

Automatic Recognition of ISAR Ship Images

S. MUSMAN
Naval Research Laboratory

D. KERR

C. BACHMANN
Naval Research Laboratory

Inverse synthetic aperture radar (ISAR) produces images of ships at sea which human operators can be trained to recognize. Because ISAR uses the ship's own varying angular motions (roll, pitch, and yaw) for cross-range resolution, the viewing aspect and cross-range scale factor are continually changing on time scales of a few seconds. This and other characteristics of ISAR imaging make the problem of automatic recognition of ISAR images quite distinct from the recognition of optical images. The nature of ISAR imaging of ships, and single-frame and multiple-frame techniques for segmentation, feature extraction, and classification are described. Results are shown which illustrate a capability for automatic recognition of ISAR ship imagery.

Manuscript received December 1, 1994; revised September 18, 1995.

IEEE Log No. T-AES/32/4/08008.

Authors' current addresses: S. Musman, Integrated Management Systems, Inc., Arlington, VA 22201; D. Kerr and C. Bachmann, Code 5362, Radar Division, Naval Research Laboratory, 4555 Overlook Ave. SW, Washington, DC 20375-5000.

U.S. Government work not protected by U.S. copyright.

0018-9251/96/\$5.00 © 1996 IEEE

I. INTRODUCTION

With microwave surveillance radar it is possible to detect and locate ships at sea in all weather, day or night, to the limits of the radio horizon. Surveillance radars do not have sufficient antenna angle resolution to allow recognition of detected ships; however, a technique known as inverse synthetic aperture radar (ISAR) can, while searchlighting the ship with its antenna beam, produce images in range and cross-range with sufficient resolution for recognition. When ISAR displays the constantly changing images of target ships to human operators for classification, the process requires trained operators to dwell for a relatively long time duration before an identification can be confirmed. ISAR produces a rapidly updating sequence of range-Doppler image frames. Range resolution gives one imaging dimension, and the resolution of differential Doppler caused by target rotation gives an orthogonal (cross-range) dimension. While the formation of ISAR imagery relies on target motion (described below), in many respects ISAR images are similar to optical presentations of the same target, but there are notable exceptions. In addition to intuitive recognition by an operator, classification of ISAR ship imagery is accomplished by a combination of three different techniques: mensuration, feature description, and shape correlation. In mensuration the locations of major target components (such as masts, superstructure breaks, cranes, weapons, etc.) along the bow-stern axis of the target are measured as a percentage of the range extent of the ship, and compared with the locations of the same components on candidate ship classes. Because ISAR imagery is not affected by parallax, these relative measurements are not affected by changes in target aspect. Feature description is the process of labeling the components of candidate targets by their degree of match to descriptive templates (such as the stern: straight, curved, or rounded; or a mast: pole, lattice, or solid). The operator compares the feature descriptions observed in an ISAR frame sequence with the feature descriptions of candidate targets for recognition. As a final confirmation of the result of the above methods of recognition, the human operator can perform a shape correlation on a single selected ISAR image frame, visually comparing the image to a wire-frame model of one or more candidate ships the operator thinks might match the unknown image. The wire-frame model is transformed by a computer to match the orientation of the target in the ISAR image. To facilitate this three-method approach to ship classification, computer-based decision aids have been developed which contain databases of ship characteristics. The automatic recognition approach described in this work emulates the computer-assisted human recognition process, and utilizes the databases constructed for the computer-based aids.

A limitation to the recognition of a large number of ships in a limited time is the performance of human image interpreters. They suffer from fatigue and require a finite amount of time to reach a decision. Thus automatic or semiautomatic classification methods are a means of overcoming operator limitations.

ISAR uses primarily the ship's own normal angular motions (roll, pitch, and yaw) to provide the aspect changes necessary for imaging. As a consequence the images change radically on time scales which are fractions of the ship's motion periods (on the order of 10 s). This variability adds a time dimension to the task of automating ISAR ship recognition. The following are some ways in which the dynamics of ISAR images affect recognition.

- 1) It may be necessary to wait as long as a ship-motion period for a favorable image aspect to appear.
- 2) The image projection-plane and cross-range (vertical) scale factor are continually changing with the ship motion.
- 3) Image frames separated by a fraction of a motion-period can reveal different target features, i.e., plan and elevation information. This allows three-dimensional information to be extracted directly from the two-dimensional images.
- 4) The rapid changing of the image means that it is not possible to improve the detectability of target features by the direct integration of successive frames by summing or averaging pixel values.
- 5) Since not all image frames are equally useful, and different target features are likely to be found in successive image frames, it may be necessary to process many frames for reliable recognition, greatly increasing the amount of pixel-processing computation required.

Apart from their ever-changing dynamics, ISAR images are quite suitable for the application of many conventional image processing techniques. The images contain relatively few pixels compared with optical images; and there is normally little background clutter, making the separation of the target from the clutter (called segmentation) relatively straightforward. Even so, due to the large dynamic range and specular nature of ISAR images, other standard processing techniques, such as edge detection and histogram-based segmentation techniques, are normally ineffective.

Since single frames of ISAR imagery may not contain all the information necessary for ship recognition, techniques have been developed for processing time sequences of ISAR image frames. The need for multiframe processing has led to the development of inexpensive processing techniques (in terms of memory and central processing unit (CPU) cycles, since they must be applied to each

frame in a sequence of ISAR images) to produce both frame-selection criteria and target-specific features. The results obtained from processing the multiple image frames are then examined to detect and locate consistent target features, which are in turn used to classify the target.

This work summarizes research by the U.S. Naval Research Laboratory's Airborne Radar Branch on basic methods for automatic ISAR ship-image recognition. Section II describes the nature of ISAR ship images, and presents some examples. Section III describes single-frame processing for segmentation, feature extraction and target modeling. Section IV describes an approach to multiple-frame feature extraction. Section V gives example of processing and classification results. Section VI presents some final observations.

