Machine Intelligent Gust Front Detection

Richard L. Delanoy and Seth W. Troxel

Techniques of low-level machine intelligence, originally developed at Lincoln Laboratory to recognize military ground vehicles obscured by camouflage and foliage, are being used to detect gust fronts in Doppler weather radar imagery. This Machine Intelligent Gust Front Algorithm (MIGFA) is part of a suite of hazardous-weather-detection functions being developed under contract with the Federal Aviation Administration. Initially developed for use with the latest-generation Airport Surveillance Radar equipped with a wind shear processor (ASR-9 WSP), MIGFA was deployed for operational testing in Orlando, Florida, during the summer of 1992. MIGFA has demonstrated levels of detection performance that have not only markedly exceeded the capabilities of existing gust front algorithms, but are competitive with human interpreters.

Gust fronts generated by thunderstorms can seriously affect the safety and efficiency of airport operations. Lincoln Laboratory, under contract with the Federal Aviation Administration (FAA), has had a significant role in the development of two Doppler radar systems that are capable of detecting low-altitude wind shears, including gust fronts, in the airport terminal control area. These systems are the Terminal Doppler Weather Radar (TDWR) and the latest-generation Airport Surveillance Radar enhanced with a wind shear processor (ASR-9 WSP).

By examining images generated by these radars, experienced human observers can reliably detect and track gust fronts. But the development of automated gust front detection algorithms having sufficiently high detection rates with few false alarms has been elusive. The gap between human and computer performance is due to several limitations of the detection algorithms. These limitations include the lack of means for handling and maintaining weak, ambiguous, and contradictory evidence, the use of multiple sequentially applied thresholds for object discrimination (such thresholds can inadvertently result in the discarding of important data), a failure to use all of the relevant information available in the input data, and the ineffective use of knowledge regarding the behavior or appearance of gust fronts under different circumstances.

Given clear, unambiguous radar gust front signatures, existing detection algorithms perform reasonably well. The challenge is in constructing algorithms that can handle the marginally detectable ambiguous cases. In such cases, various factors must be considered. For example, gust fronts can be obscured by large areas of precipitation, or gust front signatures can disappear in Doppler velocity images whenever the Doppler viewing angle is perpendicular to the direction of motion. Furthermore, gust fronts can be mimicked by other natural phenomena, such as flocks of birds, clouds of dust stirred up at construction sites, low-intensity rain, and ground clutter. And gust fronts can have very low radar cross-section densities, sometimes below the sensitivity of the radar system.

The preceding paragraph should sound familiar to those involved in the development of automatic target recognition (ATR) systems, for the issues are basically the same. In addition to the continual trade-off

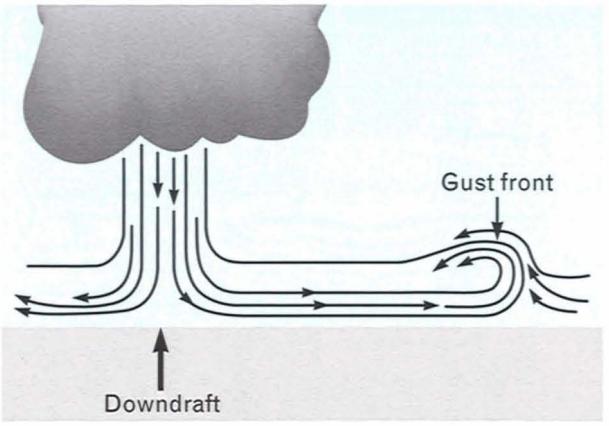


FIGURE 1. Thunderstorm downdraft and resulting gust front. The cool outflow beneath a thunderstorm spreads out in all directions. The leading edge, where the cool outflow and the warmer ambient air converge, is called the gust front.

between detection rates and numbers of false alarms, the issues for gust front detection are

1. obscuration and camouflage,
2. sensor limitations,
3. clutter and decoys, and
4. stealth.

Not surprisingly, the overall design of existing gust front detection algorithms is similar to that of most ATR systems. This traditional design is characterized by a hierarchy of modules, typically called detection, extraction (or discrimination), and classification. The detection process is essentially the application of some threshold that has been chosen to maximize the probability of detection at some acceptable level of false detections. Where signals are found that are above threshold, features are extracted, producing an abstraction, or symbolic representation, of the raw data. Given the set of extracted features, a signal is then classified as either one of the object types being sought or as clutter. In both the existing gust front detection algorithms and the traditional ATR systems, detection is generally unsophisticated: the threshold is applied either to raw radar data or to a simple transformation (such as a matched filtering) of the raw data. Sophisticated machine intelligence techniques are generally applied in the form of classifiers, e.g., by the use of neural network, statistical, or model-based classifiers.

However, the use of machine intelligence only for the classification process leads to a problem. With the application of a detection threshold, a significant amount of information is discarded, including those object signatures which are weak or ambiguous. Our belief is that increased detection reliability can be achieved by applying machine intelligence techniques prior to the application of detection thresholds.

A framework for applying machine intelligence techniques at the earliest levels of signal (image) processing is provided by the Experimental Target Recognition System (XTRS) [1], a general-purpose machine intelligence approach to ATR developed at Lincoln Laboratory. Specific techniques of knowledge-based signal processing, fuzzy set theory, and pixel-level maps of spatial evidence are all part of this approach. Based on XTRS, a Machine Intelligent Gust Front Algorithm (MIGFA) has been constructed for use with both TDWR and ASR-9 WSP imagery. Of the two radar systems, the ASR-9 presents the greatest challenge to gust front detection because of its lower sensitivity and less reliable Doppler measurements in clear air. Thus, this article will focus on the ASR-9 WSP version of MIGFA to demonstrate best the algorithm's effectiveness.

Gust Fronts

An intense thunderstorm downdraft can arise from various processes such as evaporative cooling and frictional drag between water droplets and the air. Upon impact with the ground, the downdraft is deflected horizontally (Figure 1), producing a local region of divergent winds. The downdraft feeds an outflow of outwardly expanding cool air. At the leading edge of the outflow exists a boundary where cool outflow air collides (converges) with the warmer ambient air. This leading-edge boundary, called a gust front, can grow to be many kilometers long and can propagate far away from the generating storm.

The turbulence within a gust front can be severe enough to present a danger to aircraft during takeoff and landing. And, because the prevailing winds behind a gust front can persist for a long time, the passage of a gust front over an airport often necessitates a change of active runway. When unanticipated, a gust front can delay airport operations as aircraft are

rerouted to a different runway. Aside from issues of cost and inconvenience, delays can increase the risk of potentially fatal human errors as the distance between aircraft that are taking off or landing decreases and the work load on air traffic controllers increases. With sufficient warning, though, controllers can incorporate in their plans a change in active runway at the anticipated time of a gust front's arrival, thereby minimizing the hazards and costs associated with delays.

Gust fronts can be detected in Doppler radar imagery on the basis of three physical properties: velocity convergence, thin lines, and motion. Figure 2 shows a typical gust front in both TDWR and ASR-9 WSP images.

The air within and behind a gust front converges with the ambient air ahead of the gust front. In a Doppler velocity image, this activity is observable as a boundary between regions of converging velocities. When viewed along a single radial, the convergence

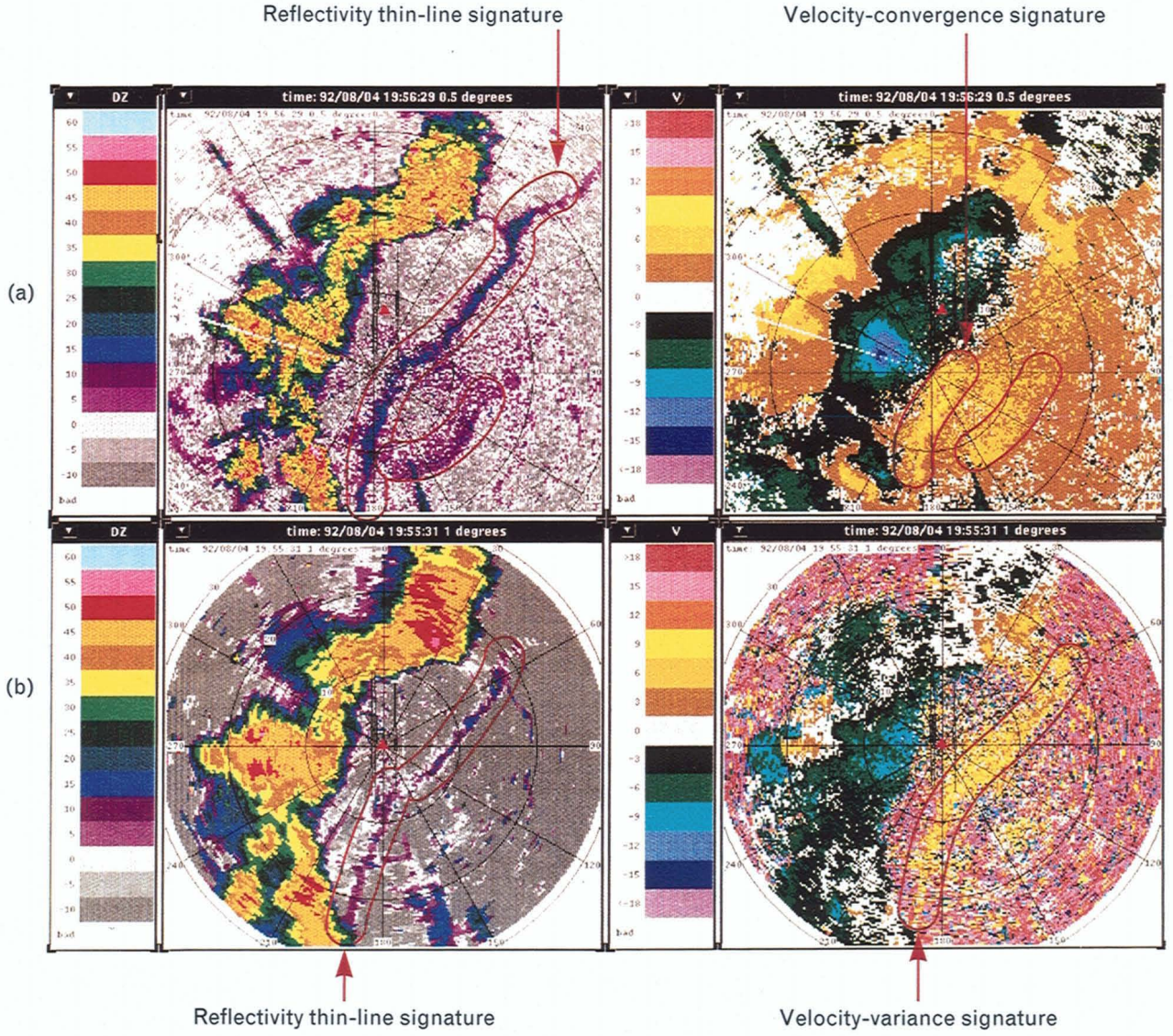


FIGURE 2. An example gust front in (a) TDWR and (b) ASR-9 WSP images. The left radar plots are reflectivity images with units in dBZ. The right radar plots are Doppler images with units in m/sec. The different signatures (see main text) of the gust front have been indicated by a human interpreter.

signature is characterized by a relatively sharp decrease in radial Doppler values with distance (Figure 3). Because Doppler radars can measure only the component of the wind that is directed along the beam, Doppler velocity measurements can often underestimate the true wind speed. In the extreme, the convergence signature of a gust front disappears completely when the direction of motion is perpendicular to the radar beam azimuth. The TDWR velocity image shown in Figure 2(a) demonstrates this problem. In the figure, the portion of the front closest to the radar site has a direction of motion that is nearly radially aligned, resulting in a pronounced convergent boundary for that area. However, at the ends of the gust front, where the direction of motion is more azimuthal, the boundary is more difficult to detect.

