UNDERSTANDING SCENE DYNAMICS

Bir Bhanu
Honeywell Systems and Research Center
3600 Technology Drive
Minneapolis, MN 55418

ABSTRACT

The objective of our work in understanding scene dynamics is to develop robust techniques for target tracking and recognition from a moving robotic vehicle. The topics currently under investigation are: decomposition of complex vehicle motion; qualitative 3-D scene modeling; target motion detection and tracking; map-based target tracking; inertial sensor integrated obstacle detection; adaptive segmentation; 3-D target model acquisition and refinement; landmark recognition; and terrain interpretation. This paper summarizes the progress made in each of these areas during the period from February 1988 to March 1989. We also present a brief discussion on scientific experiments, machine learning for target recognition, and scientific performance evaluation of vision algorithms and systems.

1. INTRODUCTION

This paper provides an overview of the research performed by our group during the past year. Our research in understanding scene dynamics is directed towards knowledge-based interpretation of scene dynamics and model-based target recognition. The key accomplishments of our work during the past year are: We have developed qualitative reasoning and "3-1/2-D" modeling techniques for detecting and tracking moving targets from a mobile platform in simple curved road scenes. The concept of fuzzy focus of expansion, which allows a very accurate determination of the instantaneous direction of a moving vehicle and camera rotations along the two axes (pan and tilt only), has been demonstrated. We have also demonstrated the "dynamic model matching" concept for landmark recognition, where the model generation and matching process dynamically changes as a function of range to the landmark and perspective as viewed by a mobile platform. In addition, we have performed initial experiments in digital map integrated target tracking.

We have investigated the following major topics:

1. Qualitative motion detection and tracking,
2. Inertial sensor integrated motion analysis,
3. Machine learning for adaptive segmentation, target model acquisition, and target model refinement,
4. Dynamic model matching for landmark recognition, and
5. Hierarchical symbolic grouping for interpretation of terrain.

The synopsis of the technical achievements in each of these technical areas is given below. We also present a brief discussion on scientific experiments, machine learning for target recognition, and scientific performance evaluation of vision algorithms and systems.

One of the primary objectives of scientific performance evaluation is the establishment of a national research database of computer vision imagery. This database will be maintained in locations accessible to all members of DARPA’s IU community through a set of uniform access procedures. Another objective is the standardization of terminology, benchmarks, and a characterization of the computer vision research infrastructure. Also, we will define a set of techniques and models for algorithm/system performance evaluation of selected matured vision algorithms.

2. QUALITATIVE MOTION DETECTION AND TRACKING

We have developed a unique approach called DRIVE (Dynamic Reasoning from Integrated Visual Evidence) based on qualitative reasoning and modeling for target motion detection and tracking.3,4,14,15,17,18 The DRIVE system performs dynamic scene understanding needed to support the application of smart weapons and autonomous navigation of robotic vehicles. Instead of refining a single quantitative description of the observed environment over time, multiple qualitative interpretations of the scene are maintained simultaneously. This technique offers considerable flexibility over traditional numerical techniques which are often ill-conditioned and noise-sensitive. With DRIVE, an autonomous
system can (i) detect and classify moving objects in the scene, (ii) estimate the vehicle’s egomotion, and (iii) derive the 3D structure of the stationary environment.

The 3-D motion of targets is obtained from displacement vectors of point features without any knowledge about the underlying 3-D structure, discovering inconsistencies between the current state of the qualitative 3-D scene model and the changes actually observed in the scene, and by detecting moving edges and regions. DRIVE uses a new algorithm for computing the region of possible focus-of-expansion (FOE) locations in image sequences, called the fuzzy FOE. This computation is accomplished in a unique manner by separating the rotational and translational components of the vehicle’s motion and using a robust method for computing the displacement vector between two images using adaptive windows.

