METRICS FOR PERFORMANCE EVALUATION OF PREPROCESSING ALGORITHMS IN INFRARED SMALL TARGET IMAGES

W.-H. Diao and X. Mao

School of Electronic and Information Engineering
Beihang University, Beijing 100191, China

V. Gui

Faculty of Electronics and Telecommunications
Politehnica Timisoara University
Timisoara 300223, Romania

Abstract—Image preprocessing is commonly used in infrared (IR) small target detection to suppress background clutter and enhance target signature. To evaluate the performance of preprocessing algorithms, two performance metrics, namely PFTN (potential false targets number) decline ratio and BRI (background relative intensity) decline ratio are developed in this paper. The proposed metrics evaluate the performance of given preprocessing algorithm by comparing the qualities of input and output images. The new performance metrics are based on the theories of PFTN and BRI, which describe the quality of IR small target image, by representing the difficulty degree of target detection. Theoretical analysis and experimental results show that the proposed performance metrics can accurately reflect the effect of the image preprocessing stage on reducing false alarms and target shielding. Compared to the traditional metrics, such as signal-to-noise ratio gain and background suppression factor, the new ones are more intuitive and valid.

Received 24 January 2011, Accepted 14 March 2011, Scheduled 18 March 2011
Corresponding author: Xia Mao (moukyoucn@yahoo.com.cn).
1. INTRODUCTION

Over the past two decades, a considerable amount of research has been done to improve the ability to detect small targets in infrared (IR) images. Generally speaking, target detection involves two stages: image preprocessing and target judging. Preprocessing is an indispensable stage, because it can reduce false-alarm rate and increase detection rate through suppressing background clutter and enhancing target signature. So far, a lot of preprocessing algorithms have been brought up, some focus on space domain and some care about frequency domain [1–5], such as two-dimensional least mean square (TDLMS) filter [6], morphological filter [7], high-pass filter [8], median filter [9], nonlinear filter [10, 11], local variance weighted information entropy (WIE) filter [12]. It is well known that performance evaluation is an essential part for an effective algorithm, so we focus on evaluating the performance of preprocessing algorithms for IR small target images.

Clearly, the reliability of evaluation result relies on specific evaluation metrics. Signal-to-noise ratio (SNR) gain and background suppression factor (BSF) [12, 13–16] are the most popular metrics to evaluate preprocessing algorithms in IR small target images. Nevertheless, SNR could not accurately describe the quality of IR small target images and clutter standard deviation does not work well in denoting the complexity of background. Thus, the mentioned two metrics (SNR gain and BSF), deriving from SNR or standard deviation are not credible enough. To solve the problem, we propose two new metrics, which evaluate the performance of a given preprocessing method by quantifying the quality difference between the processed image and the original one.

The structure of this paper is as follows. Section 1 describes the background of this research. Section 2 introduces traditional evaluation metrics for preprocessing algorithms and analyzes their deficiency. Section 3 introduces two new small target image quality descriptors: potential false targets number (PFTN) and background relative intensity (BRI). Section 4 presents two new evaluation metrics which are constructed based on the image quality descriptors defined in Section 3. Section 5 validates the proposed metrics by both theory and experiments. Finally, Section 6 draws the conclusion of this paper.

2. DESCRIPTION AND DEFECT ANALYSIS OF TRADITIONAL METRICS

In small target images, the sizes of targets are less than 0.15% of the whole image area [17, 18]. Considering that an image size is $128 \times 128$
pixels, the sizes of its small targets are ranging from $1 \times 1$ to $5 \times 5$ pixels. Due to the small size, we can only obtain little information about target shapes and textures in the image.

