Automatic target shape recognition via deformable wavelet templates

Jin Li and C.-C. Jay Kuo
Signal and Image Processing Institute and Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089-2564
Phone: (213) 740-4660, Fax: (213) 740-4651
E-mail: lijin@sipi.usc.edu and cckuo@sipi.usc.edu

ABSTRACT

In this research, we propose the deformable wavelet template (DWT) for object shape description. DWT offers not only the global information at the lower scales, but also local features at higher differential scales. Thus, it provides a natural tool for multiresolution representation and can be used conveniently in a hierarchical matching procedure. In the presentation, we address the three main processing steps in the DWT-based ATR algorithm. They are namely, 1) image preprocessing and target shape extraction, 2) shape feature normalization, 3) wavelet decomposition. The topic of multiscale matching is also touched. We demonstrate the performance of our proposed algorithm with extensive experimental results in the presentation.

Keywords: automatic target recognition, deformable wavelet template, multiresolution, hierarchical match, shape description, shift, scale and rotational invariant,

1 INTRODUCTION

Shape or template representation is one of the fundamental building blocks for target identification. Ever since World War II, the Navy has used shape identification cards to discriminate between friendly and hostile targets. The same technique is also widely used in computer vision and pattern recognition [4]. One important advantage of using the shape feature of a target is that it contains fewer parameters since it can be represented by 1-D curves. The other advantage is that the shape feature can be normalized so that it has shift-, scale- and rotation- invariant properties. These properties greatly reduce the size of the target database and the complexity of target search algorithms, since all projections of a single target with different displacements, sizes and 2-D rotations can be merged into one item.

A widely discussed shape representation tool is the Fourier descriptor [5], [6], [8]. However, the Fourier descriptor has several obvious shortcomings in shape representation. First, since the Fourier basis is not local in the spatial domain, a local variation of the shape can affect all the Fourier coefficients. Another disadvantage of the Fourier descriptor is that it does not have a multiresolution representation. Whatever the size of the target and the noise level, matching has to be performed at a designated scale. Thus, the accuracy of matching results is low and the computational complexity of the method is high. Deformable wavelet templates (DWT) offer better global shape features at the low scales as well as local detail features at high differential scales. Since DWT provides a natural multiresolution representation, small targets can be matched in a lower resolution while large targets can be matched in a higher resolution.

The paper is organized as follows. As shown in Fig. 1, there are three major steps in extracting the

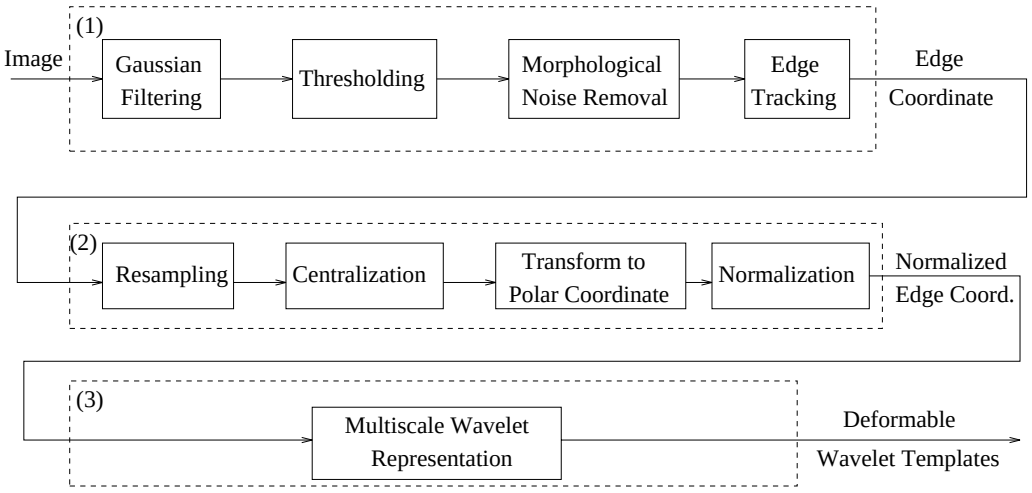


Figure 1: Block diagram of DWT extraction from an image.

DWT feature from a given image. First, in section 2, we describe the preprocessing of the image and the extraction of the target shape feature from a noisy image. Then, in section 3, we describe the normalization of the shape feature so that it can be shift-, scale-, and rotation- invariant. In section 4, the multiscale wavelet decomposition of the shape feature is described. We address the issue of multiscale target matching in Section 5. And finally, the experimental results are given in Section 6.

2 SHAPE EXTRACTION VIA EDGE DETECTION AND LINKING

In this section, we describe the process of extracting the shape information from a noisy image. The process can be further decomposed into three consecutive components: (i) morphological noise removal, (ii) edge detection and (iii) edge linking.

For (i), We convert the Gaussian-filtered gray-level image to a two-tone (black and white) image with a certain threshold T and apply a morphological closing filter in order to remove noise

$$\mathbf{B} = \mathbf{B} \bullet K = (\mathbf{B} \oplus K) \ominus K,$$

where $\oplus$ and $\ominus$ are the morphological dilation operator and the morphological erosion operator defined, respectively, by

$$\mathbf{B} \oplus K = \bigcup_{k \in K} (k + \mathbf{B}) \qquad \mathbf{B} \ominus K = \bigcap_{k \in K} (k + \mathbf{B}),$$

where K is the following 3×3 binary mask

$$K = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

For (ii), a change detector is used to identify the edge point in the binarized image:

$$E(x, y) = \begin{cases} 1 & \text{Edge point} \\ 0 & \text{None edge point} \end{cases}$$

For (iii), we adopt a clockwise edge linking algorithm to ensure that a closed contour is extracted from the image. The algorithm can be described as follows.