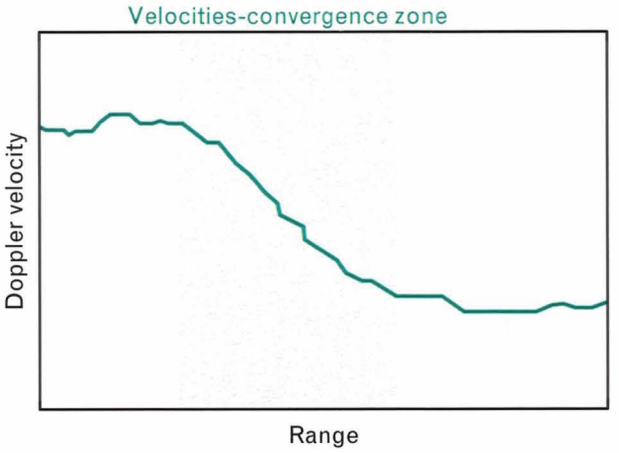


FIGURE 3. Example velocity-convergence signature associated with a gust front.

The thin-line signature is generally thought to be produced by the concentration of scatterers (dust, insects, rain droplets) along the leading edge of the thunderstorm outflow. Some gust fronts produce a distinctive cloud formation along the gust front, which can also contribute to the thin-line reflectivity. The thin line varies in width but seldom exceeds 3 km. Typical maximum reflectivities reported by the ASR-9 along gust fronts are in the range of 10 to 20 dBZ. But significant portions of many thin lines can have reflectivities as low as –5 dBZ, which is near or below the threshold of detectability for the ASR-9. (Note: The basic unit of measurement for radar reflectivity is dBZ. Reflectivities of 50 dBZ or more are typical of intense thunderstorms with heavy rain. Background typically has reflectivity values between –15 and 0 dBZ.) Because of ground-clutter obscuration, the quality of a thin-line signature often degrades at close range, and the signature can even vanish as the gust front passes over the radar. As the front moves out of the cluttered region, the signature often reestablishes itself. This type of degradation is especially troublesome for the ASR-9 because of the radar's on-airport location, which makes it more prone to detection loss when a gust front is affecting the airport.

A final key gust front signature is motion. When sequential radar scans are compared, convergence and thin-line signatures of a gust front will move conspicuously in a direction perpendicular to the orientation of the convergence boundary and reflectivity thin line. Signatures that do not move are either not gust fronts—e.g., they could be false alarms from range-ambiguous echoes (discussed in the subsection "Feature Detection"), edges of storm regions, or ground clutter—or they are gust fronts that are not operationally significant. Within limits, gust fronts tend to move uniformly as outwardly expanding curved boundaries; i.e., the propagation speed tends to be consistent along the front's length and across time. Of course, when gust fronts collide, the motion may become more erratic.

If these signatures were 100% reliable, detection would be a trivial task. For some gust fronts, however, one or more signatures may be weak, ambiguous, or entirely absent. For example, convergence signatures disappear when the radar beam is perpendicular to the wind velocity. Reflectivity thin lines and thin-line motion can disappear when a gust front is obscured by storm regions. To complicate matters further, none of these signatures are unique to gust fronts. Vertical shears, often present in severe thunderstorms, can bias low-altitude velocity estimates, producing apparent convergence signatures. Range-ambiguous echoes, ground clutter, flocks of birds, and elongated patches of low-intensity precipitation can all appear as reflectivity thin lines. Motion can be associated with anything (e.g., clouds or airborne dust) that follows the ambient wind. In short, each signature can be missing and each signature can be mimicked by other observable phenomena. Consequently, suc-

cessful discrimination requires knowledge of the circumstances for which these signatures are reliable as well as knowledge of gust front behavior. Only by weighing the quality of several signatures simultaneously can an automated system detect gust fronts with near human performance.

The task is difficult enough with TDWR data. And yet the TDWR is a pencil-beam radar, designed for weather sensing, with enough sensitivity to generate reliable Doppler values in relatively clear air and enough resolution in elevation to provide three-dimensional images of weather phenomena. In contrast, the ASR-9 is a surveillance radar that was not originally intended for weather imaging. With a fan-beam design, the ASR-9 vertically integrates signals into a single two-dimensional representation. Because the transmitted energy is distributed over a wider arc of elevation, the energy returned from a low-altitude, low-reflectivity gust front will be small relative to the energy filling the remainder of the sample volume. With this reduced sensitivity, gust front detection is much more difficult. Almost all convergence signatures are eliminated for the ASR-9 because the Doppler values are unreliable since the reflectivity returns from clear air are below the threshold of detectability for the radar. Even for cases in which gust fronts pass through regions of high reflectivity, convergence cannot be used reliably for gust front detection. For example, the signal contribution from overhanging precipitation near the edges of storms can bias the low-level wind-velocity estimate when there is vertical wind shear. Without convergence signatures, thin line and thin-line motion become the primary signatures for detecting gust fronts in ASR-9 WSP imagery. In the example ASR-9 WSP reflectivity image of Figure 2(b), the gust front is visible. But note that while the TDWR thin line is quite strong, the ASR-9 WSP thin line shows less contrast, is somewhat more fragmented, and does not extend as far as is apparent in the TDWR data.

Although a convergence signature is missing from the ASR-9 WSP velocity image of Figure 2, the gust front is still visible. The accuracy of velocity estimations degrades markedly over the range of signal-to-noise values associated with low reflectivity returns. For this reason, gust fronts are observable in ASR-9 WSP velocity images as bands of low-variance Doppler values, with high variance in the low signal-to-noise regions ahead and behind the gust front. This velocity-variance thin line is an alternative signature used in the ASR-9 WSP version of MIGFA. In addition, implicit zones of convergence can be identified. Doppler values within the gust front thin line are used to estimate winds behind the gust front. The environmental low-level wind velocity ahead of the storm can be measured by some other means—for example, from a network of anemometers at the airport. A comparison of these two wind-velocity estimates can be used to confirm that convergence exists somewhere between the gust front and the anemometer site.

Background

Automated radar gust front detection algorithms have been under development and evolution for almost ten years. H. Uyeda and D. Zrnić [2] first described an automated detection algorithm, developed for the Next Generation Weather Radar (NEXRAD), that was based solely on detecting velocity convergence along radials. The algorithm was successful in locating and tracking the strong gust fronts that commonly occur in Oklahoma during the spring.

An improved version of the initial algorithm reduces false alarms by requiring vertical association of gust front signatures from two different low-altitude elevation scans. The improved algorithm, known as the Gust Front Detection Algorithm (GFDA), also incorporates a technique for estimating horizontal winds ahead and behind detected gust fronts [3, 4]. As with its predecessor, GFDA detects velocity convergence along radials. GFDA is the algorithm currently intended for use in the initial operational deployment of TDWR systems.

Briefly described, GFDA begins with a search in each radial for runs, or segments, of decreasing radial velocity, indicating convergent shear. Segments in which the maximum shear exceeds a predetermined threshold are logically grouped into features on the basis of end-point-proximity and segment-overlap tests. The feature attributes are then tested against a number of thresholds and are kept, discarded, or combined with other features. After separately pro-

cessing each of the two full-circle scans from different altitudes, the algorithm tests for vertical continuity of the features between the scans. Features that exhibit vertical continuity and that exceed a minimum-length threshold are declared to be gust fronts. The reported location of the detected gust front is determined by fitting a curved line through the peak shear of each segment in the gust front feature. Sequential detections are associated over time to build detection histories for each gust front upon which propagation speeds are estimated and forecasts generated.

Lincoln Laboratory, in conjunction with the National Severe Storm Laboratory (NSSL), has since developed the Advanced Gust Front Algorithm (AGFA) [5, 6], which contains several enhancements, including reflectivity thin-line detection. AGFA detects thin lines by finding local maxima of reflectivity values that are consistent with the widths and intensities associated with gust fronts. Thin-line segments are generated twice: once by constructing segments over all range gates along a radial and once by constructing segments across radials along arcs of constant range. The final thin-line features consist of lists of the points connecting the centers of each of the segments. Convergence and thin-line features are fused on the basis of end-point proximity and orientation. AGFA does not use motion as a signature for detecting gust fronts. Motion is used only in heuristics that reject false features after they have been extracted.

During field testing in 1990 and 1991, a customized version of AGFA was used for gust front detection on an ASR-9 WSP [7, 8]. Because of the lack of reliable velocity-convergence features, the ASR-9 version of AGFA was configured to operate in a thin-line-only detection mode. Although the algorithm was successful in detecting gust fronts that had thin-line signatures of good quality, it had some difficulty detecting gust fronts when the reflectivity was weak or fragmented. Lacking convergence signatures to confirm the existence of gust fronts, the algorithm was prone to false alarms triggered by elongated low-reflectivity weather echoes that are sometimes associated with stratiform rain. Installing suboptimal detection thresholds to reduce the false-alarm rate further reduced the detection probabilities.

In the above study, the scoring was done against human interpretations of the same images used as input to the algorithm. The discrepancy between human and AGFA performance appears to be partially due to AGFA's not making full use of a variety of additional information that is available in the ASR-9 WSP data, including velocity thin lines and thin-line motion. Moreover, both GFDA and AGFA rely on sequentially applied thresholds to discriminate gust fronts from background. When the relevant signals are weak or ambiguous, the use of thresholds in the early stages of processing can result in the elimination of potentially relevant information, thus setting unnecessary limits on detection performance. GFDA and AGFA also rely on one-dimensional signal processing operations to locate gust fronts. The extraction of chains of points across the second dimension is done at a higher, heuristic level of processing. In contrast, two-dimensional signal processing operations can directly establish the shape of gust fronts without relying on heuristics. Finally, these early gust front algorithms have no systematic means of conditionally fusing information from various sources by taking into account the different reliabilities of the sources. Different signatures can have varying reliability depending on the situational context.

Low-Level Machine Intelligence

The conventional wisdom in computer vision/object recognition research has been to use general image processing operations, ideally devoid of object- and context-dependent knowledge, at the initial stages of processing. Such operations might include edge detection, segmentation, cleaning, and motion analysis. And yet the ideal has never really been achieved in practice. For example, some knowledge of the sensor and the expected scene contents must be implicitly encoded in the form of thresholds or other similar parameters to detect edges effectively.

From the results of such general operations, image characteristics are extracted and represented symbolically. Machine intelligence is then applied, as if by definition, only on the symbolic representations at higher levels of processing.

MIGFA has inherited the development environment, control structure, knowledge-based signal processing, and several other important attributes of

XTRS. In contrast to more conventional approaches to object recognition, sensor-, object-, and context-dependent knowledge is applied in the earliest levels of processing, i.e., at the image processing stage. As used in MIGFA, low-level machine intelligence applies knowledge in three ways.

First, knowledge of the current environment is used to choose from a library those feature detectors which are *selectively indicative* of the object being sought. Using multiple independent feature detectors, MIGFA can adapt to different contextual circumstances. At the beginning of the processing of each scan, a rule-based expert examines contextual information to select a set of feature detectors known through experience to be the most effective for a given set of circumstances. In the extreme, this process would enable MIGFA to adapt itself dynamically to changes in the environment. Currently, the only rule used by MIGFA selects between two fixed alternative sets of feature detectors, one set customized for the TDWR and the other customized for the ASR-9 WSP. Because of the redundancy inherent in the use of multiple feature detectors, MIGFA tends to be robust: the malfunction of a feature detector or even the absence of one data source does not necessarily halt processing and may have only minor effects on detection performance.

Second, knowledge is also incorporated within feature detectors through the design of matched filters that are customized to the physical properties of the sensor, the environment, and the object to be detected. A new technique of knowledge-based signal processing, called *functional template correlation* (FTC), allows the construction of customized signal processing operations that are more effective than standard operations (see the box, "Functional Template Correlation"). The output of FTC is a map of numeric values in the range [0,1] that indicate the degree of match between the pattern of pixels in an image region and the feature or object encoded in the functional template.

Finally, knowledge of the varying reliabilities of the selected feature detectors is used to guide data fusion and extraction. Conditional data fusion is simplified by using "interest" as a common denominator [9]. An interest image is a spatial map of evidence for the presence of some feature that is selectively indicative of an object being sought (the output of FTC is an interest image as long as the functional template encodes an indicative feature). Higher pixel values reflect greater confidence that the intended feature is present at that location. Using interest as a common denominator, MIGFA fuses data by combining interest images derived from various pixel-registered sensory sources. Using simple or arbitrarily complex rules of arithmetic, fuzzy logic, or statistics, MIGFA can assimilate pixel-level evidence from several coregistered sources into a single combined interest image. Clusters of high values in such combined interest images are then used to guide selective attention and can serve as the input for object extraction. If done effectively, the combined interest image provides a better representation of object shape than is evident in any single sensor modality. Using these techniques, MIGFA performs a significant amount of knowledge-based processing before the application of the first discriminating threshold. Most traditional perception systems apply one or several thresholds early in the processing as a way of quickly reducing the amount of data to be processed. However, especially with ambiguous data, each applied threshold closes off options for detecting an object. A better strategy—a strategy attempted in XTRS and MIGFA—is to apply thresholds only after evidence from many sources of information have been meaningfully fused into a single map of evidence.

