

A robust FLIR target detection employing an auto-convergent pulse coupled neural network

Maitreyee Dey^a, Soumya Prakash Rana^a, Patrick Siarry^b

^aSchool of Engineering, London South Bank University, London, UK;

^bFaculty of Science and Technology, University of Paris-Est Creteil, Paris, France

ARTICLE HISTORY

Compiled March 15, 2019

ABSTRACT

Automatic target detection (ATD) of a small target along with its true shape from highly cluttered forward-looking infrared (FLIR) imagery is crucial. FLIR imagery is low contrast in nature, which makes it difficult to discriminate the target from its immediate background. Here, pulse coupled neural network (PCNN) is extended with auto-convergent criteria to provide an efficient ATD tool. The proposed auto-convergent PCNN (AC-PCNN) segments the target from its background in an adaptive manner to identify the target region when the target is camouflaged or contains higher visual clutter. Then, selection of region of interest followed by template matching are augmented to capture the accurate shape of a target in a real scenario. The outcomes of proposed method are validated through well-known statistical methods and found superior performance over other conventional methods.

1. Introduction

Automatic target detection (ATD) is able to identify targets automatically based on the obtained data from different sensor such as RADAR, Synthetic-Aperture Radar (SAR), Forward-Looking Infrared (FLIR), etc Bhanu (1986). These sensors are mostly set very far from the original object hence there is a high possibility of affecting the data by massive clutters of the environmental hazard. Therefore, the data produced by these sensors are tremendously rigorous and hard to find out the target by human expert in a real-time manner. The ATD system using FLIR is able to locate and identify suspected objects at night, where infrared wavelength plays a pivotal role though identifying small targets from it is a challenging task Gao, Zhang, and Li (2012); Shi et al. (2018). FLIR sensor can see through in various atmospheric conditions more specifically, it captures the heat signature from an object, which even makes easier to detect camouflaged object. Thus, this system widely used in military field as most of surveillance and targeting occurs at darkness or low lighting environments.

The rapid expansion of the practical applications, traditional methods of ATD for small target detection like top hat Bai and Zhou (2010), max-mean or median filter Deshpande et al. (1999) and morphological filter Shaoa, Zhua, and Liub (2008) are vastly used to deal with background clutter. In addition, a small target detection

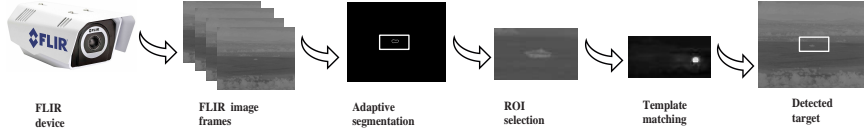


Figure 1.: The framework of proposed ATD work.

technique based on sparse ring representation (SRR) is offered in Gao, Zhang, and Li (2012). SRR can define the difference between the targets and background through its active graphical structure. Meanwhile, the feature based approaches such as, Zernike moment, multi-feature, fusion based physical relevant feature, etc. are focused on invariance properties, robustness, and powerful discriminatory of the target feature for ATD. These are mostly applied to SAR imagery Clemente et al. (2015); Kim, Song, and Kim (2018). So far, numerous methods based on neural network Cheng, Zhou, and Han (2016), and manifold learning Li et al. (2010) are applied for target detection. However, small target detection from cluttered background is still very burdensome. In recent times, bio-inspired processes like the hierarchical model Serre et al. (2007) and derived kernel Li et al. (2012) are getting more attention. Visual cortex model (VCM) is another physiological motivated information processing prototype that makes predictions and also acts like scientific eye. This helped to design a neural network named PCNN Subashini and Sahoo (2014), an effective simulation tool for learning synchronous dynamic pulse that can solve the above-mentioned challenges in an adaptive way. The pivotal challenge of the PCNN is to decide the convergence criteria through experimental and/or manual procedure, which is problem specific till now and results huge time and space complexity burden. In this paper, a robust and adaptive ATD is proposed by employing improved PCNN to solve both of the purposes: target detection along with clutter rejection. This work is introducing an auto-convergence criteria to extend the segmentation potentiality of PCNN and create powerful ATD tool from low-contrast imagery in an adaptive way. Subsequently, proposed algorithm is validated through statistical measurements and obtained significant detection performance over three other techniques.

