A Novel Bio-Inspired Vision Model for Detecting Small and Dim Objects in Infrared Imagery

Journal:	*Transactions on Geoscience and Remote Sensing*
Manuscript ID	TGRS-2023-06614
Manuscript Type:	Regular paper
Date Submitted by the Author:	22-Dec-2023
Complete List of Authors:	AL-SHIMAYSAWEE, LAITH; University of South Australia, STEM Finn, Anthony; University of South Australia STEM Academic Unit, Division of ITEE Uzair, Muhammad; The University of Adelaide - North Terrace Campus, School of Computer and Mathematical Sciences Brinkworth, Russell; Flinders University, College of Science and Engineering
Keyword List 1 (One or more are required to select):	Thermal data, Others
Keyword List 2 (One or more are required to select):	Image/signal analysis (e.g., classification, segmentation, object detection), Others
Keyword List 3 (One or more are required to select):	Others

A Novel Bio-Inspired Vision Model for Detecting Small and Dim Objects in Infrared Imagery

Laith A. H. Al-Shimaysawee, Anthony Finn, Muhammad Uzair and Russell S. A. Brinkworth

***Abstract*—This paper proposes and describes an object detection algorithm, inspired by the insect vision system, that can efficiently suppress clutter and enhance the contrast of dim objects in single frame infrared images. The technique was tested on two different scenarios: detecting small drones at long ranges using an infrared camera carried onboard a drone; and detecting different objects (including planes, helicopters, drones, boats and trucks) in a publicly available dataset called Single-frame InfraRed Small Target detection (SIRST). The proposed technique was compared against twelve existing state-of-the-art infrared object detection algorithms, which are composed of nine spatially based techniques, one neural network technique and two more recent spatio-temporal bio-inspired vision techniques. The proposed technique could detect the weak signatures in high clutter more successfully than all of the existing methods, offering 73.5% percentage object detectability–a 20.5% improvement over the next best method.**

***Index Terms*—Infrared dim target detection, infrared small target detection, drone detection, object detection, insect vision, SIRST dataset.**

I. Introduction

THE field of target detection has many applications, such as security surveillance, search and rescue, and defence [1]–[3]. Detecting targets at long range is of high importance as the earlier an object is detected, the more time is available for decision making. Recently, drones, or unmanned aerial vehicles (UAVs), have become readily available and are now being applied in diverse fields, including cinematography [4], agriculture [5], forestry [6], firefighting [7], [8], meteorology [9], surveillance and defence [10], [11]. However, misuse of drone technology also threatens the safety and security of sites and personnel [12]. For example, flying drones close to airports is dangerous as they could collide with traditional aircraft; and flying them near other facilities could allow them to collect illicit imagery. Detecting drones near important locations is thus a key task for security and safety reasons.

However, detecting small drones at long range and in high clutter environments, whilst desirable, is challenging and has recently drawn the attention of the research community. Long wave infrared (LWIR) thermal cameras have long been used for target detection [13]–[19] as they have the advantage over colour cameras of eliminating certain environmental clutter such as clouds, fog, and smoke. However, the spatial resolution of LWIR cameras is generally inferior to their visible counterparts. Thus, when the target is physically small, and has low contrast to its background, detection of its thermal signature is particularly challenging. State-of-the-art traditional [20]–[25] and neural network detection techniques [26]–[35] were primarily developed to detect physically and/or visibly large (multi-pixel) targets, such as ships, trucks, tanks, and aircraft. Because of their larger size and greater heat dissipation, even if such targets (when at range) occupy only a few pixels within a thermal image, they still tend to have a higher contrast against their surroundings than small UAVs, which are physically small and do not present large thermal signatures. As a result, detecting small drones at long ranges is challenging when using conventional approaches.

Flying insects undertake complex tasks such as detecting and chasing targets in high clutter backgrounds under a wide variety of weather and environmental lighting conditions [36], [37]. Moreover, in spite of the small size and weight of an insect brain and its limited number of neurons [36], it can perform such tasks easily and in real time. This remarkable capability has encouraged scientists to study the vision systems of insects in great depth, and to develop models that are inspired by or directly mimic them [36], [38]–[40]. Bio-inspired vision methods can outperform traditional computer vision processing techniques in complex environments [41]. Moreover, many of these bio-inspired techniques do not require training or prior knowledge of a scene's lighting conditions [41]. Consequently, such techniques have the potential to improve existing object detection algorithms when simply used as a pre-processing tool, particularly when used to enhance target contrast and/or suppress clutter [40], [42].

Existing bio-inspired models typically require temporal information [43], [44], so there is a problem if only still images are available or the movement between frames is too large to apply temporal correlation techniques: they need high frame rate image sequences to work properly. For example, to mimic the analogue signals in biology, the frame rate was $1000Hz$ in [45], $100Hz$ in [46], [47], and $90Hz$ in [42]. Although some bio-inspired vision techniques have been used to process image sequences of lower frame rates ($25-30Hz$ in [41], [48] and $30Hz$ in [40], [42], [49]), the recording cameras were mounted on stationary platforms and either the targets did not move much between consecutive images [40], [42], [49] or datasets were recorded under controlled conditions [41], [48]. Alternatively, when bio-inspired vision techniques have been used with a moving platform and a frame rate of $20Hz$ used,

Laith A. H. Al-Shimaysawee and Anthony Finn are with the University of South Australia, UniSA STEM, Mawson Lakes, SA 5095, Australia (e-mail: laith.al-shimaysawee@mymail.unisa.edu.au; anthony.finn@unisa.edu.au).

Muhammad Uzair is with the University of Adelaide, School of Computer and Mathematical Sciences, Faculty of Sciences, Engineering and Technology, Adelaide, SA 5000, Australia (e-mail: muhammad.uzair@adelaide.edu.au).

Russell S. A. Brinkworth is with the Flinders University, College of Science and Engineering, Tonsley, SA 5042, Australia (e-mail: russell.brinkworth@flinders.edu.au).

the speed of the platform was slowed to only 10 cm/s to overcome the low capture rate of the camera [50].

In this paper, a bio-inspired vision technique is proposed that detects small, dim targets in high clutter environments. It can work on any frame rate, even still imagery. This removes the requirement for high frame rate data and the technique could therefore be applied to situations where decisions need to be made from single images, such as in early warning systems [51]. The proposed technique was benchmarked against twelve existing target detection techniques, of which nine were spatially-based methods, one was a neural network based method, and two were more recent spatio-temporal, bio-inspired vision (BIV) techniques. Two datasets were examined.

The first was a publicly available dataset called Single-frame InfraRed Small Target detection (SIRST) [51]. This dataset has single images of different backgrounds and targets (e.g. planes, helicopters, drones, boats and trucks) collected from different sequences of infrared small targets [51]. The second was recorded by hovering a commercial drone platform at an altitude of 125m and capturing imagery of another small drone flying about 500m away.

In the literature, targets are generally considered small if they occupy less than 9×9 pixels [51], [52], 4×3 [53], a total of 15 pixels [54], or 2×2 or 3×3 [28], [55]. In our dataset, the maximum size of a target is 3×3, and it has very low contrast in the highly clutter environment. The main contributions of this paper are thus:

1) A biologically inspired small target detection technique that can be applied to single images independently, thereby avoiding reliance on high frame rate image sequences.
2) An analysis that shows the proposed technique significantly outperforms existing small target detection techniques by amplifying the signal to noise ratio (SNR) between the target and its surroundings whilst simultaneously suppressing any clutter surrounding it.

The rest of the paper is organised as follows: Section II provides an introduction to biologically inspired vision methods followed by a description of the proposed biologically inspired target contrast enhancement and detection methods. Section III covers the details of the experiments including the drone payload, survey site and the data collection process, comparative detection methods, the settings details of the conducted experiments, and the metrics used to evaluate the performance of the proposed approaches and the competing methods. Sections IV, V and VI presents the results, discussion, and conclusions, respectively.

II. Bio-Inspired Vision Model

The early stages of the insect vision system are composed of the photoreceptor cells in the retina region and the lamina monopolar cells in the lamina region [46]. These layers have a significant ability to enhance contrast and filter noise, i.e. improve SNR. These capabilities are essential when designing techniques for detecting dim objects. Van Hateren, et. al. [56], [57] conducted work on modelling the insect's early vision and how it responds to variations in light intensity. Then, Brinkworth, et. al. [38], [58] developed a mathematical model that mimics how the insect vision system responds to light intensity variation by conducting experiments on living insects. In the following sub-sections, the details of the initial stages (PRC and LMC) of the insect vision system and the mathematical model of these layers were explained based on the literature cited in respective subsections (II-A - II-B). It is important to highlight that the techniques presented in this paper do not propose new contributions to the mathematical models of these PRC and LMC stages, but to the way these stages are used to process the imagery, as described in the following subsections (II-C - II-D). The methods employed for contrast enhancement and detection are also explained.

A. Photoreceptor Cell (PRC) Processing

The photoreceptor cells are part of the retina and are responsible for adapting to light changes. This adaptation helps to effectively compress high dynamic range images without losing important information. It also enhances the contrast between objects of interest and their surroundings using temporal processing to enhance object separation by up to 70% [43]. The photoreceptors are composed of four layers which are adaptive temporal low pass filters, low pass filtered divisive feedback, exponential divisive feedback and a saturating non-linearity [57], [58]. This four layer pathway performs temporal, pixel-wise operations which dynamically adapt to the dark and bright image regions to suppress noise and increase the signal to noise ratio (SNR) [58]–[61].

