Dim Target Detection in Infrared Images Using Saliency Algorithms

Seyit TUNC[1], Hakkı Alparslan ILGIN[2]

[1] Roketsan, Ankara, Turkey
[2] Dept. of Electrical and Electronics Eng., Ankara University, Gölbaşı, Ankara, Turkey

seyittunc@gmail.com, ilgin@eng.ankara.edu.tr

Abstract. *Infrared (IR) target detection and tracking are commonly used in modern defense systems. Target detection is the first and very important step for several surveillance applications. Long distance between imager and targets or bad weather conditions mostly cause dim target appearance with low signal-to-noise ratio (SNR) in IR images. In this study, dim targets in IR images are enhanced and detected using saliency detection algorithms, which have not been used in IR wavelength before. Performances of the algorithms are evaluated on common IR datasets. Algorithms are compared in terms of SNR, receiver operating characteristic (ROC) and area under curve (AUC) score. Effects of parameter selection are also considered for automatic target detection. Furthermore, feasibility of the methods for real-time applications are discussed.*

Keywords

Infrared images, target detection, dim target, saliency

1. Introduction

Detection of targets in IR images is of great importance in IR search and tracking applications. Especially when targets are dim and have low signal-to-noise ratio, they are hard to differentiate from background noise, which usually causes high false alarm rates for conventional methods. IR imagers are widely used in military implementations since they sense thermal diffusion and reflection of objects. Besides their benefits, they have some drawbacks, such as low contrast between target and background and deficient spatial resolution that cause inconvenience for computer vision algorithms. Automatic target detection is an essential step for sundry surveillance applications in computer vision applications. Detection and tracking become very laborious in IR images since they yield important information about a target. Applying a threshold to an image is the most facile way to detect a target, but it is not well advised if the target has low contrast with background or has both of cold and hot components. Applying low threshold triggers high false alarm rate, whereas high threshold decreases probability of detection. A general solution to surmount the drawbacks is to enhance target signal before applying a threshold directly. Thus, undesirable results may be avoided.

In IR images, even when targets have low contrast according to their surroundings or appear in cluttered background, they are salient regions and have potential of taking visual attention. In other words, the image structure belonging to the target area differs barely or conspicuously from its local background. In some cases, it is not visually differentiable from its background. For this reason, utilizing methods that model attention of human visual system builds our motivation in order to enhance targets in IR images. A wide variety of approaches have been proposed to solve saliency detection such as frequency tuned salient region detection [1], spectral residual approach in saliency detection [2], image signature [3], computational model of visual attention called saliency maps [4], rapid scene analysis using saliency-based visual attention model [5], global contrast based salient region detection [6], context-aware saliency detection [7], graph-based visual saliency [8] and salient object detection [9]. Stable multisubspace learning method has been used to detect dim and small targets in heterogeneous scenes in [10]. Specifically, research progress of image saliency has been analyzed in [11]. Various methods for saliency detection in color images have been reviewed in [12].

We compare the performance of several saliency detection methods and configure each of them with two different parameter sets to show the effect of parameter for detecting targets in IR images having low contrast or situated in complicated scenes. Following methods designed and successfully applied for images in the visible wavelength are arranged for IR images and considered in this comparison: Frequency tuned (FT) [1], spectral residual (SR) [2] and image signature (IS) [3]. Algorithms are compared according to the enhancement, detection performance and execution time.

In the next section, saliency detection and aforementioned algorithms are described. In Sec. 3 and 4, target detection and experimental results are discussed, respectively. In the last section, conclusion is given.

2. Saliency Methods

Target regions are acknowledged as the principal areas in an image. A saliency map denotes visually dominant locations such that it represents image areas where it stimulates visual perception more than the other areas. In this work, saliency maps obtained from the mentioned three methods are utilized as enhanced target image for detection purpose.

FT method uses Lab color space and detects salient regions with color and luminance features [1]. Some frequency bands in input image are filtered by utilizing difference of Gaussians (DoG) as a band pass filter. Very low frequencies from the original image are passed in order to accentuate immense salient regions, whereas high frequencies are considered to define boundaries of the regions. In contrast to these preserved frequencies, the highest frequencies are ignored to suppress noise. In our work, since IR images are grayscale, input images are converted to Lab color space to attain performance of the pristine work, in lieu of utilizing the input directly as L luminance and ignoring a and b color components. Output saliency map S has the same size as the input image I and aims to highlight the most sizably voluminous salient regions:

$$S(x,y) = \left\| I_{\mu} - I_{w_{hc}}(x,y) \right\| \qquad (1)$$

where I_{μ} is mean intensity vector of the input image, $I_{w_{hc}}$ is Gaussian smoothed version of the original image, and $\|\cdot\|$ denotes Euclidean distance. Each pixel location is a vector in $[L,a,b]^{\mathrm{T}}$ form. Also, x and y are horizontal and vertical components, respectively.