2. Proposed methodology

Here, PCNN model has been modified to propose an automatically converged pulse coupled neural network (AC-PCNN) that provides a robust ATD tool. The flow chart of the proposed work is shown in Figure 1. The proposed ATD process has been accomplished by following these consecutive stages: data processing from the FLIR image sequence, adaptive segmentation through PCNN, ROI selection, and template matching.

2.1. Auto convergent - pulse coupled neural network

In Eckhorn et al. (1990), Eckhorn et al. first developed a neural model on cat visual cortex that produce binary images from the visual impression. Then, this network had been modified and developed to create an image by effective simulation of a synchronous behaviour Subashini and Sahoo (2014). PCNN is a third generation, single layer, two-dimensional, and laterally connected neural network. PCNN is consists of three parts: input, modulation, and pulse generator. In the input part, each neuron

receives signals through feeding ($F_{i,j}$) and linking ($L_{i,j}$) channel, where i, j stands for location of the neurons of the image in n^{th} iteration. Feeding consists of external sources ($S_{i,j}$) which is the normalized pixel value of input image. Linking is being represented by the constant synaptic weights from neuron ($w_{i,j,k,l}$) and linked with its specified neighbour neurons refer as $N_{i,j}$, where k, l indicate the location of the neighbour neuron of i, j . The modulation part, is the nonlinear combination of both feeding and linking signals through the linking coefficient β in internal activity ($U_{i,j}$). Finally, the pulse generator part produces pulse to fire neurons using an adaptive threshold variable ($T_{i,j}$) as a step function to control the firing event. After thresholding it produces the pulse output $O_{i,j}$ based on this mathematical models shown in Eqs. (1) to (5).

$$F_{i,j}[n] = S_{i,j}, \quad (1)$$

$$L_{i,j}[n] = \sum_{k,l \in N_{i,j}} w_{i,j,k,l} O_{k,l}[n-1], \quad (2)$$

$$U_{i,j}[n] = F_{i,j}[n](1 + \beta L_{i,j}[n]), \quad (3)$$

$$O_{i,j}[n] = \begin{cases} 1 & \text{if } U_{i,j}[n] > T_{i,j}[n-1], \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

$$T_{i,j}[n] = e^{-\alpha_T} T_{i,j}[n-1] + V_T O_{i,j}[n-1], \quad (5)$$

where, α_T is the threshold decay time constant and V_T is the threshold normalization constant. By handling large number of PCNN parameters, an auto convergence criterion has been introduced here to reduce the computational overhead and make simulation simpler. Target chip size has been utilized as a prior knowledge to make the convergence automatic. After each iteration, the output image of the previous iteration has been compared with the target chip to decide the cardinality of the target using the relationship, $\sum_{i=1}^r \sum_{j=1}^c O_{i,j}[n] > \text{target size}(m \times n)$, where, r and c are the number of rows and columns of the output image. It works in a way when, the threshold decreases exponentially by it causes more firing in next step. For instance, if any iteration fires more neurons than the size of target chip then network has been converged and previous iteration has been considered as the final outcomes.