MIGFA Design

The system block diagram in Figure 4 is an overview of MIGFA as configured for ASR-9 WSP data. In preparation for processing, input images V (Doppler velocity) and DZ (reflectivity) from the current radar scan are converted from polar to Cartesian representation and scaled to a useful resolution. Image SD is a map of the local standard deviations of V values. The SD and DZ images are then passed to multiple simple independent feature detectors that attempt to localize those features which are selectively indicative of gust fronts. The outputs of each of these feature detectors, most of which are based on some application of FTC, are expressed as interest images that specify evidence indicating where and with what confidence a gust front may be present. The different interest images

FUNCTIONAL TEMPLATE CORRELATION

FUNCTIONAL TEMPLATE correlation (FTC) [1, 2] is a generalized matched filter that incorporates aspects of fuzzy set theory. Consider, as a basis for understanding, the basic image processing tool *autocorrelation.* Given some input image I, an output image O is generated by matching a kernel K against the local neighborhood centered on each pixel location I_{xy}. The match score assigned to each pixel O_{xy} is computed by multiplying each element value of K by the superimposed element value in I and summing across all products. If the shape to be matched can vary in orientation, then the pixel I_{xy} is probed by K at multiple orientations. The score assigned to O_{xy} is the maximum across all orientations.

FTC is fundamentally the same operation with one important exception: whereas the kernel used in autocorrelation is an array of image values (the array is essentially a subimage of the image to be probed), the kernel used in FTC is an array of *scoring functions.* The scoring functions return scores that indicate how well the image values match the expectations of the values at each element of the kernel. The set of all returned scores are averaged and "clipped" to the continuous range [0,1]. (In the clipping process, those averaged scores which are less than zero are assigned a value of zero while those averaged scores which are greater than one are assigned a value of one.) The output of FTC is a map of these values, each of which reflects the degree that the shape or object implicitly encoded in the functional template is present at that image location.

Consider as an example the functional template implementation of a simple matched filter designed to detect gust fronts in reflectivity data (Figure A). Gust fronts are observed as thin lines of moderate reflectivity (approximately 0 to 20 dBZ) that are flanked on both sides by low reflectivity (approximately −15 to 0 dBZ). Figure A(1) shows the template kernel consisting of integers that correspond to the two scoring functions shown in Figure A(2). Elements of the kernel that do not correspond to either of the scoring functions form guard regions in which image (i.e., reflectivity) values are ignored and have no effect on match scores. Scoring function 0, corresponding to the flanking regions of low reflectivity, returns a maximal score of 1.0 for image values in the interval of −20 to −5 dBZ, a gradually decreasing score for image values in the interval −5 to 10 dBZ, and a score of −2.0 for image values larger than 10 dBZ. Scoring function 1, corresponding to the center of the kernel where moderate reflectivity values are expected, returns maximal scores in the interval 5 to 12.5 dBZ and gradually decreasing scores for both higher and lower image values. Note that although very low image values can generate scores of −1.0, a slower decline in score with a minimum score of 0.0 is returned for image values above the maximal scoring interval. This asymmetry is an attempt to mitigate the obscuring effects of storm regions and other patches of high reflectivity.

In general, by increasing or decreasing the intervals over which affirming scores (i.e., scores > 0.5) are returned, scoring functions can encode varying degrees of uncertainty with regard to which image values are allowable. In addition, knowledge of how a feature or object appears in sensor imagery can be encoded in scoring functions. The interfering effects of occlusion, distortion, noise, and clutter can be minimized by the use of various design strategies [3]. As a consequence, matched filters customized with FTC for specific applications are generally more robust than classical signal processing operations. In the thin-line matched-filter example shown in Figure A, the filter does not simply find thin lines, but selects those thin lines which have reflectivity values within a particu-

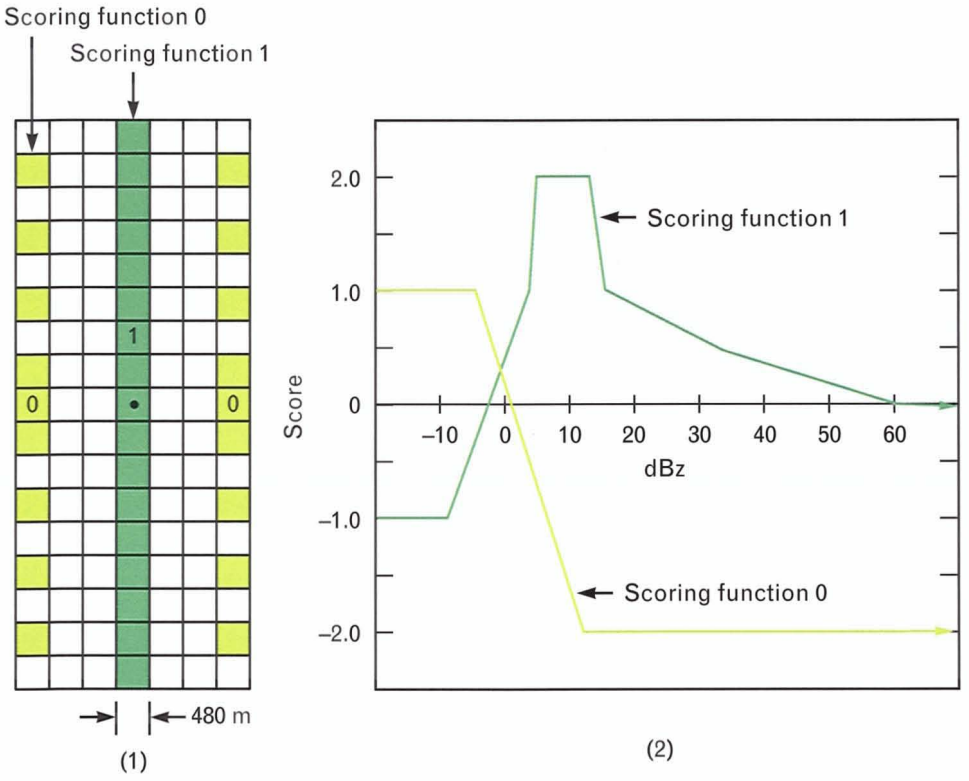


FIGURE A. Example functional template for thin-line feature detection: (1) index kernel and (2) corresponding scoring functions. By increasing or decreasing the intervals over which affirming scores (i.e., scores > 0.5) are returned, scoring functions can encode varying degrees of uncertainty with regard to which image values are allowable. In addition, knowledge of how a feature or object appears in sensor imagery can be encoded in scoring functions.

lar range. Furthermore, the matched filter can display differential tolerances to image values that are higher or lower than the expected range of values. In the automatic target recognition (ATR) systems developed at Lincoln Laboratory, FTC has been used primarily as a direct one-step means of three-dimensional object detection and extraction. In the Machine Intelligent Gust Front Algorithm (MIGFA), FTC is used more as a signal processing tool for edge detection, thin-line filtering and smoothing, shape matching, skeletonizing, and erosion.

If FTC were implemented literally as described here, the computational expense would be prohibitive for most useful tasks. But FTC is actually faster than autocorrelation if the input data are scaled to a fixed integer range (e.g., 0 to 255) and the scoring functions are implemented as a precomputed two-dimensional lookup table that is indexed by a scoring-function number and an image value.

References

1. R.L. Delanoy, J.G. Verly, and D.E. Dudgeon, "Functional Templates and Their Application to 3-D Object Recognition," *Proc. Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), San Francisco, Mar. 1992.*
2. R.L. Delanoy and J.G. Verly, "Computer Apparatus and Method for Fuzzy Template Shape Matching Using a Scoring Function," U.S. Patent No. 5,222,155 (June 1993).
3. R.L. Delanoy, J.G. Verly, and D.E. Dudgeon, "Pixel-Level Fusion Using Interest Images," *Technical Report 979,* MIT Lincoln Laboratory (26 Apr. 1993).

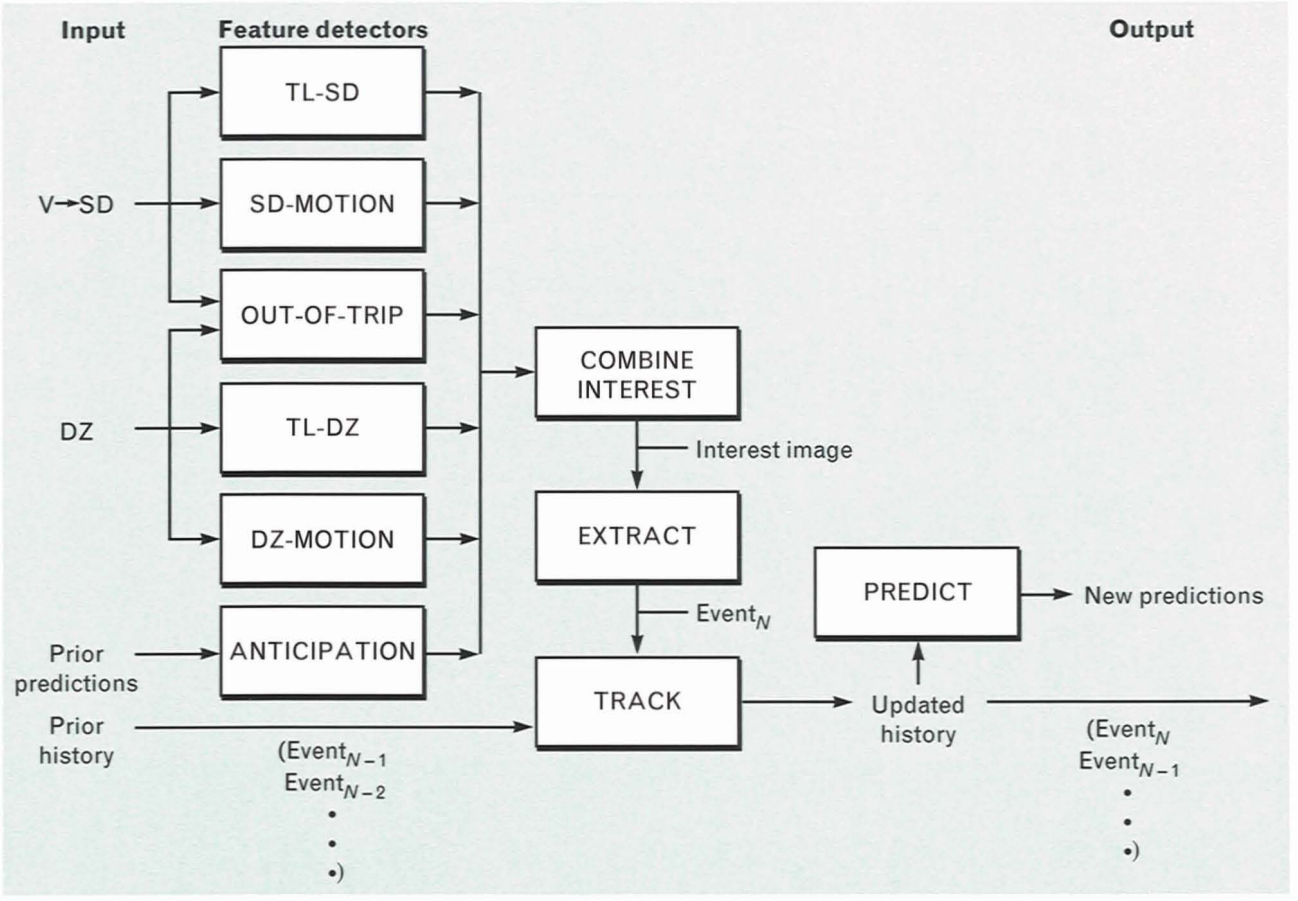


FIGURE 4. Block diagram of the Machine Intelligent Gust Front Algorithm (MIGFA). For a description of the different feature detectors, see the subsection "Feature Detection" in the main text.

are fused to form a combined interest image, thus providing an overall map of evidence indicating the locations of possible gust fronts.

