Focusing Problems in Vision Research

Steven A. Israel

Department of Surveying

University of Otago. Dunedin, New Zealand

Phone: +64 3 479-7698 Fax: +64 3 479-7586

Email: israel@albers.otago.ac.nz

Presented at the 10th Colloquium of the Spatial Information Research Centre,
University of Otago, New Zealand, 16-19 November, 1998

Abstract

This paper identifies the current processing strategies common to most vision research and suggests a new operating paradigm to extend the capabilities of new pattern recognition algorithms. The question centres on whether it is more appropriate to accept or reject an experiment or vision system in total or accept those parts that fall within the processing limits and modify those regions that fail. The application of the latter method is discussed here.

Keywords and phrases: vision research, processing paradigm, limitations of current processing strategies

1.0 Introduction

This paper contains a personal perspective of the current state of computer vision research. These comments are based upon a literature search into image analysis including, preprocessing, segmentation, matching, and classification. Jain and Dorai (1997) define computer vision as the ability to describe an object space based upon analysis of remotely sensed images. As such, it must include colour, texture, and geometric modelling. Computer vision includes the concepts of context and classification.

In the following sections, the current computer vision paradigm is identified and references are provided for how this paradigm is applied to various aspects of pattern recognition. This processing paradigm is contrasted with human processing of similar tasks. From this, a new processing paradigm is offered. This new processing paradigm was generated without biological review; however, it will be shown to make empirical sense.

2.0 Current Processing Paradigm

Common research ethos in computer vision suggest that complex systems, those that are able to recognise a large number of features, complex features, and complex scenes are viable for real world applications. This is an admirable goal but currently unrealistic. The processing goal to support human decision making or operate over a limited domain is quite different than completely automated processing and more realistic.

In this modified support role, researchers have also simplified the processing strategy. This processing strategy consists of 3 stages; preprocessing raw data, generating transfer functions between inputs and outputs, and evaluating the results. This serial linear process is called the image chain approach (Schott 1997) and is summarised in Figure 1. The focus of this processing paradigm is to develop efficiencies at each stage of the chain. By continuing to operate within this strategy, computer vision systems will not be able to attack more complex tasks such as integrated segmentation and classification or autonomously developing context.

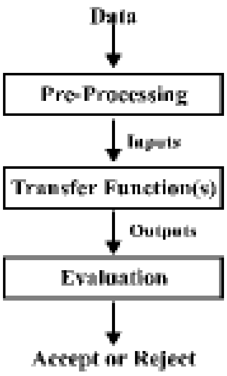

Figure 1 Current

3.0 Current Processing Paradigm

The result of the single stage focus is that best general practices are developed. Each vision task is contained as a stand-alone unit. Argialas and Harlow

(1990) separated these vision tasks into three levels of complexity (Figure 2). As yet, no unifying theory has been generated where any one vision technique could be applied to a wide range of images.

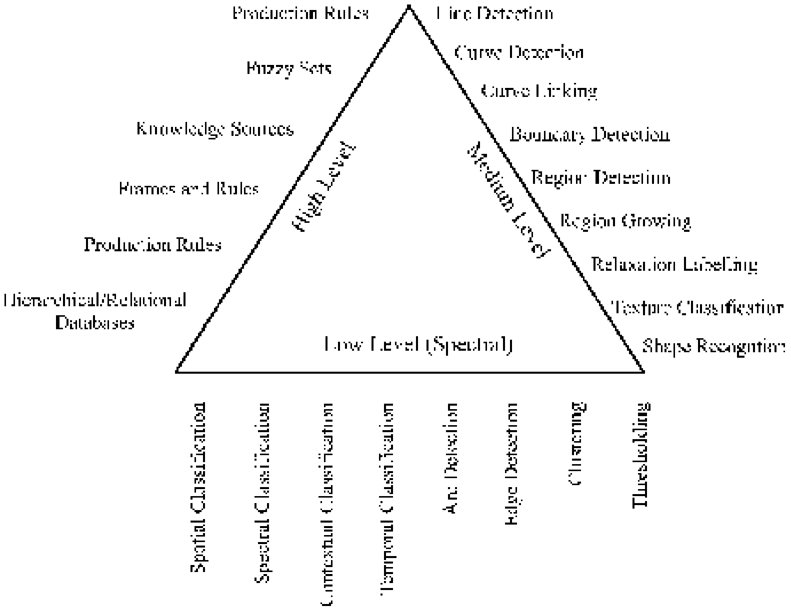

Figure 2 Argialas' (1990) Pyramid of Image Processing Functions

Before continuing, it is important to understand the processes that computers perform well and contrast them with human cognition. Computers provide excellent quantitative analysis, which includes counting and computing fact. They are also good at determining similarities. Computer models work best in context independent processing. Humans are excellent at relative measures. We are capable of determining differences and changes in textures. Humans seek to determine context for most decision making. Whether or not mimicking human perception is an appropriate function of computer vision is outside the scope of this analysis. The relationship between the individual processing stages and their constraints are identified below. The limitations of these constraints are identified.