2.2. Region of interest and template matching

This final iteration produces an image with bright intensity regions which may contain the target called region of interest (ROI). ROI has been then cropped and target template has been moved through these ROIs to identify the target location more specifically. Template matching has been done to find out best similarity within the ROI upon which the template is placed, using Euclidean similarity measure by following the formula in Eq. (6),

$$d = \sqrt{\sum_{u=1}^{T_r} \sum_{v=1}^{T_c} (\mathbf{R}[u, v] - \mathbf{T}_m[u, v])^2} \quad (6)$$

where, d represents the similarity between ROI and template, \mathbf{R} denotes ROI, \mathbf{T}_m indicates the selected image template, u and v are the coordinates of \mathbf{R} and \mathbf{T}_m . T_r and T_c are the number of rows and columns of the \mathbf{T}_m . Here, the dimension mismatch between ROI and template may occur, specifically where the template size is slightly larger than ROI. To avoid this circumstance three times larger than the target chip has been selected as the size of ROI, shown in Figure 2(a), where, m and n are the width and height of target chip, $3m$ and $3n$ are the size of ROI. Figure 2(b) shows the different sample templates of target vehicles used for this experiment. This figure depicts the various angular views of a target, where brighter portions show the vehicle's headlights and rest are the body of the target vehicles.

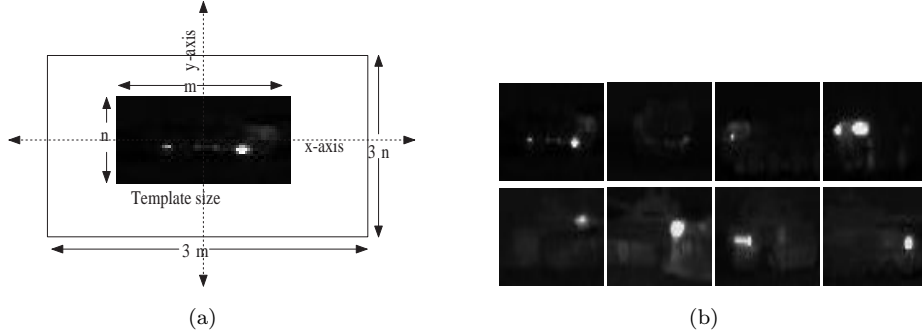


Figure 2.: Demonstration of (a) ROI selection process and (b) eight different target templates.

3. Experimental and result analysis

This section demonstrates the impact of the proposed AC-PCNN method which has been later compared with other existing methods by performing statistical measurements such as precision, recall, accuracy and receiver operating characteristic (ROC) curve Davis and Goadrich (2006). The study has been tested on a large number of FLIR images and the visual outcomes have been analysed through template matching on selected ROIs. This process has been executed for all the compared methods to evaluate the results and make fair judgement.

3.1. Data description

The FLIR dataset SENSIAC (2008) has been used for this investigation. It contains real-time videos and information about the army vehicles in ground scenario. The videos (later extracted as images) comprises situations like different weather condition, night time, engine on/off, with various aspect angles (0° to 360°), far targets, nearby targets, and camouflaged targets. Here, the experiment has been conducted on the data collected through MWIR sensor. The army vehicles were driven within a specific range from the sensor during night for data collection, detailed in Table 1. The nine types of battle tank targets have been considered and the results shows here only for the battle tank T72 is marked bold.

Forty night vision videos have been tested with two different scenes: high and low cluttered background based on the target range from sensors. Further, image frames have been extracted from these videos. Each video consists of 1024 frames that contains

Table 1.: Details of the target vehicles are included in SENSIAC dataset.

Look No.	Target Type	Scenario No.	Range (m)
1	Pickup	1923	1000
2	Sport Utility Vehicle	1925	1500
5	BTR70 Armored Personnel Carrier	1927	2000
6	BRDM2 Infantry Scout Vehicle	1929	2500
9	BMP2 Armored Personnel Carrier	1931	3000
10	T62 Main Battle Tank	1933	3500
11	T72 Main Battle Tank	1935	4000
12	ZSU23-4- Anti-Aircraft Weapon	1937	4500
13	2S3 Self-Propelled Howitzer	1939	5000

various viewpoint and different angles of a moving target. Twenty frames have been selected from them at a fixed interval to cover all possible views of a target. The original image size is 640×512 pixels with two types of target chip size: 10×5 pixels for high and 30×20 pixels for low cluttered background. This study has been conducted entirely on 800 distinct images. The experiment has been carried out using Matlab R2016b tool on an Intel(R) Core(TM) i5 processor@ 3.30GHz running Windows 7 Enterprise 64 bit operating system with a 7856 MB NVIDIA Graphics Processing Unit (GPU).