From the combined interest image, fronts are extracted as chains of points. The chains extracted from a radar scan, collectively called an event, are integrated with prior events by establishing a point-to-point correspondence. Heuristics are then applied to reject those chain points which have an apparent motion that is improbable. The updated history is used to make predictions of where points along the front will be located at some future time. Such predictions are used in the processing of subsequent images, specifically in the feature detector called ANTICIPATION. In the output of ANTICIPATION, high interest values are placed wherever fronts are expected to be, thereby selectively sensitizing the system to detect gust fronts at specific locations. ANTICIPATION is tuned so that it will not automatically trigger a detection by itself but, when its output is averaged with other interest images, it will support weak evidence that would otherwise be insufficient to trigger a detection. Figure 5 is a summary of the processing steps for an example ASR-9 WSP scan.

Image Preparation

As discussed earlier, velocity convergence is an unreliable signature for detecting gust fronts in ASR-9 WSP data. Gust fronts, nevertheless, *are* visible in velocity images. Because of the tendency for high-pass clutter-filtered pulse-pair Doppler estimates in a velocity image to have high variance in regions of low signal-to-noise ratios (SNR), the local velocity variance is higher for an area of clear air than for an area associated with slightly higher reflectivity values. This information is

translated into a usable form by transforming the velocity image V into a map of local standard deviations (the SD image). At each pixel of V, the standard deviation was computed in the surrounding 5 × 5 pixel neighborhood and assigned to the corresponding pixel in SD.

Pixel values for all images are scaled to the interval 0 to 255 to support subsequent FTC operations on the input imagery. Each image is tagged with the scaling factor and offset necessary to translate scaled values back to the original physical values.

Finally, the DZ and SD images are converted from polar arrays (240 range bins × 256 radials) to Cartesian arrays (130 × 130). Mapping is done by computing for each element of the Cartesian array the range bin and radial at which the corresponding value is to be found in the polar array. During the mapping process, an implicit subsampling of the data occurs. From an initial radial resolution of 120 m per range bin and pixel size in the azimuthal dimension decreasing from 680 m at 28 km, the final Cartesian image has a pixel resolution of 480 m per pixel.

Feature Detection

Given contextual information of the sensor being used, the location of that sensor, and the environmental conditions, a rule-based expert selects an appropriate set of feature detectors for application to the input

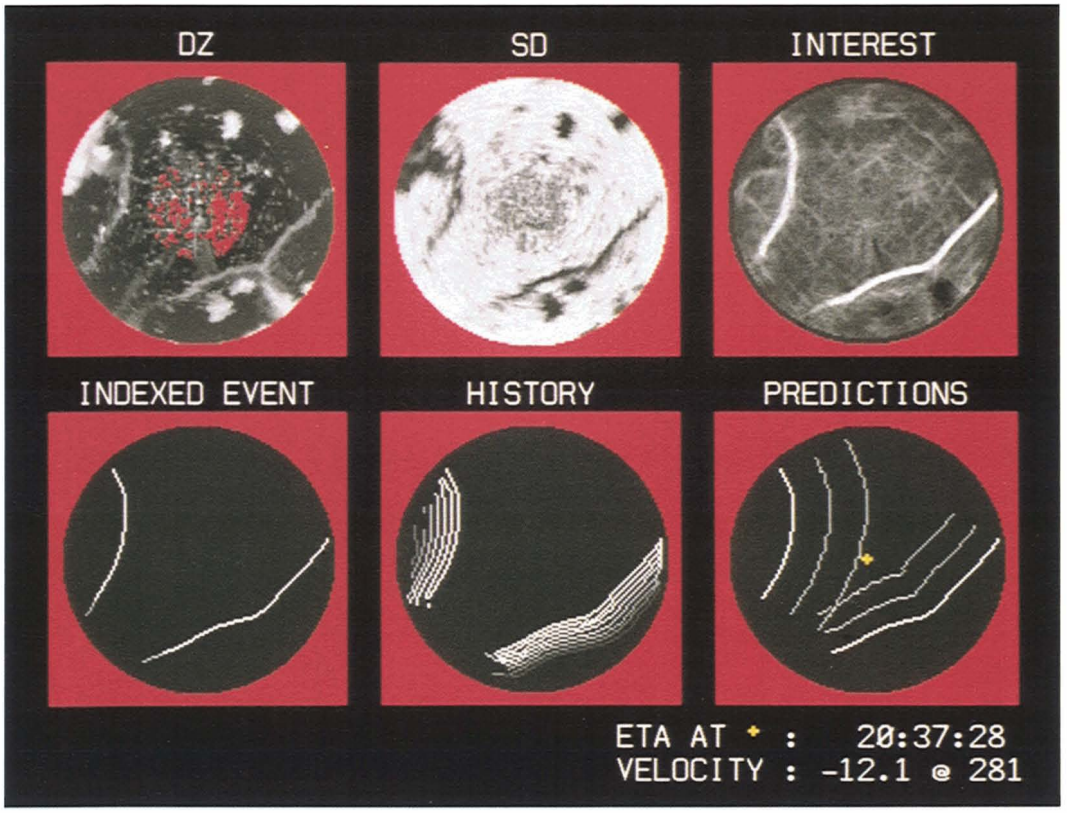


FIGURE 5. Processed scan summary. In the first row are the DZ (reflectivity) image, SD (standard deviation of velocity) image, and the combined interest image that has been computed from the DZ and SD images. The second row begins with the extracted indexed event. White pixels are those points which have been declared as part of a gust front. Gray pixels are those points which have not been tracked long enough to establish sufficient confidence. In the history frame, the current chain is shown in white and the preceding scans are shown in shades of gray (darker shades indicate more distant events in time). In the predictions frame, fat gray pixels indicate the 10- and 20-min forecasts of where the fronts are expected to be. Also shown at the bottom right corner are the estimated time of arrival of the next gust front to cross the radar site, the speed of the winds measured inside the front (in m/sec), and the direction (in degrees) from which the front is coming.

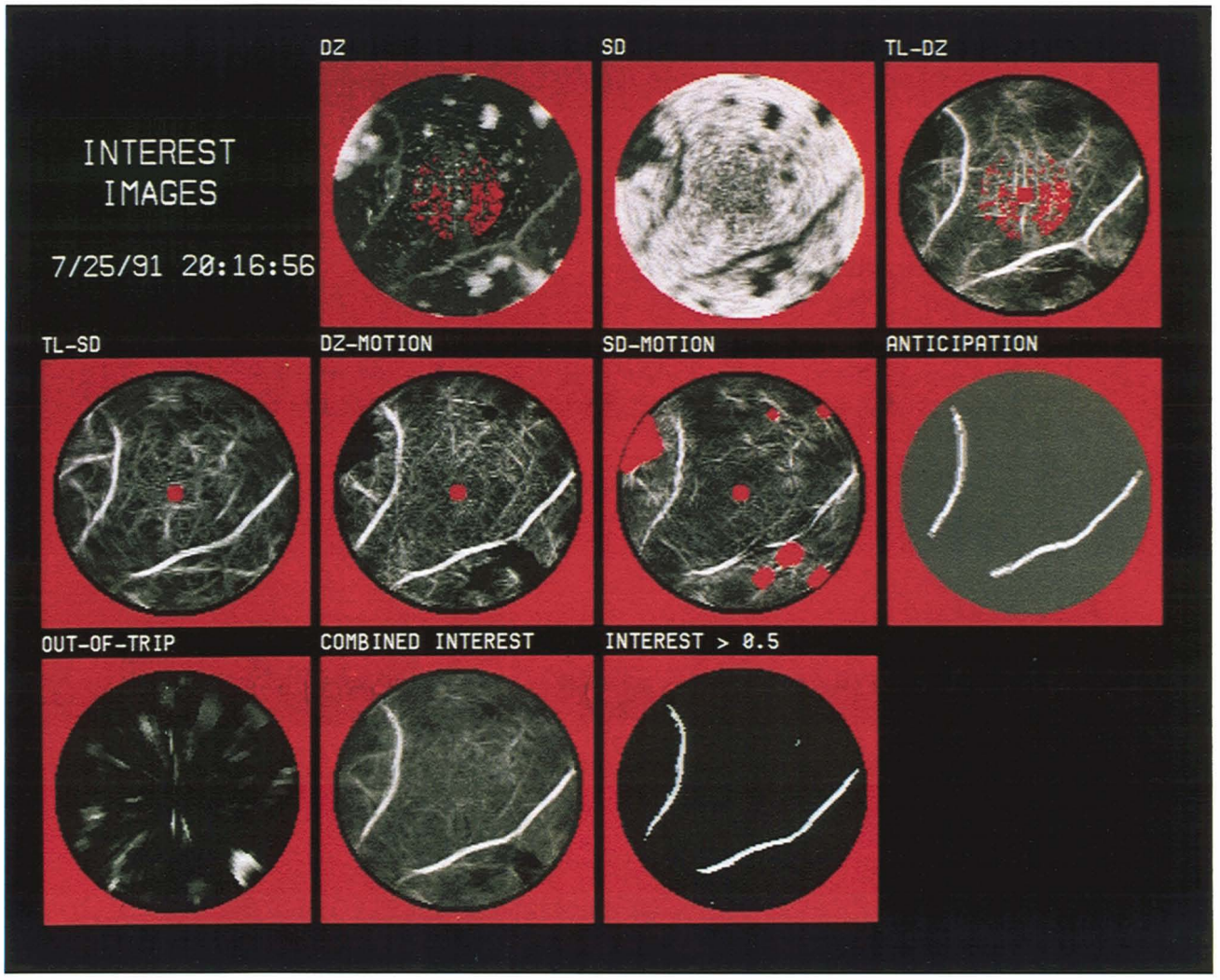


FIGURE 6. Combining interest images: strong evidence. An example DZ and SD image of two gust fronts are shown, along with the output images of the different feature detectors of Figure 4. Regions of red pixels indicate areas where specific feature detectors have not expressed an opinion regarding the presence of a gust front, deferring instead to the evidence generated by other feature detectors. The last two frames (the center and right frames of the third row) show the images that resulted from combining the different interest images.

data. For the moment, the only rule that is available chooses between two sets of feature detectors: one set customized for the ASR-9 WSP, the other for the TDWR. These two sets alone may be sufficient. As MIGFA testing is expanded, however, additional rules adapting the feature-detector set for sites other than central Florida and seasons other than midsummer may be added as required. Although we do not anticipate the need, the set of feature detectors could be designed to adapt dynamically from one scan to the next as changes in weather conditions are detected. The feature detectors that follow are the ones that were tested for the ASR-9 in Orlando during 1992.