In spectral residual (SR) approach, contrast between the original log-spectrum of the Fast Fourier Transform (FFT) and its smoothed variant, in particular the spectral residual (SR), is figured to acquire curiosity districts of an image [2]. At that point combination of phase spectrum and the spectral residual is transformed to spatial domain by utilizing inverse Fast Fourier Transform to achieve saliency map. As indicated by [2], the technique can be condensed as

$$\mathcal{A}(i,j) = \Re\big(\mathcal{F}\big[I(x,y)\big]\big), \qquad (2)$$

$$\mathcal{P}(i,j) = \Im\big(\mathcal{F}\big[I(x,y)\big]\big), \qquad (3)$$

$$\mathcal{L}(i,j) = \log\big(\mathcal{A}(i,j)\big), \qquad (4)$$

$$\mathcal{R}(i,j) = \mathcal{L}(i,j) - h_n(i,j) * \mathcal{L}(i,j), \qquad (5)$$

$$S(x,y) = g * \mathcal{F}^{-1}\Big[\exp\big(\mathcal{R}(i,j) + \mathcal{P}(i,j)\big)\Big]^2 \qquad (6)$$

where $\mathcal{A}(i,j)$, $\mathcal{P}(i,j)$, and $\mathcal{L}(i,j)$ signify amplitude, phase and log spectrum of input image I, respectively. $\Re(i,j)$ is the spectral residual and $S(x,y)$ indicates saliency map. $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the Fast Fourier Transform and Inverse Fast Fourier Transform, individually. $h_n(i,j)$ indicates the mean filter with size $n \times n$. g is a Gaussian filter

and it smooths the saliency map to show signs of improvement visual impacts. * is the convolution operation. Input image size is scaled to 64 pixels height (or width) before FFT and the saliency map is rescaled to the input image size at the last step. In the original work, color input image is converted to grayscale image before applying the first process [2]. It is not required to execute this procedure here, because input is a grayscale image itself.

In [3], a simple image descriptor, called the image signature (IS), is presented. IS is defined as the sign function of the Two-Dimensional Discrete Cosine Transform (2D-DCT) of an image I and given as below:

$$\hat{I} = DCT(I), \qquad (7)$$

$$IS(I) = \mathrm{sign}\big(\hat{I}\big) \qquad (8)$$

where $\hat{I}$ is the 2D-DCT of I. Equation (8) forsakes amplitude information and holds only phase information. For each channel of the image I_k, saliency map is obtained as below:

$$\bar{I}_k = IDCT\big(IS(I_k)\big), \qquad (9)$$

$$S = g * \sum_k \big(\bar{I}_k o \bar{I}_k\big) \qquad (10)$$

where $IDCT$ is Inverse 2D-DCT, and o represents element-wise product operator. Note that IR images have only one channel.

3. Target Detection

Saliency maps obtained using the methods in the previous section contain enhanced salient targets. After targets in IR images are enhanced, a threshold is applied to these maps generated by each of the saliency algorithms to detect targets.

Using a fixed threshold is not appropriate to label target pixels correctly, since the saliency maps have different statistical properties. Therefore, the threshold should be adaptive to the map. In [1] and [2], adaptive threshold is only related to the mean value of the obtained saliency map, while [3] gives the detection results in terms of receiver operating characteristic (ROC) and the area under the ROC (or curve), and does not apply a single threshold. Conversely, we show that the most suitable threshold value for detecting targets in a saliency map depends on both the mean and the standard deviation of the map. Thus, we introduce a new adaptive thresholding for the normalized saliency maps as

$$Th = m + \frac{(1-m)m}{2s} \qquad (11)$$

where m and s represent the mean and the standard deviation of the saliency map, respectively. This formulation is derived from $Th = m + c * s$, commonly used to binarize input images, where c is a constant. Before applying thresh-

old, each pixel value of the saliency map is normalized to [0 1]. $(1 - m)$ reflects the difference between the maximum and the mean value accordingly. Using the threshold given by (11), binarized map of salient targets is given by

$$B(x,y) = \begin{cases} 1, & \text{if } S_N(x,y) > \min(l, \ Th) \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where l is the upper limit of threshold, which is smaller than 1, and $S_N(x,y)$ is normalized saliency map. In order to have the highest target detection achievement, optimal threshold value assigns target pixels to 1, while it makes background pixels to 0 in the binarized map $B(x,y)$. In IR images, targets are salient regions because of their contrast and detailed structure. Specifically, the mean and standard deviation of normalized saliency maps obtained from IR images are low, besides the target areas are brighter than the background. An example of mean value and standard deviation of ten different normalized saliency maps are given in Tab. 1. In this table, threshold values are obtained by increasing from a lower value until the false positive rate has lower score, while the true positive rate is at higher. After that value, true positive rate starts to decrease. The highest threshold value amongst all normalized saliency maps examined in our experiments is 0.8. Therefore, setting l to 0.8 will be the safest value for upper bound in order to prevent a target area pixel to be classified as a background pixel. When comparing l to the threshold value Th of the normalized saliency map as in (12), smaller one of l and Th will be chosen, so that the pixel classification results in high sensitivity. Therefore, choosing the lesser value from l and Th increases the target detection achievement. Even though classification using the threshold (11) achieves high true positive rate and most of the threshold values are less than l, using l as an upper bound plays a role as a controller at the top level when obtaining binarized map $B(x,y)$.

4. Experimental Results and Discussion

Experiments are held for two different parameter sets for each method. Table 2 contains the parameters that may take variable values.

In FT, size and standard deviation of the Gaussian filter that determines higher cut off frequency in DoG may take numerous values. In SR, mean filter size may be adjustable, but changing the parameter value does not affect obtained residual [2]. Thus, it is considered as a fixed parameter instead of a tunable parameter here. A fixed value is also used for scaling factor since 64 pixels width (or height) as down-scaling size is recommended in [2] and [3]. The Gaussian filter is employed to smooth the output saliency map at the final stage in [2] and [3]. Changing its standard deviation and size is possible.