The 'fuzzy' FOE algorithm allows the direction of instantaneous heading of an autonomous land vehicle to be precisely determined within 1° using image information exclusively. The results obtained using ALV imagery taken at five different sites demonstrate the algorithm’s performance capabilities. This result has significant scientific importance for targeting applications. It allows the determination of self motion of moving imaging devices. Rotation in the horizontal and vertical directions (pan and tilt only) of ±5° or larger can be successfully handled by the algorithm. Moreover, it allows the use of passive approaches for surveillance activities that must detect and track moving targets and must detect and avoid obstacles using passive sensors mounted on a mobile platform.

Experiments have been carried out on 262 frames of ALV data taken at 5 different sites. Figures 1 and 2 illustrate the results.

Figure 1 shows the original image with the interesting points after edge detection and computation of the focus of expansion. Figure 2 shows the qualitative reasoning and modeling process. There are two cars in the images, one approaching the vehicle and the other receding from the vehicle. Both of the moving cars have been detected. The reasoning process is based on the changes in the expansion pattern and uses a camera model.

We have developed preliminary algorithms to integrate the DRIVE system with digital terrain elevation and land cover data. These algorithms provide information about the map location of the moving targets, the road label on which the targets are possibly traveling, and neighboring landmarks. Such information is desired for military applications and we have performed initial experiments to establish its usefulness in detecting moving targets in both high clutter and low contrast situations. Figure 3 shows an example of target detection under low contrast and high clutter situation. Target map location, the road segment on which the car is traveling, and nearby landmarks have also been detected by the algorithms which integrate map data with the motion algorithm suite.

The paper by Bhanu et. al. provides details of the interest point selection, disparity analysis, fuzzy FOE, qualitative scene model, map-based tracking, and edge/region based approaches.

3. INERTIAL SENSOR INTEGRATED MOTION ANALYSIS

Land navigation requires a vehicle to steer clear of trees, rocks, and man-made obstacles in the vehicle’s path while vehicles in flight, such as rotorcraft, must avoid antennas, towers, poles, fences, tree branches, and wires strung across the flight path. Automatic detection of these obstacles by passive sensors and the necessary guidance and control actions triggered by such detection would facilitate autonomous vehicle navigation.

Many types of existing vehicles contain inertial navigation systems (INS) which can be utilized to greatly improve the performance of several computer vision applications such as obstacle detection, target motion detection, target tracking, stereo, etc. and make them useful for practical military and civilian applications. As an example, motion analysis techniques can effectively use the output of an INS to improve interest point selection, matching of the interest points, and the subsequent motion detection, tracking, and obstacle detection.

We are using INS measurements to enhance the quality and robustness of motion analysis techniques for obstacle detection and thereby providing vehicles with new functionality and capabilities. Details of the INS integrated motion analysis for obstacle detection are given in the paper and reports by Bhanu, Roberts, and Ming.

4. MACHINE LEARNING FOR ADAPTIVE SEGMENTATION AND TARGET MODEL ACQUISITION/REFINEMENT

4.1 Adaptive Segmentation Using Genetic Algorithms

Image segmentation is typically the first, and most difficult, task of any automated image understanding process. All subsequent interpretation tasks, including feature extraction, object detection, and object recognition, rely heavily on the quality of the segmentation process. Despite the large number of segmentation techniques presently available, no general methods have been found which perform adequately across a diverse set of imagery. Only after numerous modifications to an algorithm’s control parameter set can any current method be used to process the wide diversity of
Figure 1: Direction of instantaneous heading for a moving platform is precisely given within 1°, which is a greater accuracy than human visual performance.

Significance of Honeywell's Results

Targeting Applications Perspective:
- Self-motion of moving imaging devices can be accurately obtained. This includes rotation of ±5° or larger in horizontal and vertical directions.
- Purely quantitative approach is not suitable for Target Motion Detection and Tracking from a mobile platform—we use a qualitative approach in our DRIVE system.
- Use of passive search for surveillance activities.