TL-DZ and TL-SD (Figure 4) are two simple thin-line feature detectors that are used for the input images DZ and SD, respectively. The TL-DZ feature detector is based primarily on the application of the functional template illustrated in Figure A of the box, "Functional Template Correlation." An example DZ image and the interest image generated by TL-DZ and other feature detectors are shown in Figures 6 and 7. Although the thin-line functional template applied to SD has the same kernel as the template for TL-DZ, the scoring functions are different: thin lines

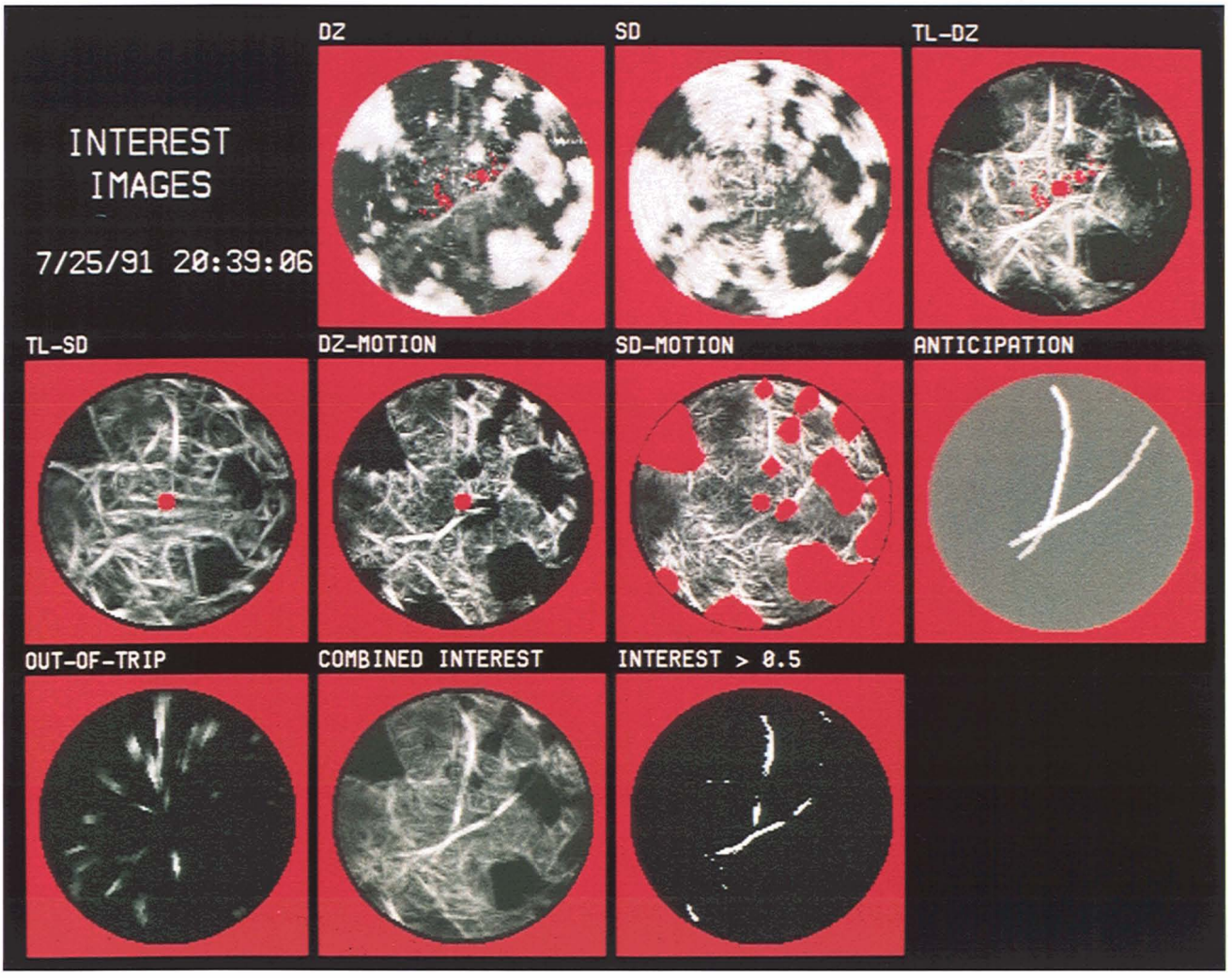


FIGURE 7. Combining interest images: weak evidence. This figure is similar to Figure 6 except that here the two gust fronts are not clearly visible in any of the single interest images except for the anticipation image. In the combined interest images, however, the gust fronts are much more apparent, illustrating how the fusion of weak evidence from multiple sources enhances gust front detectability.

associated with gust fronts have low Doppler velocity standard-deviation values within the front and high values ahead and behind the front. Consequently, the scoring function for the center strip returns maximal scores for low values while the scoring function for the flanking regions returns maximal scores for high values.

DZ-MOTION and SD-MOTION are two thin-line motion detectors that are very similar to the basic thin-line detectors TL-DZ and TL-SD. The detection of motion is based on simple differencing. For example, in DZ-MOTION the DZ image from a previous scan produced approximately 4 min before the current scan is subtracted from the current DZ image. In the differenced DZ image, gust fronts appear as white lines (positive values at the front's position in the current scan) that are trailed by parallel dark lines (negative values at the front's position in the previous scan). Although functional templates that can scan for parallel white and dark thin lines simultaneously are feasible, these types of templates have so far proven to be too computationally expensive to operate within the real-time constraints of the available computer resources. Thus the existing

DZ-MOTION simply looks for thin lines of positive values. The functional template used has a kernel that is identical to the one shown in Figure A of the box, "Functional Template Correlation," but the scoring functions are somewhat different because of the consequences of differencing. The feature-detector SD-MOTION is similar to DZ-MOTION in that SD-MOTION also applies to the difference of two sequential images a thin-line filter with customized scoring functions. With this approach, thin lines that do not move are given low interest values, reflecting the belief that a stationary thin line is either not a gust front or is a gust front that may be ignored. Because the background in differenced images is reduced to values near zero, DZ-MOTION and SD-MOTION tend to be more sensitive than TL-DZ and TL-SD.

One disadvantage of DZ-MOTION and SD-MOTION is that they tend to produce false alarms when moving storms are present because the leading edge of the storm may appear in the differenced image as a thin line of positive values. For reducing the likelihood of such false alarms, an image of storm regions is generated with a round functional template whose kernel has a diameter of 13 pixels (6.25 km). Wherever storm regions are detected with this template, interest values are decreased in DZ-MOTION and set to nil (i.e., no opinion) in SD-MOTION.

A fifth feature detector, OUT-OF-TRIP, highlights range-ambiguous echoes. Range-ambiguous echoes occur when signals are reflected by weather more distant than the maximum unambiguous range. Because the signals have traveled farther, they arrive back at the radar receiver at the same time as signals that are transmitted later and reflected from nearer weather (hence the name OUT-OF-TRIP). For these range-ambiguous echoes, the apparent range extent is maintained while the azimuthal extent is reduced proportional to the range; thus the signals have a distinctive appearance as reflectivity thin lines that are radially aligned and that are associated with high local variance in the Doppler data. Because of their thin-line appearance, range-ambiguous echoes are often inappropriately given high interest values by both TL-DZ and DZ-MOTION.

The detection of out-of-trip signals is performed by applying two functional templates simultaneously. One template looks for radially aligned thin lines in the DZ image, while the other requires that the corresponding SD values are high. The result is an interest image that highlights out-of-trip signals. After the combination of all other interest images, the out-of-trip interest image is subtracted from the combined interest image, thus selectively suppressing evidence for the presence of gust fronts where out-of-trip signals are found. Example outputs of the OUT-OF-TRIP feature detector are shown in Figures 6 and 7.

The ANTICIPATION feature detector provides a mechanism, based on situational context, for spatially adjusting the detection sensitivity of MIGFA. High anticipation values get averaged with interest values from other feature detectors to increase the likelihood of detection at specific locations. Similarly, low anticipation values suppress the likelihood of detection.

The most important use of anticipation is as a replacement for *coasting*. Simply defined, coasting is the continued tracking of a target on a radar screen for some time interval after the target has disappeared (i.e., after the target's signal has fallen below some detection threshold). Coasting assumes that the loss of a target's signal is not due to a change in the target's behavior (e.g., a change in velocity or perhaps the disappearance of the target). Gust fronts, however, do change their behavior, as in cases in which gust fronts collide. Consequently, the blind coasting of a signal after the signal's loss is a potential source of false alarms. As an alternative to blind coasting, anticipation provides a mechanism for progressively increasing the sensitivity of a detection system, supporting weak evidence that would otherwise fall below detection thresholds.

In MIGFA, prior history of the behavior of a particular gust front is used to predict where that front is expected to be in the current scan. The predictions are used to create a band of elevated interest values, typically not so high as to trigger a detection by themselves, but high enough to raise collocated weak signals above threshold. In general, as the length of time a gust front has been tracked increases, the anticipation interest values can also be increased. If absolute coasting is desired, interest values can be increased to a level high enough to trigger a detection

without any other supporting evidence. Examples of anticipation interest images are shown in Figures 6 and 7.

Anticipation can also be used to adjust the sensitivity of gust front detections on the basis of contextual knowledge. Some examples follow:

1. Many gust fronts are not observable in radar data when the fronts are directly over the radar site because of obscuration by intense ground clutter. Even with anticipation of where a gust front is expected to be, the radar system can often lose the front as the front crosses over the radar site. To prevent such a loss, absolute coasting over the radar site can be accomplished by setting interest values within 2 km of the radar site to nil (i.e., missing values) for all interest images *except* the anticipation image. Consequently, the anticipation interest image will be the only image allowed to have an opinion of what exists directly over the radar site.
2. Gust front false alarms often occur from thin, elongated bands of low-reflectivity stratiform rain. In central Florida at least, gust fronts are seldom associated with the stratiform rain that often follows intense storm activity. Hence, under such conditions, the ANTICIPATION feature detector suppresses the background anticipation interest values.
3. False alarms are rare in the absence of any precipitation. Thus, when no precipitation is visible on the radar screen, the background anticipation interest values may be safely raised, thereby increasing the likelihood of detecting an incoming gust front that is generated by a more distant storm.

Combining Evidence

During the feature-detector selection process, a *rule of combination* is also chosen to govern the combining of evidence—an example of data fusion. In principle, the rule of combination can be as simple as the averaging of pixel values across all interest images. However, for the set of ASR-9 WSP feature detectors described earlier, a somewhat more complicated rule has been used.

The four interest images generated by TL-DZ, TL-SD, DZ-MOTION, and SD-MOTION are averaged together. During the process, any missing values are ignored. The resulting averaged interest image and the anticipation interest image are combined as a weighted average: the average of the first four interest images is given a weight of 0.75 while the anticipation image is given a weight of 0.25. Finally, elements of the out-of-trip interest image are multiplied by 0.25 and subtracted from the elements of the weighted average. The resulting image is called the combined interest image.

Figure 6 shows an example ASR-9 WSP DZ image, the outputs of each feature detector, and the final interest image. In this case, strong evidence for the two fronts is visible in each of the component interest images (except, of course, for the out-of-trip image). Clearly, any one of the feature detectors acting alone would have been adequate. Now consider Figure 7, which summarizes the evidence for the presence of the two gust fronts in a later scan in which detection has become more difficult as accumulating storm regions have occluded the fronts. Note that although different parts of the gust fronts are highlighted in different interest images, the gust fronts are not unambiguously visible in any single interest image (except the anticipation image). In the combined interest image, however, the gust fronts are much more apparent. This example illustrates how evidence derived from multiple feature detectors can be combined so that the various detectors mutually support and compensate for one other.

In MIGFA, no one feature detector is meant to be a perfect, or even necessarily a good, discriminator of gust fronts and background. When used together, however, several weakly discriminating feature detectors can achieve robust performance depending on how the detector outputs are combined.

Extraction

Algorithms, such as AGFA, that track gust fronts as entities must identify gust fronts prior to tracking. The algorithms rely on the assignment of unique labels that permit the establishment of correspondence across time. Gust front statistics, such as propagation speed and location, are computed for the front

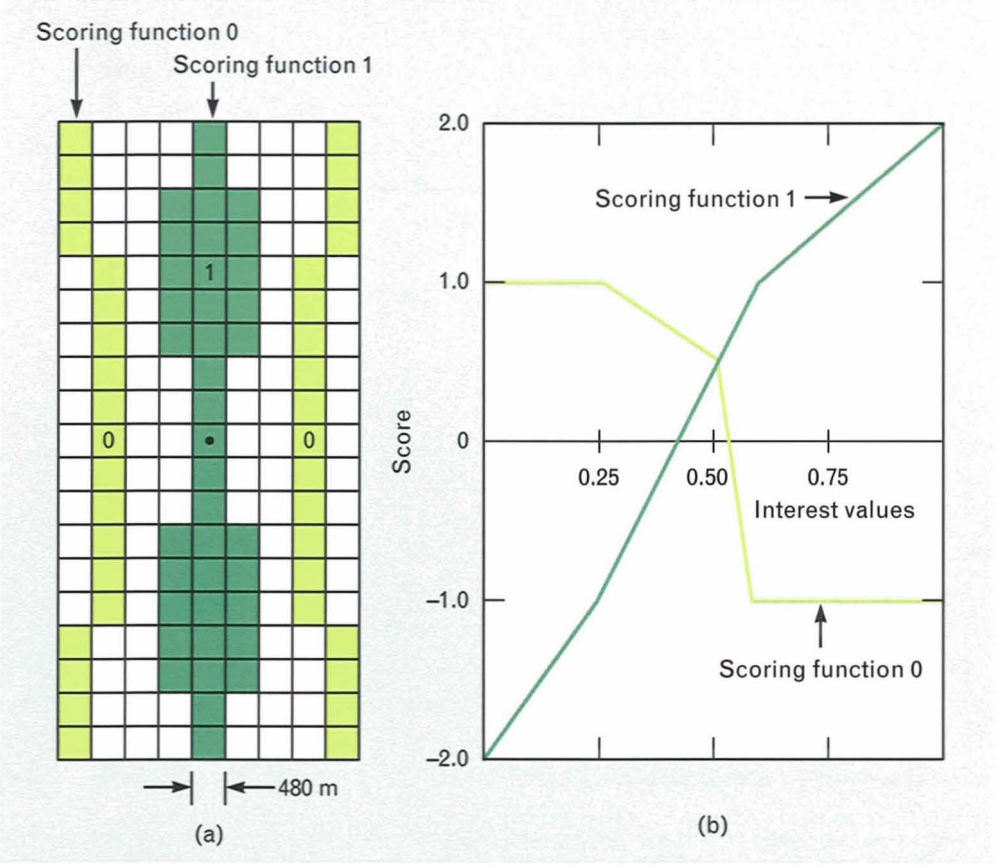


FIGURE 8. The bow-tie functional template used for thin-line smoothing: (a) index kernel and (b) corresponding scoring functions. (For an explanation of functional templates, see the box, "Functional Template Correlation.")

as a whole. This approach is adequate for simple cases. Inevitably, however, complex rules are required to handle the labeling, correspondence, and tracking for cases in which a single front breaks up into disjoint fragments or for cases in which multiple fronts merge or collide. Given the variable nature of gust front behavior, the construction of a fully comprehensive set of rules that are correct for all possible circumstances is a difficult task.