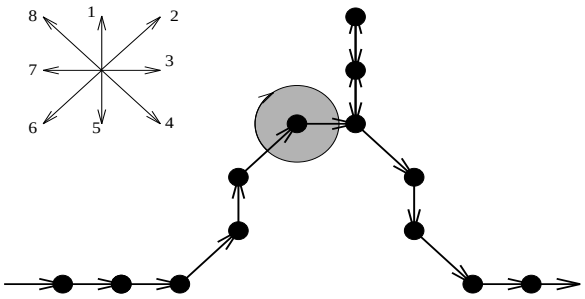
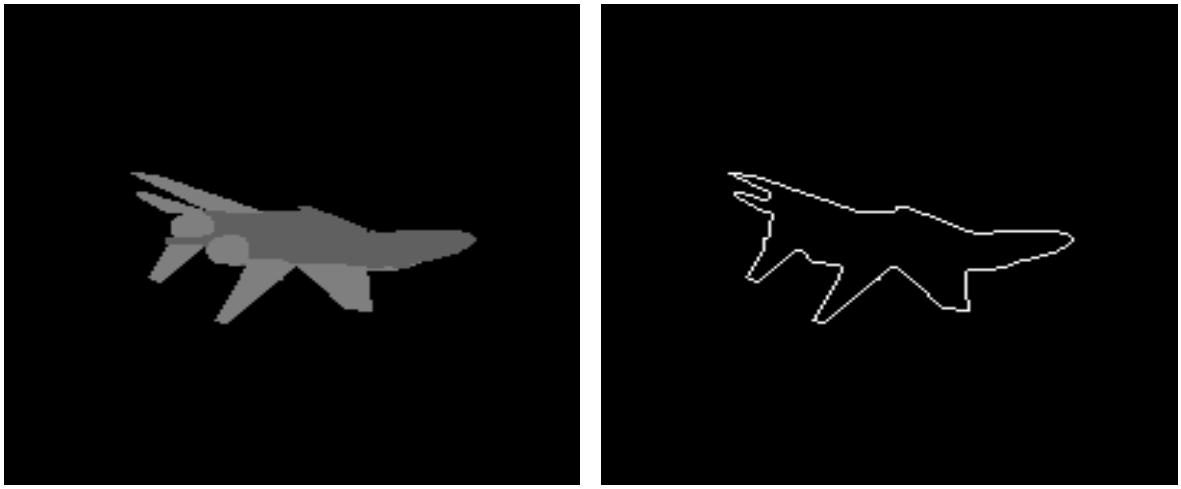


Figure 2: Clockwise search for edge linking.



(a) (b)

Figure 3: (a) Original image and (b) results obtained by edge detection and linking.

1. Initialization.
   We locate the first edge point, $E(x_s, y_s) = 1$, in the image by following the raster scan order. We set it as the starting edge point,
   $$x(0) = x_s, \qquad y(0) = y_s.$$
   Also, we set the initial direction to be $d = 3$.
2. Iteration.
   Starting from the current point and the current direction, the algorithm scans for the next edge point in a clockwise fashion. As shown in Fig. 2, the first edge point $E(x_c, y_c) = 1$ encountered in the scan is set as the next edge point,
   $$x(i) = x_c, \qquad y(i) = y_c.$$
   The direction d is also updated according to the search.
3. Termination.
   Step 2 is repeated until we return to the starting point $(x(0), y(0))$.

If there is more than one target in the image, the above procedure can be repeated. After one edge track is detected, we can remove all edge points associated with the track, and repeat the algorithm until no more edge points remain in the image. An example of edge linking is illustrated in Fig 2. If there is a non-closed branch in the shape curve, the algorithm will trace it back and forth, thus ensuring a closed

curve for each track. When multiple curves are extracted from the image, the longest (dominant) track is taken as the shape feature of the target.

A shape extraction example is shown in Fig. 3, where the original image is shown in (a) with its detected and linked edge shown in (b).

3 SHAPE NORMALIZATION

Edge coordinates are in general dependent on the displacement, scaling and rotation of the target. A normalization process is performed at the second step so that the normalized edge coordinates have 2D shift-, scale- and rotation-invariant properties. We begin with resampling the edge coordinates using a uniform number of points so that no matter how long a curve is, the number of sampling points is the same. This property is crucial for scale invariance. If there are m original edge points $(x_o(i), y_o(i))$, $i = 0, \cdots, m-1$, we can calculate the incremental curve length as:

$$\begin{aligned} l(0) &= 0, \\ l(i) &= l(i-1) + \sqrt{[x_o(i) - x_o(i-1)]^2 + [y_o(i) - y_o(i-1)]^2}, \qquad i = 1, \cdots, m. \end{aligned}$$

The total length of the curve is denoted by $L = l(m)$. We resample the curve uniformly at n points, i.e.

$$\begin{aligned} x_n(j) &= x_o(s)(1-f) + x_o(s+1)f, \\ y_n(j) &= x_o(s)(1-f) + x_o(s+1)f, \qquad i = 1, \cdots, m \end{aligned}$$

with

$$s = \lfloor L \times j/n \rfloor \qquad , f = \frac{L \times j/n - l(s)}{l(s+1) - l(s)}.$$

To normalize the representation with respect to displacement, we shift the curve so that the center of the curve is located at the origin. This is achieved by calculating the new coordinates $(x_n(j), y_n(j))$, using

$$x_n(j) = x(j) - \bar{x}, \quad y_n(j) = y(j) - \bar{y},$$

where $(\bar{x}, \bar{y})$ is the mean of the edge coordinates. To normalize the representation with respect to scaling and 2-D rotation, we transform edge points from Cartesian coordinates to polar coordinates via

$$R(j) = \sqrt{x_n(j)^2 + y_n(j)^2}, \quad \phi(j) = \arctan \frac{y_n(j)}{x_n(j)}.$$