Human Perception Perspective:
- Human observers find it difficult to determine the exact direction of heading. Average deviation of human judgment from the real direction has been reported to be as large as 10° and up to 20° in the presence of large rotation. Our algorithm can provide the direction within 1°.

* These results are based on limited 262 frames of ALV data taken at five different sites.
Original Image Sequence After Edge Detection, Point Features Are Shown by Numbers. Receding Car Is Shown by Point 24 and Approaching Car Is Shown by Point 33.

Displacement Vectors and Estimates of Vehicle Motion. Fuzzy Focus of Expansion Is Shown in the Shaded Area.

Both Car 33 and 24 Are Detected as Moving and Tracked

Qualitative 3-D Scene Model, "Closer" relations in 3-D are shown by links. Size of a node shows its distance from ALV
Target is coming down the road as shown by 0 at a distance.

Results of DRIVE Target Motion Detection and Tracking System

Qualitative 3-D Scene Model—"Closer" Relationships Are Established in 3-D. Point 391 is the Moving Vehicle

Focus of Expansion—Instantaneous Heading Is Shown by 0 in the Shaded Area

- Target Map Location
- Road Segment Identification
- Nearby Landmark Localization
images encountered in dynamic outdoor applications such as the operation of an autonomous robotic land/air vehicle, automatic target recognizer, or a photointerpretation task.

The image segmentation problem can be characterized by several factors which make parameter selection process very difficult. These factors include numerous control parameters, lack of segmentation model, and problems associated with the evaluation of segmentation.

We are using a machine learning technique known as a genetic algorithm, to perform adaptive segmentation in a closed loop feedback system. Genetic algorithms allow the segmentation process to adapt to changes in image characteristics caused by variable environmental conditions such as time of day, time of year, clouds, rain, etc. The genetic algorithm efficiently searches the enormous hyperspace of segmentation parameter combinations using a collection of search points known as a population. By combining high performance members of the current population to produce better parameter combinations, the genetic algorithm is able to locate the parameter set which maximizes the segmentation quality criteria. The paper by Bhanu, Lee, and Ming provides details of the adaptive image segmentation process.

4.2 Target Model Acquisition and Refinement

A major technology gap in state-of-the-art model-based object recognition for outdoor scenes is the process of model (natural or man made) acquisition. Generally man made object models are fixed and they do not have any learning capability; therefore, they are not adequate by themselves for object recognition in dynamic environments.

Due to recent advances in machine learning technology, some of the problems encountered in the target recognition domain seem to be resolvable. Learning allows an intelligent recognition system to use situation context in order to understand images. This context, as defined in a machine learning scenario, consists of a collected body of background knowledge as well as environmental observations which may impact the processing of the scene.

Machine learning facilitates two main breakthroughs in the target recognition domain: automatic knowledge base acquisition and continuous knowledge base refinement. The use of learning in the knowledge base construction will save the user from spending the enormous amount of time necessary to derive target models and databases. Knowledge base refinement can then be incorporated to make any necessary changes to improve the performance of the recognition system. These two modifications alone will serve to significantly advance the present abilities of most target recognition applications.

We are developing a TRIPLE: Target Recognition Incorporating Positive Learning Expertise system for automated model acquisition and refinement. The system uses a multi-strategy technique; two powerful learning methodologies are combined with a knowledge-based matching technique to provide robust target recognition. Using dynamic models, TRIPLE can recognize targets present in the database. If necessary, the models can be refined if errors are found in the representation. Additionally, TRIPLE can automatically store a new target model and recall it when that target is encountered again. Finally, since TRIPLE uses qualitative data structures to represent targets, it can overcome problems such as image noise and occlusion.