II. ISAR SHIP IMAGING

In this paper, ISAR means target-stabilized range-Doppler imaging. A pulse-to-pulse phase-coherent radar is used whose range resolution (1 m to 3 m) is much smaller than the dimensions of the target. The antenna beam of the radar searchlights the target for the duration of imaging (usually longer than the ship's roll period). Amplitude and phase echo data are collected for each of a set of range cells which span the target, and for each radar pulse during the imaging interval. The cells are stabilized in range relative to the target so that one target point remains in the same range cell for at least the duration of the time needed to collect the radar data for a single image frame. One method of range stabilization is to tie the range-sampling grid to a prominent point range tracker. Range cells form the down-range dimension of the image. The Doppler spectrum of the pulse-to-pulse data in each range cell is computed to form a cross-range (Doppler) dimension of the ISAR image frame. For a point on a rotating object (rotation is required for ISAR imaging), the Doppler frequency shift is proportional to the distance from the center of rotation, measured perpendicular to the radar line-of-sight. Thus the Doppler dimension represents cross-range. The Doppler resolution in Hz is approximately the reciprocal of the coherent processing time (in s) used to form the spectrum. The cross-range scale factor (m/Hz) is proportional to the Doppler frequency resolution, but also depends on the instantaneous rotation rate of the target (which for ship targets is generally unknown). ISAR recognition techniques thus must be tolerant to an unknown Doppler-to-cross-range scale factor. It is desirable to provide an operator with a sequence of ISAR image frames which changes smoothly as the target ship's motion changes; however, the integration time for the required Doppler resolution is usually much longer than the frame update interval for a smooth movie-like presentation. The faster frame rate is achieved by

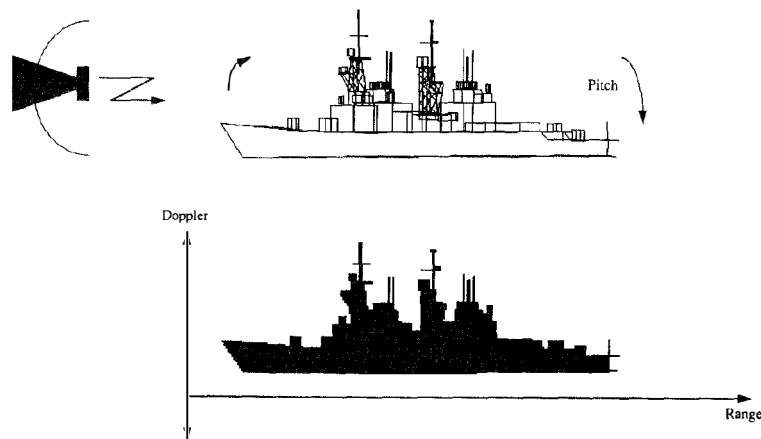


Fig. 1. Range/Doppler pitch effect.

computing the Doppler spectra for each frame on overlapping time segments of the coherent radar echo data.

The ISAR image projection plane is defined by the radar line-of-sight and the direction of rotation of the target. The image plane contains the line-of-sight, or range vector, which in this work is displayed as the image horizontal dimension. The other dimension of the projection plane can be described as the vector cross-product of the line-of-sight vector with the instantaneous angular velocity vector, that is, a vector perpendicular to both the range vector and the angular velocity vector.

Fig. 1 illustrates the connection between the target motion and the projection plane. In this example the ship is illuminated by the radar from the bow. It is subject to a pure pitch counter-clockwise rotation; and its angular velocity vector (by the right-hand rule) is perpendicular to the page. Therefore the vector cross-product is in the plane of the page, in the vertical direction. The pitch motion causes the tops of the masts to have a higher velocity, relative to the radar, than the deck, causing them to have greater Doppler shifts. This allows the vertical structure of the ship to be resolved in Doppler. If the ship pitches at a faster rate, the Doppler shifts of the masts will be greater, and the vertical image dimension enlarged. If the ship pitches in the clockwise direction, the Doppler shift will be in the opposite direction, and the ship image will be inverted. In like manner, roll motions (if the radar illumination aspect is other than parallel to the ship heading) also give ship elevation information, and yaw motions give plan information. The ISAR presentation of ship height structure, caused by roll and pitch, is known as *profiling*.

The dependence of the image projection plane and cross-range scale factor on the ship target motions causes the ISAR image dynamics. Roll and pitch are usually noncoupled oscillatory motions with varying periods on the order of 10 s, and yaw is a less

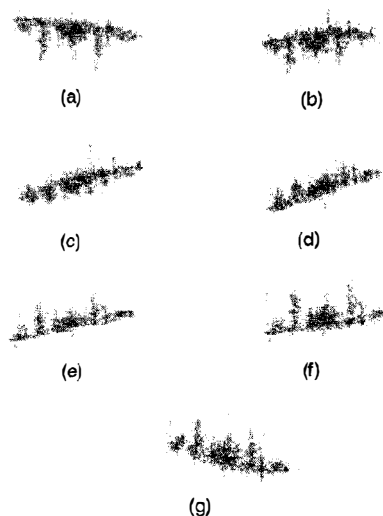


Fig. 2. ISAR images from image sequence.

structured motion depending on the steering of the ship and its interaction with the seaway. Generally, an ISAR image can be expected to change significantly in 1 s, and radically significant variations can be expected within 5 s.

Fig. 2 shows a typical sequence of ISAR images of a small (16,880 T) commercial ship. The radar used operates in X-band with 100 MHz bandwidth and a 2.5° antenna beamwidth. Successive frames are spaced in time by about 0.6 s. This image sequence is controlled by a rapidly changing roll motion, causing profiling; pitch, causing profiling and image tilt; and a slowly changing yaw motion, resulting in plan-view imaging.

1) Frame (a) contains an inverted image that is largely a vertical cut, or profile, caused by the roll and pitch motions of the ship. The vertical masts, king-posts, and central superstructure are visible. Some

pitch component is visible as a slight downward tilt of the image axis.

2) Frame (b) also shows profiling, but with more plan component.

3) In frame (c) the profile component is minimal, caused by a very small combined roll and pitch rate, but the plan information is pronounced. The apparent orientation or tilt of the image in the frame due to pitch has changed.