We further normalize $R(j)$ so that it has the unit radius. Since all targets with different sizes now have the same radius, they are scale-invariant. Finally, for rotational invariance, we rotate the coefficients so that the center of mass $\sum i\tilde{R}_k(i)$ for radius R is minimized,

$$\begin{aligned} \tilde{R}_k(i) &= \begin{cases} R(i+k), & 0 \le i < N-k \\ R(i-N+k), & N-k \le i < N \end{cases} \\ \tilde{\phi}_k(i) &= \begin{cases} \phi(i+k), & 0 \le i < N-k \\ \phi(i-N+k), & N-k \le i < N \end{cases} \\ k &= \arg\min_k \sum_{i=1}^{N} i\tilde{R}_k(i). \end{aligned}$$

4 SHAPE REPRESENTATION USING A WAVELET DESCRIPTOR

In the third step, we transform the normalized coordinates to the wavelet representation. Wavelet theory[3,2] has primarily been developed for the representation of signals in the space $L^2(R)$. A single scale discrete wavelet transform of a sequence $f[n]$ can be viewed as passing the signal through a quadrature mirror filter (QMF) consisting of a low- and high-pass filter denoted, respectively, by $h[k]$ and $g[k]$, where $g[k] = (-1)^k h[1-k]$. The forward transform can be written as

$$\begin{aligned} c_k &= \sqrt{2}\textstyle\sum_n h[n-2k]f[n], \\ d_k &= \sqrt{2}\textstyle\sum_n g[n-2k]f[n], \end{aligned}$$

while the inverse transform takes the form

$$f[n] = \sqrt{2}\left(\sum_k h[n-2k]c_k + \sum_k g[n-2k]b_k\right).$$

Additional constraints can be imposed on $h[k]$ and $g[k]$ so that the resulting wavelet representation has several nice properties [7]. While many wavelet transforms have been proposed in the literature, we find that the biorthogonal 9-7 tap spline filter [1] gives the best result.

We denote a clockwise-oriented closed curve (obtained in section 3) with parametric coordinates $R(t)$ and $\phi(t)$ by

$$\alpha(t) = \begin{bmatrix} R(t) \\ \phi(t) \end{bmatrix}, \quad t(l) = l/L, \quad 0 \le l \le L,$$

where the parameter t corresponds to the normalized arc length, l is the arc length along the curve from starting point t_0, and L is the total arc length. By applying a periodic wavelet transform to the parameterized coordinates, we obtain

$$\begin{bmatrix} R(t) \\ \phi(t) \end{bmatrix} = \begin{bmatrix} R_a^M(t) \\ \phi_a^M(t) \end{bmatrix} + \sum_{m=M_0}^{M} \begin{bmatrix} R_d^m(t) \\ \phi_d^m(t) \end{bmatrix},$$

where

$$R_a^M(t) = \sum_n R_n^M \tilde{\psi}_n^M(t), \quad \phi_a^M(t) = \sum_n \phi_n^M \tilde{\psi}_n^M(t). \tag{1}$$

these are called the approximation coefficients at scale M and

$$R_d^m(t) = \sum_n R_n^m \tilde{\psi}_n^m(t), \quad \phi_d^m(t) = \sum_n \phi_n^m \tilde{\psi}_n^m(t). \tag{2}$$

are called the detailed signals at scale m with $m = M_0$ for the finest scale and $m = M$ for the coarsest scale. We can now use the wavelet coefficients R_n^M, ϕ_n^M, R_n^m and ϕ_n^m given in (1) and (2) as the planar curve descriptor. By using only a subset of wavelet coefficients consisting of primarily coarser scale components, we can achieve different multiresolution representations of the shape.

5 HIERARCHICAL DATABASE SEARCH AND MATCHING

The multiscale target features are matched in a hierarchical way. The search strategy can be depicted with Fig. 4. Matching starts at the coarsest scale 1 and moves on to finer scales. For the coarsest scale 1, the target feature is matched with all possible entries in the database. Since the number of coarsest scale features are usually quite small, although the items to be matched are enormous, the computational complexity is limited. For each item, the matching operation attempts to make one of the three decisions:

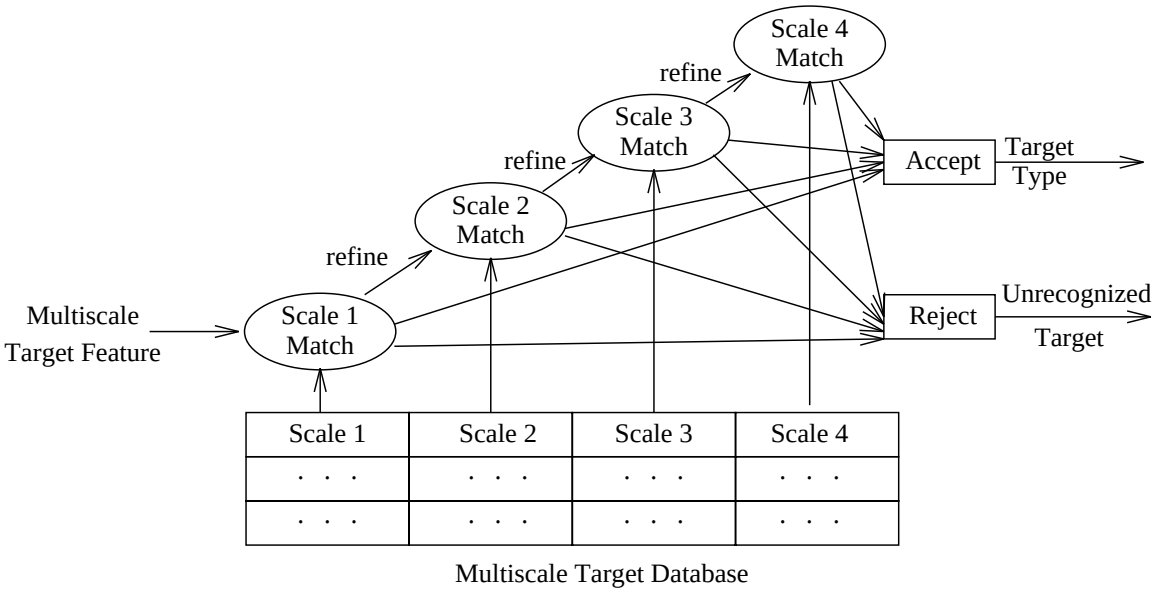


Figure 4: Multiscale hierarchical search

a) accepted the item as a valid target identification
b) rejected the item as a non match
c) marked the item as to be refined for finer scale matching