B. Lamina monopolar cell (LMC) processing

The lamina monopolar cells (LMC) are part of the Lamina and are responsible for removing spatial and temporal redundancy in the signal passed downstream by the photoreceptor cells [38], [56], [62], [63]. By removing the redundant information, the contrast of objects of interest can be enhanced in the scene leading to better object discrimination [63], [64]. The LMC are composed of four layers which are: standardise input, variable temporal high pass filter, spatial high pass filter, and compressive non-linearity [36], [41]. This four layer pathway performs a dynamic spatio-temporal adaptation depending on the scene light conditions [65], where the self adaptation has the ability to enhance object contrast in difficult lighting scenarios.

C. Proposed Bio-Inspired Object Contrast Enhancement

This section describes an extension to the existing bio-inspired vision techniques [36], [39], [40], [42], [44], [47], [66]. Unlike the approaches published previously, where a high image frame rate relative to any motion in the scene is required to avoid aliasing due to the temporal correlation stages or the inability to accurately apply low-pass filtering [38], this proposed augmentation enables processing of any image sequence, regardless of frame rate. Moreover, the proposed technique can be applied to individual images and does not require a sequence of frames (temporal information). As a result, the performance of the proposed technique is not

Algorithm 1 Proposed Object Contrast Enhancement Method: Bio-Inspired Vision Line Scanners (BIVLS)

procedure BIVLS(In, DoS)
  $\boldsymbol{In}$: Input image. $\boldsymbol{DoS}$: Direction of scanning. $\boldsymbol{vecL}$: Number of iterations in the direction of scanning (DoS).
  for $i = 1 : vecL$ **do**
    Compute $O1_t$ by PRC processing to $In(i,:)$,
    Compute $O2_t$ by LMC processing to $O1_t$,
    $Out(i,:) \leftarrow O2_t$
  end for
  return Out
end procedure

Algorithm 2 Proposed Object Detection Method: Multiscale Object Bio-Inspired Vision Line Scanners (MOBIVLS)

procedure MOBIVLS(In)
  $\boldsymbol{In}$: Input image.
  Obtain $T2B$ and $L2R$ according to Algorithm 1.
  where $\boldsymbol{T2B}$: top to bottom, $\boldsymbol{L2R}$: left to right.
  Input $ObjectScaleRange = [K_{min}, k_{max}]$.
  Input $ShiftingStep$.
  $S = \frac{K_{min}}{2} : ShiftingStep : \frac{K_{max}}{2}$.
  for $i = 1 : m$ **do**
    Shift $S4$, $S5$, $S6$ and $S7$ by S_i. Refer to Figure 1 regarding $S4$, $S5$, $S6$ and $S7$.
    $TBLR_{mul} = S4 \bigotimes S5 \bigotimes S6 \bigotimes S7$.
  end for
  where m: number of scales, $\bigotimes$: Hadamard product, $\bigoplus$: array addition.
  $Out = \sum_{i=1}^{m} O_i$
  $Output = Out \bigotimes F$
  return $Output$
end procedure

affected by jumps in frame sequence that often occur during image capture due (for instance) to sudden changes of flying direction by a drone because of weather conditions or operator intervention.

The proposed method scans each frame using one or more line scanners from different directions. Each line scanner performs a modified process of the PRC and LMC stages (see Algorithm 1 pseudo code). This technique is called: Bio-Inspired Vision Line Scanner (BIVLS). The BIVLS is part of the proposed object detection and is explained in the next section. However, the BIVLS can be used independently for object contrast enhancement applications that can comprise still images as well as a video sequence. Figure 1 shows the result of applying the BIVLS to a weak object signature from two directions. As can be seen, each line scan of the BIVLS enhances the leading edge of the object as a rising signal and its trailing edge as a falling signal.

D. Proposed Bio-Inspired Object Detection Method

The proposed object detection technique is called Multi-scale Object Bio-Inspired Vision Line Scanner (MOBIVLS). It is based on applying BIVLS from at least two directions. Figure 1 shows a block diagram describing the proposed MOBIVLS algorithm 2, and is further described in the pseudo code (Algorithm 2).

In MOBIVLS, the input image is processed using the proposed BIVLS from two directions (top to bottom and left to right). Signals from each line scanner are passed through a half-wave rectifier that divides the bipolar signal (half-wave rectification) from the LMC into a positive signal (called the ON channel) and an inverted (negative) signal (called the OFF channel). From the two line scan bipolar signals, four unipolar signals are produced, which have been shifted and multiplied as follows: The two ON channel signals were shifted in the same direction of BIVLS, while the two OFF channel signals were shifted in the opposite direction of the corresponding BIVLS. The reason for this is that each BIVLS enhances the leading edge of the object as a rising signal (the ON channel) and the trailing edge as a falling signal (the OFF channel), and the actual location of the object falls between the rising and falling signals. To determine the actual object location, the position of the leading edge (ON channel) and the trailing edge (OFF channel) need to be shifted by the size of half the object size. Moreover, by multiplying all the shifted versions as a Hadamard product [67], the locations of objects within the image are highlighted and most of the clutter suppressed.

The process of splitting the bipolar signal into ON/OFF channels is inspired by the rectified transient cells (RTC) of the insect brain [36], [44], [68] and is a key step in higher order biologically inspired object detection [47]. Since the object size in the datasets was not constant, several shifts in the range of possible object size were applied, and the output was the accumulative addition of all versions. This output was multiplied by the input image to remove the highlighted areas which were dark in the input image since these areas did not belong to the object.

The idea of shifting in the MOBIVLS is similar to the process of delaying a signal and multiplying it by its original form in an elementary motion detection (EMD) or elementary small target motion detection (ESTMD) stage of the biologically inspired vision models [36], [38], [44], [46], [47]. However, since the new methods are spatially not temporally based, it is possible to move both forwards and backwards. This means that the shifting of the proposed MOBIVLS approach is equivalent to delaying (shifting backward) and expediting (shifting forward) the signal vertically and horizontally, as shown in Figure 1.

The EMD and ESTMD processing is now described to show how they differ from the proposed MOBIVLS approach. The purpose of the EMD stage is to detect motion changes in space over time and their directions. It has two sub-units where each one multiplies a signal by a delayed version of its neighbour (obtained using a low pass filter) and then subtracts the results. The sign of the EMD output represents the motion direction, while the amplitude represents the inverse of the correlation delay, i.e. image velocity [38].

The ESTMD is a spatio-temporal filter for detecting small moving objects. The ESTMD has very similar processing steps to the EMD, but instead of processing two neighbouring signals it processes two versions of the same signal. The two versions are the positive and negative parts of the signal (ON + OFF channels) [44]. In the proposed MOBIVLS approach, there is no temporal dimension, so it is possible to shift the

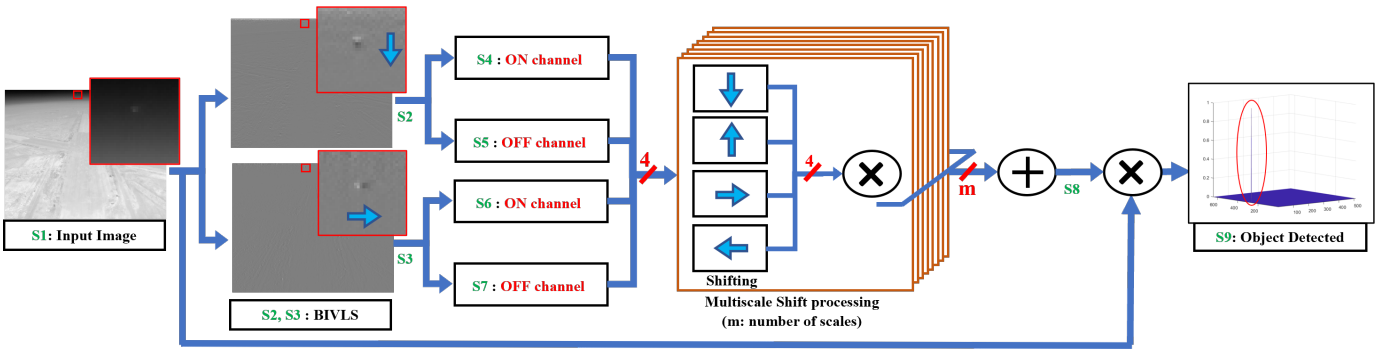

Fig. 1: Block diagram of the proposed bio-inspired object detection technique, named multiscale object of bio-inspired vision line scanners (MOBIVLS). The $\times$ symbol refers to a Hadamard product [67], while the $+$ symbol refers to normal array addition. The directions of the BIVLS processing are from top to bottom (S2), and left to right (S3). Images labelled ($S2 - S3$) show that the object contrast has been enhanced in the direction of the BIVLS line scanner.

signal in any direction. Thus, the ON and OFF channel signals are expedited (shifted forward) and delayed (shifted backward) respectively and multiplied with each other. By doing this, the proposed MOBIVLS approach can detect the centre of the object instead of its edge as per the temporally-based BIV models [44], [47]. In summary: an EMD looks for changes in space over time, an ESTMD looks for changes in a single location over time, while MOBIVLS detects changes in space at a single point in time.