For performance evaluation of preprocessing algorithms in IR small target images, SNR gain and BSF [12,13–16,19] are widely accepted metrics. They are defined as:

$$SNR \text{ gain} = \frac{SNR_{out}}{SNR_{in}}$$  \hspace{1cm} (1)

and

$$BSF = \frac{C_{in}}{C_{out}}$$  \hspace{1cm} (2)

where $SNR_{in}$ is the SNR of input image for a given preprocessing algorithm; $SNR_{out}$ is SNR of the output image obtained with the given algorithm; $C_{in}$ is the clutter standard deviation of input image; $C_{out}$ is the clutter standard deviation of output image. We can see that the two metrics are constructed by SNR and standard deviation respectively. However, SNR and standard deviation are not informative descriptors for IR small target image quality, as we will illustrate by several examples below.

2.1. Experiments for SNR

SNR is a classical method to measure the relationship between target and noise [20–22], in the field of IR target detection, it is defined as [23]:

$$SNR = \frac{I_t - I_b}{C_b}$$  \hspace{1cm} (3)

where $I_t$ is the mean gray value of the target region, $I_b$ is the mean gray value of background and $C_b$ is the standard deviation of the background. According to Equation (3), SNR mainly reflects the information of background standard deviation and the contrast between target and background. To illustrate the limitation of SNR, we give four representative images as shown in Fig. 1.

The SNR values of these four IR small target images are: (a) 2.4188; (b) 3.6732; (c) 5.8036; (d) 7.9079. According to SNR theory, image (d) is the best one, and (c) is better than (b) and (a). However, the backgrounds of (a) and (b) are simpler than (c) and (d), while the disturbing degree of background upon target detection in (c) and (d) is higher than (a) and (b), so (c) and (d) have worse image quality from the perspective of small target detection. Since the estimation result deriving from SNR is far from describing our difficulty to detect the targets, the usefulness of SNR for IR image quality assessment within the task of small target detection can be questioned.
2.2. Experiments for Standard Deviation

Standard deviation is a popular image descriptor depending on global image statistics. It is widely used for describing the background complexity of small target images. However, popularity does not mean effectiveness or sensibility. For many images, the evaluation result of this descriptor fails to provide a useful measure of small target detection difficulty. In this experiment, two real-life and two artificial generated background images, illustrated in Fig. 2, are utilized to intuitively show the limitation of the statistic measure.
In Fig. 2, image (a) is cloudy sky, and (b) is sea clutter. It is
obvious that (b) is more complex than (a), since the sea background can
bring false alarm more easily. However, the standard deviation values
of the two images are not consistent with the above judgment, as the
former's value (12.2566) is larger than the latter's (9.7361). For the two
artificial images (c) and (d), it can easily be seen that the fluctuation of
(d) is more acute and complicated while the standard deviations of the
two synthetic images are both equal to 11.3120. This is why the results
based on the standard deviation also fail to provide a useful measure in
our task. It can be concluded that the global standard deviation is not
suitable for describing the background complexity. For other image
descriptors, based on image gray level statistics, such as variance and
entropy, similar conclusions can be drawn.

Since SNR and standard deviation have inherent defect in
measuring IR small target image quality, their derivative metrics: SNR
gain and BSF, are also unreliable. Therefore, in order to find more
valid evaluation metrics for preprocessing methods, two new image
descriptors, working better than SNR and standard deviation, are
proposed in next section.

3. DESCRIPTORS FOR IR SMALL TARGET IMAGES

An infrared small target image is used to detect targets. So, a valid
image quality descriptor should reflect the difficulty degree of target
detection. We find that, no matter how the detection method is
constructed and no matter which theory it is based on, the results
of target detection disturbances are the same: false alarm, losing the
target, or failing to get the precise target position. Furthermore, we
notice that all poor detection results are due to two reasons. One is that
some background regions are similar to the target region and provide
interferential information for target detection, thus causing false alarm.
The other one is that the high gray-level background regions may shield
or obscure the target to some extent, which results in losing the target
or failing in getting the precise target position.