After searching all the available entries, if a valid target identification is found, the matching process is terminated and the target is unambiguously identified and reported. If all entries are rejected as non match, the matching process is also terminated and the target is marked as an un-identified target or false alarm. Otherwise, the matching process moves on to scale 2. We match the scale 2 differential features with the entries marked as to be refined in the scale 1 match. Again decisions are made and searching is conducted. If the target is still marked as to be refined, the matching process continues moving on to a finer scale. The program is terminated when the target is completely identified or completely rejected. The depth of matching scale is determined by the image noise level and the target size. If the matching requires refinement beyond the finest scale, the target is regarded as un-identified.

6 EXPERIMENTAL RESULTS

Extensive experiments has been conducted to demonstrate the effectiveness of the wavelet deformable templates(WDT). We first compare the wavelet descriptor with the Fourier descriptor. We show a deformed Su27 plane by the wavelet descriptor and the Fourier descriptor in Fig. 5. Both are deformed by the change of only one coefficient, the corresponding results are shown with dash curves. For the Fourier descriptor, the change affects all parts of the plane, there are no local features. In contrast, using the wavelet descriptor, there is only local change around the left tail wing of the plane.

In the following experiments, we demonstrate the normalization of the shape feature using one example of the same Su27 plane. Fig. 6 shows the original edge coordinates of the plane containing 1749 points. The curve is resampled uniformly at 1024 points as shown in Fig. 7. We then centralize the curve to achieve shift invariance and transform it to polar coordinates as given in Fig. 8. Finally, we normalize the radius and rotate the coefficients to minimize the center of mass for radius R. The resulting 2D shift-, scale- and rotation- invariant shape template is shown in Fig. 9. A set of rotationally isomorphic planes are depicted

in Fig. 10. Note that the normalized target shape template is not invariant for 3D viewing angles (see Fig. 11).

Finally, we demonstrate the multiscale wavelet representation using again the same Su27 fighter in Figs. 12 and 13. The first column of the figures gives the multiresolution radius coefficients, the second column describes the polar angle coefficients, and the third column depicts the plane at the corresponding scale. The global feature of the target is represented at lower resolutions, while the detailed features are added at higher resolutions. One can adjust the scale for target representation and matching according to image resolutions, target size and noise levels. That is, when the target is far away or the noise level is high, the detailed feature become unstable, so to avoid error existing in high scale details, only the low scale feature should be used in matching. On the other hand, when the target is near or the noise level is low, we may use detail features to enhance matching accuracy.

7 CONCLUSIONS AND EXTENSIONS

In this paper, we demonstrate the feasibility of using wavelet deformable template(WDT) for automatic target recognition. In contrast with the conventional approaches, our technology is based on the synergy of multi-resolution image processing. A hierarchical WDT descriptor is developed of which the coarsest scale components carry global approximation information while the finer scale components contain local, detailed information. The techniques of target shape extraction from the noisy image and target shape normalization for 2D shift-, scale- and rotation- invariance are also described in the paper. Reliable multiscale image denoising and enhancing techniques based on wavelet decomposition are under our current investigation. We also plan to investigate the use of WDT for handwritten cursive character recognition in the recent future.

8 REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, pp. 205–230, Apr. 1992.

[2] C. K. Chui, *An introduction to wavelets*, Academic press, 1991.

[3] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

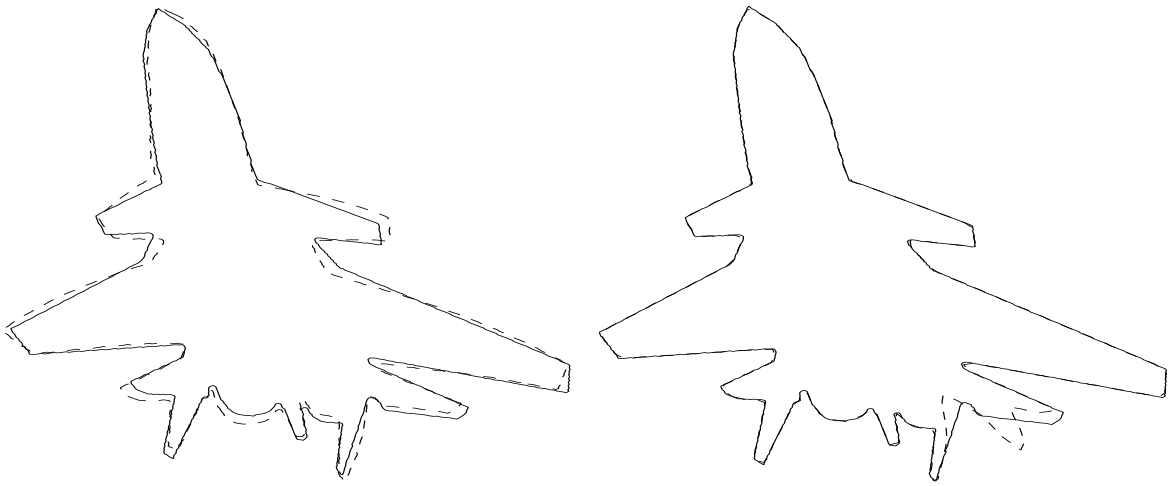
[4] R. Jain, R. Kasturi, and B. G. Schunck, *Machine vision*, New York : McGraw-Hill, 1995.

[5] H. Kauppinen, T. Seppanen, and M. Pietikainen, "An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification," *IEEE Trans. on pattern analysis and machine intelligence*, Vol. 17, pp. 201–207, Feb. 1995.