The problem is bypassed in MIGFA by making the goal of extraction the identification of all points (collectively called an *event*) that lie in *any* gust front. Certainly, some chains of points are spatially segregated or have different velocities. For purposes of reporting, such chains can be inferred to belong to separate gust fronts even though there is no concerted attempt to label or track gust fronts as entities. Instead, individual points are tracked across time; that a point belongs to one gust front or another is irrelevant to processing. MIGFA predictions are elastic in that the variable velocities of different points along the gust fronts are each used to make predictions of what the gust front appearance will be at some time in the future.

Thin lines in the combined interest image can be fragmented for gust fronts that intersect with out-of-trip weather or for fronts obscured by storm regions. To bridge gaps between collinear fragments and to suppress random unaligned high-interest values, MIGFA uses thin-line smoothing of the combined interest image. Figure 8 shows the bow-tie functional template used as the basis for thin-line smoothing. The template, inspired by the receptive field of the cooperative cell of the Boundary Contour System developed by S. Grossberg and E. Mingolla [10], has a bow-tie shape that weights the influence of the end regions over that of the center by placing more kernel elements at the ends. Consequently, the template generates high output interest scores for an image point between two collinear high-interest segments, even if

that middle point itself has a low input interest value. Because of the scoring-function design, the bow-tie filter suppresses those collinear interest values which are below the level of ambiguity (0.5), and amplifies those values which are above the level of ambiguity. With this design, the boundaries between gust fronts and background are sharpened, resulting in cleaner shapes for subsequent processing. An example of an input image of combined interest and an output smoothed image are shown in Figure 9.

A threshold of 0.5 is then applied to the smoothed image to create a binary image of candidate fronts. The lengths of resulting elongated shapes are then computed, and the elements of those binary shapes which are too short (<6 km for the ASR-9 WSP) are set to 0. The result of this process is shown in the frame labeled "match > 0.5" in Figure 9.

The bow-tie functional template also generates a map of orientations. In the orientation image, each element indicates the orientation that is associated with the highest-scoring bow tie rotated at 10° increments from 0° to 170°. Black pixels correspond with best matches at 0°; white pixels correspond with best matches at 170°. An example orientation image is shown in Figure 9.

The elongated binary shapes of the "match > 0.5" image can be thinned down to a single-pixel-width skeleton by using an FTC implementation of a modified version of S. Levialdi's homotopic thinning [11]. The result of thinning is shown in the frame labeled "marked thinned" in Figure 9.

The chains of points resulting from thinning are then extended along ridges of relatively high interest by using what is essentially a road-following algorithm. At each end point, the pixels immediately surrounding that point are examined by looking outward from the rest of the chain for the maximum-interest pixel with an orientation (found in the orientation image) that is within a specified angle from that of the initial end point. When the maximum interest score of a new point falls below 0.2 or when no new point has an orientation consistent with the initial end point, extending halts. The result of the extending process is shown in the frame labeled "extended" in Figure 9.

After the chain-extension process has been com-

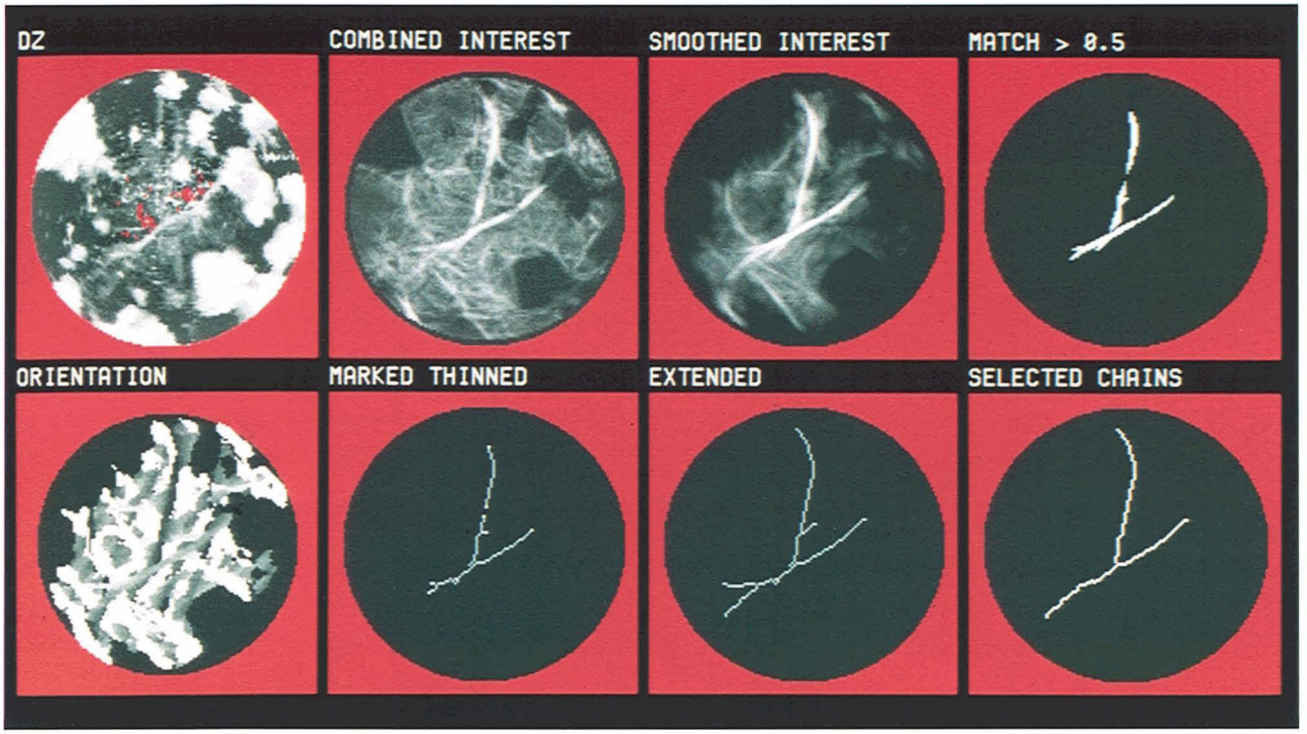


FIGURE 9. Extraction steps. Candidate gust fronts are extracted from a combined interest image. For a description of the different steps involved, see the subsection "Extraction" in the main text.

pleted, the resulting image may be highly branched and it may contain loops. For further refinement of the image, chain segments are assigned scores based on the sum of the corresponding interest values found in the smoothed interest image. In each disjoint network of chain segments, the single most interesting (usually, but not always, the longest) non-looping combination of chain segments is extracted as the candidate gust front. Once the most interesting chain has been extracted, the process is repeated on the remaining unextracted chain segments to find the next most interesting combination of chain segments. The extraction process is repeated until the most interesting remaining chain is below an empirically determined interest threshold. In Figure 9, the frame labeled "selected chains" shows the set of above-threshold combined chain segments that were extracted from the "extended" image.

Tracking/Heuristics

As stated earlier, each point in the extracted event is tracked individually. The tracking of a particular point requires that the corresponding point in the event immediately prior to the current event be found. Correspondence can be difficult to establish when several gust fronts collide; in such cases, the point in the prior event that is closest to a point in the current event might not necessarily be the correct corresponding point. Consequently, the corresponding point is chosen to be the closest point in the immediately prior event for which the orientation and speed are consistent with the given point in the current event. If no such point in the prior event is found, then the corresponding point is assumed to be the closest point. Once correspondence for a point is established, the point is indexed by creating a pointer linking that point to the corresponding point in the immediately prior event. If the distance between the two corresponding points is too large or if the distance is inconsistent with prior history, then the point is unindexed (i.e., the link is broken). Through the index links, a point can be tracked backwards in time to its first recorded instance. The number of prior events through which a point can be tracked is called the point's depth. (A depth of 0 means that the point is unindexed.) Once indexed, each point is assigned the following attributes: coordinates, distance moved, direction moved, depth, Doppler value, interest value, and propagation speed.

After indexing, each extracted chain of points is edited:

1. If the direction a single point moves is opposite (approximate difference of 180°) from its neighbors, the direction of the point is reversed.
2. Single chains may be divided into two subchains if a persistent discontinuity in velocity or a persistent change in orientation is detected at some point along the chain.
3. Various parameters such as propagation speed, Doppler value, and direction of motion are smoothed along the length of each chain.
4. Heuristics are applied that, when satisfied, unindex individual points in a chain. If more than half of any chain's points become unindexed, all points in the chain are unindexed.

The heuristics mentioned in item 4 above are based on knowledge of how false alarms can be distinguished from real gust fronts. For example, if the direction a point moves is inconsistent with the measured Doppler value, the point is unindexed. Or, if the point is approaching the radar site and moving in the same direction and no faster than the winds measured by anemometers at the radar site (i.e., there is no convergence), the point is unindexed.

In the final stage of tracking, a binary decision is made for each chain as to whether the chain should be declared a gust front. A chain's summed interest score and the depths of its constituent points are used to make the decision. For chains with high summed interest scores (reflecting a higher degree of confidence), points with lower depths may be included. On the other hand, chains that have low summed interest scores are less likely to be gust fronts and are thus required to accumulate higher depths before being included in the announced gust front detections. The frame labeled "indexed event" in Figure 5 shows the set of all extracted points. White pixels represent those points which have the sufficient depths and interest scores to be reported. Gray pixels represent those

points which will not be reported due to a lack of confidence. In the frame labeled "history," the reported points are shown in context with previously reported events.

Prediction

The current extracted event, indexed into the prior history, is used to predict the future locations of those points which have the sufficient depths and interest scores. Given the direction moved, the propagation speed, and the current coordinates of a point, a new coordinate is computed for some specified time in the future. Gaps can arise between the projected future coordinates of two adjacent gust front points when the orientations and velocities of the points are not identical. In such cases, the gaps are filled in. An example showing the reported chains and their expected locations after 10 and 20 min is shown in the frame labeled "predictions" in Figure 5.

Results

The performance of MIGFA has been scored against human interpretations of the same input radar data. Implicit in this statement is the assumption that human interpretations are 100% accurate. As we will see later, this assumption is not always correct.

