MULTISPECTRAL
IMAGE PROCESSING
AND PATTERN RECOGNITION

SERIES IN MACHINE PERCEPTION AND ARTIFICIAL INTELLIGENCE*

Editors: **H. Bunke** (Univ. Bern, Switzerland)
P. S. P. Wang (Northeastern Univ., USA)

*For the complete list of titles in this series, please write to the Publisher.

MULTISPECTRAL IMAGE PROCESSING AND PATTERN RECOGNITION

Editors

Jun Shen
Bordeaux-3 University, France

P. S. P. Wang
Northeastern University, USA

Tianxu Zhang
Huazhong University of Science & Technology, China

MULTISPECTRAL IMAGE PROCESSING AND PATTERN RECOGNITION

CONTENTS

INTRODUCTION

JUN SHEN

Image Laboratory, EGID Institute, Bordeaux-3 University
1, Allée Fernand Daguin, 33607 Pessac cedex, France

PATRICK S. P. WANG

College of Computer Science, Northeastern University
Boston, MA 02115, USA

TIANXU ZHANG

Institute for Pattern Recognition and Artificial Intelligence
Huazhong University of Science and Technology, Wuhan 430074, P.R. China

The papers of this book were selected from those presented at the First International Symposium on Multispectral Image Processing And Pattern Recognition, co-sponsored by SPIE, held on October 21–23, 1998 in Wuhan, China.

Based on the recommendations of reviewers, we selected nine papers for this book from more than 130 presentations at the conference. The authors of the nine selected papers were then invited to submit elaborate papers which were completed with meticulous theoretical deduction and/or expatiatory experimental results, for a new round of reviews. These papers were then revised again by the authors, following the comments of the reviewers.

These papers cover different topics in image processing and recognition: 3D object understanding from 2D images (P. S. P. Wang and W. R. Watkins), image characterization by geometrical, orthogonal and smoothed orthogonal moments (J. Shen, W. Shen and D. Shen), facet matching and object tracking (S. Maybank and R. Fraile), neural network techniques (H. Pan and L. Liu), and different image processing techniques and their applications, such as small target detection in infrared sea surface images (Y. S. Moon, T. Zhang, Z.-R. Zuo and Z. Zuo), vision system for ALV navigation (X. Ye, J. Liu and W. Gu), bibliography information extraction from book cover image (H. Yang, N. Onda, M. Kashimura and S. Ozawa) and radar target recognition (X. Liao and Z. Bao).

The study of 3D representation, understanding, learning and recognition is difficult and complicated because of infinite variations due to rotation, orientation and topological transformations, and this has attracted more and more attention. In particular, articulated objects are even more challenging because in addition to the above-mentioned difficulties, the object itself can change shape or structure. Yet not much research work has been done in this area because of its complexities, and most research in the past concentrates only on rigid object recognition. In the first paper, P. S. P. Wang presents a state-of-the-art of 3D object understanding from 2D images and a new method generalized from linear combination for extracting

1

features of articulated as well nonarticulated portions of a 3D object, a robust yet simple method needing only very few learning samples.

Moments, including geometric and orthogonal ones, are widely used in image processing, pattern recognition, computer vision and multiresolution analysis. To clarify the use of different types of moments, in the second paper J. Shen *et al.* present a study on different moments and compare their behaviors for image characterization and classification. They at first present geometric, Legendre, Hermite and Gaussian–Hermite moments and their calculation, and then analyze and compare their behavior in both spatial and spectral domains. They also show the orthogonality equivalence theorem and that Gaussian–Hermite moments give an approach to construct orthogonal features from wavelet analysis.

The images processed by our brain represent our window into the world. For some animals this window is derived from a single eye, for others, including humans, two or more eyes provide stereo imagery. Animal vision is truly varied and has developed over millennia in many remarkable ways. In the third paper, W. R. Watkins shows us some examples in multispectral image processing. We have learned a lot about vision processes by studying animal systems and can learn even more.

Small target detection has found its wide applications in remote sensing, surveillance, robotics, etc. The difficulty of small target detection is that the available information from targets is much less than that from the background. In the fourth paper, Y. S. Moon *et al.* present small target detection from infrared sea surface images based on multilevel filters.

In the Minimum Description Length (MDL) approach to model fitting, the length of the compressed data is the measure of the accuracy of the fit of the model to the data. MDL is applicable in cases where the usual maximum likelihood (ML) method for model fitting fails. Some applications of MDL have been made in image processing and computer vision. In the fifth paper, S. Maybank *et al.* present an MDL approach applied to model-based vehicle tracking, namely the fitting of a facet model to the image of a car, with the position and orientation of the car on the ground plane as model parameters.

Vision system has been an important topic for research on Autonomous Land Vehicle (ALV). For road following, the vision system must provide the surrounding information of a vehicle. In the sixth paper, X. Ye *et al.* present an integrated vision system for ALV navigation, including modules of 2D vision using color cameras to find road regions, 3D vision using active laser radar for obstacle detection, and information fusion to create a complete environment description. The system is composed of high-speed image processing hardware and transputer-based pipeline with high-speed data channel.

An important problem in multispectral image processing is multisensor data fusion. Bayesian networks (BN) provide a high-level generic architecture for fusing sensory observations from multiple data sources. Discrete BN's offer a general formalism for representing a joint probability distribution of multiple discrete random variables but learning discrete BN's from complete data or incomplete and soft data is still an open problem. Hybrid BN's, the most general BN's encountered

in practical applications, have not yet found a general solution for representation, inference and learning. In the seventh paper, H. Pan *et al.* present Fuzzy Bayesian Networks (FBN) as a general formalism for representation, inference and learning with hybrid BN's, and the inference algorithms of such networks.

With the development of the database and Internet, automatic input and analysis of printed documents receive more and more attention. However, in document analysis, most attention is paid to the text part, while researches on bibliography information extraction are carried out less frequently. In the eighth paper, H. Yang *et al.* present a new system for extracting and classifying bibliography regions from the color image of a book cover. The system consists of three major components: preprocessing, color space segmentation, and text regions extraction and classification. By comparing the text regions on front cover with those on spine, all extracted text regions are classified into suitable bibliography categories: author, title, publisher, and other information, without applying OCR.

Radar target recognition based on high resolution range profiles (HRRP) has received much attention in recent years. Information redundancy is quite severe in HRRP's. Since the distribution of scatterers is target-dependent and usually nonuniform, parametrized HRRP's (PHRRP) cannot be regarded as discrete signals and the normalized correlation (NC) based recognition scheme can no longer be used to classify them. In the last paper, X. Liao *et al.* present a new scheme of radar target recognition based on PHRRP's by introducing a novel generalized-weighted-normalized correlation (GWNC) for measuring the similarity between PHRRP's. With the main lobe width parameter properly chosen, the aspect sensitivity of PHRRP's can be reduced without sacrificing their discriminative power.

Finally, we would like to thank all people who have contributed to the success of this book, in particular, the authors and the reviewers for their efforts to ensure the quality of the papers. We hope you will find the book helpful.

This page is intentionally left blank

3D ARTICULATED OBJECT UNDERSTANDING, LEARNING, AND RECOGNITION FROM 2D IMAGES

PATRICK S. P. WANG

IAPR Fellow and Professor of Computer Science
Northeastern University
Boston, MA 02115, USA
E-mail: pwang@ccs.neu.edu
http://www.ccs.neu.edu/home/pwang

This paper is aimed at 3D object understanding from 2D images, including articulated objects in active vision environment, using interactive, and internet virtual reality techniques. Generally speaking, an articulated object can be divided into two portions: main rigid portion and articulated portion. It is more complicated that "rigid" object in that the relative positions, shapes or angles between the main portion and the articulated portion have essentially infinite variations, in addition to the infinite variations of each individual rigid portions due to orientations, rotations and topological transformations. A new method generalized from linear combination is employed to investigate such problems. It uses very few learning samples, and can describe, understand, and recognize 3D articulated objects while the objects status is being changed in an active vision environment.

Keywords: Articulated object recognition; feature extraction; learning; active vision; linear combination; pattern representation; thinning.

1. INTRODUCTION

Recently, the study of three-dimensional (3D) representation, description, understanding, learning and recognition of rigid objects has attracted more and more attention. It is difficult and complicated because of infinite variations due to rotation, orientation and topological transformations. *Articulated objects* are even more challenging because in addition to the above mentioned difficulties, the object itself can change shape or structure, which makes its pattern representation and matching rather difficult. Articulated objects consisting of a major portion of industrial parts play important roles in 3D pattern recognition, computer vision and industrial parts inspection.[1,2] Yet not many research work have been done in this area because of the difficulties and complexities,[3,5] and most research in the past concentrate only on rigid object recognition.[4,8−12,15,17,19,21−24] Several survey papers in this field can be found in Refs. 6, 7, 13 and 14. In Ref. 20, a new approach for analyzing and recognizing articulated objects in static environment was presented. In that method, the articulated portion of an object can be extracted by comparing two different statuses of the same view in the learning process. The observer and the object being investigated are basically stationary. This paper presents a state-of-the-art novel idea in extracting features of articulated as well as non-articulated portions of a 3D object. It is robust yet simple and needs only very few learning samples. It is also flexible in that it works in an active vision environment where the

observer moves or the object itself moves or rotates. Some examples are illustrated in the following sections. The line drawings are obtained from the original images by Hough transform and skeletonization (thinning) process, which is a very important preprocessing procedure in pattern analysis and recognition.[16] Several examples in virtual environment are illustrated. Finally, future topics of this research are discussed.

2. ANALYSIS, METHODOLOGIES AND EXPERIMENTS

We will adapt the notations and definitions in Ref. 20. A polyhedral 3D object is first transformed to a line-drawing image, which can be represented by a layered graph.

Definition 2.1. A *3D coordinated graph* (3dcg) in 3D Cartesian space is a graph $G = (V, E)$, where V is a nonempty finite set of nodes (vertices), each denoted by a Cartesian product (x, y, z), and E is a finite nonempty set of edges (branches, lines) connecting pairs of nodes in V that are neighbors. Each line is denoted by (**a, b**) if **a** and **b** are two nodes and are directly connected.

A Heuristic Algorithm for 3D object represented by 3dcg (**CGPA**).

Step 0. Unmark all input lines. Select a node **a** with maximum number of neighbors (if there are more than one such node, select the one with a longest line, else select any one of them to start).

Step 1. If **a** has n nonmarked lines connecting to $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$, then record $(\mathbf{a} - \mathbf{b}_1\ \mathbf{b}_2 \cdots \mathbf{b}_n)$. Mark all lines that have already been parsed.

Step 2. Check all $\mathbf{b}_i$ nodes *simultaneously*, where $i = 1, 2, \ldots, n$, to see if any of these $\mathbf{b}_i$ nodes have any connecting lines that are not marked yet. If no, go to **Step 3**, else, set $\mathbf{a} \leftarrow \mathbf{b}_i, \forall i = 1, \ldots, n$ such that $\mathbf{b}_i$ has nonmarked lines, and go to **Step 1**. Repeat this process till all $\mathbf{b}_i$'s have been done, and no more unmarked lines are left, then go to **Step 3**.

Step 3. Produce the parsing sequence recorded from **Steps 1** and **2** and halt.

An example is illustrated in the following Fig. 1.
An example of graph representation of the object is shown in Fig. 2.
Notice that the hinge (or joint) of a two-portion articulated object can be detected by the following algorithm in the learning process, assuming the hinge or joint is always visible amongst those learning samples in various rotations, characteristic views and statuses. Recall that a graph representation's starting nodes begin with the one with maximum weight (number of neighborhoods), and if there are more than one such node, start with any one of them[20−22]:
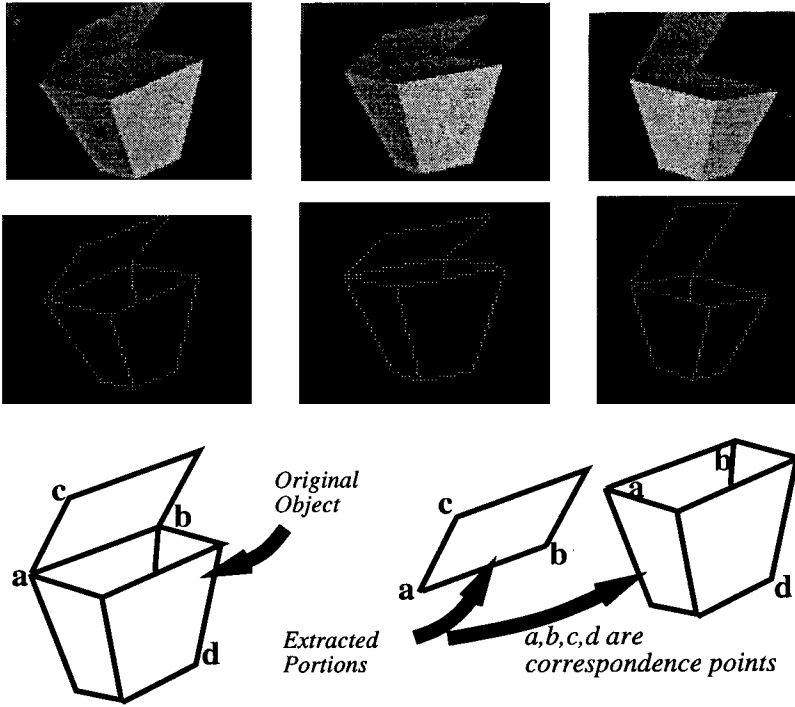
Fig. 1.   A basket with lid in various statuses and extracted portions.
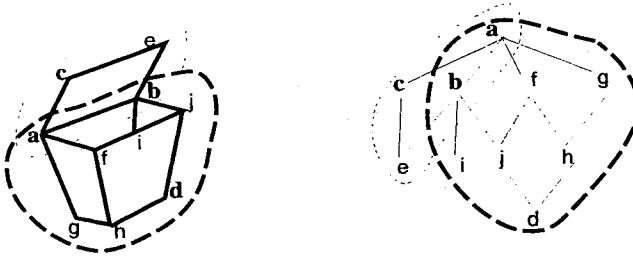
Fig. 2.   A graph representation of an *articulated object*: a basket with lid.

A Hinge Detection Algorithm

1. Find graph representation of the input line drawing, transformed from the 3D object.
2. Change status of the object by rotating the articulated portion along the hinge (joint), while the object itself may also rotate. Find its corresponding graph representation for each status and/or view.
3. Comparing with those graph representations, the hinge is the branch which has *irregular* broken successor links to its sublevel vertices (nodes).
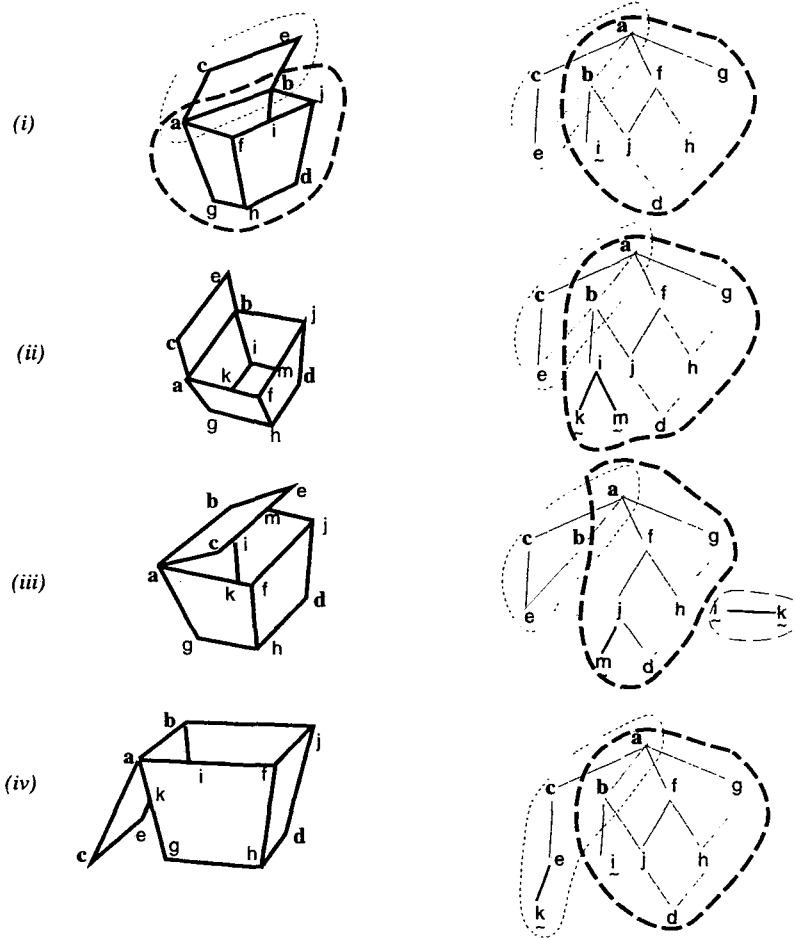
Fig. 3.   Detection of hinge (joint) a–b of an articulated object.

Extraction of Articulated Features

1. Obtain the hinge of graph representation.
2. Find a complete sequence of encircled branch from one end of the branch.
3. Output this subset of graph, and represent it as the articulated portion of the object.

    This can be illustrated in Fig. 3. There are four different views and statuses of the same articulated object, i.e. a basket with lid. Each is represented by its corresponding line drawings and graph representations, respectively. Note that each dangling node (vertex) is denoted by a s̃ign.[20]

    Once the hinge is detected, from the branch that associates with the hinge, we can find a complete sequence of branches that encircles a subset of the graph, representing the *articulated* portion of the object, as illustrated in the *light dashed encircled lines* representing **a–b–e–c**.
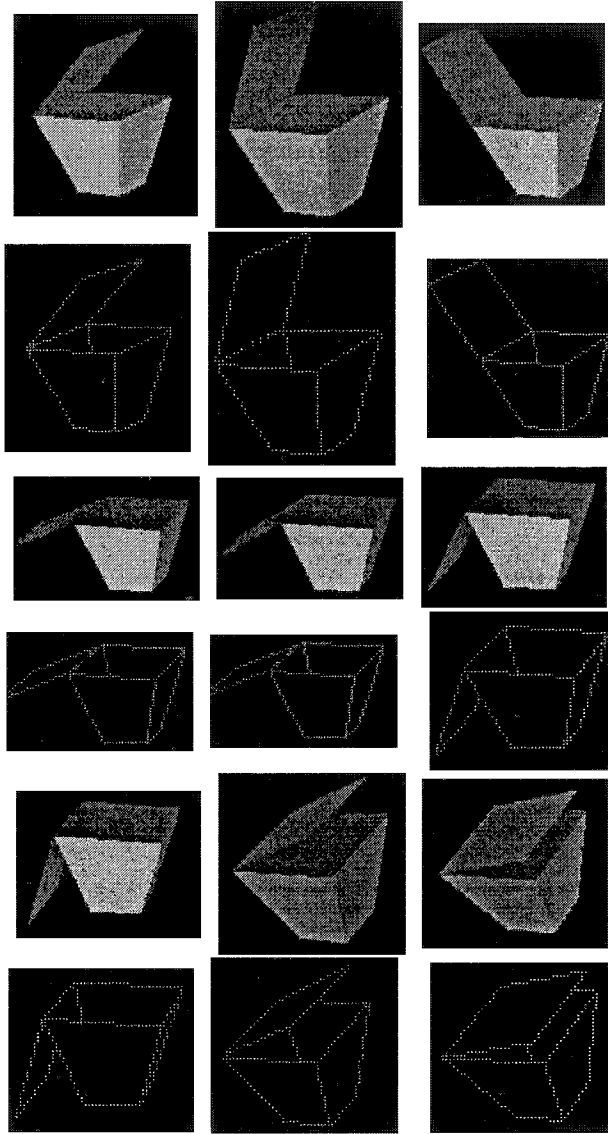
Fig. 4. Some examples of articulated object views and statuses. Dotted lines are thinned skeletons of the line drawings obtained from Hough transform of the original images.

Several partitions of characteristic views (CV's) of a basket with lid are shown in the following figures:

Some more examples of mixed characteristic views are shown below. Notice that these figures all together illustrate a continuous movement and rotation of a basket with lid, while its lid is *opening or closing*, i.e. with various statuses.

All these images will be divided into two portions, i.e. the articulated portion of the *lid*, and the main portion of the *open basket* as illustrated in Fig. 1.
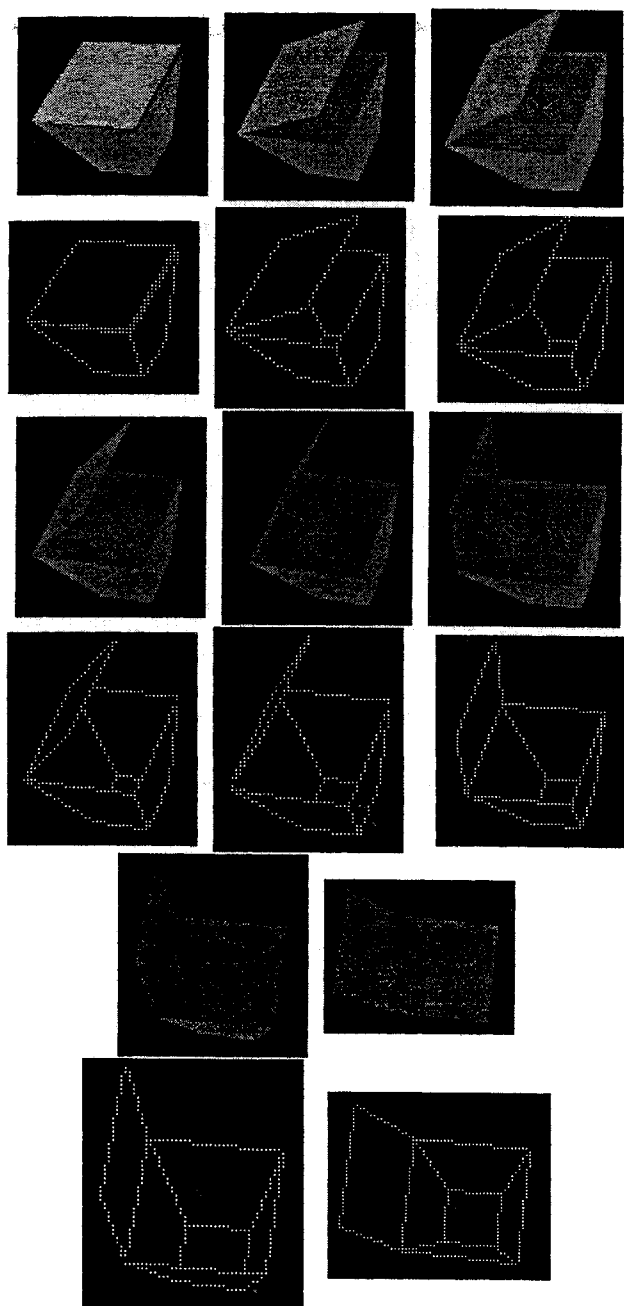
Fig. 5.   Some more examples of an articulated object with mixed characteristic views and statuses.

Note that the *articulated* portion is a relative term, i.e. one can also consider the *open basket* as the *articulated* portion, thereby, the *lid* as the *main portion*, and vice versa.
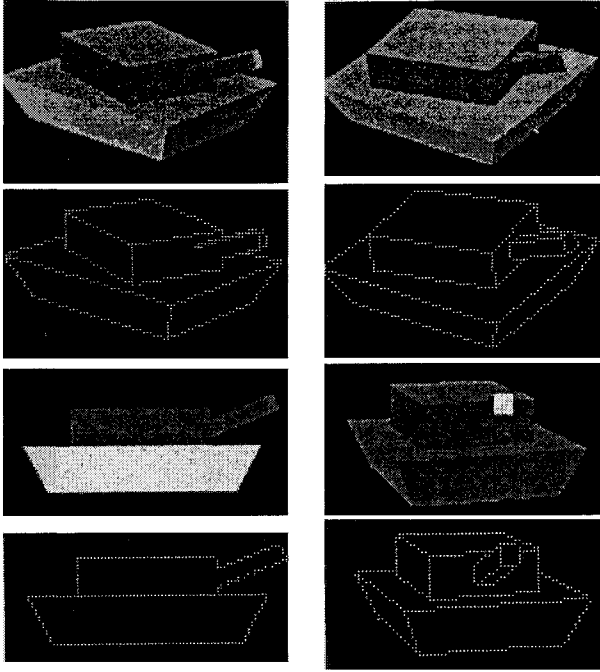
Fig. 6. Some examples of model tanks as articulated objects.

Once the *articulated* portion of the object is extracted, pattern matching can be pursued repectively in each of the portion by the *linear combination method*.[2,20,22] That is, each extracted portion shall be a linear combination of several views of the corresponding portion in the learning sample image database of the object being investigated. Notice that one of the advantages of this method is that it can save recognition time, especially for those which are not being accepted, because when the graph representations do not match with the learning samples, we do not even have to go further for any other measurement matching.[2,22] In our laboratory, we have established a set of image database of 3D articulated objects to be experimented on its 3D pattern representation, matching, and recognition, including industrial parts and military target identifications, as illustrated in the following figure, Fig. 6.

3. DISCUSSIONS, CONCLUSION AND FUTURE RESEARCH

We have presented a methodology that can extract the features of an *articulated* portion of a 3D object that consists of two portions: the *articulated* portion and the *main* portion.

Comparing with another method of extracting an *articulated* portion of an object in Ref. 22, which requires that the object being investigated is *still* or *stationary*, this new method has the flexibility of allowing either the observer's movement or the object's rotation or both, while the status of the object can vary at the same
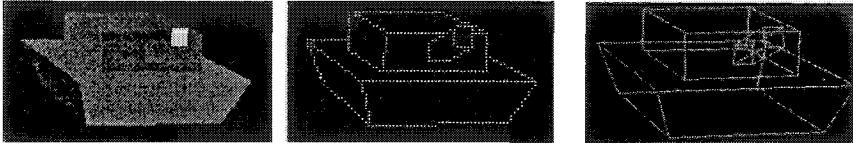
Fig. 7.   Examples of objects, line drawings and wire frames.
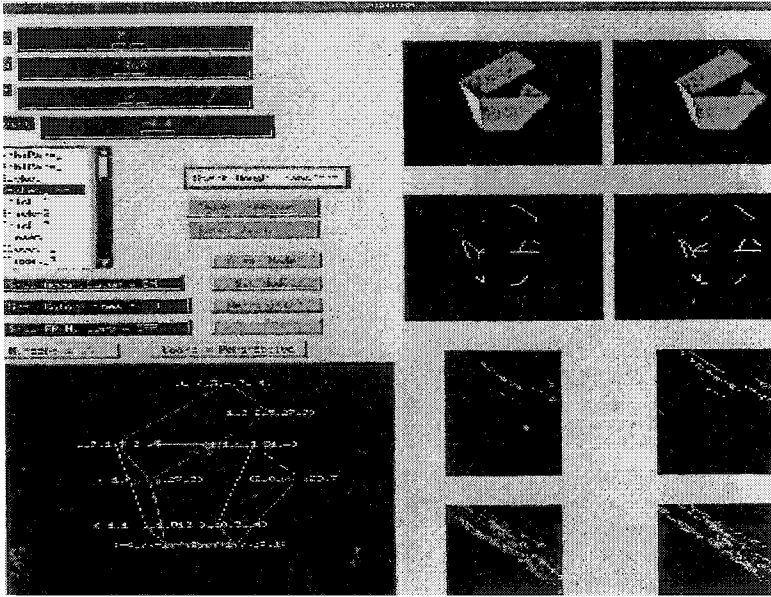
Fig. 8.   Example of an x-window 3D object recognition system.[23]

time, i.e. the articulate portion can change positions, shapes or angles relative to the *main* portion of the object. This method also requires very few learning samples in contrast to others in the literature.[3,4,7] Notice, however, that after Hough transform, sometimes a line will be broken into pieces as shown in Fig. 6. This will affect its graph representation. One way to remedy this is to "patch them" up before graph representation. There are many interesting examples of articulated objects with three or more portions (e.g. a tank). In these cases, there will be too many characteristic views although still finite. An improvement is to represent their "wire frames" which are transparent to viewers, as illustrated in Figs. 7 and 8.[23] A more advanced improved version with illustrations can be found in Ref. 24.

