Automatic target recognition using boundary partitioning and invariant features in forward-looking infrared images

Sun-Gu Sun
Sun-Gu Sun
Hyun Wook Park, MEMBER SPIE
Korea Advanced Institute of Science and Technology
Department of Electrical Engineering
373-1 Guseong-dong
Yuseong-gu, Daejeon 305-701
Korea
E-mail: hwpark@athena.kaist.ac.kr

Abstract. We propose an automatic target recognition (ATR) algorithm for recognizing nonoccluded and partially occluded military vehicles in natural forward-looking infrared (FLIR) images. The proposed algorithm consists of global and local feature extraction from partitioned boundaries of a target, and a new classification method using multiple multilayer perceptrons (MLPs). After segmenting a target, the target contour is partitioned into four local boundaries. Radial and distance functions are defined from the target contour and local boundaries, and are used to define global and local shape features, respectively. The global and local shape features are more invariant to similarity transform than traditional feature sets. Four feature vectors are composed of the global and local shape features, and are used as inputs of MLPs. The outputs of MLPs are combined to recognize nonoccluded and partially occluded targets. In the experiments, we show that the proposed features are superior to the traditional feature sets with respect to invariance and recognition performance. © 2003 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1532743]

Subject terms: automatic target recognition; classification; feature extraction; forward-looking infrared; invariant features.

Paper 020159 received Apr. 24, 2002; revised manuscript received Jul. 11, 2002; accepted for publication Jul. 15, 2002.

1 Introduction

Automatic target recognition (ATR) is a military application of pattern recognition that recognizes types of targets. In general, it executes several sequential operations such as image enhancement, segmentation, feature extraction, and classification.1–3 ATR is a very difficult task in remotely sensed FLIR images because they contain numerous artifacts, such as background clutter, target signature variability, aspect angle dependence, and partial occlusion.4–6

Many researchers have developed various ATR algorithms7–21 for FLIR images. Artificial neural networks have played important roles in the development of ATR systems. Roth7 and Rogers et al.8 surveyed the applications of neural network technology for ATR. Difficulties associated with neural networks include dimensionality and generalization problems. To learn high-dimensional data, such as images, a huge number of training data are required. Training with a small number of training data results in poor generalization of the neural networks. Therefore, feature extraction is necessary to reduce dimensionality and complexity. Lampinen and Oja22 mapped the input images into a feature space using Gabor filters and multilayer self-organizing maps. Wang, Der, and Nasrabadi10 employed feature-decomposition and data-decomposition modular neural networks to solve the dimensionality problem. Chan, Nasrabadi, and Torrieri10 examined principal component analysis (PCA), eigen separation transform (EST), and Fisher linear discriminant (FLD) as a feature extractor. DeCatrel and Surdu11 used relative distance ratios of extreme points based on target parts in the second generation FLIR images.5,6

Until now, recognition of a partially occluded target in FLIR images has been rarely studied due to variance of target signatures. Nair and Aggarwal12 studied a method that recognized nonoccluded and partially occluded targets in the second generation FLIR images. The method divided a target into several parts based on the curvature of a target boundary after segmentation. Zernike moments of the partial image were then used to classify targets by Bayesian classification. However, this method has difficulties in part decomposition because the target signatures vary and distribution of parts is not readily obtainable. In addition, high-order Zernike moments and ordinary moments, which have been popularly used to recognize 2-D objects in industrial applications, are sensitive to digitization errors, minor shape deformations, and background noise.23 These traditional feature sets can hardly be applied to ATR systems. Therefore, we need a new invariant feature set that is not sensitive to similarity transform, shape deformation, and noise.

We propose an ATR algorithm to recognize nonoccluded and partially occluded targets in ground-to-ground FLIR images. It introduces a partitioning method of the target boundary into upper, lower, left, and right regions. Also, it utilizes the proposed invariant features and a new classification method.

Section 2 briefly introduces our segmentation algorithm.24 Section 3 describes the proposed global and
local feature extraction using boundary partitioning. Section 4 describes a new classification method using multiple feature vectors and a combination of MLPs. In Sec. 5, invariance of the proposed features is analyzed under scaling, rotational transform, and noise. Using natural FLIR images of three different targets, recognition performance of the proposed ATR algorithm is compared with those of various conventional feature sets. Finally, conclusions are presented in Sec. 6.

