Conf-8903223--1

Paper presented at the Second National Symposium on Sensors and Sensor Fusion (Orlando, FA; March 27-31, 1989).

A HIERARCHICAL STRUCTURE APPROACH TO MULTISENSOR INFORMATION FUSION

Alianna J. Maren*
The University of Tennessee
Space Institute
Tullahoma, TN 37388

Robert M. Pap* and Craig T. Harston*
Accurate Automation Corporation
Chattanooga, TN 37402

* The Tennessee Center of Neural Engineering and Applications

CONF-8903223--1

DE93 002956

Keywords: Multisensor Information Fusion, Neural Networks, Scene Structure, Hierarchical Scene Structure, Image Analysis, Automatic Target Recognition

ABSTRACT

A major problem with image-based MultiSensor Information Fusion (MSIF) is establishing the level of processing at which information should be fused. Current methodologies, whether based on fusion at the pixel, segment/feature, or symbolic levels, are each inadequate for robust MSIF. Pixel-level fusion has problems with coregistration of the images or data. Attempts to fuse information using the features of segmented images or data relies on a presumed similarity between the segmentation characteristics of each image or data stream. Symbolic-level fusion requires too much advance processing (including object identification) to be useful, as we have seen in automatic target recognition tasks.

Image-based MSIF systems need to operate in real-time, must perform fusion using a variety of sensor types, and should be effective across a wide range of operating conditions or deployment environments.

We address this problem through developing a new representation level which facilitates matching and information fusion. The Hierarchical Scene Structure (HSS) representation, created using a multilayer, cooperative/competitive neural network, meets this need. The HSS is intermediate between a pixel-based (image segment) representation and a scene interpretation representation, and represents the perceptual organization of an image. Fused HSSs will incorporate information from multiple sensors. Their knowledge-rich structure aids top-down scene interpretation via both model matching and knowledge-based region interpretation.

FG07-88ER12824

DISCLAIMER

1.0 MULTISOURCE INFORMATION FUSION (MSIF): A SIGNIFICANT NEED, A CHALLENGING PROBLEM

Sensor data fusion has the potential to offer significant performance improvements in a variety of systems. Today's sensor fusion technology is no longer a "black box." The VLSI integrated circuit technology makes it possible to develop a new standard for use in the field. By integrating information from multiple sensors, we can reduce the reliance on any single sensor or sensor type. Thus, we can achieve increased system performance even under the loss of individual sensor performance.

MultiSensor Information Fusion (MSIF) systems should be robust, real-time, and fault-tolerant. There are several fundamental issues which must be addressed and understood before any technology-based consituency will fully support sensor fusion. These issues are:

- What to fuse: Focusing attention;
- When to fuse: Selecting levels for fusion,
- Where to fuse: Designing system architectures, and
- How to fuse: Selecting appropriate methodologies.

We focus on selecting the appropriate representation level, and introduce a novel "level" for fusing image information. By using this new "structural representation level," we show that many of the problems that have been major difficulties with previous fusion technologies can be overcome. New technologies, such as neural networks applied to perceptual organization, make these developments possible.

2.0 WHEN TO FUSE: SELECTING AN APPROPRIATE REPRESENTATION LEVEL

If we have images from two or more sensors, the most significant questions we can ask are:

- At what levels of representation (or single-sensor processing) should the information from two or more sensors be combined?

- Once appropriate representation level(s) for fusion have been determined, how should the information actually be combined?

The reason that these issues are the most important is that they define the nature of the fusion task. Previous work in image fusion has met with severe limitations because the representation level has not been adequate for the fusion task. If we can select the right representation level for fusion, then other issues (such as finding a way to focus attention) will fall into place. For this reason, we concentrate in this paper on the issue of selecting an appropriate representation level for image fusion.

The levels which have been proposed thus far are:

* Sensor data level fusion,
* Segment/feature level fusion, and
* Symbolic information fusion.

2.1 SENSOR DATA LEVEL FUSION

Some researchers propose fusing information directly at the pixel level; that is, direct image fusion [e.g. Evans & Stromberg, 1983; Welch & Ehlers, 1987]. There are a lot of problems with this approach, beginning with problems of coregistration. If the images are taken from similar sensors, and are taken from the same locations, then pixel-level fusion is feasible. Even then, there is some question as to what each "new" pixel value (resulting from fusion of pixels from two or more sensors) actually means. If the sensors are from different locations, or if they have very different spectral responses, then simply overlaying pixels from one sensor onto another will not work. It defeats the purpose of gaining real information from the different sensors.