This will significantly reduce the number of graph representations and simplify the complexities of pattern matching. Also the hinge or joint may not be so clear or visible like a tank. How can it be represented and extracted? How can our method described in this paper be extended to those objects? How about curved objects in a more complicated occluded environment? These are interesting topics for future research. Notice that in our method the correspondence points are chosen manually and the success depends on the choice of threshold values. As shared with
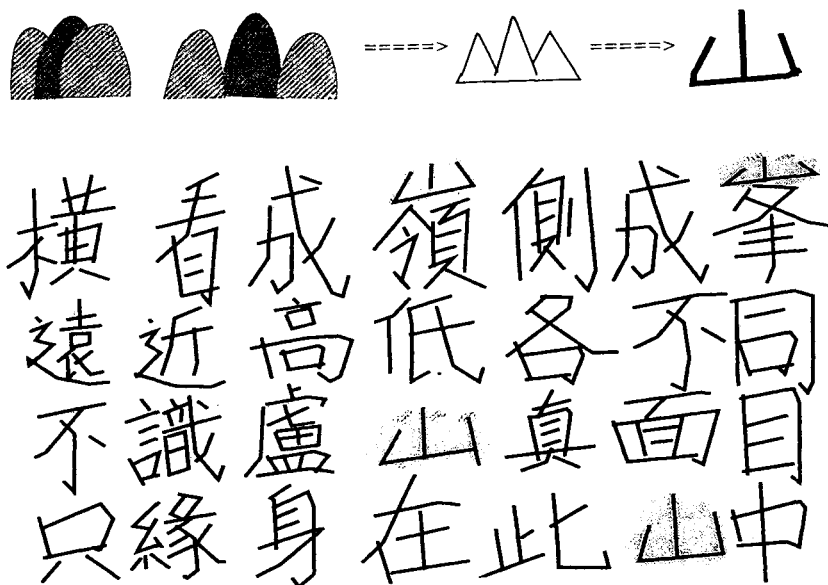
Looking from one side, mountains, from the other side, ranges
Far, near, high, low, all different
I can hardly recognize the true face of the Lu Shan mountain
Simply because I am in the middle of it

Fig. 9. A famous poem by a well-known Chinese poet Su Dong-po (11th Century).

some drawbacks of LC method,[2] certain objects are misrecognized or incorrectly rejected. In future, we would like to explore automatic determination of *correspondence points*, and overcome some difficulties of misrecognition due to poor choice of threshold values. An inherent difficulty of recognizing 3D objects seems due to its so many *variations*, as many as *infinite*, from various angles, sizes and statuses. Interesting enough, this basic natural phenomenon happens to coincide with a famous poem written by a famous Chinese poet Su Dong-po about one thousand years ago, as shown in Fig. 9.

A more recent new version of improved learning and recognition system using internet and JAVA, with live illustrations can be found in the author's homepage at Refs. 23, 24. Readers can conduct their own recognition tests from their own computers using any browser such as PC-IE or Netscape Communicator, as long as it supports JDK 1.1.6 or above.

ACKNOWLEDGMENTS

REFERENCES

1. L. Baird and P. S. P. Wang, "3D object recognition using gradient descent and the universal 3D array grammar," *SPIE on Intelligent Robots and Computer Vision X: Algorithms and Techniques*, ed. D. Casasent, Vol. 1607, 1992, pp. 711–719.
2. R. Basri, "Viewer-centered representations in object recognition — a computational approach," *Handbook of Pattern Recognition and Computer Vision*, eds. C. Chen, L. Pau and P. S. P. Wang, World Scientific, 1993, pp. 863–882.
3. A. Beinglass and H. Wolfson, "Articulated object recognition, or how to generalize the generalized Hough transform," *Proc. ICCVPR*, 1991, pp. 461–466.
4. P. Besl and R. Jain, "3D object recognition," *ACM Comput. Surv.* **17**, 1 (1985) 75–154.
5. A. F. Bobick and R. C. Bolles, "The representation space paradigm of concurrent evolving object descriptions," *IEEE–PAMI* **14**, 2 (1992) 146–156.
6. I. Chakravarty and H. Freeman, "Characteristic views as a basis for 3D object recognition," *SPIE: Robot Vis.* **336** (1982) 37–45.
7. R. T. Chin and C. R. Dyer, "Model-based recognition in robot vision," *ACM Comput. Surv.* **18**, 1 (1986) 67–108.
8. S. I. Dickinson, A. P. Pentland and A. Rosenfeld, "3D shape recovery using distributed aspect matching," *IEEE–PAMI* **14**, 2 (1992) 174–197.
9. Y. Leclerc and M. Fishler, "An optimization-based approach to the interpretation of simple line drawings as 3D wire frames," *IJCV* **9**, 2 (1992) 113–136.
10. D. B. Lysak Jr. and R. Kasturi, "Interpretation of engineering drawings of polyhedral and nonpolyhedral objects," *Proc. IEEE-ICDAR'91 (Int. Conf. Document Analysis and Recognition)*, Saint-Malo, France, September 1991, pp. 79–87.
11. T. Marill, "Emulating the human interpretation of line-drawings as 3D objects," *IJCV* **6**, 2 (1991) 147–161.
12. E. Marti, J. Regincós, J. López-Krahe and J. Villanueva, "A system for interpretation of hand line drawings as 3D scene for CAD input," *Proc. ICDAR'91*, Saint-Malo, France, September 1991, pp. 472–481.
13. I. V. Nagendra and U. G. Gujar, "3D objects from 2D orthographic views — a survey," *Comput. Graph.* **12**, 1 (1998) 111–114.
14. S. Negahdaripour and A. K. Jain, "Challenges in computer vision: future research directions," *IEEE-CVPR'92*, 189–198.
15. A. Parent and P. Wang, "Distance recovery of 3D objects from stereo images using Hough transform," *SPIE Intelligent Robots and Computer Vision XIII*, Vol. 2354, 1994, pp. 348–356.
16. C. Y. Suen and P. S. P. Wang (eds.), *Advances of Thinning Method for Pattern Recognition*, World Scientific, 1994.
17. K. Sugihara, *Machine Interpretation of Line Drawings*, MIT Press, Cambridge, 1986.
18. P. S. P. Wang, "3D object understanding from 2D images," *Proc. SPIE Int. Conf. ISMIP*, 1998, pp. 33–43.
19. P. S. P. Wang, "3D line image analysis — a heuristic parallel approach," *Int. J. Inform. Sci.* **81**, 3 (1994) 155–175.
20. P. S. P. Wang, "Visualization and understanding of 3D objects — a model-based approach," *SPIE Intelligent Robots and Computer Vision XIII*, Vol. 2353, 1994, pp. 10–21.
21. P. S. P. Wang, "A heuristic parallel algorithm for line-drawing object pattern representation and recognition," *Advances in Image Processing*, ed. E. Dougherty, Marcel Dekker, 1994, pp. 197–222.
22. P. S. P. Wang, "Perception and visualization of line images," *Proc. IEEE-Int. Conf. Image Processing*, 1995.

23. P. S. P. Wang, http://www.ccs.neu.edu/home/pwang/ObjectRecognition, January, 1999.
24. P. S. P. Wang, http://www.ccs.neu.edu/home/pwang/3dpr, February, 2000.

Patrick S. P. Wang is an IAPR Fellow and has been tenured full professor of computer science at Northeastern University since 1983, a research consultant at MIT Sloan School since 1989, and adjunct faculty master of computer science at Harvard University Extension School since 1985. He received his Ph.D. in computer science from Oregon State University, his M.S. in I.C.S. from Georgia Institute of Technology, his M.S.E.E. from National Taiwan University and his B.S.E.E. from National Chiao Tung University. He was on the faculties of the University of Oregon and Boston University, and senior researcher at Southern Bell, GTE Labs, and Wang Labs prior to his present position. In addition to his research experience at MIT AI Lab, Prof. Wang has been a visiting professor and has been invited to give lectures, do research and present papers in more than a dozen countries in Europe and Asia and many universities and industries in the U.S.A. and Canada. He has published over 120 technical papers and 18 books in pattern recognition, AI and imaging technologies and has three OCR patents by US and Europe Patent Bureaus. As IEEE senior member he has organized numerous international conferences and workshops and served as reviewer for many journals and NSF grant proposals. Prof. Wang is currently founding Editor-in-Charge of the *International Journal of Pattern Recognition and Artificial Intelligence*, Editor-in-Chief of the series on *Machine Perception and Artificial Intelligence* by World Scientific Publishing Co., and elected chair of IAPR-SSPR (International Association of Pattern Recognition). In addition to his technical interests, he has also written several articles on the operas of Verdi, Puccini, and Wagner, and the symphonies of Mozart, Beethoven, and Tchaikovsky. Dr. Wang has been honored as "Otto-von-Guericke" Distinguished Guest Professor of Magdeburg University of Germany, Fall 1996.

This page is intentionally left blank

ON GEOMETRIC AND ORTHOGONAL MOMENTS

JUN SHEN

Image Laboratory, Institute EGID, Bordeaux-3 University
Allée F. Daguin, 33607 Pessac, France
E-mail: shen@egid.ubordeaux.fr

WEI SHEN

Department of Electronics
Zhongshan University, Canton, P.R. China

DANFEI SHEN

Ecole Nationale Supérieure des Télécommunications
Paris, France

Moments are widely used in pattern recognition, image processing, computer vision and multiresolution analysis. To clarify and to guide the use of different types of moments, we present in this paper a study on the different moments and compare their behavior. After an introduction to geometric, Legendre, Hermite and Gaussian–Hermite moments and their calculation, we analyze at first their behavior in spatial domain. Our analysis shows orthogonal moment base functions of different orders having different number of zero-crossings and very different shapes, therefore they can better separate image features based on different modes, which is very interesting for pattern analysis and shape classification. Moreover, Gaussian–Hermite moment base functions are much more smoothed, they are thus less sensitive to noise and avoid the artifacts introduced by window function discontinuity. We then analyze the spectral behavior of moments in frequency domain. Theoretical and numerical analyses show that orthogonal Legendre and Gaussian–Hermite moments of different orders separate different frequency bands more effectively. It is also shown that Gaussian–Hermite moments present an approach to construct orthogonal features from the results of wavelet analysis. The orthogonality equivalence theorem is also presented. Our analysis is confirmed by numerical results, which are then reported.

Keywords: Image characterization; classification; moments; orthogonal moments; behavior analysis; frequency analysis; orthogonality equivalence.

1. INTRODUCTION

Moments, such as geometric moments and orthogonal moments, are widely used in pattern recognition, image processing, computer vision and multiresolution analysis. In order to better represent local characteristics of images, smoothed orthogonal moments were also recently proposed.

To clarify and to guide the use of different types of moments, an analysis of their behavior both in spatial and frequency domains would be necessary. In this paper, we analyze and compare the behavior of different moments, in particular, geometric moments, Legendre moments, Hermite moments and Gaussian–Hermite moments. Our paper is organized as follows. In Sec. 2, we present geometric, Legendre and Hermite moments, and their computation. We analyze and compare the behavior

of these moments in Sec. 3, in spatial domain and in Sec. 4, in the frequency domain. In Sec. 5, we introduce the Gaussian–Hermite moments, their calculation, the relation between Gaussian–Hermite moments and the wavelet analysis, and the orthogonality equivalence theorem. A comparison with other moments is also reported. The paper is terminated by some concluding remarks, which are presented in Sec. 6.

2. GEOMETRIC, LEGENDRE AND HERMITE MOMENTS AND THEIR CALCULATION

2.1. Geometric Moments and Their Calculation

Geometric moments and their calculation have found wide applications in pattern recognition, linear filtering, edge detection, image segmentation, multiresolution analysis, texture analysis, and other domains of image processing and computer vision.[1,3,7,9,14−16,18,21,27,28] Geometric moments of a 1D signal $S(x)$ are defined by

$$M_n(x) = \int_{-w}^{w} S(x+t)t^n \, dt \qquad n = 0, 1, 2, \ldots \qquad (1)$$

where $M_n(x)$ is the moment of order n calculated from a window of size $(2w+1)$ pixels centered at the point x. Geometric moments of a 2D image $I(x, y)$ are defined similarly, i.e.

$$M_{m,n}(x, y) = \int_{-w1}^{w1} \int_{-w2}^{w2} I(x+u, y+v)u^m v^n \, du \, dv \qquad m, n = 0, 1, 2, \ldots \qquad (2)$$

where $M_{m,n}(x)$ is the moment of order (m, n) calculated from a window of size $(2w_1 + 1) \times (2w_2 + 1)$ pixels centered at the pixel (x, y).

Due to the large computational complexity of moment calculation, fast algorithms for geometric moment computation and their application have been proposed.[8,10−13,29] As is well known, one of the advantages of geometric moments is that they have an explicit geometrical and statistical significance. In fact, the geometrical moments used in pattern recognition and image processing are exactly those defined in statistics. If we consider the geometric moments from the viewpoint of functional analysis, they are projections of the input signal $S(x)$ onto the polynomial function space, taking the monomial functions $1, x, x^2, \ldots, x^N, \ldots$ as bases.

2.2. Orthogonal Legendre Moments and Their Calculation

Obviously, when one projects a signal onto the axes of a space, generally, one had better use orthogonal bases. The advantage of using orthogonal bases is that the orthogonality between the bases, helps in reducing the calculation; the error is easier to estimate when a limited number of projections is used, and reconstruction could also be more simple.[5,6,26] So orthogonal moments are proposed and their performance is analyzed.[19,20,22−24] Taking the polynomial space to analyze the signal, the orthogonal bases are known as Legendre polynomials.[26]

Legendre polynomials are defined as

$$P_n(x) = \begin{cases} (d^n/dx^n)(x^2 - 1)^n/(2^n \cdot n!) & \text{for } x \in [-1, 1] \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $n = 0, 1, 2, \ldots$, is the order of the Legendre polynomial and $n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 1$. Based on the Legendre polynomials defined above, we can define the scaled Legendre polynomials as

$$L_n(x) = \begin{cases} [(d^n/dx^n)(x^2 - w^2)^n]/[(2w)^n \cdot n!] & \text{for } x \in [-w, w] \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where n is the order of the scaled Legendre polynomial, $n = 0, 1, 2, \ldots$, and $w > 0$, the scale factor.

The functions $L_n(x)$, $n = 0, 1, 2, \ldots$, are orthogonal polynomials on the interval $[-w, w]$, i.e.

$$\int_{-w}^{w} L_m(x) L_n(x) dx = 0 \qquad \text{for } m \neq n. \tag{5}$$

Given a signal $S(x)$, by use of the scaled Legendre polynomial $L_n(x)$ above, we can define the nth order orthogonal Legendre moment $M_n(x)$ as follows

$$M_n(x) = \int_{-w}^{w} S(x + t) L_n(t) \, dt$$

$$= \langle L_n(t), S(x + t) \rangle \tag{6}$$

where $\langle f_1(t), f_2(t) \rangle$ means the inner product of f_1 and f_2.

By use of the orthogonality, we have shown that the nth order orthogonal Legendre moments can be calculated by using the following theorem:[22-24]

Theorem of recursive calculation of Legendre moments. The nth order orthogonal Legendre moments of $S(x)$, calculated from the window $[x - w, x + w]$, can be computed from the $(n-1)$th and $(n-2)$th order Legendre moments as follows

$$M_0(x) = \langle L_0(t), S(x + t) \rangle = S_1(x + w) - S_1(x - w) \tag{7}$$

$$M_1(x) = \langle L_1(t), S(x + t) \rangle = [S_1(x + w) + S_1(x - w)] - \langle L_0(t), S_1(x + t) \rangle/w \tag{8}$$

$$M_n(x) = \langle L_n(t), S(x + t) \rangle = \langle L_{n-2}(t), S(x + t) \rangle$$
$$- [(2n - 1)/w] \langle L_{n-1}(t), S_1(x + t) \rangle \quad \text{for } n \geq 2 \tag{9}$$

where

$$\langle L_0(t), S_i(x + t) \rangle = S_{i+1}(x + w) - S_{i+1}(x - w) \tag{10}$$

$$\langle L_1(t), S_i(x + t) \rangle = [S_{i+1}(x + w) + S_{i+1}(x - w)]$$
$$- \langle L_0(t), S_{i+1}(x + t) \rangle/w \tag{11}$$

$$\langle L_n(t), S_i(x + t) \rangle = \langle L_{n-2}(t), S_i(x + t) \rangle - [(2n - 1)/w]$$
$$\times \langle L_{n-1}(t), S_{i+1}(x + t) \rangle \quad \text{for } n \geq 2 \tag{12}$$

with

$$S_0(t) = S(t)$$

and

$$S_i(t) = \int_{-\infty}^{t} S_{i-1}(y)\,dy \qquad \text{for } i = 1, 2, 3, \ldots \tag{13}$$

As $S_i(t)$ can be easily calculated from $S_{i-1}(t)$ by use of the recursive sum-box technique,[2,17] the recursive calculation of Legendre moments according to the above theorem reduces the computational complexity considerably.

In 2D cases, given an image $I(x, y)$, the 2D Legendre moment of order (p, q) in a 2D window centered at (x, y) is defined as

$$M_{p,q}(x, y) = \int_{-k_x}^{k_x} \int_{-k_y}^{k_y} I(x + t, y + v) L_p(t) L_q(v)\,dt\,dv \tag{14}$$

Obviously, 2D Legendre moments are separable, so the calculation of 2D Legendre moments can be decomposed into the cascade of 1D Legendre moment calculation, and the recursive algorithm applies. Table 1 shows a comparison between the computational complexity of the recursive method and that of the classical method using 2D masks. Readers can see Refs. 22–24 for more details.

Table 1.   Comparison of computational complexity per pixel for calculating all 2D Legendre moments of order (p, q) with $p + q \leq N$.

Window size $(2k_x + 1)$ $\times (2k_y + 1)$	Max. moment order N $(p + q \leq N)$	Classical algorithm		Recursive algorithm	
		Nb. of ×	Nb. of +	Nb. of ×	Nb. of +
	3	490	490	5	124
7 × 7	5	1715	1715	22	299
	7	4116	4116	65	570
	3	810	810	5	124
9 × 9	5	2835	2835	22	299
	7	6804	6804	65	570
	3	1210	1210	5	124
11 × 11	5	4235	4235	22	299
	7	10164	10164	65	570
	3	2250	2250	5	124
15 × 15	5	7875	7875	22	299
	7	18900	18900	65	570
	3	4410	4410	5	124
21 × 21	5	15435	15435	22	299
	7	37044	37044	65	570
	3	9610	9610	5	124
31 × 31	5	33635	33635	22	299
	7	80724	80724	65	570
	3	26010	26010	5	124
51 × 51	5	91035	91035	22	299
	7	218484	218484	65	570

2.3. Hermite Moments

Another family of orthogonal polynomial functions is Hermite polynomials,[6,26] which can be used as moment base functions. A scaled Hermite polynomial of order n is defined as

$$P_n(t) = H_n(t/\sigma) \tag{15}$$

with

$$H_n(t) = (-1)^n \exp(t^2)(d^n/dt^n) \exp(-t^2). \tag{16}$$

The 1D nth order Hermite moment $M_n(x, S(x))$ of a signal $S(x)$ can therefore be defined as

$$M_n(x, S(x)) = \int_{-\infty}^{\infty} P_n(t) S(x+t)\, dt \qquad n = 0, 1, \ldots \tag{17}$$

where $P_n(t)$ is the scaled Hermite polynomial function defined in Eq. (15).

This moment definition can be easily generalized for 2D cases, and the 2D Hermite moments centered at (x, y) of an image $I(x, y)$ are defined as

$$M_{p,q}(x, y, I(x, y)) = \iint_{-\infty}^{\infty} H_{p,q}(t/\sigma, v/\sigma) I(x+t, y+v)\, dt\, dv \tag{18}$$

with

$$H_{p,q}(t/\sigma, v/\sigma) = H_p(t/\sigma) H_q(v/\sigma). \tag{19}$$

Obviously, by use of the separability of 2D Hermite base functions, 2D Hermite moments can be calculated by the cascade of 1D moments, which reduces the computational complexity.

3. COMPARISON OF SPATIAL BEHAVIOR OF GEOMETRIC, LEGENDRE AND HERMITE MOMENTS

In this section, we analyze and compare the spatial domain behavior of geometric, Legendre and Hermite moments. Figure 1 shows the graphs of different 1D moment base functions of different orders. In order to better show the zero-crossings and the oscillations of Hermite base functions, the pseudologarithm is used for geometric and Hermite moment base functions to visualize the function values. The pseudologarithm of $f(x)$ is defined by $\mathrm{sign}(f(x)) \log(1 + |f(x)|)$, such that 0 will be mapped to 0, and the visualized value reserves the sign of the original value.

From Fig. 1, we see that for geometric moment base functions, the graphs have almost similar shapes independent of the moment order, i.e. all base functions of odd orders are monotone, and all base functions of even orders are monotone with respect to the absolute value of x. The moments of a signal can be considered as projections of the signal onto the base function space, so if we want to characterize images of different spatial modes by the moments, the geometric moments would not be very efficient. On the contrary, for Hermite and Legendre moment base functions, we see many oscillations of the functions depending on the moment order.
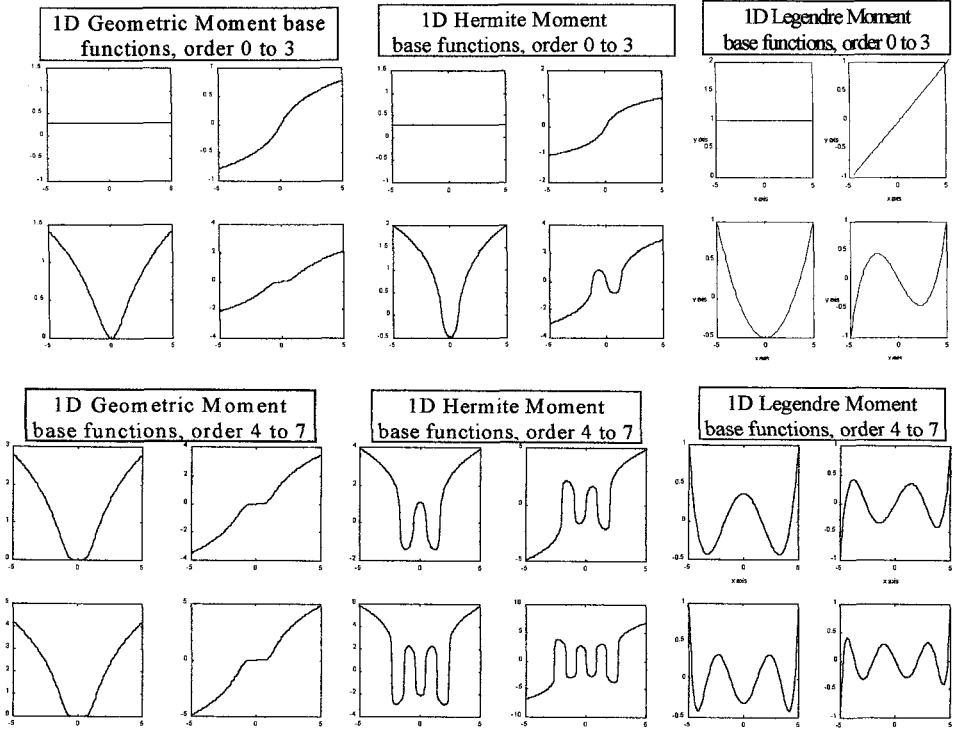
Fig. 1.   Graph of different 1D moment base functions of different orders.

This means that when we characterize the signal by the use of these moments, we extract efficiently the characteristics of different spatial modes of the signal. The reason for this efficiency can be easily understood from the properties of orthogonal polynomials. In fact, an orthogonal polynomial of order n has always n different real roots, i.e. n different zero-crossings, moments of different orders therefore correspond to different spatial modes. Moreover, we see that the oscillations exhibited in Hermite moment base functions are much less important than that of Legendre ones. So the Hermite moments represent different spatial modes less effectively than Legendre ones. This is because the Hermite bases themselves are not really orthogonal if we do not use a weight function, as is the case for Hermite moment calculation. In other words, we can say that the Legendre bases are really orthogonal, the Hermite bases are closer to orthogonal than the geometric moment bases which are not at all orthogonal. The same conclusion holds in 2D cases, which can be seen in Fig. 2.

Of course, from the viewpoint of functional analysis, because all these families of base functions are complete, the initial signal given can be reconstructed from the moments, no matter which kind of moments are used. So all the geometric, Hermite or Legendre moments could be used to characterize a signal. But from the practical application viewpoint, if we want to characterize different spatial modes,
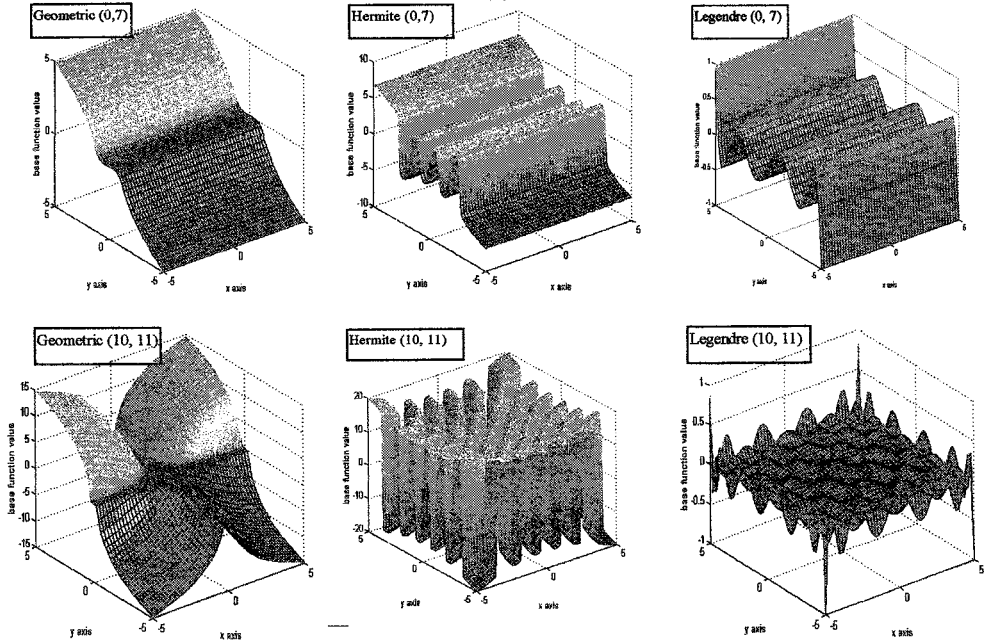
Fig. 2.   Graph of different 2D moment base functions of different orders.
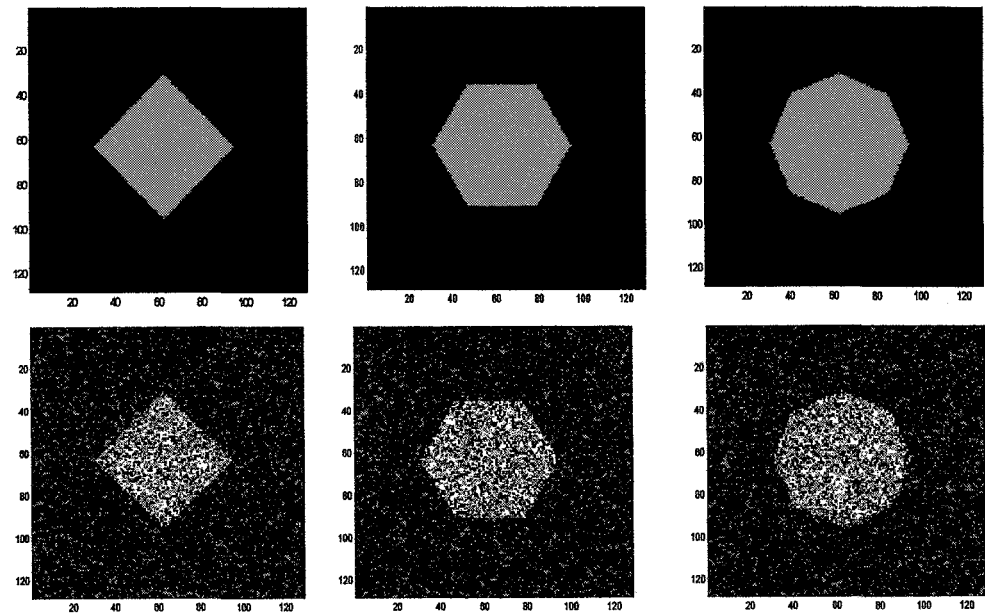

Fig. 3.   Three reference shape images and examples of noisy images.

such as to distinguish different forms for classification, Legendre moments would be more effective.

To verify the conclusion above, we performed the following experiment. The objective is to classify noisy images by using different moments. Three different reference shape images: quadrilateral, hexagon and octagon, are used as models. From these reference images, we generate noisy images by adding random noises of different standard deviations. Figure 3 shows the three different reference shape images: quadrilateral, hexagon and octagon, and an example of their noisy images.