2 Target Segmentation

An overall block diagram of the proposed ATR system is shown in Fig. 1. The first half of the block diagram is for target segmentation. An accurate representation of a target boundary is very important in a feature-based ATR system because the performance of the ATR system critically depends on the segmentation results. Natural FLIR images, which are obtained from battlefield, may have much background noise. The noise should be eliminated in the segmentation process to get a more accurate target boundary. The authors have proposed a segmentation algorithm of an FLIR image using fuzzy thresholding and edge detection, as shown in Fig. 1. The segmentation algorithm utilizes the results of Canny edge detection that include the Gaussian smoothing process to reduce noise effect. Figure 2 shows examples of FLIR images that are used in this work. Figure 3(a) shows the segmented image of Fig. 2(c), which is segmented by the proposed segmentation algorithm.

When soldiers identify military vehicles, the main gun of the vehicle can be a good feature for target recognition. However, the segmentation procedure usually misses the main gun because it is difficult to accurately extract the main gun from a noisy image. Thus, we remove the main gun by a binary morphological opening operation to extract stable features only. In the opening operation, the structuring element is a rectangular window with 5 × 5 pixels. An optimum size of the structuring element is analyzed in Sec. 5. After eliminating the main gun, a target boundary is extracted by a conventional boundary following the algorithm. Figure 3(b) shows the target boundary without the main gun of Fig. 2(c).

3 Proposed Feature Extraction

Feature extraction is a process by which a binary image is transformed into a new small feature vector. Recognition of a target is usually performed by using features that have the following requirements.

1. The distinction between the features of two classes that have different shapes should be as large as possible.
2. The variance of the feature in the same class has to be as small as possible.
3. The features have to be insensitive to digitization error, minor shape deformation, and noise.
4. The number of features used in classification should be as small as possible.

We try to find a new invariant feature set that is not sensitive to similarity transform, shape deformation, and noise. To find more robust features, many geometric shape features are extracted from radial and distance functions of a target boundary. The extracted shape features are tested to see whether they are invariant to the similarity transform and noise. Section 5 describes how to select invariant features against similarity transform and noise. Then, a sequential forward selection (SFS) algorithm is applied to the invariant features to classify the features having larger interclass differences. In the SFS algorithm, the leave one out (LOU) test, using a minimum distance classifier, is ap-

![Fig. 1 Overall block diagram of the proposed ATR algorithm.](image1)

![Fig. 2 Examples of three military vehicles at an aspect angle of 90 deg: (a) target A, (b) target B, and (c) target C.](image2)

![Fig. 3 (a) Segmentation result of Fig. 2(c), and (b) target boundary of Fig. 2(c) after removing the main gun.](image3)
plied to the features of training set. The following subsections describe the selected invariant features having larger interclass differences.

### 3.1 Global Features

The target boundary is represented by the Euclidean distance between the boundary pixels and the centroid of the target boundary. Let the N boundary pixels be described by ordered pairs \([x(i), y(i)], i = 1, 2, ..., N\). The Euclidean distance between the centroid and a boundary pixel \([x(i), y(i)]\) represents a 1-D radial function of the contour as follows,

\[
z(i) = \left[ (x(i) - \bar{x})^2 + (y(i) - \bar{y})^2 \right]^{1/2},
\]

where \([\bar{x}, \bar{y}]\) is the centroid of the target boundary. Figure 4 is an example of \(z(i)\) of Fig. 3(b). We define four global features from the radial function \(z(i)\), which are all invariant to the similarity transform, as follows.

**Amplitude variation.** Amplitude variation, \(G_1\), is a dispersion statistic of the radial function of Fig. 4 and is defined by

\[
G_1 = \frac{1}{N} \sum_{i=1}^{N} \left[ z(i) - \bar{z} \right]^2,
\]

where \(\bar{z}\) is the mean of \(z(i)\).

**Skewness.** Skewness of \(z(i)\) is defined with respect to its mean by

\[
G_2 = \frac{1}{N} \sum_{i=1}^{N} \left[ z(i) - \bar{z} \right]^3.
\]

If \(G_2\) is zero, the radial function is symmetric with respect to \(\bar{z}\).

### 3.2 Partitioning of a Target Boundary

As a human visual system usually uses global and local information to recognize an object, our approach utilizes global and local features to recognize a target. We intend to recognize nonoccluded targets by global and local features, and partially occluded targets by local features from nonoccluded regions.