In this work, not only the methods are compared, but also the effects of the Gaussian filters in each method are examined in enhancement, detection and required computa-

Mean	Standard Deviation	Threshold
0.03	0.07	0.35
0.04	0.07	0.30
0.29	0.15	0.80
0.05	0.06	0.39
0.04	0.04	0.58
0.07	0.07	0.57
0.07	0.07	0.48
0.05	0.07	0.63
0.08	0.09	0.43
0.05	0.07	0.42

Tab. 1. Mean and standard deviations of some normalized saliency maps and corresponding highest threshold values.

FT	SR	IS
Filter size	Filter size	Filter size
Filter std. dev.	Filter std. dev.	Filter std. dev.
	Scaling factor	Scaling factor
	Mean filter size	

Tab. 2. Tunable parameters of the methods.

Scene	Resolution	Target Size	Ratio (%)
S1	[640 512]	[16 8]	[2.5 1.6]
S2	[640 512]	[18 5]	[2.8 1]
S3	[112 112]	[5 4]	[4.5 3.6]
S4	[112 112]	[5 4]	[4.5 3.6]
S5	[640 512]	[14 12]	[2.2 2.3]
S6	[640 512]	[22 7]	[3.4 1.4]
S7	[640 512]	[11 5]	[1.7 1]
S8	[112 112]	[5 5]	[4.5 4.5]
S9	[112 112]	[4 2]	[3.6 1.8]
S10	[112 112]	[3 2]	[2.7 1.8]
L1	[640 512]	[58 34]	[9.1 6.6]
L2	[640 512]	[78 31]	[12.2 6.1]
L3	[112 112]	[47 52]	[42 46.4]
L4	[112 112]	[18 16]	[16 14.2]
L5	[640 512]	[56 22]	[8.8 4.3]
L6	[640 512]	[47 20]	[7.3 3.9]
L7	[640 512]	[44 29]	[6.9 5.7]
L8	[112 112]	[34 39]	[30.4 34.8]
L9	[112 112]	[29 19]	[25.9 17]
L10	[112 112]	[36 16]	[32.1 14.3]

Tab. 3. Specifications of the representative scenes.

tion time manner. We use two different parameters for the Gaussian filters in order to reflect performance of the methods in arduous conditions having various target size, spatial image resolution, target contrast and background complexity. At the first set, Set1, the filter parameters, which are standard deviation and size, are arranged to 0.5 and 3×3, respectively. At the second set, Set2, standard deviation and filter size are set to 2.2 and 9×9 in that order. The filter size is determined minimally to demonstrate the major characteristic of the Gaussian filter at that standard deviation. Test scenarios are collected from commonly used IR dataset, AMCOM [13] and SENSIAC [14], and grouped into two different target size scales, i.e., small and large. Each group also consists of two different resolutions. These scenarios are given in Tab. 3. The letters S and L in the scene column denote scenes containing small and large target, respectively. While some of the small and large scenes have high resolution, others have low resolution. Resolution and target size of each scene are given in terms

of pixel in the first and the second column, and their percentages (target size/resolution) are shown in the ratio column of Tab. 3, respectively.

In order to evaluate enhancement results, visibility of input and output target signals are compared using

$$SNR = \frac{|I_T - \mu_B|}{\sigma_B} \qquad (13)$$

where I_T represents mean intensity value of the target pixels in an image. μ_B and σ_B are mean intensity value and standard deviation of the image, respectively.

SNR results for small and large target scenes are given in Tab. 4 and Tab. 5, respectively. Average SNR value is given at the last row in the tables as SNR_μ. For both Set1 and Set2 parameters, the best results for small target scenes are obtained by IS and SR in Tab. 4. Specifically, both methods give higher SNR values than the input. Namely, they improve the input for target detection. However, FT is inappropriate for small target scenes. For three of the methods, higher standard deviation and bigger filter size of Set2 result in lower SNR than Set1. Analogously, methods with Set1 parameters enhance the results better than with Set2, since Set1 has low standard deviation and filter size that are appropriate for small targets.

SNR results for large target scenes are shown in Tab. 5. In this case, FT enhances the input better for uniform targets such as L3 and L4. IS and SR also improve the input. Three of the methods with Set2 parameters obtain slightly higher SNR values for large targets than with Set1, since standard deviation and filter size are high in Set2. In addition to SNR, enhancement and target detection results for four small and four large scenes are shown in Fig. 1 and Fig. 2 to be able to make subjective evaluation.

Detection results are given in terms of ROC in Fig. 3 and Fig. 4 for small and large targets, respectively. In these figures, best detection results are achieved by SR and IS for small target scenes in SNR. High true positive rate against low false positive rate in ROC graph indicates a high target detection achievement. AUC scores are given in Tab. 6 and Tab. 7 for small and large targets, respectively. Average AUC is given as AUC_μ at the last rows of the tables. AUC measures target detection achievement with the highest value of 1. The higher the AUC is, the better the detection result is. Its relationship with ROC is that if true positive rate increases so does AUC, since it is measured by the area under the ROC curve. Therefore, high AUC value indicates more accurate target detection result. Parallel to the SNR results, IS and SR give the best AUC results for both small and most of the large target scenes as shown in Tab. 6 and Tab. 7.

