Target Recognition Using Multi-Scale Gabor Filters

Xing Wu and Bir Bhanu
College of Engineering
University of California, Riverside, CA 92521

Abstract
This paper presents a model-based target recognition approach that uses a hierarchical Gabor wavelet representation. The approach combines the global and local Gabor-based measures for target indexing into the model database. A Gabor grid, a topology-preserving map, efficiently encodes both signal energy and structural information of a target in a sparse multi-resolution representation. The Gabor grid subsamples the Gabor wavelet decomposition of a target model and is deformed to allow the indexed target model match with the image data. Flexible matching between the model and the image minimizes a cost function based on local similarity and geometric distortion of the Gabor grid. Grid erosion and repairing is performed whenever a collapsed grid, due to target occlusion, is detected. The results on infrared imagery are presented, where targets undergo rotation, translation, scale, occlusion and aspect variations.

1 Introduction

Target recognition in real-world situations is difficult because a robust solution has to consider multiple factors such as, contrast, signature, scale, and aspect variations; noise and spurious low resolution sensor data; and high clutter, target occlusion and articulation. Current approaches using the model-based paradigm rely on extracting shape primitives, silhouette and contours, colors, and invariant object features for matching. The performance of these methods is acceptable when targets are well defined, have very high contrast, and are at close range. However, it does not gracefully degrade when competitive clutter and target shape distortion are present in the input data. Another associated problem is the large number of hypothesized target and clutter types that need to be verified. The problem is compounded when the image contains complicated background and many targets.

In this paper, we will show that the multi-scale decomposition and elastic Gabor grid based flexible matching is a robust method for target recognition. Our method requires no explicit use of shape features and contours which cannot be reliably detected under low contrast, noise, high clutter and occlusion. The flexible matching mechanism can tolerate significant amount of target distortions due to viewing scale, aspect and environment.

![Figure 1: Target recognition is an iterative process of matching based on Gabor grid.](image)

2 Gabor Wavelet Representation

Both model and image data are represented by magnitude and phase responses of Gabor wavelet filters. A complex Gabor function [1, 2] is given as,

\[ G(x, \psi) = \exp \left( -\frac{\psi^2 x^2}{2\sigma^2} \right) \cdot \exp(i\psi x), \]  

(1)
where \( \vec{\psi} \) is the 2-D wave propagation vector defined in (2). It is determined by the modulation frequency \( \omega \) and orientation \( \phi \) of a Gabor kernel. The parameter \( \sigma \) regulates the size of the Gaussian window, and \( x \) is the coordinate of the filter kernel relative to the spatial centroid of the Gaussian window. In this work, Gabor wavelet filters \( G_\vec{\psi} \) are defined by 7 logarithmically spaced center frequencies \( \omega_k \) and 8 orientations \( \phi_l \). It samples the frequency domain into 56 bandpassed spectral channels. The log-polar sampling of \( \vec{\omega} \) is given as,

\[
\vec{\omega}_{k,l} = \omega_k e^{i\phi_l}, \quad \omega_k = \rho^k \omega_0, \quad \phi_l = l \phi_0,
\]

where \( \rho = \sqrt{2}, \quad \omega_0 = \pi/16, \quad \phi_0 = \pi/8 \), and the size of the Gaussian window varies inversely with center frequency \( \omega \).

Decomposition of an image \( I(x) \) is obtained by convolving it with the complex Gabor wavelet filter kernels \( G(x, \vec{\psi}) \). The magnitude of this decomposition at a spatial location \( x \) in the image forms a vector of length \( k \times l \) which is referred to as a Gabor probe \( P(x) \),

\[
P(x) = \left[ |(I * G_{\vec{\psi}})|[x], \ldots \right] \quad \forall k, l.
\]

Each Gabor probe can extract localized signal energy and structural patterns in a neighborhood whose extent is defined by the size of the Gabor kernel and is inversely proportional to its modulation frequency.

The Gabor wavelet decomposition of a target is an iconic multi-resolution template. To reduce the inter-pixel redundancy, subsampling this template forms an elastic Gabor grid \( G_D \) which covers the whole target with fixed distance \( D \) and \( N \times M \) nodes in the \( x \) and \( y \) directions respectively. Gabor probe \( P(x) \) together with its spatial location \( x \) are stored at each grid node \( v_j \). An edge \( e_j \) which links between probes serves as a spatial constraint, and can be deformed like a spring to make a model probe match with a distorted target probe.