III. Experiments

This section describes the equipment used in the field trials and datasets examined in this study. The implementations of the existing object detection techniques used for comparison against the proposed MOBIVLS, together with any experimental settings and performance evaluation metrics, are also described.

A. Equipment

The infrared camera used to record imagery of the second dataset (IREST) was an ICI-8640 P-series with spectral band 7 - 14µm [69]. The pixel depth (dynamic range) was 14 bits and the payload captured high dynamic range (HDR) raw images. The spatial resolution was 640 $\times$ 512 pixels. The lens model was a 12.5mm manual focus with field of view (FOV) (50° $\times$ 37.5°). The drone carrying the payloads was a commercial platform designed for professional aerial photography and industrial applications. The camera was mounted on a commercial gimbal to provide stability. The computer used to process the data was an Alienware M15 Laptop model with Core i7-9750H 2.6GHz CPU, 16GB DDR4 memory, and NVIDIA GeForce RTX 2060 GPU. The MATLAB software version was R2020b.

B. Infrared Images Datasets

The first dataset examined was a publicly available one called Single-frame InfraRed Small Target detection (SIRST) [51]. This dataset has single infrared images of different backgrounds and objects (planes, helicopters, drones, boats, and trucks) collected from different sequences of small objects [51]. The dataset is composed of 427 images. Figure 2(the first two images) shows samples of the dataset.

The second dataset (IREST) was recorded by hovering the drone platform at an altitude of about 125m and capturing another small drone while it took off from the ground and rose to an altitude of about 80m. The capture frame rate was 30Hz and the distance from the drone platform to the drone target about 500m. The characteristic dimension of the object of interest is about 30cm, but its primary heat source derives from four approximately 2cm x 2cm x 2cm electric engines that drive its propellers, one located at each corner of the drone. By either measure, the drone is sub-pixel. However, the inset images of Figure 2(the second two images) show that the object of interest occupies around 3×3 pixels, albeit with low contrast to the background. This expanded dimension is caused by the camera optics. There is also notable thermal radiation emanating from the ground, which adds considerable complexity to the detection of an object if false alarms are to be minimised. The dataset is composed of 495 images, 246 of which have a single object in them while the rest 249 images do not have any object of interest. Throughout the field trial the location of the drone was measured using a small standard positioning service (SPS) GPS receiver that had an update rate of 1Hz. The accuracy of the GPS receiver locations was thus $\pm$3m and this was used to assist with the visual/manual location of targets within the LWIR images of the dataset, which was undertaken using software tools written in MATLAB.

C. Comparative Methods

Twelve state-of-the-art small target detection techniques were used as a benchmark to evaluate the performance of the proposed MOBIVLS approach. Nine of these techniques were spatially-based methods, one was a neural network based method and the last two were more recent spatio-temporal, bio-inspired vision techniques. The spatially-based techniques were the average absolute grey difference (AAGD) [16], the improved average absolute grey difference (IAAGD) [18], the high boost multiscale local contrast measure (HB-MLCM)

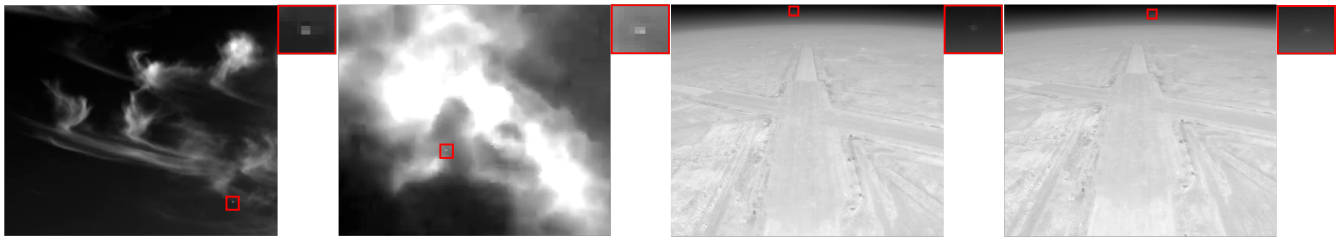

Fig. 2: The leftmost two images show examples from the SIRST dataset [51] and the rightmost two images show examples from the IREST dataset.

[17], the improved local contrast measure (ILCM) [14], the multiscale patch contrast measure (MPCM) [15], the multiscale local contrast measure (MLCM) [13], the novel local contrast measure (NLCM) [70], the relative local contrast measure (RLCM) [71], and the multiscale top hat morphological transform (TopHat) [72]. In general, each of these methods computes the contrast between an object of interest and its surrounding by a sliding window around the input image vertically and horizontally, where the window centre is meant to be the object location [13]–[18], [70], [71]. The TopHat [72] enhances objects of interest using several morphological filters of different scales. The neural network based technique used was the recent asymmetric contextual modulation (ACM) [51]. This network takes advantage of both the top-down high level semantic feedback and the reverse bottom up contextual modulation mechanism to preserve the fine details and pass them into deeper network layers [51]. The ACM network was developed for single-frame small object detection [51].

The proposed MOBIVLS was also compared against two of the latest spatio-temporal biologically inspired vision techniques. These are the centre surround total difference index (CSTDI) [42] and the bio-inspired vision model (BIV2021) [49]. CSTDI uses a temporal process, inspired by the PRC and LMC cells, applying a spatial centre-surround differential filter to the output of the LMC. This suppresses clutter and enhances object detection [42]. The BIV2021 detector uses four stages of spatio-temporal processing inspired by the PRC, LMC, RTC and ESTMD neurons [49]. It is important to highlight that these spatio-temporal methods are not really applicable to single image detection tasks so they were not tested on the first dataset as they can only be applied to high frame rate image sequences, such as the second dataset.

D. Experimental Settings

For comparison purposes, MATLAB code for the AAGD, IAAGD, HB-MLCM, ILCM, MLCM, MPCM, NLCM, TopHat and RLCM were implemented based on the information of the papers proposing these techniques. The ACM, CSTDI and BIV2021 detector implementations were obtained from the respective authors. Parameter values were tuned empirically based on the object of interest size in the two datasets (see Table I for parameter settings).

TABLE I: Parameter setting of different methods.

No.	Methods	Parameter settings
1	**AAGD** [16]	Inner window scales: 3, 5, 7, 9, 11
2	**IAAGD** [18]	pixels. Outer window scales: 21, 21,
3	**HB-MLCM** [17]	21, 21 pixels
4	**ILCM** [14]	Window scale: 10 pixels, step size: 1 pixel
5	**MLCM** [13]	Object window scales: 3, 5, 7, 9, 11
6	**MPCM** [15]	
7	**NLCM** [70]	Window scale: 12 pixels, number of maximal grey values (K): 3, step size: 1 pixel
8	**TopHat** [72]	Structuring element scales = [3, 5, 7, 9, 11, 13]
9	**RLCM** [71]	number of maximal grey values for the centre cell (K_1): [2, 5], number of maximal grey values for the other 9 cells (K_2): [4, 9], window scale (N): pixels
10	**ACM** [51]	host network: U-Net [73], Backbone: ResNet-20 [74], batch size: 8, epoch: 300, learning rate: 0.05
11	**CSTDI** [42]	Scales=[8, 16, 24], median scale fusion
12	**BIV2021** [49]	The BIV detector has a self adaptation capability to the characteristic of image sequence [49]
13	**MOBIVLS**	Object scales 3 – 11 pixels

E. Performance Metrics

To assess the performance of object enhancement and clutter suppression element of the proposed MOBIVLS and the baseline methods, a signal to clutter ratio gain (SCRG) and background suppression factor (BSF) were used [75], [76]:

$$SCRG = \frac{SCR_{out}}{SCR_{in}}, \qquad SCR = \frac{|Av_t - Av_b|}{\sigma_b}$$
$$BSF = \frac{\sigma_{in}}{\sigma_{out}} \tag{1}$$

Where SCR_{out} and SCR_{in} are the signal to clutter ratio (SCR) of the output and input images respectively, Av_t and Av_b are the average brightness of the target and background, and σ_b, σ_{in}, and σ_{out} are the standard deviations of the background, input image, and output image, respectively. For the SIRST dataset, as there are different shaped objects and a relatively accurate set of masks publicly available with the dataset, the mean of the mask area was used to compute the

average brightness of the objects, and the remainder of the image used to compute the average background brightness. For the second dataset (IREST), as the target is very small (covers less than 3×3 pixels), a mask of radius 6 pixels was placed around the centre of each object and the mean computed. Once again, the rest of the image was then used to calculate the background average.

To assess the performance of the detection element of the proposed approach and the comparative methods, the well-known receiver operating characteristic (ROC) curve was used [49], [75]. The ROC curve was created by plotting the true positive rate (TPR) against the false positive rate (FPR) for different values (range: 0 - 1) of a global detection threshold applied to the final output (saliency) map. The TPR and FPR are defined below 2:

$$\begin{aligned} True\ Positive\ Rate\ (TPR) &= \frac{TP}{A} \\ False\ Positive\ Rate\ (FPR) &= \frac{FP}{N} \end{aligned} \tag{2}$$

Where TP is the number of true positives (correct detections) in the dataset, A is the total number of actual objects of interest in the dataset, FP is the number of false positives (incorrectly detected pixels) in the dataset, and N is the total number of pixels in the dataset. For the SIRST dataset, a detection was deemed to occur correctly (TP) if the distance between the centroid of the detection and the ground truth was within the ground truth mask provided for each object. For the second dataset, since there is no segmentation mask and the objects of interest are of low contrast against their surroundings, a detection was deemed to occur correctly (TP) if the distance between the centroid of the detection and the ground truth was less than or equal to 6 pixels (which is another reason why a mask of radius 6 pixels was placed around each object). The best technique was the one with the highest TPR and lowest FPR. In other words, the best technique has its line closest to the top left-hand corner in the ROC curve [77]. However, as the ROC curve can be susceptible to outliers, the area under the ROC curve (AUROC) was also computed. The AUROC evaluates the detection rate over a range of FPR rather than just at a single value. A logarithmic x-axis (FPR) was used in the AUROC computation to give more weight to the region of the curve with low false positive values. The FPR range was limited to 0 - 10^{-4}.