Number of blobs for IR images with four small and four large targets located at the target or the background region are shown in Tab. 8 and Tab. 9, individually. The letters T and B represent "at the target region" and "at the background region", respectively. The symbol - denotes no blob detected at this region. Higher blob number at the target and lower blob number at the background mean better

Scene	Input	FT		IS		SR	
		Set1	Set2	Set1	Set2	Set1	Set2
S1	0.41	1.51	0.10	3.78	1.41	7.39	2.90
S2	0.89	0.07	0.48	5.83	9.06	4.27	6.97
S3	0.64	0.47	0.84	18.01	12.78	14.42	11.58
S4	1.53	1.46	0.45	15.74	11.21	15.53	11.63
S5	0.69	1.70	0.60	23.55	7.79	22.56	8.30
S6	1.57	2.27	0.86	5.85	3.16	6.11	3.14
S7	0.65	0.47	0.78	7.94	2.58	5.10	1.94
S8	0.94	0.27	0.27	8.48	11.45	8.05	11.04
S9	0.42	0.50	0.63	26.71	13.64	19.66	12.10
S10	0.60	0.19	1.06	19.17	12.40	19.15	12.68
SNRμ	0.83	0.89	0.61	13.51	8.55	12.22	8.23

Tab. 4. SNR results of the small target scenes.

Scene	Input	FT		IS		SR	
		Set1	Set2	Set1	Set2	Set1	Set2
L1	0.51	0.99	0.83	4.31	4.18	4.77	4.87
L2	0.05	1.30	0.98	4.85	4.67	4.77	4.83
L3	1.17	0.92	0.92	0.42	0.97	0.27	0.76
L4	2.14	3.25	3.67	2.06	4.03	2.05	4.03
L5	0.56	1.24	0.86	5.68	4.60	6.04	4.93
L6	0.20	1.35	0.80	9.96	7.78	9.60	7.50
L7	0.59	0.92	0.63	4.57	3.99	6.49	5.38
L8	0.26	1.04	0.95	0.92	1.90	0.85	1.91
L9	1.84	1.53	1.89	1.51	2.87	1.36	2.77
L10	0.31	0.80	0.24	1.31	2.51	0.99	2.08
SNRμ	0.76	1.33	1.18	3.56	3.75	3.72	3.91

Tab. 5. SNR results of the large target scenes.

Scene	Input	FT		IS		SR	
		Set1	Set2	Set1	Set2	Set1	Set2
S1	0.583	0.743	0.461	0.960	0.897	0.986	0.983
S2	0.744	0.522	0.377	0.994	0.998	0.993	0.995
S3	0.741	0.402	0.239	1.000	1.000	0.998	1.000
S4	0.828	0.731	0.405	0.997	1.000	0.994	1.000
S5	0.652	0.768	0.610	1.000	1.000	1.000	1.000
S6	0.784	0.705	0.643	0.962	0.995	0.969	0.996
S7	0.642	0.336	0.200	0.997	0.982	0.991	0.956
S8	0.831	0.570	0.554	0.958	1.000	0.942	1.000
S9	0.579	0.310	0.370	1.000	1.000	0.999	1.000
S10	0.659	0.352	0.033	0.998	1.000	0.997	1.000
AUCμ	0.704	0.544	0.389	0.987	0.987	0.987	0.993

Tab. 6. AUC results of the small target scenes.

Scene	Input	FT		IS		SR	
		Set1	Set2	Set1	Set2	Set1	Set2
L1	0.592	0.675	0.666	0.924	0.997	0.940	0.997
L2	0.523	0.768	0.746	0.951	0.999	0.933	0.998
L3	0.124	0.766	0.776	0.615	0.817	0.572	0.714
L4	0.807	0.766	0.811	0.887	0.993	0.895	0.992
L5	0.365	0.745	0.680	0.856	0.996	0.973	0.996
L6	0.548	0.790	0.700	0.994	0.999	0.992	0.999
L7	0.611	0.753	0.674	0.942	0.999	0.979	0.999
L8	0.528	0.676	0.646	0.748	0.958	0.748	0.943
L9	0.895	0.807	0.886	0.823	0.991	0.807	0.974
L10	0.562	0.625	0.587	0.827	0.974	0.816	0.957
AUCμ	0.556	0.737	0.717	0.857	0.972	0.866	0.957

Tab. 7. AUC results of the large target scenes.

detection results. Furthermore, only one blob at the target is preferable to the piecewise blobs since the entire blobs at the target belong to single target.

