PERGAMON

Pattern Recognition 32 (1999) 753–771

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

Silhouette recognition using high-resolution pursuit

Seema Jaggi[a,*], W. Clem Karl[b], Stéphane G. Mallat[c,d], Alan S. Willsky[a]

[a] *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA*
[b] *Department of Electrical and Computer Engineering and the Department of Biomedical Engineering, Boston University, Boston, MA, USA*
[c] *Ecole Polytechnique, Paris, France*
[d] *Courant Institute, New York, USA*

Received 9 September 1997

Abstract

This paper introduces a simple new approach to object recognition from silhouettes. This new approach utilizes features extracted using an adaptive approximation technique called high-resolution pursuit (HRP). In this work, a comparatively small set of HRP features and a simple recognition scheme are used. We demonstrate the strengths of the HRP-based recognition scheme by discriminating among 17 military aircraft. The HRP-based algorithm matches the performance of a widely studied method based on Fourier descriptors in the presence of boundary, scale and orientation variations, and surpasses the performance of the Fourier descriptor-based algorithm in the presence of occlusion and localized silhouette variations.

Keywords: High-resolution pursuit; Silhouette recognition; Object recognition; Centroidal distance profile; Adaptive approximation

1. Introduction

Object recognition from silhouettes is a problem with substantial literature and history in computer vision [1–7]. The ideal recognition system is one that is robust to orientation variations, scale variations, and boundary perturbations as well as localized distortions. Localized distortions, such as the presence or absence of underwing stores on an airplane, have been especially difficult for systems based on global features (e.g. Fourier descriptors) [8]. In this paper, we develop an alternative approach to object recognition from silhouettes. In contrast to current approaches in model-based silhouette recognition, our approach utilizes features extracted using an adaptive approximation technique called high-resolution pursuit (HRP) [9,10]. The set of HRP features are local, and therefore, robust to localized silhouette variations and occlusion, yet the recognition algorithm based on HRP features is remarkably simple. Indeed, in addition to contributing a highly competitive yet simple algorithm for this particular recognition application, an equally important objective of this paper is the demonstration of the potential of a new approach to feature extraction that has a number of very attractive properties. As we will show, the set of HRP features have a clear physical interpretation in terms of the objects being recognized. In other words, each of the elements in the HRP representation are directly related to the geometric (e.g. size and location of subparts) characteristics of the object. The HRP-based features are parsimonious in that a

*Corresponding author.

0031-3203/99/$– See front matter
PII: S0031-3203(98)00112-5

comparatively small set of such features capture all of the information needed for recognition; that is, the feature extraction procedure effectively focuses information. In addition, the statistical variability of the features for each object class is modest, allowing us to easily characterize probability density functions which describe the variability of the features. Finally, the extraction of HRP-based features is a computationally simple task.

In this paper, we describe and evaluate this HRP-based feature extraction and object recognition scheme. Our proposed approach is to use the HRP algorithm to extract a set of features which quantify the size and location of the subparts of the object. Since a number of researchers [1,2,3,10] have based their recognition schemes on a 1-D representation of the silhouette, we do as well. In particular, we use HRP to extract features from a 1-D representation of the silhouette, namely, the centroidal distance profile (CDP) [2,4,10]. The HRP algorithm is used to construct a feature vector of low dimension, and recognition is performed using this feature vector and an M-ary hypothesis testing scheme.

We demonstrate the HRP-based algorithm by discriminating among the seventeen military airplanes shown in Fig. 1. Similar data sets have been used in a number of other articles [6,7,10–12]. These types of silhouettes may be generated from gray scale images, such as the one shown in Fig. 2, using any one of a number of edge detectors [13–16], segmentation algorithms [17], or active contour models [18]. This data set is a fairly challenging one since it contains some planes which are very similar. For example, Planes # 2, 9, and 10 are similar to one another and Planes #4, 15 and 16 are similar to one another. Planes # 7, 12 and 17 are considered swept wing aircraft. Our experimental results show that this approach using a low-dimensional HRP-based feature vector is robust to boundary perturbations, scale and orientation variations. Further, the local nature of the HRP-based features allows us to deal easily with moderate levels of occlusion and localized silhouette variation without a significant change in the algorithm.

For comparison, we present recognition results on the airplane dataset using one of the more widely-studied methods based on Fourier descriptors (FD) [1,10]. The notable benefits of the FD-based algorithm include its small computational burden, its robustness to orientation and scale variation, and its rejection of boundary noise. These are benefits that are matched in the HRP-based algorithm. However, the FD-based algorithm uses an order of magnitude more features than the HRP-based algorithm to achieve the same level of performance. Moreover, since the features in the FD-based algorithm are nonlocal, the algorithm performs poorly in the presence of localized silhouette variations and occlusion. In fact, even moderate levels of occlusion require a significant change in the algorithm [6], and the performance of the modified algorithm to date is at a low level of granularity. That is, this modified algorithm is only able to distinguish some broad classes of objects rather than individual shapes. In contrast, the HRP-based algorithm significantly surpasses the performance of the FD-based algorithm in the presence of localized silhouette variations and occlusion.

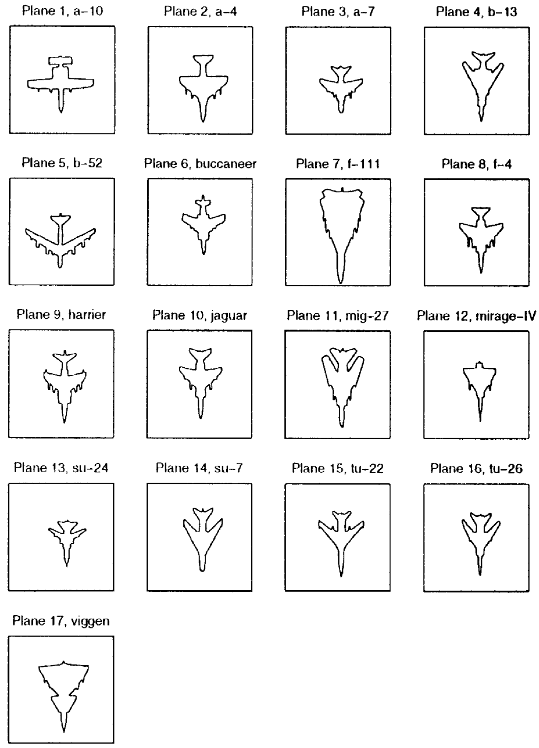


Fig. 1. Silhouettes of the military airplanes which make up the data set.

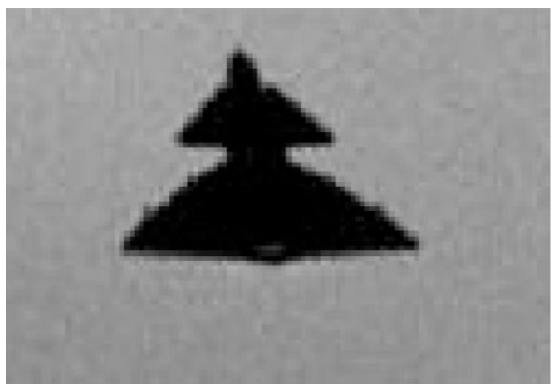


Fig. 2. Grayscale image of a viggen aircraft.

The outline of this paper is as follows. Section 2 describes the CDP, the FD-based recognition algorithm, and the general HRP algorithm. Section 3 details the HRP-based recognition algorithm and Section 4 demonstrates the strengths of this HRP-based recognition algorithm using both real and simulated data. Finally, Section 5 presents some conclusions and areas for future work.

2. Preliminaries

In this section, we begin by describing a 1-D representation of the silhouette, some models used to create noisy silhouettes, and a classification technique based on Fourier descriptors. Then, we describe the HRP algorithm as a general adaptive approximation technique.

2.1. Centroidal distance profile

For this work, we will use the centroidal distance profile (CDP) [2,10] as a 1-D representation of the airplane silhouettes. Specifically, if $(x(n), y(n))$ are an ordered sequence of boundary points which are equidistant along the boundary, then the CDP $f(n)$ is given by

$$f(n) = \sqrt{(x(n) - X_c)^2 + (y(n) - Y_c)^2}, \tag{1}$$

where (X_c, Y_c) is the centroid of the object [2]. Fig. 3 shows the CDP, f, for Plane #1. This CDP starts at the tip of the nose and proceeds in a counterclockwise direction. From this figure, we note that the physical features of the plane are clearly identifiable in the CDP. The four peaks in the profile correspond to the nose, wing #1, tail, and wing #2, respectively.

One drawback of the CDP is that it essentially discards angular information. Clearly, it is not possible to reconstruct the original boundary from the CDP alone. The additional information needed to construct the original boundary is the angular profile, $\theta(n)$. For boundary points $(x(n), y(n))$, the angular profile $\theta(n)$ is given by

$$\theta(n) = \arctan\left(\frac{y(n) - Y_c}{x(n) - X_c}\right). \tag{2}$$

The CDP and the angular profile are simply the representation of the boundary in polar coordinates centered at the centroid of the silhouette. This angular information will be useful in Section 3.

2.2. Silhouette Variations

For an algorithm for recognizing objects from silhouettes to be of practical importance, the algorithm must degrade gracefully in the presence of scale variation, orientation variation, boundary perturbations, and variation due to occlusion. In this section, we illustrate some of these possible variations, their effects on the CDP, and describe models used in the literature to create these silhouette variations.

The simplest variations are those due to a change in scale, which results in a change in amplitude of the CDP,

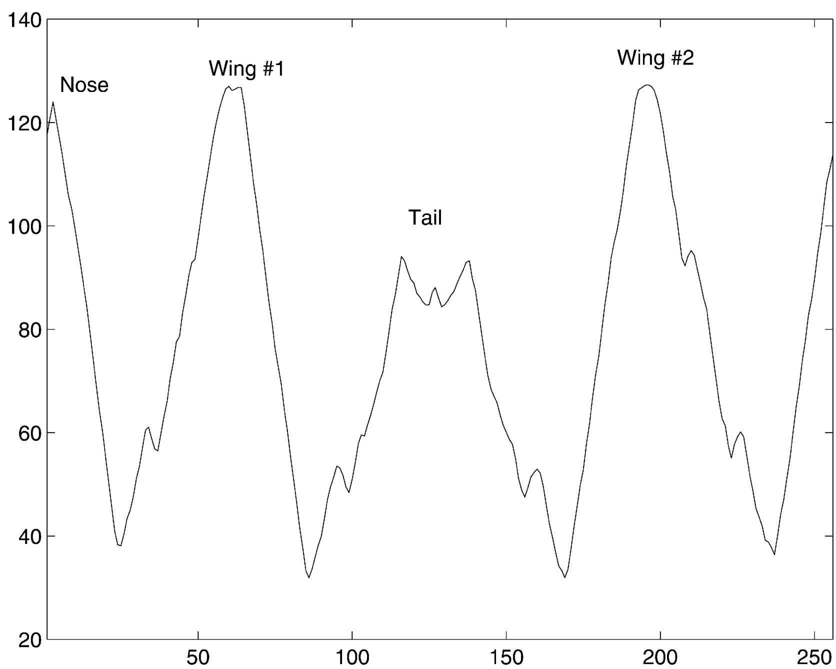


Fig. 3. The CDP, f, for Plane #1 calculated at equally spaced points along the boundary.