The two main learning components of the TRIPLE system are Explanation-Based Learning (EBL) and Structured Conceptual Clustering (SCC). Explanation-based learning provides the ability to build a generalized description of a target class using only one example of that class. Structured conceptual clustering allows the recognition system to classify a target based on simple, conceptual descriptions rather than using a predetermined, numerical measure of similarity. While neither method, used separately, would provide substantial benefits to a target recognition system, they can be combined to offer a consolidated technique which employs the best features of each method and is very robust. The paper by Bhanu and Ming provides more details of the TRIPLE system for target model acquisition and refinement.

5. DYNAMIC MODEL MATCHING FOR LANDMARK RECOGNITION

We have developed a technique called PREACTE (Perception REAsoning ACTion and Expectation) based on dynamic model matching for landmark recognition from a mobile platform. The technique can recognize landmarks and other objects from partial and complete views in dynamic scenarios. It relies on the generation of multiple landmark descriptions from 3D models based on different estimated ranges and aspect angles. These descriptions are a result of feature, spatial, and geometric models of a single landmark. Expectations about the landmarks (appearances) vary dynamically as the autonomous robot approaches the landmark. Dynamic Model Matching also includes the generation of specific landmark recognition planning strategies whereby different features of different landmarks are emphasized at varying ranges. It is an expectation driven, knowledge-based approach and uses limited map information for updating the ALV's location in the map.

Figure 4 illustrates an example of dynamic model matching for landmark recognition from a mobile platform. Note that landmark recognition allows the determination of the ALV's position within 3 feet compared to an inertial position error of 105 feet over a distance of one mile. Figures 5 and 6 provide three examples of landmark recognition.
(a yellow gate and two wooden gates) taken at two different times at the Martin Marietta test site. It is to be noted that segmentation results affect the recognition results. Being very close to a landmark does not necessarily mean that its segmentation will be better than the segmentation of the image acquired at a greater range.

6. HIERARCHICAL SYMBOLIC GROUPING FOR INTERPRETATION OF TERRAIN

An autonomous robotic vehicle must be able to navigate not only on the roads, but also through terrain in order to execute its missions of surveillance, search and rescue, and munitions deployment. To do this the vehicle must categorize the terrain regions it encounters as to the traffickability of the regions, the land cover of the regions, and region-to-map correspondence. Our approach for terrain interpretation employs a robust texture-based algorithm and a hierarchical region labeling scheme for ERIM 12 channel Multi-Spectral Scanner data. The technique, called HSGM (Hierarchical Symbolic Grouping for Multi-spectral data), is specifically designed for multi-spectral imagery, but is appropriate for other categories of imagery as well. For this approach, features used for segmentation vary from macro-scale features at the first level of the hierarchy to micro-scale features at the lowest level. Examples of labels at the macro-level are sky, forest, field, mountain, road, etc. Figure 7 shows texture gradient images, and the final region boundaries for large regions. These regions are labeled using a knowledge-based classifier.

For each succeeding level of the hierarchy, the identified regions from the previous stage are further subdivided, if appropriate, and each region’s labeling is made more precise. The process continues until the last stage is reached and no further subdivision of regions from the preceding stage appears to be necessary. Examples of region labels for this level of the hierarchy are gravel road, snowberry shrub, gambel oak tree, rocky ledge, etc. Further development of the technique will employ multiple sources of a priori information such as land cover, terrain elevation map information, range data, seasonal information, and time of day.

Details of the HSGM technique with results and examples from real imagery are given in papers by Bhanu and Symoscek.11,12

7. VISION-BASED TARGETING EXPERIMENTS

As discussed earlier, we have developed two key algorithm suites, called DRIVE (Dynamic Reasoning from Integrated Visual Evidence) and PREACTE (Perception, REASONing, ACTion and Expectation). DRIVE accomplishes target motion detection and tracking while PREACTE performs landmark recognition. We plan to advance this research by performing a set of scientific experiments directed towards a practical mobility and targeting application of a robotic combat vehicle.

We plan to conduct scientific experiments in two areas: landmark recognition for path traversal and target motion detection and tracking. Two series of experiments are planned, one in each of these areas. Each experimental series begins with data collection and proceeds through progressively more difficult scenarios. The final experiments in the series will be characteristic of practical mobility scenarios for a robotic combat vehicle. For both series of experiments, the vehicle will be in continuous motion.