4) In frame (d) the profile information has begun to change sense, beginning to give an upright image; with the plan scale and tilt mostly unchanged. The sharpness of the focus is reduced, a result of nonuniform rotational motion.

5) Frame (e) has well-defined right-side-up profile information.

6) Frame (f) once again has sharply focused profile information.

7) Frame (g) shows significant plan component in addition to the profiling.

The time elapsed between frames (a) and (g) in Fig. 2 is 3.7 s. This figure illustrates the dynamics of ISAR imaging, requiring the use of special recognition techniques.

III. SINGLE FRAME PROCESSING AND FEATURE EXTRACTION

Since there is no *a priori* indication of classification utility for an ISAR image, each image frame in a time sequence must be processed. Some frames will be of better image quality than others, and some will contain more target classification information than others because of the target presentation. (In this paper, the term *presentation* refers to the appearance and orientation of the target in the ISAR image, while *aspect* is reserved for the physical orientation of the target to the radar line-of sight.) Thus the processing of each frame consists of *segmenting* the target from the rest of the image, and then extracting *features* and a *frame selection criterion*. Computing the frame selection criterion yields extra information about the ISAR image to indicate the reliability of the other features which have been extracted.

Although the techniques described are implemented to be applicable to multiframe processing, the last part of this section demonstrates the accuracy of the single-frame feature extraction process by using the features in classification. The locations of the extracted features are used to define ISAR transformations of wire-frame line drawings of candidate ships to make them fit the ISAR image. This demonstration is of a simple model-based single-frame classifier which both demonstrates the effectiveness of the feature extraction algorithms, and also mimics the shape correlation part of the human approach to ship classification.

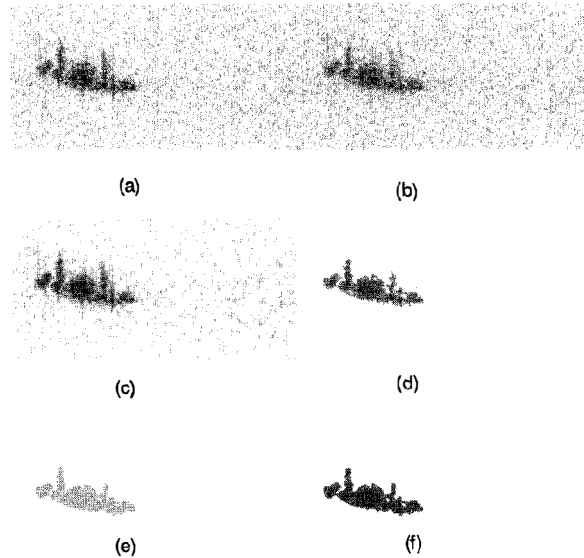


Fig. 3. Example of segmentation process. (a) Original image. (b) Unstreaked image. (c) Smoothed image. (d) Thresholded image. (e) Clustered image. (f) Segmented image.

A. Segmentation

The first step is extracting the target from the rest of the image (sea-clutter or noise). While it is commonly believed within the computer-vision community that recognition schemes relying on a robust segmentation process are generally less successful than other approaches, the characteristics of ISAR imagery make such a segmentation process reliable. Although it is not practical to create a perfect segmentation process for ISAR imagery (which would always produce perfect results and be scalable to a real-time implementation), the process described below performs well and is computationally amenable to a real-time implementation. Highlights of the basic segmentation approach for ISAR images, illustrated in Fig. 3, are as follows.

1) Select a threshold to remove the background noise and clutter from the original image (Fig. 3(a)), but do not apply it until later.

2) Detect and reduce vertical streaks in the image. These vertical streaks, only a few pixels wide, are object-related image features arising from either internal motion on the ship (e.g., moving machinery) or multiple-bounce scattering from the moving sea surface. Apply a low-pass filter to the image to remove noise spikes (Fig. 3(c)).

3) Apply the threshold found in Step 1 to the smoothed image (Fig. 3(d)).

4) Perform morphological region-growing and region-filling [1] to the remaining image regions.

5) Apply a geometric clustering algorithm to remove any remaining nontarget regions (Fig. 3(e)).

6) Overlay the clustering result on the original unprocessed image to obtain target intensity information (Fig. 3(f)).

The goal of the segmentation process is to leave only the target areas in the image. The idea behind the processing steps listed above is to find all image regions bright enough to be part of the target, and then to use a geometric clustering algorithm (described below) to eliminate any bright regions which are not located in image positions that are characteristic of the shape of a target. Although many processing steps are applied during the segmentation process, each step can be performed independently of the others. This allows for an inherently parallel set of pipelined processes, which could minimize the frame-to-frame processing time.

Most of the segmentation steps use standard image-processing operations [1]. The exceptions are the streak-elimination procedure and the clustering algorithm. *Streaks* are thin vertical image features caused by transient or rapid motions internal to a target, and thus are not focused in Doppler by the ISAR processing which assumes rigid-body rotation of the whole target. Streak elimination uses the threshold selected in Step 1 to count the number of "bright" (i.e., intensities above the threshold) cross-range pixels for each range cell in the image. Streaks are considered to be range cells with more than 66% of their Doppler pixels above this threshold. Each range cell containing a streak is then processed by comparing each Doppler (vertical) bin in the range cell to neighboring Doppler bins (in adjacent range cells) to find isolated bright image pixels surrounded by nonbright image pixels. If a pixel is bright and its neighbors are not, then it is considered to be a streak pixel and its intensity is reduced by 30%. The application of streak removal reduces the visibility of very thin vertical structures on the target; however, because most of the target image elements occupy several range cells, most of the target information is generally not adversely affected.

The geometric clustering algorithm is specifically designed for the expected shape of ship targets. Ship ISAR images tend to be long shapes oriented along the range dimension of the image. The application of the morphological region-growing and region-filling algorithms, done in Step 5 of the segmentation process, makes the resultant image input to the clustering algorithm consist of disjoint image regions. Each region contains image areas bright enough to be part of the target. The clustering algorithm performs region-labeling. Statistics such as size, boundaries, and centers of mass are collected for each region. The algorithm either retains or discards a labeled region based on its size and perpendicular distance to the estimated centerline of the target. Large regions which are close to the centerline are retained, while small regions further away from the center line are eliminated.