3.2. Parameter fitting

Proposed AC-PCNN algorithm includes some parameters that have been fixed by the method of successive estimation. The parameters are threshold decay constant (α_T), normalization constant (V_T), and linking strength (β) denoted in Section 2. The value of V_T is set as 20 because it must be high enough to prevent the neurons from firing instantly. Experimentally the value of β and α_T are being set to 3 and 0.2. This parameter selection phase makes the algorithm able to find out the correct position of a target with less computational complexity. Therefore, the quantitative analysis of parameters is considered to achieve optimum performance by the network for true target detection. The parameters of AC-PCNN have been tuned to deliver best results within the considered data. The selection of parameters make the model capable to achieve optimal results and is essential for the correctness of the proposed ATD.

3.3. Result analysis and comparison

AC-PCNN produces segmented images after each iteration based on the intensity level of the input images. The obtained results from the iterations of AC-PCNN have been shown in Figure 3. In first iteration all neurons are fired, as the variables are initialized with zero except $F_{i,j}$. Then, the threshold normalization constant V_T have increased the $T_{i,j}$. Thus, this high threshold value restricts the immediate firing of neurons. Associated decay parameter α_T reduced $T_{i,j}$ exponentially over the iterations that the minor changes in resultant binary images could be identified significantly. It has been observed that, the network takes approximately twenty-one iterations to fire the neuron again after first iteration as shown in Figure 3. It has been found that, the output of the 23rd iteration contains the brighter intensity values which became greater than the threshold and consequently neurons started firing. The resultant images from iteration 23 to 29 illustrate the increment of the target region along with firing of neuron. Over the iterations, $T_{i,j}$ has decayed and the 30th iteration shows more background portions. As a result, it produces insignificant boundary regions, which include the background

or clutter. Therefore, the 29th iteration met the convergence criteria (described in Section 2.1) and has considered as final output. All frames have been processed in the same way to reach their converged iterations.

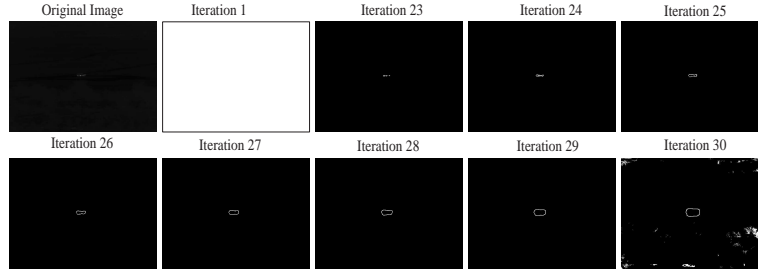


Figure 3.: Outcomes from PCNN iterations for a single image.

The other established detection methods such as, Otsu thresholding Otsu (1979), centre surround difference (CSD) Sun et al. (2005), saliency extraction Hou and Zhang (2007) and fuzzy energy based active contour (FEAC) Mondal, Ghosh, and Ghosh (2016) have been applied to the FLIR data to compare them with the results of PCNN. The performance comparison with AC-PCNN and CSD have been displayed in Figure 4. Here, two low and two high cluttered FLIR images have been chosen for presenting the difficulty of handling noisy input images and prove the robustness of the proposed method.