To recognize a partially occluded target, we assume that the elevation angle of a target is nearly zero as previous researches assumed. This assumption is reasonable because targets of this application are ground vehicles. At first, two extreme points are selected to divide a target into upper (turret) and lower (track) regions. The left extreme point is a boundary point that is on the left-hand side of the centroid and has the longest distance from the centroid. The right extreme point is defined in a similar way on the right-hand side of the centroid. These left and right extreme points are named \(P_{cl}\) and \(P_{cr}\), respectively, as shown in Fig. 5(a). A line defined by the two extreme points, \(P_{cl}\) and \(P_{cr}\), divides the target boundary into upper and lower re-
regions, as shown in Figs. 5(b) and 5(c). To partition the boundary contour into left and right regions, we define the uppermost point \( P_u \) that is a peak point of the target boundary as shown in Fig. 6(a). The far-left point \( P_{ul} \) from the uppermost point \( P_u \) is a boundary point that is on the left-hand side of \( P_u \) and has the longest distance from \( P_u \). The far-right point \( P_{ur} \) from \( P_u \) is defined in the similar way on the right-hand side. As shown in Figs. 6(b) and 6(c), the left and right region boundaries are contours from \( P_u \) to \( P_{ul} \) and from \( P_u \) to \( P_{ur} \), respectively. Figures 5 and 6 show four local boundaries of Fig. 3(b). Sensitivity of extreme points to noise is analyzed in Sec. 5.

### 3.3 Local Features

When a target is partially occluded, features that are extracted from nonoccluded regions can be robust to the partial occlusion. Figure 7 shows some examples of partially occluded targets. To make some features robust to partial occlusion, we partition the target boundary into several local boundaries. Because the same features are extracted from each partitioned boundary, we explain how to extract the local features of the upper region as shown in Fig. 5(b). When \( M \) contour pixels of the upper region are described by an ordered pair \([x(i),y(i)], i = 1,2,\ldots,M\), as shown in Fig. 8, the Euclidean distance between the line drawn by two end points and a boundary pixel \([x(i),y(i)]\) forms a 1-D sequence. When the line equation of two end points is given as \( ax + by + c = 0 \), the distance function \( d(i) \) is defined by

\[
d(i) = \frac{|ax(i)+by(i)+c|}{(a^2+b^2)^{1/2}}.
\]

An example of \( d(i) \) is shown in Fig. 9, which is obtained

---

**Fig. 6** (a) Definition of \( P_u \), \( P_{ul} \), and \( P_{ur} \), (b) partitioned boundary of left region, and (c) partitioned boundary of right region.

**Fig. 7** Examples of partially occluded targets.

**Fig. 8** Definition of distance function for local feature extraction.
Amplitude variation. Amplitude variation of \( d(i) \) is defined by
\[
L_1 = \frac{1}{M} \sum_{i=1}^{M} [d(i) - \bar{d}]^2, \tag{7}
\]
where \( M \) is the number of local boundary pixels and \( \bar{d} \) is the mean of \( d(i) \).

Skewness. Skewness of \( d(i) \) with respect to its mean is defined by
\[
L_2 = \frac{1}{M} \sum_{i=1}^{M} [d(i) - \bar{d}]^3. \tag{8}
\]

Flatness. Flatness of the military vehicles is defined by
\[
L_3 = \frac{D_{\text{width}}}{M}, \tag{9}
\]
where \( D_{\text{width}} \) is the distance between two extreme points. When \( M \) is large compared with \( D_{\text{width}} \), the shape of the partitioned boundary of the vehicles is not flat.

The ratio of \( D_{\text{width}} \) and maximum of \( d(i) \). It is defined by
\[
L_4 = \frac{\max[d(i)]}{D_{\text{width}}}. \tag{10}
\]