An important consideration for the segmentation algorithm is to keep bright superstructure features on the target well defined, while retaining the low intensity pixels which usually define the target end-points. In practice this is hard to achieve. Because of the very large dynamic range of image intensities of ship components, the intensity of image processing artifacts (Doppler sidelobes and focus errors) around the superstructure regions is similar to the actual target intensity near the ends of the ship. Thus, the bright features can be sharpened by increasing the threshold and excluding artifacts, but at the expense of eliminating the target end-points; while extraction of the end-point and target extremity information can be achieved with a lower threshold, but fine details associated with the target superstructure may be lost.

The threshold selection is a one-pass, rather than an iterative, algorithm. While a noniterative approach may not perform as well as an iterative method, it produces satisfactory results and requires much less computation. A threshold is chosen by examining the statistics of an area of the image background which contains no target energy. It is an intensity level above the noise and processing artifact floor of the image background. Histogram-based segmentation techniques (which look for bimodal characteristics in the intensity histogram) are unsuitable because the background energy in the image has a Rayleigh-like distribution, and the target intensity information is only slightly above the background floor.

The intensity-based processing (Steps 1-4) is sometimes sufficient for segmenting an image. Usually however, other nontarget image areas have exactly the same intensity characteristics as the target. To eliminate spurious bright image pixels, the geometric clustering algorithm is used. Before this operation is performed, some preprocessing is performed (Step 5) to fill in the small gaps between close but nontouching image regions.

B. Feature Extraction and Projection Plane Determination

In order for a human operator to accomplish the shape-correlation stage of classification described in the Introduction, the projection plane of the image must be determined. An operator, using a computer-based interpretation aid, does this by designating the apparent positions in the image of the bow, the stern, and the height of the bridge on the ship. While this is a simple task for a human operator, it requires extensive processing to accomplish automatically. In this Subsection, the feature extraction of projection plane parameters is described, and Subsection C describes the performance of an automatic shape-correlation classifier module. The projection-plane parameters define the physical projection of a 3-dimensional target model into the

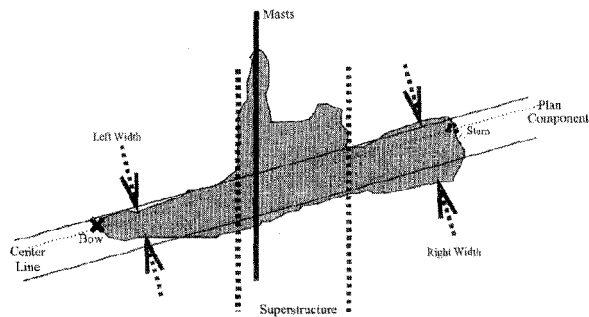


Fig. 4. Features automatically extracted to deduce ISAR projection parameters.

2-dimensional ISAR image plane. This allows the automatic warping of target wire-frame models to match the orientation of the target within the image frame.

The features extracted to deduce the orientation of the target include the following.

- 1) Centerline.
- 2) End-points (bow and stern).
- 3) Plan component (how much plan information is present).
- 4) Profile component (how much profile information is present).
- 5) Width of the plan component at either end (at bow and stern).
- 6) Width of the outline at either end.
- 7) Superstructure breaks and extent.
- 8) Major uprights (mast and crane locations).

Fig. 4 is a sketch of an image showing these features. In extracting them, a large portion of the processing time is spent in making a consistently robust estimation of the centerline of the target (described below). After the centerline has been found, much of the information used to extract it also can be used in the extraction of the other target features. In the interest of processing speed, only approximations to the actual features are found. For example, the real centerline of the target is the line connecting the point of the bow through the center of the stern. These points are often not identifiable in the imagery and are only obtainable by inference from other target features, using prior knowledge of symmetry and expected target shape. Thus, the estimated centerline obtained from the automatic processing does not always coincide with the real centerline of the target.

Since ships in ISAR images are long thin shapes, the centerline of a target can be estimated by searching for the maximum peak of the Hough transform [1] of a partially segmented image. This technique differs from that of Drazovich [3] who used a least-square-fit line through the target to estimate the centerline. The least-squares method is less effective than the Hough transform because the amount of target height profile

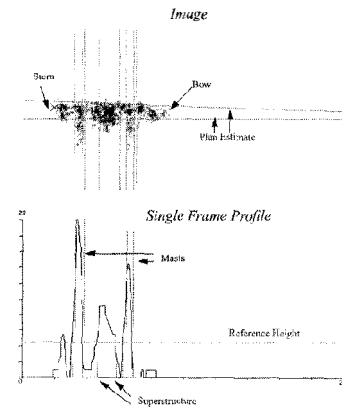


Fig. 5. Single frame processing of ISAR image.

in the imagery will tend to bias the least-squares line-fit towards the side of the image (top or bottom) on which the height features appear.

The plan component is estimated by fitting straight lines to each deck edge by searching the Hough transform of the segmented image boundary for peaks at an angle parallel to the angle of the centerline. Most ships have long smoothly curving sides (known as a "fair curve"), so approximation of the deck edges as straight lines is a reasonable procedure. The plan component is estimated by the Doppler separation of the two deck edges. Estimation of the plan component is an important part of the automatic ISAR processing used to differentiate between height and width information.

The remaining features are then extracted. Target end-points are obtained by examining image intensities near the centerline, and the width estimates are found by examining values in the Hough transform at angles 90 deg from the centerline. By examining histograms of the number of target pixels extending in the Doppler direction beyond either side of the deck edges for each range cell, it is possible to determine the direction of any target height profile (right-side-up or inverted) within the image frame. If there is enough height profile, a peak-decomposition procedure [7] can be applied to the profile histogram to find the peaks likely to be the uprights (mast locations) and superstructure blocks. The locations of such features can be used to classify a ship target (by mensuration) when they are made invariant to the ship's aspect by measuring their percentage distance along the ship's length from the bow to the stern. Fig. 5 illustrates the feature extraction described above when applied to an ISAR image.