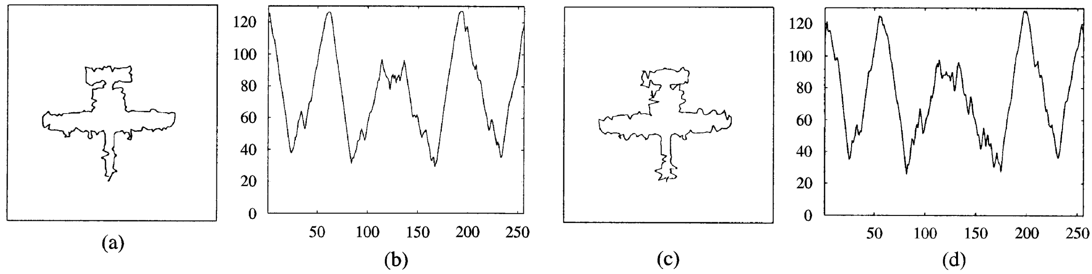


Fig. 4. Generating Gaussian perturbations perpendicular to the object contour. (a) Noisy boundary with $p = 40\%$ and $s = 0.9$. (b) The corresponding noisy CDP. (c) Noisy boundary with $p = 40\%$ and $s = 2.1$. (d) The corresponding noisy CDP.

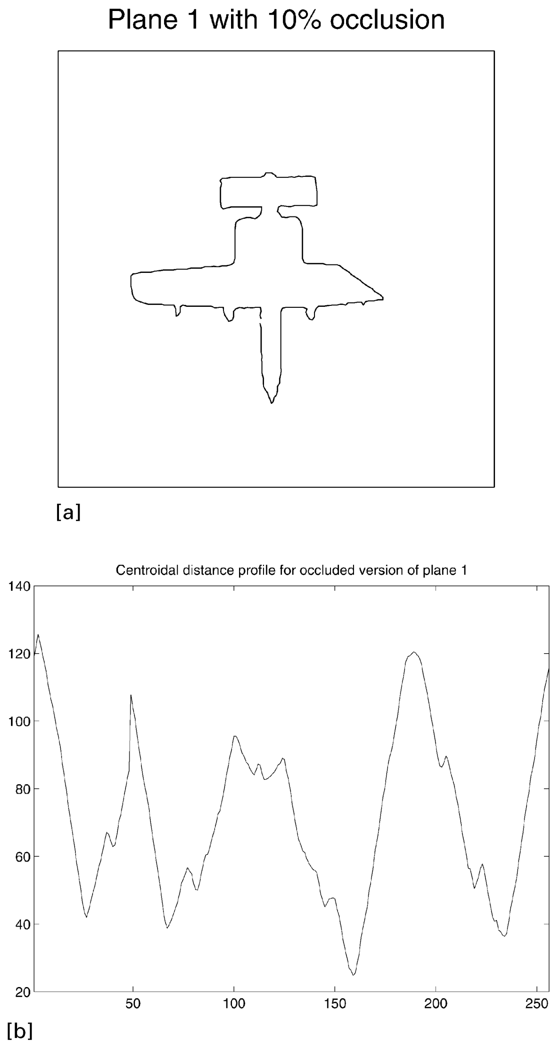


Fig. 5. Plane #1 with occlusion and the corresponding CDP.

and a change in orientation in the viewing plane, which results in a circular shift of the CDP. Imaging conditions such as lighting, reflectance, and haze may cause less predictable perturbations in the boundary curve itself. We will model these boundary perturbations using the technique adopted by several other researchers [2,10,19] where p percent of the boundary points are perturbed by Gaussian noise which is perpendicular to the boundary. That is, the nth boundary point goes from $(x(n), y(n))$ to $(x_c(n), y_c(n))$ where

$$x_c(n) = x(n) + dr\cos(\xi(n)), \tag{3}$$

$$y_c(n) = y(n) + dr\sin(\xi(n)), \tag{4}$$

where d is the distance between boundary points n and $n + 1$, r is a random variable chosen from $N(0, s^2)$, and ξ is the angle between the x-axis and the direction normal to the boundary at point n. Fig. 4 shows a noisy version of Plane #1 with $p = 40$ and $s = 0.9$ and the corresponding CDP.

Occlusion may be caused by changes in lighting, reflectance, haze, or the presence of other objects in the image. In the article by Liu [12], occluded boundaries are created by replacing a random, consecutive set of q percent of the boundary points in a silhouette with a straight line. Fig. 5 shows an occluded version of Plane # 1 with $q = 10$ and the corresponding CDP. Occlusion causes two types of distortion in the CDP. First, since a portion of the boundary is missing more boundary points will be devoted to the remaining airplane features. As a result, the scale of the remaining features has changed. For example, comparing Figs. 3 and 5b, we note that as a result of the occlusion the second wing in Fig. 5b appears wider than the second wing in Fig. 3. Second, the occlusion also causes a shift in the centroid which leads to warping in the CDP.

2.3. Fourier descriptors

One well-studied class of techniques used to recognize silhouettes is based on Fourier descriptors [1,5,10]. The benefits of these techniques is their small computational burden, robustness to scale and orientation variation, and relative robustness to boundary perturbations. On the other hand, these techniques deteriorate rapidly in the presence of localized changes in the silhouette and occlusion.

In an article by Kauppinen [10], the author constructs a feature vector based on the normalized magnitude of the Fourier transform of the CDP of a silhouette. If the CDP of a silhouette has P samples and is given by f, and the discrete Fourier transform of f is given by F, then the feature vector $\mathbf{y}$ is given by

$$\mathbf{y} = \left[\frac{|F_1|}{|F_0|} \cdots \frac{|F_{P/2}|}{|F_0|}\right], \tag{5}$$

where F_i denotes the ith component of F. The classification of silhouettes based on their Fourier descriptors is done using a K nearest neighbor procedure [20]. In this work, the number of points in the CDP will be $P = 256$, and the corresponding FD feature vector will consist of 128 elements. In contrast, the HRP-based recognition scheme will use an order of magnitude fewer features.

2.4. HRP algorithm

Recently, adaptive approximation techniques [9,21–23] have become popular for obtaining parsimonious representations of large classes of signals. In these adaptive approximation techniques, the goal is to find a representation of a function f as a weighted sum of elements from an overcomplete dictionary. That is, f is represented as

$$f = \sum_{\gamma \in \Gamma} \lambda_\gamma g_\gamma, \tag{6}$$

where the set $\{g_\gamma | \gamma \in \Gamma\}$ is a redundant dictionary spanning the space of possible functions. Since many possible decompositions of f exist over this redundant dictionary, the "optimal" decomposition is often application dependent. Several techniques have been developed to find the "optimal" decomposition of f, including matching pursuit [22], basis pursuit [23], and high-resolution pursuit [8,9]. Matching pursuit chooses one element at a time out of the dictionary such that the L^2 norm of the mth residual is minimized at each step, where the mth residual is defined to be the difference between f and the first $m-1$ terms in the decomposition and is denoted $R^m f$ where

$$R^m f = f - \sum_{n=0}^{m-1} \lambda_n g_{\gamma_n}. \tag{7}$$

As a result, the matching pursuit algorithm is a fast but greedy algorithm which often sacrifices local fit (and therefore physically meaningful features) for global fit. Further, the matching pursuit algorithm may introduce artifacts, namely significant new features at finer scales because of its greedy approach. The basis pursuit algorithm chooses all elements in the decomposition at once by minimizing the ℓ^1 norm of the coefficients λ_γ in Eq. (6). Decompositions obtained using basis pursuit yield physically meaningful features but at an extreme computational price. Finally, high-resolution pursuit (HRP) chooses one element at a time out of the dictionary by optimizing an L^2-like metric subject to a set of constraints which are designed to balance local fit with global fit. In contrast to the other two techniques, HRP is a fast algorithm, but is still able to extract physically meaningful features; therefore, HRP is the adaptive approximation technique that will be used in this work.

The dictionary used with HRP is user specified and should reflect prior knowledge about the types of signals to be decomposed. HRP may be implemented with many different dictionaries, such as wavelets and wavelet packets. One dictionary that is well matched to the CDPs and therefore will be particularly useful for this work is the cubic b-spline dictionary. A cubic b-spline is a box spline convolved with itself three times. Fig. 6 shows a box spline $b(x)$ and the resulting cubic b-spline $g_u(x)$. A cubic b-spline at scale j and translation t will be denoted $g_{j,t}(x)$ and is given by

$$g_{j,t}(x) = \sqrt{2^j} g(2^j(x-t)), \tag{8}$$

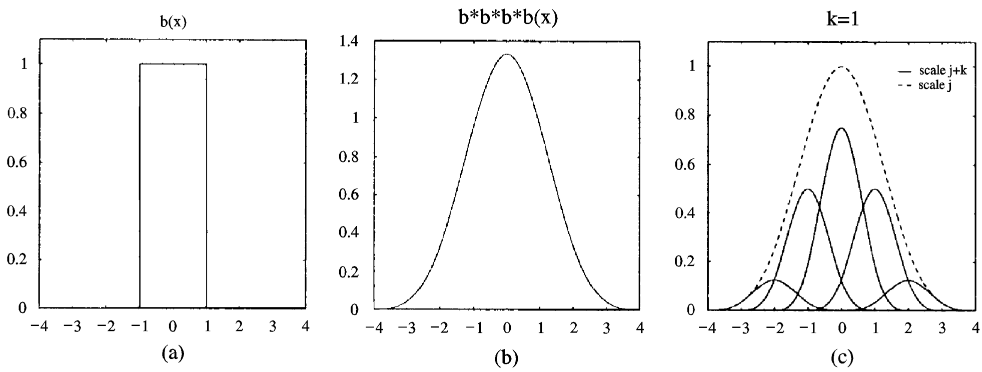


Fig. 6. (a) A box spline, $b(x)$. (b) A cubic box spline, $g_u(x) = b*b*b*b(x)$. (c) Any cubic b-spline is the weighted sum of finer scales cubic b-splines at scale $j+k$ as illustrated for $k = 1$.

where $g(x) = g_u(x)/\|g_u(x)\|$. The parameter j controls the size of $g_{j,t}$ such that increasing j corresponds to a finer scale, more localized function. In this work, we will often use γ to denote the pair (j, t) so that $g_\gamma(x) = g_{j,t}(x)$. The cubic b-spline dictionary then contains a set of functions $g_{j,t}(x)$ over an appropriate range of scales and translations. Note that any cubic b-spline may be written as the sum of finer scale cubic b-splines which also happen to be dictionary elements. For example, $g_{j,t}$ may be written as the weighted sum of finer scale cubic b-splines which are all at the same scale, $j + k$, as is illustrated in Fig. 6c for $k = 1$.