4 Proposed Classification Method Using Multiple MLPs

When a target is not occluded, global and local features are all valid. However, when a target is partially occluded, local features of the nonoccluded regions are valid and can be used to classify the target. Partially occluded targets comprise incorrectly segmented targets and occluded targets. The feature vector used for nonoccluded targets is defined by
\[
F_1 = \{G1,G2,G3,G4,L1_A,L2_A,L3_A,L4_A,L3_B,L4_B\}, \tag{11}
\]
where subscripts \( A \) and \( B \) mean that local features are extracted from the upper and lower regions of the partitioned target boundaries, respectively, as shown in Fig. 5. Feature vectors used for partially occluded targets are defined as follows,
\[
F_2 = \{L1_A,L2_A,L3_A,L4_A\}, \tag{12}
\]
\[
F_3 = \{L1_C,L2_C,L3_C,L4_C\}, \tag{13}
\]
\[
F_4 = \{L1_D,L2_D,L3_D,L4_D\}, \tag{14}
\]
where subscripts \( C \) and \( D \) mean that local features are extracted from the left and right regions of the partitioned...
target boundaries, respectively. Feature vector $F_2$ is robust to occlusion of the lower region, whereas $F_3$ and $F_4$ are robust to occlusion of the right and left sides of a target, respectively.

Three MLPs in neural networks are used to classify targets. In the training phase, MLP1 and MLP2 are trained by feature vectors $F_1$ and $F_2$ of the training data, respectively. MLP3 is trained by feature vectors $F_3$ and $F_4$. Each MLP has three layers as shown in Fig. 10(a). The desired output of the correct target class is set to one, and the other outputs are set to zero. MLP1 has 10 input neurons, 20 hidden neurons, and 3 output neurons. MLP2 has 4 input neurons, 25 hidden neurons, and 3 output neurons. MLP3 has 4 input neurons, 30 hidden neurons, and 3 output neurons. Individual neurons of each MLP have a sigmoid function. The input features are normalized by Gaussian distribution with zero mean and unit variance. The Levenberg-Marquardt algorithm is used for learning, where the learning rate is 0.01 and initial weights are randomly assigned. The proper choice of the parameters can affect the training times, but accuracy of the proposed ATR algorithm is not sensitive to these parameters.

In the test phase, four feature vectors extracted from a test image are applied to each MLP as shown in Fig. 10(b). Final outputs are computed by the average of the four MLP outputs. Then, the classification result is a target having the maximum among the final outputs. Averaging of the four MLP outputs makes one output dominant.

### 5 Experimental Results and Discussions

#### 5.1 Invariance of the Proposed Features for Similarity Transform

Because radial and distance functions are independent of translation, the proposed features that are defined from the radial and distance functions are invariant to translation. To demonstrate the invariance of the features for rotational and scaling transforms, the proposed feature set is compared with Zernike moment invariants, affine moment invariants, and affine Fourier descriptors. Ten rotated images and ten scaled images are generated from Fig. 3(b). The rotational angles range between $-50$ to $50$ deg and each rotated image is obtained at 10-deg intervals. The scaling ratios are ranged between 0.5 to 2.0. Each scaled image is obtained from 50 to 90\% scaling from the original image with 10\% intervals, and from 120 to 200\% with 20\% intervals.

To examine the performance of invariance more clearly, a figure of merit $E$ is defined by a mean square error (MSE) criterion as follows,

$$E = \frac{1}{N_s} \sum_{n=1}^{N_s} [(f(n) - \bar{f})^2], \tag{15}$$

where $f(n)$ is a normalized feature value of $n$'th rotated or scaled images, $\bar{f}$ is the mean value of $f(n)$, and $N_s$ is the number of rotated or scaled images. In Eq. (15), feature values are normalized by the maximum value of each feature so that every feature value ranges from zero to one.

Tables 1 and 2 show feature variances of ten scaled and rotated images, respectively. Four global features of an overall target boundary and four local features of a partitioned upper boundary, as shown in Fig. 5(b), are selected for comparison with conventional features. In the Zernike moment invariants, four Zernike moments are derived from the silhouette and the other four are from the boundary, which are $S_{S_1}, S_{S_2}, S_{S_3}, S_{S_4}, S_{B_1}, S_{B_2}, S_{B_3}$, and $S_{B_4}$. The subscript $S$ means that target silhouettes are used to compute the moment invariants, whereas $B$ means that target boundaries are used to compute the moment invariants. In the affine moment invariants, four affine moments are derived from the target silhouette and the other four are from the boundary, which are $I_{S_1}, I_{S_2}, I_{S_3}, I_{S_4}, I_{B_1}, I_{B_2}, I_{B_3}$, and $I_{B_4}$. In the affine Fourier descriptors, equidistance sampling is applied to a target boundary and the sampling points are 256 points. The four lowest frequency coefficients are selected from each side of the frequency spectrum, which are $W_2, W_3, W_4, W_5, W_{-2}, W_{-3}, W_{-4}$, and $W_{-5}$. Table 1 shows that the proposed global and local

### Table 1 Comparison of feature variances of ten scaled images. ZMI are Zernike moment invariants. AFI are affine moment invariants, AFD are affine Fourier descriptors, and GL are proposed global and local features. (Unit: 10$^{-3}$ MSE.)