2.2 FEATURE LEVEL FUSION

Most researchers doing MSIF with images favor fusing extracted image segments. This whole approach is based on the premise that a segment in one image can be matched on a one-to-one basis with a corresponding segment in a different image [Marr & Poggio, 1976; Drumheller, 1986; Allen & Bajcsy, 1985; Mitiche & Aggarwal, 1986; Nandhakumar & Aggarwal, 1986, Magee & Aggarwal, 1983]. There are several difficulties with this approach.

First, segments and features from different sensors may not match. This may be due to the intrinsic nature of what each sensor reponds to, or it may be due to differences in the way the segmentation algorithms work. Other differences may be due to temporal or spatial dislocations in the reponses of sensors.

Second, the feature level itself may not be a useful level for information fusion. There is no potential at the feature level itself to represent patterns of features, whether spatial, temporal, or spatio-temporal. It may be that the information needed for accurate fusion resides at a pattern level which is one level more abstract than feature extraction.

2.3 SYMBOLIC LEVEL FUSION

Some researchers advocated fusion at the symbolic level of data representation [Rearick, 1987 (a) & (b)]. However, a problem with fusion at the symbolic level is that it presumes that interpretation of data from each sensor has already been done. This is contrary to the purported goal of MSIF, in which data from multiple sensors is used to create a symbolic interpretation. Thus, we see that each of the representation levels currently in use has inadequacies for MSIF.

3.0 BIOLOGICAL MULTISOURCE INFORMATION FUSION SYSTEMS SERVE AS INSPIRING MODELS

The human brain is capable of addressing problems such as "what, when, where and how" data fusion should occur. The nervous system functions by integrating different types of information, controlling what data is fused, resolving spatial discontinuities, integrating views from different angles, and fusing data from different sensors. In the brain, sensory data is channeled from a bed or network of neurons called a nucleus to the next network. Each network processes or refines the data into more meaningful or abstracted concepts. The result is passed on to higher levels of the brain for further processing. Conceptually organized sensory fusion begins to occur as information is passed from one cortical area to another [Churchland, 1986].

In human visual processing, features such as color, movement, edges and orientation result from neuronal activity in the eyes, the lateral geniculate nuclei, and the primary visual cortex [Churchland & Sejnowski, 1988]. Obviously, these features are not meaningful in themselves. To be meaningful, considerable activity at the cortical level is required. At each cortical step, the visual information is processed, associated, or fused with information from other sensors.

By the time the multisensory-fused information gets to the parietal cortex, the object has been located in the visual field. In the parietal cortex, the object becomes fused or associated with attentional importance [Wise & Desimone, 1988].

Interestingly, identification of the object is not involved with the parietal cortex. It is in the temporal associative cortex where fusion or association with object identification occurs. This suggests that meaning is the result of association or fusion of neural activities from different areas.

There are two uses for sensory information. One is nonspecific or motivational in nature and is used to activate or alert the brain to the new activity. Much of this work is done in the brain stem's reticular activation system. The nature of this information is not specifically meaningful but is used as a general motivation. It is fused with sensory features in the cortex to help provide motivation for attention and movement.

The other use of sensory information is specific in nature. Information is moved from one layer (ie., nucleus) of networks to the next. Ultimately, the information is processed by the cortex. Here the raw data has been featurized so there is some meaning. Meaning at this level consists of movement detection, color or edge detection. While this information is important, it is also fundemental. Little high level conceptual meaning is evident at this point. Principally, the sensory processing is restricted to columnar organization with little fusion.

4.0 NEW APPROACH TO MSIF: FUSING SCENE STRUCTURES

We have developed a robust, generic, and powerful approach to MSIF that works by using a new representation level for fusion. This new representation level is called a Hierarchical Scene Structure (HSS). The power of our HSS is based upon region clustering done by neural network technology. Our neural network technology uses a multilayer, cooperative/competitive paradigm [Minsky & Maren, 1989; Maren et al., 1988]. This technology draws upon neuroscience principles and allows for discrimination of significant perceptual objects from background, noise, or clutter.

The multilayer architecture allows the system to identify the most "perceptually salient" regions in an image. This capability could be combined with a "novelty detector" and an adaptive filter to focus attention on meaningful objects. This approach draws inspiration from the biological models which we discussed earlier.