The objective of HRP is to combine the speed of matching pursuit with the ability of basis pursuit to extract physically meaningful features. In particular, the HRP algorithm guards against the introduction of artifacts or spurious features that may be introduced by the matching pursuit algorithm. The basic idea behind HRP is to decompose the function as the sum of the most significant elements from the dictionary set. At each iteration, the most significant element is defined to be the one which maximizes the HRP similarity measure, which has been designed to emphasize local fit accuracy. To achieve the desired sensitivity to local fit, HRP requires that each dictionary element, g_γ, be associated with a user-specified set or subfamily, I_γ, which should be constructed of finer scale functions which somehow capture the local structure of g_γ. These subfamilies are an integral part of the HRP algorithm and we will shortly discuss them in more detail. The HRP similarity measure between f and a particular dictionary element g_γ will be denoted $S(f, g_\gamma)$. To be sensitive to local fit, $S(f, g_\gamma)$ should reflect not only the similarity between f and g_γ, but also the similarity between f and finer scale members of the subfamily I_γ which capture the important *local* structures of g_γ. One might expect that some combination of $\{\langle f, g_i\rangle_{g_i \in I_\gamma}\}$ would yield a similarity measure which is sensitive to local fit. Intuitively, the HRP similarity measure should be dominated by the worst local fit over the set of functions in I_γ. For example, the minimum of $\{\langle f, g_i\rangle_{g_i \in I_\gamma}\}$ would be dominated by worst local fit. In fact, the HRP similarity measure, which is given in Appendix A, is the minimum of $\{\langle f, g_i\rangle_{g_i \in I_\gamma}\}$ and therefore is dominated by worst local fit. Appendix A also outlines the complete HRP algorithm. Note the important role played by the subfamilies I_γ in the HRP algorithm. These subfamilies are used enforce local fit accuracy, and the scale of the functions in I_γ specifies the scale to which local fit accuracy will be enforced. That is, HRP will tolerate local mismatches as long as they are at a finer scale than the functions included in I_γ.

For the cubic b-spline dictionary, one logical choice for I_γ is the set of finer scale dictionary elements at scale $j + k$ (i.e. at a relative scale k levels finer than g_γ itself) which when properly weighted and summed yields g_γ. That is,

$$I_\gamma(k) = \left\{ g_{j+k,t_i} \,\middle|\, g_\gamma = \sum_{i=1}^{L} c_i g_{j+k,t_i} \right\} \cup \{g_\gamma\}. \tag{9}$$

Note that we have used the notation $I_\gamma(k)$ to indicate the dependence of this set on the relative scale parameter k. As k increases, the subfamily, $I_\gamma(k)$, includes finer scale, more localized functions. Correspondingly, as k increases, HRP will tend to enforce finer scale local fit.

The depth parameter, k, is a tuning parameter which is used to incorporate prior knowledge about the relative scales of important features and provide robustness in the HRP algorithm. The parameter k can be held constant or change at each step of the HRP algorithm. To illustrate the effect of k, consider the following decompositions of a noisy CDP of Plane #1, where the noise has been added as described in Section 2.2 with $p = 40\%$ and $s = 0.9$. Fig. 7a shows a noisy CDP for Plane #1 and the first four elements of the decomposition obtained using HRP with cubic b-splines dictionaries where the subfamilies used are as given in (9) with $k = 1$. For this value of k, the HRP algorithm enforces local fit on a fairly coarse scale and is therefore more tolerant of local structural mismatch. As a result, the HRP decomposition consists of coarse scale dictionary elements which are not well matched to the local structures of the CDP. In contrast, Fig. 7b shows the decomposition of the same noisy CDP using HRP with $k = 2$. For this value of k, the subfamilies I_γ consist of finer scale elements and therefore HRP is less tolerant of local mismatch. The HRP elements for $k = 2$ correspond to the physical features of the plane. That is, from left to right, the HRP elements correspond to the nose, the first wing, the tail, and the second wing. Qualitatively, we might say that the elements in Fig. 7b corresponding to the nose and the wings are a good fits to the CDP, but the element corresponding to the tail is a poor fit since it is not well matched to the profile over the entire tail. One can also imagine letting k vary as a function of HRP element number. Suppose we first use HRP with $k = 3$ to extract two elements, then use $k = 2$ to extract a third element, and then use $k = 1$ to extract a fourth element. The results of this decomposition are shown in Fig. 7c. This type of variation scheme for k yields a desirable description of the CDP since each element of the decomposition corresponds to a physical feature of the object and accurately describes that feature.

The HRP decomposition shown in Fig. 7c, where k changes as a function of HRP element number, illustrates some of the properties which make the HRP algorithm attractive for feature extraction. First, by using such a scheme to choose and vary k, the elements in the HRP decomposition have a clear and accurate physical interpretation in terms of the underlying object. Second, the HRP decomposition provides a parsimonious

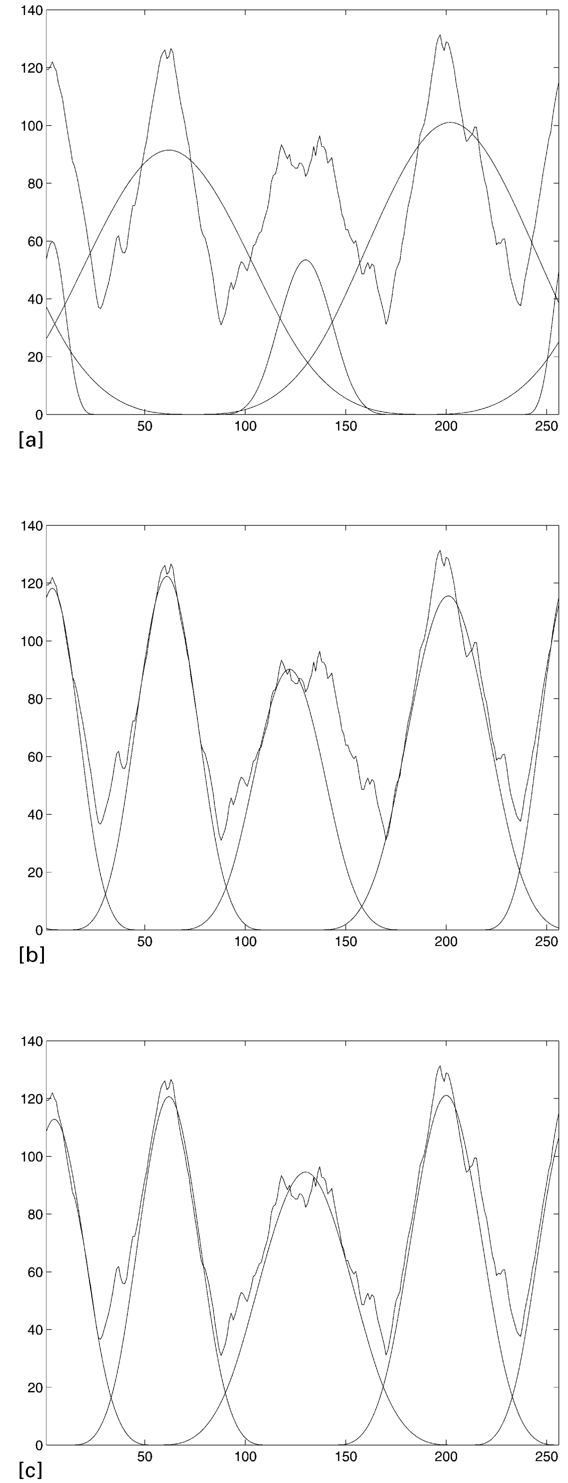


Fig. 7. Effect of k on the HRP decomposition. These graphs show HRP decomposition of a noisy CDP from Plane #1 using different values of k, the depth parameter. (a) $k = 1$ (b) $k = 2$ (c) k varying at each step.

representation of the underlying signal. In the example in Fig. 7c, a qualitatively accurate description of the noisy CDP is obtained using only four dictionary elements. Each HRP element is described by three parameters (scale, translation, and magnitude) and therefore, one can imagine constructing a feature vector of low dimension which gives a qualitative description of the CDP. In contrast, the traditional method based on FD utilizes a 128-dimensional feature vector. Third, the elements of the HRP decomposition shown in Fig. 7c are localized. That is, the elements corresponding to adjacent physical plane structures do not interfere with one another.

3. HRP-based recognition

Having illustrated some of the attractive properties of features extracted by the HRP algorithm, we now turn our attention to describing a feature extraction and recognition algorithm which takes advantage of these properties. Again, our basic approach is to use the HRP algorithm to construct a feature vector of low dimension which can then be used as the basis for recognition. In this section, we begin by describing the HRP-based feature extraction process which is used for both for training the algorithm and for recognition of unknown profiles. Then, we describe the particulars of the training phase of the algorithm. Finally, we describe the recognition algorithm which is based on M-ary hypothesis testing and the generalized likelihood ratio test.

3.1. *HRP-based feature extraction*

Since our underlying goal is to extract features which are suitable for object recognition, we must extract features which are robust and statistically significant. There are a number of questions to address in order to use HRP to extract features which can be used to discriminate among the planes in our data set. The particulars of the HRP algorithm must be specified, including the number of HRP elements to extract and the appropriate value(s) of the tuning parameter k. Further, we examine a means to incorporate the angular information, which is essentially discarded by the CDP, to obtain a more robust feature set. Finally, we touch briefly on the issue of ordering the parameters extracted by HRP into a feature vector.

3.1.1. *HRP implementation*

Drawing from the examples in Section 2.4, we can specify the number of HRP elements and the appropriate values of k. First, as we have already seen the four peaks in the CDPs correspond to the four physical features of the aircraft, namely, the nose, the wings, and the tail. Given that CDPs for most planes in the data base exhibit four similar robust features in the presence of noise, one logical approach would be to extract four HRP elements from each profile. Since each HRP element can be described by three parameters (i.e. scale, translation, magnitude), this approach leads to a 12-dimensional feature vector, which is an order of magnitude smaller than the feature vector used in the FD-based algorithm.[1] Second, the tuning parameter k should be used to incorporate prior knowledge so that HRP elements correspond to the important physical features of the object and provide robustness in the HRP decompositions. As the example in Fig. 7c showed, allowing k to vary at each step in HRP algorithm yields a physically meaningful and accurate description of the CDP used in the example. Through experimentation, we have found that using this same pattern for k (i.e. $k = 3$ for the first two elements, $k = 2$ for the third element, and $k = 1$ for the fourth element), in general, yields robust and accurate decompositions for the CDP's from each of the planes in the data set. Further, our experimentation results show that this pattern of values for k yields qualitatively accurate descriptions in the presence of several different values of boundary noise and occlusional noise. Therefore, for both the training and testing data, k will be 3 for the first two elements, 2 for the third element and 1 for the fourth element.