<table>
<thead>
<tr>
<th></th>
<th>$S_{S_1}$</th>
<th>$S_{S_2}$</th>
<th>$S_{S_3}$</th>
<th>$S_{S_4}$</th>
<th>$S_{B_1}$</th>
<th>$S_{B_2}$</th>
<th>$S_{B_3}$</th>
<th>$S_{B_4}$</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZMI</td>
<td>108.6</td>
<td>115.9</td>
<td>109.6</td>
<td>113.6</td>
<td>109.3</td>
<td>115.8</td>
<td>106.7</td>
<td>107.6</td>
<td>110.9</td>
</tr>
<tr>
<td>AMI</td>
<td>$I_{S_1}$</td>
<td>$I_{S_2}$</td>
<td>$I_{S_3}$</td>
<td>$I_{S_4}$</td>
<td>$I_{B_1}$</td>
<td>$I_{B_2}$</td>
<td>$I_{B_3}$</td>
<td>$I_{B_4}$</td>
<td>Average</td>
</tr>
<tr>
<td>AFD</td>
<td>$W_2$</td>
<td>$W_3$</td>
<td>$W_4$</td>
<td>$W_5$</td>
<td>$W_{-2}$</td>
<td>$W_{-3}$</td>
<td>$W_{-4}$</td>
<td>$W_{-5}$</td>
<td>Average</td>
</tr>
<tr>
<td>GL</td>
<td>$G_1$</td>
<td>$G_2$</td>
<td>$G_3$</td>
<td>$G_4$</td>
<td>$L_1$</td>
<td>$L_2$</td>
<td>$L_3$</td>
<td>$L_4$</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>18.9</td>
<td>1.0</td>
<td>0.6</td>
<td>1.7</td>
<td>34.6</td>
<td>1.7</td>
<td>0.1</td>
<td>7.4</td>
</tr>
</tbody>
</table>
features are much more invariant to the scaling transforms than conventional feature sets. Table 2 shows that the variance of the proposed features is slightly larger for the rotational transform than that of Zernike moment invariants. However, the difference is very small.

5.2 Noise Sensitivity of the Proposed Features

To analyze the noise sensitivity, we generated noisy boundaries from the original training images. Gaussian noise with zero mean is added to the target boundaries. Its standard deviation \((\sigma_x, \sigma_y)\) ranges from 0.0 to 2.0 with 0.4 intervals. At each noise level, feature values are computed from 150 original boundaries and 150 noisy boundaries with various standard deviations. MSE criterion is used to evaluate noise sensitivity. Figure 11 shows examples of the noisy target boundaries. Noise sensitivity of the proposed features is compared with the traditional feature sets by experiments. Among the traditional feature sets, such as affine Fourier descriptors, affine moment invariants, and Zernike moment invariants, the second and third order Zernike moment invariants are most insensitive to noise. Figure 12(a) shows the comparison with the proposed features and the second order Zernike moment invariants for noisy boundaries. The proposed global and local features, except for \(G_2\) and \(L_2\), are more or less sensitive to the second order Zernike moment invariants. Figure 12(b) shows the comparison with the proposed features \(G_2\), \(L_2\), and the third order Zernike moment invariants. \(G_2\) and \(L_2\) are more or less sensitive to the third order Zernike moment invariants. By the same experiments, we know that the noise sensitivity of \(L_1\) to \(L_4\). In conclusion, the noise sensitivity of the proposed features is comparable to that of the low order Zernike moment invariants.