IV. Results

A. Ablation Study: Performance Analysis of the MOBIVLS Stages

Figure 3 ($S1 - S9$) shows the normalised responses of the MOBIVLS stages to the horizontal and vertical pixel vectors passing through the centre of the object of interest located in the image shown in the bottom left of Figure 2 at pixel location (290, 15). It can be seen that, after each stage of the MOBIVLS, the contrast of the object with respect to the background has increased and the clutter has been suppressed; and the object of interest in the final output can be detected with high confidence by applying a simple threshold.

To evaluate the performance of each stage of the MOBIVLS algorithm, and to quantify their contribution to the final detection results, a ROC curve was generated after each stage of processing for the IREST dataset (see Figure 4). Table II shows the TPR and AUROC for each stage of the MOBIVLS method. TPR were computed at a FPR of 10^{-5} while the AUROC were computed for FPR and TPR ranges of (0 - 10^{-4}) and (0 - 1). As can be seen, each stage contributes positively to the final performance of MOBIVLS, with stage $S8$ contributing the most–a 49% and 32% improvement to the TPR and AUROC, respectively.

During the stage S8, most of the clutter is suppressed by the additive accumulation of multiple shifts and multiplications. In the previous stages (S4 & S5 and S6 & S7), the object of interest was enhanced and clutter suppressed along the vertical and horizontal dimensions, respectively; the outputs of these stages then combined during stage S8. Stage S8 cannot be applied on its own as it needs the output from the previous stages. In addition, from Figure 4 and Table II, it can seen that the scanning in the horizontal direction has provided more improvement than that in vertical direction. This was because the change between the sky and ground in a vertical slice. In other words, when scanning in the horizontal direction, there is less background change. The ROC curve for $S1$ was computed by directly applying a threshold to the input image, and this clearly shows that detections cannot be reliably obtained by simply applying a threshold to the input images when the objects of interest are dimly contrasted against their backgrounds.

TABLE II: True positive rates (TPR), area under the curve (AUROC) and incremental improvement for each stage of the proposed MOBIVLS when computed usinf the IREST dataset. The best result for each metric is displayed in bold font. The TPR were computed at a false positive rate (FPR) of 10^{-5}, while the AUROC were computed over FPR and TPR ranges of 0 - 10^{-4} and 0 - 1. Stage S8 provides the greatest incremental improvement as it suppresses most of the clutter using a process of additive accumulation of multiple shifts and multiplications. The location of the stages ($S1 - S9$) in the MOBIVLS process are depicted in Figure 1.

Stages of MOBIVLS	TPR (%)	AUROC (%)
S1	0	0
Avg. (S2, S4, S5)	0	4.78
Avg. (S3, S6, S7)	1.72	24.76
S8	50.42	56.57
S9	**57.63**	**63.31**

B. Object Detection Results

Figures 5a - 5c show the ROC curve for the MOBIVLS technique when compared against the twelve existing state of the art object detection techniques. Figure 5a shows the results for the SIRST dataset when using the original size of the images, where the average object size and the standard deviation are 44 and 60 pixels, respectively. Figure 5b shows

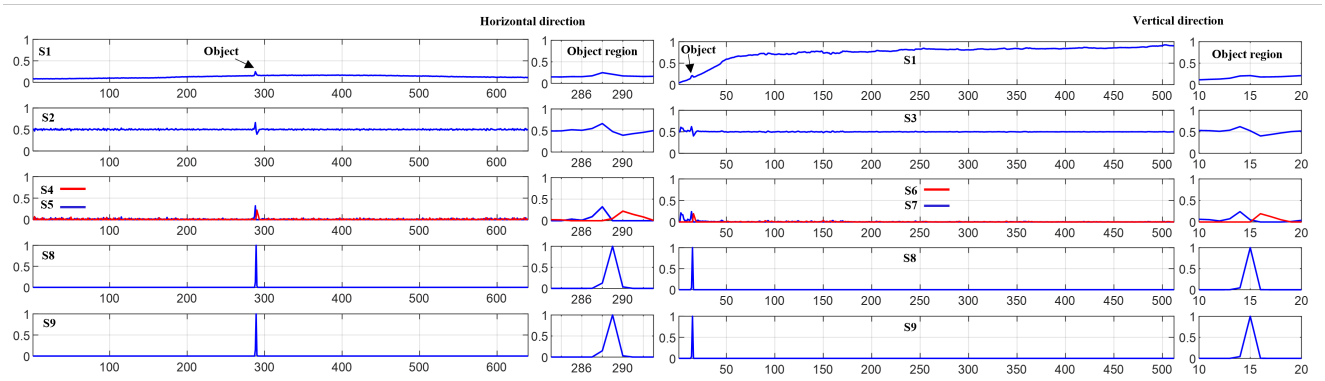

Fig. 3: Horizontal and vertical profiles of normalised responses observed at each of the MOBIVLS stages. Stages 4 and 5 are shown on a single graph. The horizontal and vertical pixel vectors pass through the centre of the object located in the image shown at the bottom left of Figure 2, at pixel location (290, 15). $S1$ shows the target input intensity. $S2$, $S3$ show the output of the BIVLS in the directions top to bottom, and left to right, respectively. $S4$, $S5$ show the ON and OFF channels of $S2$. $S6$, $S7$ show the ON and OFF channels of $S3$. $S8$ shows the output of the addition of the multi-scale shifting process. $S9$ shows the output of the Hadamard product between the input intensity $S1$ and the output of the multi-scale shifting process $S7$. $S9$ shows the final output of MOBIVLS. The location of the $S1$ – $S9$ stages in the process of the MOBIVLS are depicted in Figure 1. The object of interest region is also displayed in an enlarged format (the small images in the centre and to the right of each row) to show how its saliency has been enhanced by each stage relative to any surrounding clutter.

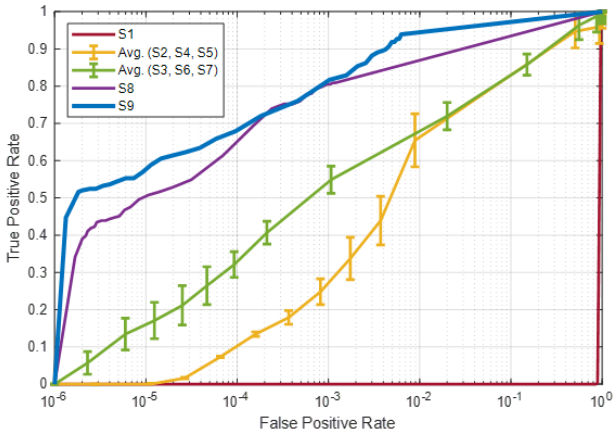

Fig. 4: Receiver operating characteristic (ROC) curves for each stage of the proposed MOBIVLS when computed using the IREST dataset. The location of the MOBIVLS stages $S1 - S9$ are depicted in Figure 1. $S1$ represents the input intensity when a threshold is applied directly to the input images. This clearly shows that no detection can reliably be obtained by simply applying a threshold. This also indicates how dim the objects of interest are.

the results for the SIRST dataset when down-sampling the images individually so that the size of the object of interest does not exceed 3×3 pixels.

Figure 5c shows the average results for the SIRST dataset from both Figures 5(a and b). The SIRST dataset has 427 images, so for the ACM network technique 50% and 20% of the images were used for training and validation, respectively. The remaining 30% (128 images) were use for testing, with the ROC curves for all techniques generated based on evaluation against this data. As the dataset was downsampled to generate a more challenging (3×3 pixel) dataset to examine the performance of the proposed and state-of-the-art techniques against lower spatial resolution imagery, it is important to note that the whole process of training and validation was repeated for the ACM technique on the down-sampled version of the dataset before applying it to the split of test images.

Figure 5d shows the ROC curves from the second dataset (IREST). The objects of interest in the second dataset are of similar size to objects of interest in the downsampled version of the IREST dataset but have lower contrast in most images. Once again, the proposed technique significantly outperformed all existing methods, both in terms of higher detection and lower false alarm rates.

Table III shows the TPR at an FPR of 10^{-5} for the two datasets. The proposed MOBIVLS outperformed all state-of-the-art object detection techniques, offering TPR of 69.1%, 77.8%, and 57.6% in the SIRST (original and downsampled images) and the IREST datasets, respectively. This compares to a TPR of 72.4% for IAAGD in the original size SIRST dataset, 65.1% for the downsampled version, and 31.1% (for CSTDI) in the IREST dataset.