The current HSS process completes the formation of an HSS for the entire image before interpretation begins. However, HSS clusters are formed from the most perceptually salient regions first. These clusters of regions have significant distinctions from their surround. It would be possible to modify the HSS approach so that analysis begins as soon as perceptually distinctive clusters of regions are extracted. Further, knowledge-based intepretation can be invoked to search for expected region correlations as soon as hypotheses are made.

In multisensor fusion applications, the output of different sensors, or of sensors in different locations, could be processed locally by the HSS paradigm. Novel or perceptually salient image regions would be identified by each HSS processing system. These novel or salient regions would appear as distinct clusters in the HSSs created from each sensor output.

Different sensors will produce images which have different types of perceptually salient features. For example, a tank gun barrel may be a perceptually distinctive feature in a visual image, whereas an exhaust trail may be distinctive in the IR image. If perceptually salient clusters can be identified in similar locations in different images, then these clusters can be fused very early in the HSS process. It would not be necessary to complete HSSs for each image in order for fusion to occur.

Fusion will proceed top-down, and focus attention on matching novel or salient clusters first. This will facilitate real-time operation. This fusion could be accomplished by modifying the current HSS paradigm as described in Minsky & Maren [1989]. Our neural network HSS system can use modifiable weights to learn to identify certain types of high-priority perceptual groupings, such as would occur with man-made objects. A connectionist system could then perform intelligent correlation with stored decision points and models.

We are developing an analog of the HSS approach for representing the structure of temporally-varying signals. This Hierarchical Data Structure (HDS) can similarly be a basis for fusing temporal information, such as is found in seismic or sonar signals [Maren et al, 1989].

4.1 HIERARCHICAL SCENE STRUCTURES REPRESENT THE PERCEPETUAL ORGANIZATION OF SEGMENTED IMAGES

We have developed a unique Hierarchical Scene Structure (HSS) method to represent for the perceptual organization of segmented regions of an image. The primary advantage of using an HSS is that image information is represented in a structured manner. The HSS explicitly encodes valuable high-level information. This high-level information includes the relationships between the segmented regions in an image. (In related research, we are using Hierarchical Data Structures (HDSs) to structure the segmented events in time-varying signals.) This relationship information, including such "perceptual features" as proximity between segments, similarity of intensity or amplitude, and other features, may be valuable in both characterizing the nature of a structured cluster of segments, and in facilitating matches between structures.

By using Hierarchical Scene Structures, we introduce a new representation level into the typical low-level to high-level approach to image processing and interpretation. The HSS level is intermediate, as is shown in Figure 1.

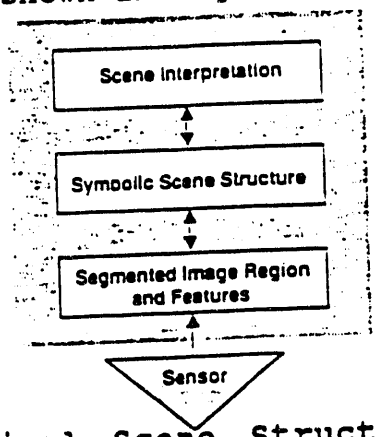

Figure 1. Hierarchical Scene Structures introduce a new representation level into image processing systems.

In these Hierarchical Structures, the "top node" of the structure contains information about the entire structure, such as its total size, average intensity or amplitude, and other features which describe globally the entire set of segments which make up the structure. Lower-level branch nodes similarly contain information about all nodes subordinant to them. Thus, by examining only the top layers of a structure, it is possible to extract a great deal of information about the structure and its components.