Landmark recognition experiments include laboratory landmark recognition tests using off road data; non-real time landmark recognition in off road traversal by the ALV; real time dynamic landmark recognition in off road traversal by the ALV; and dynamic landmark model learning with return path traversal. Motion detection and tracking experiments involve verifying motion results against land navigation data; non-real time detection of multiple moving objects while maintaining reasonable rotation components of the vehicle; real time detection of multiple moving objects; integrating ETL map data with target motion detection and tracking; and advanced experiments carried out under more difficult visual scenes involving low contrast and high clutter.

We also plan to develop a flexible software architecture and the associated software for “real time” instrumentation and evaluation of the landmark recognition and the motion detection and tracking algorithms. Some of the important aspects of this work involve defining the criteria for evaluation and acquiring, retrieving, and presenting the desired information in meaningful ways so as to provide insight into the associated vision algorithms.

8. MACHINE LEARNING FOR TARGET RECOGNITION

Present target recognition systems are unable to adapt to changes in environmental conditions, target variations, and the unexpected appearance of new targets. Each of these situations affects the appearance of the targets in the image, which in turn, degrades the overall performance of current generation recognition system.

One of the key challenges to automating the target recognition process is that of automatically responding to changes occurring in the targets seen in an image. We address this problem at every stage of the multi-level vision problem by a unique multi-strategy machine learning approach not available in any current model-based recognition system. We want to show that significant benefits can accrue through applying machine learning technology to
Figure 4: Example of Dynamic Model Matching for landmark recognition from a mobile platform.
Figure 5: Original image sequences from three ALV test sites.
Figure 6: Dynamic model matching results for the three sites shown in Figure 5. Recognition accuracy for each landmark and the resulting site uncertainty are indicated.
Figure 7: Texture gradient images and region boundaries using multispectral imagery
automatically acquire new target models and update their descriptions; to learn new target features based on perceptual cues; and to adapt segmentation parameters using genetic algorithms.\(^7\)

Through an in-depth analysis, performed by Honeywell\(^2\) on the applicability of state-of-the-art machine learning technology to model-based vision, we have developed the concepts for a novel machine learning system, called ORACLE (Object Recognition Accomplished through Consolidated Learning Expertise). The ORACLE system incorporates explanation-based learning, structured conceptual clustering, genetic algorithms, and learning by examples into a multi-strategy learning approach for automated target recognition. At the high level of computer vision, ORACLE utilizes the characterization and aggregation capabilities of explanation-based learning, structured conceptual clustering, and similarity-based learning in the target recognition and learning process. By combining these three learning systems, ORACLE overcomes the inherent limitations of the individual techniques and provides solutions to practical problems in model-based vision technology such as the need for automated model acquisition and refinement. During the intermediate level vision processing, ORACLE uses explanation-based learning with a perceptual cue database to acquire new target features. Finally, at the low level of vision, ORACLE uses genetic algorithms for parameter adaptation capability. Thus, the ORACLE system provides a learning capability at all the three levels of vision: low, intermediate and high.

9. SCIENTIFIC PERFORMANCE EVALUATION OF VISION ALGORITHMS AND SYSTEM

At present, very little work has been performed in the area of evaluation for image understanding algorithms and systems. In the DARPA-sponsored image understanding research, a wide variety of algorithms and systems are being developed for photointerpretation, navigation, manufacturing, cartography, and targeting applications. Scientific (both quantitative and qualitative) performance evaluation of diverse vision algorithms and systems will help in advancing the computer vision field at a faster pace, which in turn will lead to the most rapid fielding of computer vision technology. Effective performance evaluation will allow the measurement of not only the qualitative advancements in the computer vision field, but will also help to quantify the progress in the field. Most importantly, scientific performance evaluation will provide more rapid technology transfer (see Figure 8) to DoD applications by reducing the time needed to develop and validate robust vision algorithms. Figure 9 and 10 show algorithm evolution cycle and the evolution of algorithms. The four phases of the evolution cycle are conceptualization, generation, evaluation and adaptation. Once the algorithms have shown their potential value, they can be subjected to automatic evaluation.