5.3 Structuring Element for Elimination of Main Gun

To obtain stable features against noise, the proposed algorithm uses a morphological opening operator with a \(5 \times 5\) structuring element to eliminate main gun. Size of the

<table>
<thead>
<tr>
<th></th>
<th>(S_{s1})</th>
<th>(S_{s2})</th>
<th>(S_{s3})</th>
<th>(S_{s4})</th>
<th>(S_{b1})</th>
<th>(S_{b2})</th>
<th>(S_{b3})</th>
<th>(S_{b4})</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZMI</td>
<td>0.1</td>
<td>0.2</td>
<td>0.6</td>
<td>1.6</td>
<td>0.2</td>
<td>0.6</td>
<td>23.1</td>
<td>13.8</td>
<td>5.0</td>
</tr>
<tr>
<td>AMI</td>
<td>5.9</td>
<td>12.7</td>
<td>8.6</td>
<td>21.7</td>
<td>23.2</td>
<td>133.5</td>
<td>35.4</td>
<td>37.5</td>
<td>34.8</td>
</tr>
<tr>
<td>AFD</td>
<td>71.5</td>
<td>79.3</td>
<td>81.9</td>
<td>71.9</td>
<td>78.0</td>
<td>86.3</td>
<td>73.0</td>
<td>82.3</td>
<td>78.0</td>
</tr>
<tr>
<td>GL</td>
<td>(G_1)</td>
<td>(G_2)</td>
<td>(G_3)</td>
<td>(G_4)</td>
<td>(L_1)</td>
<td>(L_2)</td>
<td>(L_3)</td>
<td>(L_4)</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>21.3</td>
<td>0.2</td>
<td>1.0</td>
<td>1.1</td>
<td>37.6</td>
<td>2.5</td>
<td>0.2</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Fig. 11 Examples of the noisy target boundaries: (a) \(\sigma_x=\sigma_y=1.2\), and (b) \(\sigma_x=\sigma_y=2.0\).

Fig. 12 Noise sensitivity of the proposed features: (a) MSE of global and local features except for \(G_2\) and \(L_2\), and (b) MSE of \(G_2\) and \(L_2\).
The structuring element is determined experimentally. We applied various sizes of the structuring elements to nonoccluded targets, obtained feature values from the smoothed target boundaries, and classified them by using MLP1. Figure 13 shows the classification result, which describes the optimum size of the structuring element as $5 \times 5$ for our database. In the experiment, the distance between the targets and the imaging sensors was about 1100 to 1200 m and the image size was $256 \times 256$. Since we assume the distance between imaging sensor to a target could be measured by a laser range finder, we can adjust the size of the structuring element according to the measured distance.

### 5.4 Classification Experiments in FLIR Images

The FLIR images to be trained and tested are gathered by an infrared thermal sight installed on a military vehicle. The targets of interest are three military vehicles, whose examples are shown in Fig. 2. In the database, the aspect angle of the targets $\theta$ ranged between $0 < \theta < 360$ deg and the elevation angle was nearly zero. The imaging sensor was located at a distance of about 1100 to 1200 m from the target. The image size is $256 \times 256$ pixels with 256 gray levels. A training set consists of 50 images per target. A test set consists of images of the same vehicles as the training set, but it is independent of the training set. The test set consists of 100 nonoccluded images and 100 partially occluded images per target. The partially occluded images are randomly generated by artificial occlusion for the nonoccluded test images. As shown in Fig. 7, occlusion areas are randomly selected and generated by Paintshop software. However, the top area of the turret is not occluded in the experiment. Occlusion of the turret can happen in the case of disguise only, where tank classification is very difficult due to the similar hull shape. In the experiment, the occlusion ratio is about 21.0%, which is defined by

$$R_o = \frac{A - A_v}{A} \times 100\%,$$

where $R_o$ is the occlusion ratio, $A$ is area of the nonoccluded target, and $A_v$ is the visible area of a partially occluded target.

To evaluate the classification efficiency, the proposed method is compared with Zernike moment invariants,32 affine moment invariants,33 and affine Fourier descriptors.34 Zernike and affine moment invariants are computed and aligned from low to high order moments for the boundary and silhouette. Affine Fourier descriptors are computed and aligned from low to high frequency coefficients of a target boundary. A minimum distance classifier is used to classify targets for three conventional feature sets. Figure 14 shows the classification results of the nonoccluded and partially occluded images.