Each shape (including reference and noisy shapes) is then characterized by twelve 2D moments of orders $(0,0), (0,1), \ldots, (0,5), (1,0), (1,1), \ldots, (1,5)$, i.e. each shape is represented by a moment vector of twelve dimensions. Geometric, Hermite and Legendre moments are tested respectively. The classification is done by comparing the moment vector of a noisy shape with those of the three reference shapes. Obviously, in order to well characterize the shapes, for each noisy shape, the difference (distance in moment space) between the moment vector and its reference shape should be as little as possible, and the distance between the moment vector and the other reference shapes should be as great as possible. Table 2 shows the experimental results of the distances when using different kinds of moments. Notice that in order to facilitate the comparison of performances, the distances listed in Table 2 are normalized with respect to the distance to the reference shape to which the noisy shape originally belongs.

From Table 2, we see that Hermite moments better characterize the noisy shapes than the geometric ones, and Legendre moments best characterize the noisy shapes.

4. BEHAVIOR OF GEOMETRIC HERMITE AND LEGENDRE MOMENTS IN FREQUENCY DOMAIN

In this section, we analyze and compare the behavior of geometric, Hermite and Legendre moments in frequency domain. Figure 4 shows the Fourier transform amplitude of these moments of orders 0 through 7.

From Fig. 4, we see that almost all the geometric moment base functions are typical low-pass kernel, no matter how large the order is. The amplitude of the Fourier transform is always monotonically decreased when the frequency is increased from frequency 0 or a frequency very close to 0. As to Hermite moment base functions, when the order is increased, the maximum Fourier amplitude position moves more and more to the right and the frequency domain behavior becomes more and more similar to a band-pass kernel. If this property is not well shown in Fig. 4 because of the very limited graph size, one can see it better in Fig. 5 where Fourier transform amplitude of higher order moment base functions is shown. Legendre moment base functions exhibit best the band-pass characteristics. From Figs. 4 and 5, we see that except for functions of extremely low orders, all the Legendre base functions rather represent band-pass kernels than low-pass ones. The higher the order is, the more to the right the pass-band moves.

We can therefore conclude that Legendre moments separate characteristics in different frequency bands better than Hermite moments, which in turn separate

Table 2. Experimental results of distance between noisy and reference shapes when using different moments.

	Noise standard deviation	Distance to Ref. shape by use of geometric moment vector			Distance to Ref. shape by use of Hermite moment vector			Distance to Ref. shape by use of Legendre moment vector		
		Reference quadrilateral shape	Reference hexagon shape	Reference octagon shape	Reference quadrilateral shape	Reference hexagon shape	Reference octagon shape	Reference quadrilateral shape	Reference hexagon shape	Reference octagon shape
Noisy quadrilaterals	10	1	1.0111	1.0425	1	1.8521	3.4395	1	69.5791	100.872
	30	1	1.0002	1.0007	1	1.0202	1.0888	1	16.3091	23.9741
	50	1	1.0004	1.0014	1	1.0323	1.1210	1	22.4044	32.4188
	70	1	1.0006	1.0022	1	1.0399	1.1601	1	12.3208	17.7907
Noisy hexagon	10	1.5636	1	1.6236	3.7325	1	4.5302	39.8008	1	19.4541
	30	1.0000	1	1.0000	1.0045	1	1.0048	20.2906	1	9.5398
	50	1.0007	1	1.0007	1.0482	1	1.0632	17.3557	1	7.3535
	70	1.0000	1	1.0001	1.0043	1	1.0177	6.9162	1	4.2629
Noisy octagon	10	1.0058	1.0014	1	1.5838	1.1900	1	211.769	66.9128	1
	30	1.0005	1.0001	1	1.0475	1.0125	1	20.5961	5.9715	1
	50	1.0017	1.0004	1	1.1890	1.0575	1	21.2934	7.2966	1
	70	1.0000	1.0000	1	1.0037	1.0011	1	12.5002	3.9836	1

FT of geometric mmt, order 0 to 3   FT of Hermite mmt, order 0 to 3   FT of Legendre mmt, order 0 to 3

FT of Geometric mmt, order 4 to 7   FT of Hermite mmt, order 4 to 7   FT of Legendre mmt, order 4 to 7
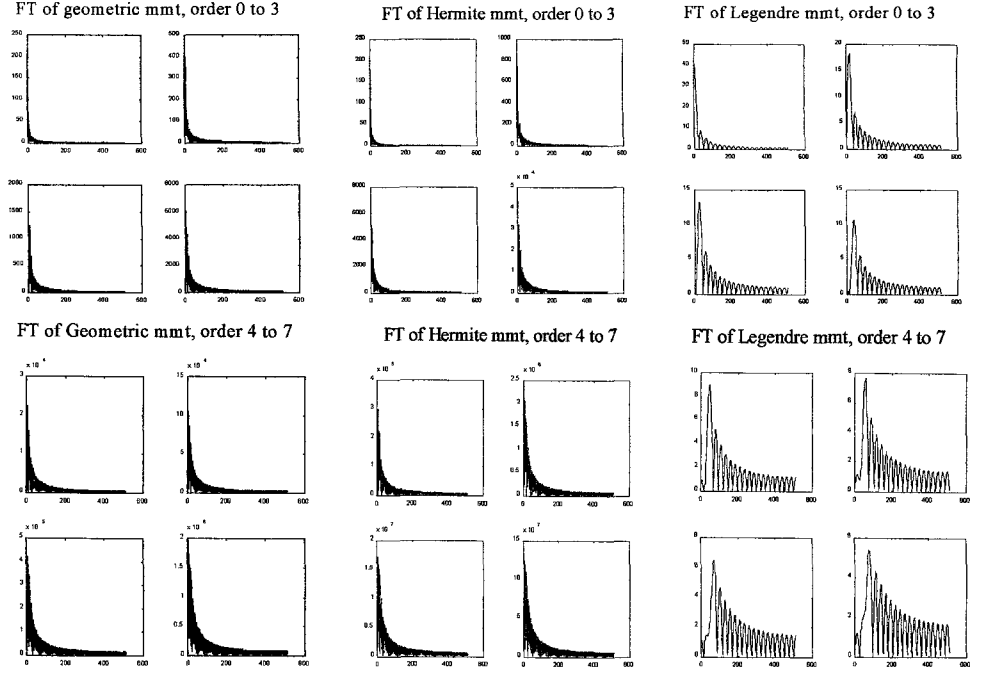
Fig. 4. Fourier transform amplitude of geometric, Hermite and Legendre moment base functions.

FT of Geometric mmt, order 22 to 25   FT of Hermite mmt, order 22 to 25   FT of Legendre mmt, order 22 to 25
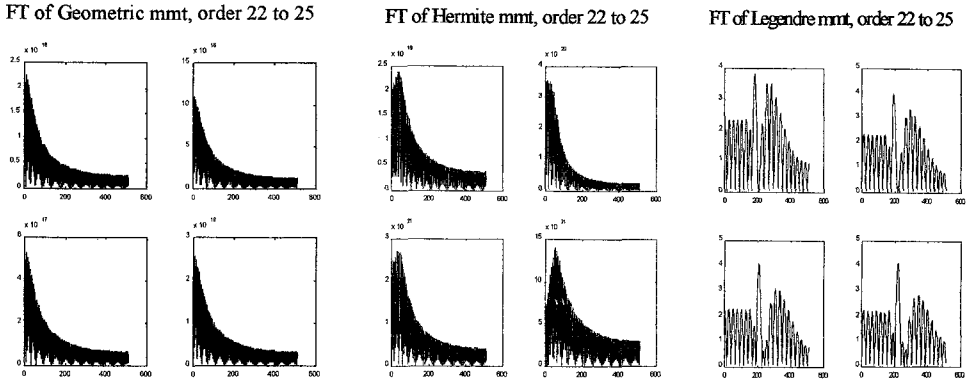
Fig. 5. Fourier transform amplitude of high order geometric, Hermite and Legendre moment base functions.

different frequency bands better than geometric ones. So from the frequency analysis viewpoint, it seems that the orthogonal Legendre moments characterize images more efficiently, which is also supported by the experimental results in Table 2.

On using the conventional frequency window analysis technique representing frequency characteristics by the band center and band width, which are widely used in short-time Fourier analysis and wavelet analysis, we can better see the difference between the moment base functions. Table 3 shows these characteristics

Table 3. Frequency window characteristics of geometric, Hermite and Legendre moment base functions.

Base function order	Geometric moment base function		Hermite moment base function		Legendre moment base function	
	Center	Effective bandwidth	Center	Effective bandwidth	Center	Effective bandwidth
0	0.40	2.11	0.40	2.11	0.40	2.11
1	1.25	3.87	1.25	3.87	1.25	3.87
2	1.59	4.71	1.64	4.77	1.94	4.90
3	2.32	5.74	2.42	5.86	2.60	5.74
4	2.61	6.23	2.79	6.45	3.25	6.46
5	3.28	7.00	3.56	7.30	3.92	7.15
6	3.52	7.36	3.93	7.78	4.59	7.77
7	4.15	7.98	4.69	8.50	5.29	8.41
8	4.36	8.27	5.10	8.94	6.01	9.01
9	4.95	8.79	5.89	9.59	6.77	9.64
10	5.14	9.03	6.36	10.03	7.55	10.23

for orders from 0 through 10. In the computation, the center and effective bandwidth of frequency window are defined as

$$\text{Center } \omega_0 = \left[\int_0^{+\infty} \omega |F(\omega)|^2 d\omega \right] \Big/ \left[\int_0^{+\infty} |F(\omega)|^2 d\omega \right] \qquad (20)$$

$$\text{Effective bandwidth} = \left\{ \left[\int_0^{+\infty} (\omega - \omega_0)^2 |F(\omega)|^2 d\omega \right] \Big/ \left[\int_0^{+\infty} |F(\omega)|^2 d\omega \right] \right\}. \qquad (21)$$

Similar conclusion holds in 2D cases, which is shown in Fig. 6 where Fourier transform amplitude of 2D moment base functions is illustrated.

FT of 2D Geometric Moment Base
order (2, 3)

FT of 2D Hermite Moment Base
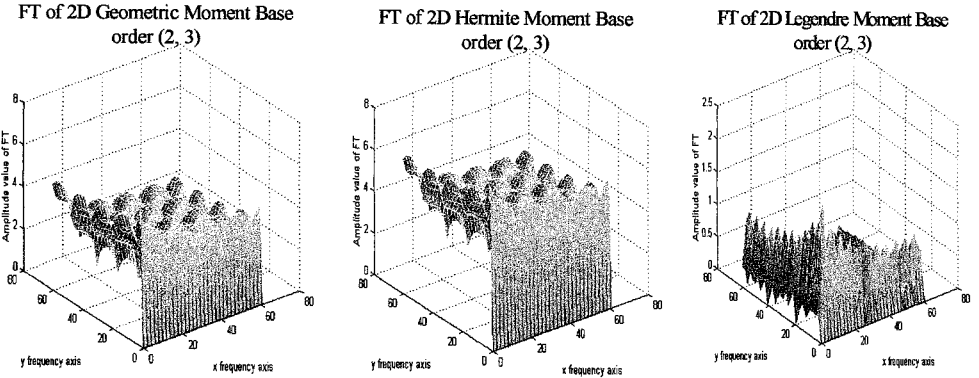order (2, 3)

FT of 2D Legendre Moment Base
order (2, 3)

Fig. 6. Fourier transform amplitude of 2D moment base functions (amplitude visualized in pseudolog value: visual$(x) = \log(1 + |x|)$).

5. GAUSSIAN–HERMITE MOMENTS AND ORTHOGONALITY EQUIVALENCE THEOREM

From the analysis of spatial domain behavior of geometric, Hermite and Legendre moments above, we see that these moment base functions exhibit a great discontinuity at the window boundary. In order to better represent local characteristics of images, particularly for noisy images, one should use orthogonal moments with a smoothing window function. Notice that the use of smoothing window functions is one of the fundamental ideas in signal and image processing, such as in short-time Fourier transform. Taking the well-known Gaussian functions as smoothing kernel, smoothed orthogonal Gaussian–Hermite moments were recently proposed.[25] In this section, we analyze smoothed orthogonal Gaussian–Hermite moments and compare them with geometric, Hermite and Legendre moments.

Given the Gaussian smoothing function $g(x, \sigma)$ with

$$g(x, \sigma) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2) \tag{22}$$

the nth order smoothed Gaussian–Hermite moment $M_n(x, S(x))$ of the signal $S(x)$ is defined as

$$M_n(x, S(x)) = \int_{-\infty}^{\infty} B_n(t)S(x + t)\, dt \qquad n = 0, 1, \ldots \tag{23}$$

with

$$B_n(t) = g(t, \sigma)P_n(t) \tag{24}$$

where $P_n(t)$ is a scaled Hermite polynomial function of order n defined as

$$P_n(t) = H_n(t/\sigma) \tag{25}$$

with

$$H_n(t) = (-1)^n \exp(t^2)(d^n/dt^n) \exp(-t^2)\,. \tag{26}$$

We have shown[25] that the smoothed Gaussian–Hermite moments are orthogonal moments and they can be recursively calculated as follows:

$$M_n(x, S^{(m)}(x)) = 2(n - 1)M_{n-2}(x, S^{(m)}(x))$$
$$+ 2\sigma M_{n-1}(x, S^{(m+1)}(x)) \quad \text{for } m \geq 0 \text{ and } n \geq 2 \tag{27}$$

with

$$M_0(x, S^{(m)}(x)) = g(x, \sigma)^* S^{(m)}(x) \qquad \text{for } m \geq 0 \tag{28}$$

$$M_1(x, S^{(m)}(x)) = 2\sigma d[g(x, \sigma)]/dx^* S^{(m)}(x) \qquad \text{for } m \geq 0 \tag{29}$$

and in particular,

$$M_0(x, S(x)) = g(x, \sigma)^* S(x) \tag{30}$$

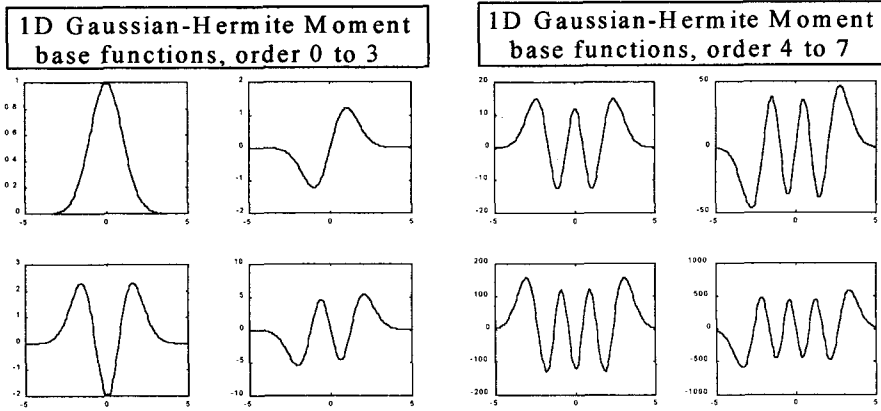$$M_1(x, S(x)) = 2\sigma d[g(x, \sigma)^* S(x)]/dx \tag{31}$$

Fig. 7.   Spatial behavior of 1D Gaussian–Hermite base functions.

where

$$S^{(m)}(x) = d^m S(x)/dx^m \,,$$

$$S^{(0)}(x) = S(x) \,,$$

and * denotes the convolution operator.

Now we analyze the spatial domain behavior of smoothed Gaussian–Hermite moment base functions. Because the nth order Hermite polynomial $H_n(x)$ has n different real roots, the Gaussian–Hermite base function $g(x,\sigma)H_n(x/\sigma)$ will also have n different real roots. Therefore the nth order Gaussian–Hermite base function will change its sign n times. In other words, it consists of n oscillations. So Gaussian–Hermite moments can well characterize different spatial modes as other orthogonal moments. Figure 7 shows the spatial behavior of Gaussian–Hermite base functions of different orders.

As to the frequency domain behavior, since Gaussian–Hermite base functions comprise more and more oscillations when the order n is increased, they will thus contain more and more high frequencies. From the spectral analysis viewpoint, the orthogonal Gaussian–Hermite moments efficiently separate the signal features in different frequency bands. Figure 8 shows the Fourier transform amplitude of some Gaussian–Hermite base functions. On comparing with Figs. 4 and 5, we see that the orthogonal Gaussian–Hermite moments better separate different frequency bands. Table 4 shows the frequency window characteristics of Gaussian–Hermite moment base functions. Comparing Table 4 with Table 3, we see that the frequency window quality factor $Q = $ (Center)/ (Effective bandwidth) of Gaussian–Hermite moment bases is larger than that of other moment bases. Gaussian–Hermite moments therefore separate different bands more efficiently.

Moreover, from the recursive calculation of Gaussian–Hermite moments, we see that these moments are in fact linear combinations of the derivatives of the signal filtered by a Gaussian filter. As is well known, derivatives are important features widely used in signal and image processing. Because differential operations are
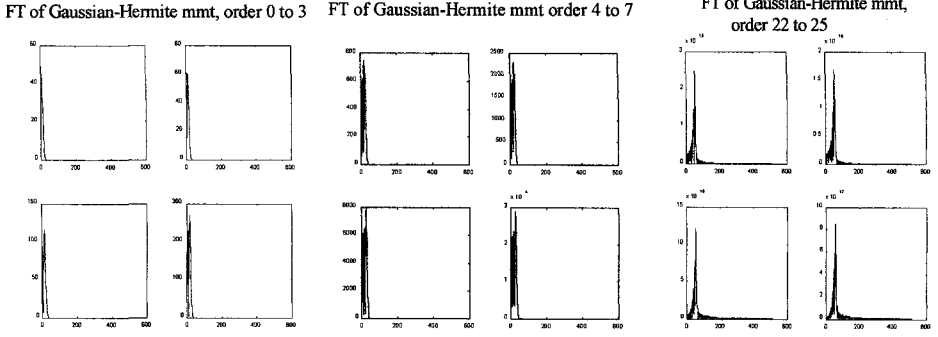
FT of Gaussian-Hermite mmt, order 0 to 3   FT of Gaussian-Hermite mmt order 4 to 7   FT of Gaussian-Hermite mmt, order 22 to 25

Fig. 8.   Fourier transform amplitude of Gaussian–Hermite base functions.

Table 4.   Frequency window characteristics of Gaussian–Hermite moment base functions.

Order	0	1	2	3	4	5	6	7	8	9	10
Center	0.53	1.13	1.36	1.69	1.86	2.12	2.24	2.47	2.59	2.81	2.97
Effective bandwidth	0.44	0.48	0.75	0.80	0.97	1.01	1.15	1.19	1.33	1.41	1.66

sensitive to random noise, a smoothing is in general necessary. The Gaussian–Hermite moments just about meet this demand because of the Gaussian smoothing included. In image processing, one often needs derivatives of different orders to effectively characterize the image,[2,17] but how to combine them is still a difficult problem. Gaussian–Hermite moments show a way to construct orthogonal features from different derivatives.

The derivatives of Gaussian functions are widely used as mother wavelets for signal and image analysis,[4] and different order derivatives of Gaussian filters define different wavelets, so the derivatives of an input signal filtered by Gaussian filters of different standard deviations represent the decomposition of the signal into wavelets. Therefore the smoothed orthogonal Gaussian–Hermite moments also offer a solution to construct orthogonal features from the wavelet analysis results.

2D orthogonal Gaussian–Hermite moments of order (p, q) of an input image $I(x, y)$ can be defined similarly

$$M_{p,q}(x, y, I(x, y)) = \iint_{-\infty}^{\infty} G(t, v, \sigma) H_{p,q}(t/\sigma, v/\sigma) S(x + t, y + v) dt\, dv \qquad (32)$$

where $G(t, v, \sigma)$ is the 2D Gaussian function, and $H_{p,q}(t/\sigma, v/\sigma)$, the scaled 2D Hermite polynomial of order (p, q), with

$$H_{p,q}(t/\sigma, v/\sigma) = H_p(t/\sigma) H_q(v/\sigma). \qquad (33)$$

Obviously, 2D orthogonal Gaussian–Hermite moments are separable, so the recursive algorithm in 1D cases can be applied for their calculation. Figure 9 shows two Gaussian–Hermite moment base functions and the Fourier transform amplitude of a 2D Gaussian–Hermite base function.
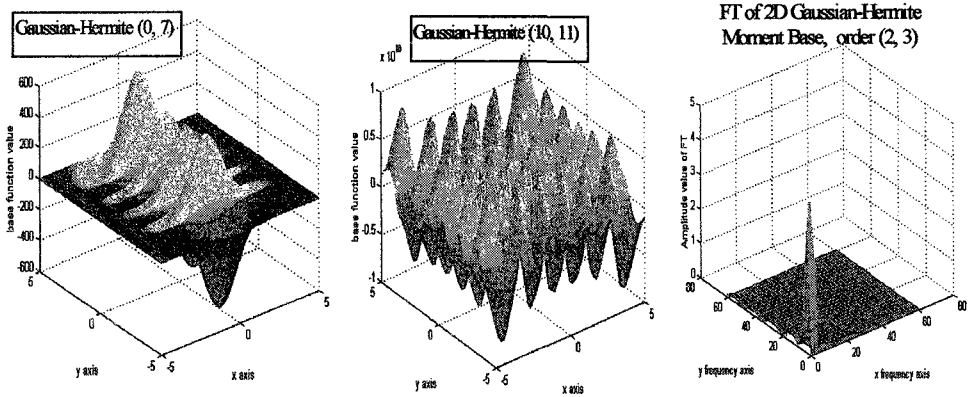
Fig. 9.   2D Gaussian–Hermite moment base functions and Fourier transform amplitude.

To compare the performance of 2D Gaussian–Hermite moments with others, we have to test the sensibility of the moments to noise. By using noise-free images and noisy ones with additional random noise, we calculated the moment vectors $(m_{0,0}, m_{0,1}, \ldots, m_{0,5}, m_{1,0}, m_{1,1}, \ldots, m_{1,5})$, and the normalized distances between noise-free images and noisy ones are shown in Table 5.

We see from Table 5 that the Gaussian–Hermite moments are less sensitive to noise than the other moments.

As is shown above, the orthogonality of moment base functions is defined in the space domain, so one can easily understand their good performance in the spatial domain, for example, they can well separate different image modes. But why do they have a good performance in the frequency domain as well? To better understand the good performance of orthogonal moments in both spatial and frequency domains, we propose the following orthogonality equivalence theorem.

Orthogonality equivalence theorem. Orthogonal moment base functions are not only orthogonal in spatial domain but also in frequency domain.
Let $f_0(x)$, $f_1(x)$, $\ldots$, $f_n(x)$, $\ldots$ be a family of orthogonal polynomial functions defined on the interval $[a, b]$ and taking $w(x)$ as the weight function, with $w(x) \geq 0$, $\forall x \in [a, b]$, the interval $[a, b]$ can be $[a, +\infty)$, $(-\infty, b]$ or $(-\infty, +\infty)$. According to the definition of orthogonal polynomial functions, we have

$$\int_a^b w(x) f_m(x) f_n(x) \, dx = 0 \qquad \text{for } m, n = 0, 1, \ldots \text{ and } m \neq n \qquad (34)$$

$$\int_a^b w(x) f_m(x) f_n(x) \, dx$$

$$= \int_a^b w(x) f_m{}^2(x) \, dx > 0 \qquad \text{for } m, n = 0, 1, \ldots \text{ and } m = n. \qquad (35)$$

Because $w(x) \geq 0$, $\forall x \in [a, b,]$, we can define

$$\rho(x) = \sqrt{w(x)}, \qquad x \in [a, b]. \qquad (36)$$

Table 5.   Normalized distance between noise-free and noisy images when different moments are used.

Moment order	(0,0)	(0,1)	(0,2)	(0,3)	(0,4)	(0,5)	(1,0)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	Total in 12D moment space
Geometric	0.0663	0.0825	0.0414	0.0764	0.0319	0.0693	0.0846	0.0072	0.0638	0.0070	0.0504	0.0064	0.1961
Hermite	0.0663	0.0825	0.0397	0.0751	0.0300	0.0666	0.0846	0.0072	0.0620	0.0070	0.0476	0.0062	0.1928
Legendre	0.0663	0.0825	0.0128	0.0008	0.0065	0.0064	0.0846	0.0072	0.0012	0.0009	0.0043	0.0009	0.1367
Gaussian–Hermite	0.0215	0.0164	0.0134	0.0122	0.0055	0.0196	0.0201	0.0001	0.0141	0.0002	0.0128	0.0004	0.0473

Orthogonal moment base functions using the orthogonal function family is

$$B_n(x) = \rho(x)f_n(x), \qquad n = 0, 1, 2, \ldots \tag{37}$$

where n is the order of the function.

Now we show that the orthogonal function family is orthogonal in frequency domain as well. Let $F_n(\omega)$ be the Fourier transform of $B_n(x)$, with

$$F_n(\omega) = \int_{-\infty}^{\infty} B_n(x) \exp(-j\omega) \, dx$$

$$= \int_{a}^{b} \rho(x)f_n(x) \exp(-j\omega) \, dx. \tag{38}$$

The inner product between $F_m(\omega)$ and $F_n(\omega)$ in the frequency domain will be

$$\langle F_m(\omega), F_n(\omega) \rangle$$

$$= \int_{-\infty}^{\infty} F_m(\omega)\overline{F_n(\omega)} \, dw$$

$$= \int_{-\infty}^{\infty} F_m(\omega) \int_{a}^{b} \rho(x)f_n(x) \exp(j\omega) \, dx \, d\omega$$

$$= \int_{a}^{b} \rho(x)f_n(x) \left[\int_{-\infty}^{\infty} F_m(\omega) \exp(j\omega) \, dw \right] \, dx$$

$$= \int_{a}^{b} \rho(x)f_n(x)\rho(x)f_m(x) \, dx$$

$$= \int_{a}^{b} w(x)f_n(x)f_m(x) \, dx. \tag{39}$$

From the definition of the orthogonal polynomial functions above, we obtain

$$\langle F_m(\omega), F_n(\omega) \rangle \begin{cases} = 0 & \text{for} \quad m, n = 0, 1, 2 = \ldots, \quad \text{and} \quad m = n \\ > 0 & \text{for} \quad m, n = 0, 1, 2, \ldots \quad \text{and} \quad m \neq n \end{cases} \tag{40}$$

which means that $F_m(\omega)$ and $F_n(\omega)$ are orthogonal in the interval $(-\infty, +\infty)$ in frequency domain.

In particular, taking $w(x) = 1$ and the interval $[a, b] = [-T, +T]$, where $2T$ is the window size, we obtain Legendre moment base functions which are orthogonal in both spatial and frequency domains. Taking $w(x) = \exp(x^2/\sigma^2)$ and the interval $(-\infty, +\infty)$, we obtain smooth Gaussian–Hermite moment base functions which are also orthogonal in both spatial and frequency domains.

6. CONCLUDING REMARKS

In this paper, we analyzed and compared the behavior of different moments such as geometric, Legendre, Hermite and Gaussian–Hermite moments.

After an introduction to geometric, Legendre and Hermite moments, we first analyzed their behavior in spatial domain. Our analysis showed: (1) All these moments allow the reconstruction of the image from moments, so they can all be

used for image analysis. (2) The general variation trend of the values of geometric moment base functions of different orders is the same, so these moments do not very well characterize different spatial modes of image functions. (3) Legendre and Hermite moment base functions of different orders have different number of zero-crossings and very different shapes, therefore they could better separate image features based on different modes, which is very interesting for pattern analysis and shape classification. (4) Legendre moment base functions are orthogonal to each other, but Hermite moment base functions for a finite window size are not. So by using Legendre moments, the calculation could be reduced, the error is easier to estimate when moments of a limited order are used, and the image reconstruction from these moments could also be simpler. (5) All geometric, Legendre and Hermite moment base functions exhibit a great discontinuity at the window boundary, so to some extent, they would be sensitive to noise. These results are confirmed by experimental results for shape characterization.

To characterize noisy images, Gaussian–Hermite moments are then introduced. We showed (1) The Gaussian–Hermite base functions are also orthogonal, so they preserve all the advantages of orthogonal moments such as Legendre ones. Moreover they are much more smoothed than other moment base functions and do not exhibit a discontinuity even at the window boundary. Gaussian–Hermite moments are therefore much less sensitive to noise and avoid the artifacts introduced by window function discontinuity. (2) The orthogonality of Legendre and Gaussian–Hermite moment base functions ensures the development of the image in an orthogonal function space and therefore facilitates afterwards the image comparison and classification.

The spectral behavior of these moments in frequency domain was also analyzed. Our analyses showed that Legendre and Gaussian–Hermite moments of different orders separate different frequency bands more effectively than other moments. In particular, the quality factor (Q) of frequency window of Gaussian–Hermite moment base functions is better than other moments. This conclusion is confirmed by experimental results on noisy images. We also show that Gaussian–Hermite moments present an approach to construct orthogonal features from the results of wavelet analysis, which is very important for image analysis and classification. To better understand the good performance of orthogonal moments, the orthogonality equivalence theorem was also presented.