Targets in input sensor data often appear at different scales when viewed from different aspect and distance. Realizing that our Gabor grid and probe representation is invariant to scale changes, we can achieve the goal of using a single model set to recognize targets having scale variations from the model. A Gabor probe can be seen as a set of concentric multi-scale disk platters indexed by the modulation frequency and orientation of the Gabor wavelet. Probes of a scaled target will then have similar 'platters' to the model with proper displacement along the frequency axis.

If a target scale \( s \) is a multiple of the modulation frequency spacing \( \rho \), a new model grid \( G_D' \) at this scale can be generated by scaling down edges of the Gabor grid the same factor \( s \), and shifting Gabor probe \( P_j \) at each node \( v_j \) from the corresponding frequency index \( \omega \),

\[
e_j \rightarrow e_j/(\sqrt{2})^s, \\
P_j(\omega_s, \phi_l) \rightarrow P_j(\omega_{(k-s)}, \phi_l).
\]

In case \( s \) is not a multiple of \( \rho \), we round the scale factor \( s \) to the closest multiple of \( \rho \), and the subsequent flexible matching can overcome this small scale distortion without the need for interpolation.

### 3 Flexible Matching

Target Recognition using a Gabor grid is an iterative process of matching. It first finds the optimal global placement of the grid over an image while the grid is kept rigid. Estimation of target scale is necessary when sensor information is not available. Then, deformation of the grid allows model probes match with local features of the distorted image. Grid erosion and repairing are performed whenever a collapsed grid is detected and target occlusion is suspected. Finally, the correct target aspect is selected by the evaluation criteria and rules.

#### 3.1 Flexible Model Matching

Correlations between an image decomposition and all model templates (Gabor grid) are performed to find the location and index of a target in an input image. To improve efficiency, this process is restricted to areas where high Gabor magnitude responses are observed. The peak response in the correlation result is selected as the target location and used as indexing into the target database. When the target scale is unknown, model grids in multiple scales are first used in correlations with the target decomposition to estimate the target scale before a target aspect and location hypothesis is generated.

After location indexing, flexible matching starts to further verify the hypothesis, in which nodes of the model grid can be moved locally and independently to find the best matched image probes. The process minimizes a cost function \( C \) which is balanced between grid distortions \( D \) and local similarities \( S \).

\[
C = \lambda \sum_i^N D(v_i) - \sum_i^N S(P_i^I, P_i^M)
\]

During matching, grid nodes are visited in random order and take a random step. A move of a node is accepted when either,

- the global cost \( C \) is reduced, or
- \( \Delta C \) satisfies a probability \( \exp(-\Delta C/T) \).

However, the topological property of the grid has to be preserved for a rigid target.

The flexible matching is a dynamic process controlled by the following parameters. The flexibility parameter \( \lambda \) controls the degree of deformation allowed for the Gabor grid. The temperature \( T \) controls the probability of finding the best matched Gabor probes inside the target data. The process starts with a higher temperature, which helps it to quickly converge to the stable state, and cools down generally at a constant rate \( \gamma \) until it is stable or reaches freezing temperature. Underestimating the value of \( \lambda \) will produce collapsed grid during matching. Overestimating the value of \( \lambda \) will keep the process from generating the optimal result because the grid is too rigid.

#### 3.2 Local Similarity and Grid Distortion

Local similarities between a model and a target is measured between Gabor probes and the Gabor decompos-
tion of a target at each spatial location of a grid node.

\[ S(\vec{r}_p, \vec{r}_q) = \frac{1}{\eta} \frac{\vec{r}_p \cdot \vec{r}_q}{|\vec{r}_p| \cdot |\vec{r}_q|} \min \left( \frac{|\vec{r}_p|}{|\vec{r}_q|}, \frac{|\vec{r}_q|}{|\vec{r}_p|} \right), \tag{6} \]

here \( \eta \) is a normalization term to suppress difference in target contrast and signature.

When a grid is deformed, it comes with two kinds of distortions, length and angular distortion as shown in Figure 2. They are measured by comparing the current grid with its original structure which has fixed length \( D \) and a rectilinear grid. The distortion for a node \( v_k \) is then calculated as:

\[ D(v_k) = \frac{1}{D^2} \sum_{i=0}^{3} (d_i - D)^2 + \sum_{i=0}^{3} \frac{\left( a_i - \sqrt{d_i^2 + d_{i+1}} \right)^2}{d_i^2 + d_{i+1}^2}, \tag{7} \]

where \( D \) is the fixed length of the model grid. The first term in (7) measures the grid length distortion, while second term measures the angular distortion of the grid.