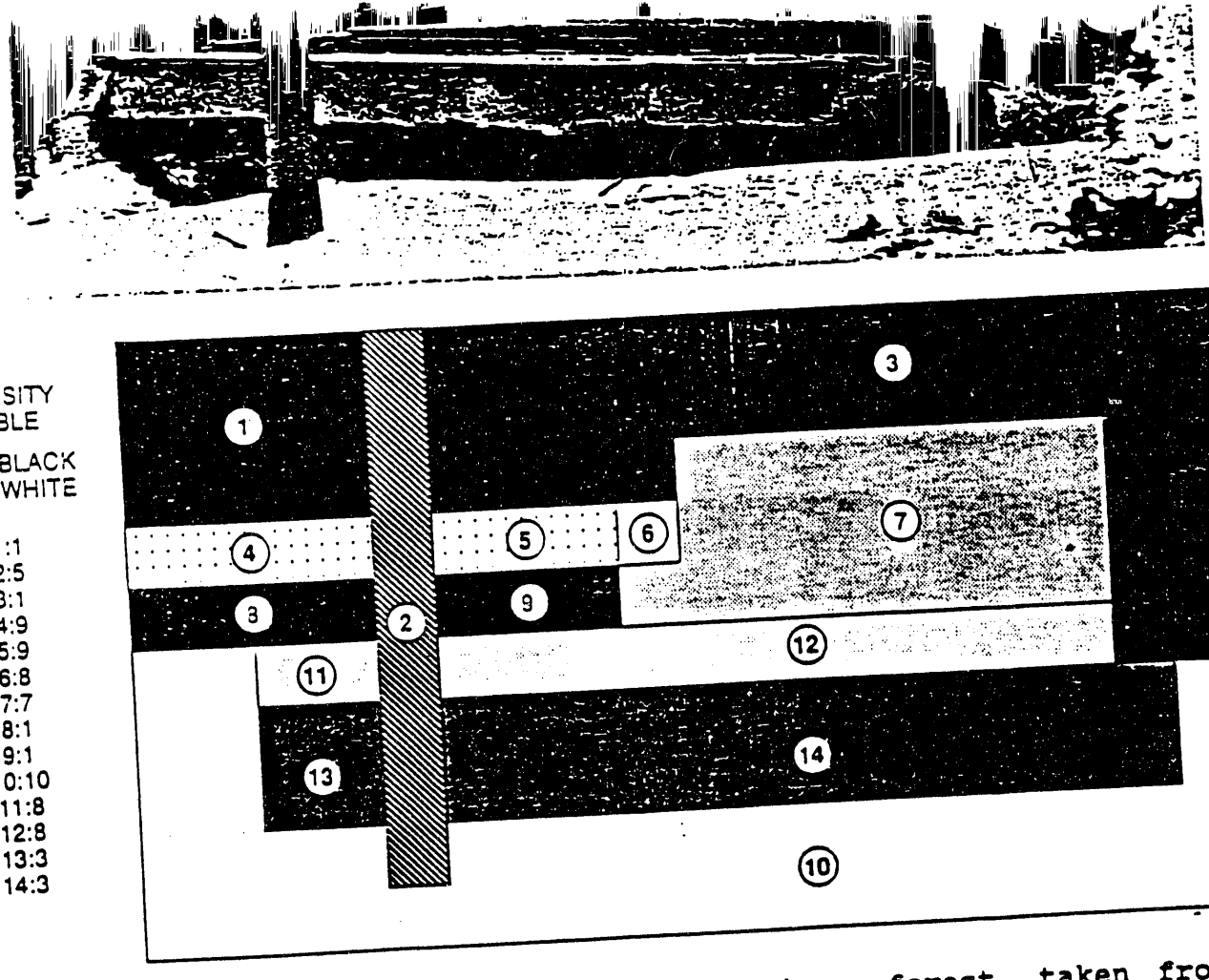

Figure 2. (a) Image of a Soviet tank in a forest, taken from Soviet Military Power (1988). (b) Stylized segmented version of the tree and tank, using a large-pixel synthetic image.

Figure 3(a) shows a Hierarchical Scene Structure created from data corresponding to the segmented tank in Figure 2(b). This HSS was created using a multilayer cooperative/competitive network, which is fully described in [Minsky & Maren, 1989]. An earlier cooperative/competive network for performing this type of region grouping was described in [Maren, Minsky, & Ali, 1988].