Table III shows the AUROC for all techniques for an FPR integration range of 0 - 10^{-4}. MOBIVLS provided an AUROC of 76.2%, 81.0%, and 63.3% in the SIRST (original and downsampled images) and IREST datasets, respectively. This compares to AUROC values of 77.1% for IAAGD in the SIRST-original size dataset, 69.7% for ACM in the downsampled version of the dataset, and 38.5% for CSTDI in the IREST dataset. It is important to mention that CSTDI and BIV2021 are spatio-temporal methods, cannot be usefully applied to single image detection tasks, i.e. they can only be applied to high frame rate image sequences such as in the second dataset.

From Table III, it can also be seen that there were some

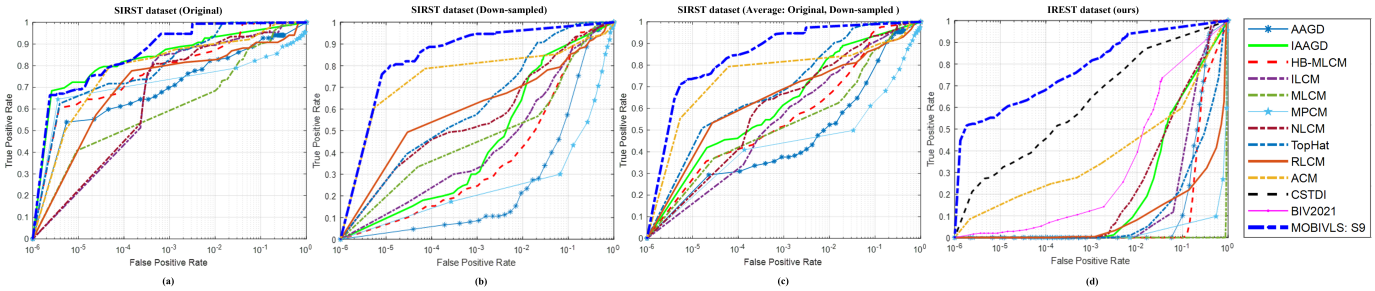

Fig. 5: Receiver operating characteristic (ROC) curves for the proposed MOBIVLS method, nine existing spatially-based single-frame object detection methods (AAGD, IAAGD, HB-MLCM, ILCM, MLCM, MPCM, NLCM, TopHat and RLCM), the neural network based method (ACM), and two spatio-temporal bio-inspired vision techniques (CSTDI and BIV2021). Figure (a) shows the ROC curves for the SIRST dataset with original image sizes, (b) shows the ROC curves for the down-sampled version of the SIRST dataset, where the size of the objects of interest do not exceed 3×3 pixels, (c) shows the average ROC curves for the SIRST dataset when using original and down-sampled image sizes, and (d) shows the ROC curves from the IREST dataset. It is important to highlight that the CSTDI and BIV2021 made use of the temporal nature of the IREST image sequence and hence they only shown in Figure (d) as they cannot process single images. It can also be seen that the MOBIVLS consistently ranked highest in all tests with the exception of the SIRST dataset (original size) at low FPR, where it was second.

improvements in the performance of some techniques when comparing the downsampled and original size images from the first dataset, i.e. 5.6% in TPR and 14.0% in AUROC for NLCM, 6.3% in TPR but 0.67% decrease in AUROC for ACM, and 8.7 % in TPR and 4.8 % in AUROC for MOBIVLS. This was because in the downsampled versions of the images, some false positives were not detected by these techniques. However for most techniques there was no improvement between the downsampled and original imagery as they missed a significant number of detections due to the contrast-diminishing effect of downsampling. It can also be seen that all of the baseline techniques of the spatial only saliency target enhancement category (AAGD, IAAGD, HB-MLCM, ILCM, MLCM, MPCM, NLCM, TopHat, RLCM) achieved 0% TPR and AUROC ¡ 4×10^{-2}), highlighting just how challenging the second dataset scenario is.

Table III shows the average processing time for each implementation of the comparative techniques versus the proposed MOBIVLS method. The average processing time for the unoptimised MATLAB implementation of MOBIVLS was 2.136 microseconds per pixel (ms/p), quicker than several techniques but generally slow. It should be noted, however, that the BIVLS and spatial shift function of MOBIVLS could be coded in such a manner as to exploit parallel processing pipelines (the current code does not do this), and this parallelisation would significantly improve processing times. It is important to mention that as the size of the images differ for each dataset, it was more convenient to present the processing times in microseconds per pixel (μs/p).

The first two columns of Figure 6 show the results of processing the two representative image samples from the SIRST dataset (Figure 2, leftmost two images) using the proposed MOBIVLS, the nine comparative spatial single-frame based detection methods (AAGD, IAAGD, HB-MLCM, ILCM, MLCM, MPCM, NLCM, TopHat and RLCM), and the ACM neural network. It can be seen that the object of interest was completely missed in the detection maps of the AAGD, IAAGD, HB-MLCM, ILCM, MPCM, and NLCM techniques. Moreover, many false alarms were generated by these methods. The other existing methods either missed or could not clearly detect the objects of interest, and generated a large number of false alarms. In contrast, the proposed MOBIVLS technique detected the objects of interest and suppressed the clutter in all the images, providing clearly visible detections against minimal background clutter.

The third and fourth columns of Figure 6 show the results of processing the two representative image samples of the IREST dataset (Figure 2, rightmost two images) using the proposed MOBIVLS, the nine comparative spatial single-frame based detection methods (AAGD, IAAGD, HB-MLCM, ILCM, MLCM, MPCM, NLCM, TopHat and RLCM), the ACM neural network, and the two spatio-temporal techniques, CSTDI and BIV2021. Again, the proposed MOBIVLS technique detected the objects of interest and suppressed the clutter in both images, while the other techniques either completely missed them or detected them with low confidence, i.e. with a large number of false alarms.

C. Object Contrast Enhancement Results

To evaluate the performance of MOBIVLS and the existing object detection techniques in terms of background suppression and signal versus background clutter improvement, the background suppression factor (BSF) and signal-to-clutter ratio gain (SCRG) were computed for each technique using all of the images in both datasets. The results are presented in Table IV. The MOBIVLS technique outperformed all other methods against both metrics, with a SCRG and BSF of 19.5 and 42.2 (in the SIRST-original size dataset), 17.3 and 26.5 (in the SIRST-down-sampled version dataset), and 5.7 and 71.3 (in the IREST dataset), respectively. This compares to SCRG of

TABLE III: True positive rates (TPR), area under the curve (AUROC) and processing time (seconds per frame) for the object detection techniques examined in this study. The data was computed using the SIRST and IREST datasets. The best result for each metric is highlighted using an underline and a bold font. The second best result uses just bold font. The TPR were computed at a false positive rate (FPR) of 10^{-5}, while the AUROC were computed over FPR and TPR ranges of 0 - 10^{-4} and 0 - 1, respectively. As the sizes of frames differ for each dataset, processing time was presented in units of microseconds per pixel (μs/p).

No.	Methods	SIRST dataset				IREST dataset		Time (ms/p)
		Original size		Down sampled				
		TPR (%)	AUROC (%)	TPR (%)	AUROC (%)	TPR (%)	AUROC (%)	
1	**AAGD** [16]	54.3	56.88	3.1	3.96	0	9.34×10^{-4}	0.122
2	**IAAGD** [18]	**72.37**	**77.11**	9.8	13.77	0	5.47×10^{-3}	0.397
3	**HB-MLCM** [17]	62.52	64.55	7.33	10.76	0	7.66×10^{-5}	0.153
4	**ILCM** [14]	22.5	22.28	11.51	16.7	0	3.27×10^{-4}	0.336
5	**MLCM** [13]	40.79	38.8	20	24.94	0	5.32×10^{-4}	0.854
6	**MPCM** [15]	64.94	66.15	7.34	7.06	0	1.73×10^{-5}	0.448
7	**NLCM** [70]	21.7	21.48	27.33	35.47	0	9.2×10^{-3}	2.838
8	**TopHat** [72]	63.84	68.08	27.33	36.1	0	3.9×10^{-3}	4.425
9	**RLCM** [71]	41.4	56.72	33.81	43.5	0	3.22×10^{-2}	15.02
10	**ACM** [51]	58.8	70.34	**65.1**	**69.67**	15.2	19.03	1.221
11	**CSTDI** [42]	n/a	n/a	n/a	n/a	**31.13**	**38.49**	0.977
12	**BIV2021** [49]	n/a	n/a	n/a	n/a	2.4	4.11	0.427
13	**MOBIVLS** (ours)	**69.1**	**76.24**	**77.82**	**81.04**	**57.63**	**63.31**	2.136

18.0 for IAAGD in the SIRST original dataset, 11.5 for RLCM in the SIRST downsampled dataset, and 2.1 for CSTDI in the IREST dataset. The next best BSF results are 27.1 and 10.8 for IAAGD in the SIRST original and downsampled datasets, and 27.3 for BIV2021 in the IREST dataset.

TABLE IV: Results of the signal to clutter ratio gain (SCRG) and background suppression factor (BSF) for different object detection techniques when computed over the two datasets. The best result for each metric is highlighted using an underline and a bold style. The second best result is written in bold style alone.