### 3.3 Recognizing Occluded Targets

The existence of a collapsed grid after matching helps to detect potential target occlusion. A target can be occluded by either a background, a natural target, or another target. The occlusion may be more than 50% of the target size. After matching, grid nodes which belong to the occluded part may fold together, while nodes which belong to the non-occluded part can still find good matches. Then, grid erosion process starts by cutting collapsed nodes from the grid one by one, removing isolated sub-grid, and redo the grid replacement until either (1) the collapsed part of the grid is completely removed, or (2) the whole grid moves to a new location and no collapses are detected. This involves possible iterations among the following processes —correlation/ flexible matching/model refinement. After erosion, evaluation on the remaining grid is performed to find the correct aspect.

Although graph matching is an NP-complete problem, and collapsed grids come with various shapes, we only work in a restricted context where heuristics are used. This makes the problem solvable in polynomial time. Two graphs have been used to detect target occlusion and refine the grid placement. They are the connected-grid which is defined in the same manner as the elastic Gabor grid, and the complementary-grid which is created by connecting the center of each block in the connected grid. They are shown as thin and thick lines in Figure 3.

### 4 Evaluation of Matching

Regardless of whether or not a corresponding target is in the image, the process of matching always yields the best matching result. The match, corresponding to the correct target category, should have the lowest cost and smallest distortion among all the tested model grids. However, the expected result may not come out distinctively when large aspect variations and signature changes are presented in the input data. Sophisticated evaluation criteria are required to distinguish them.

**Flexible-matching cost \( \varsigma \):** It is given by (5). \( \varsigma \) combines the similarity measure between probes and grid distortions. To suppress background probes, the similarity measure term is multiplied by the minimum amplitude of either the model or the target probe.

**Dissimilarity \( \varepsilon \):** It is defined as the difference between a perfect matching and the actual matching. The similarity is 1 for perfectly matched Gabor probes and should be less than 0.5 for randomly matched probes.

\[ \varepsilon = \sum_i N \left[ 1 - S(P_i^l, P_i^m) \right]^2; \tag{8} \]

**Displacement \( \delta \):** It is defined as the local translational displacement between matched Gabor probes. Given that a model probe \( P_m \) matches with an image probe \( P_i \), the localized phase difference along the direction of modulation \( \phi_i \) can be used to estimate this displacement and evaluate the matching error,

\[ d(\omega_k, \phi_i) = \frac{\Delta \delta(\omega_k, \phi_i)}{\omega_k} \tag{9} \]

Since phase difference is restricted to \((-\pi, \pi]\), the maximal displacement estimated is limited to \((-\pi/\omega_k, \pi/\omega_k]\). Therefore, lower center frequency channels in a Gabor wavelet can estimate larger displacement with less accuracy, while higher ones can estimate small displacement with higher accuracy. The displacements estimated by different filter frequency bands for a probe \( p \) are combined in a similar way as \([4]\) to overcome the bandpass noise,

\[ d_p(\phi_i) = \frac{\sum_k w(\omega_k, \phi_i) d(\omega_k, \phi_i)}{\sum_k w(\omega_k, \phi_i)} \tag{10} \]

and local magnitude responses of the model and the image are used as weight,

\[ w(\omega_k, \phi_i) = \min \left[ \frac{m_m(\omega_k, \phi_i)}{m_i(\omega_k, \phi_i)}, \frac{m_i(\omega_k, \phi_i)}{m_m(\omega_k, \phi_i)} \right]. \]
An average of the amplitude of the local displacement over all grid nodes is used as the matching error $\delta$. Since grid nodes are not necessarily located at high Gabor magnitude response, a group of points, selected from the target with high Gabor magnitude responses, are used and back-projected onto the matched model. More accurate phase measures can be obtained using these projected pairs than using grid nodes.

5 Target Recognition Experiments

A series of experiments have been carried out to test our algorithms and three image databases have been created.

1. Synthetic FLIR (forward looking infrared) images are generated by computer that simulate a given environmental conditions. They are used to find the relationship between the recognition performance and the effect of environmental changes on target signatures in order to anticipate and adapt the performance degradation in real applications.

2. Selected targets from the second generation FLIR images.

3. FLIR images with occluded targets.

We have also done experiments on TV images and similar performance was obtained.