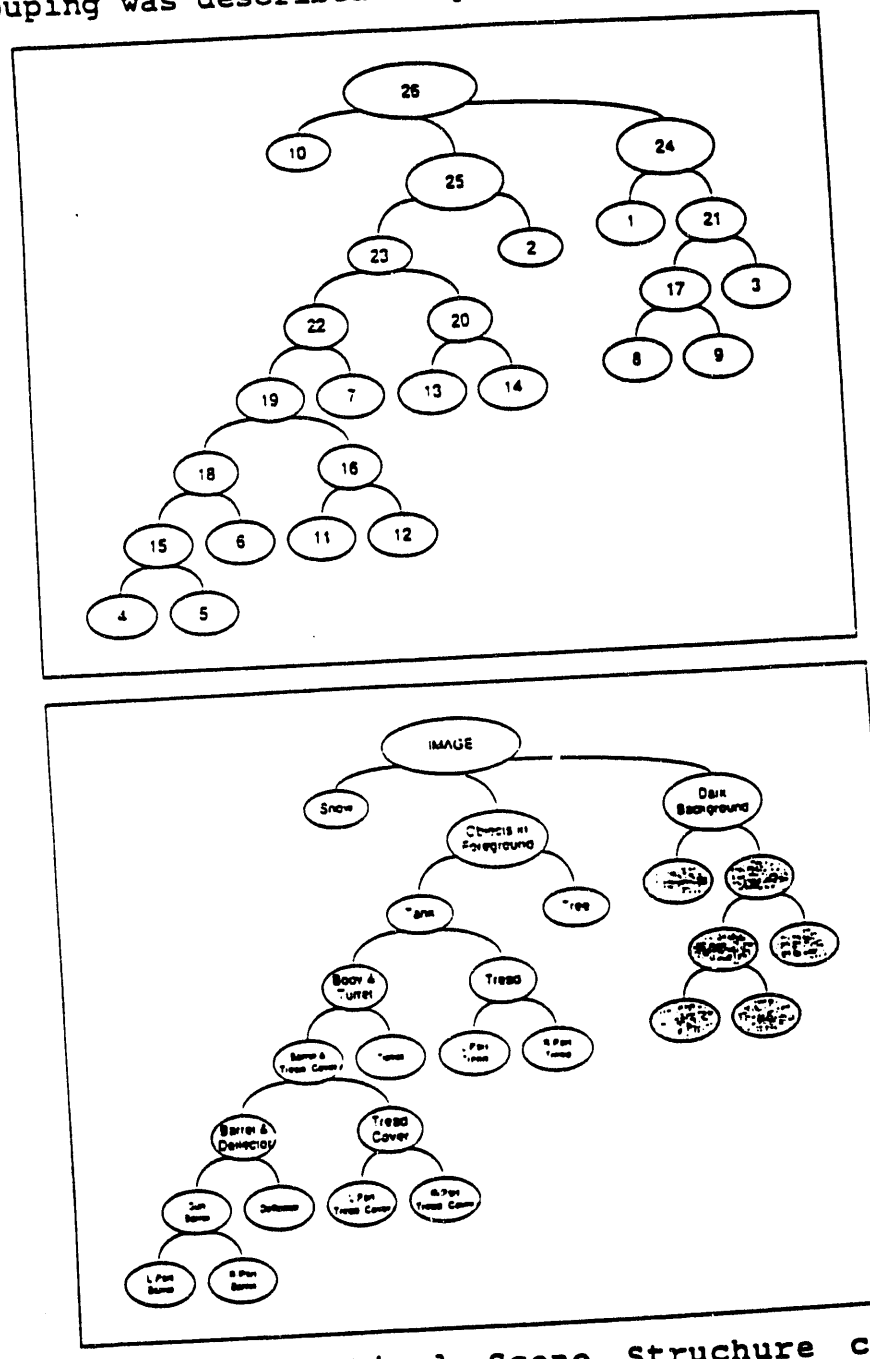

Figure 3 (a) Hierarchical Scene Struchure created using perceptual relations calculated between segments shown in Figure 2(a). (b) A knowledge-based system could traverse the HSS and identify major components using both perceptual features and segment-descriptive features, both of which are stored in the HSS. For details, see [Minsky & Maren, 1989; Maren & Ali, 1988, & Naren et al., 1988].

A knowledge-based system could interpret the structure shown in Figure 3(a) to yield interpretation of both the objects and the different parts of the objects, as shown in Figure 3(b). There are two ways in which a knowledge-based system could perform this interpretation. These methods correspond to object model matching and knowledge-based region interpretation. Both approaches rely on the fact that every branch node in an HSS contains information that describes to aggregate properties of all nodes which descend from that branch. Thus, the properties of a branch node can be used to generate tentative object matches or hypotheses about the entire group of regions denoted by that node.

4.2 MULTICOURCE HIERARCHICAL SCENE STRUCTURES ARE A ROBUST WAY TO REPRESENT FUSED INFORMATION

Each HSS represents the perceptual organization of a segmented scene.. By fusing the HSSs made from different images, we can create a new, information-rich Multisource Hierarchical Scene Structure (MHSS). This structure captures high-confidence components (image segments) from multiple sources, along with knowledge of significant relationships between components, features describing them, and confidence measures. This structured representation is amenable to top-down image analysis.

The representation levels for an MSIF system are shown in Figure 4. There are two new levels in this system; an HSS level for each sensor, and a fusion level, occuring just above the HSS level, to represent the Multisensor-fused Hierarchical Scene Structure.