[6] S. Mahmoud, "Arabic character-recognition using fourier descriptors and character contour encoding," *Pattern Recognition*, Vol. 27, pp. 815–824, June 1994.

[7] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Review*, pp. 614–627, Dec. 1989.

[8] S. Wang, P. Chen, and W. Lin, "Invariant pattern-recognition by moment fourier descriptor," *Pattern Recognition*, Vol. 27, pp. 1735–1742, Dec. 1994.



(a) (b)

Figure 5: Deformed shape template obtained by changing one coefficient in (a) the fourier descriptor and (b) the wavelet descriptor.

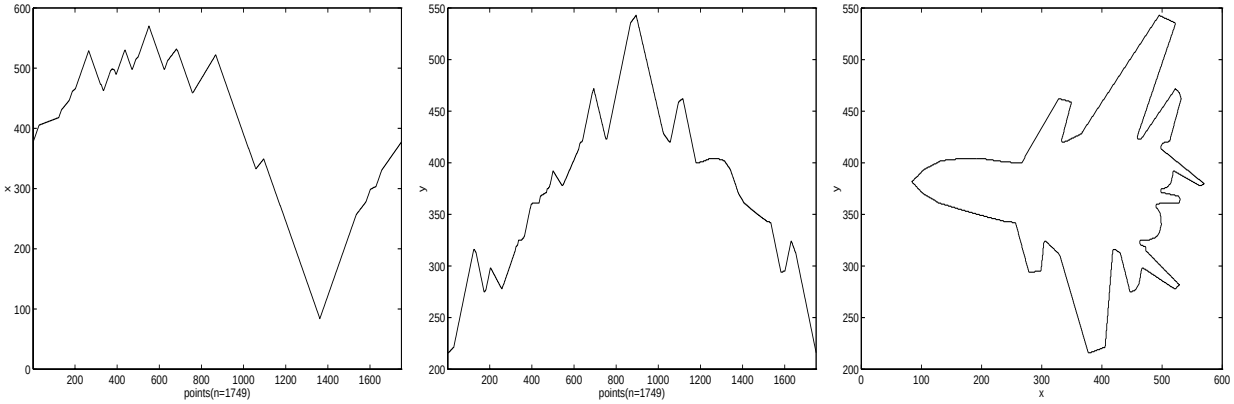


Figure 6: Original edge coordinates after linking with 1749 points.

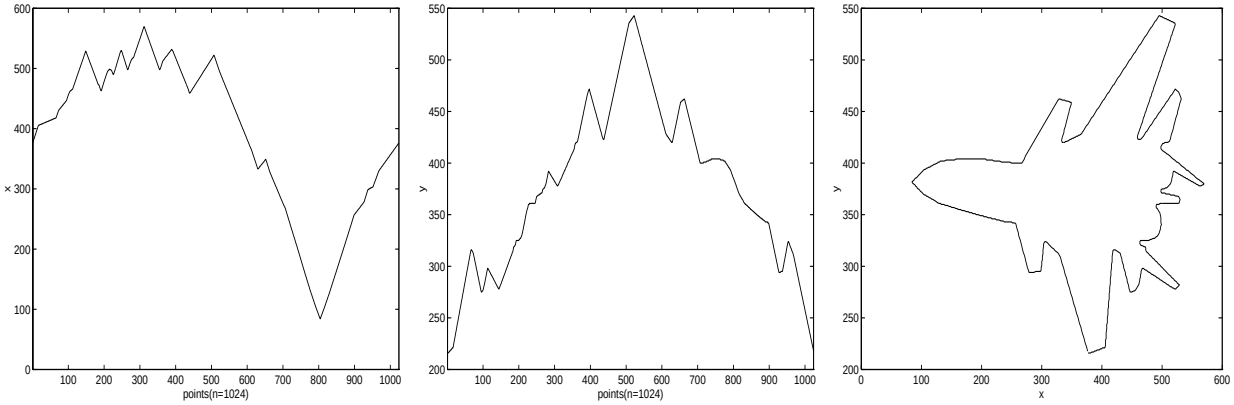


Figure 7: Resampled edge coordinates with 1024 points.

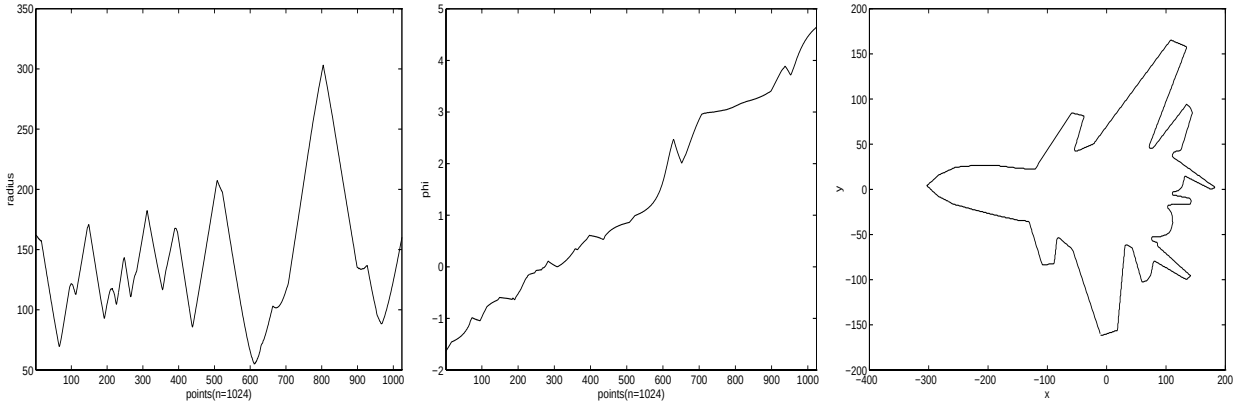


Figure 8: Polar edge coordinates which are shift invariant.