Since the moments are widely used in image analysis, multiresolution analysis and pattern recognition, we hope that the analysis presented in this paper will give a global view of different moments and can help moment-based image analysis.

REFERENCES

1. Y. S. Abu-Mostafa and D. Psatlis, "Recognitive aspect of moments," *IEEE Trans.* **PAMI-6** (1984) 698–706.
2. S. Castan and J. Shen, "Box filtering for Gaussian-type filters by use of the B-Spline functions," *Proc. 4th Scandinavian Conf. Image Analysis*, Norway, 1985 pp. 235–243.

3. L. H. Chen and W. H. Tsai, "Moment preservation curve detection," *IEEE Trans.* **SMC-18** (1988) 148–158.

4. Ch. K. Chui (ed.), *Wavelets: A Tutorial in Theory and Applications*, Academic Press Inc., Boston, 1992.

5. P. Davis, *Methods of Numerical Integration*, 1975.

6. C. Fröberg, *Introduction to Numerical Analysis*, 1965.

7. R. M. Haralick, "The facet approach to optic flow," *Proc. Image Understanding Workshop*, Arlington, 1983, pp. 84–93.

8. M. Hatamian, "A real-time two-dimensional moment generating algorithm and its single chip implementation," *IEEE Trans.* **ASSP-34**, 3 (1986) 546–553.

9. M. K. Hu, "Visual pattern recognition," *IRE Trans. Inform. Th.* **IT-8** (1962) 179–187.

10. B. C. Li and J. Shen, "Fast computation of moment invariants," *Patt. Recogn.* **24**, 8 (1991) 807–813.

11. B. C. Li and J. Shen, "Pascal triangle transform approach to the calculation of 3D moments," *CVGIP: Graph. Mod. Imag. Process.* **54**, 4 (1992) 301–307.

12. B. C. Li and J. Shen, "Range-image-based calculation of three-dimensional convex object moments," *IEEE Trans. Robot. Automat.* **8**, 4 (1993) 484–490.

13. B. C. Li and J. Shen, "Two-dimensional local moment, surface fitting and their fast computation," *Patt. Recogn.* **27**, 6 (1994) 785–790.

14. S. T. Lio and W. H. Tsai, "Moment preservation corner detection," *Patt. Recogn.* **23**, 5 (1990) 441–446.

15. C. H. Lo and H. S. Don, "3D moment forms: their construction and application to object identification and positioning," *IEEE Trans.* **PAMI-11**, 10 (1989) 1053–1064.

16. E. P. Lyvers, "Subpixel measurement using a moment-based edge operator," *IEEE Trans.* **PAMI-11**, 12 (1989) 1293–1308.

17. M. J. McDonnell, "Box-filtering techniques," *CGIP* **17**, 1 (1981) 65–70.

18. H. H. Nagel, "On the estimation of optical flow: relations between different approaches and some new results," *Artif. Intell.* **33** (1984) 299–324.

19. M. Pawlak and S. X. Liao, "On image analysis by moments," *IEEE Trans.* **PAMI-18**, 3 (1996) 254–266.

20. M. Pawlak and X. Liao, "On image analysis by orthogonal moments," *Proc. 11th ICPR*, The Hague, August 30–September 3, 1992, pp. 549–552.

21. F. A. Sadjadi, "3D moment invariants," *IEEE Trans.* **PAMI-2**, 2 (1960) 127–136.

22. J. Shen and D. F. Shen, "Image characterization by fast calculation of low-order Legendre moments," *Proc. IEEE Int. Conf. SMC'96*, Beijing, October 1996, pp. 1144–1149.

23. J. Shen and D. F. Shen, "Image characterization by fast calculation of Legendre moments," *Proc. European Symp. Satellite Remote Sensing'96*, Italy, September 1996, pp. 295–306.

24. J. Shen and D. F. Shen, "Orthogonal Legendre moments and their calculation," *Proc. ICPR'96*, Vienna, August 1996, pp. B241–B245.

25. J. Shen, "Orthogonal Gaussian–Hermite moments for image characterization," *Proc. SPIE, Intelligent Robots and Computer Vision XVI: Algorithms, Techniques, Active Vision, and Materials Handling*, Pittsburgh, October 15–17, 1997, pp. 224–233.

26. G. Szegö, *Orthogonal Polynomials*, N.Y., 1959.

27. C. H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans.* **PAMI-10**, July (1988) 496–512.

28. N. Yokaya, "Range image segmentation based on differential geometry: a hybrid approach," *IEEE Trans.* **PAMI-11**, 6 (1989) 643–649.

29. M. F. Zakaria, L. J. Vroomen, P. J. A. Zsombar-Murray and J. M. H. M. Van Kessel,
     "Fast algorithm for the computation of moment invariants," *Patt. Recogn.* **20**, 6 (1987)
     634–643.
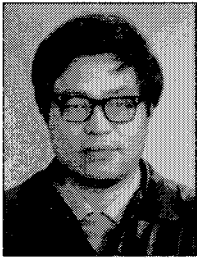
Jun Shen is a professor in computer science and head of Image Laboratory at EGID Institute, Bordeaux-3 University, France. He received the B.S. degree in radio and electronic engineering from Tsinghua University, Beijing, China in 1968, and the Ph.D. and "Doctorat d'Etat" degrees in computer science from Paul Sabatier University, Toulouse, France in 1982 and 1986, respectively. He is author or co-author of more than 100 publications in image processing and computer vision. He received an "Outstanding Paper Honorable Mention" from IEEE Computer Society in 1986.

His current research interests include image processing and analysis, computer vision, pattern recognition and neural network application.

Danfei Shen is with the National Higher Institute of Telecommunications (Ecole Nationale Supérieure des Télécommunications), Paris, France. He received the B.S. and M.S. degrees in applied mathematics from the University of Bordeaux-1, France, in 1996 and 1997, respectively.

His current research interests include signal and image processing and computer vision.

Wei Shen is an associate professor at Zhongshan University, Canton, China. He received the B.S. degree in mechanics from the Polytechnic Institute of Jiangsu, China in 1982, and the M.S. degree in robotics and Ph.D. from University of Poitiers, France, in 1989 and 1996, respectively.

His research interests include image processing, computer vision, pattern recognition and neural network application.

MULTISPECTRAL IMAGE PROCESSING: THE NATURE FACTOR

WENDELL R. WATKINS

U.S. Army Research Laboratory
ATTN: AMSRL-SL-EA
White Sands Missile Range, NM, 88002–5513, USA

The images processed by our brain represent our window into the world. For some animals this window is derived from a single eye, for others, including humans, two eyes provide stereo imagery, for others like the black widow spider several eyes are used (8 eyes), and some insects like the common housefly utilize thousands of eyes (ommatidia). Still other animals like the bat and dolphin have eyes for regular vision, but employ acoustic sonar vision for seeing when their regular eyes do not work such as in pitch black caves or turbid water. Of course, other animals have adapted to dark environments by bringing along their own lighting such as the firefly and several creatures from the depths of the ocean floor. Animal vision is truly varied and has developed over millennia in many remarkable ways. We have learned a lot about vision processes by studying these animal systems and can still learn even more.

Keywords: Image processing; human vision; depth perception; wide baseline stereo; target acquisition; navigation; optical turbulence.

1. INTRODUCTION

There have been many advances in the twentieth century, but few can rival those in the area of image processing. The miniaturization and increased speed and memory of computers have opened the door for exploitation of a wide variety of applications in the area of multispectral image processing. We now have imaging systems that reconstruct 3D images of human body organs beneath our skin through computer tomography (CT) scans, 3D images of aircraft approaching an airport can be displayed inside a crystal for air traffic control, photographs are digitally stored and enhanced before printing, and robots can navigate and perform complex object manipulation using processed stereo imagery. Yet, in some ways we have not taken full advantage of the evolution of the imaging systems developed by animals including humans.

What is vision anyway? The brain receives signals from our optic nerves that represent the images formed on the retinas of our eyes. Our brain with only about a 125-terabyte storage capacity is able to translate these image signals into human vision with incredible speed. We perform object recognition using color, shape and texture. We perform range mapping to determine the relative position of objects. We even fill in image data voids based on previous image processing experience. We have begun to replicate some of the human vision processing through the development of mathematical algorithms and techniques such as neural networks, wavelets, fuzzy logic, etc. Yet, our understanding has a way to go. We have begun to use the information produced simultaneously by several images for robotic depth perception

purposes, but it is difficult to imagine what the vision of the common housefly must be with its 4000 image forming elements. More importantly, however, we have become comfortable with less than optimum methods of image display.

During the course of a day, we are presented a tremendous amount of visual information to process. We read newspapers, magazines and books. We read our email messages off our computer monitors. We even digest visual information from television displays. We have become accustomed to processing a large majority of our visual information as a 2D representation that can be easily replicated in digital form. It is much easier to process information and produce algorithms for a 2D array. It is to their credit that the great sculptors of the past were able to break out of the 2D representation mold and produce 3D works of art. It is this understanding of depth perception and the use of our vision capability that has not been fully exploited in many areas of multispectral image processing. We will look at some examples in this area and other areas in this paper.

2. MULTISPECTRAL IMAGE PROCESSING

Just as an oasis is a delightfully refreshing change of pace for the desert traveler, multispectral image processing provides the scientist and engineer an opportunity to apply a diversity of disciplines to a myriad of applications in a world where specialization and repetition have become the norm. Multispectral image processing encompasses the disciplines of mathematics, computer sciences, electrical and mechanical engineering, physics, photogrammetry and biological sciences. Bernd Jähne gives a nice summary of multispectral image processing tasks in his recent book entitled *Practical Handbook on Image Processing for Scientific Applications*.[12] He includes geometric measurements of gauging and counting, radiometric measurements of revealing the invisible, depth measurements of exploring the 3D space, and velocity measurements of exploring dynamic processes. Although we have come a long way in terms of being able to analyze imagery using sophisticated physical and mathematical algorithms that drive our computer vision measurement analysis, we still cannot match human vision that is able to easily solve very difficult recognition and classification tasks.[18]

The complexity of the tasks for which we now use multispectral image processing truly challenges our scientific and engineering ingenuity. To optimize the results of our image processing we must consider the scene or object illumination and optical properties for proper contrast and texture. The radiation field, which constitutes the image, may be altered by the atmosphere between the scene or object and the camera system collecting the image. The spectral, spatial and temporal characteristics of the camera itself may not accurately collect all of the important details contained in the scene. Our image analysis tools may also introduce artifacts or distort the characteristics of the scene. Finally, the means used to display the results may not allow important aspects of the scene to be accurately portrayed.

We can enhance the aircraft pilot's ability to land in adverse weather using infrared and millimeter imagery overlaid on the pilot's visible vision using heads-up displays.[22] We can even see through intense explosive plumes using laser-gated

imagery;[26] but, we still cannot safely drive automobiles under foggy conditions. We can diagnose diseases using sophisticated techniques such as CT scans. Yet, orthodontists routinely have to compare dental X-rays taken over a period of months to determine tooth positioning. Our visible cameras can now match the sensitivity of the human eye,[23] and we will soon have the capability of true 3D display of aircraft positions for air traffic controllers.[24] Still, we obtain most of our daily news and information from flat displays, pictures and written text. We have become too comfortable with a 2D representation of our 3D world.

3. IMAGE PROCESSING TOOLS

Machine or computer processing requires that the images be input in digital form. Conversion from a radiance field, which is sensed by the imaging system used, to the required digital format introduces several problems such as sampling, aliasing and data fidelity that must be resolved.[11] Unless atmospheric degradations like aerosol scattering and optical turbulence distortion are eliminated using techniques like guide star deformable mirror corrections during image collection,[21] the image content must be restored using processing techniques such as boundary contour systems.[1] Before an image processing task such as target detection can be performed, the imagery data must be enhanced and restored for optimum results. This process typically entails filtering and/or averaging of some kind and can include spatial, temporal and spectral image characteristics. Although there are many references on image processing techniques, Anil K. Jain's book entitled *Fundamentals of Digital Image Processing* continues to be the definitive text in this area.[13] Once the images have been prepared for a particular image processing task, the real challenging work begins. Features must be extracted from the imagery in the form of regions, edges and lines, depth and motion, and scale and texture. These features are then used to create the masks that represent the objects of interest through segmentation, size and shape, and finally classification.

It should again be stressed that the machine or computer vision process really only measures quantities related to the content of imagery databases. Often what is desired in an image processing task requires analysis that verges on the qualitative recognition aspect that is performed by animal vision or, in essence, we want intelligent robotic vision. It is in this regime that the Nature Factor is important. Human vision, for example, uses a sophisticated hierarchy of organized massively parallel processing. Machine or computer processing is still in its infancy in terms of performing the qualitative image processing tasks. Animals rely for their survival on how well they blend into the terrain, yet we have only begun to be able to model the camouflage process in cluttered environments for image processing tasks.[10,30] Part of the problem relates to how we perceive objects based on texture and color variations and how these processes can be applied to imagery in spectral bands outside of the visible.[2,4,25] Finally, neural nets and fuzzy logic are beginning to be used as the image processing speed available in personal computers has attained the hundreds of MHz central processing unit clock speed.[11]

How is the Nature Factor used? To begin, the sensors we use to collect the imagery to be analyzed as well as the rectangular or square shape of the digitally stored image pixels does not correspond to the hexagonal pattern of close-packed cones used for human color vision. Nature's hexagonal close-pack translates into a pattern where each individual color cone, for example red, is surrounded by three green and three blue cones alternately spaced at 60° intervals. This is not a random distribution but optimum for minimizing the Moiré-effect. That is why color printers with dot matrices have their three color axes rotated by 60° with respect to each other. The spectral content of the imagery can also be an issue. The bow and arrow deer hunter with his human-designed camouflage can be in for a surprise. Deer are nocturnal and, hence, have enhanced scotopic vision. In fact, they see ultraviolet (uv) fluorescence that humans do not. Not only will they see the clothing washed with phosphate detergent but even the bowstring will glow unless it is coated with uv-blocking wax. Many manmade objects have vertical and horizontal edges that differ from Nature's more curved shapes so why not use a bar-pattern to measure the human vision's modulation transfer function. It was not until very recently that anything else was even tried for characterizing the vision process.[3,27] Finally, Nature has given us many excellent examples of vision processes. Most animals have two eyes. The hunters have two forward-looking eyes with excellent visual acuity for discerning primarily motion. The preys, on the other hand, have two opposite looking eyes to give hemispherical coverage of approaching danger. Some birds that have opposite looking eyes rock forward and backward to obtain depth perception information about the location of small crustaceans hidden under the moving sands of the surf on the oceans' shores. Yet, for the most part, we have discarded the value of stereoscopic depth perception for use in addressing many common image processing tasks.[34]

4. THE NATURE FACTOR

The basis for exploiting the Nature Factor began with the development of a field research tool called the Mobile Imaging Spectroscopy Laboratory (MISL). The basic concept behind its use was the matching of multiple spectral band imaging systems on a target plane from a close up and far away location. Typically the ratio of the far to close ranges was 10 to 17. By comparing the imagery that had the fields of view in the target plane optically matched with appropriate lenses, the effects of the propagation of the scene radiation could be characterized.[32] The MISL was used to characterize atmospheric effects on target contrast. Changes in solar loading due to the passage of clouds that blocked the sun during the day were especially pronounced in the thermal infrared spectral region.[5] Contrast transmission and optical turbulence distortion of the propagated imagery could be characterized as well.[14,33] Because the contrast transmission is dependent upon the product of spectral radiance times atmospheric transmission times the detector response, this led to bandpass optimization for the use of thermal imaging for long path applications.[17] These effects were summarized in a paper presented at the 1993 Society of Photo-Optical Instrumentation Engineers Symposium in Orlando, FL.[31]

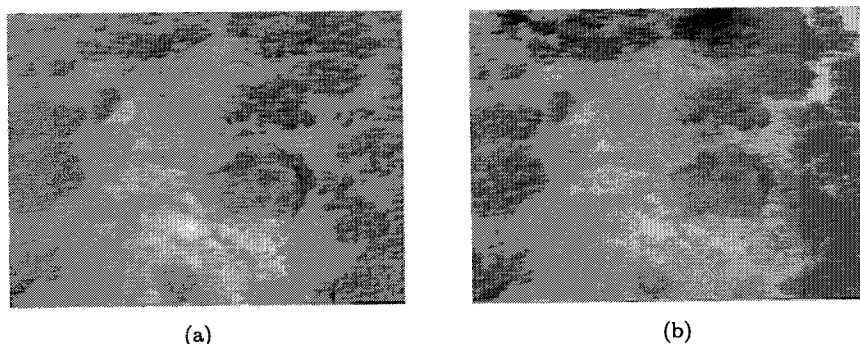<div align="center">(a)   (b)</div>

Fig. 1.   (a) Right view (b) Left view of camo nets, 3 m separation.

One of the most significant outgrowths of the concept of matching the near and far views was the visualization of scene dynamics. Some of the effects like solar loading changes could be seen from both views, while others like optical turbulence distortions were only seen in the distant view. Almost by accident the benefits of using wide baseline stereo imaging was conceived when the near and far imaging system live monitor imagery was fused with cross-eyed stereo viewing. The optical distortions that were occurring on the far imagery seemed to disappear when fused with the undistorted near imagery. Several experiments were performed to illustrate the benefits of real-time viewing with wide baseline stereo. The first example was that of the application of the depth perception aspect of stereo vision applied to the detection of camouflage netting shown in Fig. 1

Figure 1 shows what can be done with hyperstereo (wide baseline, compared to the human interocular distance of 6-to-7 cm, stereo) vision when applied to camouflage netting viewed from a 45° look-down angle at a range of about 55 m.[34] The figure gives a cross-eyed stereo view with 3 m baseline separation. When viewed in stereo two camouflage nets (one obvious forest net and one very difficult-to-see desert net) leap out of the scene. Care must be taken in the display of the stereo images to be effective. In this example, the camera separation was 3 m with the camera fields of view of 20° by 15° (width/height). This corresponds to images with 19 m lateral extent and ranges between the bottom and top of the scene of 46 m to 69 m or a vertical extent of 27.5 m along the ground. With a typical image display width of 11.5 cm and a 2 cm separation between them, the stereo images can easily be fused at a viewing distance of 30 cm to 100 cm. This represents an angular viewing extent between 7° to 20°. A problem can occur when too much magnification is used which corresponds to increasing the viewing extent too far beyond the actual scene field of view. When a 6 m camera baseline was used for horizontal viewing, there was a problem with too much parallax to fuse magnified fields of view of about 2° in width for ranges less than about 300 m. The problem is similar to trying to see all of your fingers when you spread them out at different ranges right in front of your face. You can only view them one at a time. An additional benefit of wide baseline stereo viewing that is not experienced when just

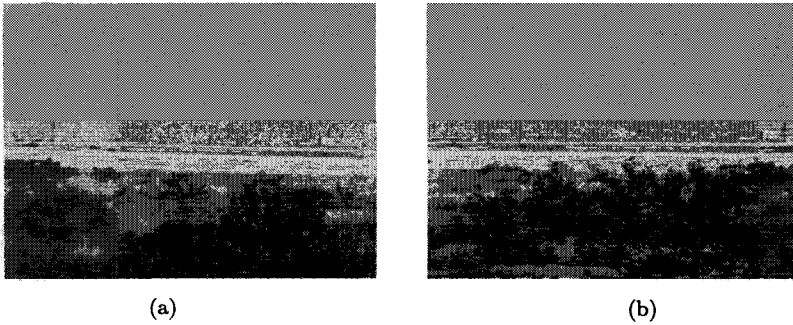(a)   (b)

Fig. 2.   (a) Right view (b) Left view of optical turbulence distortion.

looking at a stereo pair of still images is dynamic cueing. When the wind is blowing the motion of the individual branches of the bushes can be readily distinguished from the back and forth undulations of the flat surface camouflage net even at much longer ranges.

The second example was that of the application of the stereo fusion process for mitigating optical turbulence distortions. Figure 2 shows a cross-eyed stereo view with a 10 m baseline separation of a scene with ranges of one-to-several kilometers to distinct features that can be stereo vision fused. In this stereo pair the line of telephone poles about half way up in the scene is at a range between one and two kilometers away. The lower portion contains features at ranges of much less than a kilometer. With the 10 m platform separation and a line of sight centered at about 2 kilometers, stereo fusion of most of these features is not possible because the same features are not even in both fields of view. The features in the upper half of the stereo pair can be fused and give good depth perception up to multiple kilometers. These images were taken from VHS videotapes that show appreciable optical turbulence distortion along both stereo lines of sight. This can be seen as different distortions of the edges of the individual high contrast features between the two scenes shown in Fig. 2. When viewed as single line of sight video the individual objects would dance around and change shapes. Because the platform separation was wider than the typical 1-to-2 m coherence length for the optical turbulence present along each path, the distortions of the features out to several kilometers were different for each line of sight; and, hence, the distortions were decoupled from each other. When the video from each line of sight was viewed in stereo, the stereo fusion process filtered out most of the dancing around and changes in shape as noise. The features became more distinct in shape with only a little residual high frequency jitter of the edges. It is in the area of dealing with scene dynamics that the Nature Factor can have significant impact.

5. DEPTH VERSUS MOTION CUES

To understand how the Nature Factor can be used, consider for the moment what it must be like to see using the common housefly's vision system with its compound eyes of thousands of individual imaging elements or ommatidia. By simultaneously

analyzing the imagery from patches of ommatidia with a brain the size of the head of a pin, you would be able to discern from motion or depth cues dangers that are present in any direction. Motion cues are derived from the change in position of an object with respect to its background. The range to an object can be derived by determining the angular offset of the object with respect to its background along two or more different lines of sight and, then, solving the feature correspondence between the different image pairs. In essence, motion and depth cues are treated the same mathematically. Currently, motion cues are derived from temporal sets of images but depth information is typically not used. Consider what the world would look like if your eyes were placed multiple meters apart and equipped with magnifying optics. When such a test was conducted the results were that the same human vision analysis processes that were used to obtain depth information around us upto tens of meters provided depth information at multiple kilometer ranges (see Fig. 3).[34] In order to illustrate how the depth mapping can be performed, views of this scene were taken along three different lines of sight and digitized to form 512 by 512 pixel arrays. There was a central line of sight, a left line of sight with a baseline separation of 19.6 m, and a right line of sight with a baseline separation of 27.2 m. The three scenes were aligned at a central point and a 256 by 256 pixel area was extracted. Four distinct stationary features were picked from the central scene and the pixel disparity between left and center and center and right were calculated. The disparity was measured by applying pixel template area matching. Disparities were calculated from the sum of squared differences for each distinct feature in the two pairs of stereo images.[28] This approach uses a small area of interest from one stereo pair image that is matched to the second stereo pair image within a larger region whose extent is based on expected disparity limits to obtain the best template match. The ratio of the camera baseline separations is 1.39 (left-center/center-right). The results are shown in Fig. 4.[29] An entire scene depth map could not be derived because the images were taken at different times and scene feature changes due to lighting and other effects such as motion due to wind in general caused large errors in texture correspondence especially for the forest canopy. For whole scene depth mapping all the different lines of sight imagery must be taken at the same time.

As mentioned above, the use of wide baseline stereo vision has other benefits. The optical turbulence distortions were automatically mitigated through the human vision processing of the imagery. Evidently what happens is the turbulence distortions being seen by each eye become decoupled by the angular separation; and, hence, the human vision processing averaged out the distortions as noise. This occurs because the human stereo fusion process is based on relatively coarse spatial content within images as evidenced by image compression research.[9,15] In addition, when wide baseline stereo was applied to an underwater robotic task of picking up objects in turbid water, the depth perception obtained allowed the object of interest to be separated from the floating murky debris that constituted severe clutter.[20] Yet, the algorithms currently used for target acquisition are based on statistics from digitized single line of sight images with no depth perception cues. Using the Nature

Right View   Left View
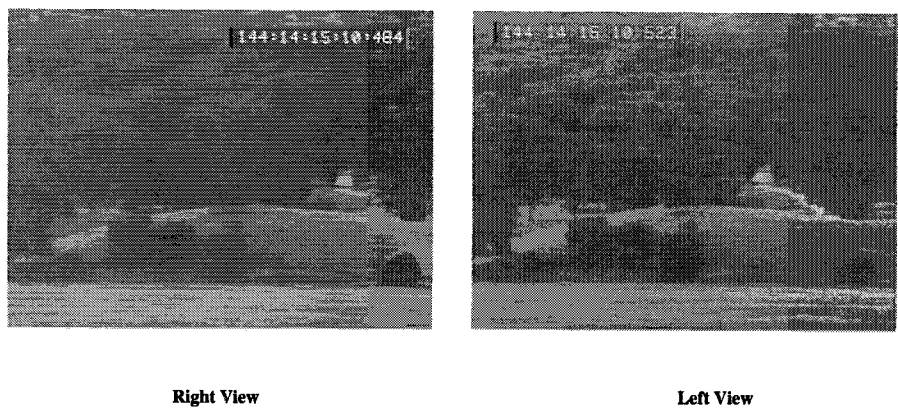
Fig. 3. Cross-eyed stereo pair of rural terrain at 2 km range collected with two visible cameras with 10X lens and a 50 m baseline separation. Additional details of the imagery and method of viewing are found in Ref. 34.
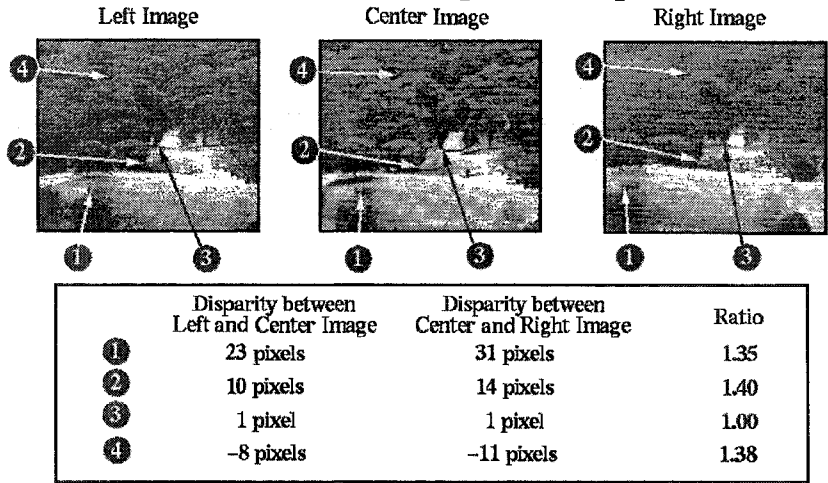
Stereo Algorithm: Depth Computation

Left Image   Center Image   Right Image

	Disparity between Left and Center Image	Disparity between Center and Right Image	Ratio
❶	23 pixels	31 pixels	1.35
❷	10 pixels	14 pixels	1.40
❸	1 pixel	1 pixel	1.00
❹	–8 pixels	–11 pixels	1.38

Fig. 4. Set of three lines of sight for the same terrain as in Fig. 3 aligned on the center pixel. The disparities between four stationary locations in the scenes are shown for the left-center and center-right stereo pairs. The ratio between disparities (left-center/center-right) is calculated at each location for comparison with the camera platform separation ratio of 1.39.

Factor of replicating the stereo vision depth perception that animal vision utilizes to detect targets, there are several cues that can be applied such as depressed trails in grassy terrain that lead to the vehicle producing them. In addition, much of the clutter that is detected in single line of sight imagery disappears because volume and surface contour can be used as discriminating factors. Finally, the ground and tree lines are better defined showing where real targets cannot be and where hidden targets may be.

6. NAVIGATION

A final example of how the Nature Factor can be used is in the area of navigation. Before delving into how wide baseline stereo vision can be used to enhance navigation, a digression is in order to explain why it is not already being used. It is not as if stereo vision or even wide baseline stereo vision was just discovered. The first documented use of stereo vision was by Euclid; and, subsequently, it was represented in drawings by Giovanni Battista della Porta around 1600. More recently the Brewster Stereoscope was in widespread use from the mid-1800s to mid-1900s. It was not until very recently that technology has provided the tools to allow easy viewing of high spatial resolution stereo video imagery. The other missing element in making wide baseline stereo vision practical was the computational speed now available to enhance the video inputs to the stereo imagery display. The cost of obtaining the necessary processing and display equipment is now only a few thousand instead of tens of thousands of dollars.