Based on the factors identified above, two image descriptors,
namely PFTN and BRI, are developed to measure the difficulty
degree of target detection in IR small target images. Here, PFTN
represents the ability of the background to provide fake targets, and
BRI represents the ability of the background to shield the target.
Sections 3.1 and 3.2 will define and explain them in detail.
3.1. PFTN

As false alarm is caused by the target-like regions in the background, it is intuitive that the more potential false targets exist, the stronger the ability of background to disturb target detection is. As shown in Fig. 3, it is obvious that image (c) has the highest probability of false alarm, while the quality of image (a) is the best with its clean background. Given the close relationship between the number of false alarms and potential false targets number, it is necessary to take this measure (PFTN) as a descriptor of IR small target image.

The value of this image descriptor can be obtained by subtracting the number of real targets from that of possible candidate targets. In order to pick up the possible candidate targets from the image as more as possible, morphologic methods are firstly used to increase the contrast between candidate target regions and others. Then three restrictive conditions are applied to judge the candidate targets, including four-connected rule, gray threshold and the limited size of small targets. The detail of this algorithm is as follows.

Step 1: Add the original image to the image transformed by morphologic top-hat operator, then subtract the image transformed by morphologic bottom-hat operator. For a $128 \times 128$ IR small target image, the size of these two operators is $5 \times 5$ which is selected according to the definition of small target mentioned above (the size of small target is not larger than $5 \times 5$ in a $128 \times 128$ image). The structuring element used in this work has a diamond shape.

Step 2: Binarize the image by means of an adaptive gray threshold, which is obtained by Otsu method [24]. In this method, the Otsu threshold is the one obtained with minimal weighted within-class variance.

Step 3: Label the connection region complying with Four-connected rule (only the pixels connected in horizontal or vertical orientation could be seen as the same object).

Step 4: Compute the weights of the labeled regions which are
candidate targets. Use only the labeled regions with size less than or equal to 0.015% of the whole image size, otherwise, set the weight of the region to 0.

Step 5: Count the number of possible candidate targets and potential fake targets.

An instance of calculating the potential fake targets number is shown in Fig. 4. The background of the original image is sea clutter. The result shows that the method nearly picks up all the possible candidate targets.

![Figure 4. An instance of calculating potential fake targets number.](image)

### 3.2. BRI

In this section, we will define a new image descriptor, named background relative intensity, which describes the degree of background shielding the target. It is commonly accepted that the higher the gray-level of the background region is, the stronger the probability of that region to shield a target is. However, we find that the distance between the background region and the target also influences the probability of shielding the target. See Fig. 5 for an illustration of the idea.

In image (a), the target shelter performance is the strongest, while the background region in image (c) almost can not shelter the target at all. The target shelter performance of image (b) is in between the cases (a) and (c). It can be concluded that the above results are caused by the different distances between the background region and the target in the three images.

Based on the example above, it can be concluded that two factors need to be combined in BRI calculation:

1) Each pixel of the background has a weight concerning its gray value. The greater gray value it has, the greater weight it is assigned. The rule of gray weight assignment is defined by the equation:

$$w_1(i) = \begin{cases} 
1 & (I_i \geq I_t) \\
\frac{I_t}{I_i} & (I_i < I_t) 
\end{cases} \quad (4)$$
Figure 5. Three artificial small target images with different distances between background region and target.

Figure 6. Function relation between the weight and the distance.

When the gray value of a background pixel is not lower than that of the target, its weight is set as 1, otherwise it is set as $\frac{I_t}{I_i}$, where $I_t$ is the target mean gray value, and $I_i$ denotes the gray value of pixel $i$ in the background.

2) Each pixel of the background has a distance weight, depending on its distance to the target. The shorter the distance to the target, the higher the distance weight assigned to that pixel is. More specifically, the weight is set as a decreasing function of the distance to the target. To ensure that the weight equals to 1 when the distance is 1 (the minimum value of the distance in digital images) and is close to 0 when the distance is big enough (not less than 10), we use an exponential function, to define the second weight as,

$$w_2(i) = e^{-(d(i) - 1)} \quad (e = 2.71828)$$  \hspace{1cm} (5)

where $w_2(i)$ is the weight of pixel $i$, $e$ is the constant 2.71828 and $d(i)$ denotes the distance between the pixel $i$ and the target. This function is plotted in Fig. 6.