### Table 3: Classification results from conventional features.

<table>
<thead>
<tr>
<th>Test images</th>
<th>Zernike moment invariants</th>
<th>Affine moment invariants</th>
<th>Affine Fourier descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonoccluded images</td>
<td>277/300 = 92.3%</td>
<td>269/300 = 89.7%</td>
<td>265/300 = 88.3%</td>
</tr>
<tr>
<td>Occluded images</td>
<td>154/300 = 51.3%</td>
<td>149/300 = 49.7%</td>
<td>152/300 = 50.7%</td>
</tr>
<tr>
<td>Total images</td>
<td>431/600 = 71.8%</td>
<td>418/600 = 69.7%</td>
<td>417/600 = 69.5%</td>
</tr>
</tbody>
</table>

### Table 4: Confusion matrix for nonoccluded test images by using the proposed method.

<table>
<thead>
<tr>
<th>Input</th>
<th>Target A</th>
<th>Target B</th>
<th>Target C</th>
<th>Correct classification rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target A</td>
<td>94</td>
<td>5</td>
<td>1</td>
<td>94.0</td>
</tr>
<tr>
<td>Target B</td>
<td>4</td>
<td>94</td>
<td>2</td>
<td>94.0</td>
</tr>
<tr>
<td>Target C</td>
<td>4</td>
<td>4</td>
<td>90</td>
<td>90.0</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>92.7</td>
</tr>
</tbody>
</table>
occluded test images. As shown in Fig. 14, the correct classification rate of conventional feature sets is almost satu-
rated when the number of features is larger than eight. Thus, the lowest eight features for each conventional feature set are selected and used for inputs to the MLP classifier. The same MLP structure is used for all conventional features, which has 8 input neurons, 20 hidden neurons, and 3 output neurons. They are separately trained and tested. The Levenberg-Marquardt algorithm is used for learning, where the learning rate is 0.01 and initial weights are randomly assigned. All neurons have sigmoid functions and all feature values are normalized by Gaussian distribution with zero mean and unit variance. Classification results from three conventional feature sets using the MLP are listed in Table 3.

The classification results from the proposed method are listed in Tables 4 and 5 in the form of a confusion matrix. Table 4 is for nonoccluded targets and Table 5 is for partially occluded targets in FLIR images. It utilizes global and local shape features recognizing nonoccluded and partially occluded targets in FLIR images. It utilizes global and local shape features recognizing nonoccluded and partially occluded targets in FLIR images. It utilizes global and local shape features recognizing nonoccluded and partially occluded targets in FLIR images. It utilizes global and local shape features recognizing nonoccluded and partially occluded targets in FLIR images.

Table 5 Confusion matrix for partially occluded test images by using the proposed method.

<table>
<thead>
<tr>
<th>Input</th>
<th>Target A</th>
<th>Target B</th>
<th>Target C</th>
<th>Correct classification rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target A</td>
<td>63</td>
<td>18</td>
<td>19</td>
<td>63.0</td>
</tr>
<tr>
<td>Target B</td>
<td>18</td>
<td>74</td>
<td>8</td>
<td>74.0</td>
</tr>
<tr>
<td>Target C</td>
<td>13</td>
<td>16</td>
<td>71</td>
<td>71.0</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>69.3</td>
</tr>
</tbody>
</table>

6 Conclusions

We propose an efficient ground-to-ground ATR algorithm recognizing nonoccluded and partially occluded targets in FLIR images. It utilizes global and local shape features extracted from partitioned target boundaries. The proposed features are much more invariant to scaling and rotational transforms than conventional features. The proposed classification method using multiple feature vectors and MLPs is highly applicable to recognizing nonoccluded and partially occluded targets. The experimental results demonstrate the superiority of the proposed ATR algorithm. We expect that the proposed algorithm can be applied to current and future ATR systems.

References

Sun-Gu Sun received the BS and MS degrees in electrical engineering from Han- yang University, Korea, in 1987 and 1989, respectively. He has been a researcher and senior member at the Agency for De- fense Development (ADD) in Korea since 1989. He is currently studying for his PhD degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea. His current re- search interests include image processing, pattern recognition, computer vision, and ATR.

HyunWook Park received the BS degree in electrical engineering from Seoul Na- tional University, Korea, in 1981, and the MS and PhD degrees in electrical engi- neering from the Korea Advanced Institute of Science and Technology (KAIST), Seoul, Korea, in 1983 and 1988, respectively. He has been a professor in the electrical engi- neering department at KAIST since 1993. He was a research associate at the Univer- sity of Washington from 1989 to 1992 and was a senior executive researcher at the Samsung Electronics Company, Limited, from 1992 to 1993. He is a senior member of IEEE and a member of SPIE. His current research interests include image computing system, image compression, medical imaging, and multimedia system.