The output of the HRP-based feature extraction procedure described up to this point is a set of four vectors, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$, where each $\mathbf{e}_i$ consists of the three parameters (scale, translation, and magnitude) which describe each HRP element. In the next section, we replace the scale and translation parameters with more robust angular equivalents.

3.1.2. Incorporating angular information

The angular profile, essentially discarded in the CDP representation of a profile, is important for obtaining robust HRP features. Recall that each of the four HRP elements is represented by three parameters, scale, translation and magnitude. While the translation parameter extracted using HRP can be highly variable due to noise and occlusion, the equivalent angular position is much more stable. The HRP scale parameters are somewhat affected by boundary noise and most definitely affected by occlusion, but rescaling the HRP scale parameters to angular extent rather than translational extent along the boundary leads to more stable features. By replacing the HRP translation parameters with their angular equivalents and rescaling the HRP scale parameters using the slope of the angular profile, we obtain a more robust feature vector in the face of boundary noise and occlusion as well as variations in orientation and overall scale.

Replacing the HRP translation parameters with their angular equivalents is straightforward. If $\theta(n)$ is the angular profile accompanying the CDP, $f(n)$, then the translational parameter t of each HRP element is replaced by $\theta(t)$.

To illustrate the variability of the HRP scale parameters in the presence of occlusion and to motivate our rescaling of the HRP scale parameters, consider the following example. Suppose we are given an occluded silhouette of Plane #1 as shown in Fig. 8 and the corresponding CDP as shown in Fig. 9a. Fig. 9a also shows the HRP element which corresponds to the wing feature in the occluded profile. For comparison, Fig. 9b shows the CDP for an unoccluded version of Plane #1 and the HRP element which represents the wing feature. We note that the scale parameter of the HRP element which represents the wing feature in the occluded profile is quite different from that of the HRP element which represents the wing feature in the unoccluded profile. Thus, the occlusion has caused a shift in the scale parameter extracted by HRP. However, we also observe that the occlusion has also caused a change in the slope of the angular profile over the support of the element corresponding to the wing. The question we would like to answer is: How can this change in the local slope of the angular profile be used to account for the change in scale in the HRP wing element in order to produce a robust measure of "angular scale?"

To see how this rescaling can be accomplished, consider a single HRP feature. Specifically, suppose that $g_0(n)$ is a cubic b-spline at scale j extracted from the CDP $f(n)$ and that $a_0(n)$ is the best fit line to the angular profile $\theta(n)$ over the support of $g_0(n)$. Then, the polar coordinates $(g_0(n), a_0(n))$ trace out an approximation to a portion of the boundary curve described by the polar coordinates $(f(n), \theta(n))$. Actually, the approximating curve, $(g_0(n), a_0(n))$, may be represented by many different pairs of polar coordinates of the form $(g(n), a(n))$ where g is a cubic b-spline and a is a line. For example, the curve

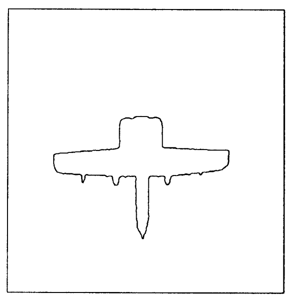


Fig. 8. The boundary corresponding to Plane #1 with the tail occluded

[1] One could also imagine extracting more HRP elements from each profile. While these subsequent HRP elements may provide a better fit to the CDP, they are also less robust to noise and therefore exhibit a greater variability.

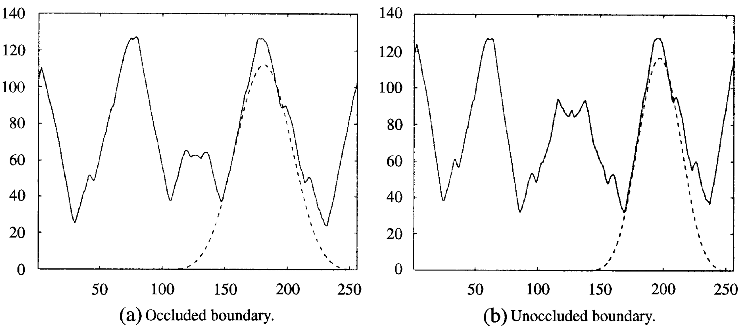


Fig. 9. A comparison of HRP elements corresponding to the wing of Plane #1 for the occluded and unoccluded profiles.

$(g_0(n), a_0(n))$ may also be represented by $(g(n), a(n))$ where g is a cubic b-spline with half the support of g_0 (i.e. a cubic b-spline at scale $j+1$) and a is a straight line with double the slope of a_0. Generalizing this intuition, we obtain a relationship for normalizing the scale of any HRP element based on the local slope of the angular profile. Given that the scale of an HRP element extracted from a CDP, f, is j and the slope of the corresponding angular profile over the support of the element is s, then the slope-adjusted scale will be given by $B(j)$ where

$$B(j) = j + \log_2\left(\frac{s_0}{s}\right), \tag{10}$$

where $s_0 = 2\pi/P$ and P is the number of points in the CDP. The reference scale s_0 is the average slope of the angular profile over the entire profile.

3.1.3. Feature vector ordering

The output of the feature extraction procedure described up to this point is a set of four vectors, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$, where each $\mathbf{e}_i$ consists of three components, which have been extracted by HRP and then modified to include angular information. One remaining topic for discussion is: how should these 12 parameters be organized in a feature vector suitable for recognition? Ideally, we would like to organize the HRP parameters in a feature vector according to the physical features that they describe. That is, we would like to construct our feature vector $\mathbf{y}$ to have the following structure:

$$\mathbf{y} = [\mathbf{e}_{nose}\ \mathbf{e}_{wing1}\ \mathbf{e}_{tail}\ \mathbf{e}_{wing2}]^{\mathrm{T}}, \tag{11}$$

where

$$\mathbf{e}_{\mathrm{nose}} = [B(j_{nose})\ \theta(t_{nose})\ S_{nose}]^{\mathrm{T}},$$

$$\mathbf{e}_{\mathrm{wing1}} = [B(j_{wing1})\ \theta(t_{wing1}) - \theta(t_{nose})\ S_{wing1}]^{\mathrm{T}},$$

$$\mathbf{e}_{\mathrm{tail}} = [B(j_{tail})\ \theta(t_{tail}) - \theta(t_{nose})\ S_{tail}]^{\mathrm{T}},$$

$$\mathbf{e}_{\mathrm{wing2}} = [B(j_{wing2})\theta(t_{wing2}) - \theta(t_{nose})\,S_{wing2}]^{\mathrm{T}},$$

where j_{nose}, t_{nose}, and S_{nose} are the scale, translation, and magnitude, respectively, of the HRP element corresponding to the nose feature in the CDP, and $B(j)$ is as defined in Eq. (10). Thus, $\mathbf{y}$ is a vector of 12 components which are derived from the parameters of the HRP elements. In the training procedure, we know truth and therefore we can order to achieve the ordering given in Eq. (11) more easily. In Section 3.2, we describe how this was done and, in particular, we will see that different orderings were required for different aircraft. However, in the recognition algorithm, since the true aircraft is unknown, we must define an ordering procedure that works on-line and reflects possible variability in ordering and we describe this in Section 3.3.

Note that we will occasionally refer to $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ as the elemental components of $\mathbf{y}$. In contrast, when we refer to the components of $\mathbf{y}$ we mean the individual parameters of the vector $\mathbf{y}$ as enumerated in Eq. (11). Further, note that we have used the difference between the angular parameter of the nose and all the other elements to construct $\mathbf{y}$. In doing so, we have used the stability of the nose feature to circularly shift the other features and obtain a level of orientation invariance.

3.2. Training procedure

Training data are used to construct the probability density functions that will be incorporated in the recognition phase of the algorithm. The statistical variability of the HRP-based features is modest, and therefore, we are easily able to characterize the probability density functions which accurately describe the variability of the HRP-based features. In fact, as we will illustrate in this section, the HRP features for each object class are well modeled by the Gaussian and exponential cosine [24] distributions. The only assumption made about the training data is that the orientation is known so that all the CDPs start at the tip of the nose and proceed counter-clockwise.

For the training data, HRP-based features are extracted exactly as described in Section 3.1 and a feature vector of form Eq. (11) is constructed in the following manner. We have found that for all the aircraft in the data set, the desired ordering can be obtained using either the translation or the magnitude parameters extracted by the HRP algorithm, depending upon the aircraft. Specifically, feature vectors extracted from all aircraft except Planes #7 and #17 can be ordered according to the HRP translation parameters. Specifically, the nose is taken to be the HRP element which has a translation $t \in [1,.1P] \cup [0.9P,P]$, where P is the number of points in the CDP, the element with the next largest t is taken to be the first wing, the element with the next largest t is taken to be the tail, and the remaining element is taken to be second wing. Feature vectors extracted from Planes #7 and #17 can be ordered according the HRP magnitude parameters. That is, the nose is taken to be the HRP element with the maximum magnitude, the tail is taken to be the HRP element with the next largest magnitude, and the wings are taken to be the elements with the smallest magnitudes. Note that since we have assumed that all the profiles are oriented to begin at the tip of nose, $\theta(t_{\text{nose}}) = -\pi/2$ for all the planes.

For each object class, the variability of the components of the feature vector $\mathbf{y}$ is well modeled by the Gaussian and exponential cosine distributions. Note that we assume the components of $\mathbf{y}$ are independent so that the conditional probability densities for each object class are given by

$$p_{\mathbf{y}|\mathrm{H}_m}(\mathbf{Y}|\mathrm{H}_m) = \prod_{i=1}^{12} p_{y_i|\mathrm{H}_m}(Y_i|\mathrm{H}_m), \tag{12}$$

where each hypothesis H_m corresponds to a particular plane from the data set. For each object class, the Gaussian distribution provides a reasonably accurate representation for the variability of all the components of $\mathbf{y}$ except for the angular parameters. That is, the variability of components #1,3,4,6,7,9,10,12 of $\mathbf{y}$ is well approximated by

$$p_{y_i|\mathrm{H}_m}(Y_i|\mathrm{H}_m) = \frac{1}{\sqrt{2\pi}\sigma_{m,i}} \exp\left(-\frac{(Y_i - E_{m,i})^2}{2\ \sigma_{m,i}^2}\right), \tag{13}$$

where $E_{m,i}$ and $\sigma_{m,i}$ are the mean and variance for component i under hypothesis m. For each object class, $E_{m,i}$ and $\sigma_{m,i}$ are obtained from the sample means and sample variances from the training data. To illustrate that the Gaussian is a suitable model for the variability of the HRP features, Fig. 10 shows a histogram of the training data for component #4 of $\mathbf{y}$ for Plane #1 and the corresponding Gaussian distribution model. The remaining components (i.e. #2,5,8,11) of $\mathbf{y}$ are well modeled by the exponential cosine distribution [24], which is given by

$$p_{y_i|\mathrm{H}_m}(Y_i|\mathrm{H}_m) = \frac{\exp(\alpha_{m,i}\cos(Y_i - E_{m,i}))}{2\pi I_0(\alpha_{m,i})}, \tag{14}$$