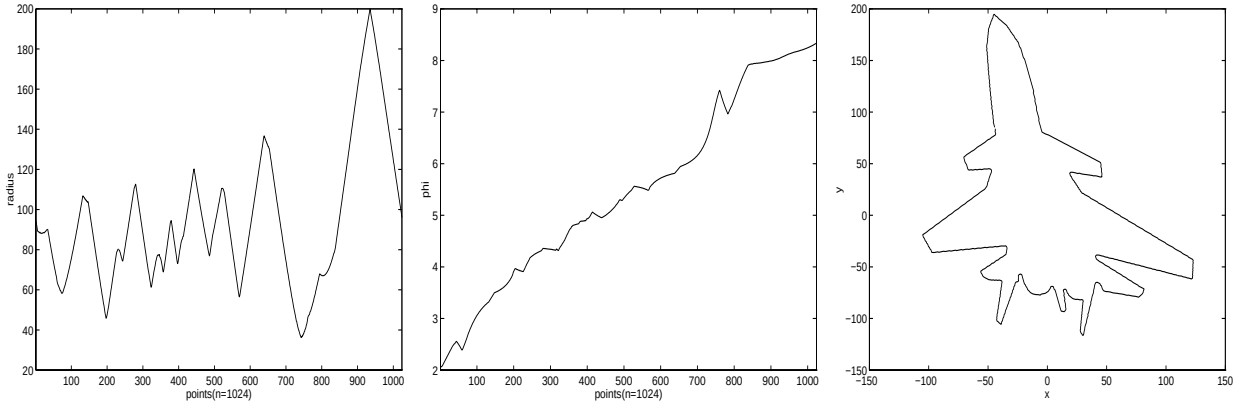


Figure 9: Normalized polar edge coordinates which are shift, scale and rotational invariant.

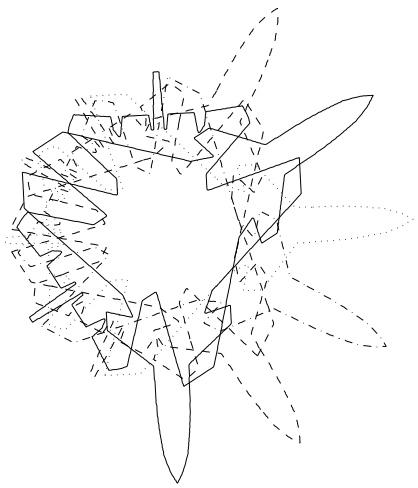


Figure 10: Rotational isomorphic templates.