Figure 4 demonstrates the results of low and high cluttered images. Figure 4a(i)-(iv) is the original input image and the associated spatial distributions have been displayed in Figure 4b(i)-(iv) which has reflected maximum of 40 intensity value approximately. The 3D spatial distribution with consequent colour bar indicates the different intensity level of thermal variabilities of target and background. Figure 4a(i) and 4a(iii) have the similar spatial characteristics in terms of homogeneous 3D perspective view. Although a sharp peak has been detected in both the cases, due to the head light of the tank though the detection of the shape of the whole target is the most challenging task of this detection process. In case of, Figure 4a(ii) and 4a(iv) both contain more visible regions than the other two. But, the thermal variability of these images makes it more imperceptible for target selection from the immediate background. The distribution edges of the object and background have been realized from Figure 4b(ii) where no clearly separate peak has been found for target, but the colour bar indicates the presence of the target by high intensity values (shown in yellow colour).

The obtained results from CSD have been shown in Figure 4c(i)-(iv) and corresponding 3D histograms are plotted in Figure 4d(i)-(iv). The peaks of the distribution indicate the tentative target region, which contains huge amount of false alarms. Progressive sub-sampling and low pass filtering have created the difference between fine and coarse scales of pixel towards the intensity feature map. However, due to the intensity overlap, it could not reveal contrasts between the target tanks and back bush properly which have the similar intensity and reflectivity. Often it missed the actual target position because of the prominent background, which might be occurred due to the thermal inconsistency of the input images. As a result, it has produced binary output images with target along with more number of false alarms per scene. Figure 4e(i)-(iv) shows AC-PCNN results which converged in 27th, 23rd, 20th, and 21st iteration respectively (from left to right). The intensity distribution have been shown in Figure 4f(i)-(iv), where the target regions are found accurately and realized from the peak. . Thus, it implies that the high difference between target and background

intensity (in case of high cluttered images) insists the rate of firing neuron and converge the proposed method earlier. Successively, to acquire the true shape of target, ROI has been cropped and template matching has been being implemented using the target chip in an overlapping manner. The lowest distance between ROI and target chip has been considered as best matching region. Then the final results have been shown in Figure 4g(i)-(iv) by using a bounding box on the target region. In addition, histogram equalization have been performed to enhance the visibility of the output images which are the final results of the proposed ATD.

3.4. *Performance evaluation*

Following baseline methods: Otsu thresholding Otsu (1979), CSD Sun et al. (2005), saliency extraction Hou and Zhang (2007), and fuzzy energy based active contour (FEAC) Mondal, Ghosh, and Ghosh (2016) have been explored to compare and assess the results of the proposed method. Furthermore, ROI and template matching have been included for all baseline methods in final ATD process for justified evaluation. Experimental results have been validated through confusion matrix, precision, recall, and accuracy measurement. Confusion matrix has been determined by the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) measures. Potential targets have been categorized in two cases, actual target (shown in column) and predicted target (shown in row).

The results of proposed method have been compared with different methods and tabulated in Table 2, where it can be seen that the CSD is able to detect 421 targets accurately among 800 targets. CSD performs the adaptive image thresholding by using the Gaussian pyramid as progressive low-pass filter and sub-sampling of an input image. The number of false alarms obtained in each scene is quite high thus, the detection rate is very low compared to other ATD method. In the case of Otsu thresholding and saliency extraction, both detected 434 number of true targets but the type I and type II errors are different because, they use different approach for detecting targets. Otsu uses the thresholding for segmentation whereas, saliency employs feature mapping to detect targets. The selection of threshold in Otsu is solely depends on the nature of the pixel intensity. So, if the global distribution of target image and background varies widely the target detection performance degrades automatically. However, in case of saliency, it employs feature mapping at a given scale by using feature vectors for each pixel, instead of using separate saliency maps for scalar values of each feature. But it fails to detect the targets from FLIR imagery due to the low contrast and thermal variability between the target and background. Also, FEAC is an active contour based method has been implemented for comparing the result of proposed work. This method is focused on fuzzy energy based active contour by considering local spatial information. It is capable to deal with noisy, blurred, discontinuous edged, and high intensity inhomogeneity images. However, it struggled extensively to segment small size targets from low-contrast and high cluttered FLIR images. FEAC has detected 484 true targets from the given FLIR images. Thus, this method is also failed to compete with the obtained performance from proposed work in terms of ATD. Therefore, Table 2 infers that the proposed method has detected highest number of accurate targets 718 out of 800 targets, which is the acquired optimal performance from this experiment. The accuracy has been measured for all the methods by measuring total truly detected and not detected targets among the complete experimental dataset. The detection