A determination of which end of the target corresponds to the bow is based on the width measurements, superstructure locations, and the upright locations. A single decision is made for an image frame when features indicate unambiguously which end of the target is the bow; but in cases where

the features give an inconclusive determination, it is necessary to pursue two separate hypothesis for each target, one with the bow on the left and one with the bow on the right.

C. Classification by Shape Correlation

A single-frame model-based classification module was developed for good quality ISAR images which correlated them with projections of wire-frame models from a 50-ship database. For each model, the set of aspect and motion parameters is computed from the features extracted, and these are used to geometrically transform the wire-frame model into the same 2-D projection as the image. The aspect is determined from the range-extent measured between the bow and stern features of the image and the length of the model ship. An estimate of the yaw rate is calculated from the Doppler difference between the bow and stern, and the length of the ship. An effective roll and pitch rate, which arbitrarily apportions the vertical velocity between roll and pitch, is calculated from the Doppler extent beyond the deck edges of the height features. It is not possible to calculate the real roll and pitch values using a single image frame because the equations represent an underdetermined system.

Using the aspect and motion parameters, the wire-frame models are rotated and scaled in the Doppler dimension to match the orientation of the target in the image; and are then correlated with the shape of the target (as defined by the segmented version of the original image). The shape of the target is compared with the shape of the wire-frame model by solidifying the wire-frame model, and then masking the segmented image over the model to count the number of pixels that show through. This procedure is then repeated by masking model over the image. A good match is indicated by low pixel counts for both masking operations. Ranking the results of the correlation demonstrates the classification potential of model-based predictions and also shows the effectiveness of the single-frame feature extraction process. If the single-frame feature extraction process is successful, then the correct target should be ranked near the top of the correlation list, and should be appropriately scaled and rotated to allow an easy confirmation of the choice by operator inspection. Fig. 6 illustrates the best ranked result of automatically fitting a wire-frame model to a selected ISAR image.

Wire-frame models were utilized because they are easy to obtain and they are computationally cheap to use. Further application of model-based information for realistic interpretation problems would require target models which include shadowing and radar intensity estimates.

Despite the limitations of the wire-frame models, they show the benefits of using a model-based

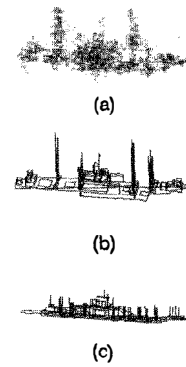


Fig. 6. Automatic shape correlation. (a) ISAR image. (b) Best matching overlay. (c) Second best matching overlay.

approach, having a fundamental advantage over data-driven classification approaches which would require data samples of each target at all aspects and orientations before it could be added to a data-driven classifier.

IV. MULTI-FRAME PROCESSING

The temporal characteristics of ISAR images make classification decisions based on single image frames unreliable. Classification should be based on information obtained from multiple image frames in a time series of ISAR images. Thus single-frame image processing should be as simple as possible to minimize processing expense. The processing of single image frames, by the methods described in this work, does not use iterative algorithms. The complete processing (including segmentation) of a single frame takes under 10 s on a uni-processor workstation, even using many LISP functions and unoptimized code. It should be possible to achieve real-time processing rates using specialized hardware. The speed of processing is achieved at the expense of a tradeoff in algorithm effectiveness and consistency. Because of the compromises, the algorithms will not always return results for each single frame of the quality shown in Fig. 6. To increase the performance of the feature extractors it would be necessary either to increase the amount of preprocessing applied to the data before the algorithms are applied, or to devise entirely new feature extraction algorithms to work with the raw data. This might involve a many-fold increase in processing time. The goal of this processing was to achieve a high feature-detection performance, while maintaining a very low false alarm rate.

To achieve multiframe classification, feature extraction processing (which we call feature detection) and frame selection are performed on each of many individual ISAR image frames; and then frame selection criteria and multiframe "temporal" tests are applied to distinguish between real and false feature

detections. This means that if the feature detectors perform at low false alarm rates, only real target features will be consistently present over the temporal history that is built up as a sequence of images is processed.

The use of multiple-frame processing has allowed the use of single-frame processing results even though they are not perfect. To clarify any ambiguity that exists, for the remainder of this work the feature extraction process that is performed for each individual image frame is referred to as *feature detection*, while the multiple-frame feature extraction process (which utilizes the frame selection and temporal relations) is called *feature extraction*.

A. Frame Selection

Since detection of the features discussed earlier relies on height profile information, a selection criterion is applied to the information extracted from each frame to determine whether the desired profile component is present. As part of the feature detection process, both an estimate of the plan component, and the height (which may also include a superimposed plan component) of the profile are extracted. For best extraction of the height features, the target presentation should be mostly profile with little or no plan component. The amount of plan component is easily measured as the separation between the two straight-line approximations of the ship sides.

The amount of height profile is obtained from a histogram of the number of target pixels per range cell on the side of the target with the most pixels outside the plan component (see Fig. 5: Single Frame Profile). Indications of the amount of profile present can be obtained by a variety of measures: the standard deviation of the profile histogram; the number of pixels in the histogram; the maximum height of the histogram; the median height of the histogram; and the ratio of the number of target pixels found either side of the plan component. Furthermore, a significant number of pixels outside the plan boundary on the side opposite to the height profile side may indicate a frame unsuitable for reliable feature extraction. In devising a frame selection criterion from the feature extraction algorithms, it is important to obtain an absolute estimate of the profile height, rather than a relative measurement compared with the minimum or maximum values observed during a sequence. This is because some complete sequences of ISAR imagery may contain few or no profile images. When this happens, the resulting confidences in profile feature detections are much lower than for sequences with a larger number of good quality profiles. The criteria described below were devised after an evaluation of the effectiveness of the feature detection algorithms for many different image sequences.

1) The plan-side pixel count outside the plan boundary opposite the profile side should be less than a threshold (200 pixels for our imagery).