The human interpreter had access to both Doppler and reflectivity images for an entire sequence of ASR-9 WSP scans, which could be viewed separately or in sequence as a movie. For each scan, a description of "truth" (i.e., the interpretation of the scan by a human) was stored in a table as a list of coordinates marking the gust front end points and an intermittent sampling of points in between. For categorization of results, the estimated maximum wind shear in the zone of convergence was also stored. This scoring exercise was intended to measure MIGFA's detection performance, not the end-to-end gust front detection

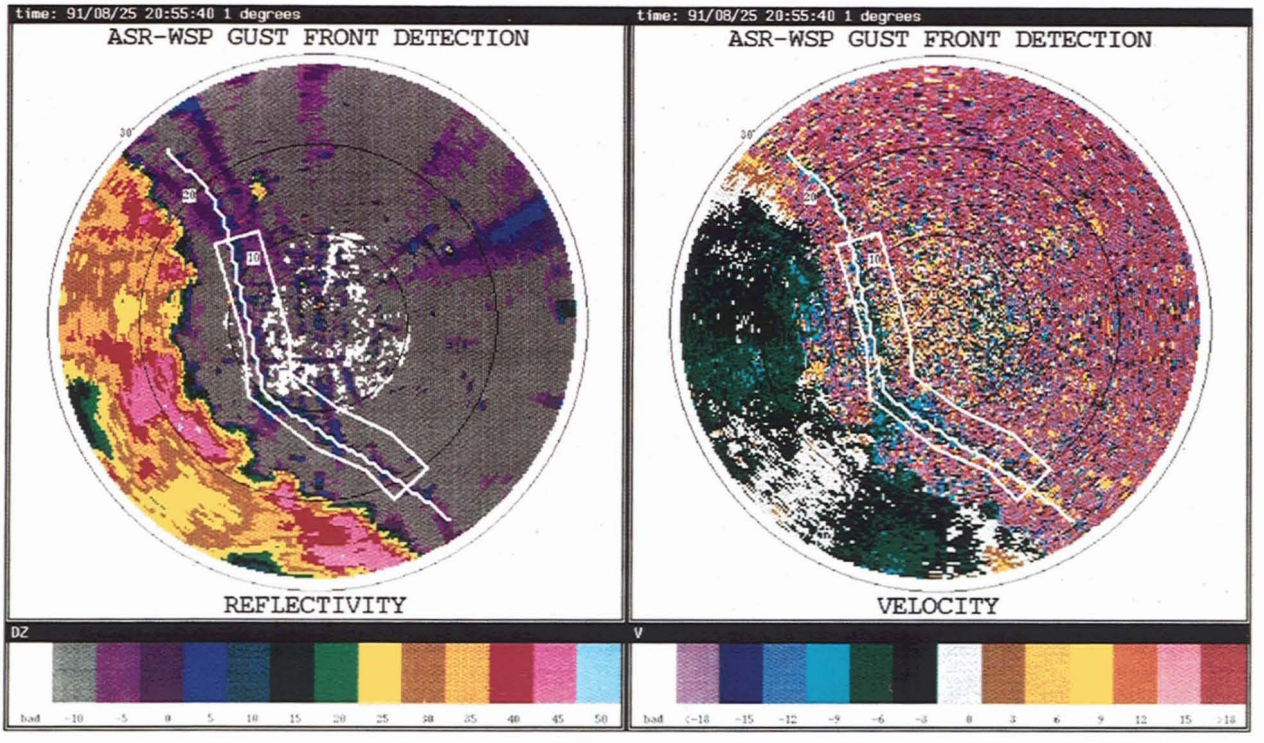


FIGURE 10. Human versus MIGFA interpretation of ASR-9 WSP data. The 5-km-wide box denotes a region where a human interpreter has detected a gust front. The single line represents a detection by MIGFA. Note that the human interpreter did not include the extreme ends of the front because the ends were nearly radially aligned and had weak reflectivity values—characteristics of out-of-trip weather. However, because the extended thin line moved consistently with the center of the front and because the variance of Doppler velocity values associated with the thin line was too low to be out-of-trip weather, MIGFA probably gave the more likely interpretation. The reflectivity is given in dBZ, and the velocity in m/sec.

capability for the ASR-9 WSP. Consequently, the human interpreter was restricted to including in the truth set only those gust fronts which had some visible signature, however subtle. Other data sources, such as matching TDWR data and anemometer measurements of winds over the radar site, were used to confirm or deny the existence of gust fronts that had an ambiguous appearance in ASR-9 WSP data. The interpreter, however, did not use these other data sources to define gust fronts in the absence of visible ASR-9 WSP signatures. For cases in which MIGFA detections in ASR-9 WSP data were scored against a human interpreter looking at TDWR data, the same procedures were used to generate the TDWR truth tables.

An automatic scoring procedure, described in detail by D. Klingle-Wilson et al. [12], compares computed gust front detections with human-generated truth (see Figure 10). Briefly described, the scoring algorithm draws lines that connect the sequence of coordinates encoding the human-estimated limits of a gust front. The lines are then expanded to a 5-km-wide region that is called, in this article, a truth box. Computed gust front detections overlapping with some portion of the truth box are counted as successful detections while those not overlapping are counted as false alarms. A probability of detection (POD) is computed by dividing the number of successfully detected fronts by the number of fronts identified by the human interpreter. The probability of a false alarm (PFA) is the number of false alarms divided by the total number of algorithm-generated detections. (Note: In this article, POD and PFA values will be expressed as percentages.) In addition to the hit-or-miss POD and PFA scores, scoring is also done in terms of the percent overlap of computer-generated detections and truth boxes. The percent length detected (PLD) is the number of points in an algorithm-generated detection that fall within a truth box divided by the length of that truth box (in pixels). The percent false length detected (PFD) is the number of points in an algorithm-generated detection that fall outside any truth box divided by the total number of algorithm-generated gust front points.

One improvement to this method is the use of a MAYBE category of truth. Often gust fronts or parts of gust fronts are only marginally detectable, forming a gray area in which the human observer is undecided or uncertain. If an algorithm detects a weak gust front associated with an ambiguous signature, the detection should not count as a false alarm. Similarly, if the algorithm misses a gust front that is too weak to have any operational significance, the miss should not affect the POD and PLD scores. Radar image features that are categorized as MAYBE are omitted from scoring.

Table 1 compares the performance of MIGFA against the latest version of AGFA, which uses more conventional methods of signal processing and computer vision. The test set of ASR-9 WSP data collected in Orlando, Florida, during field testing in 1991 contained nine different moderately strong gust fronts tracked through 15 hours (372 images). A human interpreter looking at the same data detected 280 instances of the nine gust fronts. The first two columns of Table 1 indicate that MIGFA increased

Table 1. AGFA and MIGFA Performance* on ASR-9 WSP Data

	Gust Fronts		*Gust Front Length*	
	Probability of Detection (POD)**	Probability of a False Alarm (PFA)**	Percent Length Detected (PLD)	Percent False Length Detected (PFD)
Baseline (AGFA)	56.7	4.6	38.9	12.9
MIGFA	88.1	0.6	86.2	33.4

* As scored against human interpretations of ASR-9 WSP data

** Expressed as a percent

Table 2. AGFA and MIGFA Performance* on ASR-9 WSP Data

	Gust Fronts		*Gust Front Length*	
	Probability of Detection (POD)**	Probability of a False Alarm (PFA)**	Percent Length Detected (PLD)	Percent False Length Detected (PFD)
Baseline (AGFA)	42.6	3.2	21.0	4.2
MIGFA	75.1	0.0	58.7	6.4

* As scored against human interpretations of matching TDWR data

** Expressed as a percent

by more than 50% the number of fronts detected by AGFA, while decreasing the false-alarm rate. Similarly, the PLD scores (column 3) indicate an improvement in detection performance. The increase in PFD (from 12.9% to 33.4%), however, appears to suggest that MIGFA is not as good as AGFA at discriminating the extent of individual fronts.

For a better understanding of why MIGFA was extending fronts beyond what the human interpreter believed appropriate, we examined several cases in which the PFD was high. In most of those cases, we found the extra points that MIGFA included in the gust front detections were believable. For example, Figure 10 shows a gust front truth box that overlays a MIGFA-generated detection. The human interpreter was reluctant to include the extreme ends of the front because the ends were nearly radially aligned and had weak reflectivity values—characteristics of out-of-trip weather. However, because the extended thin line moved consistently with the center of the front and because the variance of Doppler velocity values associated with the thin line was too low to be out-of-trip weather, MIGFA probably gave the more likely interpretation of the scene.

To substantiate such anecdotal observations, we took the gust fronts that MIGFA and AGFA had detected in ASR-9 WSP data and scored the fronts against human interpretations of TDWR data that had been taken at the same time. Although the resulting scores (Table 2) support the general trend of the first three columns of Table 1, the PFD for MIGFA (6.4%) is now roughly the same as that for AGFA (4.2%). Because gust fronts are more readily observable in TDWR imagery, we assume that the TDWR truth (i.e., the TDWR data as interpreted by a human) is more accurate than the ASR-9 WSP truth (i.e., the ASR-9 WSP data as interpreted by a human). Thus the difference between the PFDs as scored against ASR-9 WSP and TDWR truths crudely approximates the percentage of detected gust front points missed by the human interpreter. For MIGFA, this difference (33% – 6% = 27%) added to the PLD scored against the ASR-9 WSP truth (86%) is 113%; i.e., MIGFA's performance was 13% better than that of the human interpreter. For AGFA, the comparable result is 13% – 4% + 39% = 48%.

MIGFA was installed at the ASR-9 WSP site at Orlando International Airport in the spring of 1992 and was part of a formal operational test from 8 July to 20 September. During this time, gust front detections and predictions were relayed to air traffic controllers for their use in planning air traffic operations. During the early part of the summer, several minor problems and algorithm deficiencies were identified, and several fixes and enhancements were added during the middle of July. Careful interpretation, or "truthing," of the ASR-9 WSP data by a human was done from 1 August to 20 September.

As with the off-line testing described earlier, the on-line performance was scored against human interpretations of the same data. Table 3 shows the performance statistics for the test period. In general, the on-line test results substantiate the off-line results. Not surprisingly, the POD (75%) and PLD (81%) were somewhat lower than for the off-line test results shown in Table 1. Most of this differ-

Table 3. MIGFA Results* on ASR-9 WSP Data

	Gust Fronts		*Gust Front Length*	
	Probability of Detection (POD)**	Probability of a False Alarm (PFA)**	Percent Length Detected (PLD)	Percent False Length Detected (PFD)
MIGFA	75.4	1.8	80.8	21.1

* Results are scored against human interpretations of the same ASR-9 WSP data

** Expressed as a percent

Note: The data are for the period 1 August to 20 September 1992 in Orlando, Florida

ence can be explained by two problems.

First, several gust fronts had reflectivity values at or below the sensitivity limits of the ASR-9. Of course, those fronts with reflectivity values below the ASR-9 limits were missed by both MIGFA and the human interpreter. But there were a few cases of marginal contrast in which the human could detect a gust front while MIGFA had not accumulated enough confidence to declare an alarm. Note that, unlike MIGFA, the human interpreter had the opportunity to examine the sequence of radar images repeatedly and could use information from scans late in the sequence to confirm or deny the existence of the gust front in early scans. Not much can be done to overcome the sensitivity limits of the ASR-9. In most (but not all) cases, however, gust fronts with marginal reflectivity levels were associated with weak wind shears. Because these weak fronts had a minimal impact on airport operations, a failure to detect such fronts was not a significant liability.

The second problem was that several gust fronts were missed due to obscuration. In these cases, storm regions or out-of-trip weather were extensive enough to hide or fragment the thin-line signatures so that some gust fronts were detected late, dropped early, or sometimes missed altogether.

The PFA (1.8%) represents 19 false detections out of 1080 total detections generated by MIGFA in more than 14,000 scans processed. The high PFD (21.1%) is almost entirely the result of MIGFA's extending gust fronts beyond the ends delimited by the human interpreter. With the use of anticipation based on prior tracking data, MIGFA was able to extend the detected gust front lengths through areas where the signatures appeared ambiguous. As was seen with the off-line testing described earlier, a case-by-case analysis indicates that most of these extensions were in fact justified even though they were inappropriately scored as false lengths. Rescoring the results against TDWR data should improve the PFD score.

Another way to assess detection performance is to score only those gust fronts which had an impact on airport operations. From 20 July to 20 September, 14 convergent wind shears of greater than 15 kn were recorded on the anemometer network at the airport. Two of the wind shears were the result of short-lived localized winds beneath storm regions that were directly over the airport. The cause of a third wind shear could not be determined for certain, but was probably due to a microburst that was reported at the south end of the airport just as the wind shear was recorded. In none of these three instances could human interpreters find evidence of gust fronts in the ASR-9 data.