With $n$ denoting the number of the pixels in the background
region, the BRI is defined as follows:

\[ BRI = \sum_{i=1}^{n} w_1(i) \times w_2(i) \]  

(6)

3.3. Experiments for PFTN and BRI

In the first experiment, we calculated the PFTN and BRI of the images in Fig. 1. The values of PFTN are: (a) 0; (b) 1; (c) 72; (d) 1095. The values of BRI are: (a) 17.6807; (b) 20.9315; (c) 18.6142; (d) 26.4548. The above values indicate that the quality of (c) is worse than the quality of (a) and (b). Although the ability of the background to shield the targets in (c) almost equals that in (a) and (b), the ability of the background (c) to provide the fake targets is stronger than that in (a) and (b); the quality of (d) is the worst of all, because the background could provide a lot of fake targets during target detection and also the BRI is very high, given small distances to potential targets. It means that the probability of false alarm in image (d) is very high. We can see that, for all four images where the SNR has trouble in effectively measuring the complexity of small target detection task, the results estimated by the new descriptors are consistent with the real difficulty of detection.

Generally speaking, image quality evaluation results should be associated with the ATR algorithm performance. Furthermore, a good image descriptor should have a nearly monotone relationship with the ATR algorithm performance [25, 26]. Therefore, to validate these two image descriptors further, IR small target images with different backgrounds are used as samples for analyzing the relationship between our descriptors and actual performance of detection algorithms. Fig. 7 shows some samples used in this experiment. Here, we select two classic target detection algorithms: Victor algorithm [27] and Reed algorithm [28].

The relationship between PFTN values and false alarm is shown in Fig. 8, and the relationship between BRI values and probability of detection is shown in Fig. 9. The experimental results prove that the proposed image descriptors have a nearly monotone relationship with target detection algorithm performance. Therefore, we can conclude that the new descriptors are valid to measure IR small target image quality.
Figure 7. Samples of IR small target images with different backgrounds.

Figure 8. Relationship between PFTN and false alarm.

Figure 9. Relationship between BRI and probability of detection.
4. NEW METRICS FOR PERFORMANCE EVALUATION OF PREPROCESSING ALGORITHMS

In this study, we construct two new evaluation metrics for preprocessing algorithms based on the proposed image descriptors, PFTN and BRI. We evaluate the algorithm performance by comparing the quality of processed image with that of original image, as illustrated in Fig. 10.

![Flowchart for performance evaluation of preprocessing algorithms.](image)

The PFTN and BRI measures are used to describe the quality of both, the input and output images. Further on, to quantitatively measure the performance of given preprocessing algorithm, the PFTN decline ratio and BRI decline ratio are developed. The two metrics are defined as follows:

\[
PFTN\,\text{decline ratio} = \frac{PFTN_{in} - PFTN_{out}}{PFTN_{in}} \tag{7}
\]

\[
BRI\,\text{decline ratio} = \frac{BRI_{in} - BRI_{out}}{BRI_{in}} \tag{8}
\]

where \(PFTN_{in}\) is potential false targets number of original image, \(PFTN_{out}\) is potential false targets number of the image after preprocessing; \(BRI_{in}\) denotes background relative intensity within original image, \(BRI_{out}\) equals to background relative intensity within preprocessed image.

As shown in formula (7), \(PFTN\,\text{decline ratio}\) can not exceed 1. For most cases, its value is positive, which means that the quality of the processed image is better than that of original one, for there are less possible fake targets in the output image. If the value is negative, it indicates that the tested preprocessing algorithm makes the image quality worse than before. When comparing or sequencing the performance of various preprocessing algorithms, in view of reducing false alarm, the greater the value of this metric, the better the corresponding preprocessing method is.
According to Equation (8), the \textit{BRI decline ratio} has the same range as the \textit{PFTN decline ratio}. When the value is positive, it can be concluded that the preprocessing method does well in suppressing the background clutter. However when its value is less than 0, the corresponding preprocessing algorithm makes target detection harder than before. Specifically, 1 corresponds to a perfect preprocessing method, which has cleared up completely the background clutter and 0 implies that the method contributes nothing to enhancing the contrast between target and background. The greater the value of this metric, the better the corresponding preprocessing method is, in view of increasing the detection rate.