Methods	SIRST dataset				IREST dataset	
	Original size		Down sampled			
	SCRG	BSF	SCRG	BSF	SCRG	BSF
AAGD [16]	12.24	8.53	2.47	3.08	0.1	10.84
IAAGD [18]	**18**	**27.08**	8.07	**10.81**	0.04	12.65
HB-MLCM [17]	16.35	17.17	7.21	6.48	0.09	10.56
ILCM [14]	12.83	7.25	6.83	2.28	0.21	4.5
MLCM [13]	4.23	2.44	5.66	1.47	0.86	1.05
MPCM [15]	4.6	5.57	1.39	2.48	0.87	3.98
NLCM [70]	9.39	10.98	7.44	3.43	0.56	5.46
TopHat [72]	9.29	6.35	3.64	3.13	0.5	1.44
RLCM [71]	15.71	6.43	**11.47**	2.7	0.6	1.61
ACM [51]	3.24	6.24	2.03	5.04	0.3	5.87
CSTDI [42]	n/a	n/a	n/a	n/a	**2.1**	8.97
BIV2021 [49]	n/a	n/a	n/a	n/a	0.09	**27.29**
MOBIVLS (ours)	**19.54**	**42.22**	**17.32**	**26.52**	**5.69**	**71.32**

V. Discussion

The results show that the proposed MOBIVLS small target detection technique was able to successfully detect small objects in low contrast and high clutter scenarios. Moreover, this approach substantially outperformed all other state-of-the-art single frame based small dim object detection methods examined in this study, as well as two spatio-temporal bio-inspired vision techniques applied to high frame sequence imagery.

Spatially based object detection techniques, such as the AAGD [16], IAAGD [18], HB-MLCM [17], ILCM [14], MLCM [13], MPCM [15], NLCM [70], TopHat [72], and RLCM [71], have been extensively tested in the literature. They have been shown to be capable of detecting objects such as planes, drones and boats, even when small and of low contrast. However, although these objects were dim, the background intensities were relatively even and did not comprise heavy clutter. Relative to the results reported in the literature, the performance of the algorithms appears to have decreased drastically when tested on our IREST dataset, which contained both heavy clutter and very small targets. Indeed, all of these algorithms provided negligible AUROC values (¡ 4×10^{-2}) compared to 63.3% for the proposed MOBIVLS. In the overall results (i.e. the average of the IREST and original and downsampled images from the first dataset), the best performing spatial technique (TopHat) achieved only 34.7% AUROC compared to 73.53% for the proposed MOBIVLS; and while the ACM neural network showed better results than the spatially based approaches, it only achieved an overall AUROC of 53.0%.

Biologically inspired vision techniques have proven their ability to detect small, even micro (single pixel), objects in low contrast, high clutter environments [36], [39], [40], [42], [44], [47] whist simultaneously dynamically adapting to their environments. However, while the dynamic adaptation within the existing bio-inspired techniques is based on the temporal variation of an image sequence, the adaptation in MOBIVLS is based solely on each individual frame. Thus, there is no longer a need for high frame rate image sequences to obtain a high probability of detection at low rates of false alarm. Moreover, even for high frame rate data, if the recording platform moves somewhat erratically, as is often the case for drones flying in windy weather, the recorded image sequences can still suffer from optical flow jitter, which can degrade the performance of existing bio-inspired vision techniques as they require smooth optic flow sequences to perform optimally [39], [40], [42], [47]. That said, it should be noted that the purpose

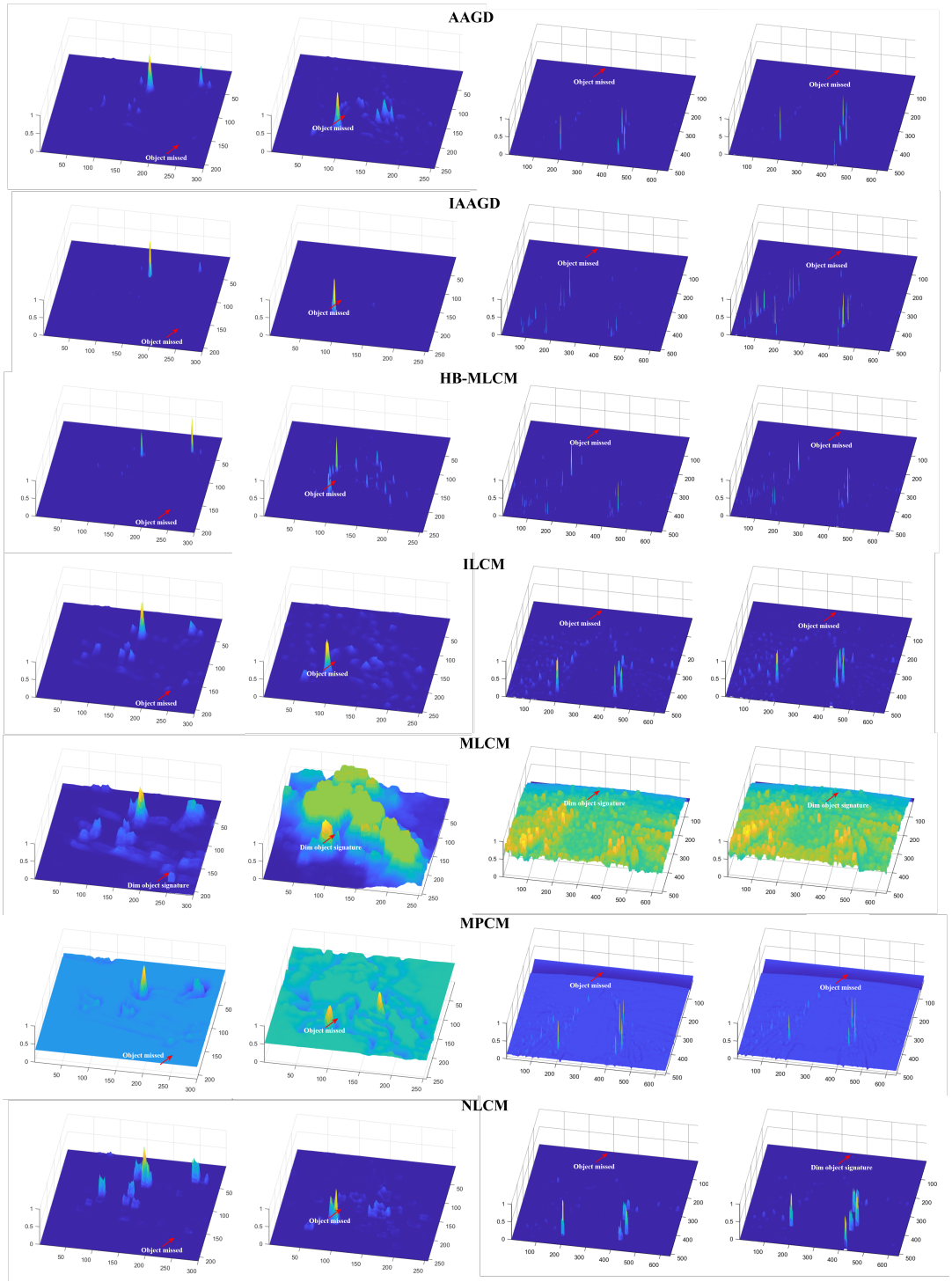

Fig. 6: Detection maps generated using thirteen state-of-the-art object detection techniques, nine spatial single-frame-based object detection methods (AAGD, IAAGD, HB-MLCM, ILCM, MLCM, MPCM, NLCM, TopHat and RLCM), an ACM neural network, two comparable spatio-temporal biologically inspired vision techniques (CSTDI and BIV2021), and the proposed MOBIVLS method. The results are for the four images shown in Figure 2. For each detection method, more false detections were generated and/or the object of interest response was of lower contrast than the proposed MOBIVLS method (Images best viewed enlarged in digital format).

of MOBIVLS is not to replace techniques such as CSTDI and BIV2021 as these techniques have shown their effectiveness in terms of detecting even smaller objects (less than 3×3 pixels), including those recorded from stationary platforms [42], [49].