We take advantage of our depth perception extensively in navigating over rough terrain. During the day we have little problem avoiding obstacles when the visibility is good. But at night avoiding obstacles becomes a real problem. Recent tests at the U.S. Army Research Laboratory by V. Grayson CuQlock-Knopp et al. have compared monocular, biocular and binocular night vision goggles.[6,7] Their findings were that the depth perception of the binocular goggles provided the best navigation as well as target detection capabilities. Hyperstereo vision has also been used in robotics applications including navigation.[8,16,19] To illustrate how wide baseline stereo vision could be adapted to enhanced night time vehicular navigation a set of experiments were performed using a pair of Inframetrics, Inc. Model 610 infrared imagers with 2.5 m baseline separation. In order to illustrate how a driver could use a stereo display headset for navigation, a passenger in the vehicle wore a pair of stereoscopic display goggles designed for viewing asynchronous video signals on a see through LCD display shown in Fig. 5. The cost of the unit was less than one thousand dollars. The type of stereo imagery display that was viewed is shown in Fig. 6 with additional details given elsewhere.[34] The important finding of the experiment was that the stereoscopic vision indeed could be used to navigate the vehicle even in rough terrain. For military applications of driving without headlights, navigation in this case would have been impossible. As in the case of applying wide baseline stereo vision to search and target acquisition tasks, care must be taken in presenting the correct stereoscopic field of view for the particular task. In this case some magnification can be used without creating problems with the distance perception needed for driving a vehicle. The cameras used were set for a 75% field of view magnification to lessen the problem of fusing the entire field of view of the individual cameras. Additionally, the camera lines of sight need to be adjusted depending upon whether the terrain is on-road or off-road. The on-road terrain requires the camera lines of sight to cross at about 150 m because the driving speeds are typically higher with less terrain hazards than the off-road case where the lines of sight should cross at a range of about 30 m. In this way the stereoscopic depth of field is optimized for detecting the most probable location of navigation
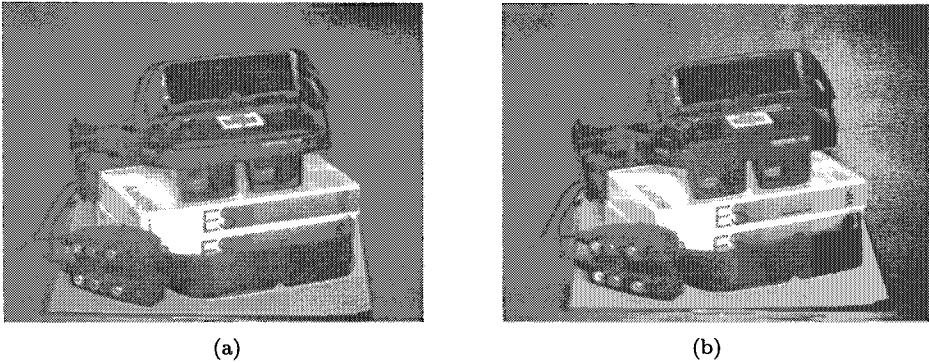
(a)   (b)

Fig. 5.   (a) Right view (b) Left view of stereo pairs showing a modified pair of stereo goggles for
asynchronous video signal viewing.

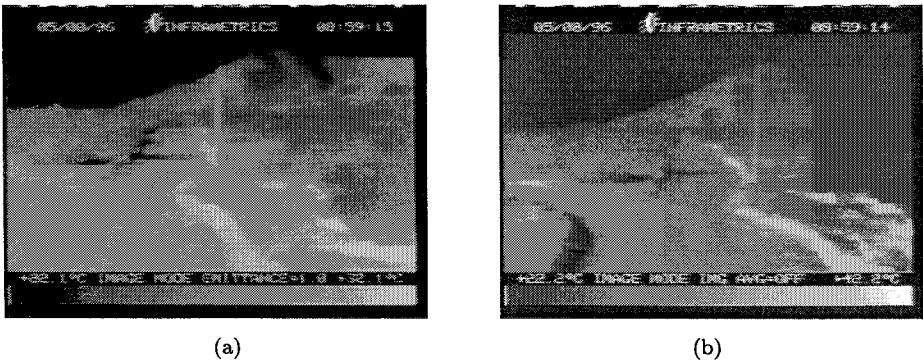(a)   (b)

Fig. 6.   (a) Right (b) Left IR image of road with 75% fov.

hazard. Figure 6 shows a dark shadow in the road that could be interpreted as a
dip whereas the actual dips are often not seen at all without using stereoscopic
vision. Our current research is now focusing on using wide baseline stereo vision
with chopped active source illumination for navigation through fog.

7. CONCLUSIONS

As the applications of multispectral image processing expand into the realm where
quantitative measurements are being used to make qualitative assessments, we must
rely more heavily on the recognition processes already developed in Nature. The
human vision process must be better understood and modeled. The inclusion of this
Nature Factor will not only change the mathematical approach we use to analyze
the multispectral image processing task, but also the manner in which the imagery
data to be used is obtained. The example of the basic omission of depth perception
in image processing analysis was highlighted, but there are many more ways that
we can benefit from vision systems that have been developed naturally.

REFERENCES

1. E. Ardizzone, D. Molinelli and R. Pirrone, "A fast and robust BCS application to the stereo vision," *Proc. 7th Italian Workshop on Neural Nets*, Vietri sul Mare, Salerno, Italy, 18–20 May 1995, pp. 215–225.
2. P. Bao and B. Leung, "Wavelet transform image coding based on bisual perception modeling," *Proc. SPIE* **3391** (1998) 649–658.
3. P. Bijl and M. Valeton, "TOD: a new method to characterize electro-optical system performance," *Proc. SPIE* **3377** (1998) 182–193.
4. A. C. Copeland and M. M. Trivedi, "Texture perception in humans and computers: models and psychophysical experiments," *Proc. SPIE* **2742** (1996) 436–446.
5. S. B. Crow, W. R. Watkins, F. R. Palacios and D. R. Billingsley, "Technique for measuring atmospheric effects on image metrics," *Proc. SPIE* **1486** (1991) 333–344.
6. V. G. CuQlock-Knopp, W. Torgerson, D. E. Sipes, E. Bender and J. O. Merritt, "A comparison of monocular, biocular and binocular night vision goggles for traversing off-road terrain on foot," U.S. Army Research Laboratory, Technical Report ARL-TR-747, 2800 Powder Mill Rd., Adelphi, MD 20783–1190, March 1995.
7. V. G. Cu-Qlock-Knopp, D. E. Sipes, W. Torgerson, E. Bender and J. O. Merrritt, "Human off-road mobility, preference and target-detection performance with monocular, biocular and binocular night vision goggles," U.S. Army Research Laboratory, Technical Report ARL-TR-1170, 2800 Powder Mill Rd., Adelphi, MD 20783–1190, August 1996.
8. A. K. Dalmia and M. Trivedi, "Acquisition of 3D structure of selectable quality from image streams," *Proc. Second IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, 5–7 December 1994, pp. 289–296.
9. I. Dinstein, G. Guy and J. Rabany, "On the compression of stereo images: preliminary results," *Sig. Process.* **17** (1989) 373–382.
10. R. Hecher, "Camaeleon — camouflage assessment by evaluation of local energy, spatial frequency and orientation," *Proc. SPIE* **1687** (1992) 342–349.
11. G. C. Holst, *Sampling, Aliasing, and Data Fidelity*, JCD Publishing, Winter Park, FL and SPIE Optical Engineering Press, Bellingham, WA, 1998.
12. B. Jähne, *Practical Handbook on Image Processing for Scientific Applications*, CRC Press, Boca Raton, FL, 1997.
13. A. K. Jain, *Fundamentals of Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.
14. J. B. Jordan, W. R. Watkins, F. R. Palacios and D. R. Billingsley, "Simulated dynamic effects of atmospheric turbulence on IR digital imagery," *Infrared Phys. Tech.* **37** (1996) 607–617.
15. B. Julesz, "Stereoscopic vision," *Vision Res.* **26**, 9 (1986) 1601–1612.
16. T. Kanade, "Development of a video-rate stereo machine," *Proc. 94 ARPA Image Understanding Workshop*, Monterrey, CA, November 14–16, 1994.
17. F. T. Kantrowitz and W. R. Watkins, "Bandpass optimization for low-altitude long-path infrared imaging," *Opt. Eng.* **33**, 4 (1994) 1114–1119.
18. M. Levine, *Vision in Man and Machine*, McGraw-Hill, New York, 1985.
19. S. B. Marapane and M. M. Trivedi, "Region-based stereo analysis for robotic applications," *IEEE Trans. Syst. Man Cybern.* **19**, 6 (1989) 1447–1464.
20. J. O. Merritt, "Stereoscopic display applications issues," video from Short Course #SC06, *SPIE's 12th Annual Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*, Orlando, FL, 13–17 March 1998.
21. V. I. Polejaev and M. A. Vorontsov, "Adaptive active imaging system based on radiation focusing for extended targets," *Proc. SPIE* **3126** (1997) 216–220.
22. P Proctor, "FSI pushing IR technologies into broader-based markets," *Aviat. Week Space Tech.*, April (1997) 49–53.

23. A. Rose and P. K. Weimer, "Physical limits to the performance of imaging systems," *Phys. Today*, September (1989) 24–32.

24. P. Soltan *et al.*, "Laser-projected 3D volumetric display system (Second Generation)," *Proc. SPIE* **3057** (1997) 496–506.

25. A D. Stocker *et al.*, "Analysis of infrared hyperspectral measurements by the joint multispectral program," *Proc. SPIE* **2469** (1995) 587–602.

26. M. M. Tilleman and K. Krishnaswami, "Flame penetrating high-speed photographic system," *Program and Abstracts*, The 10th Meeting on Optical Engineering in Israel, 2–6 March 1997, Jerusalem, Israel.

27. A. Toet, P. Bijl, F. L. Kooi and M. Valeton, "Quantifying visual target distinctness," *Proc. SPIE* **3375** (1998) 152–163.

28. M. M. Trivedi, "Computer and human 3D vision in outdoor environments," U.S. Army Res. Off. Sci. Services Prog. Report, Contract No. DAAL03-91-C-0034, TCN No. 95:050, December 1997.

29. M. M. Trivedi, G. Ravichandran and W. R. Watkins, "A novel multiple, widebaseline stereo approach for target ranging," *Proc. SPIE* **2742** (1996) 236–244.

30. M. M. Trivedi and M. V. Shirvaikar, "Quantitative characterization of image clutter: problem, progress and promises," *Proc. SPIE* **1967** (1993) 288–299.

31. W. R. Watkins, "Environmental bugs invade EO imaging systems," *Proc. SPIE* **1969** (1993) 42–53.

32. W. R. Watkins, S. B. Crow and F. T. Kantrowitz, "Characterizing atmospheric effects on target contrast," *Opt. Eng.* **30**, 10 (1991) 1563–1575.

33. W. R. Watkins and R. W. Dutro, "Contrast transmission predictions versus far-infrared imager measurements for low-altitude horizontal paths," *Opt. Eng.* **33**, 8 (1994) 2588–2593.

34. W. R. Watkins, J. B. Jordan and M. M. Trivedi, "Novel applications of hyperstereo vision," *Proc. SPIE* **3110** (1997) 72–88.

Wendell R. Watkins received his B.S. in physics from Louisiana State University in New Orleans in 1968 and his M.S. in physics from the Georgia Institute of Technology in 1970. He was drafted into the U.S. Army in 1970 and spent two years at the Atmospheric Sciences Laboratory, White Sands Missile Range, New Mexico, as an army scientist and engineer. In 1971 he received the Army Commendation Medal for his research in the area of laser absorption by the atmosphere. Since completing his military obligation, he has worked at the Atmospheric Sciences Laboratory, now Army Research Laboratory (ARL), as a civil service employee.

He has performed a wide range of research at the ARL dealing with the propagation and detection of electro-magnetic radiation through the atmosphere resulting in many open literature and government publications as well as six patents. He has been an active National Research Council adviser in the area of image processing since 1996 and has chaired a conference on Targets and Backgrounds at the SPIE AeroSense Symposium since 1990 in Orlando, Florida.

His current research includes the uses of wide baseline stereo vision in a variety of areas.

DETECTION OF SEA SURFACE SMALL TARGETS IN INFRARED IMAGES BASED ON MULTILEVEL FILTER AND MINIMUM RISK *BAYES* TEST*

Y.-S. MOON

Department of Computer Science and Engineering
The Chinese University of Hong Kong, Hong Kong
E-mail: ysmoon@cse.cuhk.edu.hk

ZHANG TIANXU,[†] ZUO ZHENGRONG and ZUO ZHEN

Institute for Pattern Recognition and Artificial Intelligence
Huazhong University of Science and Technology
State Key Lab. for Image Processing and Intelligent Control
Wuhan 430074, P.R. China
E-mail: txzhang@mail.hust.edn.cn

This paper discusses the research in small target detection in infrared images with heavy clutter background. For most infrared images, ship objects are rather dim in the relative dark sea surface background. The existence of scan line disturbance and noise also increases the difficulty in proper detection. Dim objects must be distinguished from a dark background. On the other hand, the small targets must also be distinguished from clutters. Through analysis of the targets and background, we build characteristic models of small ship objects, noise and sea backgrounds respectively, and indicate their differences in spatial and frequency domains among them. Based on the principles of signal processing, pattern recognition and artificial intelligence, we propose a combined algorithm for detecting sea surface small targets. In this algorithm, components of background and noise are first suppressed by a multilevel filter designed accordingly, meanwhile enhancing the target ones of interest. The pixels of the candidate targets are then discriminated by minimum risk *Bayes* test. Finally, according to *a priori* knowledge about the targets such as the ranges of their sizes, the targets of interest can be detected. In particular, the related probability distributions used by statistic decision are obtained by offline learning of typical training samples. Experiments show that the algorithm is excellent for such kinds of target detection and is robust to noise.

Keywords: Target detection; infrared imaging; multilevel filter; minimum risk *Bayes* test; pattern recognition; modeling and learning.

1. INTRODUCTION

Small target detection has found its wide applications in such areas as remote sensing, surveillance, environment monitoring, aerospace, medical diagnosis, robotics, and so on.[1,2,4–9] The crux of small target detection lies in the fact that available information from the targets are generally much less than that from the background so that the solution algorithms must explore the largest possible information inherent to the problem and utilize such information very effectively. Unfortunately, no general solution to the problem exists. For this reason, heuristic solutions to

the problems will not only have important application values but also theoretical significance in image analysis and signal processing. A few researchers have used wavelets and fractals as tools to attack the issues of small target detection. Casasent et al.[1] applied *Gabor* transform to detecting vehicles on the ground. Pell[4] used fractal dimension singular changing as a sign indicating candidate man-made objects on the ground in remote-sensed images. Recently, Wang and Zhang[6] made use of fractal dimension and fractal constant abrupt variation for detecting small ship targets. Despite their success, the following problems remain to be resolved: (1) automatic determination of optimal threshold for real-time detection, (2) construction of adaptable models of the targets, background and noise for effective target detection, and (3) mining of adequate inherent information to the problem for improving detection performance instead of only relying on a single signal processing technique. In view of the fact that detecting sea surface small targets in infrared images is a typical case, this work proposes and tests a new detecting approach, featuring synthesized techniques from signal processing, pattern recognition and artificial intelligence. The remainder of the paper is organized as follows. Section 2 gives an analysis of the characteristics of target, noise and background as well as their modeling. Section 3 presents multilevel filter for enhancing targets and suppressing background clutter. In Sec. 4 optimum threshold real-time estimation and target detection are discussed. Finally, experiments conducted on practical infrared image data and conclusion are reported in Sec. 5.

2. ANALYSIS OF TARGET AND BACKGROUND CHARACTERISTICS AND MODELING

Figure 1 shows one of the images we studied. The image was taken by an infrared imaging instrument which scans the scenes from left to right. We can clearly discern the horizontal scan lines. Because the infrared imaging instrument is easily disturbed by the environment, such as climate, vibrating, occlusion and so on, noises which are the randomly distributed bright dots in the image are obvious.
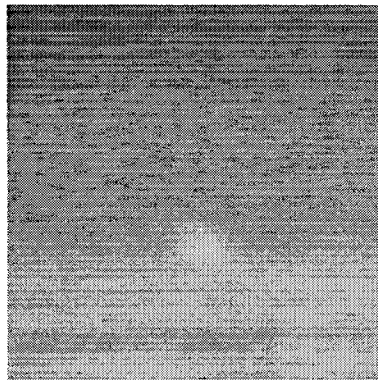

Fig. 1.   An infrared image of sea surface small target.

The background is the relative dark sea surface and the targets are the comparatively brighter parts which might be the ship objects. Even if the ships' gray level is in the small target domain, one can distinguish them from the nearby comparatively darker background. But general speaking, we cannot identify them only by their gray values because the domains occupied by the ships are small and their gray value distributions are often mixed with those of the background. Ships are continuous parts in the image when compared with dot noise. We can tell their difference by considering the difference of their space-domain appearance. For example, the ships' sizes in the studied images are in the range of 6 to 14 pixels in length, 6 to 10 pixels in height. The noise size is usually below 3×3 pixels.

Without losing generality, we assume that targets, noise and background clutter are all approximated respectively by cubes with different sizes or cube mixtures of different sizes. Let a standard cube $r_0(j, k)$ be represented by

$$r_0(j, k) = \begin{cases} 1, & |j| <= T_j, \quad |k| <= T_k \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $2T_j$ and $2T_k$ are the size of definition domain of r_0. Then any cube $r_{x,y}(j, k)$ in x–y plane can be denoted as

$$r_{x,y}(j, k) = f(x, y) r_0(x - \alpha j, y - \beta k) \tag{2}$$

where α, β are dilation coefficients in x–y coordinates, respectively, and $f(x, y)$ is considered as a dilation coefficient in the gray level axis.

In another view, we can study the characteristics of the background, target and noise in the frequency domain. Let us first describe the background, target and noise model in one dimension according to Eqs. (1) and (2). See Fig. 2.

This is an ideal model. We only consider the gray value and size attributes of background, target and noise, ignoring their relative positions since the position shift in space domain corresponds to the phase shift in frequency domain. Let G_n, G_t, G_b be the gray value of noise, target, background, respectively. $2T_n, 2T_t, 2T_b$ represent the covering range (nonzero size) of noise, target, background. C_n, C_t, C_b are the central positions of noise, target, background covering ranges, respectively.
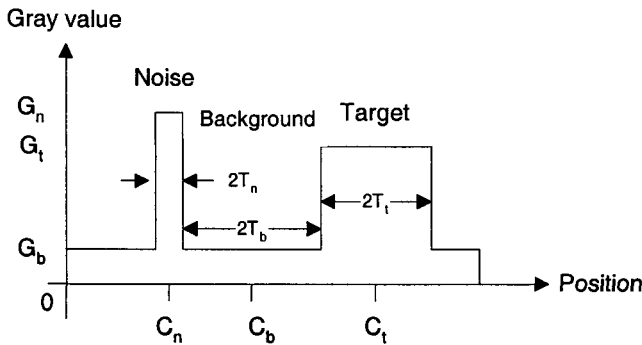
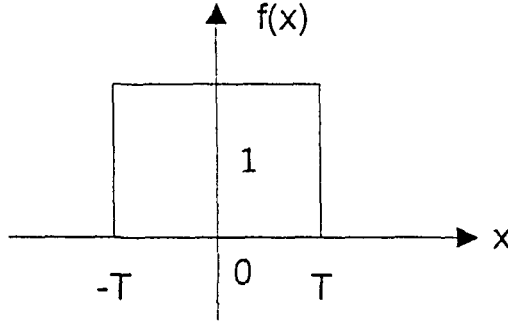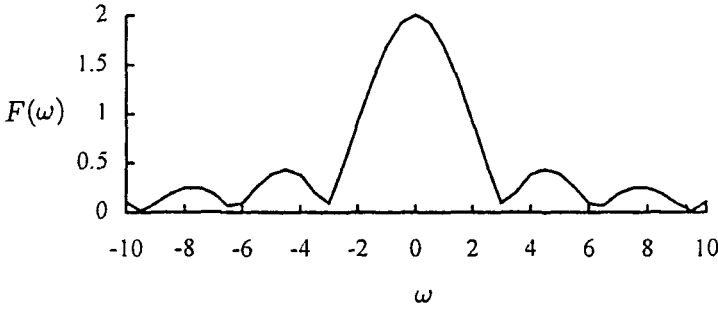

Fig. 2.   Ideal model of background, target and noise.

Fig. 3.   Gate function.

Fig. 4. $F(\omega)$ is amplitude spectrum of $f(x)$.

We find the three parts are similar in shape in Fig. 3 with variations of gray value, covering range and central position.

Figure 3 shows the gate function. Its *Fourier* transformation is given by

$$\int_{-\infty}^{+\infty} f(x)e^{-j\omega x}dx = \int_{-T}^{+T} e^{-j\omega x}dx = \frac{2\sin\omega T}{\omega} = \frac{2T\sin\omega T}{\omega T}. \tag{3}$$

This is a *sinc* function. Its modulus $F(\omega)$ is $|\frac{2T\sin\omega T}{\omega T}|$, as shown in Fig. 4.

$F(\omega)$ gets its maximal value at $\omega = 0$. At $\omega = \frac{k\pi}{T}$, $k = \pm 1, \pm 2, \ldots$, $F(\omega) = 0$, and at $\omega = \frac{k\pi+\pi/2}{T}$, $k = \pm 1, \pm 2 \ldots$, $F(\omega)$ gets its local maximal value $F(\frac{k\pi+\pi/2}{T}) = \frac{4T}{(2k+1)\pi}$. Clearly, the decadence of $F(\omega)$ along the ω axis is dependent on T value. In other words, we can say the energy distribution of $F(\omega)$ is dependent on the T value considering $F^2(\omega)$ represents signal energy. When T increases, the zero points and local maximal points of $F(\omega)$ compress along the ω axis, implying that the signal energy is mainly distributed in the low frequency part. When T decreases, the zero points and local maximal points of $F(\omega)$ expand along the ω axis, implying that the signal energy in the high frequency part cannot be neglected.

Large T corresponds to the background part (including scan lines) in our ideal model. Small T corresponds to the target part and very small T corresponds to the noise. The amplitude of target spectrum in the middle and high frequency parts is

relatively larger than the background. Noise has more frequency components in the high frequency part. Although this is only the ideal model, the same features also occur in real life images.

One can extend this conclusion to the two-dimensional image. The *Fourier* transformation of a cube function defined on a rectangular region is a two-dimensional *sinc* function, which has the similar property as the one dimension *sinc* function. Therefore, we get the same conclusion: background is relatively dominant in low frequency part, target is viewable in middle frequency part and noise is more obvious in high frequency part in the frequency domain. Thus, two-dimensional *sinc* function can be viewed as a frequency-domain modeling of targets, noise and background.

3. MULTILEVEL FILTER FOR TARGET ENHANCEMENT

Relying on the above space-frequency domain modeling as the first step for detecting targets, we try to use a set of filters to enhance the target and suppress background clutter. It seems that using a multilevel filter can solve the problem.

We first apply a Low Pass Filter (LPF) to the image. If the truncate threshold of the LPF is well chosen, the background part will be enhanced while the target and noise will be inhibited. Let F represent the amplitude spectrum of an original image f and Lp represent the amplitude spectrum of LPF. The result of f passing through the LPF can be written as $F \times Lp$. By subtracting the LPF result from the original image, the background part, written as $F - F \times Lp = F(1 - Lp)$. Similarly, we can inhibit the noise and enhance the target. This is then followed by another LPF whose amplitude spectrum is Lq written as $F(1 - Lp)Lq$. After this series of processing, the target part can be easily detected from the image.

The next problem is the determination of the truncate threshold of the LPF. We know, if a LPF is followed by the LPF, the total amplitude spectrum is $Lp \times Lp$. The total spectrum contracts when compared with Lp. This means that by concatenating two LPF, we get a narrower band than LPF. For this reason, we can begin with a wider band LPF, followed by another, to test if the band width is proper. This is equivalent to manual checking whether the target stands out or not. Similarly, Lq can also be constructed by this concatenation method. We can achieve good detect results if we properly adjust the number of concatenates in the stage of training and learning of the system. Figure 5 shows this multilevel filter system used for small target detection. To construct the system easily, we can use the same LPF for all levels. This approach can help to cut down the costs when constructing the system using hardware.
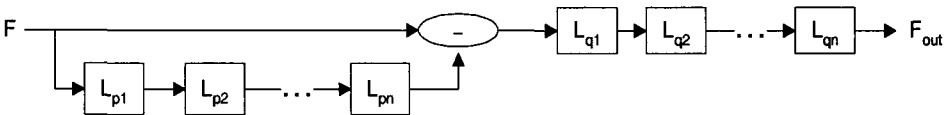

Fig. 5. Multilevel filter for detection.

We can also directly construct the LPF in space domain. To illustrate it, we can use a five-point average LPF as an example. The width of the filter is 5, noted as $W_f = 5$. If the width of the signal is W_s and $W_f > W_s$, when the signal passes through LPF, it will be seriously weakened. If $W_f \ll W_s$, changes only occur at the edge of the signal. We know the width of two successive concatenation filters $W(2) = W(1) + W(1) - 1$, where $W(n)$ means the width of n concatenated filters, it can be calculated based on the above equation. For the use in n-level filter detection, we should choose n, such that $W(n) >= Wt$, and $W(1) \gg Wn$, where Wt means the width of target and $W(1)$ is the basic filter unit. This gives relationships among the basic parameters of the filters and sizes of targets, noise and background clutter. For example, if the width of target is 9, we can use $W(1) = 5$, and $n = 2$ to denote that two concatenated LPF are used. In practice, we have found that the use of multilevel filter gives good detection result.

4. OPTIMUM THRESHOLD REAL-TIME ESTIMATE AND TARGET DETECTION

Due to departure of practice from ideal modeling as well as no perfect filtering available, the above filter is unable to isolate the targets of interest without some remaining clutter. Hence, new detection algorithm should be implemented in at least two steps. First as mentioned above, we can apply multilevel filter to the image to inhibit noise, background and enhance target. Next, pattern recognition method is used to distinguish targets from the remaining clutter. Figure 6 shows the procedure in the form of a flowchart.

After carefully studying the image characteristics, we can choose proper basic filter parameters, structure and the level number of filter used. An original image f first passes through the series of Lp-based multilevel filter, the filtered image f_1 is then subtracted from f. Next, the resultant image f_2 passes through another Lq-based multilevel filter. Finally, we can use the contrast of the filtered image f_3 as a feature for small target detection.

Define the filtered image contrast as

$$C = \frac{|f_3 - a|}{a} \tag{4}$$

where f_3 represents the gray level of filtered image, a is contrast average of the image. Assume that contrast x of the filtered image is a random variable. Let $P(R_1)$ and $P(R_2)$ be the occurrence probabilities of the background and target pixels, respectively. Also, let $p(x/R_1)$ and $p(x/R_2)$ be the probability distribution functions of the contrast values of the background and target pixels of the filtered image, respectively. In this way, the detection of targets can be regarded as a pattern classification problem involving two classes[3]: target and background so that we can use the minimum risk *Bayes* test method to choose an optimal classification discrimination threshold t.