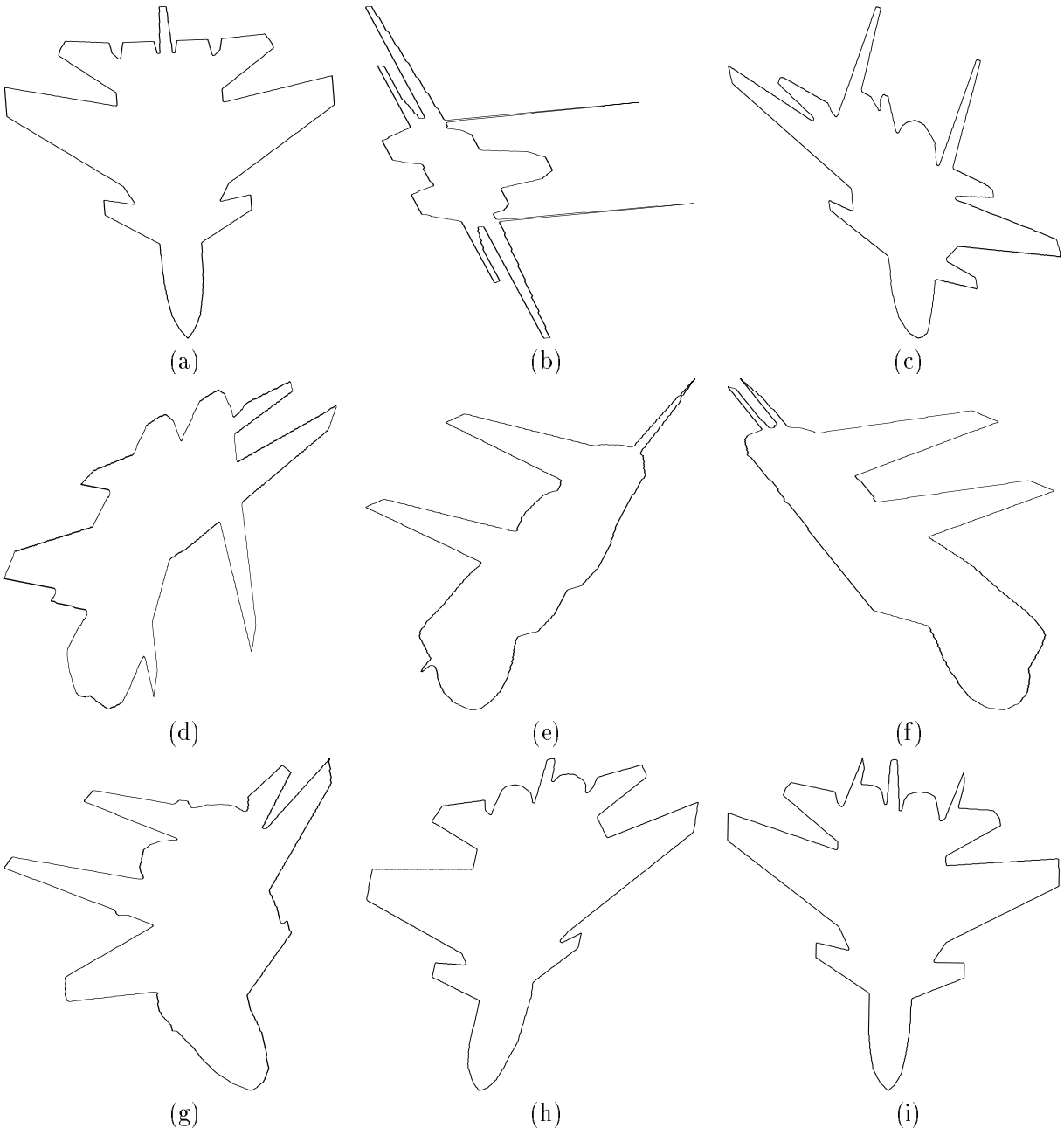
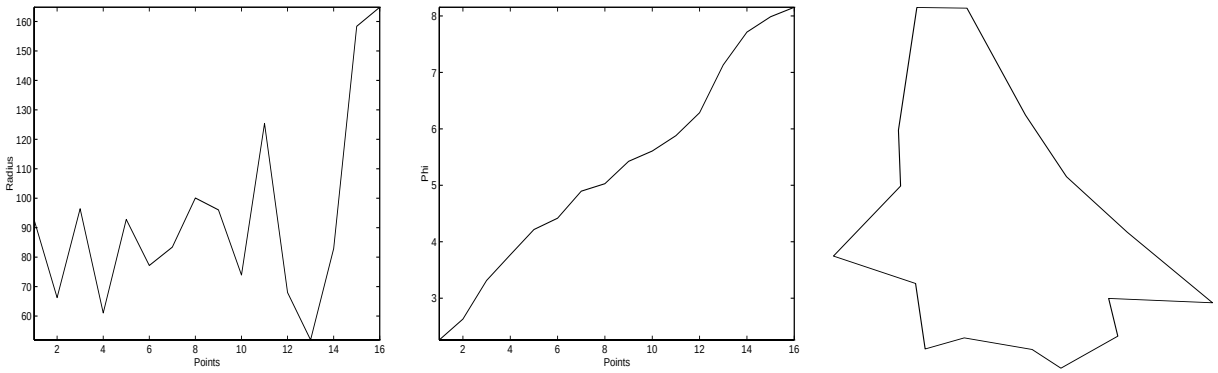
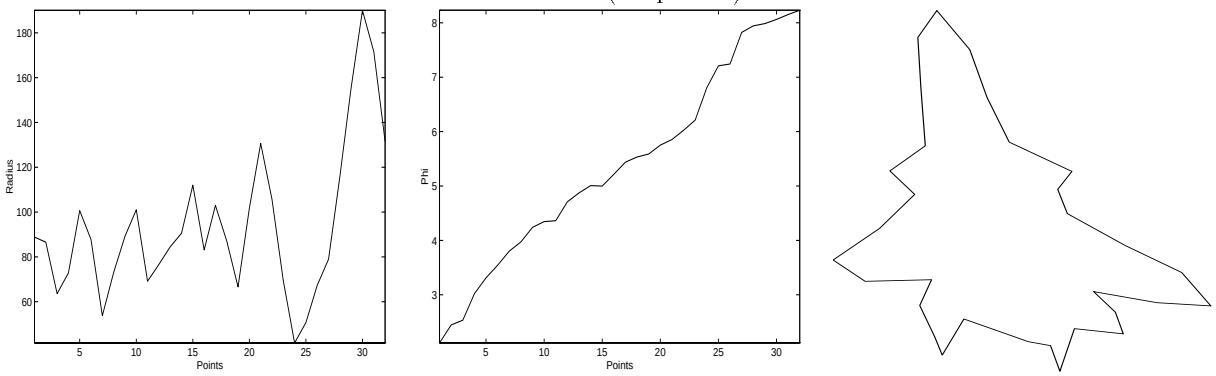


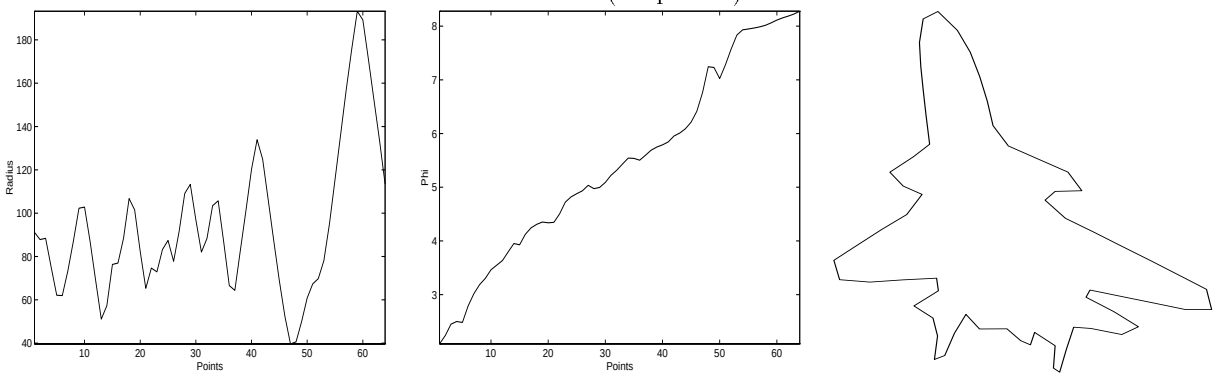
Figure 11: The same Su27 viewed from different 3D angles