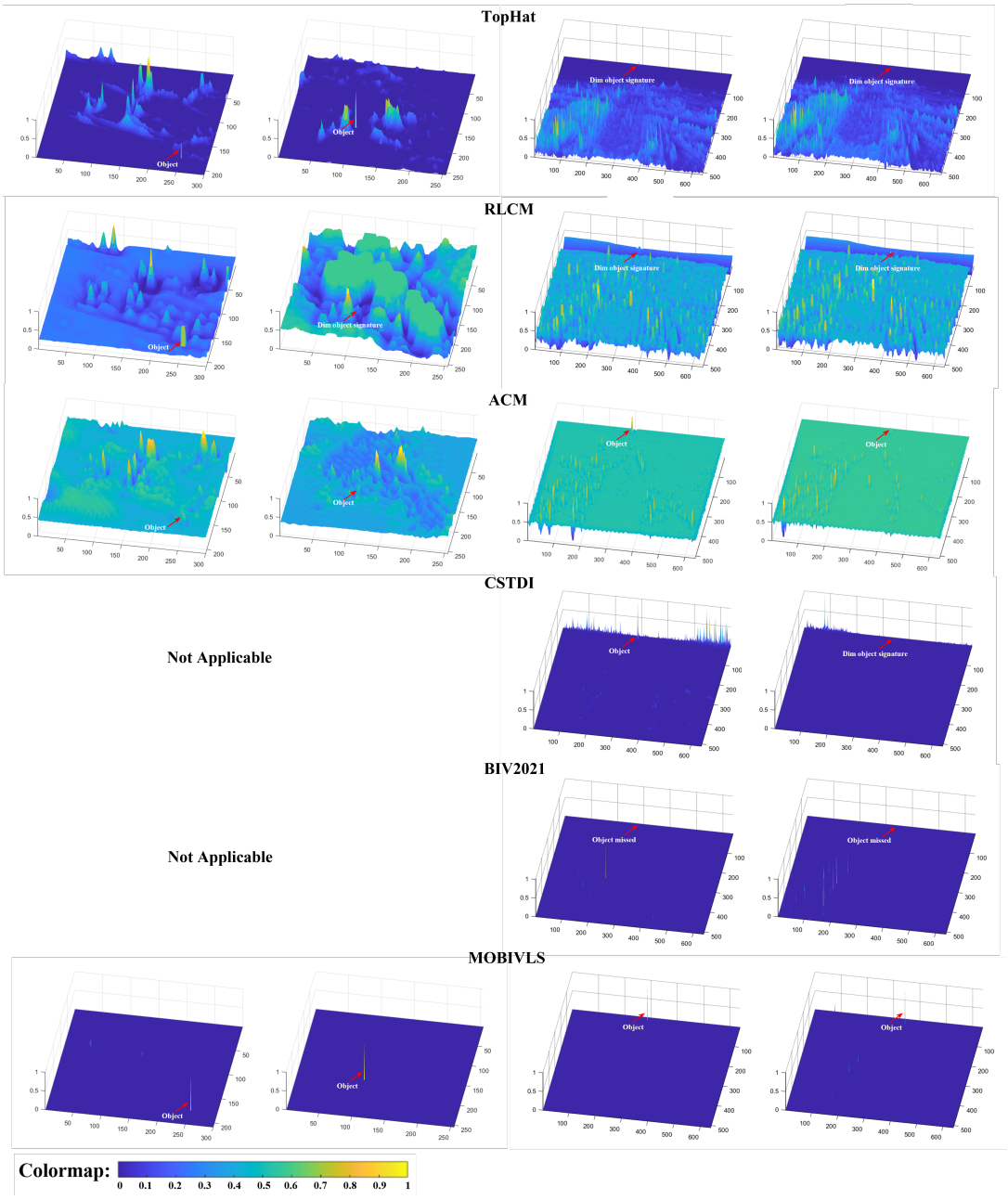

Fig. 6: (continued)

The main goal of this research was to develop a small target detection technique that can perform successfully in scenarios that CSTDI and BIV2021 are either not applicable to, such as single frames, or not good at, such as when the relative motion between an object of interest and the recording platform is too high and the temporal information lost between successive frames is too great, i.e. where these techniques rely heavily on high frame rate temporal information.

In [49], results show that BIV2021 and CSTDI provide TPR of 63% and 44% at FPR of 10^{-5}. This is a 19% increase in TPR for the BIV2021 over CSTDI. However, when testing BIV2021 and CSTDI on the IREST dataset, the performance of CSTDI was better than BIV2021 by 29% TPR. The reason is because CSTDI is less dependent on temporal information and thus less impacted by the effect of optical flow jitter in the image sequences of the IREST dataset. In [49] there was no optical flow jitter as the recording platform was stationary. It is also important to highlight that all of the stages (PRC, LMC, RTC and ESTMD) of the BIV2021 model require temporal information, whereas in CSTDI only the first two stages (PRC and LMC) need temporal information; and the final output is generated by applying a spatial filter to the output of the LMC stage.

VI. Conclusion

This paper proposes a biologically inspired vision processing technique for detecting low contrast objects in high clutter environments. The method is known as multi-scale object bio-inspired vision line scanners or MOBIVLS. It is inspired by the early stages of the insect vision system, specifically the PRC, LMC, RTC and a computational model of the EMD derived from the BIV2021 approach. Unlike other existing bio-inspired techniques, MOBIVLS can process an image sequence of any frame rate and even still imagery. To evaluate the technique's performance two infrared datasets were used, each containing different environments and objects. The results showed the proposed technique significantly outperformed all other dim/small object detection methods examined, whether based on spatial, spatio-temporal (bio-inspired), or neural network processing. The MOBIVLS offered an AUROC of 73.53% overall as opposed to the next best technique, neural networks (53.01%), the next best spatial method (34.73%), or the next best spatio-temporal method (38.49%). MOBIVLS could be adapted to a range of different fields and applications that require objects to be detected in complex environments such as medical imaging, seismology, and astronomy, where detection of low contrast objects in cluttered environments is a priority.

Acknowledgement

Laith Al-Shimaysawee is supported by the University of South Australia (UniSA) President's Scholarship and Research Training Program. The authors would like to thank the Australian Department of Defence for allowing access to the site where some of the data was recorded. The authors would also like to thank everyone involved in the field trials, especially the Defence team and Steven Andriolo of EyeSky.

References

[1] B. Mishra, D. Garg, P. Narang, and V. Mishra, "Drone-surveillance for search and rescue in natural disaster," *Computer Communications*, vol. 156, pp. 1–10, 2020.

[2] M. Dogariu, L.-D. Stefan, M. G. Constantin, and B. Ionescu, "Human-object interaction: Application to abandoned luggage detection in video surveillance scenarios," in *2020 13th International Conference on Communications (COMM)*. IEEE, 2020, pp. 157–160.

[3] Z. Domozi, D. Stojcsics, A. Benhamida, M. Kozlovszky, and A. Molnar, "Real time object detection for aerial search and rescue missions for missing persons," in *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*. IEEE, 2020, pp. 000 519–000 524.

[4] I. Karakostas, I. Mademlis, N. Nikolaidis, and I. Pitas, "Uav cinematography constraints imposed by visual target tracking," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 76–80.

[5] L. M. Dang, S. I. Hassan, I. Suhyeon, A. kumar Sangaiah, I. Mehmood, S. Rho, S. Seo, and H. Moon, "Uav based wilt detection system via convolutional neural networks," *Sustainable Computing: Informatics and Systems*, 2018.

[6] L. Wallace, A. Lucieer, C. Watson, and D. Turner, "Development of a uav-lidar system with application to forest inventory," *Remote sensing*, vol. 4, no. 6, pp. 1519–1543, 2012.

[7] B. Aydin, E. Selvi, J. Tao, and M. J. Starek, "Use of fire-extinguishing balls for a conceptual system of drone-assisted wildfire fighting," *Drones*, vol. 3, no. 1, p. 17, 2019.

[8] A. N. Jensen and M. Jensen, "A research platform for drone assisted firefighting with ar path visualisations," 2020.

[9] K. Rogers and A. Finn, "Three-dimensional uav-based atmospheric tomography," *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 2, pp. 336–344, 2013.

[10] M. Kratky and J. Farlik, "Countering uavs-the mover of research in military technology." *Defence Science Journal*, vol. 68, no. 5, 2018.

[11] J. Nygårds, P. Skoglar, M. Ulvklo, and T. Högström, "Navigation aided image processing in uav surveillance: Preliminary results and design of an airborne experimental system," *Journal of Robotic Systems*, vol. 21, no. 2, pp. 63–72, 2004.

[12] D. Gettinger and A. H. Michel, "Drone sightings and close encounters: An analysis," *Center for the Study of the Drone, Bard College, Annandale-on-Hudson, NY, USA*, 2015.

[13] C. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, 2013.

[14] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2168–2172, 2014.

[15] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognition*, vol. 58, pp. 216–226, 2016.

[16] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 60–72, 2016.

[17] Y. Shi, Y. Wei, H. Yao, D. Pan, and G. Xiao, "High-boost-based multiscale local contrast measure for infrared small target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 33–37, 2017.

[18] S. Aghaziyarati, S. Moradi, and H. Talebi, "Small infrared target detection using absolute average difference weighted by cumulative directional derivatives," *Infrared Physics & Technology*, vol. 101, pp. 78–87, 2019.

[19] T. Liu, Q. Yin, J. Yang, Y. Wang, and W. An, "Combining deep denoiser and low-rank priors for infrared small target detection," *Pattern Recognition*, vol. 135, p. 109184, 2023.

[20] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 4996–5009, 2013.

[21] Y. Qin, L. Bruzzone, C. Gao, and B. Li, "Infrared small target detection based on facet kernel and random walker," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7104–7118, 2019.

[22] Y. Lu, L. Dong, T. Zhang, and W. Xu, "A robust detection algorithm for infrared maritime small and dim targets," *Sensors*, vol. 20, no. 4, p. 1237, 2020.

[23] S. Kim and J. Lee, "Small infrared target detection by region-adaptive clutter rejection for sea-based infrared search and track," *Sensors*, vol. 14, no. 7, pp. 13 210–13 242, 2014.

[24] G. Chen and W. Wang, "Target recognition in infrared circumferential scanning system via deep convolutional neural networks," *Sensors*, vol. 20, no. 7, p. 1922, 2020.

[25] B. Wang, Y. Motai, L. Dong, and W. Xu, "Detecting infrared maritime targets overwhelmed in sun glitters by antijitter spatiotemporal saliency," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5159–5173, 2019.

[26] T. Wu, B. Li, Y. Luo, Y. Wang, C. Xiao, T. Liu, J. Yang, W. An, and Y. Guo, "Mtu-net: Multi-level transunet for space-based infrared tiny ship detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[27] H. Wang, L. Zhou, and L. Wang, "Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8509–8518.

[28] F. Chen, C. Gao, F. Liu, Y. Zhao, Y. Zhou, D. Meng, and W. Zuo, "Local patch network with global attention for infrared small target detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 5, pp. 3979–3991, 2022.