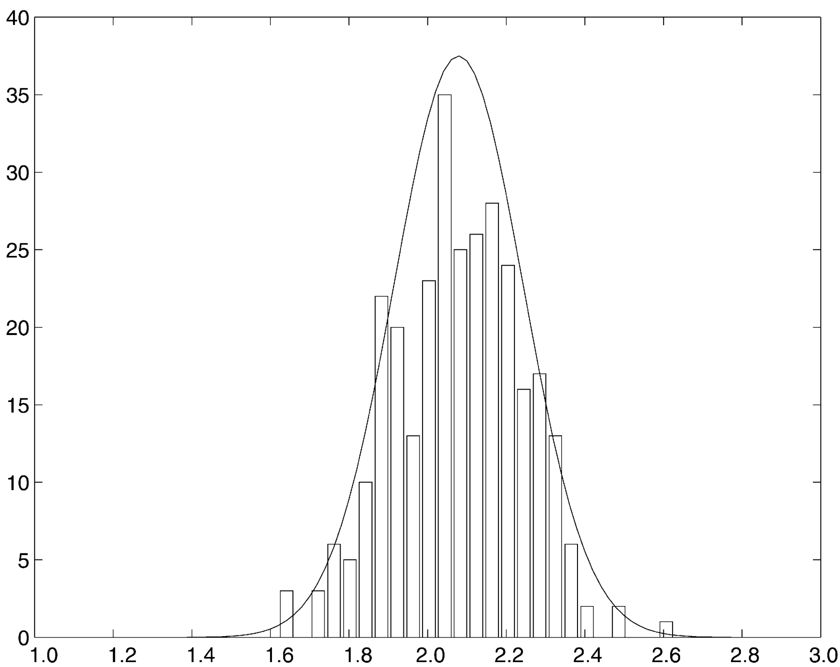
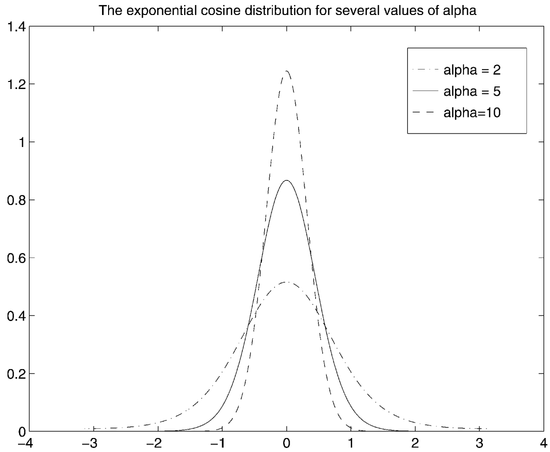
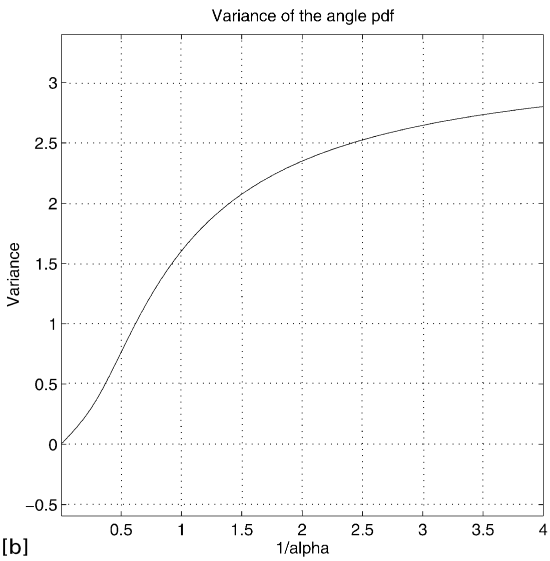


Fig. 10. A histogram of training data for component #4 for Plane #1, and the corresponding Gaussian distribution model.



[a]



[b]

Fig. 11. The exponential cosine distribution and variance as a function of α. (a) The exponential cosine distribution for several values of α. (b) Variance of the exponential cosine distribution as a function of α.

where I_0 is a zeroth-order modified Bessel function of the first kind, $E_{m,i}$ is the mean of component i under hypothesis m, and $\alpha_{m,i}$ is a parameter related to the variance of component i under hypothesis m. Fig. 11a shows this exponential cosine distribution for $E_{m,i} = 0$ and several values of α. For large values of α, the exponential cosine distribution looks very similar to the Gaussian distribution. The variance of this distribution is given by

$$\sigma^2 = \frac{\pi}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n I_n(\alpha)}{n^2 I_0(\alpha)}. \tag{15}$$

Fig. 11b shows the relationship between α and σ. The parameter $E_{m,i}$ is approximated as the sample mean of y_i of the training data for plane m. The parameter $\alpha_{m,i}$ is determined from the sample variance of y_i of the training data for plane m, together with Eq. (15).

3.3. Recognition algorithm

The basic recognition algorithm is an M-ary hypothesis testing scheme which uses the conditional probability distributions Eq. (12) obtained from the training procedure described in Section 3.2. This basic hypothesis testing algorithm is amended to include the generalized likelihood ratio test [25] (GLRT) to accommodate the possibility of switches in the elemental components of the feature vector (due to errors in ordering the HRP features) and the possibility that portions of the data vector are corrupted because of occlusion (so that one or more true features are not present in the silhouette and, consequently, one or more of the HRP features do not represent useful features). The only assumption made in the recognition phase is that coarse knowledge of the orientation is available so that each CDP begins within $\pm 30^\circ$ of the tip of the nose.

In the recognition phase, HRP-based features are extracted from the CDP's as described in Section 3.1 and ordered according to the translation parameters of the HRP elements in the following manner. The feature vector $\mathbf{y}$ is ordered according to translation so that the nose is taken to be the HRP element which has a translation $t \in [1, 0.1P] \cup [0.9P, P]$, where P is the number of points in the CDP, and the remaining elements are ordered according to translation. In constructing the feature vector $\mathbf{y}$, we set $\theta(t_{nose}) = -\pi/2$ and thereby adjust for the coarse estimate of orientation. As a result, the algorithm is insensitive to orientation variation of less than 30° in magnitude (i.e. to variations consistent with our coarse level knowledge of orientation).

The basic recognition algorithm is an M-ary hypothesis testing scheme which uses the conditional probability distributions developed in Section 3.2. The most likely hypothesis is given by

$$H_4^* = \arg\max_{\mathrm{H}_m} p_{\mathrm{y}|\mathrm{H}_m}(\mathbf{Y}|\mathrm{H}_m), \tag{16}$$

where we have included the subscript "4" to indicate that this is a decision based on the four elemental components in $\mathbf{y}$ and $p_{\mathrm{y}|\mathrm{H}_m}(\mathbf{Y}|\mathrm{H}_m)$ is as given by Eqs. (12)–(14). If the underlying conditional distributions $p_{y_i|\mathrm{H}_m}(Y_i|\mathrm{H}_m)$ were all Gaussian, then maximizing $p_{\mathrm{y}|\mathrm{H}_m}(\mathbf{Y}|\mathrm{H}_m)$ is equivalent to minimizing a weighted distance. Since some of the underlying conditional distributions are exponential cosine instead of Gaussian, maximizing $p_{\mathrm{y}|\mathrm{H}_m}(\mathbf{Y}|\mathrm{H}_m)$ is not precisely equivalent to minimizing a weighted distance. Intuitively, however, since the exponential cosine may be thought of as a Gaussian for large values of α, we can think of maximizing $p_{y|H_m}(\mathbf{Y}|H_m)$ as being very close to minimizing a weighted distance. This basic procedure is amended in light of the following important caveats.

3.3.1. Feature vector ordering

For recognition, rather than training, we will always be presented with a feature vector $\mathbf{y}$ that has been ordered according to translation. For most aircraft, this ordering is robust and stable so that the elements in the feature vector correspond, in order, to the nose, the first wing, the tail, and the second wing. For two aircraft (Planes #7 and #17), this ordering does not produce a stable feature vector. For these aircraft, the probability density functions are modified to include a nuisance variable, β, corresponding to a random permutation of the elemental components of $\mathbf{y}$. For example, when evaluating the likelihood for hypothesis #7 in the recognition procedure, the GLRT is used to determine which permutation of the elemental components is most likely, as

$$\beta^* = \max_{\beta} p_{y|H_7,\beta}(\mathbf{Y}|H_7,\beta). \tag{17}$$

Then, the quantity $p_{y|H_7,\beta}(\mathbf{Y}|H_7,\beta^*)$ is used to compute the likelihood ratios in the M-ary hypothesis test. The procedure for hypothesis #17 is analogous.

3.3.2. Occlusion

Further, as a result of occlusion, one or more of the true object features (nose, wings, tail) maybe missing and as a consequence one or more of the elemental components of $\mathbf{y}$ may not represent a meaningful object feature. However, since the HRP algorithm is locally based, the remaining elemental components still give an accurate description of the physical features of the underlying plane. Thus, a reasonable model in such a case is to include another nuisance variable corresponding to one or more of the four features being essentially meaningless for recognition. Again, the GLRT is used to determine which if any of the features are meaningless due to occlusion.

For example, suppose precisely one elemental component has been affected by occlusion. In this case, the GLRT is used to determine, for each aircraft hypothesis H_m, the element τ^* which is least likely to correspond to a meaningful feature. Specifically,

$$\tau_m^* = \arg\max_{\tau} p_{y|H_m,\tau}(\mathbf{Y}|H_m,\tau), \tag{18}$$

where

$$p_{y|H_m,\tau}(\mathbf{Y}|H_m,\tau) = \begin{cases} \prod_{i\in\{2,3,4\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if} \tau = 1, \\ \prod_{i\in\{1,3,4\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if} \tau = 2, \\ \prod_{i\in\{1,2,4\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if} \tau = 3, \\ \prod_{i\in\{1,2,3\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if} \tau = 4, \end{cases} \tag{19}$$

where $\mathbf{E}_i$ refers to the elemental components of $\mathbf{Y}$. Then, the most likely hypothesis is chosen by excluding the component which is least likely to be meaningful for each hypothesis, as follows:

$$H_3^* = \arg\max_{H_m} p_{y|H_m,\tau}(\mathbf{Y}|H_m,\tau_m^*), \tag{20}$$

where we have included the subscript "3" to indicate that this is a decision based on three elemental components of $\mathbf{y}$.