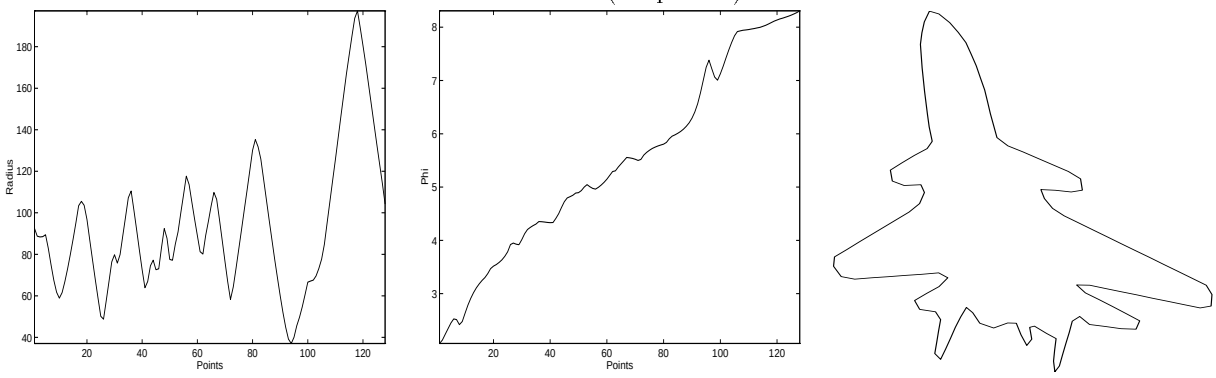
Scale 6 (16 points)



Scale 5 (32 points)

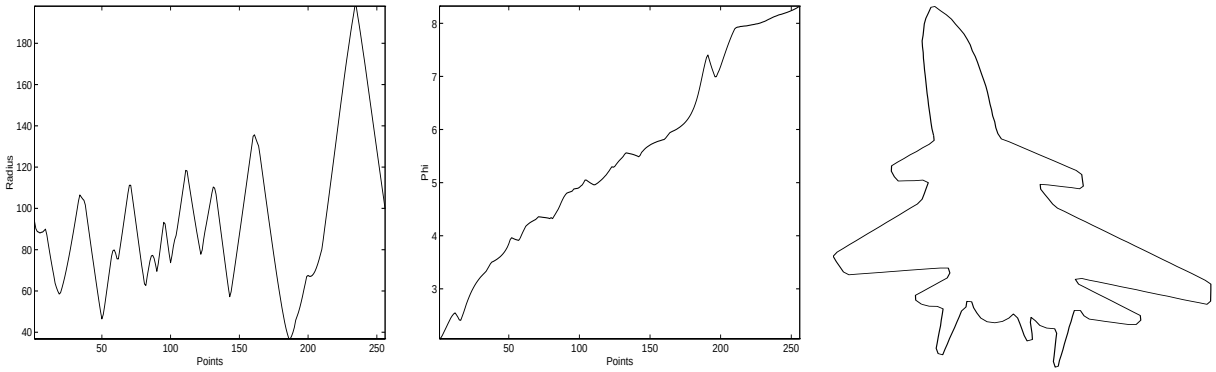


Scale 4 (64 points)

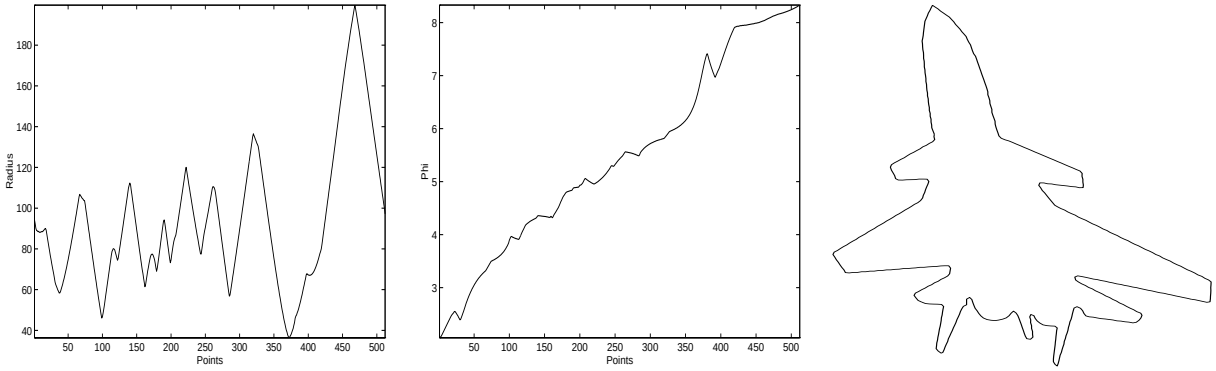


Scale 3 (128 points)

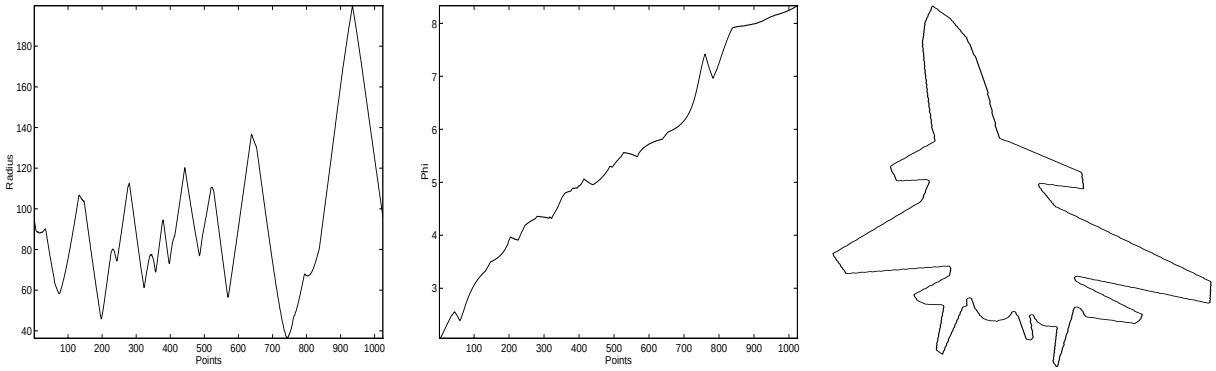
Figure 12: Multiscale target representation by the stochastic deformable wavelet template (DWT). The numbers in parentheses denote the number of points used in the representation.



Scale 2 (256 points)



Scale 1 (512 points)



Scale 0 (1024 points)

Figure 13: Multiscale target representation by the DWT (continued).