2) Plan width must be less than a threshold (10 Doppler cells for our imagery).

3) The median height of the profile histogram must exceed a threshold (greater than 5 Doppler cells for our imagery).

4) The profile histogram standard deviation should exceed another threshold (greater than 4.0 for our imagery).

An important characteristic of this frame selection technique is that it is based on both a plan and a profile measure. Since the ISAR presentation of a target is dependent on the combination of its roll, pitch, and yaw motions, the amounts of plan view and profile view can vary independently of each other. Thus the image in a sequence containing the most profile may also have a lot of plan component. Since the appearance of plan component features can affect the extraction of profile features, it is important to be able to detect the presence of a plan component and take into account its effect on a given profile feature detector. Techniques which measure the amount of profile only by measuring the Doppler extent are inadequate because two images with exactly the same amount of profile (caused by roll and pitch motion) may contain radically differing amounts of plan component (caused by yaw motion).

B. Multiframe Feature Extraction

Because of the time-varying characteristics and quality of the imagery, many image frames are processed before making a determination of a feature detection (approximately 30 profile view frames). Feature detections are made independently for each image frame, and the consistency of location in the range dimension is used to select the features for further classification processing. Figs. 7 and 8 show the processing results accumulated over many images of a sequence: the first 22 frames for Fig. 7, and 99 frames for Fig. 8. The figures show both a single image frame and a series of histograms of feature locations in range. In each figure the image represents the current image frame being examined with the detected features indicated as follows.

1) The image at the top left of each figure is a single ISAR image frame with the end-points, plan component, superstructure breaks, and upright locations, as determined by our feature detectors, marked.

2) The "Single Frame Profile" shows the height in Doppler cross-range resolution cells of the profile component (versus range) of the target in the illustrated image, and shows how features have been detected from the peaks of the profile histogram.

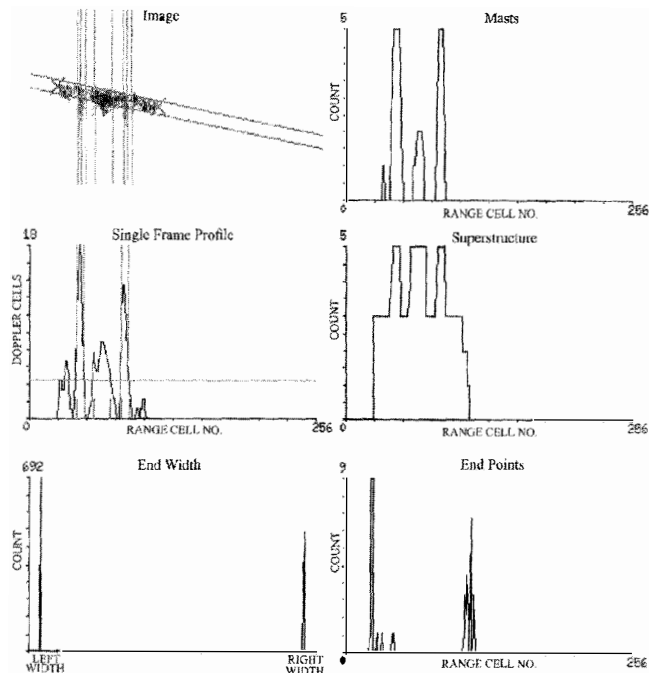


Fig. 7. Results of processing first 22 images.

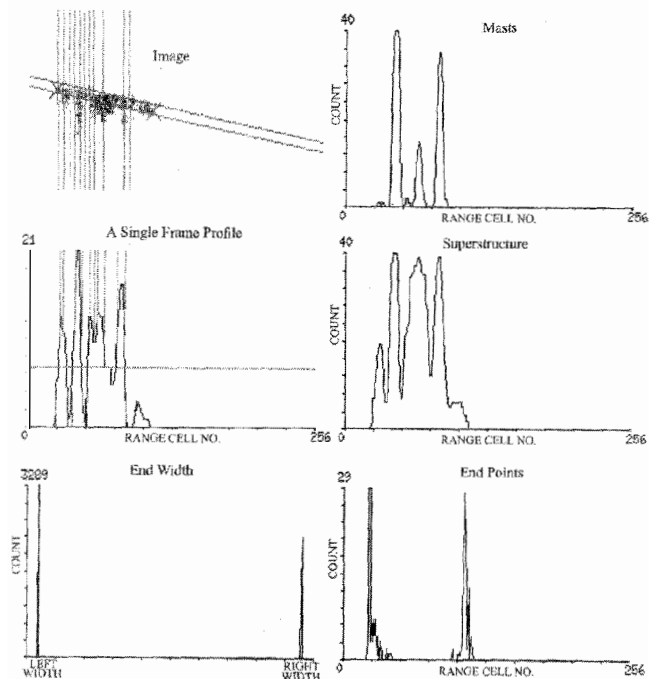


Fig. 8. Results of processing 99 images.

3) The other four plots in each figure are histograms of the cumulative feature detections over the image sequence.

4) The plot at the lower left of Figs. 7 and 8 shows the sums over all processed frames of the left-width and the right-width as defined in Fig. 4. The

larger of the two often indicates the stern end of the ship.

The three histograms on the right-hand side on each of Figs. 7 and 8 display the number of detections of each feature for each range bin in the image, accumulated over the sequence of images.

The histogram at the lower right of the figures showing the end-point detections gives an indication of how much the location of the detected end-point of the target can vary from frame to frame. Although it is necessary to measure the locations of other features as percentages of the number of range cells between the bow and stern, it is not possible to convert the location of each detected feature into a percentage for each individual frame because of the variability in the single-frame end-point locations. Instead it is necessary to use the most frequent end-point range positions over many frames to determine the bow/stern reference locations.