A comparison about feasibility for real-time applications and required execution time is provided in Tab. 10 for
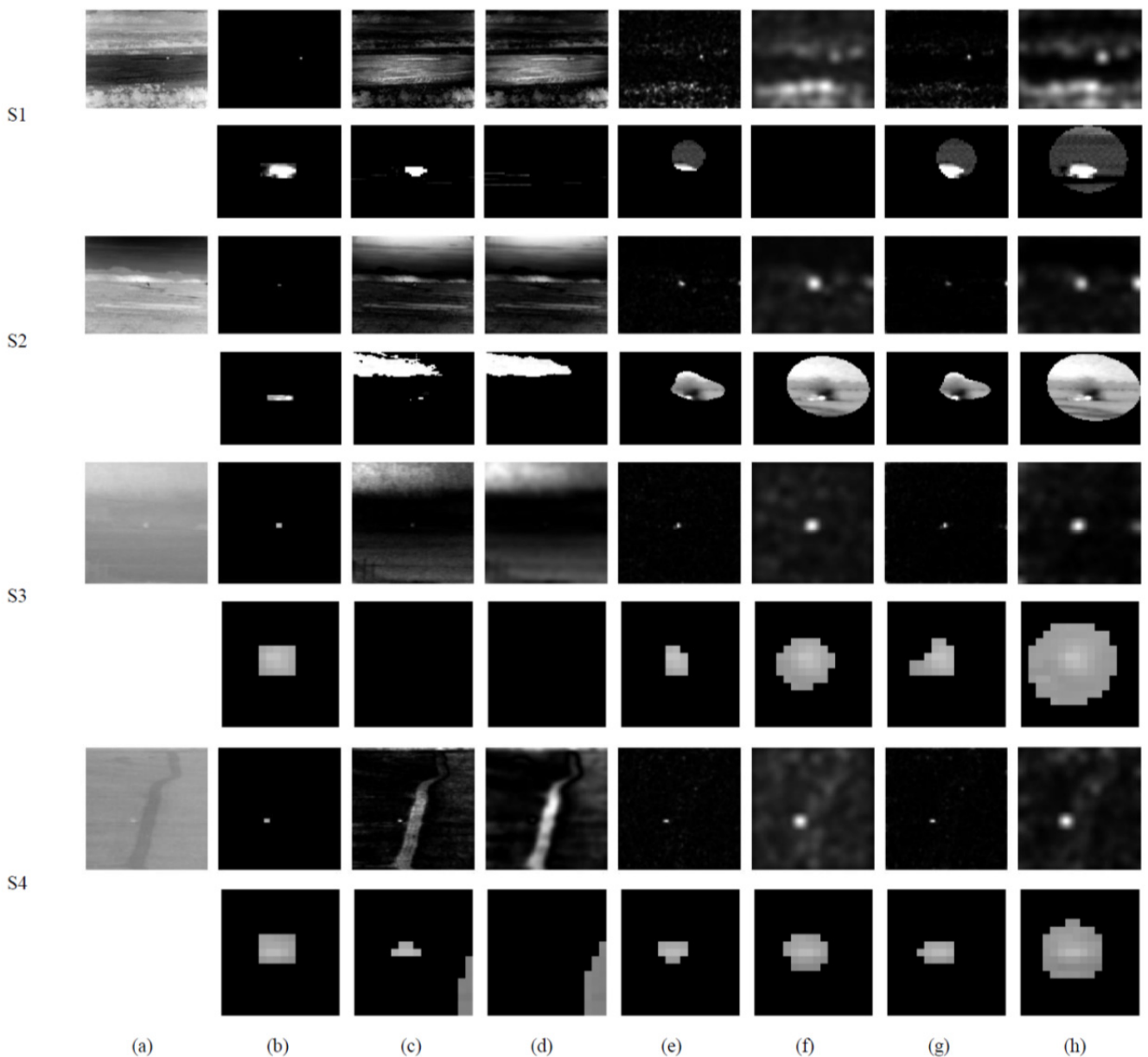
Fig. 1. Detection results of the enhanced small target scenes, S1, S2, S3 and S4: (a) The scene, (b) Above: masked target region, below: zoomed target area, (c)-(h) Above: enhancement result of FT Set1, FT Set2, IS Set1, IS Set2, SR Set1 and SR Set2, respectively, below: magnified target areas found from the method-set combination above.

each method. Each of the methods is processed 100 times for all scenes and the average computation time is calculated for each of them. Algorithms are performed on a computer with 3.50 GHz Intel® Xeon® CPU E3-1270 v3 and 32 GB RAM. A threshold calculated via the method mentioned in the previous section is applied to each enhanced target images and accordingly detection results are obtained. Detection results for the methods FT, IS, SR with the sets Set1 and Set2 around the magnified target region for both small and large target scenes are shown at the second row for each scene in Fig. 1 and Fig. 2, respectively. The targets are aligned at the center of the region exactly. As visually perceived in the given results in the figures and the tables, FT is not capable at the diminutive target scenes especially having the cluttered background such as S1 and S2 for both of the sets. FT with Set1, which