3.1 Data Acquisition

From the collective experience with imaging sensors (Dereniak and Boreman 1996) and applications (Schowengerdt 1997) optimised viewing and object space characteristics have been observed. Most vision models perform well with Lambertian surfaces with a smooth, continuous, nondeformable object space. The diffuse lighting is evenly distributed across the field-of-view (Marr 1982; Lemmens 1988; Brown 1992). Colvocoresses (1983) showed that the shape and orientation of the sensor constrains the registration process to improve reliability. Finally, several researchers reviewed the effects of information content in the input stream varies as a function of spatial and spectral resolution, contrast, and scene complexity (Duggin 1985; Everitt *et al.* 1996; Jones 1997).

In natural scenes, the illumination and object space conditions vary considerably. For example, water and metallic targets are spectral reflectors. Natural confusion exists between urban areas and beach sand. The concept of target area is quite subjective. Often the success of a computer vision system is determined on an application basis.

3.2 Pre-processing/Attribute Selection

As image data is a surrogate for physical features, raw data may require processing. It is generally agreed that attribute selection is application dependent. The curse-of-dimensionality problem offsets a continuous search for a pre-processing technique that may provide negligible improvements to vision performance (Kohonen 1997). At this stage, the vision techniques separate into their individual components. Registration and segmentation algorithms extract high spatial frequency information and classification generally low spatial frequencies. Pavlidis (1978) noted that the subtleties of the specific attribute extraction technique were less important than the class of images to which it was applied. Haralick and

Shapiro (1985) reviewed a number of image segmentation techniques. Gong and Howarth (1990) looked into how these segmentation and texture features could be applied to improve spectral classification accuracy. Stehman (1992) summarised methods for dividing the input stream into training and test datasets to best represent class populations. Heath *et al.* (1998) compared edge detector performance for segmentation and feature extraction.

The major problem facing the current processing strategy is the difference between required information and the information produced by the attribute selection and preprocessing operators. For example, edge detectors find areas of a local intensity gradient. The expectation is that these gradients are outline target boundaries. The difference is significant. Edge detection is a function of the width of the edge detector, brightness difference, and brightness gradient. Human operators are able to identify the spectral reflection of a light source, called a highlight, within a natural scene and discount it from analysis (Figure 3). Edge detectors identify this local gradient as a target.

3.3 Generation of Transfer Functions

Several registration and classification transfer functions have been developed for computer vision. The most common algorithms currently used are connectionist and statistical. The former is used because of its ability to adapt to the datastream without assuming prior probabilities. The latter has the benefits of fast efficient processing. Lippmann (1987) compared a number of neural network algorithms. Bezdek (1993) expanded on this work by comparing neural networks to statistical and fuzzy algorithms for pattern recognition. Guelch (1988) compared the performance of several image registration algorithms for general use. These comparisons were performed for isolated processing. Common human processing is an integrated approach. For example, we use high frequency information to obtain context and segment targets. Once segmented, the targets are recognised as belonging to a group, which allows better segmentation to improve identification of the target as belonging to a specific kind. This processing may be too good in some instance, as context may not allow the observer to recognise fact.

Figure 3 Spectral Target within Field-of-View: Reflection of the overhead light is visible in the crystal in the centre of the frame.

In Figure 4, human cognition observes two interleaved triangles. However, objective analysis shows that neither triangle actually exists. No pattern recognition system would be able to identify this pattern either. Whether this type of performance is beneficial is outside the scope of the paper, but the issue must be addressed during development, especially integration of tasks.

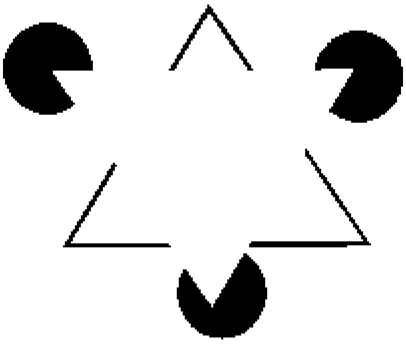

Figure 4 Visual Teaser of Interleaved Triangles, or Not?

3.4 Evaluation of Results

Quantitative evaluation is essential for all computational systems. Mohan *et al.* (1989) and Forstner (1984) produced a number of relative measures to evaluate automated image registration techniques.