Life cycle of any technology consists of four phases as shown in Figure 11. The maturity of any area (application areas or low, intermediate or high levels) of image understanding is related to the degree to which agreement on benchmarks can be reached and performance evaluation can be conducted.

Objective of performance evaluation is not to find out "the best" algorithm, but quantitative/qualitative understanding of capabilities of algorithms and systems. Evaluation works best when it is not tailored for a particular implementation; time to run the test is short; no new systems are designed; it has extreme cases especially those that cause known algorithms to fail; it has an extensive set of test images; anyone can submit results; and researchers perform the test on their own system.

It is extremely important to develop quantitative performance criteria for image understanding algorithms for several reasons:

1. To compare various "matured" algorithms and systems and to predict their performance in a given scenario and/or for a specific application,
2. To study the behavior of an application system and its components under different conditions and parameter settings, so as to be able to find the optimum performance achievable and the performance bounds of its components,
3. To understand the characteristics of the imagery that affect the performance of the algorithms,
4. To find common functional elements for an application among the algorithms currently in use,
5. To help the algorithm developer choose the appropriate algorithms for his/her application and research, and
6. To provide an objective and complete evaluation methodology for standardization purposes.

Performance evaluation allows performance analysis (strengths/weaknesses), sensitivity analysis, performance models. All these lead to prediction of algorithms and prediction is an important element of science.

The critical ingredients for scientific performance evaluation are:

a. Image database groundtruth,
Figure 8: Scientific approach to technology evaluation.

Figure 9: Algorithm evolution cycle.

Figure 10: Evolution of algorithms.

Figure 11: Life cycle of technology development.
(b) Techniques for performance evaluation,
(c) Common system environments (KBVision and others).

Some of the problems with performance evaluation are:
1. Lack of appropriate database,
2. Slightly different problem statements and assumptions need different data sets,
3. Difficulty to quantify algorithm performance, many facets of the problem,
4. Interpretation of evaluation results (who?)

There are two solutions: natural evolution or concrete action to promote the maturation of IU tech base. We concentrate on the second solution.

We firmly believe that the effective characterization and prediction of algorithm performance is an essential step in transforming computer vision from an art to a science. The ability to successfully predict performance depends on a clear understanding of the complex relationship among the input data, algorithm operations, and produced results.

Through active interaction with the DARPA IU/SC community, the following objectives are pursued for scientific performance evaluation.

(1) National research database of computer vision imagery,
(2) Characterization (taxonomy) of vision research,
(3) Benchmarks for performance measurement of algorithms/systems (what to measure?),
(4) Techniques and models for "matured" algorithms and systems for performance evaluation (how to measure?),
(5) Workshop of DARPA IU community on performance evaluation.

The details of the above objectives are now discussed.

9.1 NATIONAL RESEARCH DATABASE OF COMPUTER VISION IMAGERY

The objective of establishing a national research image database is to promote the orderly development and dissemination of image information to serve the needs of DARPA IU algorithms/systems developers. This encompasses the standards for data interchange and activities for data collection, data organization, and data distribution.

The important considerations for these databases are: ground truth data requirements (site, sensor characterization, sensor platform, objects of interest, meteorological conditions), ground truth recording procedures, and database quantity/quality/variety requirements. The ground truth information is very critical and many times is not available or is too expensive to capture. Whenever the ground truth information is available, imagery should be partitioned into two categories: For some imagery, the ground truth is supplied to the researcher so that he/she can use them in the development of vision algorithms; the other category should be the imagery for which ground truth is sequestered and used to evaluate the robustness of the algorithms after development.