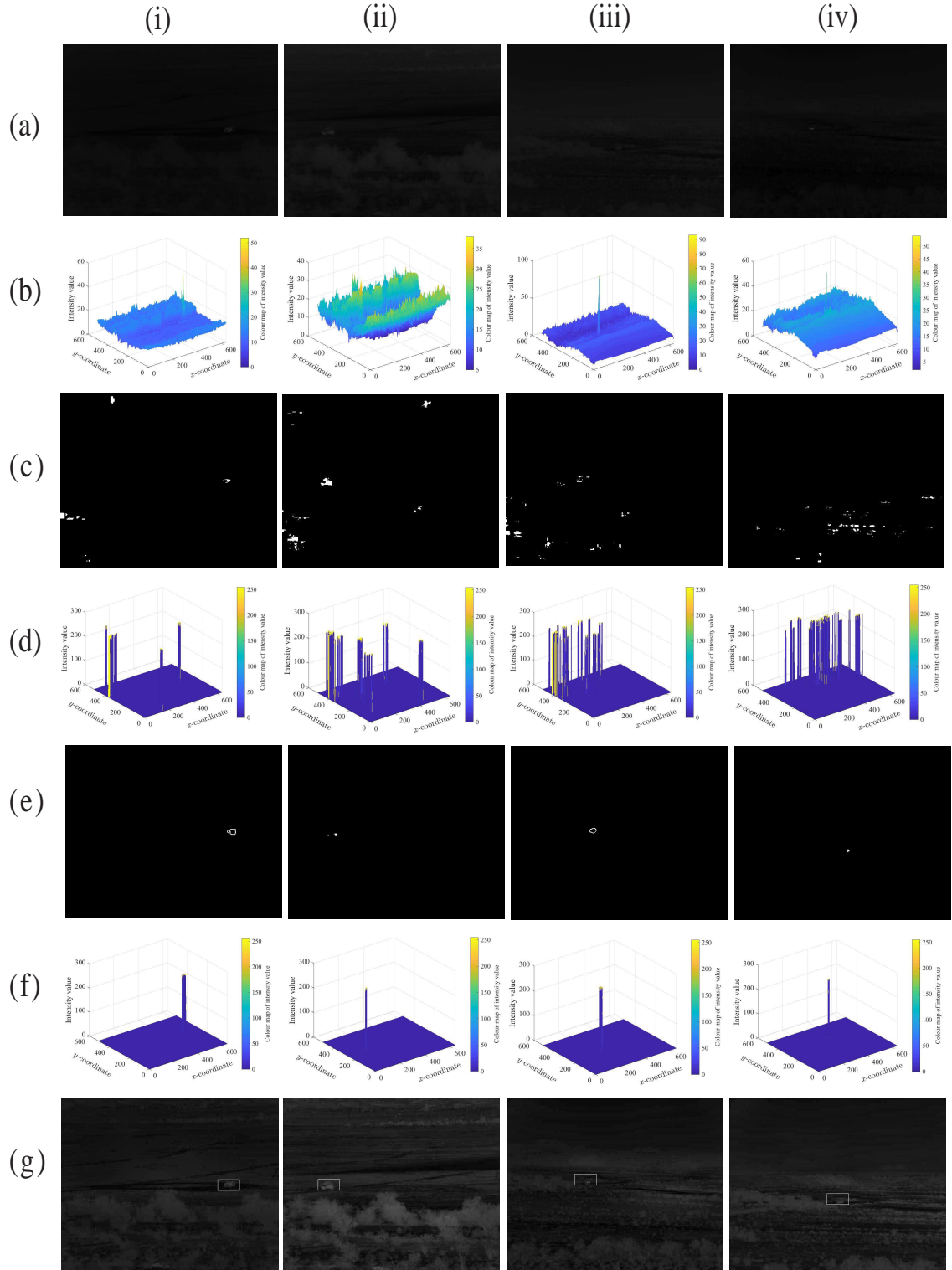


Figure 4.: Results of two low and two high cluttered images from (i) to (iv), (a) original images (b) 3D spatial histogram distribution of original images (c) resultant images of CSD (d) 3D spatial histogram distribution after applying CSD (e) output images of proposed AC-PCNN (f) 3D spatial histogram distribution after applying AC-PCNN (g) final detected result of proposed method.

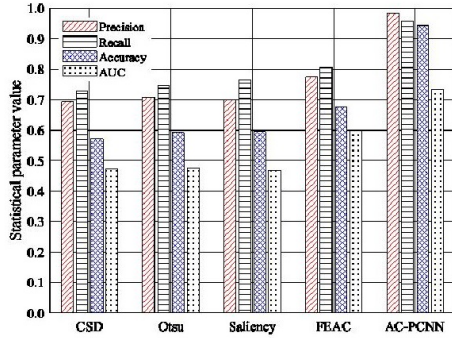
rate and the probability of accuracy have been calculated via precision and recall measurements, tabulated in Table 3 and the values have been plotted in a bar graph, shown in Figure 5(a). It has been observed from the results that the proposed method achieved high precision, recall, accuracy, and area under the curve (AUC) values (mark in bold in Table 3) compared to the existing methods. In addition, ROC curve has been displayed in Figure 5(b) to demonstrate the performance of proposed method in a graphical way. It infers that, the overall performance of our proposed method is superior than the other baseline methods explored in this work.

Table 2.: Confusion matrix to demonstrate the performance of the proposed method over other existing methods.

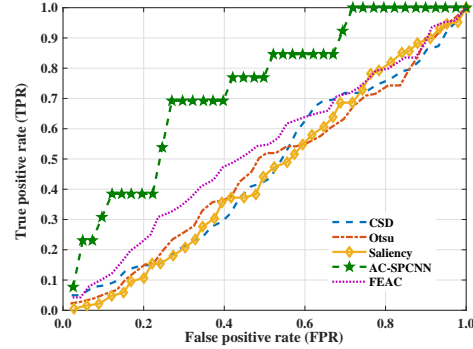
Methods		Actual	Not Target
