPC and Research Support Unit, Queensland University of Technology, Brisbane, Australia.

References

- [Arnold et al., 1993] J. Arnold, S. W. Shaw, and H. Pasternack, "Efficient target tracking using dynamic programming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 44-56, Jan. 1993.
- [Barnett et al., 1993] J. T. Barnett, B. D. Billard, and C. Lee, "Nonlinear morphological processors for point-target detection versus an adaptive linear spatial filter: a performance comparison," presented at the Signal and Data Processing of Small Targets, Orlando, FL, October 22, 1993.
- [Barniv, 1985] Y. Barniv, "Dynamic programming solution for detecting dim moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-21, pp. 144-156, Jan. 1985.
- [Barniv and Kella, 1987] Y. Barniv and O. Kella, "Dynamic programming solution for detecting dim

- moving targets part II: Analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, pp. 776-788, Nov. 1987.
- [BASI, 1991] "Limitations of the See-and-Avoid Principle," Australian Transport Safety Bureau, Apr. 1991.
- [Billingsley, 1995] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.
- [Braga-Neto et al., 2004] U. Braga-Neto, M. Choudhary, and J. Goutsias, "Automatic target detection and tracking in forward-looking infrared image sequences using morphological connected operators," *Journal of Electronic Imaging*, vol. 13, pp. 802-813, Oct. 2004.
- [Bruno, 2004] M. G. S. Bruno, "Bayesian methods for multiaspect target tracking in image sequences," *IEEE Trans. Signal Process.*, vol. 52, pp. 1848-1861, Jul. 2004.
- [Bruno and Moura, 1999] M. G. S. Bruno and J. M. F. Moura, "The optimal 2D multiframe detector/tracker," *Int. J. Electron. Commun. (AEU)*, vol. 53, pp. 1-17, Dec. 1999.
- [Bruno and Moura, 2001] M. G. S. Bruno and J. M. F. Moura, "Multiframe detector/tracker: optimal performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, pp. 925-945, Jul. 2001.
- [Carnie et al., 2006] R. Carnie, R. Walker, and P. Corke, "Image processing algorithms for UAV "sense and avoid"," presented at the IEEE Int. Conf. on Robotics and Automation, Orlando, FL, May, 2006.
- [Casasent and Ye, 1997] D. Casasent and A. Ye, "Detection filters and algorithm fusion for ATR," *IEEE Trans. Image Process.*, vol. 6, pp. 114-125, Jan. 1997.
- [Davey et al., 2008] S. J. Davey, M. G. Rutten, and B. Cheung, "A comparison of detection performance for several track-before-detect algorithms," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1-10, Oct. 2008.
- [Deshpande et al., 1999] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," presented at the Signal and Data Processing of Small Targets, Denver, CO, October, 1999.
- [Dougherty and Lotufo, 2003] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing*. Bellingham, Wash.: SPIE Optical Engineering Press, 2003.
- [Elliott et al., 1995] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. Berlin, Germany: Springer-Verlag, 1995.
- [Forney, 1973] G. D. Forney, Jr., "The viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268-278, Mar. 1973.
- [Gandhi et al., 2003] T. Gandhi, M.-T. Yang, R. Kasturi, O. Camps, L. Coraor, and J. McCandless, "Detection of obstacles in the flight path of an aircraft," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, pp. 176-191, Jan. 2003.
- [Gandhi et al., 2006] T. Gandhi, M.-T. Yang, R. Kasturi, O. I. Camps, L. D. Coraor, and J. McCandless, "Performance characterization of the dynamic programming obstacle detection algorithm," *IEEE Trans. Image Process.*, vol. 15, pp. 1202-1214, May. 2006.
- [Gonzalez et al., 2004] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Morphological Image Processing," in *Digital Image Processing using MATLAB*. Upper Saddle River, NJ: Pearson Prentice Hall, 2004, pp. 334-377.
- [JiCheng et al., 1996] L. JiCheng, S. ZhengKang, and L. Tao, "Detection of spot target in infrared clutter with morphological filter," presented at the National Aerospace and Electronics Conference, Dayton, OH, May 20-23, 1996.
- [Johnston and Krishnamurthy, 2000] L. A. Johnston and V. Krishnamurthy, "Performance analysis of a track before detect dynamic programming algorithm," presented at the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Istanbul, June 5-9, 2000.
- [Moura and Balram, 1992] J. M. F. Moura and N. Balram, "Recursive structure of noncausal Gauss-Markov random fields," *IEEE Trans. Inf. Theory*, vol. 38, pp. 334-354, Mar. 1992.
- [Moura and Balram, 1993] J. M. F. Moura and N. K. Balram, "Statistical algorithms for noncausal Gauss-Markov fields," in *Handbook of Statistics: Signal Processing and Its Applications*, vol. 10. Amsterdam; New York: North-Holland, 1993, ch. 15, pp. 623-691.
- [Rabiner, 1989] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257-286. 1989.
- [Schaefer and Casasent, 1995] R. Schaefer and D. Casasent, "Nonlinear optical hit-miss transform for detection," *Appl. Opt.*, vol. 34, pp. 3869-3882, Jul. 1995.
- [Soille, 2003] P. Soille, "Opening and Closing," in *Morphological Image Analysis: Principles and Applications*, 2nd ed. Berlin, Germany: Springer, 2003, ch. 4, pp. 105-137.
- [Tom et al., 1993] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," presented at the Signal and Data Processing of Small Targets, Orlando, FL, October 22, 1993.
- [Tonissen and Evans, 1996] S. M. Tonissen and R. J. Evans, "Performance of dynamic programming techniques for Track-Before-Detect," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 1440-1451, Oct. 1996.
- [Warren, 2002] R. C. Warren, "A Bayesian Track-before-Detect Algorithm for IR Point Target Detection," Defence Science and Technology Organisation - Aeronautical and Maritime Research Laboratory DSTO-TR-1281, Feb. 2002.
- [Xie et al., 2005] L. Xie, V. A. Ugrinovskii, and I. R. Petersen, "Probabilistic distances between finite-state finite-alphabet hidden Markov models," *IEEE Trans. Autom. Control*, vol. 50, pp. 505-511, Apr. 2005.
- [Yu et al., 2003] N. Yu, H. Wu, C. Wu, and Y. Li, "Automatic target detection by optimal morphological filters," *J. Comput. Sci. Technol.*, vol. 18, pp. 29-40, Jan. 2003.
- [Zeng et al., 2006] M. Zeng, J. Li, and Z. Peng, "The design of Top-Hat morphological filter and application to infrared target detection," *Infrared Physics & Technology*, vol. 48, pp. 67-76, Apr. 2006.
- [Zhu et al., 2000] Z. Zhu, Z. Li, H. Liang, B. Song, and A. Pan, "Gray-scale morphological filter for small-target detection," presented at the Infrared Technology and Applications XXVI, San Diego, CA, December 15, 2000.