Similarly, suppose that precisely two elemental components have been affected by occlusion, then using the GLRT we can determine for each hypothesis H_m which two elemental components are least likely to correspond to meaningful features. That is, find δ_m^* where

$$\delta_m^* = \arg\max_{\delta} p_{y|H_m,\delta}(\mathbf{Y}|H_m,\delta) \tag{21}$$

and

$$p_{y|H_m,\delta}(\mathbf{Y}|H_m,\delta) = \begin{cases} \prod_{i\in\{1,2\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if } \delta = 1, \\ \prod_{i\in\{1,3\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if } \delta = 2, \\ \prod_{i\in\{1,4\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if } \delta = 3, \\ \prod_{i\in\{2,3\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if } \delta = 4, \\ \prod_{i\in\{2,4\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if } \delta = 5, \\ \prod_{i\in\{3,4\}} p_{e_i|H_m}(\mathbf{E_i}|H_m) & \text{if } \delta = 6. \end{cases} \tag{22}$$

Then, determine the most likely hypothesis by excluding the components which are least likely to be meaningful for each hypothesis, as follows:

$$H_2^* = \arg\max_{H_m} p_{y|H_m,\delta}(\mathbf{Y}|H_m,\delta_m^*), \tag{23}$$

where we have included the subscript "2" to indicate that this is a decision based on two elemental components of $\mathbf{y}$.

The question now arises if we are given a unknown a feature vector $\mathbf{y}$ extracted from CDP a silhouette which may or may not be distorted due to occlusion, how does one determine if zero, one or two elemental components correspond to meaningless features and what is the most likely hypothesis given the data. In other words, if $H_4^* \neq H_3^* \neq H_2^*$, which hypothesis should we choose? Our approach is to choose among H_2^*, H_3^* and H_4^* based on the likelihoods associated with each of these results. For convenience, let us define $p_{y|H_2^*}(\mathbf{Y}|H_2^*)$, $p_{y|H_3^*}(\mathbf{Y}|H_3^*)$, and $p_{y|H_4^*}(\mathbf{Y}|H_4^*)$ as

$$p_{y|H_2^*}(\mathbf{Y}|H_2^*) = \max_{H_m} p_{y|H_m,\delta}(\mathbf{Y}|H_m,\delta_m^*), \tag{24}$$

$$p_{y|H_3^*}(\mathbf{Y}|H_3^*) = \max_{H_m} p_{y|H_m,\tau}(\mathbf{Y}|H_m,\tau_m^*), \tag{25}$$

$$p_{y|H_4^*}(\mathbf{Y}|H_4^*) = \max_{H_m} p_{y|H_m}(\mathbf{Y}|H_m). \tag{26}$$

Recall our intuition that the M-ary hypothesis test based on Gaussian and exponential cosine conditional densities is roughly equivalent to minimizing a weighted distance. Intuitively, since the distance between n elements will be larger than the distance between $n-1$ elements, we expect

$$p_{\mathbf{y}|\mathrm{H}_2^*}(\mathbf{Y}|\mathrm{H}_2^*) > p_{\mathbf{y}|\mathrm{H}_3^*}(\mathbf{Y}|\mathrm{H}_3^*) > p_{\mathbf{y}|\mathrm{H}_4^*}(\mathbf{Y}|\mathrm{H}_4^*). \tag{27}$$

Thus, our modified decision rule becomes

$$\mathrm{H}^* = \arg\max(p_{\mathbf{y}|\mathrm{H}_4^*}(\mathbf{Y}|\mathrm{H}_4^*), p_{\mathbf{y}|\mathrm{H}_3^*}(\mathbf{Y}|\mathrm{H}_3^*)/\mathscr{P}, p_{\mathbf{y}|\mathrm{H}_2^*}(\mathbf{Y}|\mathrm{H}_2^*)/\mathscr{P}^2), \tag{28}$$

where $\mathscr{P}$ is an empirically determined term to penalize the exclusion of signal features.

3.3.3. Scale invariance

Up until this point, we have not discussed the effect of scale on our M-ary hypothesis testing plus GLRT recognition scheme. Recall from Section 2.2 that scale variation in the silhouette leads to a change in amplitude in the CDP. Also, note that amplitude variation in a signal will only effect the magnitude parameter of the elements of the HRP decomposition of that signal. Certainly, the algorithm as we have described it will not be robust to variations in the scale of the silhouette. However, robust classification in the presence of scale variation of the silhouette can be accomplished with a very simple change in the basic algorithm. The basic M-ary hypothesis testing plus GLRT algorithm can be made insensitive to scale variation by normalizing each CDP by its average value. That is, for both the training and testing data sets, each CDP is normalized by its average value and the recognition proceeds as previously outlined.

4. Results

In this section, we describe experiments to test our HRP-based recognition algorithm. We begin by demonstrating our approach on a grayscale image which has been perturbed with white noise. Then, to obtain more statistically significant and exhaustive results, we demonstrate our approach on simulated data. The results using simulated data indicate that this approach is robust to boundary perturbations, scale variation, orientation variations, distortions due to occlusion, and to localized variations in the boundary.

For comparison, we also show results based on Fourier descriptors (FD) and the K nearest neighbor test with $K = 3$. This approach based on FD was described in Ref. (10) and was summarized in Section 2.3. Note that this approach uses $P/2$ features where P is the number of points in the CDP. In the experiments we describe in this section $P = 256$ so the FD technique uses 128 features. In contrast, the HRP based technique uses a feature vector with only 12 components.

4.1. Gray-scale image

We begin by demonstrating the HRP-based method on silhouettes extracted from a gray-scale image perturbed by white noise. The gray-scale image in Fig. 2 shows a Viggen aircraft which is Plane #17 in our data set. This image was perturbed by additive white Gaussian noise and silhouette contours were extracted using a snakes algorithm [26]. The recognition algorithm, however, was not trained on the real imagery but rather on noisy simulated data using the model described in Section 2.2 with $p = 40\%$ and $s = 0.9$.[2] The classification results using the HRP-based algorithm and the FD-based algorithms are shown in Table 1. For the sample sizes used in these tests, the estimate of percent correct classification have a standard deviation of 8%. These results indicate that the performance of the HRP-based algorithm far surpasses that of the FD-based algorithm in the stressing case of high noise. In contrast, the HRP-based algorithm does not degrade at all in the presence of high noise.

However, to test the full merits of our algorithm in a statistically meaningful way, we need a large and rich set of samples spanning all classes of objects (i.e. all 17 planes) and a variety of different types of conditions. Therefore, in the following five sections, we turn to simulated data to demonstrate the strengths of our HRP-based approach.

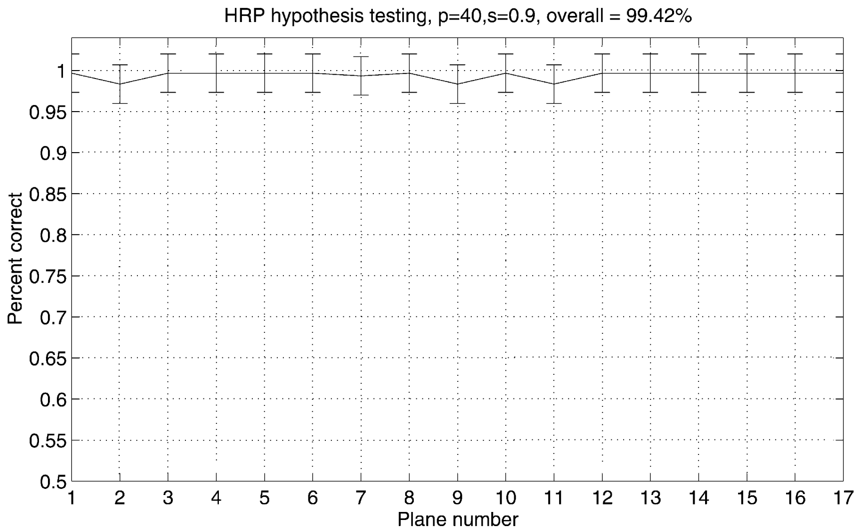
4.2. Boundary perturbations

Our first simulated experiment tests the performance of the HRP-based algorithm in the presence of boundary perturbations. We construct training and testing data sets using the boundary perturbation model described in Section 2.2 for $p = 40\%$ and for several values of s. In

Table 1
Percent correct classification of noisy versions of the grayscale image shown in Fig. 2. The standard deviation of the estimate of percent correct classification is about 8%

SNR per sample (d/s)	FD %	HRP (%)
10	96.30	96.30
0	37.04	96.30

[2] This is exactly the same training used in the Monte Carlo simulations in the following sections.



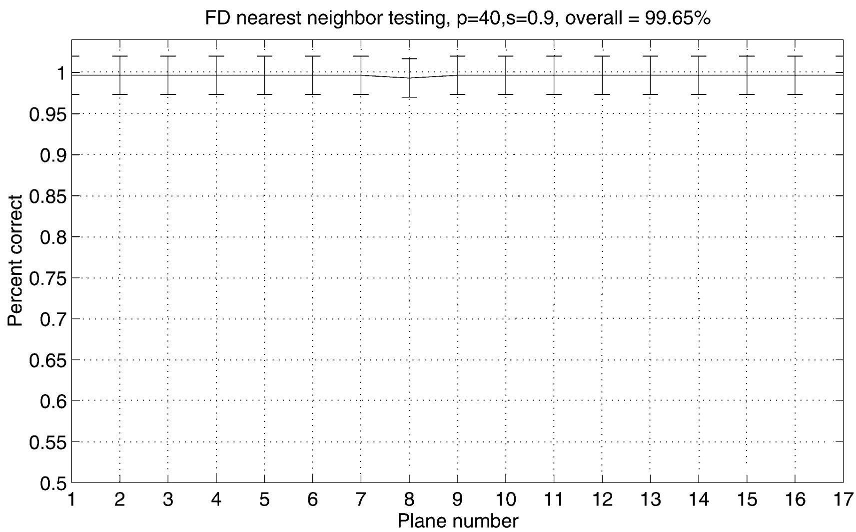


Fig. 12. Classification results for noisy data generated using $p = 40\%$ and $s = 0.9$. For this experiment, the noise level in the training and testing data was the same and there was no scale or orientation variation. This figure shows the percent correct classificationh of the testing data for each plane and the one standard deviation error bars.

this section, there is no scale variation, orientation variation, or variation due to occlusion in either the testing or training data sets.

For the first experiment, both the training and testing data sets consisted of noisy profiles generated using $p = 40\%$ and $s = 0.9$. That is, the noise level was the same for both the training and testing data. Fig. 12 shows the percent correct classification of the testing data as a function of plane number for both our HRP-based technique and the FD nearest-neighbor classifier. Fig. 12 also shows the one standard deviation error bars for the estimate of percent of correct classification for each plane. Both techniques perform very well. The HRP technique gives 99.42% correct classification overall and the FD technique gives 99.65% correct classification overall. Given that the standard deviation of the estimate of the percent correct classification is about 2.5% for each plane, we conclude that this difference in overall performance is not statistically significant. These results are included in the first row of Table 2, which summarizes all of the comparative results from this and subsequent sections.