Let the contrast threshold be t, the background region be R_1 and the target region be R_2. If the threshold t splits R_1 into R_{11} and R_{12} where R_{11} and R_{12} are
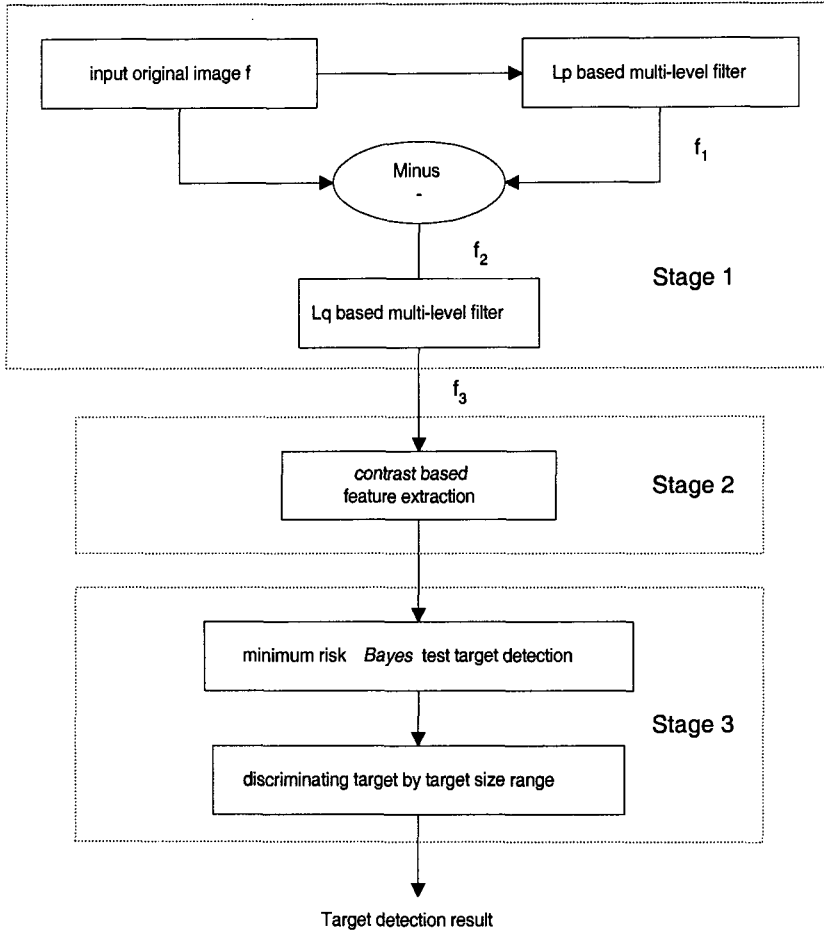
Fig. 6. A combined algorithm for small target detection.

the respective right and wrong classifications of background region, respectively. Let the cost for R_{11} be $C_{11} = 0$. Similarly, let the cost of R_{12} be C_{12}. Threshold t also splits R_2 into R_{21} and R_{22} which are the respective wrong and right classifications of target region. Let the cost for R_{22} be $C_{22} = 0$. Also, R_{21} is the wrong classification of target region into background class and its cost is C_{21}. Then the total cost $r(t)$ can be calculated in discrete condition as

$$r(t) = \sum C_{12} P(R_1) p(x/R_1) + \sum C_{21} P(R_2) p(x/R_2) \qquad (5)$$

where $r(t)$ can be minimized by adjusting t. The $\hat{t}$ value corresponding to the minimal r is the proper contrast based segmentation threshold, i.e.

$$r(\hat{t}) = \min r(t). \qquad (6)$$

These formulas of probability and distribution are determined by a supervised offline learning and training of the detection system on a large sample set of infrared images, with man-machine interface. A flow chart of the training procedure is
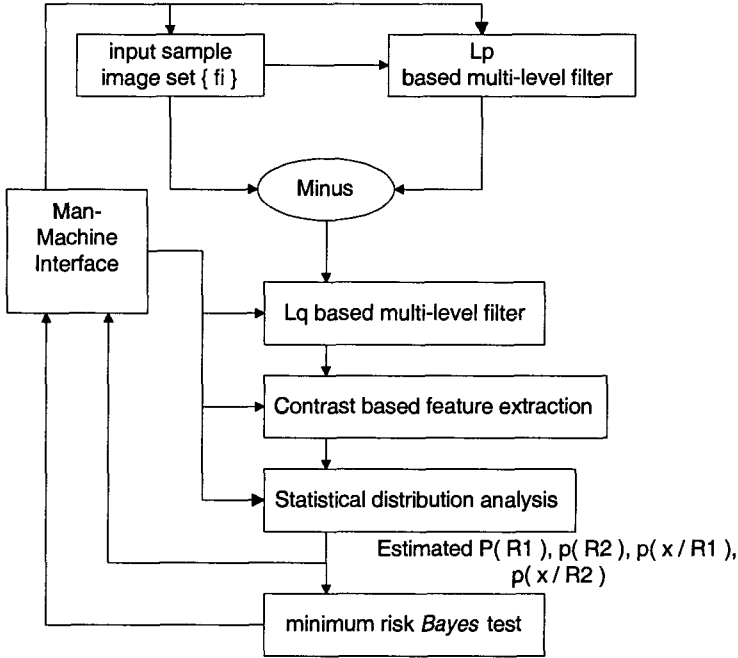
Fig. 7.   A supervised procedure of learning and training.

shown in Fig. 7. Despite it being common to assume these distributions are normal distribution, practical situations are often a serious departure from it. Hence the training and learning on samples is necessary for effective detection of targets.

The final detection step utilizes *a priori* knowledge of targets to further improve the performance of the methodology. In this experiment, the candidate target regions whose sizes are less than a given threshold are discarded.

5. EXPERIMENTAL RESULTS AND CONCLUSION

Through learning and training on a set of large sample images, we obtained Fig. 8 showing the contrast distribution of background and target. This forms the foundation of using minimal risk *Bayes* test to get the right classification threshold. The *priori* contrast conditional distributions of target and background were learned from the sample set containing 80 sample images. Figure 9 shows several sample images. We first marked all target pixels in the sample images manually, and then, calculated the contrast images according to the processing steps described previously. Finally, the contrast distribution of object and background can be estimated from the contrast images with the known position of the target pixels. The calculation of contrast images and the estimation were carried on the small window with size 50×32, which slide along the images.

The functions fitting to the statistical data of Figs. 8(a) and 8(b) are $p_1(x) = \lambda e^{-\lambda x}$ and $p_2 = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$ respectively, shown in Fig. 8(c). For the distribution curve shown in Fig. 8(a) , the parameters are: $\lambda = 1.9192$. For the distribution
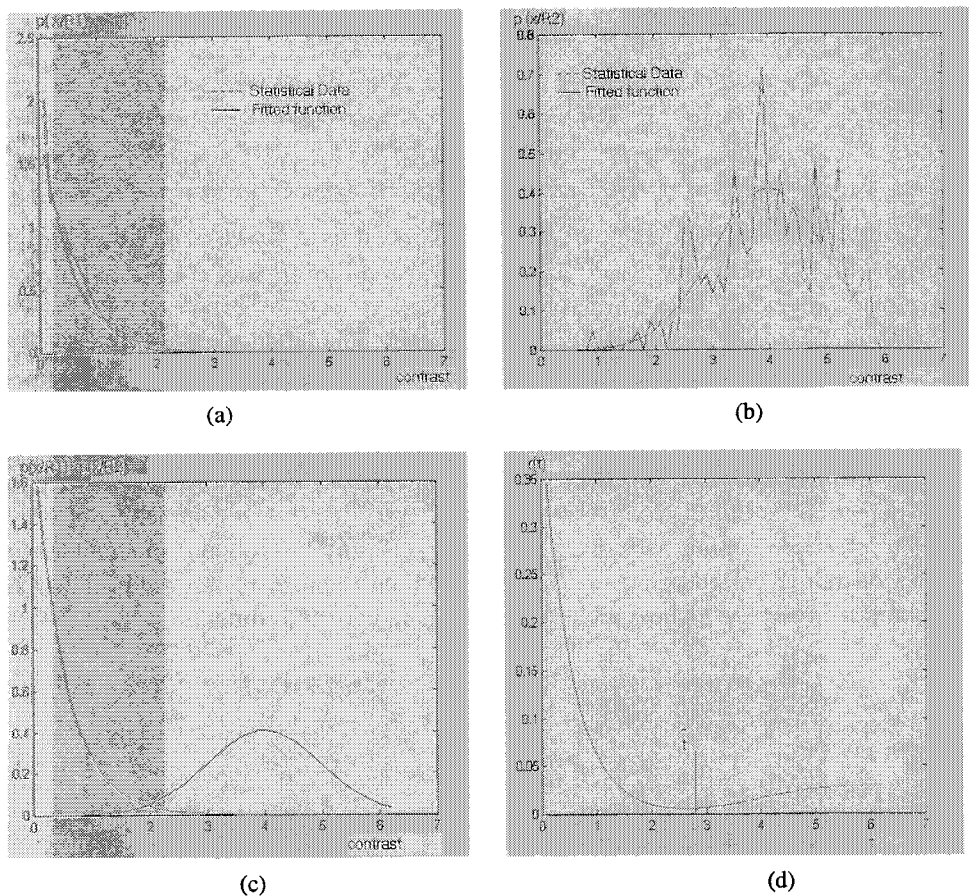
(a)

(b)

(c)

(d)

Fig. 8. (a) Contrast distribution of background $p(x/R_1)$. (b) Contrast distribution of target $p(x/R_2)$. (c) Fitted contrast distributions of background and target. (d) Plotting of $r(\hat{t})$.
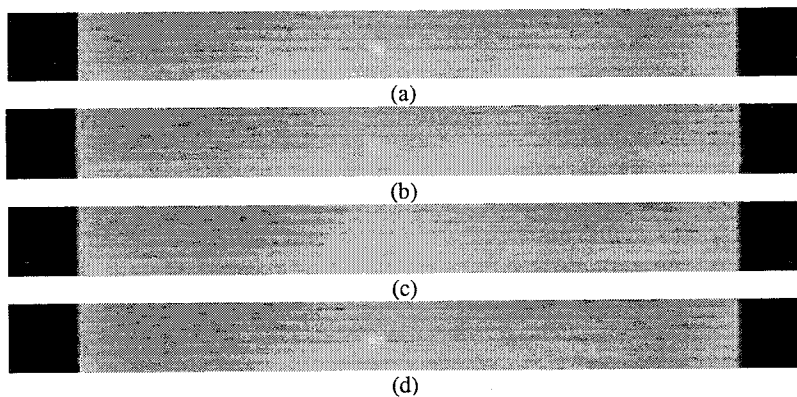

(a)

(b)

(c)

(d)

Fig. 9. Some sample images used for estimating the *priori* conditional distribution of the target and background.

Table 1. Optimal threshold $\{\hat{t}\}$ under varying cost (C_{12}, C_{21}).

C_{12}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
C_{21}	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\hat{t}$	2.1	2.4	2.5	2.6	2.8	2.9	3.0	3.2	3.6

curve shown in Fig. 8(b), the parameters are: $\mu = 3.9811, \sigma = 0.9780$. It is obvious that the fitted probability distribution of background contrast of the filtered images deviates significantly from normal distribution.

Table 1 lists different optimal thresholds under changing cost (C_{12}, C_{21}).

Figure 10 shows the experimental results of a sequence of images (generated using the above discussed methodology).

We have done many experiments on a series of sea surface infrared images using the above method. The excellent results demonstrate accurate target detection. Our detection method is robust to noise. However, due to the complexity of small target detection, further work should include: (1) stability problem[9] of small target detection under various kinds of clutters; (2) enhancement of the adaptability or robustness of the algorithms under random changing environments.
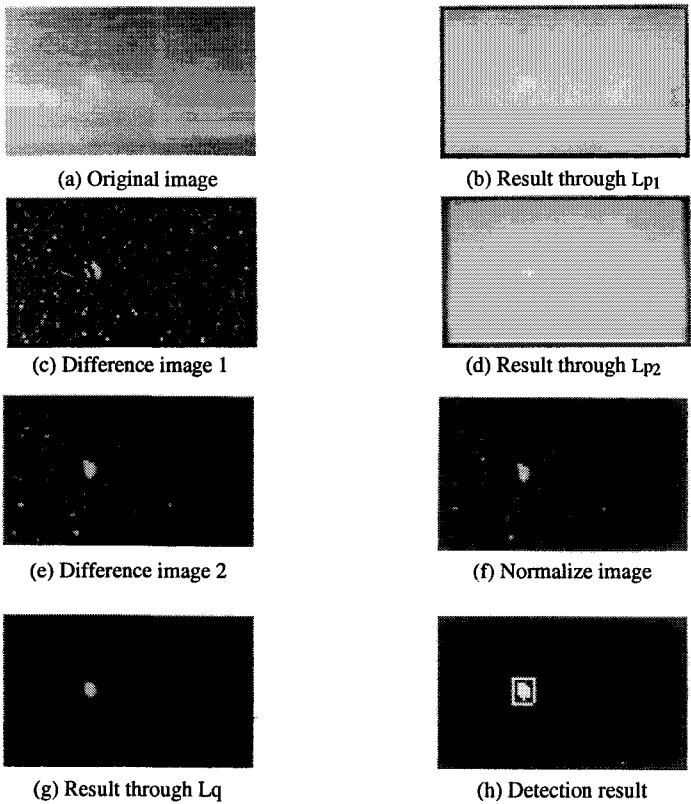

(a) Original image

(b) Result through Lp1

(c) Difference image 1

(d) Result through Lp2

(e) Difference image 2

(f) Normalize image

(g) Result through Lq

(h) Detection result

Fig. 10.   Experiment result of a sequence of sea surface infrared images with heavy clutters.

ACKNOWLEDGMENT

REFERENCES

1. D. P. Casasent, J. Smokelin and A. Ye, "Wavelet and Gabor transforms for detection," *Opt. Eng.* **31**, 9 (1992) 1893–1898.
2. O. E. Drummond, "Signal and data processing of small targets 1994," *Proc. SPIE*, Vol. 2235, Orlando, 1994.
3. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1972.
4. T. Pell, "Multiscale fractal theory and object characterization," *J. Opt. Soc. Am. Ass.* **7**, 6 (1990) 1401–1412.
5. R. N. Strickland and H. I. Hahn, "Wavelet transform method for object detection and recovery," *IEEE Trans. IP* **6**, 5 (1997) 724–735.
6. G. Wang, T. Zhang, L. Wei and N. Sang, "Efficient method for multiscale small target detection from a natural scene," *Opt. Eng.* **35**, 3 (1996) 761–768.
7. G. Wang, J. Li and T. Zhang, "Approach to target detection based on scale changing fractal signature," *Proc. SPIE*, Vol. 3062, 1997.
8. T. Zhang, N. Sang, G. Wang and X. Li, "An effective method for identifying small objects on a complicated background," *Artif. Intell. Eng.* **10**, 4 (1996) 343–349.
9. T. Zhang and X. Xie, "An investigation on the stability of object extraction," *IEEE Trans. AES* **33**, 3 (1997) 1051–1060.

Yiu-sang Moon received his B.Sc. (Honours) degree from the University of Manitoba in 1973, M.Sc. and Ph.D. degrees from the University of Toronto in 1974 and 1979 respectively.

Dr. Moon joined the Computer Science and Engineering Department of the Chinese University of Hong Kong as an assistant lecturer in 1978 when the department first started its major computer science programme. He became a lecturer in 1979 and a senior lecturer in 1989.

His research interests range from numerical computering to computer graphics. In recent years, he has been focussing research activities on practical computer problems, especially in Oriental outline fonts, digital color publishing and Chinese computing.

He has published about 60 papers and had received research grants from sources such as the UPGC and the Chinese University, etc. Dr. Moon also actively participates in computer education programmes of the Hong Kong Secondary Schools by serving for more than ten years in various computer studies committees set up by the Hong Kong Examination Authority and the Department of Education of the Hong Kong Government.
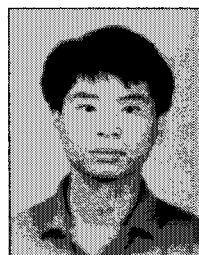
Zuo Zhengrong received the M.S. degree from the China University of Geoscience in 1993, and worked as a lecturer in the China University of Geoscience from 1993 to 1998. Now, he is a Ph.D. candidate in pattern recognition and intelligent systems at Huazhong University of Science and Technology.

His research interests are image processing and pattern recognition.
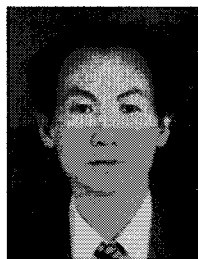
Zhang Tianxu received the B.S. degree from the University of Science and Technology of China, M.S. degree from Huazhong University of Science and Technology, Ph.D. degree from Zhejiang University, China. He has been a Professor and Director since 1993 at the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, P.R. China. He has published more than 90 academic and technical papers. In 1997, he served as Visiting Professor at Bordeaux III University, France. In 1999, he served as a Research Scholar at the Department of Computer Science and Engineering, Chinese University of Hong Kong. In 1998, he served as Co-chair of SPIE' International Symposium of Multispectral Image Processing, held in Wuhan, China.

Dr. Zhang is a member of the IEEE, and senior member of China Society of Image and Graphics. He is Deputy Editor-in-Chief on the editorial board of *J. Infrared and Laser Engineering*.

His research interests include image processing, computer vision, pattern recognition and parallel processing.

Zuo Zhen received the B.S. degree from the University of Science and Technology of China, M. S. degree in science and technology from Huazhong University of Science and Technology in 1993 and 1998 respectively.

His interests include image processing, image compression and pattern recognition.

Photograph of Z. Zhen is not available.

MINIMUM DESCRIPTION LENGTH METHOD FOR FACET MATCHING

STEPHEN MAYBANK* and ROBERTO FRAILE†

Department of Computer Science, The University of Reading
Whiteknights, Reading, Berkshire, RG6 6AY, UK

The Minimum Description Length (MDL) criterion is used to fit a facet model of a car to an image. The best fit is achieved when the difference image between the car and the background has the greatest compression. MDL overcomes the overfitting and parameter precision problems which hamper the more usual maximum likelihood method of model fitting. Some preliminary results are shown.

Keywords: Model fitting; facet model; minimum description length; MDL; log* code; Kolmogorov complexity; code length; tracking.

1. INTRODUCTION

If a set of data is a good fit to a model, then loss free data compression can be achieved by first coding the model and then coding any deviations of the data values from those predicted by the model. If the model is simple and the deviations are small, then a large compression can be achieved. In the Minimum Description Length (MDL) approach to model fitting the length of the compressed data, measured in bits, is the measure of the accuracy of the fit of the model to the data. If there are a number of models, then the one which achieves the shortest length of compressed data is chosen.[7-9,12,13]

MDL overlaps three scientific disciplines: statistics, computer science and information theory. In statistics MDL provides a way of assigning probabilities to complex models,[12,13] in computer science it is an approximation to Kolmogorov complexity,[8,9] and in information theory it provides a method of coding where storage space is reduced at the cost of increased computation.[9] The theoretical justification for MDL is found in the theory of Kolmogorov complexity.

At first sight, it is surprising that a measure as simple as bit length can provide a useful basis for model selection. The theory underlying MDL shows that if an agent seeks to make good hypotheses about incoming data, and if it has a classical computer but not any other more powerful device, then data compression is a good way of choosing hypotheses. Compression is not always optimal, but it is guaranteed never to make a bad choice of hypothesis.

The Kolmogorov complexity of a string s is the length in bits of the shortest program which can compute s from zero input. The models in MDL correspond to different programs for compressing s. Unfortunately, Kolmogorov complexity is not computable in that there is no algorithm which can take an arbitrary string

s and return its Kolmogorov complexity. In applications of MDL computability is achieved by restricting the candidate programs to those considered "natural" or "appropriate" in the context of the application.

Let s be a bit string which records the data and let θ be a vector of parameters which specify the model. Both the number of components in θ and the values of the different components may vary. The vector θ is coded as a string $M(\theta)$ and the deviations of the data from the model are coded as a string $C_\theta(s)$. Here, all strings are bit strings, i.e. strings of 0s and 1s only. The string s is coded as $M(\theta).C_\theta(s)$ where "." is string concatenation. The original string s is recovered from $M(\theta).C_\theta(s)$ by first reading $M(\theta)$, then constructing the model and finally using the information in $C_\theta(s)$ to adjust the model until it coincides with s. The codes for $M(\theta)$ and $C_\theta(s)$ must be prefix codes, firstly so that the end of $M(\theta)$ in $M(\theta).C_\theta(s)$ can be detected and secondly to ensure that code lengths correspond to probabilities.

Let $s \mapsto \lambda(s)$ be the length function defined on strings. The model θ is a good fit to s if $\lambda(M(\theta)) + \lambda(C_\theta(s)) \ll \lambda(s)$.

The MDL approach to model fitting has the following advantages.

(i) It has a good theoretical foundation in Kolmogorov complexity.
(ii) It can accommodate an infinite number of models.
(iii) Overfitting is discouraged, because the additional components of θ increase the length of $M(\theta)$ without decreasing the length of $C_\theta(s)$.
(iv) The optimal precision of the components of θ can be found by minimizing $\lambda(M(\theta)) + \lambda(C_\theta(s))$ over a range of precisions.

MDL is applicable in cases where the usual maximum likelihood (ML) method for model fitting fails. For example, ML cannot determine the precision of the parameters or detect overfitting. Various extensions of ML exist in which overfitting is discouraged by attaching a cost to the number of parameters,[1,6] but these extensions require strong statistical assumptions about the errors in the data. The MDL cost of a given set of parameter values is easy to compute, and it is possible to compare the costs of radically different models in a principled way.

MDL is often combined with a probabilistic approach,[5,7,12,13] in which the data x is drawn, at least nominally, from a sample space carrying a parameterized density $p(x|\theta)$. The length of the bit string $C_\theta(s)$ is set equal to $-\log(p(x|\theta))$. The rationale is that this gives the shortest expected code length over all possible realizations of the data: data with high probabilities are assigned short codes. In this paper we avoid any explicit use of probability distributions and seek only to find a short code for the current data, without invoking a hypothetical sample space of all the data which might have been observed.

Some applications of MDL have been made in computer vision.[4,5,7,10,11] The most striking application is due to Leclerc,[7] who describes an image segmentation algorithm based entirely on MDL. More usually, MDL occurs as part of a larger algorithm or computational scheme.

In this paper MDL is applied to a central task in model based vehicle tracking,[14,15,18] namely the fitting of a facet model to the image of a car. The model

parameters are the position and orientation of the car on the ground plane. Potential applications include vehicle counting, detection of hazardous situations or anomalous driver behavior, and vehicle classification. The major advantage of MDL over previous fitting methods is that it allows a rational comparison of different hypotheses about the shape, appearance and position of the car without having to construct detailed probabilistic models.

2. MDL FOR A FACET

We describe two algorithms for coding the gray levels in an image region. The first achieves compression when the gray levels vary smoothly with position. The second achieves compression when the gray levels vary randomly. In an application to model fitting the region R would be the projection of a model facet, as shown for example in Fig. 1.
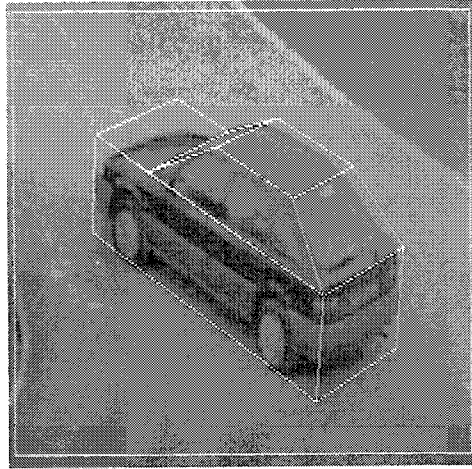
Fig. 1.   Facet model projected to the image.

2.1. First Algorithm

The gray levels g_{ij}, $(i,j) \in R$ are listed in a standard order, for example lexicographic in i and j. A polynomial of degree e is fitted to the g_{ij} by choosing coefficients r_{pq} such that the following sum is minimized,

$$\sum_{(i,j)\in R} \left(g_{ij} - \sum_{p+q\leq e} r_{pq} i^p j^q \right)^2 . \tag{1}$$

The coefficients r_{pq} are infinite precision real numbers, and thus unsuitable, as part of a short description of the gray levels in R. They are replaced by ratios of integers. Let $d_0, \ldots d_e$ be levels of precision, and define ratios of integers a_{pq}/b_{pq} by

$$a_{pq}/b_{pq} = \texttt{Rationalize}[r_{pq}, d_{p+q}] \qquad 0 \leq p+q \leq e . \tag{2}$$

Define integers δ_{ij} by

$$\delta_{ij} = \text{Floor}\left[g_{ij} - \sum_{p+q\leq e} a_{pq}i^p j^q / b_{pq}\right] \qquad (i,j) \in R. \tag{3}$$

In (2) and (3), `Rationalize` and `Floor` are Mathematica functions.[17] `Rationalize[r,d]` returns the rational number q with the smallest denominator such that $|r - q| < d$. `Floor[r]` returns the greatest integer n such that $n \leq r$.

The code for the δ_{ij} is based on Rissanen's[13] universal prior density for the positive integers. Rissanen's density $n \mapsto p_n$ is an appropriate default density when there is little or no prior information about the integers that might be encountered. Its most striking characteristic is that p_n decreases very slowly as n increases.

To define p_n, let log be the logarithm to base 2 and let $\log^*$ be the function defined by

$$\log^*(x) = \log(x) \qquad\qquad 1 \leq x \leq 2$$
$$\log^*(x) = \log(x) + \log^*(\log(x)) \qquad 2 < x.$$

The above definition of $\log^*$ follows Baxter,[3] but differs from the $\log^*$ function defined by Li and Vitányi in Ref. 9. The density $n \mapsto p_n$ is given by $p_n = c^{-1}2^{-\log^*(n)}$, $n \geq 1$ where c is a normalizing constant chosen such that

$$\sum_{i=1}^{\infty} p_n = 1.$$

The density $n \mapsto p_n$ has an associated prefix code such that the code $d(n)$ for the integer n has length $\log^*(n) + O(1)$. A program to compute $d(n)$ is given by Baxter.[3] The following code is an implementation of Baxter's coding algorithm in Mathematica:

```
logstar[n_] := Block[{ls,k},
    If[n == 1, Return[{0}]];
    ls = {IntegerDigits[n, 2]};
    k = Length[ls[[1]]]-1;
    While[k >= 2,
        ls = Join[{IntegerDigits[k,2]},ls];
        k = Length[ls[[1]]]-1];
    Join[Flatten[ls],{0}]
]
```

Let abs(n) be the absolute value of the integer n. A prefix code $n \mapsto c(n)$ is defined on the integers, both positive and negative, by

$$c(n) = d(2n) \qquad\qquad n \geq 1$$
$$c(n) = d(2\,\text{abs}(n) + 1) \qquad n < 1.$$

δ_{ij} are coded using c. Briefly, the pixels in R are listed in the standard order. Let the induced order on the δ_{ij} be $\delta_1, \ldots, \delta_{n(R)}$, where $n(R)$ is the number of pixels in R. The gray levels g_{ij}, $(i, j) \in R$ are coded as

$$d(e).c(a_{00}).c(b_{00}) \ldots c(a_{0e}).c(b_{0e}).d(n(R)).c(\delta_1) \ldots c(\delta_{n(R)}). \tag{4}$$

The precisions $d_0 \cdots d_e$ are varied until the code (4) has a minimum length. If this length is less than $\lambda(d(n)) + 8n(R)$ bits, then the bivariate polynomial of degree e is a possible model for the gray levels in R.

2.2. Second Algorithm

The second algorithm is based on Huffman coding.[2] Suppose that the pixels in R take m distinct gray level values, h_i, $1 \leq i \leq m$ and let f_i be the frequency of h_i, $1 \leq i \leq m$. The f_i are used as probabilities for a Huffman code. A histogram of the gray levels obtained for a typical facet is shown in Fig. 2.
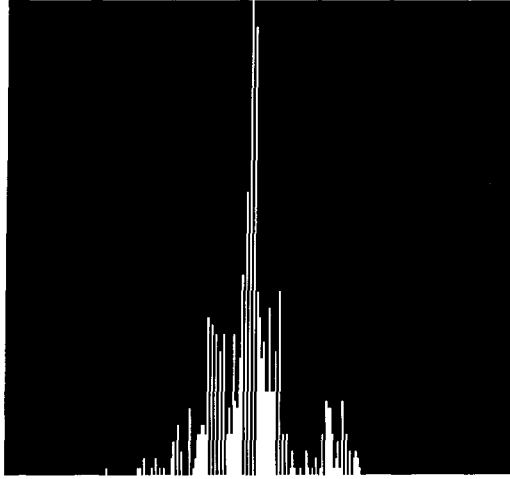
Fig. 2. Gray level histogram of a facet.

Let c_i be the code word corresponding to f_i, $1 \leq i \leq m$ and let $h \mapsto k(h)$ be the function that assigns the code word $k(h)$ to the gray level h. The gray levels in R are listed in a standard order $g_1, \ldots, g_{n(R)}$, and coded by the string

$$d(m).h_1.c_1. \ldots .h_m.c_m.d(n(R)).k(g_1). \ldots .k(g_{n(R)}).$$

To achieve compression, the number $n(R)$ of pixels has to be large enough to ensure that the compression of the data compensates for the cost of recording the coding table $h_i \leftrightarrow c_i$, $1 \leq i \leq m$.

3. MODEL FITTING

The car is modeled by seven facets, as shown in Fig. 1. The model is provisional in that it does not include those surfaces of the car not visible in Fig. 1. There are

three parameters: the position (x, y) and the orientation θ of the car in the ground plane. The position of the ground plane relative to the camera and the internal parameters of the camera are determined offline, using the method described by Worrall *et al.*[18] With this information the projection of the model to the image can be calculated for any set of values x, y, θ.

The coding is applied to the gray levels in an image window measuring 400 pixels horizontally by 300 pixels vertically. Let B be a fixed background image not containing the car and let I be an image containing the car. It is assumed that B is known, both to the coder and the decoder. Let x, y, θ be hypothetical values for the position and orientation of the car on the ground plane. The model is projected into I and two regions R_1, R_2 are defined, where R_1 is the set of pixels of I included in the projection of the model, and R_2 is the remainder $R_2 = I \setminus R_1$.

Let J be the difference image defined by $J_{ij} = I_{ij} - B_{ij}$. If the parameter values x, y, θ are well chosen, then the gray levels J_{ij} for $(i, j) \in R_2$ will be small. The image J is coded using the log* code for the gray levels of pixels in R_2 and the normal 8 bit code for the gray levels of pixels in R_1.