5.1 Synthetic FLIR Data

To find the relationship of recognition performance with changing environmental conditions, we have investigated the effect of air temperature and solar energy on target signatures. The air temperature changes from 12°C to 26°C for a period of 23 hours in July 1984 at Grayling, M. is plotted in Figure 5. A total of 18 Models are synthesized at 4pm of the day, with depression angles of 0° and 20°, and aspect angles from 0° to 180° with 22.5° separation. A total of 138 target signatures are generated for testing. They have depression angles of 10° and 30°, and aspect angles of 60°, 90° and 120° and were obtained during 23 hours. Six target signatures including background, with 60° aspect angle and 10° depression angle at each 4 hours time segment, are

![Figure 4: Target signature variations under environmental condition depicted in Figure 5(a). They are sampled from the 138 images used in the experiment.](a) 7am (b) 11am (c) 3pm (d) 7pm (e) 11pm (f) 3am

Figure 5: Recognition performance for 23 infrared signatures under different weather conditions. The recognition power is defined as the difference of matching cost between the best and the second best match shown in Figure 4. Change in signature from frame to frame are clearly observable.

The recognition performance shown in Figure 5 is obtained from experiments on 23 infrared target signatures with 3 aspect angles and 0° depression angle. The correlation between curves of the recognition performance and the environmental parameter curve is clearly observable. Table 1 shows the results of total 207 recognition experiments using the synthesized FLIR models and images mentioned above. A successful recognition rate of 98% was achieved when the three described evaluation criteria were used together.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Images</th>
<th>Failures</th>
<th>Successful Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>207</td>
<td>79</td>
<td>61.8%</td>
</tr>
<tr>
<td>$\gamma+\varepsilon$</td>
<td>207</td>
<td>14</td>
<td>93.2%</td>
</tr>
<tr>
<td>$\gamma+\varepsilon+\delta$</td>
<td>207</td>
<td>5</td>
<td>97.6%</td>
</tr>
</tbody>
</table>

5.2 Second Generation FLIR Data

The second generation FLIR data has shown significant improvement in image resolution, contrast and noise. In the following two experiments, we show examples of target recognition with scale and aspect variations (Figure 6), and different target signatures (Figure 7). In these experiments, the distortion values correspond to 105m in viewing distance, 11° in depression angle, and 52° in aspect angle.

Recognizing a target with unknown scale involves tasks of, (1) estimating target scale, (2) location indexing, and (3) flexible matching. To find the correct target aspect, matching results are evaluated based on the three criteria discussed earlier and following rules in order.

1. Select the one having the lowest matching cost and the smallest dissimilarity.
2. Select the one having the smallest dissimilarity with conditions that its matching cost and dissimilarity are both lower than a defined threshold.
3. Select the one having the smallest displacement measure.
5.3 Target Occlusion

Grid erosion is performed whenever a collapsed grid is detected and target occlusion is suspected. The recognition is performed in the following way,

1. Detect collapsed matching result.

2. Create the connected and complementary grid defined, find and mark the collapsed grid nodes in both graphs.

3. Remove collapsed grid nodes and corresponding edges according to their spatial locations and relationship.

Figure 7: A target model is matched with three targets having scale, aspect variations, and different signatures.

Figure 8: An occluded target and result after grid erosion.

4. Remove isolated subgrids and reevaluate matching result based the remaining grid.

A target with an artificial 30% occlusion is shown in Figure 8(a). The initial matching results in a collapsed Gabor grid as shown Figure 8(b). Two subgrids survived after grid erosion are shown in Figure 8(c). After repairing, the remaining grid matched perfectly with the non-occluded part of the target shown in Figure 8(d).

6 Conclusions

Signature variations of 20° in depression angle, 22.5° in aspect and up to 50% occlusion are tolerable in our matching experiments reported here. Extensive evaluation will be carried out in the future.

Acknowledgment

This work was supported by ARPA/AFOSR grant F49620-93-1-0624 and ARPA grant MDA972-93-1-0010. The contents of the information does not necessarily reflect the position or the policy of the Government.

References


1994 Image Understanding Workshop

Proceedings
Hyatt Regency Hotel
13 - 16 November
Monterey, California
Volume I

Image Understanding Research Program
University of California at Riverside
Host of 1994 Image Understanding Workshop

Sponsored by:
Advanced Research Projects Agency
Software and Intelligent Systems Technology Office
Image Understanding Workshop

Proceedings of a Workshop
held in
Monterey, California

November 13 - 16, 1994

Volume I

Sponsored by:
Advanced Research Projects Agency
Software and Intelligent Systems Technology Office

This document contains copies of reports prepared for the ARPA Image Understanding Workshop. Included are Principal Investigator reports and technical results from both the basic and strategic computing programs within ARPA/SISTO sponsored projects and certain technical reports from selected scientists from other organizations.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the United States Government.