Next, noisy profiles were generated using $p = 40\%$ and $s = 1.5$. Again, the noise level was the same for both the training and testing data sets. The overall results for the HRP-based and FD-based algorithm are listed in the second row of Table 2. The overall percent correct classification is 96.53% for the HRP-based technique and 97.64% for the FD-based technique. Again, given that the standard deviation of the estimate of the percent correct classification 2.5% for each plane, we conclude that this difference in overall performance is not statistically significant. Similarly, noisy profiles were generated using $p = 40\%$ and $s = 2.1$, and results are included in the third row of Table 2. Fig. 13 shows the percent correct classification of the testing data as a function of plane number for both our HRP-based technique and the FD nearest neighbor classifier. The overall percent correct classification is 94.51% for the HRP-based technique and 91.37% for the FD technique. Once again, the standard deviation of the estimate of percent correct classification for each plane is 2.5%. Thus, these results indicate that the HRP-based technique has slightly better overall performance, at a modest level of statistical significance.

Table 2
Summary of the experimental results for Sections 4.2–4.6. For all the results listed here the standard deviation for the estimate of percent classification for each plane is 2.5%

Experimental description	Section Number	FD (%)	HRP (%)
Boundary noise, $p = 40\%$, $s = 0.9$	4.2	99.65	99.42
Boundary noise, $p = 40\%$, $s = 1.5$	4.2	97.64	96.53
Boundary noise, $p = 40\%$, $s = 2.1$	4.2	91.37	94.51
Bdy noise (40%, 2.1) + rotation variation	4.3	91.27	94.45
Bdy noise (40%, 2.1) + scale variation	4.4	91.37	90.01
Occlusion ($q = 10\%$)	4.5	89.66	94.74
Bdy noise (40%, 0.9) + localized silhouette variation	4.6	12.58	92.72

4.3. Rotational noise

Next, we construct testing data which have unknown rotational variation between $\pm 10^{\circ}$ plus boundary perturbations generated with $p = 40\%$ and $s = 2.1$, the largest level of boundary perturbations. The training data used was generated without rotation and with boundary perturbations using $p = 40\%$ and $s = 2.1$. The overall performance of the two techniques is included in the fourth row of Table 2. The overall percent correct classification is 94.45% for the HRP-based technique and 91.27% for the FD technique. Once again, the standard deviation of the estimate of percent correct classification for each plane is 2.5%. Thus, these results indicate that the HRP-based technique has slightly better overall performance, at a modest level of statistical significance.

4.4. Scale variation

As we mentioned in Section 3.3.3, the basic M-ary hypothesis testing plus GLRT algorithm can be made insensitive to scale variation by normalizing each CDP by its average value. That is, for both the training and

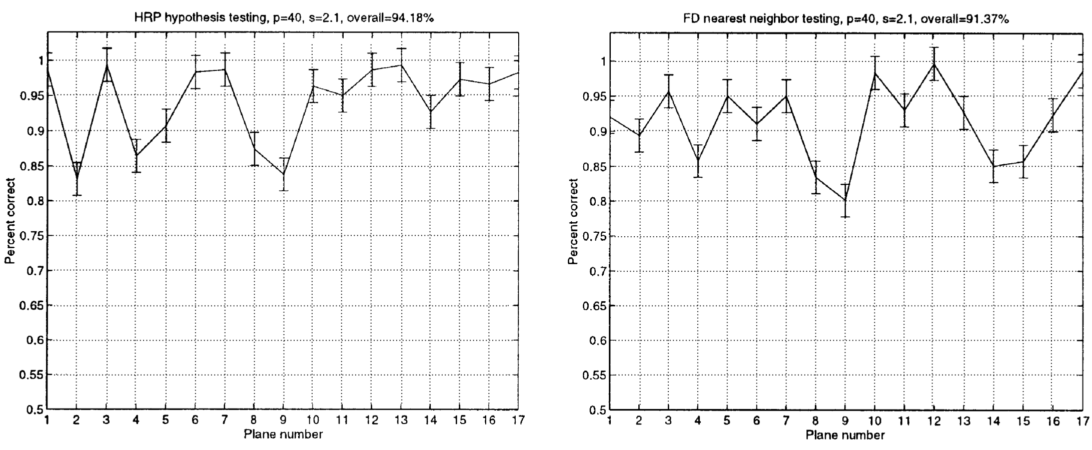


Fig. 13. Classification results for noisy data generated using $p = 40\%$ and $s = 2.1$. For this experiment, the noise level in the training and testing data was the same and there was no scale or orientation variation. This figure shows the percent correct classificationh of the testing data for each plane and the one standard deviation error bars.

testing data sets, each CDP is normalized by its average value and the recognition proceeds as previously outlined. The performance of this scale insensitive version of HRP is illustrated in the following experiment. For this experiment, noisy CDP's are generated as described in Section 2.2 with $p = 40\%$ and $s = 2.1$ and normalized by their average value. The overall performance of the two techniques for this experiment is included in the fifth row of Table 2. These results indicate that both techniques perform similarly and the difference in the overall performance is not statistically significant.

4.5. Occlusion

Next, we investigate the performance of the M-ary hypothesis plus GLRT recognition approach in the presence of occlusion. For this experiment, the training data set consists of CDPs with the smallest level of boundary noise, i.e. $p = 40\%$ and $s = 0.9$. For the testing data set, we construct CDPs corrupted by occlusion as described by Liu [12] and summarized in Section 2.2 with $q = 10\%$. That is, for the testing data set 10% of the boundary is replaced by a straight line. The overall performance of the two techniques for this experiment is included in Table 2. The HRP technique shows 94% correct classification while the FD technique shows only 88% correct classification, a statistically significant difference in which the classification errors have been cut in half.

4.6. Localized silhouette variation

Next, we investigate the performance of the HRP-based feature extraction and recognition scheme in the presence of localized variation in the silhouette. Fig. 14a shows the silhouette of Plane #8 in the database and Fig. 14b shows a silhouette of Plane #8 carrying an alternate load. The difference between these two silhouettes is slight and is concentrated along the wings of the aircraft. For this experiment, the training data is generated using the boundary perturbation model described in Section 2.2 with $p = 40\%$ and $s = 0.9$ and the version of Plane #8 is the same one used in the previous sections, namely the one shown in Fig. 14a. The testing data, however, consisted of noisy boundaries ($p = 40\%$, $s = 0.9$) of the version of Plane #8 shown in Fig. 14b which carries an alternate load. In this case, the HRP-based recognition scheme gave 92.72% correct classification and the standard deviation of the estimate of the percent correct classification was about 2.5%. In contrast, the FD-based technique yielded correct classification of less than 15%. Thus, in the presence of localized silhouette variation the performance of the HRP-based recognition scheme far exceeds that of the FD-based recognition scheme.

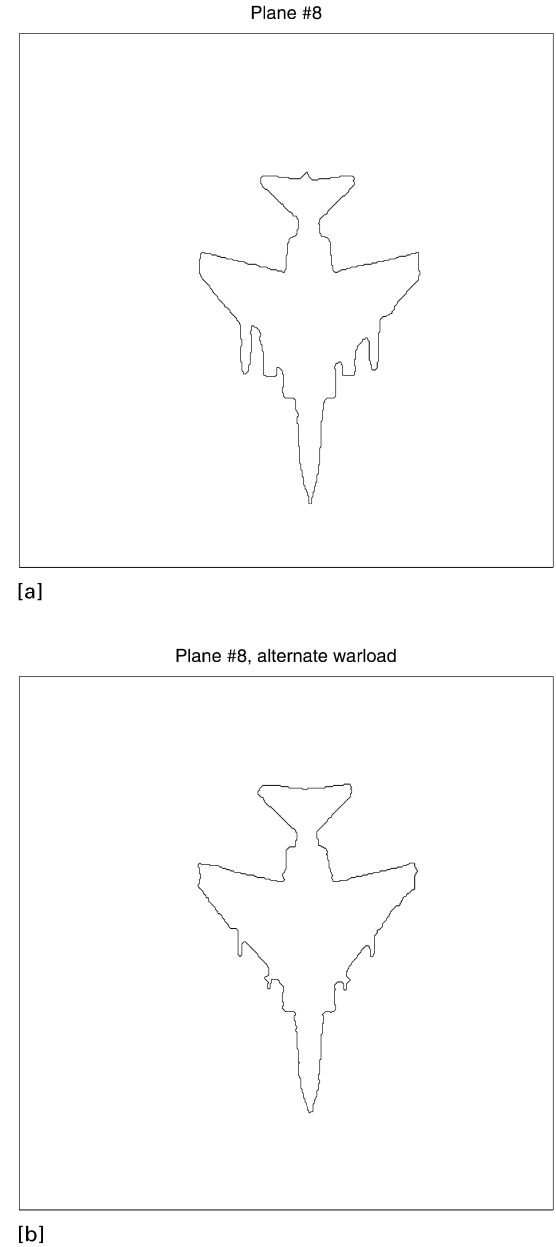


Fig. 14. (a) The version of Plane #8 used in the training data. (b) The version of Plane #8 used in the testing data.

5. Conclusions

In this paper, we have illustrated a new algorithm for feature extraction and object recognition based on high-resolution pursuit (HRP). In the object recognition context, the elements extracted by HRP are a new class of features that describe the geometric (i.e. size and location) properties of subparts of the object. The HRP-based recognition approach demonstrated in this paper is fast and robust to many different types of variations (boundary perturbation, orientation variation, scale variation, occlusion, and localized silhouette variation). We have demonstrated that this HRP-based approach is a highly competitive algorithm for this particular recognition problem. In fact, in the presence of variations due to occlusion and localized variation in the silhouette, the performance of the HRP-based method far surpasses the performance of the FD-based method. An equally important contribution of this paper is that it demonstrates one of the first uses of the emerging methods of adaptive function approximation to object recognition and clearly shows the promise of such approaches.

The promising results of this work may be extended in several different directions to obtain a more powerful recognition system. One area for further research is the development of more sophisticated coupled probabilistic models of the HRP features. Even though the simple probabilistic models using an independence assumption provided promising results, the performance of the HRP-based recognition scheme may be enhanced by more sophisticated probabilistic models. In addition, one might imagine developing Markov chain models which capture the idea that the choice of HRP elements depends on the previous elements extracted from the CDP. Further, this work has already shown that the using the coarsest features in the hierarchy of features extracted by HRP leads to a competitive recognition algorithm, but more HRP elements could be extracted from the CDPs. The incorporation of additional HRP features may be used for finer classification, although admittedly they will exhibit greater uncertainty. Finally, the simple paradigm of setting the tuning parameter k independent of object class may be modified to obtain features which are adapted to each object class.

6. Summary

This paper introduces a simple new approach to object recognition from silhouettes. In contrast to current approaches in model-based silhouette recognition, this new approach utilizes features extracted using an adaptive approximation technique called high-resolution pursuit (HRP). HRP is an attractive technique for feature extraction because it yields features that have a physical interpretation, and does so in a computationally efficient manner. Further, the HRP-based features effectively focus information so that a comparatively small set of such features capture all the information needed for recognition and a very simple algorithm may be used in the recognition process. Finally, the set of HRP features are local, and therefore, robust to localized silhouette variations and occlusion.