The advantages of the evaluation metrics developed here could be mainly embodied in three aspects.

First, the new metrics has a better theoretical basis. The new metrics are based on the image descriptors PFTN and BRI, while the traditional ones rely on the theory of SNR and standard deviation. Since in our work PFTN and BRI outperformed SNR and standard deviation in describing IR small target image quality, we conclude that \textit{PFTN decline ratio} and \textit{BRI decline ratio} are more reliable metrics within our target detection purpose.

Secondly, the new metrics are more intuitive to measure the performance of preprocessing algorithms, since: (1) we can qualitatively judge the performance of a given preprocessing algorithm easily, according to the positive and negative attributes of the metrics and (2) The value range of both is \([0, 1]\) under normal circumstances. Values normalized to 1 are easier to be used and understood than non-normalized ones. Therefore, the new metrics can be conveniently used to compare the performance of different preprocessing methods. (3) By comparing the actual value of the new metrics with the ideal value 1, the metrics can indicate in a straightforward manner whether the preprocessing algorithm can be improved or not.

Lastly, the new metrics provide more meaningful information. When preprocessing the image, some researchers are only interested in reducing false-alarm rate, while others may focus on increasing the detection rate. These different requirements can be both satisfied by our metrics. Specifically, we can get the two needed answers via analyzing the values of \textit{PFTN decline ratio} and \textit{BRI decline ratio} respectively.
5. EXPERIMENTS WITH THE NEW EVALUATION METRICS

Three preprocessing performance evaluation instances are given to depict the new evaluation metrics more clearly.

The experiment results of instance 1 and instance 2 are shown in Table 1 and Table 2 respectively. In these two evaluation instances, original images are two IR small target images with complicated sea-clutter backgrounds. Though all the three preprocessing methods used here include frequency domain high-pass filter and gray scale stretch, the corresponding cutoff frequencies of filters are different. Specifically, cutoff frequency of algorithm I = 1.5; cutoff frequency of algorithm II = 1.7; cutoff frequency of algorithm III = 2.5.

According to the performance evaluation results, for the image shown in Table 1, the values of BRI decline ratio show that algorithm III is the best one to make the true target clearly detectable and the following are algorithm II and algorithm I. Since the BRI decline ratio value of algorithm III is 0.9262, which is close to 1, we can believe that this preprocessing method is excellent for increasing detection rate. Meanwhile, as the results of PFTN decline ratio are all negative, we know that all the three preprocessed images can arouse false alarm more easily than the original one. Especially after processing by algorithm III, the background is the most cluttered and many potential fake targets appear in preprocessed image. After carefully observing the preprocessed images, we can see that the estimation results mentioned above are consistent with human evaluation. Taking the image processed by algorithm III for an example, here the true target is very clear as the background clutter is far from clear. At the same time, compared to other images, there are more background regions similar to the target region in this image. This judgment agrees with the conclusion drawn from the evaluation metrics.

For the image shown in Table 2, BRI decline ratio of algorithm III equals to 0.8089, and it is the greatest among the three algorithms. This indicates that algorithm III performs best in decreasing the degree of background shielding the target. According to BRI decline ratio values of algorithm I and II, we can see that algorithm II is better to make the target clearly detectable than algorithm I. Moreover, PFTN decline ratios of the three algorithms are −1.1262, −2.8421, and −2.2097 respectively. Judging by this metric, we can conclude that the three preprocessed images can arouse false alarm more easily than the original one, and algorithm II is the worst in view of reducing false-alarm rate. After analyzing the images processed by the three
Table 1. Evaluation results of three preprocessing algorithms using new metrics.