The use of feature detections which are consistent in range depends on certain assumptions: 1) the azimuth aspect of the target does not change significantly during the time of observation and, 2) the image is stabilized in range throughout the image sequence and does not drift. In practice these assumptions often hold, but not always. To compensate for range drift, it is necessary to "range register" each frame. The particular image frame from which the feature was detected must pass the frame-selection height profile criterion before a feature detection is added to the multiframe histograms. Only one frame selection criterion (described above) is used for all of the features shown. Improved performance might be obtained by using different frame selection criteria for each feature detector, based only on the consistency and performance of that particular detector.

As Fig. 8 shows, after a fair number of profile (40 out of 99) frames have been found, the upright and superstructure locations can be determined easily. The detection algorithms and the multiframe extraction have not been tested sufficiently to determine confidence levels for each detector. However, the example histograms indicate that the determination of binary (feature or not-feature) decisions for each range cell is easy when enough profile images have been processed. After further testing of the feature extraction algorithms, it should be possible to partition the number of detections into confidence levels. It has not been necessary to use sophisticated statistical techniques for these decisions about feature extraction; and this has been taken as an indication of the reliability of the methods.

V. MULTIFRAME CLASSIFICATION

A. Multiframe Procedure

In this research, classification techniques have been kept simple. The focus has remained on the feature extraction process. The classification algorithm used is a continuation of the ship-recognition work developed by Booker and Hota [2] and more recently Musman, Chang, and Booker [6]. It utilizes the database of a decision-aid developed at the Naval Research

Laboratory (NRL) to assist human interpreters when classifying ISAR images. The basic classification operation of the demonstration system utilizes Bayesian belief networks [2] to combine subjective uncertainty estimates associated with the automatically extracted target features. These features, described in the previous two sections, are the following.

- 1) target range extent (a minimum length for the observed target),
- 2) locations of superstructure breaks (all target locations where the superstructure noticeably changes height) given as a percentage of the range extent of the image,
- 3) percentage of the target range extent covered by superstructure,
- 4) the number of major uprights on the target (these are normally the masts),
- 5) upright locations as a percentage of the range extent,
- 6) mast/reference ratio (the ratio of the highest mast compared with the median height of the profile, needed to compute the target projection parameters),
- 7) profile shape correlation (a correlation of an aggregation of the 5 best profiles with a stored characteristic profile shape for each possible ship class).

The classifier utilizes the feature locations (as percentages of the distances between bow and stern) stored in the decision-aid database and, with the exception of profile shape correlation, it measures the difference between each automatically extracted location and the corresponding one for each target class in the database. It allocates an error "point" for each 1% that the observed feature location differs from one in the database. Thus an exact percentage match gives 0 points, a difference of 3% off gives 3 points, etc.

The upright locations are matched with an error score of up to 10 "points" (10% error, relative to the length of the ship), and 0, 5 or 10 "points" are allocated for the number of uprights (depending on whether the number detected equals, differs by 1, or differs by 2 from the actual number of uprights for the target in the database). The error measurements and penalties are then turned into subjective likelihoods, which represent how well the given error value may match a known target. The likelihood values were developed by extrapolating the results of testing the feature extraction algorithms on a limited set of test data, generalizing those results to produce the subjective likelihood estimates. For each feature and each target, the values represent the likelihood that the observation matches a specific type of target as compared with matching random noise. The advantage of this approach is that the classification results yield an intuitive indication of how well the observed features match each specific target: if the results are close

Automatically Extracted Features

7 Best Classification Scores

- 1) Range = 44.5 ft.
- 2) Masts at 30% and 74%.
- 3) Superstructure Breaks at 40% and 63%.
- 4) Reference Ratio = 3.51.

- 1) Com/Aux 5 = 0.918
- 2) Com/Aux 3 = 0.643
- 3) Lg Cbt 8 = 0.642
- 4) Com/Aux 4 = 0.573
- 5) Com 3 = 0.357
- 6) LP 2 = 0.0
- 7) Med Cbt 6 = 0.0

Comparing Automatic Features vs Top 3 Ships

<i>Com/Aux 5 : (Correct Ship)</i>	<i>Com/Aux 3 :</i>	<i>Lg Cbt 8 :</i>
Masts at (766.3 32 32) gives a 4% error	Masts at (45 65 72) gives a 17% error	Masts at (38 54 65) gives a 17% error
Superstructure at (5 17 23 31) 39 62 89 95) gives a 2% error	Superstructure at (37 41 51) 58 68 71 73) gives a 1% error	Superstructure at (24 33 44) 58 64 72 78) gives a 5% error
Superstructure width gives a 3% error	Superstructure width gives a 4% error	Superstructure width gives a 22% error
Mast/Superstructure Ratio is 3.58 gives a 0.07 error.	Mast/Superstructure Ratio is 2.36 and gives a 1.15 error.	Mast/Superstructure Ratio is 3.33 and gives 0.18 pts.
Profile Shape gives 13 pts.	Profile Shape gives 2 pts.	Profile Shape gives 12 pts.

Fig. 9. Classification results for Fig. 8.

to 1.0, then the observed features exactly match the target; if the results are 0.5, then the observed features are just as likely to be random noise, as opposed to a specific target; and if the results are 0.0, then the observed features are believed not to match the specific target.

The mast/reference ratio is the height of the highest mast compared with the median height of the profile; and 1 error "point" is scored for each 0.2 difference between the results of the image sequence and an entry in the database, once again limited to 10 "points" for the worst possible match. This score represents the design characteristic of ships which utilizes the fact that some ships have very tall masts, and some have much shorter masts.

The 1-D profile height correlation is a square-error correlation of the average of the 5 best height-normalized Doppler profiles in the image sequence (as determined by the frame selection criterion) compared with a characteristic target height profile for each possible target class stored in the decision-aid database. This score is ranked from 0 "points" for an exact match down to 50 as the match gets worse. In practice no ship ever gets a perfect score for the profile correlation; and while the correct ship may not be the best ranked ship, it is usually ranked in the top 10 in a 50-class problem. This profile correlation is distinctly different from the 2D-shape correlation reported in Section III; but is designed to achieve the same basic results using only one-dimensional shape (Doppler extent) information.