As is demonstrated in this paper, the HRP-based feature extraction and object recognition scheme is a highly competitive algorithm for this particular recognition problem. We demonstrate the strengths of the HRP-based recognition scheme by discriminating among 17 military aircraft. For comparison, we present recognition results on the airplane dataset using one of the more widely studied methods based on Fourier descriptors. The HRP-based algorithm matches the performance of the Fourier descriptors-based algorithm in the presence of boundary, scale and orientation variations, and surpasses the performance of the Fourier descriptor-based algorithm in the presence of occlusion and localized silhouette variations.

Acknowledgements

This research was conducted with support provided in part by AFOSR under grant F49620-95-1-0083, MURI under grant GC123913NGD, and DARPA under grant F49620-93-1-0604.

Appendix A. the hrp similarity measure and algorithm

The basic idea behind HRP is to decompose the function as the sum of the most significant elements from the dictionary set. At each iteration, the most significant element is defined to be the one which maximizes the HRP similarity measure, which has been designed to emphasize local fit accuracy. The HRP similarity measure between f and any dictionary element g_γ is defined to be

$$S(f, g_\gamma) = m(f, g_\gamma)s(f, g_\gamma) \tag{A1}$$

$$s(f, g_\gamma) = \min_{g_i \in I_\gamma(k)} \frac{|\langle f, g_i \rangle|}{|\langle g_i, g_\gamma \rangle|} \tag{A2}$$

$$m(f,g_\gamma) = \begin{cases} +1 \text{ if } \dfrac{\langle f, g_i \rangle}{\langle g_i, g_\gamma \rangle} > 0 \text{ for all } g_i \in I_\gamma(k), \\ -1 \text{ if } \dfrac{\langle f, g_i \rangle}{\langle g_i, g_\gamma \rangle} < 0 \text{ for all } g_i \in I_\gamma(k), \\ 0 \quad \text{otherwise,} \end{cases} \tag{A3}$$

where $I_\gamma(k)$ is as defined in Eq. (9). The first element chosen by HRP is denoted g_{γ_0} and is given by

$$g_{\gamma_0} = \arg\max_{g_\gamma} |S(f,g_\gamma)|. \tag{A4}$$

The function f is then decomposed as

$$f = S(f, g_{\gamma_0})g_{\gamma_0} + Rf, \tag{A5}$$

where Rf is the residual. Subsequent elements are chosen similarly to be the best fit to the previous residual.

That is,

$$g_{\gamma_n} = \arg\max_{g_\gamma} |S(R^n f, g_\gamma)| \tag{A6}$$

where $R^n f$ is the nth residual. This yields a cumulative decomposition of

$$f = \sum_{n=0}^{m-1} S(R^n f, g_{\gamma_n}) g_{\gamma_n} + R^m f. \tag{A7}$$

References

[1] C. Richard, H. Hemami, Identification of three-dimensional objects using Fourier descriptors of the boundary curve, IEEE Trans. Systems Man Cybernet. SMC-4 (4) (1974) 371–378.

[2] G.N. Bebis, G.M. Papadourakis, Object recognition using invariant object boundary representations and neural network models, Pattern Recognition 25 (1) (1992) 25–44.

[3] Gil Ettinger, Hierarchical object recognition using libraries of parameterized model sub-parts, Master's Thesis, MIT, 1987.

[4] S. Dubois, F. Glanz, An autoregressive model approach to two-dimensional shape classification, IEEE Trans. Pattern Analy. Mach. Intell. PAMI-8 (1986) 55–66.

[5] C.T. Zahn, R.Z. Roskies, Fourier descriptors for plane closed curves, IEEE Trans. Comput. C-21 (3) (1972) 269–281.

[6] C.C. Lin, R. Chellappa, Clasification of partial 2D shapes using fourier descriptors, IEEE Trans. Pattern Analy. Mach. Intell. PAMI-9 (5) (1987) 686–690.

[7] J. Gorman, R. Mitchell, F. Kuhl, Partial shape recognition using dynamic programming, IEEE Trans. Pattern Analy. Mach. Intell. 10 (2) (1988) 257–266.

[8] H. Kauppinen, T. Seppanen, M. Pietikainen, An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification, IEEE Trans. Pattern Analy. Mach. Intell. 17 (2) (1995) 201–207.

[9] S. Jaggi, Multiscale geometric feature extraction and object recognition, Ph.D. Thesis, Massachusetts Institute of Technology, 1997.

[10] S. Jaggi, C. Karl, S. Mallat, A. Willsky, High resolution for feature extraction, Appl. Comput. Harmonic Anal., 1996, submitted.

[11] S.O. Belkasim, M. Shridhar, M. Ahmadi, Pattern recognition with moment invariants: A comparative study and new results, Pattern Recognition 24 (1991) 1117–1138.

[12] H.-C. Liu, M.D. Srinath, Partial shape classification using contour matching in distance transformation, IEEE Trans. Pattern Analy. Mach. Intell. 12 (11) (1990) 1072–1079.

[13] D. Marr, E. Hildreth, Theory of edge detection, Proc. Roy. Soc. London (1980) 182–217 .

[14] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Analy. Mach. Intell. 8 (6) (1986) 679–698.

[15] R. Deriche, Fast algorithms for low-level vision, IEEE Trans. Pattern Analy. Mach. Intell. 12 (1) (1990) 78–87.

[16] P. Fua, Y. Leclerc, Model driven edge detection, Mach. Vision Appl. 3 (1990) 45–56.

[17] Z. Wu, R. Leahy, An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation, IEEE Trans. Pattern Analy. Mach. Intell. 15 (11) (1993) 1101–1113.

[18] A. Amini, T. Weymouth, R. Jain, Using dynamic programming for solving variational problems in vision, IEEE Trans. Pattern Analy. Mach. Intell. 12 (9) (1990) 855–867.

[19] Z. You, A.K. Jain, Performance evaluation of shape matching via chord length distribution, Comput. Vision Graphics and Image Process. 28, 1984.

[20] R. Duda, P. Hart, Pattern Classification and Scene Analysis, J. Wiley, New York, 1973.

[21] I. Daubechies, Ten Lectures on Wavelets. SIAM Philadelphia, 1992.

[22] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Signal Process. 1993.

[23] S. Chen, D. Donoho, Atomic decomposition by basis pursuit, Technical report, Statistics Dept., Stanford University, May 1995. Available via ftp at playfair.stanford.edu.

[24] A. Viterbi, Phase-locked loop dynamics in the presence of noise by Fokker–Plank techniques, Proc. IEEE (1963).

[25] H.V. Trees, Detection, Estimation and Modulation Theory. Wiley, New York, 1968.

[26] C. Bregler, M.S. Snakes, Technical Report IRC 1995-017, Interval Research Corporation, 1995. Also utilized accompanying software.

About the Author—SEEMA JAGGI received the S.M. and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge in 1992 and 1997, respectively. She received the B.S. in Electrical Engineering from Purdue University in 1990. Since her graduation from MIT, Dr Jaggi has been working for the Boston Consulting Group as a consultant in the area of high-technology.

About the Author—WILLIAM C. KARL recieved the Ph.D. degree in Electrical Engineering and Computer Science in 1991 from the Massachusetts Institute of Technology, Cambridge, where he also received the S.M., E.E., and S.B. degrees. He held the position of Staff Research Scientist with the Brown-Harvard-M.I.T. Center for Intelligent Control Systems and the M.I.T. Laboratory for Information and Decision Systems from 1992 to 1994. He joined the faculty of Boston University in 1995, where he is currently Assistant Professor of Electrical, Computer, and Systems Engineering. Science January 1996 he has also held a joint of appointment in the department of Biomedical Engineering. He is a guest editor of the Int'l J. Patt. Recog. and Artificial Intelligence special issue on "Processing, Analysis, and Understanding of MR Images of the Human Brain" He is also associated editor of the Transactions on Image Processing. Dr Karl's research interests are in the areas of multidimensional and multiscale signal and image processing, geometric estimation, detection, and medical signal and imge processing. Prof. Karl is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.

About the Author—STEPHANE MALLAT Mallat was born in Paris, France. He graduated from Ecole Polytechnique in 19984 and from Ecole Nationale Superieure des Telecommunications, Paris, in 1985. He received the Ph.D. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 1988.

In 1988, he joined the Computer Science Department of the Courant Institute of Mathematical Sciences at New York University, and became Associate Professor in 1993. In the fall 1994, he was a visiting Professor in the Electrical Engineering Department at MIT and in the spring 1994 in the Applied Mathematics Department at the University of Tel Aviv. Since 1995, he has been a Professor in the Applied Mathematics Department at Ecole Polytechnique, Paris, France. His research interests include computer vision, signal processing and diverse applications of wavelet transforms. He is the author of a recent book *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

Dr Mallat received the 1990 IEEE Signal Processing Society's paper award, the 1993 Alfred Sloan fellowhip in Mathematics, the 1997 Outstanding Achievement Award from the SPIE Optical Engineering Society, and the 1997 Blaise Pascal Prize in applied mathematics, from the French Academy of Sciences.

About the Author—ALAN S. WILLSKY received both the S.B. degree and the Ph.D. degree from the Massachusetts Institute of Technology in 1969 and 1973 respectively. He joined the M.I.T. faculty in 1973 and his present position is Professor of Electrical Engineering. From 1974 to 1981 Dr Willsky served as Assistant Director of the M.I.T. Laboratory for Information and Decesion Systems. He is also a founder and member of the board of directors of Alphatech, Inc. In 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. Dr Willsky has held visiting positions at Imperial College, London, L'Universiti de Paris-Sud, and the Institut de Recherche en Informatique et Systhmes Aliatoires in Rennes, France. He was program chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals including the IEEE Transactions on Automatic Control, has served as a mebmer of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society, was program chairman for the 1981 Bilateral Seminar on Control Systems held inthe People's Republic of China, and was special guest editor of the 1992 special issue of the IEEE Transactions on Information Theory on wavelet transforms and multiresolution signal analysis. Also in 1988 he was made a Distinguished Member of the IEEE Control Systems Society. Dr Willsky has given several plenary lectures at major scientific meetings including the 20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the SIAM Conf. on Applied Linear Algebra, 1991, the 1992 Inaugural Workshop for the National Centre for Robust and Adaptative Systems, Canberra, Australia, the 1993 IEEE Symposium on Image and Multidimensional Signal Processing, Cannes, France; and the 1997 SPIE Wavelets and Applications Symposium, San Diego.

Dr Willsky is the author of the research monograph Digital Signal Processing and Control and Estimation Theory and is co-author of the undergraduate text Signals and Systems. He was awareded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph. Dr Willsky's present research interests are in problems involving multidimensional and multiresolution estimation and imaging, statistical image and signal processing, image reconstruction, and computer vision.