4. RESULTS

The results obtained so far are provisional in that the two coding algorithms described in Sec. 2 have not yet been applied to the model fitting. Instead, as noted in Sec. 3, we use a simpler but less efficient coding in which the background gray levels in the difference image have the log* code, and the car gray levels have the usual fixed length 8 bit code.

The orientation θ of the car is held constant. Let $f(x, y)$ be the code length for the difference image, as a function of position (x, y) on the ground plane. A graph of f is shown in Fig. 3. The graph is constructed by taking 169 sample values (x_i, y_j) for $1 \leq i \leq 13$, $1 \leq j \leq 13$. The intervals $|x_{i+1} - x_i|$, $|y_{j+1} - y_j|$ are each of length 25 cm. The labelings on the x and y axes in Fig. 3 correspond to i and j
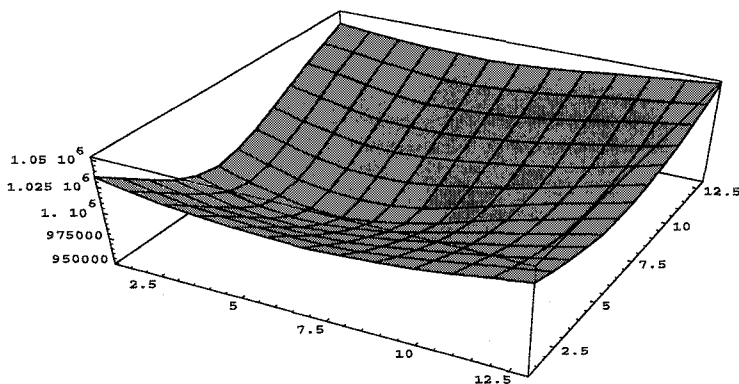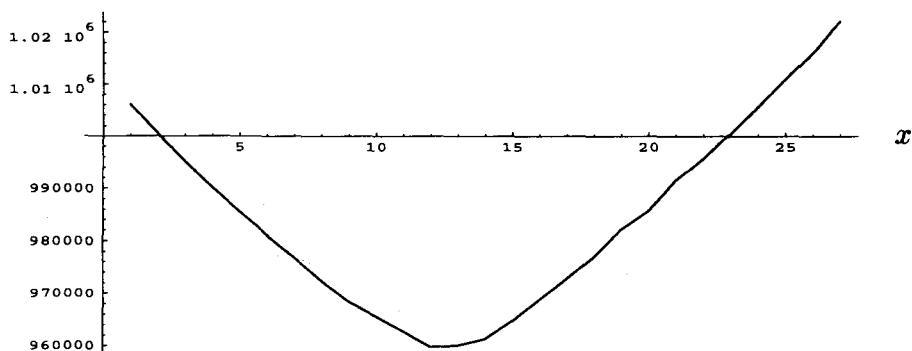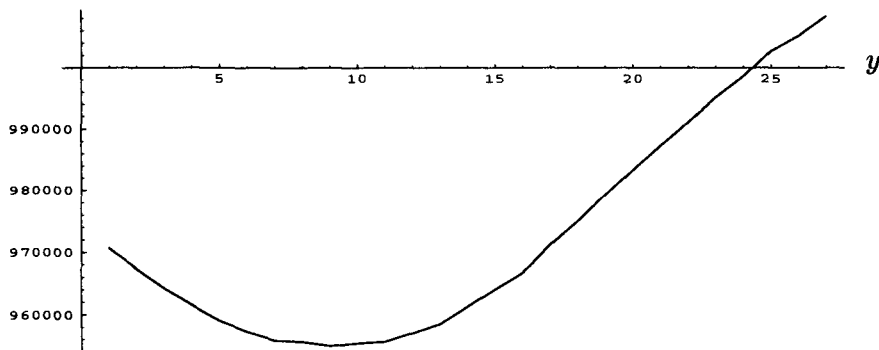

Fig. 3.   Graph of code length as a function of position (x, y).

code length

Fig. 4.   Graph of code length as a function of x.

code length


Fig. 5.   Graph of code length as a function of y.

respectively, i.e. they run from 1 to 13. The surface has a clearly defined minimum, located which we suppose, near to $(13, 9)$. The function $x \mapsto f(x, 9)$ is shown in Fig. 4 and the function $y \mapsto f(13, y)$ is shown in Fig. 5.

5. CONCLUSION

An application of MDL to the task of fitting a facet model to the image of a car has been described. It is apparent from Figs. 3–5 that the car can be located in spite of the discrepancies between the shape of the model and the shape of the car. In future work, the model will be made more accurate and the coding will be made more efficient, firstly by using a Huffman code for the background gray levels and secondly by using polynomial approximations to model the gray levels on the different facets of the model for the car.

REFERENCES

1. H. Akaike, "A new look at statistical model identification," *IEEE Trans. Automat. Contr.* **19** (1974) 716–723.
2. R. Ash, *Information Theory*, Interscience Tracts in Pure and Applied Mathematics, Vol. 19, John Wiley, N.Y., 1965.
3. R. Baxter, "Minimum message length inductive inference — theory and applications," Ph.D. Thesis, Department of Computer Science, Monash University, Australia, 1996.
4. T. Darrell and A. Pentland, "Cooperative robust estimation using layers of support," *IEEE Trans. Patt. Anal. Mach. Intell.* **17** (1995) 474–487.
5. P. Fua and A. J. Hanson, "An optimization framework for feature extraction," *Mach. Vis. Appl.* **4** (1991) 59–87.
6. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, 1996.
7. Y. G. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. J. Comput. Vis.* **3** (1989) 73–102.
8. M. Li and P. M. B. Vitányi, "Inductive reasoning and Kolmogorov complexity," *J. Comput. Syst. Sci.* **44** (1982) 343–384.
9. M. Li and P. M. B. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Graduate Texts in Computer Science, 2nd edition, Springer, Berlin, 1997.
10. T. Lindeberg and M.-X. Li, "Segmentation and classification of edges using minimum description length approximation and complementary junction clues," *Comput. Vis. Imag. Understand.* **67** (1997) 88–98.
11. A. Pentland, "Part segmentation for object recognition," *Neural Comput.* **1** (1989) 82–91.
12. J. Rissanen, "Stochastic complexity," *J. R. Stat. Soc. Series B* **49** (1987) 223–239.
13. J. Rissanen, "A universal prior for the integers and estimation by minimum description length," *Ann. Statist.* **11** (1983) 416–431.
14. G. D. Sullivan, "Visual interpretation of known objects in constrained scenes," *Phil. Trans. R. Soc. Lond. Series B* **337** (1992) 361–370.
15. G. D. Sullivan, "Model-based vision for traffic scenes using the ground plane constraint," *Real-Time Computer Vision*, eds. D. Terzopoulos and C. Brown, CUP, Cambridge, 1994.
16. C. S. Wallace and P. R. Freeman, "Estimation and inference by compact coding," *J. R. Stat. Soc. Series B* **49** (1987) 240–265.
17. S. Wolfram, *The Mathematica Book*, 3rd edition, Cambridge University Press, Cambridge, 1996.
18. A. D. Worrall, G. D. Sullivan and K. D. Baker, "A simple intuitive camera calibration tool for natural images," *British Machine Vision Conf.* **2**, ed. E. Hancock, 1994, pp. 781–790.
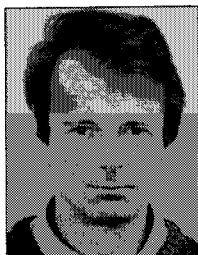
Steve Maybank obtained a first degree in mathematics from King's College Cambridge and a doctorate in computer science from Birkbeck College, London. He is currently a Reader in computer vision, with research interests in visual surveillance, people modeling and tracking, 3D scene reconstruction and statistical methods for inferring models from data. He has published over 65 articles and one book.

Roberto Fraile received his licentiate degree in algebra and geometry from University of Valladolid, Spain, in 1996. He is currently a Ph.D. student at University of Reading, England.

His current research interests include three dimensional computer vision and remote sensing, and applications of algorithmic complexity theory.

This page is intentionally left blank

AN INTEGRATED VISION SYSTEM FOR ALV NAVIGATION

YE XIUQING*, LIU JILIN and GU WEIKANG

Department of Information Science and Electronic Engineering
Zhejiang University Hangzhou 310027, P.R. China
* E-mail: yexq@isee.zju.edu.cn

In this paper an integrated vision system for autonomous land vehicle is described. The vision system includes 2D and 3D vision modules and information fusion module. The task of 2D vision is to provide the physical and geometry information of road, and the task of 3D vision system is to detect the obstacles in the surrounding. Fusion module combines 2D and 3D information to generate a feasible region provided for vehicle navigation.

Keywords: ALV; obstacle; fusion; road segmentation; vision system.

1. INTRODUCTION

Autonomous vehicle is an active field of research. Several prototypes have been developed.[2,3] In 1990s, intelligent vehicles became an important part of research in Intelligent Transportation System and Intelligent Vehicle Highway System in some countries. Vision system plays an important role for an intelligent vehicle, but using vision to guide an autonomously driving vehicle is still a challenging subject. Many researchers are working on relative subjects.[1,4]

A powerful robot vision system, together with fast image processing algorithm, which enables an autonomous land vehicle (ALV) to act in the real-world environment, is a project of research in our group.

For road following, a vision system must provide the surrounding information of a vehicle for the planning system. This means the perception must inform where is the road region and where is the obstacle. We have designed and developed a vision system for this purpose. This system includes 2D vision, 3D vision and information integration. The sensor of 2D vision is a color camera, and the sensor of 3D vision is an active laser radar. 2D vision is used for finding the road region in the color image, and 3D vision is used for obstacle detection. Fusion module integrates 2D information and 3D information to create a complete environment description. We designed and built two kinds of hardware equipment for vision system operating at near real time, one is a high-speed image processing system and the other is a pipeline equipment based on Transputer with high-speed data channel.

2. 2D VISION

In a real road environment, the possible road for a vehicle may be divided into two broad categories: structural road with road markers and unstructured road without special markers. We propose two kinds of algorithm for the two-catalogue road

71

situations, one is region and edge information integrated algorithm for unstructured road and the other is white line following algorithm for structured road.

2.1. Region and Edge Information Combination Algorithm

Due to the complexity of outdoor environment, segmentation following in ALV is made very difficult. Segmentation methods can be divided into three types. One is based on feature space classification, the second is region based method which includes region growing and split-merge methods, the third is based on edge methods. There are global and local conflicts in image segmentation. According to the analysis of small neighborhood, the region consistency can be determined. The confidence of region consistency will be higher when larger neighborhood is used. But using over-large neighbor will increase the possibility of data inconsistency. To avoid this problem the local edge information and global region consistency information are combined.

In consideration of real-time operating requirement, we do not use time-consuming complex algorithms. For the robustness of the algorithm, the region segmentation and edge detection information are combined to reach the segmentation purpose.

Since a color image has richer information than intensity image, most of the outdoor robots use color images instead of monochrome, but it is time consuming to use R-G-B color image directly and a large amount of redundant information exists among the R-G-B color space. We use classical Fisher's linear discriminate technique to calculate the optimal projection transform $G(i, j) = w_1 R(i, j) + w_2 G(i, j) + w_3 B(i, j)$, where $W = (w_1 w_2 w_3)^t = S_w^{-1}(m_1 - m_2)$, and calculate the optimal transform value for some typical road surface and set a look-up table for high-speed selection.

Among the variety of image segmentation techniques, the threshold method is popular. This is actually a two-class classifier with a single feature. There are lots of threshold selection algorithms in literature. But for our purpose, it is best to use road knowledge. We assume that the area in front of the vehicle is always the road area. A trapezoid window on the bottom of an image is selected and the feature value of the window is calculated. This value is used to guide threshold selection. The histogram of $G(i, j)$ is smoothed with Gaussian filter, then the convex hull of the histogram is constructed. The concavities are found by taking the difference between the convex hull and the histogram. The maximum of the concavities is used as candidate threshold site. Then the feature value guides the optimal threshold selection. Double thresholds are used in the algorithm. Due to the complexity of road environment, the global threshold method is not appropriate to some situations, regions may appear under or over segmented. So the edge detection information is combined to modify the segmentation results.

A Canny-like edge detector, which includes four directional (vertical, horizontal, 45°, 135°) second-order derivative mask, is used in our algorithm. Some post-processing procedures are processed after edge operating. Edge tracking is used to connect the edge points, and the short edge lines are eliminated. The remaining edge

lines are used to search and compare with the boundary determined by threshold based segmentation method. On the basis of neighborhood and direction property, most of the edges which are not road edges can be discarded.

In the two-class problem, two kinds of mistakes may appear. One, the regions which actually belong to road regions be assigned to nonroad regions. Another is that the regions belonging to nonroad regions be assigned to road regions. According to some rules, for example, if an edge point detected by the edge detector falls into the road region, then set the value of that point on the region as zero. After verifying all edge points, morphology and connectedness operators are processed. Another example is that if road regions lie on the same side of edge line then the statistical value calculated and the region growth proceed. Through some combination rules, the satisfied segmentation results can be obtained. We tested the algorithm in summer and autumn from morning to late afternoon on a campus road. Figures 1 and 2 show the processed results.
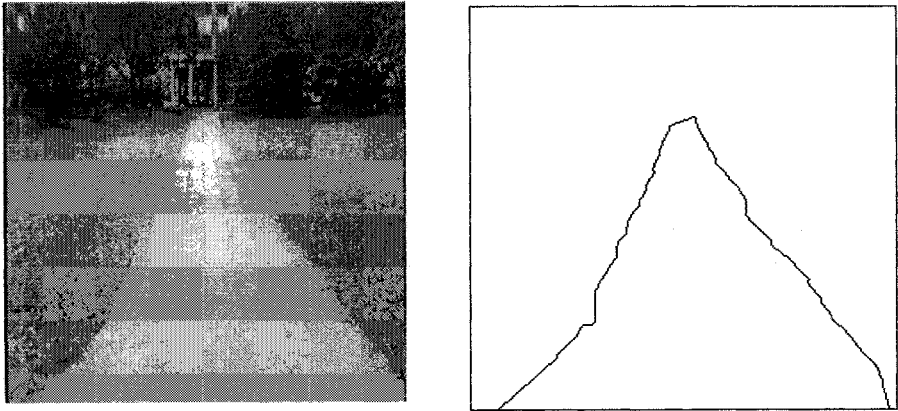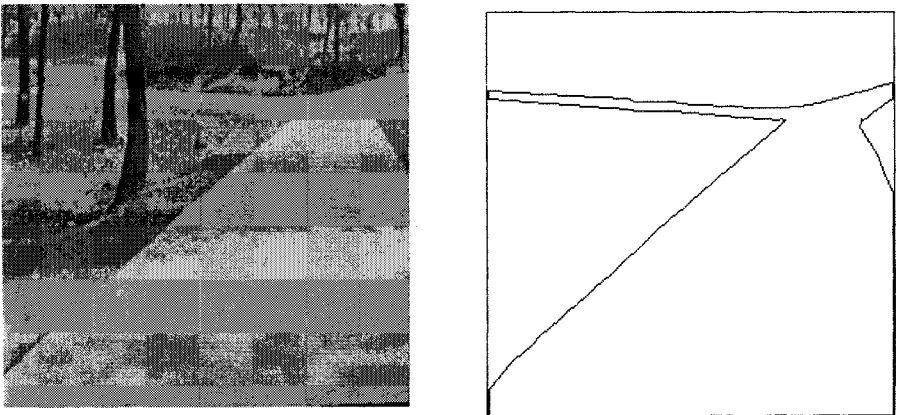

Fig. 1.   Nonuniform light condition.

Fig. 2.   Road with shadow.

2.2. White Line Following Algorithm

For white line following, there are two important tasks: one is to search for starting points of two white lines (a left one and a right one), the other is line following from the just found starting point. In the algorithm, white line starting points are searched by a local thresholding and hypothesis verification method, and the white line following is based on window searching scheme.

The changing conditions of light, shadows on ground, and the interference from nearby objects often make it difficult to find out the starting points correctly. We enhance the performance of the searching algorithm by three steps: enlarge searching area, use as many features of the white line as possible, feedback line following results to starting point searching and remove illegal starting points. The left line starting point is searched in the left part and right line starting point in the right part. The searching begins at the bottom row of the image and goes upward row by row. First, change the gray level image to a binary image (with only "1" or "0") according to:

$$B[i,j] = \begin{cases} 1 & \text{if}(I[i,j] > \text{ave}[i] + \theta_1) \wedge (I[i,j] > \theta_2) \wedge (I[i,j] > \max[i] - \theta_3) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $I[i,j]$ represents the gray level value of a pixel in the image, θ_1 and θ_3 are two thresholds, which are set according to road model, θ_2 is the average gray level of the road surface calculated from the previous image; $\max[i]$ and $\text{ave}[i]$ are maximum and average values of intensity in a row. There are several line segments with gray level "1" in each row of binary image. Find out the line segments that are comparable in width with known white lines and give a mark for each of them. Each line segment marked is a candidate for the starting point, then a hypothesis verification procedure is processed. Assume a candidate P as shown in Fig. 3, make the rays starting from P with slope ϕ as R_ϕ, in practice we take $\phi = 0°$, 11.25°, 22.5°, 33.75°, 45°, 56.25°, 67.5°, 78.75°, 90°. For each ray, count the number of pixels with value
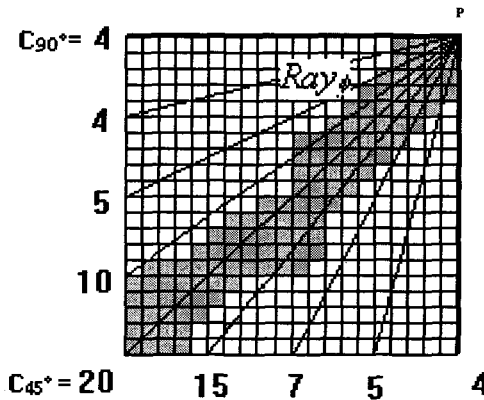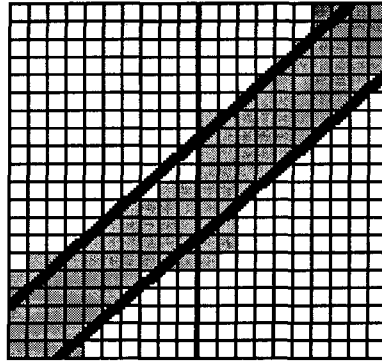

Fig. 3.   White line starting point searching.

Fig. 4.   Two discontinuity edge of white line in a window.

1 on the ray in a window and pick up the maximum count number C. If $C \geq T$, accept P as a white line starting point. If not, continue the searching process.

Window following follows a white line in the image in both upward and downward directions from the starting point in a window with size 20 × 20. If there is a white line segment in a window, there will be a gray level discontinuity edge from the background to white line and a gray level discontinuity edge from white line to the background as shown in Fig. 4. We use multidirectional local projection to find out the two gray level discontinuity edges and calculate the projection of sixteen directions in a window. In each direction, there will be a rising intensity and a dropping intensity. The maximum of rising intensity corresponds to the discontinuity edge from the background to white line and the maximum of dropping intensity corresponds to the discontinuity edge from white line to the background. When the position, orientation and intensity of the two discontinuity edges have been calculated, we can determine whether there is a white line segment in the window. If there is, one then forecast the position of the next window on the basis of the position and orientation of the white line calculated in the current window and repeat the steps until the line following has been finished according to some rules. Actually when the vehicle makes a turn or avoids an obstacle, we can only see one white line in an image sometimes. For deciding that a white line is the left edge line or the right one, the knowledge of processed results of the previous images is used to easily handle the problem. Figure 5 is the experiment result.

3. 3D VISION

The purpose of 3D vision here is to detect obstacles on the road. A laser range sensor (LRS) made by our university is used as a sensor device in 3D vision. LRS is amplitude-modulated continuous-wave laser radar that senses the range to the target by measuring the shift in phase between an emitted beam and its echo. Every half-second, LRS gives an image, which contains 64 row by 256 column of range value corresponding to 30° by 80° field of view. For the outdoor scene, usually there are two methods for obstacle detection, one is surface slope algorithm;
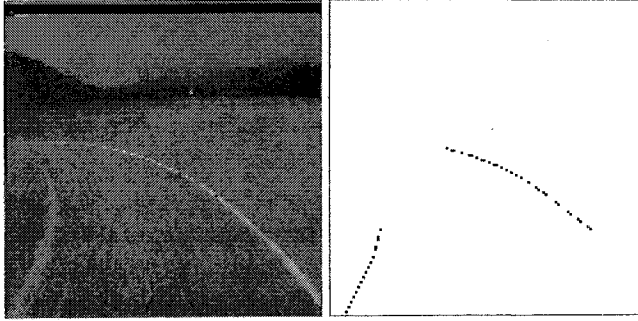
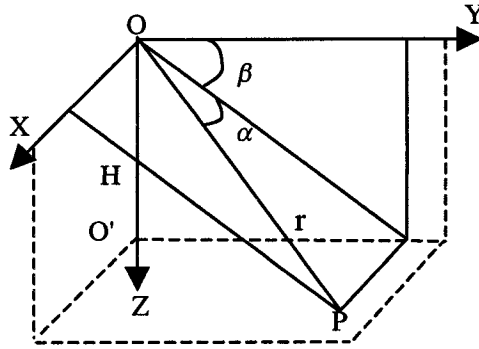Fig. 5.   White line following result.

Fig. 6.   Range coordinate system.

the other is elevation difference algorithm. We use the latter which is sufficient for
detection purpose. The idea of elevation difference segmentation is to transfer the
original range image from spherical coordination of sensor to Cartesian coordina-
tion of ground to get the elevation map, then to process the elevation map to get
the obstacle region. Figure 6 shows the range image coordinate system. Points in
the range image are described by the spherical coordinates (α, β, r), where r is the
range value of a pixel representing the distance from sensor to the object, α and β
are scanning angles in horizontal and vertical directions, respectively. Range sensor
is mounted on the ALV above the origin of a local vehicle coordinate $(0, 0, H)$. Local
vehicle coordinate system (X_V, Y_V, Z_V) is a moving Cartesian coordinate system,
where the positive direction of axis of Y_V is the same as the vehicle traveling di-
rection, Z_V is the axis directly upward. From the above geometry relation, we can
transfer points from sensor coordinate system to local vehicle coordinate system by
the following equations:

$$\begin{cases} X_V = r \sin \alpha \\ Y_V = r \cos \alpha \cos \beta \\ Z_V = H - r \cos \alpha \sin \beta \end{cases} \tag{2}$$

In consideration of the movements and posture changing of the vehicle, we must complete transformation from local vehicle coordinate to world coordinate (X_W, Y_W, Z_W) according to parameters given by the inertial navigation system. The inertial navigation system delivers six parameters for every line of range data. Three of the parameters represent absolute orientation with respect to inertial frame and the other three represent instantaneous acceleration with respect to the inertial reference frame. Assume that the world coordinate coincides with the initial reference frame, then point $\mathbf{P}_V$ is transformed from local vehicle coordinate to the world coordinate by following equation:

$$\mathbf{P}_W(t) = \mathbf{S}_W(t) + \mathbf{R}\mathbf{P}_V(t) \tag{3}$$

where $\mathbf{S}_W(t)$ is the position of ALV in world coordination at time t, $\mathbf{R}$ is transformation matrix of three posture angles. After all of the above transformation, obstacle position can be obtained by thresholding the tranformed image in which the intensity of each point represents the height of that point above the ground plane. But by using global threshold wrong information of the obstacle position is possible. So a local threshold row by row is used to segment the transformed image. Figure 7 shows the original range image and transformed image. In order to eliminate the noise on the range image, a 3*3 media filter is used for noise smoothing before transformation.

4. FUSING OF 2D AND 3D INFORMATION

2D vision and 3D vision have their special function; each of them can only give part of the information about road scenes. For obtaining the complete environment description it is necessary to fuse 2D and 3D information.
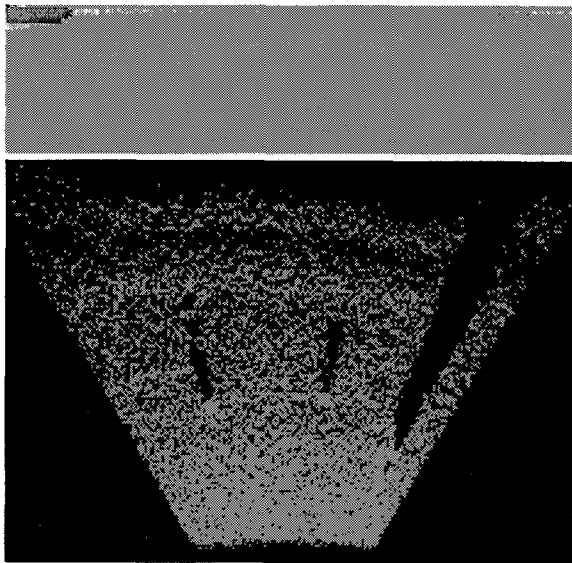

Fig. 7. Range image and transformed image.

As an important theory of uncertainty inference, Dempster–Shafer theory has extensive applications in multisensors fusion. In this theory the probability is assigned to proposition (i.e. to subsets of a frame of discrimination). A basic probability assignment is a function $m : 2^{\Theta} \to [0,1]$ where $m(\phi) = 0$, $\sum_{A \subseteq \Theta} m(A) = 1$. Dempster's rule of combining states that two bpa's m_1 and m_2, corresponding to two independent sources of evidence, can be combined to yield a new bpa $m : m_1 \oplus m_2(c) = K^{-1} \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)$, where $k = 1 - \sum_{A_i \cap B_j = \Phi} m_1(A_i) m_2(B_j)$.

Dempster's rule of combination is an important tool for belief value updating. However the condition required by Dempster's rule is that the bodies of evidence to be combined must be independent. This constraint condition may not be satisfied for some applications. And the form of normalization used in the rule leads to some counterintuitive results in some situations. To deal with the above problems, Wu and Yang et al.[5] proposed a generalized evidence combination formula relaxing the requirement of evidence independence, and Zhang and Gu[6] presented a modification of Dempster's rule of combination in our sensors fusion application. The procedure of fusing 2D and 3D information includes:

- Sensors calibration.
- Define the frame of discernment $\Theta = \{F, O\}$, F represents free region on the road and O represents obstacle region.
- Select evidence from features of 2D and 3D.
- Calculate each evidence bpa and calculate belief measure bel and plausibility measure pls of a proposition.
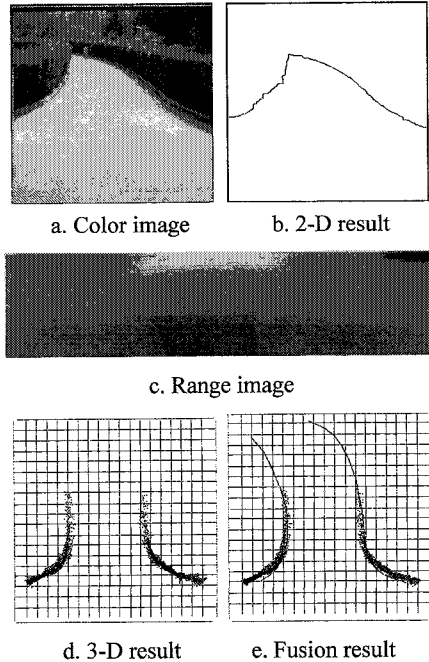

a. Color image   b. 2-D result

c. Range image

d. 3-D result   e. Fusion result

Fig. 8.   The result of 2D and 3D fusion.

- According to modification formula of D–S combination rule calculate combined bpa of all evidence.
- According to some decision rules, select possible hypothesis, which has maximum support provided by joint evidence.

Figure 8 shows the result of 2D and 3D information fusion.

5. HARDWARE ARCHITECTURE

Figure 9 is an overview of our image processing system, which provides 256×256 red, green and blue images with 8 bits of intensity per image. The field of view and focus of the camera are kept fixed. The system contains several special purpose processors (SP in Fig. 9), each of them performs a standard image processing operation (such as look-up table, convolution, morphology, connecting components labeling) with video frame rate (1/25 second). The system also contains a general purpose processing unit (GP in Fig. 9) based on TMS320C30 microprocessor. It can perform some algorithms at high speed. Each processor communicates with each other through an image data bus and also communicates with the host computer through a PC bus. The processors can be programmed by the host computer and can be organized as a pipeline structure suitable for some image processing and understanding tasks. For ALV road following, the system is available for both unstructured and structured roads.
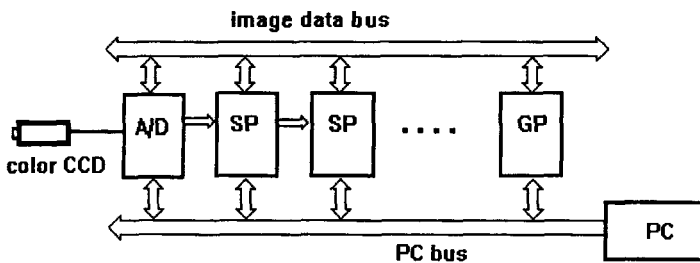

Fig. 9.   2D hardware architecture.