[29] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9813–9824, 2021.

[30] H. Fang, M. Xia, G. Zhou, Y. Chang, and L. Yan, "Infrared small uav target detection based on residual image prediction via global and local dilated residual networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[31] B. Li, C. Xiao, L. Wang, Y. Wang, Z. Lin, M. Li, W. An, and Y. Guo, "Dense nested attention network for infrared small target detection," *IEEE Transactions on Image Processing*, 2022.
[32] T. Zhang, L. Li, S. Cao, T. Pu, and Z. Peng, "Attention-guided pyramid context networks for detecting infrared small target under complex background," *IEEE Transactions on Aerospace and Electronic Systems*, 2023.
[33] L. Huang, S. Dai, T. Huang, X. Huang, and H. Wang, "Infrared small target segmentation with multiscale feature representation," *Infrared Physics & Technology*, vol. 116, p. 103755, 2021.
[34] C. Yu, Y. Liu, S. Wu, Z. Hu, X. Xia, D. Lan, and X. Liu, "Infrared small target detection based on multiscale local contrast learning networks," *Infrared Physics & Technology*, vol. 123, p. 104107, 2022.
[35] R. Li and Y. Shen, "Yolosr-ist: A deep learning method for small target detection in infrared remote sensing images based on super-resolution and yolo," *Signal Processing*, p. 108962, 2023.
[36] S. D. Wiederman, R. S. Brinkworth, and D. C. O'Carroll, "Bio-inspired target detection in natural scenes: optimal thresholds and ego-motion," in *Biosensing*, vol. 7035. International Society for Optics and Photonics, 2008, p. 70350Z.
[37] J. R. Serres and S. Viollet, "Insect-inspired vision for autonomous vehicles," *Current opinion in insect science*, vol. 30, pp. 46–51, 2018.
[38] R. S. Brinkworth and D. C. O'Carroll, "Robust models for optic flow coding in natural scenes inspired by insect biology," *PLoS Comput Biol*, vol. 5, no. 11, p. e1000555, 2009.
[39] D. Griffiths, T. Scoleri, R. S. Brinkworth, and A. Finn, "Pixel-wise infrared tone remapping for rapid adaptation to high dynamic range variations," in *Electro-Optical and Infrared Systems: Technology and Applications XVI*, vol. 11159. International Society for Optics and Photonics, 2019, p. 111590V.
[40] M. Uzair, R. S. Brinkworth, and A. Finn, "Insect-inspired small moving target enhancement in infrared videos," in *2019 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2019, pp. 1–8.
[41] S. Poursoltan, R. Brinkworth, and M. Sorell, "Biologically-inspired pre-compression enhancement of video for forensic applications," in *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*. IEEE, 2013, pp. 1–6.
[42] M. Uzair, R. S. Brinkworth, and A. Finn, "A bio-inspired spatiotemporal contrast operator for small and low-heat-signature target detection in infrared imagery," *Neural Computing and Applications*, pp. 1–14, 2020.
[43] R. S. Brinkworth, E.-L. Mah, J. P. Gray, and D. C. O'Carroll, "Photoreceptor processing improves salience facilitating small target detection in cluttered scenes," *Journal of vision*, vol. 8, no. 11, pp. 8–8, 2008.
[44] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "A model for the detection of moving targets in visual clutter inspired by insect physiology," *PloS one*, vol. 3, no. 7, p. e2784, 2008.
[45] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Properties of neuronal facilitation that improve target tracking in natural pursuit simulations," *Journal of The Royal Society Interface*, vol. 12, no. 108, p. 20150083, 2015.
[46] P. S. Skelton, A. Finn, and R. S. Brinkworth, "Consistent estimation of rotational optical flow in real environments using a biologically-inspired vision algorithm on embedded hardware," *Image and Vision Computing*, vol. 92, p. 103814, 2019.
[47] A. Melville-Smith, A. Finn, and R. S. Brinkworth, "Enhanced micro target detection through local motion feedback in biologically inspired algorithms," in *2019 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2019, pp. 1–8.
[48] S. Poursoltan, R. Brinkworth, and M. Sorell, "Biologically-inspired video enhancement method for robust shape recognition," in *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*. IEEE, 2013, pp. 1–6.
[49] M. Uzair, R. S. Brinkworth, and A. Finn, "Detecting small size and minimal thermal signature targets in infrared imagery using biologically inspired vision," *Sensors*, vol. 21, no. 5, p. 1812, 2021.
[50] Z. M. Bagheri, B. S. Cazzolato, S. Grainger, D. C. O'Carroll, and S. D. Wiederman, "An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments," *Journal of neural engineering*, vol. 14, no. 4, p. 046030, 2017.
[51] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 950–959.
[52] M. Long, S. Cong, H. Shanshan, W. Zoujian, W. Xuhao, and W. Yanxi, "Sddnet: Infrared small and dim target detection network," *CAAI Transactions on Intelligence Technology*, 2023.
[53] Y. Bai, R. Li, S. Gou, C. Zhang, Y. Chen, and Z. Zheng, "Cross-connected bidirectional pyramid network for infrared small-dim target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
[54] H. Fang, Z. Liao, X. Wang, Y. Chang, and L. Yan, "Differentiated attention guided network over hierarchical and aggregated features for intelligent uav surveillance," *IEEE Transactions on Industrial Informatics*, 2023.
[55] K. Wang, S. Du, C. Liu, and Z. Cao, "Interior attention-aware network for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
[56] J. Van Hateren, "Processing of natural time series of intensities by the visual system of the blowfly," *Vision research*, vol. 37, no. 23, pp. 3407–3416, 1997.
[57] J. Van Hateren and H. Snippe, "Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells," *Vision research*, vol. 41, no. 14, pp. 1851–1865, 2001.
[58] R. S. Brinkworth, E.-L. Mah, and D. C. O'Carroll, "Bio-inspired pixel-wise adaptive imaging," in *Smart Structures, Devices, and Systems III*, vol. 6414. International Society for Optics and Photonics, 2007, p. 641416.
[59] M. Juusola, E. Kouvalainen, M. Järvilehto, and M. Weckström, "Contrast gain, signal-to-noise ratio, and linearity in light-adapted blowfly photoreceptors." *The Journal of general physiology*, vol. 104, no. 3, pp. 593–621, 1994.
[60] R. V. Frolov and I. I. Ignatova, "Electrophysiological adaptations of insect photoreceptors and their elementary responses to diurnal and nocturnal lifestyles," *Journal of Comparative Physiology A*, vol. 206, no. 1, pp. 55–69, 2020.
[61] I. I. Ignatova and R. V. Frolov, "Distinct mechanisms of light adaptation of elementary responses in photoreceptors of dipteran flies and american cockroach," *Journal of Neurophysiology*, vol. 128, no. 1, pp. 263–277, 2022.
[62] J. H. van Hateren, "A theory of maximizing sensory information," *Biological cybernetics*, vol. 68, no. 1, pp. 23–29, 1992.
[63] M. S. Drews, A. Leonhardt, N. Pirogova, F. G. Richter, A. Schuetzenberger, L. Braun, E. Serbe, and A. Borst, "Dynamic signal compression for robust motion vision in flies," *Current Biology*, vol. 30, no. 2, pp. 209–221, 2020.
[64] M. V. Srinivasan, R. Pinter, and D. Osorio, "Matched filtering in the visual system of the fly: large monopolar cells of the lamina are optimized to detect moving edges and blobs," *Proceedings of the Royal Society of London. B. Biological Sciences*, vol. 240, no. 1298, pp. 279–293, 1990.
[65] M. Juusola, M. Weckstrom, R. Uusitalo, M. Korenberg, and A. French, "Nonlinear models of the first synapse in the light-adapted fly retina," *Journal of neurophysiology*, vol. 74, no. 6, pp. 2538–2547, 1995.
[66] K. Haltis, M. J. Sorell, and R. Brinkworth, "A biologically inspired smart camera for use in surveillance applications," in *Crime Prevention Technologies and Applications for Advancing Criminal Investigation*. IGI Global, 2012, pp. 188–201.
[67] R. A. Horn, "The hadamard product," in *Proc. Symp. Appl. Math*, vol. 40, 1990, pp. 87–169.
[68] S. Wiederman, R. S. Brinkworth, and D. C. O'Carroll, "Performance of a bio-inspired model for the robust detection of moving targets in high dynamic range natural scenes," *Journal of Computational and Theoretical Nanoscience*, vol. 7, no. 5, pp. 911–920, 2010.
[69] "infrared camera ici 8640-p-series," https://infraredcameras.com/product/8640-p-series/, accessed: 2022-11-29.
[70] Y. Qin and B. Li, "Effective infrared small target detection utilizing a novel local contrast method," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1890–1894, 2016.
[71] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, "Infrared small target detection utilizing the multiscale relative local contrast measure," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 612–616, 2018.
[72] 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, no. 1, pp. 67–76, 2006.
[73] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[75] P. Lv, S. Sun, C. Lin, and G. Liu, "A method for weak target detection based on human visual contrast mechanism," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 261–265, 2018.
[76] C. Gao, L. Wang, Y. Xiao, Q. Zhao, and D. Meng, "Infrared small-dim target detection based on markov random field guided noise modeling," *Pattern Recognition*, vol. 76, pp. 463–475, 2018.
[77] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.