CSD	Predicted	421	186
	Not Target	158	35
Otsu	Predicted	434	179
	Not Target	147	40
Saliency	Predicted	434	186
	Not Target	133	42
FEAC	Predicted	484	143
	Not Target	117	56
AC-PCNN	Predicted	718	13
	Not Target	32	37

Table 3.: Performance comparison of different methods through statistical metrics.

Methods	Precision	Recall	Accuracy	AUC
CSD	0.693	0.727	0.570	0.473
Otsu	0.707	0.746	0.592	0.475
Saliency	0.700	0.765	0.595	0.467
FEAC	0.772	0.805	0.675	0.601
AC-PCNN	0.982	0.957	0.943	0.735



(a)



(b)

Figure 5.: Performance measurements by (a) accuracy comparison method and (b) ROC curve.

4. Conclusions and future work

Proposed method demonstrates the ability of the AC-PCNN to segment the highly cluttered FLIR images by an efficient adaptive thresholding method. In addition, this work contributes to convert the PCNN to an auto-convergent PCNN (AC-PCNN) towards ATD process. The combination of the ROI and template matching for this work gives a good approximation of the accurate target shape. The promising results have been obtained from the proposed work compared to the existing methods. Although, this work is limited to identification of a target image and underway to extend for target recognition towards making a full automatic target recognition (ATR) system.