Of the 11 remaining wind shears, which were all verified later as gust fronts by human interpreters, MIGFA correctly tracked eight at least up to (but not always over) the airport. In the eight cases, air traffic controllers were given initial warnings from 18 to 79 min prior to the arrival of the front. Of the three missed gust fronts, one was occluded by fast-moving storm regions that were trailing the front. The second missed gust front had a very weak fragmented thin-line signature that was missed both by MIGFA and the human operators at the radar site who were logging weather and system activity. The third missed

case was a young gust front that had been generated by a large microburst only 5 km away from the runways. Because of its youth, the gust front had not yet developed a thin-line signature. Human interpreters who studied the radar scans after the testing was completed could find no evidence of this particular gust front in the ASR-9 WSP data, but could see a small zone of convergence without a corresponding thin-line signature in the data from TDWR. In summary, although MIGFA correctly detected and tracked (up to the airport) 8 out of 11, or 73%, of the gust fronts that had an impact on airport operations (wind shear > 15 kn), human operators working at the radar site were able to log 9 out of 11, or 82%, of the same gust fronts.

False gust front detections that are reported to be approaching an airport can also adversely affect airport operations. If a false alarm were trusted, inappropriate changes in airport operations planning might be made and the resulting delays could be just as bad as when a gust front is missed. During the test period, three incoming events—covering a combined time of 24 min (12 scans)—were scored as false alarms. Only one event generated a false wind-shear hazard alert (wind shear > 15 kn). All three were probably the result of thin lines from stratiform rain. None of these false alarms should have influenced airport operations planning because in each case tracking was dropped when the estimated time of arrival at the airport was more than 40 min.

Evaluation

Using the same input ASR-9 WSP data, we have shown by direct comparison that MIGFA provides a substantial improvement over AGFA in detection performance. We have also provided indirect evidence suggesting that, given the same input data, MIGFA may be nearly as good as human interpreters. However, the absolute reported POD scores for MIGFA (88% when scored against ASR-9 truth and 75% when scored against TDWR truth) are potentially misleading and should be regarded with caution because the dataset used for comparison testing was relatively small and from only one season at one site. Thus the off-line test probably did not contain a good representative sampling of gust fronts. The test did, however, provide a reasonable basis for comparing MIGFA against the older algorithm.

The results for the operational test period should be more representative of MIGFA performance. In the on-line testing, the POD and PLD scores remained high (in fact, the scores were only somewhat lower than those reported for the off-line testing), but an apparent problem in the relatively high PFD score (21%) persisted. Again, as was shown in the initial off-line testing, many of the false detections were in fact weak gust fronts or parts of gust fronts that the human interpreter had overlooked. Although these results have not been rescored against TDWR truth, the existences of gust fronts were established for several cases by the examination of matching TDWR or anemometer data.

An analysis of results accumulated during the 1992 operational test period has identified three main classes of failure modes for the ASR-9 WSP version of MIGFA. The failures within the first class are a direct result of the limited sensitivity of the ASR-9. Some gust fronts that were visible in TDWR data and that had an impact on the Orlando airport with moderate wind shear had reflectivity returns below the sensitivity of the ASR-9. Like MIGFA, experienced human observers using ASR-9 data did not see such gust fronts, although with the benefit of hindsight the observers could sometimes detect above-threshold fragments of what must have been the approaching front. In general, gust fronts with thin-line signatures that have reflectivity levels at or below the sensitivity limits of the ASR-9 usually (but not always) exhibit weak wind shears, making them operationally less significant.

The second failure mode was due to a lack of reliable Doppler estimates of velocity in clear air. Because of the unreliability of these values, the ASR-9 version of MIGFA had to rely on thin-line signatures for detecting gust fronts. As discussed earlier, however, not all thin lines are caused by gust fronts. For example, elongated low-reflectivity storm echoes associated with extensive areas of stratiform rain moving with the ambient wind were a source of false alarms in the operational testing. Because the reflectivity levels of light-rain echoes overlap with the range of reflectivity levels exhibited by gust fronts, the

thin-line feature detectors produced high interest values. In most of these cases, the thin-line features associated with the stratiform rain were transient and did not accumulate enough confidence through time for the system to declare a gust front. Some false alarms could be dismissed because of the lack of implicit convergent wind shears, which were computed by comparing the radar-measured winds in incoming candidate gust fronts with the winds measured by the airport anemometers surrounding the radar site. In at least one case, however, a false alarm could not be rejected with this criterion. The winds at the airport were variable and not representative of the winds immediately in front of the feature, which was 15 km away from the airport.

The third failure mode was caused by obscuration. During the 1992 operational test period, several gust fronts were either detected late, prematurely lost, or not detected at all due to obscuration by patches of high reflectivity that were caused by storms, range-ambiguous echoes, or ground clutter. Even in places where the thin-line features were visible, such patches of high reflectivity had sometimes fragmented the features into short segments. One missed gust front is known to have had an impact on the airport with a wind shear greater than 15 kn.

Experience gained from the operational test period has led to the implementation of a partial solution to the obscuration problem. The solution uses anticipation and the system's ability to detect obscuring weather patterns. Given a sequence of images, there often exists some time interval when a significant part of the gust front is not obscured and tracking can be initiated. Once sufficient confidence has accumulated, the system begins to anticipate where the gust front ought to be in the next scan. In normal operation, the thin lines of increased interest in the anticipation interest image are used to boost weak signals that would otherwise be below threshold for detection. (During the operational testing, obscuration suppressed all interest, eliminating any signals for anticipation to confirm.) In the modified system, when obscuring weather is found to overlap the anticipated gust front locations, the anticipation interest values can be increased to a level at which detection is triggered regardless of how weak the other evidence is. In other words, when obscuration is detected, the anticipation interest image becomes absolute, resulting in spatially restricted coasting.

Summary

The identifying signatures for gust fronts—thin lines of increased reflectivity, boundaries of converging Doppler values, and motion perpendicular to the thin lines and convergence boundaries—are conceptually easy to define and exploit as the basis of detection algorithms. And yet, although several research groups have worked collectively for nearly 10 years to develop reliable automatic gust front algorithms, none of the algorithms has demonstrated performance comparable to the ideal of human performance.

The problem is that automatic gust front detection, like other applications in computer vision, is deceptively much more difficult than the task of simply finding one or more signatures. Human observers use a variety of perceptual skills that have been notoriously and surprisingly difficult to implement in computer-vision systems. For example, humans have a talent for dealing with uncertain, ambiguous, and even contradictory evidence. Humans use specific knowledge of the object being sought and the context of observation as well as the object's spatial and temporal context. Unlike most other computer-vision and automatic target recognition (ATR) methodologies, the Experimental Target Recognition System (XTRS) and the Machine Intelligent Gust Front Algorithm (MIGFA) do not rely on machine intelligence only at the higher symbolic levels of processing. XTRS provides a framework for applying knowledge at the level of raw data by using specialized techniques for knowledge-based signal processing and pixel-level processing of evidence. The fact that MIGFA performance is competitive with that of human observers is at least partially due to this use of low-level machine intelligence.

Acknowledgments

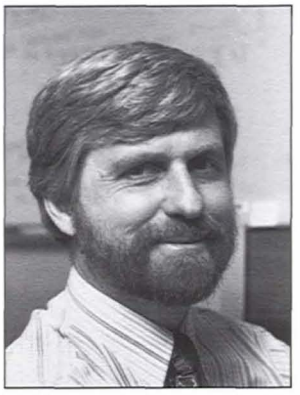
The authors would like to thank the staff of the FL-3 ASR-9 radar site in Orlando, Florida, who were responsible for running MIGFA, recording data, and logging results during the operational test period in 1992. These people include Wes Johnston, Craig

McFarland, Jeff Boisseau, and Cindy Meuse. The authors would also like to thank Joe Cullen for generating "truth" from ASR-9 WSP data, against which all algorithm results have been scored. XTRS, the prototype object recognition system upon which MIGFA is based, has evolved over several years of collaboration with Jacques Verly and Dan Dudgeon.

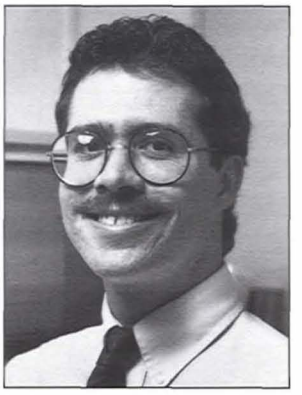
This work was sponsored by the Federal Aviation Administration.

REFERENCES

1. J.G. Verly, R.L. Delanoy, and D.E. Dudgeon, "Machine Intelligence Technology for Automatic Target Recognition," *Linc. Lab. J.* **2,** 277 (1989).
2. H. Uyeda and D.S. Zrnić, "Automated Detection of Gust Fronts," *J. Atmos. Oceanic Tech.* **3,** 36 (1986).
3. S. Smith, A. Witt, M. Eilts, L. Hermes, D. Klingle-Wilson, S. Olson, and J.P. Sanford, "Gust Front Detection Algorithm for the Terminal Doppler Weather Radar Part I: Current Status," *Proc. 3rd Intl. Conf. on the Aviation Weather System, Anaheim, CA, Jan. 1989,* p. 31.
4. L. Hermes, A. Witt, S. Smith, D. Klingle-Wilson, D. Morris, G. Stumpf, and M. Eilts, "The Gust Front Detection and Wind Shift Algorithms for the Terminal Doppler Weather Radar System," *J. Atmos. Oceanic Tech.* (to be published).
5. M. Eilts, S. Olson, G. Stumpf, L. Hermes, A. Abrevaya, J. Culbert, K. Thomas, K. Hondl, and D. Klingle-Wilson, "An Improved Gust Front Detection Algorithm for the TDWR," *Proc. 4th Intl. Conf. on the Aviation Weather System, Paris, June 1991,* p. J37.
6. M.W. Merritt , D. Klingle-Wilson, and S.D. Campbell, "Wind Shear Detection with Pencil-Beam Radars," *Linc. Lab. J.* **2,** 483 (1989).
7. T.A. Noyes, S.W. Troxel, M.E. Weber, O.J. Newell, and J.A. Cullen, "The 1990 Airport Surveillance Radar Wind Shear Processor (ASR-WSP) Operational Test at Orlando International Airport," *Project Report ATC-178,* MIT Lincoln Laboratory (July 1991), AD-239852 (NTIS only).
8. M.E. Weber, "Airport Surveillance Radar (ASR-9) Wind Shear Processor: 1991 Test at Orlando, FL," *Project Report ATC-189,* MIT Lincoln Laboratory (June 1992), AD-252246 (NTIS only).
9. R.L. Delanoy, J.G. Verly, and D.E. Dudgeon, "Pixel-Level Fusion Using Interest Images," *Technical Report 979,* MIT Lincoln Laboratory (26 Apr. 1993).
10. S. Grossberg and E. Mingolla, "Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations," *Perception and Psychophysics* **38,** no. 2, 141 (1985).
11. S. Levialdi, "Parallel Pattern Processing," *IEEE Trans. Syst. Man Cybern.* **1,** 292 (1971).
12. D.L. Klingle-Wilson, M.F. Donovan, S.H. Olson, and F.W. Wilson, "A Comparison of the Performance of Two Gust Front Detection Algorithms Using a Length-Based Scoring Technique," *Project Report ATC-185,* MIT Lincoln Laboratory (May 1992).



RICHARD L. DELANOY
is a staff member of the Machine Intelligence Technology Group. His work spans the fields of computer vision, machine learning, and construction of object-recognition systems. From 1980 to 1983, he was a research scientist at the University of Virginia Department of Psychology, where he investigated the biochemical correlates of learning and the effects of stress-related hormones on electrophysiological models of memory. Before joining Lincoln Laboratory in 1987, he worked for GE Fanuc Automation N.A., Inc., as a software engineer developing numerical and programmable controllers for manufacturing automation. Dick received a B.A. degree in biology from Wake Forest University in 1973, a Ph.D. degree in neuroscience from the University of Florida College of Medicine in 1979, and an M.S. degree in computer science from the University of Virginia in 1987. He was a National Science Foundation Predoctoral Fellow and a National Institute of Mental Health Postdoctoral Fellow.



SETH W. TROXEL
received a B.S. degree in meteorology from San Jose State University, California, in 1983 and went on to work as a meteorologist in the Atmospheric Lidar Group of the NOAA Wave Propagation Laboratory in Boulder, Colorado. In 1987, he joined Lincoln Laboratory as a software engineer and meteorologist with the Weather Sensing Group. Since coming to Lincoln Laboratory, he has been involved in the testing and development of hazardous-weather-detection capabilities for the Airport Surveillance Radar (ASR-9) and Terminal Doppler Weather Radar (TDWR). Seth's primary research interests are in the areas of remote sensing, software engineering, and design of algorithms for aviation weather products.