has smaller standard deviation, highlights the target region as well as the largest uniform area in the scene that is a road at S4. In addition to poor performance in enhancement, FT usually loses the targets and offers very high false alarm rates (compare the number of blobs detected at the target and at the background in Tab. 8). The performance of FT is additionally not adequate at the large targets locating at the cluttered background such as L1 and L2 in Tab. 9. If a target is larger and nearly uniform at a scene like L3 and L4, FT may perform better and is capable to detect blobs placed at the target region and may provide no false alarm. Moreover, the other methods fail visually in particular at L3, as shown in Fig. 2. Consequently, FT generally emphasizes immense uniform areas. If a target suits these conditions in an application, FT with larger standard deviation may have successful results. According to
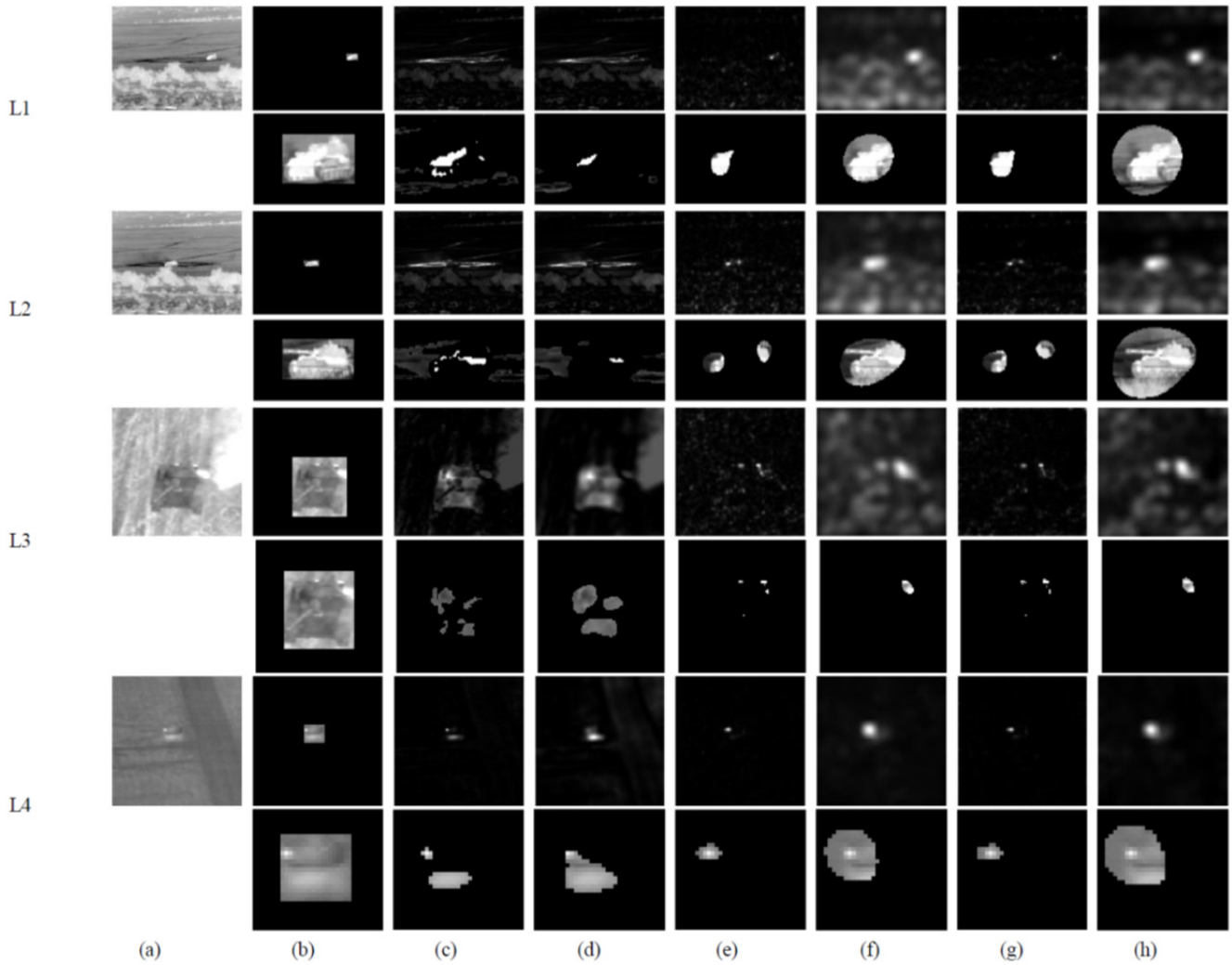
Fig. 2. Detection results of the enhanced large target scenes, L1, L2, L3 and L4: (a) The scene, (b) Above: masked target region, below: zoomed target area, (c)-(h) Above: enhancement result of FT Set1, FT Set2, IS Set1, IS Set2, SR Set1 and SR Set2, respectively, below: magnified target areas found from the method-set combination above.

Scene	Blob	FT		IS		SR	
		Set1	Set2	Set1	Set2	Set1	Set2
S1	T	4	1	1	-	1	1
	B	558	35	24	3	10	3
S2	T	2	-	1	1	1	1
	B	45	2	1	-	4	1
S3	T	-	-	1	1	1	1
	B	12	1	-	-	5	2
S4	T	1	-	1	1	1	1
	B	13	3	-	-	-	-

Tab. 8. Number of blobs detected in the small target scenes.

Scene	Blob	FT		IS		SR	
		Set1	Set2	Set1	Set2	Set1	Set2
L1	T	9	2	1	1	1	1
	B	47	12	2	-	2	-
L2	T	11	4	2	1	2	1
	B	51	5	-	-	3	1
L3	T	4	3	4	1	5	1
	B	-	-	-	-	1	-
L4	T	2	1	1	1	1	1
	B	-	-	-	-	-	-

Tab. 9. Number of blobs detected in the large target scenes.

the tables and the figures, SR and IS yield very close performance. It is conceivable that utilizing phase information in both methods can cause that homogeneous attribute. Both of the methods are very successful to stick out diminutive targets even when the background is highly cluttered, if the standard deviation is set to a small value as in Set1. Amount of false alarms may be quite a lot at S1 if the applied threshold is not high enough. Here it is noted that the target is located around the most proper region in the enhanced images obtained by both methods, that is to say that the threshold is convenient to be increased to get lower false alarms. Increasing threshold is also applicable to the other scenes (S2, S3 and S4), but it is noted that there is a tradeoff between detection probability and false alarm rate. As shown in Fig. 1 for both SR and IS, Set1 estimates the size of the small target more accurately than Set2.