References

- Bai, X., and F. Zhou. 2010. "Infrared small target enhancement and detection based on modified top-hat transformations." *Computers & Electrical Engineering* 36 (6): 1193–1201.
- Bhanu, B. 1986. "Automatic target recognition: state of the art survey." *IEEE Transactions on Aerospace and Electronic Systems* 22 (4): 364–379.
- Cheng, G., P. Zhou, and J. Han. 2016. "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images." *IEEE Transactions on Geoscience and Remote Sensing* 54 (12): 7405–7415.
- Clemente, C., L. Pallotta, I. Proudler, A. De Maio, J. J. Soraghan, and A. Farina. 2015. "Pseudo-Zernike-based multi-pass automatic target recognition from multi-channel synthetic aperture radar." *IET Radar, Sonar & Navigation* 9 (4): 457–466.
- Davis, J., and M. Goadrich. 2006. "The relationship between Precision-Recall and ROC curves." In *Proceedings of the 23rd international conference on Machine learning*, 233–240. ACM.
- Deshpande, S. D., H. E. Meng, R. Venkateswarlu, and P. Chan. 1999. "Max-mean and max-median filters for detection of small targets." In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, 74–83. International Society for Optics and Photonics.
- Eckhorn, R., H. J. Reitboeck, M. Arndt, and P. Dicke. 1990. "Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex." *Neural Computation* 2 (3): 293–307.
- Gao, C., T. Zhang, and Q. Li. 2012. "Small infrared target detection using sparse ring representation." *IEEE Aerospace and Electronic Systems Magazine* 27 (3): 21–30.
- Hou, X., and L. Zhang. 2007. "Saliency detection: A spectral residual approach." In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 1–8. IEEE.
- Kim, S., W.-J. Song, and S.-H. Kim. 2018. "Double Weight-Based SAR and Infrared Sensor Fusion for Automatic Ground Target Recognition with Deep Learning." *Remote Sensing* 10 (1): 72.
- Li, H., Y. Wei, L. Li, and Y. Y. Tang. 2010. "Infrared moving target detection and tracking based on tensor locality preserving projection." *infrared physics & technology* 53 (2): 77–83.
- Li, H., Y. Wei, L. Li, and Y. Yuan. 2012. "Similarity learning for object recognition based on derived kernel." *Neurocomputing* 83: 110–120.
- Mondal, A., S. Ghosh, and A. Ghosh. 2016. "Robust global and local fuzzy energy based active contour for image segmentation." *Applied Soft Computing* 47: 191–215.
- Otsu, N. 1979. "A threshold selection method from gray-level histograms." *IEEE transactions on systems, man, and cybernetics* 9 (1): 62–66.
- SENSIAC. 2008. "Military Sensing Information Analysis Center, ATR Algorithm Development Image Database." Georgia, USA. US Army Night Vision and Electronic Sensors Directorate (NVESD).
- Serre, T., L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. 2007. "Robust object recognition with cortex-like mechanisms." *IEEE transactions on pattern analysis and machine intelligence* 29 (3): 411–426.
- Shaoa, Z., X. Zhua, and J. Liub. 2008. "Morphology infrared image target detection algorithm optimized by genetic theory." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37: 1299–1304.
- Shi, Y., Y. Wei, H. Yao, D. Pan, and G. Xiao. 2018. "High-Boost-Based Multiscale Local Contrast Measure for Infrared Small Target Detection." *IEEE Geoscience and Remote Sensing Letters* 15 (1): 33–37.
- Subashini, M. M., and S. K. Sahoo. 2014. "Pulse coupled neural networks and its applications." *Expert Systems with Applications* 41 (8): 3965–3974.
- Sun, S. G., D. M. Kwak, W. B. Jang, and D. J. Kim. 2005. "Small target detection using center-surround difference with locally adaptive threshold." In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA 2005)*, Zagreb, Croatia, 402–407.