<table>
<thead>
<tr>
<th>Original image</th>
<th>Performance evaluation results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PFTN decline ratio</td>
</tr>
<tr>
<td>Result of preprocessing algorithm I</td>
<td>−0.5509</td>
</tr>
<tr>
<td>Result of preprocessing algorithm II</td>
<td>−0.4209</td>
</tr>
<tr>
<td>Result of preprocessing algorithm III</td>
<td>−1.3351</td>
</tr>
</tbody>
</table>

preprocessing algorithms, it can be concluded that the above inference is creditable. Taking the image processed by algorithm III for an example, here the true target is the most clear, as the target to background contrast is the lowest in the three preprocessed images. In addition, by observing the three images, we can see that the image processed by algorithm II has the most candidate fake targets and the image processed by algorithm III is next in the row. This judgment is also in agreement with the conclusion drawn from the evaluation metrics.
Table 2. Evaluation results of three preprocessing algorithms using new metrics.

<table>
<thead>
<tr>
<th>Original image</th>
<th>Performance evaluation results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( PFTN ) decline ratio</td>
</tr>
<tr>
<td>Result of preprocessing algorithm I</td>
<td>-1.1262</td>
</tr>
<tr>
<td>Result of preprocessing algorithm II</td>
<td>-2.8421</td>
</tr>
<tr>
<td>Result of preprocessing algorithm III</td>
<td>-2.2097</td>
</tr>
</tbody>
</table>

The results of the third group of experiments are shown in Table 3. In this case, the original image is also an IR small target image with sea clutter background and the tested preprocessing algorithms include space domain high-pass filter and gray scale stretch. The templates of the three filters are shown in Fig. 11.

By considering the \( BRI \) decline ratio values of the tested algorithms, we can conclude that algorithm III is the best one to increase the contrast between target and background and the following
Table 3. Evaluation results of three preprocessing algorithms using new metrics.

<table>
<thead>
<tr>
<th>Original image</th>
<th>Performance evaluation results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$PFTN$ decline ratio</td>
</tr>
<tr>
<td>Result of preprocessing algorithm I</td>
<td>0.3472</td>
</tr>
<tr>
<td>Result of preprocessing algorithm II</td>
<td>0.8750</td>
</tr>
<tr>
<td>Result of preprocessing algorithm III</td>
<td>0.5139</td>
</tr>
</tbody>
</table>

are algorithm II and algorithm I. Meanwhile, we know that all the three preprocessed images are less prone to provide false alarm than the original one, as the $PFTN$ decline ratio values are all positive. Especially after processing by algorithm II, the least potential fake targets appear in the preprocessed image. After carefully observing the preprocessed images, we can see that the evaluation results are reasonable.
\[
\begin{bmatrix}
-1 & -1 & -1 \\
-1 & 8 & -1 \\
-1 & -1 & -1 \\
\end{bmatrix}
\begin{bmatrix}
-1 & -2 & 5 & -1 \\
-2 & 5 & 5 & -2 \\
-1 & -2 & -2 & -1 \\
\end{bmatrix}
\begin{bmatrix}
-1 & -1 & -1 & -1 \\
-1 & -1 & -1 & -1 \\
-1 & -1 & 24 & -1 \\
-1 & -1 & -1 & -1 \\
-1 & -1 & -1 & -1 \\
\end{bmatrix}
\]

(a) algorithm I  (b) algorithm II  (c) algorithm III

**Figure 11.** Filter templates of three tested preprocessing algorithms.

### 6. CONCLUSION

In this study, two new metrics for preprocessing algorithms performance evaluation in IR small target images, namely \textit{PFTN decline ratio} and \textit{BRI decline ratio}, are proposed. The theoretical analysis and the experiments show that the new metrics are more valid and easy-to-use than the traditional ones, SNR gain and BSF. In addition, the proposed metrics can provide more information concerning the decline of the ability of the background to provide false alarm and to shield the target in IR small target images.

### ACKNOWLEDGMENT

This paper was supported by the Innovation Foundation of BUAA for PhD Graduates.

### REFERENCES


18. Yonoviz, D., “Tunable wavelet target extraction preprocessor,”