Foody (1992) compared classification accuracies to chance results for the same data. Rosenfield and Fitzpatrick-Lins (1986) evaluated the suitability of thematic map accuracy using different sampling techniques. These techniques are divided by how the computer vision task separates the feature space and the amount of supervision during sampling. Supervised sampling allows quantitative analysis because the output contains both *a priori* and *a posteriori* labels. Automated or unsupervised processes only have the assignment labels.

3.5 Accept or Reject

The problems of computer vision are highlighted during the decision of whether or not to accept or reject the analysis. No single theory of goodness exists for any vision task or application. Cordella *et al.* (1995) suggest that distance thresholds should be developed to recognise samples that are sufficiently different from the centres of the developed acceptance regions in the feature spaces. Those samples outside this region are unlabelled. However, the logic for determining what the level should be remains unknown. Jain and Dorai (1997) and Casasent and Neiberg (1995) define a reject rate for detectors to reduce the collection of false positive samples. However, where to assign this threshold remains ambiguous

4.0 Case Study

The above problems in the system are highlighted in the following example. Suppose a computer vision system was developed to separate targets on the basis of colour. If the system ran into problems by assigning the known colour green to red, the following processes would have to be performed under the current processing paradigm. The sensor would be calibrated against all known colours. Then, the pre-processor would be evaluated for how well it segmented the input datastream into examples. The individual transfer functions would be either retrained or recreated.[1] All of this is computationally demanding. The system would then be evaluated based upon the number of correct responses and the system would be considered suitable if the errors were within a certain tolerance.

Human processing of the above example would proceed as follows. A parent asks his or her child the colour of the grass. If the child responds with ‘red’ as in the above case, the parent corrects the child with the correct label of ‘green’. Then the parent may add that the bricks on the house are red. In this model, the parent both excites the correct response, and inhibits the incorrect response. In the human case, existing processing rules were modified based upon experience, or in this case, a modification of the input stream. This integration of fixed rules or prior knowledge with adaptive training maximises the flexibility of the vision system to process a wider stream of examples and minimise processing time and required sampling (Perlovsky 1998).

4.0 New Processing Paradigm

The simple and elegant solution of human modification of colour labelling may be applied to computer vision. This new processing paradigm uses the confusion in-formation developed in the evalua-tion stage as a means for adapt-ing the transfer functions. I recommend the use of connectionist algorithms specifically fuzzy neural networks because they have the ability to adapt from examples and encode prior knowledge. Figure 5 shows the schematic for a new processing paradigm. The recursiveness occurs between processing stages rather than within a single stage, as is the case for most recursive function.

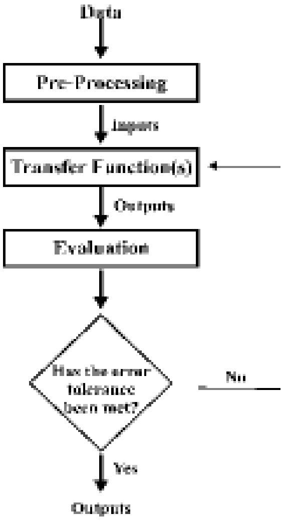

Figure 5 New Processing Paradigm

---

[1] Interesting here. A statistical function separating the colours would have to be regenerated. The samples for the colours of green and red edited so that a greater separation existed between the groups. A connectionist function would be retrained using the entire dataset until the system returned the correct results. If after a sufficient amount of time, the answers were still incorrect, the dataset would then be edited and transfer functions recreated. Fuzzy logic would require the analyst to develop rules to better separate between the red and green labels. Rules would be modified until the correct results were obtained.

This technique has been applied to both unsupervised or automated data extraction and supervised classification (Israel and Kasabov 1996; Israel and Kasabov 1997). In both cases, a statistically significant improvement was observed. This improvement was realised with a limited number of iterations and a reduced example set. The other advantage of this processing paradigm is computational savings during the creation of the transfer functions. Since these techniques are extremely efficient, the individual transfer functions require only a limited number of iterations to identify where the confusion exists.

Acknowledgements

Image exploitation was partially funded through the Public Good Science Fund (New Zealand) Foundation for Research, Science, and Technology (FRST)-96-UOO-606 and FRST-95-UOO-S19-4052. The concepts developed here are based upon my PhD research and observing the movements of my newborn child, his two siblings, and how they have taught me.

References

Argialas, D.P. and Harlow, C.A. (1990) Computational Image Interpretation Models: An Overview and a Perspective, *Photogrammetric Engineering and Remote Sensing* 56(6): 871-886.

Bezdek, J.C. (1993) A Review of Probabilistic, Fuzzy, and Neural Models for Pattern Recognition, *Journal of Intelligent and Fuzzy Systems* 1(1): 1-25.

Brown, L.G. (1992) A Survey of Image Registration Techniques, *ACM Computing Surveys* 24(4): 325-376.