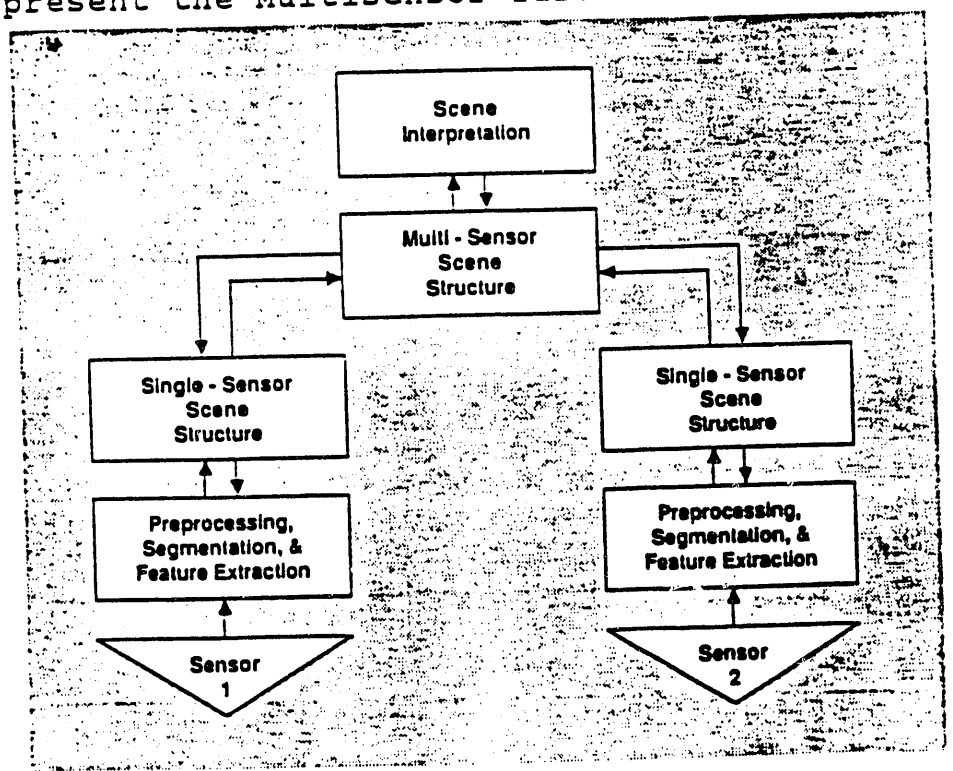

Figure 4. Major representation levels for multisensor information fusion system.

Hierarchical Scene Structures, created from each of the input images or data streams, are an appropriate starting point for MSIF because of two factors. First, Hierarchical Scene Structures contain unique perceptually-based information (e.g., proximity, similarities) which can be invaluable in matching sets of segments from one image or data stream to sets of segments from another source.

Second, the unique encoding of Hierarchical Scene Structures facilitates rapid identification of significant and/or strongly differentiated areas of interest in each image. The fusion process begins with these significant areas, enhancing the probability of a useful match and concentrating processing power on those groups of regions or data segments which are most perceptually distinctive or significant.

Figure 5 shows how single-sensor Hierarchical Scene Structures can be used to provide a basis for fusing multisensor images. Unlike feature-mapping approaches, the fusion here takes place at the scene structure representation level. Stylized visible and IR images in Figures 5(a) and (b) each yield a HSS, shown in Figures 5(c) and 5(d). Fusion occures by matching and merging the single-sensor HSS's in Figure 5(e). This high-confidence MHSS provides a robust basis for scene interpretation, as illustrated in Figure 5(f).

The Hierarchical Scene Structure method can be modified to be used for fusing temporally-varying signals, including sonar, radar, and seismic data. This is vital in areas where a large amount of information from multiple sensors needs to be analyzed many times per second and per sensor. Neural networks have the potential to process this type of data and compare it with other information. This information can then be discriminated against other information to provide a viable recognized pattern.

Our current work focuses on extending our HSS method to representing temporally-varying signals, such as would be observed in sensor data readouts. We are also extending our cooperative/competitive method for creating initial HSS's to create a robust method for matching HSSs against existing models and for fusing HSSs to create an MHSS.

5.0 HOW CAN STRUCTURE-BASED MSIF BE USEFUL?

The benefits which can be achieved through high-level, symbolic fusion of multisource information include increased accuracy of object/feature recognition under both controlled and natural conditions, greater specificity in characterizing object/feature attributes, and improved functionality of autonomous and semi-autonomous systems. Previous work has pointed the way to the benefits which could be achieved, but has also shown how difficult the task of MSIF truly is. The technical approach offered here provides a robust framework for symbolic information integration and for making the fused information accessible and useful.