One potential use of an accepted imagery database would be for evaluating various "matured" algorithms that perform the same function (e.g., stereo, segmentation, motion detection, object recognition in range images, etc). Each year, the results of this evaluation, which have a well defined objective and scientific experiment, can be publicized at the Image Understanding Workshop and there would be recognition for researchers who demonstrate the best results using the "matured" algorithms for a given set of images (for a given application). More importantly, the overall progress made by the IU community would be made apparent.

Our short-term objective is to define and make available a standard set of images to be used by the DARPA IU community for algorithm development and evaluation. Some of the common functions are: stereo, motion algorithms, object recognition in outdoor/indoor scenes using TV and range images, etc.

Some of the currently available databases that may meet our needs are:

* Martin's Collage 1 and Collage 2 ALV Database:* Contains a large number of color TV, multispectral, and range images. Ground truth information is limited.

* USC Database:* Contains a large number of texture, aerial, and color images. Ground truth is limited.
CMU: NavLab, Calibrated Imaging Lab

Martin Marietta: ALV database

Utah Range Database: Contains four sets of 33 images (5.108 Megabytes). Data is available in Unix TAR format, 1600 BPI tape. This data includes:

University of Utah Images: The set is encoded in White Scanner format. Consists of images of bottles, cylinders, polygons, and various parts including the Renault auto part from INRIA. The parameters and details of the scanning systems are known for the set of images scanned at Utah.

SRI Images: Grapes and space shuttle images.

North Carolina State University: PC board, the image of Victor Hugo, and others.

ENST Paris: Victor Hugo image obtained in one degree aspect increments.

The sample image sets can be separated by class as well. In the case of stereo imagery, the photogrammetry community has distributed a very good stereo database (not epi-polar constrained) with known ground truth. Imagery acceptable for motion research is available from SRI, which has some image sequences that have good imaging information and some partial depth ground truth.

Note that database is closely tied up with evaluation (see Figure 12) Database should also be viewed as extensible, not the finished product. In summary some of the issues related with database are:

1. Models for evaluation,
2. Requirements of database,
3. Collection of database,
4. Organization of database,
5. Maintenance of database,
6. Access and usage expectations,
7. Groundtruth information - sensor, map, other ancillary information,
8. Standards for imagery and non-imagery information,
9. Types of data,
10. Specific IU algorithms, systems and applications.
We are working on a detailed plan for data acquisition and accessibility. We are identifying core data set and plan to expand it in a systematic way.

9.2 Characterization of Vision Research Infrastructure or Taxonomies of Vision Research

A current detailed taxonomy of vision research is desired which is based on diverse criteria such as:

- Applications (navigation, terminal homing, remote sensing, etc.),
- Class of sensors (TV, FLIR, Range, SAR, etc.),
- Use of multisensors (TV & range, FLIR & Range, etc.),
- Functionalities (segmentation, feature extraction, texture analysis/synthesis, etc.),
- Principles (top down, bottom up, etc.),
- Reasoning processes (qualitative, perceptual, etc.),
- Hardware architectures (systolic, array, cellular, etc.),
- Implementation techniques (VHSIC, VLSI, discrete, etc.), and
- Use of auxiliary information (digital elevation map data, land cover data, etc.).

The rationale for characterization is to help in the organization and development of image database, definition of benchmarks and methodologies for evaluation. This characterization will provide a common framework of terminology and description to promote improved communication among the members of the vision community and between technology developers and applies. Since the computer vision field is still quite young and undergoing rapid evolution, the proposed taxonomy should be viewed as a "snapshot" of the field today and will likely need to undergo significant modifications and extensions as the field progresses. After the development of the proposed taxonomy, the development of the other three related goals will be pursued: a common image database, general vision system benchmarks, and an effective methodology for performance evaluation. One can think of a very deep tree whose leaf nodes are very specific (for example, the segmentation of tank targets at "close" distances in range images for terminal homing applications). We associate the specific database, benchmarks, and methodology with these leaf nodes for performance evaluation.