Casasent, D.P. and Neiberg, L.M. (1995) Classifier and Shift-invariant Automatic Target Recognition Neural Networks, *Neural Networks* 8(7/8): 1117-1129.

Colvocoresses, A.P. (1983) The Relationship of Acquisition Systems to Automated Stereo Correlation, *Photogrammetric Engineering and Remote Sensing* 49(4): 539-544.

Cordella, L.P., De Stefano, C., Tortorella, F. and Vento, M. (1995) A Method for Improving Classification Reliability of Multilayer Perceptrons, *IEEE Transactions on Neural Networks* 6(5): 1140-1147.

Dereniak, E.L. and Boreman, G.D. (1996) *Infrared Detectors and Systems*. New York, John Wiley & Sons, Inc.

Duggin, M.J. (1985) Factors Limiting the Discrimination and Quantification of Terrestrial Features using Remotely Sensed Radiance, *International Journal of Remote Sensing* 6(1): 3-27.

Everitt, J.H., Escobar, D.E., Alaniz, M.A. and Davis, M.R. (1996) Comparison of Ground Reflectance Measurements, Airborne Video, and SPOT Satellite Data for Estimating Phytomass and Cover on Rangelands, *Geocarto International* 11(2): 69-76.

Foody, G.M. (1992) On the Compensation for Chance Agreement in Image Classification Accuracy, *Photogrammetric Engineering and Remote Sensing* 58(10): 1459-1460.

Forstner, W. (1984) Quality Assessment of Object Location and Point Transfer Using Digital Image Correlation Techniques, *International Archives of Photogrammetry and Remote Sensing* 25(3): 197-219.

Gong, P. and Howarth, P.J. (1990) The Use of Structural Information for Improving Land-Cover Classification Accuracies at the Rural-Urban Fringe, *Photogrammetric Engineering and Remote Sensing* 56(1): 67-73.

Guelch, E. (1988) Results of Test on Image Matching of ISPRS WG III/4, *International Archieves of Photogrammetry and Remote Sensing* 27(3/3): 254-271.

Haralick, R.M. and Shapiro, L.G. (1985) Image Segmentation Techniques, *Computer Vision, Graphics, and Image Processing* 29: 100-132.

Heath, M., Sarkar, S., Sanocki, T. and Bowyer, K. (1998) Comparison of Edge Detectors: A Methodology and Initial Study, *Computer Vision and Image Understanding* 69(1): 38-54.

Israel, S.A. and Kasabov, N.K. (1996) Improved Learning Strategies for Multimodular Fuzzy Neural Network Systems: A Case Study on Image Classification, *Australian Journal of Intelligent Information Processing Systems* 3(2): 61-68.

Israel, S.A. and Kasabov, N.K. (1997) Statistical, Connectionist and Fuzzy Inference Techniques for Image Classification, *SPIE Journal of Electronic Imaging* 6(3): 337-347.

Jain, A.K. and Dorai, C. (1997) Practicing Vision: Integration, Evaluation and Applications, *Pattern Recognition* 30(2): 183-196.

Jones, G.A. (1997) Constraint, Optimization, and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features, *Computer Vision and Image Understanding* 65(1): 57-78.

Kohonen, T. (1997) *Self-Organizing Maps*. Berlin, Springer-Verlag.

Lemmens, M.J.P.M. (1988) A Survey on Stereo Matching Techniques, *International Archives of Photogrammetry and Remote Sensing* 27(B8): V11-V23.

Lippmann, R.P. (1987) An Introduction to Computing with Neural Nets, *IEEE Transactions on Acoustic, Speech, and Signal Processing* April: 4-22.

Marr, D. (1982) *Vision*. New York, W.H. Freeman and Company.

Mohan, R., Medioni, G. and Nevatia, R. (1989) Stereo Error Detection, Correction, and Evaluation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(2): 113-120.

Pavlidis, T. (1978) A Review of Algorithms for Shape Analysis, *Computer Graphics and Image Processing* 7: 243-258.

Perlovsky, L.I. (1998) Conundrum of Combinatorial Complexity, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(6): 666-670.

Rosenfield, G.H. and Fitzpatrick-Lins, K. (1986) A Coefficient of Agreement as a Measure of Thematic Classification Accuracy, *Photogrammetric Engineering and Remote Sensing* 52(2): 223-227.

Schott, J.R. (1997) *Remote Sensing: The Image Chain Approach*. New York, Oxford University Press, Inc.

Schowengerdt, R.A. (1997) *Remote Sensing Models and Methods for Image Processing*. San Diego, Academic Press.

Stehman, S.V. (1992) Comparison of Systematic and Random Sampling for Estimating the Accuracy of Maps Generated from Remotely Sensed Data, *Photogrammetric Engineering and Remote Sensing* 58(9): 1343-1350.