B. Multiframe Results and Discussion

Fig. 9 is an example of the classifier's application to the ISAR image sequence of Fig. 2. Despite its rather simplistic approximation of uncertainty measurements and the small number of features used, the correct ship is ranked at the top of the final classification list. The

results of similar testing on images of many other targets confirms the validity of the result shown.

The features used to obtain the classification scores are compared with feature locations from the operator-aid database. This contains locations that a human interpreter would be likely to detect. Although the automatic detectors are designed to detect these same features, each automatic detector has its own characteristic way of recognizing a feature. Thus, the automatically extracted superstructure breaks are not exactly the same as those that would be extracted by a human interpreter. If a database of feature locations were created from a characteristic evaluation of each automatic detector, even better classification results would be achieved.

The classifier described here has used only a small number of the potential features that can be extracted from the ISAR imagery. Others might include: bow curvature, stern shape, mast type, and superstructure shape. The advantage and appeal of this approach to automating the ISAR ship classification process is that the features extracted from the imagery are the same ones as those used by human interpreters. This makes verification of the performance of the feature extraction algorithms very easy.

This work has concentrated on the development of cheap, robust algorithms which can be tested on representative samples of data. Testing has demonstrated that the multiframe processing techniques are reliable enough to permit the automatic selection of good quality image frames.

The estimation of a target's ISAR projection plane described in Section III is an important prerequisite to the approach of Miltonberger, et al. [5], in which a multihypothesis decision space is used to refine successive estimates of target aspects for each possible target. In the work described here, no attempts have been made to use such an approach. The use of a multihypothesis decision space is difficult to extend into the multiframe domain, and can only realistically be used on single frames after an evaluation of the quality of the image frames has been made. The approach described here specifically attempts to address this image quality issue by developing a robust frame selection measure.

VI. DISCUSSION

An approach for automatic ISAR ship recognition has been described. It has been shown how the processing of single ISAR image frames is limited by the time variability of the ISAR imagery. To combat this, multiframe processing appropriate to ISAR's temporal characteristics has been devised. The techniques are computationally simple enough to allow extensive testing of the algorithms on general purpose computers. This work is an example of a bottom up (extract features and then classify) process,

learning enough from the image sequences to reduce significantly the number of candidate ships. This allows a subsequent analysis of a few single frames, automatically selected and processed in greater detail. Processing these selected image frames can then utilize the time history information obtained from the whole image sequence, but then utilize more computationally expensive feature extraction techniques and perhaps even model-based predictions computed from complex models capable of predicting shadowing and radar scattering intensities.

ACKNOWLEDGMENT

The body of work summarized in this paper owes a great deal to the pioneering work of David L. Drake when he was an NRL employee, and subsequently as a consultant to NRL. He has had an essential role in developing the classification of ships from ISAR images. He first insisted that ISAR images are classifiable, and then demonstrated this with controlled human interpreter tests. He was a continual contributor and consultant to the research reported here.

REFERENCES

- [1] Ballard, D., and Brown, C. (1982) *Computer Vision*. Englewood Cliffs, NJ: Prentice Hall, 1982, 102-105.
- [2] Booker, L. B., and Hota, N. (1986) Probabilistic reasoning about ship images. In *Proceedings of the AAAI-86 Workshop on Uncertainty in AI*, Seattle, WA, Aug. 1986.
- [3] Drazovich, R. T. (1980) The feasibility of artificial intelligence techniques for radar target classification. Final Report, AI&DS, Mountain View, CA, 1980.
- [4] Lynn, P. A. (1982) *An Introduction to the Analysis and Processing of Signals*. London: Macmillan, 1982, 229-232.
- [5] Miltonberger T., Morgan D., and Orr, G. (1988) Multi-sensor object recognition from 3D models. Presented at *SPIE Symposium on Optical and Optoelectronic Engineering*, Cambridge, MA, Nov. 1988.
- [6] Musman, S. A., Chang, L. W., and Booker, L. B. (1990) A real-time control strategy for Bayesian belief networks with application to ship classification problem solving. Presented at the Tools for AI Conference, IEEE Computer Society, Washington, DC, Nov. 1990.
- [7] Sankar, P. V., and Rosenfeld, A. (1979) Hierarchical representation of waveforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1, 1 (Jan. 1979), 73-80.
- [8] Ullman, S. (1987) Analysis of visual motion by biological and computer systems. In M. Fischler and ● Firschein (Eds.), *Readings in Computer Vision*. Los Altos, CA: Morgan Kaufmann, 1987.



Scott Musman was born in New York City in 1963. He received a B.Sc. in electrical engineering and computer science from the University of Sussex, England, in 1984. He obtained an M.S. in computer science from The Johns Hopkins University, Baltimore, MD, in 1988.

In 1985 he joined the Naval Research Laboratory and began working in the area of radar image understanding, target modelling and human decision aids. In late 1992 he joined the MITRE Corporation. Since joining MITRE he has continued researching issues associated with spatial and temporal reasoning, with applications to real-time military sensor and weapons application. In late 1995 he started the Intelligent Systems Division of IMSI. He continues to work in the area of multiframe image understanding, classification systems and resource allocation.



David W. Kerr received the BSEE degree from Texas College of Arts and Industries, Kingsville, TX, and the MSEE and Ph.D. from The Johns Hopkins University, Baltimore, MD.

He is Imaging Radar Coordinator for the Airborne Radar Branch of the Radar Division at the Naval Research Laboratory in Washington, D.C. He has been working with ISAR for ship recognition since 1976.



Charles M. Bachmann received the Ph.D. in 1990 and the Sc.M. in 1986, both in physics from Brown University, Providence, RI, and the A.B. in 1984 in physics from Princeton University, Princeton, NJ.

While completing his Ph.D., he was involved in interdisciplinary research in the Brown University Center for Neural Science. He joined the Airborne Radar Branch at the Naval Research Laboratory as a Research Physicist in 1990. Since 1994, he has been a Supervisory Research Physicist and Head of the Target Classification Section. His research interests include machine learning, neural network algorithms, projection pursuit, wavelets, and the application of these techniques to automatic classification of remote sensing imagery, especially radar and multispectral visible and infra-red imagery.