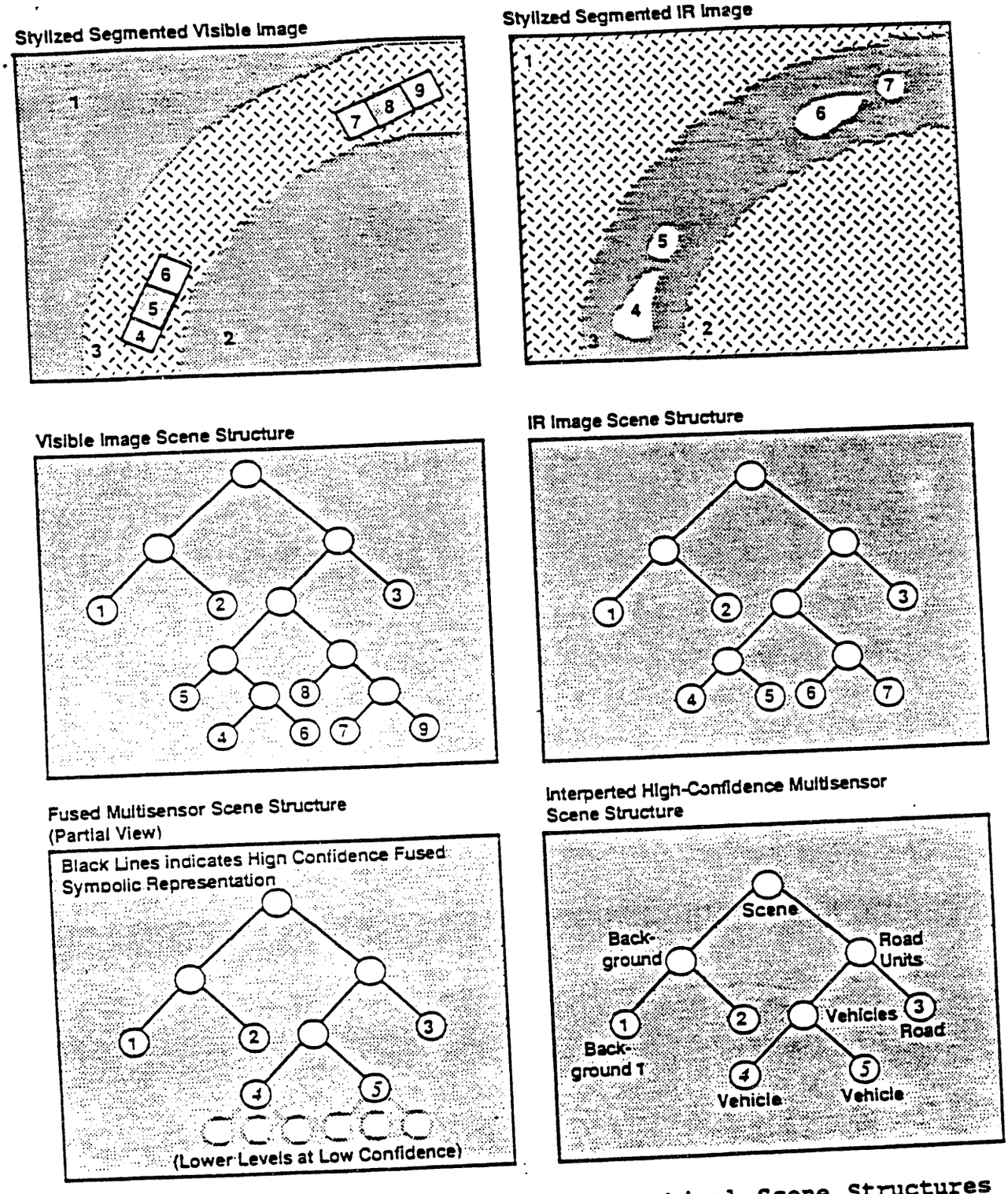

Figure 5. Illustration of how Hierarchical Scene Structures can aid multisensor information fusion. (a) & (b), stylized synthetic images of a convoy as taken from visual and IR cameras. (c) & (d), Hierarchical Scene Structures created from (a) and (b), respectively. (e), Multisensor-fused Hierarchical Scene Structure (MHSS). (f), interpreted structure.