Additionally, the target regions have pixels with the highest values in all enhanced images achieved by utilizing Set1. However, utilizing Set2 does not provide pixels with the highest value for all target regions conversely (see S1 in Fig. 1, the brightest region does not refer to the target).
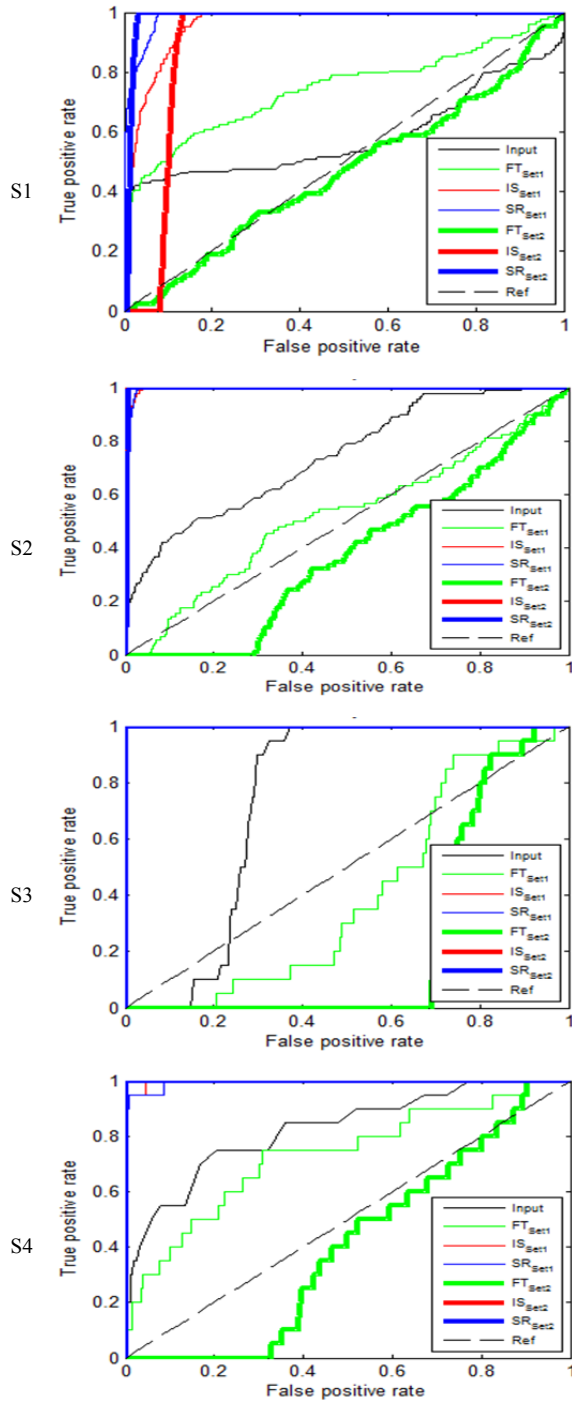
Fig. 3. Detection results (ROC) of the small target scenes.

Fig. 4. Detection results (ROC) of the large target scenes.

On the other hand, the detected blobs by using Set1 are biased according to the center of the region at the second row at S1 and S2 in Fig. 1. This implies that Set1 has disadvantage on determining exact location of the targets. For the large targets, Set2 works well in contrast to Set1.

Both of parameter sets show poor performance at L3 that has large target size-image resolution ratio (almost 50% in one dimension). At large target scenes, Set1 may induce partial detected blobs at the target region and false alarms at the background, which are unpropitious. The execution times given in Tab. 10 indicate that SR and IS meet
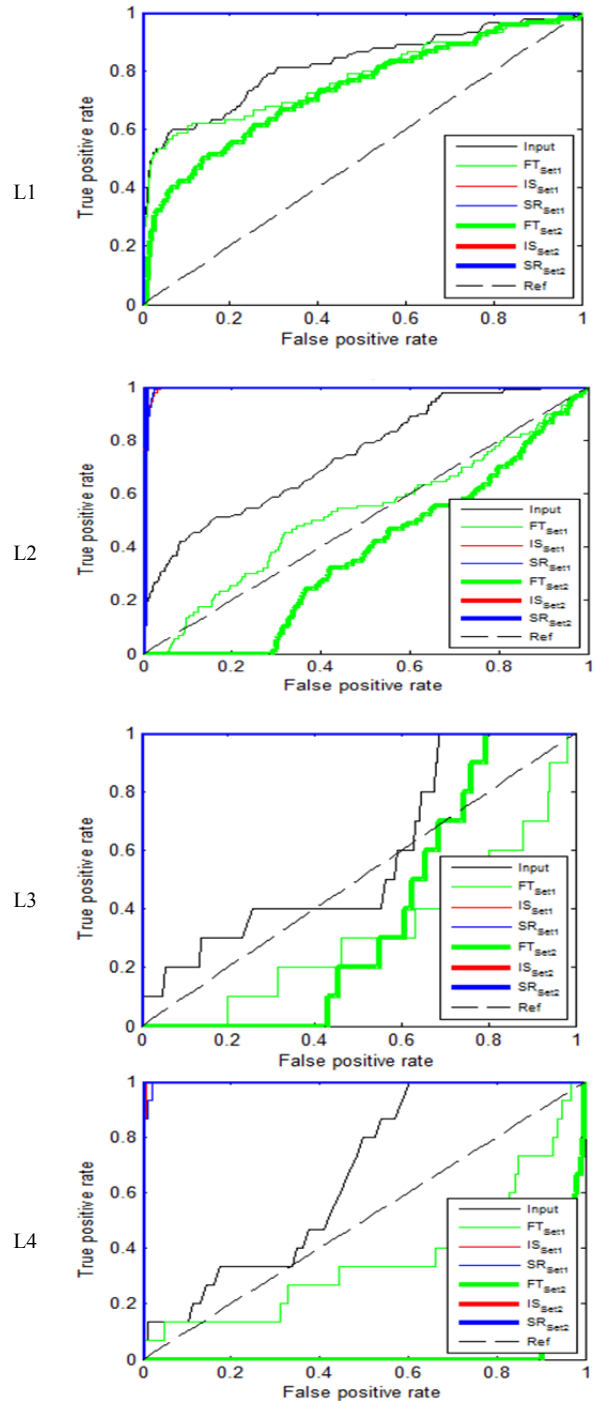
	FT		IS		SR	
	Set1	Set2	Set1	Set2	Set1	Set2
Time (ms)	79.4	78.7	5.79	5.83	5.76	5.86

Tab. 10. Average execution time.

requirements for real-time applications since they may be performed on more than 150 frames for each second.