9.3 Terminology and Benchmarks for Performance Evaluation

It is important to have common terminology and benchmarks for performance evaluation. Subtle differences in meaning can be very important for evaluation. Even terms such as "ground truth" mean different things to different people. A lexicon that establishes standard terminology and standard benchmarks will provide uniformity in carrying out scientific experiments for performance evaluation.

9.4 Scientific Methodology and Models for Performance Evaluation

Since one of the goals of computer vision is to build machines that can solve real world problems, we need to define the systematic methods and models for performance evaluation of individual vision algorithms (segmentation, feature computation, texture measurement, etc.) and systems (object recognition, vision-based navigation, etc.) for a particular application (terminal homing, surveillance, etc.). We need to thoroughly understand the practical experimental designs and the errors of observation and their treatment. As an example, the results of segmentation are still evaluated qualitatively. They need to be evaluated, in part, on the basis of how well the implicit or explicit model in the technique is able to predict performance. In other words, the quality of an algorithm should depend not only on certain test performance results, but on the accuracy of the model that predicts algorithm performance over a diverse database of imagery. If an algorithm performs well over a narrow database (a few images), but a resulting performance prediction model proves to be inaccurate over a larger database, the proper conclusion is that the overall algorithm performance is deficient. In this framework, evaluation of an algorithm consists of two components: an algorithm and the associated performance prediction model. We can refer to the combination of these two components as a generalized algorithm. Simple quantitative measures (which may or may not have intuitive physical significance), such as the number of pixels misclassified with respect to the true object, the correlation coefficient between the true and extracted object, mean square error between the true and extracted object, object-to-background contrast, and the metric based on these criteria, can be effectively used for segmentation evaluation.

Carrying this evaluation process a step further, we need an evaluation methodology for the evaluation of complete systems such as an object recognition system. The system performance should be evaluated on the basis of the task it is able to perform in a given environment considering such factors as sensor type, resolution of data, type of objects, and complexity and information content of the scene. It is logical to assume that to obtain the optimum
performance of the system, it is essential to achieve the maximum attainable performance of each of its components. However, it may or may not satisfy the goals of the system performance, since most of the image understanding components are nonlinear. Here a top-down approach for evaluation may be more meaningful than a bottom-up approach.

In summary, the emphasis of performance evaluation is on computer vision problems, scientific experimental design and interfaces between vision components and functions. We need to define a performance metric for each of the image understanding algorithms as well as a performance metric for the system as a whole. This can be done for the specific matured algorithms/systems being pursued by the Image Understanding community.

ACKNOWLEDGEMENTS

The work summarized in this report is supported by DARPA under contract DACA 76-86-C-0017, by the NASA under contract NAS2-12800, and other sponsors. I would like to thank Bob Simpson, Takeo Kanade, Ram Nevatia, Keith Price, Ed Riseman Tom Binford, Steve Shafer, Chuck Thorpe, Karen Olin, Bob Bolles, and Steve Savitt for helpful comments and suggestions on Scientific Performance Evaluation.

The following researchers have contributed to the work described in this report: P. Symosek, H. Nasr, W. Burger, J. Ming, B. Roberts, J. Kim, H. Lai, S. Schaffer, and D. Robohm.

REFERENCES


PROCEEDINGS:

Image Understanding Workshop

POWervision (c) 1989 ADS

OBJECT HYPOTHESES → OBJECT RECOGNITION

OBJECT MODEL

FEATURE EXTRACTION

OBJECT

Sponsored by:
Defense Advanced Research Projects Agency
Information Science and Technology Office

May 1989
Image Understanding Workshop

Proceedings of a Workshop
Held at
Palo Alto, California

May 23-26, 1989

Sponsored by:
Defense Advanced Research Projects Agency
Information Science and Technology Office

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

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

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