We are investigating the use of associative neural networks to perform interpretation of fused information. The goal of this work is to associate fused information with meaningful concepts, which can be symbolically represented as words. Our approach to this problem, using an associative neural network cortical model, is patterned after the brain and is used to guide these research efforts. The cooperative/competitive component of the system is a proven concept. Additional work continues on the organization of the association process. Current activity is involved with extending the systems'range of applicable input, including digital data representations and data structures.

Certain of today's VLSI integrated circuit technologies allows for interface with microprocessor and eye pattern generator. These could be used to directly capture on-site or remote imaging sensor data for an MSIF/HSS processor. The possibility exists that incoming sensor data could be interpreted in real-time as it is received. Conceptually, distortion correction, digitization, reduction and magnification as well as image signal enhancement can be developed for this system in a relatively small and fieldable unit. The unit would maintain electronic files and be capable of byte transfer as well as high speed transfer modes. Neural network technology is becoming available in integrated circuits that could then interface into the system.

The MSIF/HSS is becoming a technology that can be adapted to use in many future applications.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the work performed on developing the refinements of the HSS concept by Veronica Minsky, who created the Hierarchical Scene Structure shown in Figure 3(a). Work was performed under DoE contract DEF07-88-ER12823, NSF grant EET8901748, and under internal research and development programs of Accurate Automation Corporation and the University of Tennessee Space Institute.

REFERENCES

Allen, P.K.; & Bajcsy, R. "Integrating sensory data for object recognition tasks." In O.D. Faugeras & R. Kelly (Eds), Computer Vision for Robots (SPIE Proc. 595, 1985, Cannes, France), 225-232.

Churchland, P.S., Neurophilosophy (Cambridge, MA: MIT Press, 1986).

Churchland, P.S., & Sejnowski, T.J., "Perspectives on cognitive neuroscience," Science, 242 (1988), 741 - 745.

Drumheller, M. "Connection machine steromatching," Proc. of AAAI-86 (Aug. 11-15, 1986, Philadelphia, PA), 748-753.

Evans, D., & Stromberg, B. "The use of multisensor images for earth science applications," Proc. of the National Telesystems Conference (San Francisco, CA, November 14-16, 1983), 271 - 275.

Magee, M.J., & Aggarwal, J.K. "Using multisensory images to derive the structure of three dimensinal objects - a review," Computer Vision, Graphics, and Image Processing, 82 (1985), 145 - 157.

Maren, A.J., Pap, R.M., Harston, C., "A Hierarchical Data Structure representation for fusing multisensor information." Paper to be presented at the 1989 SPIE Technical Symposia on Aerospace Sensing (Orlando, FA; March 27-31, 1989).

Maren, A.J., Minsky, V., & Ali, M., "A multilayer, cooperative /competitive method for creating hierarchical scene structures by clustering maximally-related nodes," Proc. Second IEEE Int'l. Conf. on Neural Networks (San Diego, CA; July 24-27, 1988).

Maren, A.J., & Ali, M., "Hierarchical scene structure representations to facilitate image understanding," Proc. First Int'l. Conf. on Indus. & Eng. Appl. of AI & Expert Sys. (Tullahoma, TN; Jun. 1-3, 1988).

Marr, D., & Poggio, T., "Cooperative computation of stereo disparity," Science, 194 (Oct. 15, 1976), 283 - 287.

Minsky, V., & Maren, A.J., "Representing the perceptual organization of segmented scenes." (Manuscript submitted for review to Spatial Vision.)

Mitiche, A., Aggarwal, J.K. "Multiple sensor integration fusion through image processing: A review," Optical Engineering, 25 (March 1986), 380-386.

Nandhakumar, N., & Aggarwal, J.K. "Integrating information from thermal and visual images for scene analysis," Proc. of SPIE Conference, 635 (April 1--3, 1986, Orlando, FA.), 132-142.

Rearick, T.C. "Multi-source information fusion for autonomous landing," Proc. of the Autonomous All-Weather Navigation and Landing Workshop, AFWAL (Monteray, CA: 1987).

Rearick, T.C., "Knowledge-based multi sensor image fusion," Lockheed Horizons, 25 (Dec., 1987), 22-30.

Welch, R., & Ehlers, M. "Merging multiresolution SPOT HRV and Landsat TM data," Photogrammertic Engineering and Remote Sensing, 53 (March 1987), 301-303.

Wise, S.P., & Desimone, R., "Behavioral neurophysiology: Insights into seeing and grasping," Science, 242 (1988), 736 - 741.

DATE
FILMED
02/02/93