In general speaking, Set1 with smaller standard deviation is more felicitous for very small targets, whereas Set2 with larger standard deviation is more opportune for large targets. FT may be well-accomplished if a target is virtually uniform and occupies a sizably voluminous region in

an image, otherwise SR and IS with congruous parameter sets outperform FT. The performances of SR and IS are proximate to each other.

One of the most consequential and vital goal of automatic target detection algorithms is to find pixel coordinates of the targets in an image. Lower scaling factor in SR and IS should be employed with the intention of getting target position in high precision, particularly at the diminutive targets. Furthermore, SR and IS are also acceptable for real-time applications.

5.   Conclusion

Three different saliency detection algorithms commonly utilized in visible wavelength are used newly for IR images and compared in this work. Each of them is designed with two diverse parameter sets to examine the results of parameter choice on target enhancement and detection under sundry conditions such as different target size, target contrast, image resolution and background complexity. Size of the target is the most foremost limitation while selecting a well-suited parameter set. IS and SR with right parameter sets give adequate results in all conditions aside from the scenes where target-image resolution ratio is too high. FT shows better performance in these scenes by highlighting large uniform target areas. This designates that the methods are preferable to each other according to different conditions. An exertion can be expended to develop new methods that need no prior information about scene and are influenced less from varieties of the conditions in the future. Also, an SNR method that is more coherent with human visual perception system than the existing objective evaluation algorithms can be studied to evaluate the detection results.

Acknowledgment

References

[1] ACHANTA, R., HEMAMI, S., ESTRADA, F., et al. Frequency tuned salient region detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Miami Beach (FL, USA), 2009, p. 1594–1604. DOI: 10.1109/CVPR.2009.5206596

[2] HOU, X., ZHANG, L. Saliency detection: A spectral residual approach. In *IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis (USA), 2007, p. 1–8. DOI: 10.1109/CVPR.2007.383267

[3] HOU, X., HAREL, J., KOCH, C. Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2011, vol. 34, no. 1, p. 194–201. DOI: 10.1109/TPAMI.2011.146

[4] BULLING, A., ALT, F., SCHMIDT, A. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *SIGCHI International Conference on Human Factors in Computing Systems*. Austin (TX, USA), 2012, p. 3011–3020. DOI: 10.1145/2207676.2208712

[5] ITTI, L., KOCH, C., NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1998, vol. 20, no. 11, p. 1254–1259. DOI: 10.1109/34.730558

[6] CHENG, M., ZHANG, G., MITRA, N., et al. Global contrast based salient region detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*. Colorado Springs (USA), 2011, p. 409–416. DOI: 10.1109/CVPR.2011.5995344

[7] GOFERMAN, S., ZELNIK-MANOR, L., TAL, A. Context-aware saliency detection. In *IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco (CA, USA), 2010, p. 2376–2383. DOI: 10.1109/CVPR.2010.5539929

[8] HAREL, J., KOCH, C., PERONA, P. Graph-based visual saliency. In *Proceedings of Neural Information Processing Systems (NIPS)*. Canada, 2006, p. 545–552.

[9] BORJI, A., SIHITE, D., ITTI, L. Salient object detection: A benchmark. In *European Conference on Computer Vision*. Florence (Italy), 2012, part 2, p. 414–429. DOI: 10.1007/978-3-642-33709-3_30

[10] WANG, X., PENG, Z., KONG, D., et al. Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, vol. 55, no. 10, p. 5481–5493. DOI: 10.1109/TGRS.2017.2709250

[11] MA, R., YU, Y., YUE, X. Survey on image saliency detection methods. In *IEEE International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. Xian (China), 2015, p. 329–338. DOI: 10.1109/CyberC.2015.98

[12] GARG, A., NEGI, A. A survey on visual saliency detection and computational methods. *International Journal of Engineering and Technology*, 2017, vol. 9, p. 2742–2753. DOI: 10.21817/ijet/2017/v9i4/170904406

[13] AMCOM-database [Online] Available at: http:://www.cs.ucf.edu/~vision/projects/meanshift/means hift.html

[14] SENSIAC-database [Online] Available at: https://www.dsiac.org/resources/research-materials/cds-dvds-databases/atr-algorithm-development-image-database

About the Authors …

Seyit TUNÇ received B.S., M.S. and Ph.D. degrees in Electronics Engineering from Ankara University, Ankara, Turkey, in 2006, 2010 and 2016, respectively. He is currently with Roketsan in Ankara, Turkey. His research interests are signal processing, target tracking and video processing.

Hakkı Alparslan ILGIN (corresponding author) received the B.S. and M.S. degrees in Electronics Engineering from Ankara University, Ankara, Turkey, in 1993 and 1997, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Pittsburgh, Pittsburgh, PA, USA in 2004. He worked for two years as Research Assistant and one year as Teaching Assistant for the University of Pittsburgh. He is currently working as Associate Professor in the Dept. of Electrical and Electronics Engineering, Ankara University, Ankara, Turkey. His research interests are signal, image and video processing.