Figure 10 shows the 3D vision hardware architecture. Operation of 3D vision includes inertial navigation parameter acquirement, range data acquirement, correction, coordinate transformation, data modification according to the change of moving vehicle posture, data merge and data transmission. The most significant feature is that data throughput is very large in the operating procedure. In order to satisfy the above requirements the hardware of 3D vision was designed. That is based on Transputer pipeline architecture, which includes five Transputer with high-speed data channel and input module. One PC is used as host. Input module receives range data at rate 2 frame/s. Each frame includes 256×64 pixels with 8 bits/pixel. High-speed data channel can provide 33 Mbytes/s throughput. An important point in this system is that the high-speed data channel provides conditions for high-speed
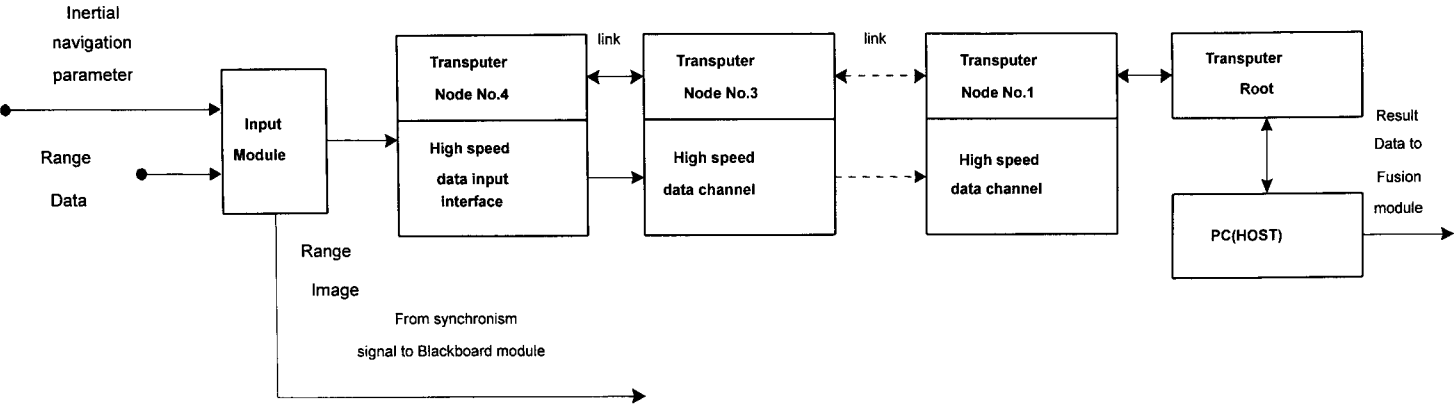
Fig. 10. 3D hardware architecture.

data exchanging among Transputers. A parallel implementing software has been developed for carrying out algorithms on multinode Transputer. The parallelism in vision processing can be divided into two types: pipelining and parallel decomposition (also called data decomposition). The pipelining operation can be represented as $Y = F(X) = F_n(F_{n-1}(F \ldots (F_2(F_1(X))) \ldots))$, and the parallel decomposition can be represented as $Y = F(X) = F(X_1) \parallel F(X_2)) \parallel \ldots \parallel F(X_{n-1}) \parallel F(X_n)$, where X is the input image or sub-region of image, Y the output image, F the processing operator, $\parallel$ the parallel operator. In the obstacle detection algorithm the parallel program is developed based on such parallelisms.

6. CONCLUSION

An integrated vision system for ALV is presented in this paper. The experimental results show that the vision system works well on the testing site. Now we are working on multiinformation combination on different processing level for adapting to complex road situations. We are also constructing an image processing system based on advanced DSP chip in order to increase program flexibility and to reduce the system's size.

REFERENCES

1. G. Foresti, V. Murno, C. S. Regazzoni and G. Vernazza, "A distributed approach to 3D road scene recognition," *IEEE Trans. Vehicular Technol.* **43**, 2 (1994) 389–406.
2. D. Kuan, G. Phipps and A. C. Hsueh, "Vision and navigation for Carnegie–Mellon Navlab," *IEEE Trans.* **PAMI-3** (1988) 362–373.
3. M. Maurer, R. Behringer and D. Dickmanns *et al.* "VaMoRs-P—An advanced platform for visual autonomous road vehicle guidance," *Proc. SPIE Conf. Mobile Robots*, IX, Vol. 2352, Boston, 1994, pp. 239–248.
4. K. Uchimura and Z. Hu, "Recognition of shape models for general roads," *Proc. Asian Conf. Computer Vision*, Vol. 1, Hong Kong, China, 1998, pp. 152–159.
5. Y. J. Wu, J. Y. Yang and K. Liu, "On the evidence inference theory," *Inform. Sci.* **89** (1996) 245–260.
6. Q. Zhang and W. K. Gu, "Combining camera and laser radar for ALV navigation," *Proc. Asian Conf. Computer Vision*, Vol. 1, Hong Kong, China, 1998, pp. 144–151.
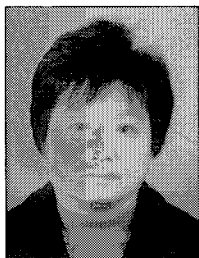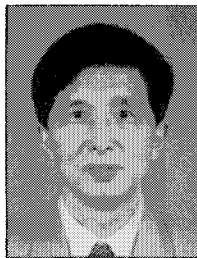
Ye Xiuqing is a Professor at the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China.

Her current research interests include machine vision, real-time image processing, intelligent transportation system and face processing.
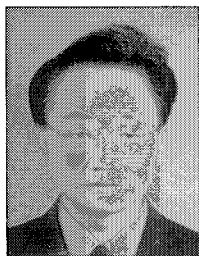
Gu Weikang is a Fellow of the Electronic Association of China and a Professor in the Electronics and Information Department, Zhejiang University, China.

His research areas are in computer vision, real-time image processing, machine vision and intelligent control.

Liu Jilian is a Professor at the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He is currently the Deputy Director of the Institute of Information and Communication Engineering. He is currently a member of the editorial board of *J. Data Acquisition and Processing* and he is a member of New York Academy of Sciences. He has been teaching graduate and undergraduate courses in digital signal processing and digital image processing since he became a professor in 1990.

One of his main research areas focuses on ITS (Intelligent Transportation system), especially on video navigation for intelligent vehicle. He has authored over 60 publications.

FUZZY BAYESIAN NETWORKS — A GENERAL FORMALISM FOR REPRESENTATION, INFERENCE AND LEARNING WITH HYBRID BAYESIAN NETWORKS

HEPING PAN

Digital Earth Research Centre
Wuhan Technical University of Surveying and Mapping
129 Luoyu Road, Wuhan 430079, P.R. China
E-mail: panhp@wtusm.edu.cn

LIN LIU

University of South Australia
Cooperative Research Centre for Sensor Signal and Information
SPRI Building, Technology Park Adelaide
Mawson Lakes, SA 5095, Australia
E-mail: lliu@cssip.edu.au

This paper proposes a general formalism for representation, inference and learning with general hybrid Bayesian networks in which continuous and discrete variables may appear anywhere in a directed acyclic graph. The formalism fuzzifies a hybrid Bayesian network into two alternative forms: the first form replaces each continuous variable in the given directed acyclic graph (DAG) by a partner discrete variable and adds a directed link from the partner discrete variable to the continuous one. The mapping between two variables is not crisp quantization but is approximated (fuzzified) by a conditional Gaussian (CG) distribution. The CG model is equivalent to a fuzzy set but no fuzzy logic formalism is employed. The conditional distribution of a discrete variable given its discrete parents is still assumed to be multinomial as in discrete Bayesian networks. The second form only replaces each continuous variable whose descendants include discrete variables by a partner discrete variable and adds a directed link from that partner discrete variable to the continuous one. The dependence between the partner discrete variable and the original continuous variable is approximated by a CG distribution, but the dependence between a continuous variable and its continuous and discrete parents is approximated by a conditional Gaussian regression (CGR) distribution. Obviously, the second form is a finer approximation, but restricted to CGR models, and requires more complicated inference and learning algorithms. This results in two general approximate representations of a general hybrid Bayesian networks, which are called here the *fuzzy Bayesian network (FBN)* form-I and form-II. For the two forms of FBN, general exact inference algorithms exists, which are extensions of the junction tree inference algorithm for discrete Bayesian networks. Learning fuzzy Bayesian networks from data is different from learning purely discrete Bayesian networks because not only all the newly converted discrete variables are latent in the data, but also the number of discrete states for each of these variables and the CG or CGR distribution of each continuous variable given its partner discrete parents or both continuous and discrete parents have to be determined.

Keywords: Hybrid Bayesian networks; Gaussian mixture models; fuzzy logic; inference; learning.

1. INTRODUCTION

Multispectral image processing has entered the era of multisensor data fusion. Bayesian networks also called causal probabilistic networks, belief networks, or influence diagrams, provide a high-level generic architecture for fusing sensory observations from multiple sensors and nonsensory data from multiple data sources. From the view point of probability theory, Bayesian networks are computational architectures that guarantee the consistency and coherence of a causal probabilistic model of a given problem domain and maintain the probabilistic equilibrium of the model upon arrival of new information. Inference using Bayesian networks is very flexible, in the sense that new information can be input into any section of the model and propagated throughout the rest of the network. There is no need to distinguish between forward chaining or backward chaining as referred to in logical production systems. Bayesian networks are particularly relevant to multisensor data fusion which normally involve disparate and noncommensurate random variables corresponding to disparate information sources.

Discrete Bayesian networks provide a general formalism for representing a joint probability distribution of multiple discrete random variables. Exact general-purpose inference algorithms with discrete Bayesian networks exist and are well developed, such as the well-known junction tree inference algorithm.[8,15] Learning discrete Bayesian networks from complete data or incomplete and soft data is generally a NP-hard problem and still has to be considered unsolved yet, but a number of suboptimal solution components[2,11,23−25] have been developed, which may be integrated to provide an operational solution.[21]

As a big contrast with discrete Bayesian networks, general hybrid Bayesian networks in which continuous and discrete variables may appear anywhere in a directed acyclic graph (DAG) have not found a general solution for representation, inference and learning. However, hybrid Bayesian networks are the most general form of Bayesian networks encountered in the vast variety of practical applications. A general operational solution for representation, inference and learning with general hybrid Bayesian networks is in great demand by real-world modeling practitioners.

In the literature, some constrained classes of hybrid Bayesian networks have been investigated. One of these is the class of networks with *Conditional Gaussian* (*CG*) density functions.[12,14,16] In these networks, the conditional distribution of the quantitative continuous variables given the qualitative discrete variables is assumed to be multivariate Gaussian. The CG model can be extended to allow continuous variable parents for continuous variables, and becomes the CG regression model. Exact probabilistic inference with the CG model or CG regression model or their close relatives[4,16] is possible and turns out to be polynomial in the number of continuous variables. Parametric learning with CG regression models from incomplete data is shown[19] to be possible using EM algorithm. However, there is a strict limitation that continuous variables cannot be parents of discrete variables. This asymmetry between the continuous and discrete variables obstruct the applications of hybrid networks in general practice. A further extension is to approximate an arbitrary conditional probability distribution by a sum of weighted CG (SWCG)

distribution.[1,3] However, the dark side of this approach is that there is no general efficient calculus for division and multiplication of SWCG models[18] which is required by the message passing in the junction tree inference algorithm. In fact, the number of terms of such a SWCG in the initial approximation of a continuous probability distribution can be large. The computational complexity is even worse when message passing is considered: the multiplication of SWCG makes the number of terms in the product grow exponentially with the path length, and the worst is the division of two SWCG models for which there is simply no closed-form solution yet. All these difficulties make the probabilistic inference computationally intractable.

Therefore, we must resort to approximate representations which are computable. Recently, Kozlov and Koller[10] proposed an approach of nonuniform dynamic discretization for hybrid networks. Rather than discretizing each variable separately, they discretized a continuous function on its entire multidimensional domain at once by a nonuniform adaptive discretization using a binary split partition tree. Like most discretization approaches used by probabilists, their discretization is defined to produce mutually exclusive quantizations. Pan and McMichael[22] proposed an alternative approach to crisp exclusive quantization using fuzzification in which each discrete state of a newly converted discrete variable corresponds to a fuzzy set of the original continuous variable. Defuzzification from the discrete states to a crisp continuous value can also be done if required. However, no explicit inference and learning schemes were given in Ref. 22.

This paper takes a large step further from that fuzzification and proposes the precise formalism of fuzzy Bayesian networks in two alternative forms for representation, inference and learning with general hybrid Bayesian networks. In both forms, instead of using fuzzy sets for mapping between continuous variables and their partner discrete variables, we use Conditional Gaussian (CG) models. The first form replaces each continuous variable in the given directed acyclic graph (DAG) by a partner discrete variable and adds a directed link from the partner discrete variable to the continuous one. The mapping between two variables is not crisp quantization but is approximated (fuzzified) by a conditional Gaussian (CG) distribution. The CG model is equivalent to a fuzzy set but no fuzzy logic formalism is employed. The conditional distribution of a discrete variable given its discrete parents is still assumed to be multinomial as in discrete Bayesian networks. The second form only replaces each continuous variable whose descendants include discrete variables by a partner discrete variable and adds a directed link from that partner discrete variable to the continuous one. The dependence between the partner discrete variable and the original continuous variable is approximated by a CG distribution, but the dependence between a continuous variable and its continuous and discrete parents is approximated by a conditional Gaussian regression (CGR) distribution. Obviously, the second form is a finer approximation, but restricted to CGR models, and requires more complicated inference and learning algorithms. This results in two general approximate representations of a general hybrid Bayesian networks, which are called here the *fuzzy Bayesian network (FBN)* form-I and form-II. For the two forms of FBN, general inference algorithms exists, exact for the form-I and

approximate for the form-II, which are extensions of the junction tree inference algorithm for discrete Bayesian networks.

Learning fuzzy Bayesian networks from data is different from learning pure discrete Bayesian networks because not only all the newly converted discrete variables are latent in the data, but also the number of discrete states for each of these variables and the CG or CGR distribution of each continuous variable given its partner discrete parents or both continuous and discrete parents have to be determined. However, based on the previous work on learning discrete Bayesian networks, it is possible to extend an integrated approach for discrete network learning[21] to hybrid network learning using the formalism of fuzzy Bayesian networks. In this paper, we will focus on the representation forms and inference algorithms of the proposed fuzzy Bayesian networks. Some thoughts on learning fuzzy Bayesian networks from data are also presented.

2. FUZZY BAYESIAN NETWORKS AS APPROXIMATE REPRESENTATION OF HYBRID BAYESIAN NETWORKS

First of all, we like to point out that fuzziness indeed exists not only in human cognitive processes and communication languages, but also in the peripherals of discrete Bayesian networks. It makes the point more obvious if we take a look at an example Bayesian network for automobile diagnostics which is popularized by the Hugin system. Figure 1 shows a simple discrete Bayesian network for the problem of automobile diagnostics. In this network, although all the variables are defined to be discrete, a subset of the variables is genuinely continuous, as shown in Fig. 2.
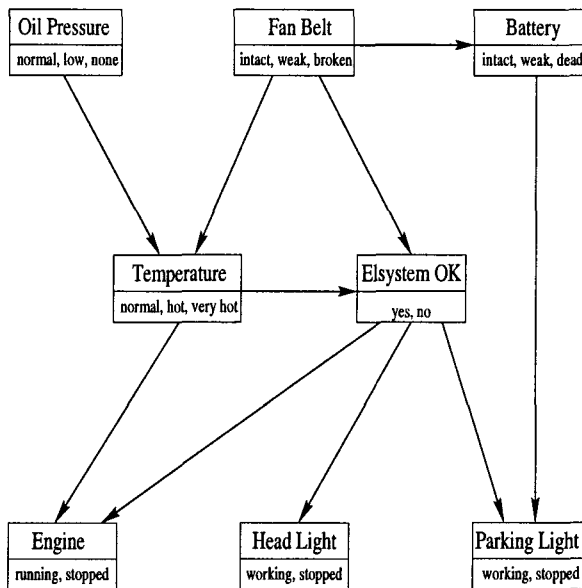

Fig. 1.   An example of discrete Bayesian networks: each variable is discrete, with their discrete states shown under its name.
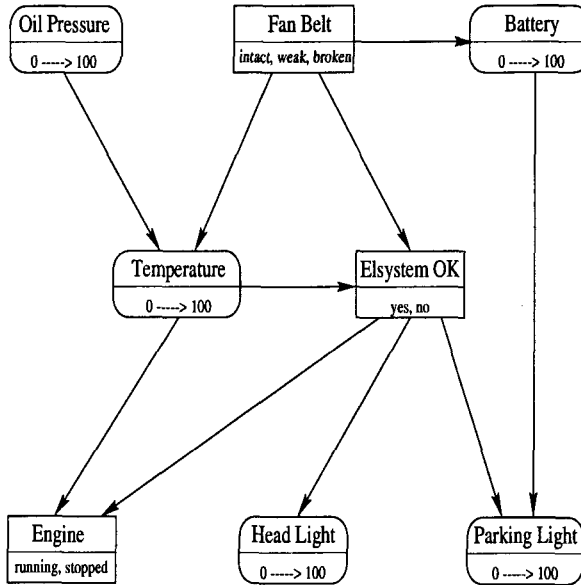
Fig. 2. Some variables shown in rounded boxes should be genuinely continuous, their frames are normalized to a real line segment [0,100].

For a discrete variable which is genuinely continuous, a mapping is implicitly involved whenever an originally continuous observation arrives. Such an observation is normally called *evidence* in the Bayesian network terminology. According to the axioms of the probability theory, the discrete states of an originally continuous variable must be mutually exclusive and collectively exhaustive. Consequently, the mapping from the continuous value domain to the discrete state frame is realized through a quantization. However, not every continuous variable can be sensibly quantized, as the cut between two neighboring discrete states is often hard to define. In reality, we often have no sound reason to set a clear-cut boundary between neighboring states if its discrete states are commensurate. For example, the variable "Temperature" is inherently continuous, and it is discretized to three discrete states: "normal", "hot", and "very hot". However, there is no meaningful way to set the boundary between "normal" and "hot", and between "hot" and "very hot".

We recognize that fuzziness is a sort of vagueness and uncertainty which is indeed different from the axiomatic definition of probability. Briefly, probability is a measure of the undecidability in the outcome of clearly defined and randomly occurring events, while fuzziness is concerned with the ambiguity or undecidability inherent in the description of the event itself. Nevertheless, this does not necessarily rule out the possibility of using probabilistic calculus for handling fuzziness. It does remind us that we as probabilists should at least face the real-world complexity with an open mind. On the other hand, we also consider fuzzy logic as an approximate reasoning formalism which may be easy to use and possibly sufficient in many ordinary applications. It is quite possible that fuzzy Bayesian networks as proposed

here may realize anything fuzzy logic can do and will inherit all the rigor, flexibility and other superior properties of probabilistic approaches.

Let us now face the reality of general hybrid Bayesian networks with the following definition.

Definition 1. *Hybrid Bayesian Networks*

A general hybrid Bayesian network (HBN) is a directed acyclic graph (DAG) representing the joint probability distribution of a given set of variables $\mathbf{V}$, including discrete variables $\mathbf{X}$ and continuous ones $\tilde{\mathbf{Y}}$

$$HBN = (\mathbf{V}, \mathbf{L}, \mathbf{P}) = (\mathbf{X}, \tilde{\mathbf{Y}}, \mathbf{L}, \mathbf{P}) \tag{1}$$

where $\mathbf{X} \subseteq \mathbf{V}$ denotes a set of discrete variables, $\tilde{\mathbf{Y}} \subseteq \mathbf{V}$ a set of continuous variables, $\mathbf{L}$ a set of directed links between variables, and $\mathbf{P}$ a set of conditional probability distributions

$$\mathbf{L} \subset \mathbf{V} \times \mathbf{V} \tag{2}$$

$$P = \{P(V|\Gamma_V^+, \Psi_V^+); \quad V \in \mathbf{V}\} \tag{3}$$

where the notation is defined as follows: we use upper-case letters such as V for a variable name, and corresponding lower-case letters such as v for a value of a variable (of V). We use bold-face letters such as $\mathbf{V}$ for sets and $|\mathbf{V}|$ for the cardinality of $\mathbf{V}$. All the continuous variables are denoted by capping with $\tilde{a}$ such as $\tilde{Y}$. If a continuous variable $\tilde{Y}$ is discretized, the corresponding discrete variable is denoted by the original letter Y with the cap removed. Γ_V^+ and Ψ_V^+ denote the set of discrete and continuous parents respectively for a variable V. Configurations of Γ_V^+ and Ψ_V^+ are denoted by γ_V^+ and ψ_V^+ respectively. The set of discrete and continuous children of V are denoted by Γ_V^- and Ψ_V^- respectively. A configuration of Γ_V^- and Ψ_V^- are denoted by γ_V^- and ψ_V^- respectively. Note also that V can be either discrete or continuous. Figure 3 shows a *HBN* abstracted from the example of Fig. 1.
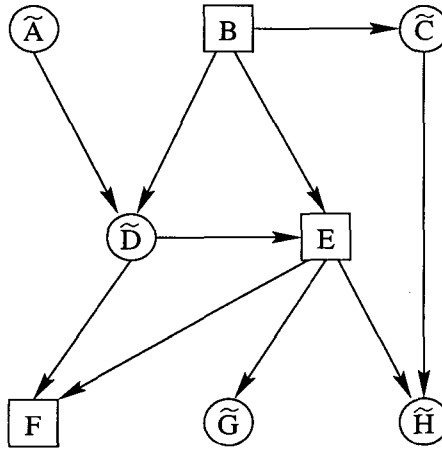

Fig. 3.   A hybrid Bayesian network abstracted from the example of Fig. 2.

Note that in the above definition of HBN, we have no restriction on the dependency topology among all the variables. A continuous or discrete variable can have continuous and/or discrete variables as parents or children. We believe that this ultimate generality of HBN's is most demanded in practical applications and it will give nontechnical users a peaceful mind even though the fuzzy Bayesian networks as defined below are approximate representations. However, approximation is absolutely necessary because we have no general representation of arbitrary probability distributions of hybrid variables: $P(V|\Gamma_V^+, \Psi_V^+)$. Fortunately, once a fuzzy Bayesian network is constructed by transforming a given HBN, the inference algorithm can be exact (as shown later).

The first form of the fuzzy Bayesian network for a given hybrid network HBN is defined as follows:

Definition 2. *Fuzzy Bayesian Network Form-I*

Given a hybrid Bayesian network HBN as defined in the form of (1)–(3), the first form of its corresponding fuzzy Bayesian network FBN is defined by

$$FBN^1 = (\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{L}^1, \mathbf{P}^1) \tag{4}$$

where $\mathbf{Y}$ is the set of discrete variables defined by discretizing $\tilde{\mathbf{Y}}$, i.e.

$$Y \in \mathbf{Y} \leftrightarrow \tilde{Y} \in \tilde{\mathbf{Y}} \tag{5}$$

we call Y the fictitious discrete variable or fuzzy variable of the continuous variable Y. $\mathbf{L}^1$ is the set of directed links for FBN^1 obtained by the fuzzification transformation of HBN

$$\forall \tilde{Y} \in \tilde{\mathbf{Y}}, \text{ replace } \tilde{Y} \text{ in } HBN \text{ by } Y \text{ and create a new directed link } Y \to \tilde{Y} \tag{6}$$

$\mathbf{P}^1$ is the set of conditional probability distributions

$$\mathbf{P}^1 = \mathbf{P_X} \cup \mathbf{P_Y} \cup \mathbf{P_{\tilde{Y}}} \tag{7}$$

where

$$\mathbf{P_X} = \{P(X|\Gamma_X^+), \quad X \in \mathbf{X}\} \tag{8}$$

$$\mathbf{P_Y} = \{P(Y|\Gamma_Y^+), \quad Y \in \mathbf{Y}\} \tag{9}$$

$$\mathbf{P_{\tilde{Y}}} = \{P(\tilde{Y}|Y), \quad \tilde{Y} \in \tilde{\mathbf{Y}}\}. \tag{10}$$

Note that for each originally discrete variable $X \in \mathbf{X}$, all its parents are now discrete and still denoted by Γ_X^+, but this Γ_X^+ contains the original Γ_X^+ in the HBN and the fictitious discrete variables of X's continuous parents Ψ_X^+ in the HBN. Similarly for the parents Γ_Y^+ of each fictitious discrete variable Y converted from $\tilde{Y}$. However, each continuous variable $\tilde{Y}$ now has only one discrete parent Y and no continuous parent at all. With this transformation, the conditional probability distribution for each discrete variable X and each fictitious discrete variable Y can be assumed to be multinomial which is general for discrete variables while the distribution for each continuous variable $\tilde{Y}$ can be approximated by a conditional Gaussian (CG)

$$P(\tilde{Y} = \tilde{y}|Y = y) = \frac{1}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y} - \mu_y)^2}{2\sigma_y}\right\} \tag{11}$$

where y denotes a discrete state of Y and $\tilde{y}$ denotes any continuous value of $\tilde{Y}$. Note that there is no one-to-one correspondence between y and $\tilde{y}$. However, for each state y, there is a parameter vector $\theta_y = (\mu_y, \sigma_y)$. It is equivalent to say that the original conditional distribution $P(\tilde{Y}|\Gamma_{\tilde{Y}}^+, \Psi_{\tilde{Y}}^+)$ of a continuous variable $\tilde{Y}$ given its original discrete parents $\Gamma_{\tilde{Y}}^+$ and its original continuous parents $\Psi_{\tilde{Y}}^+$ in *HBN* is approximated by a conditional Gaussian mixture (CGM) model

$$p(\tilde{y}|\gamma_{\tilde{Y}}^+, \psi_{\tilde{Y}}^+) = \sum_y \frac{P(y|\gamma_Y^+)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y} - \mu_y)^2}{2\sigma_y}\right\}. \tag{12}$$

The marginal probability of the continuous variable $\tilde{Y}$ is also represented by a Gaussian mixture (GM) model

$$p(\tilde{y}) = \sum_y \frac{P(y)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y} - \mu_y)^2}{2\sigma_y}\right\}. \tag{13}$$

It is commonly accepted[18] that Gaussian mixture models are general approximations to arbitrary probability distributions. Therefore, we can say that the first form of the fuzzy Bayesian network FBN^1 defined by (4)–(11) provides a general approximate representation to a general hybrid Bayesian network HBN defined by (1)–(3). The accuracy of this approximation relative to the original HBN can be made sufficiently high if a sufficient number of discrete states $\{y\}$ for each fictitious discrete variable Y is used. Figure 4 shows the first form of the fuzzy Bayesian network for the hybrid Bayesian network of Fig. 3.
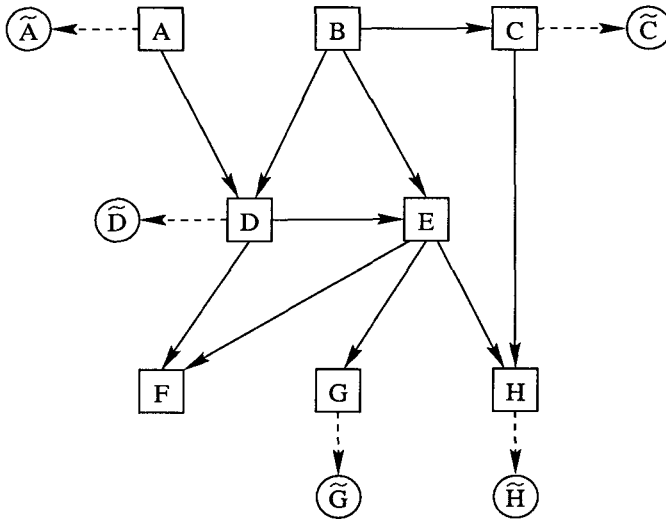

Fig. 4.   Form-I of fuzzy Bayesian network for the hybrid Bayesian network.

The second form of the fuzzy Bayesian networks differentiates those continuous variables whose descendants do not include discrete variables from other continuous variables:

Definition 3. *Fuzzy Bayesian Network Form-II*

The second form of the fuzzy Bayesian network FBN for a given hybrid Bayesian network HBN is obtained by applying the fuzzification transformation as defined by (6) only to those continuous variables $\tilde{\mathbf{Z}} \subseteq \tilde{\mathbf{Y}}$ whose descendants in HBN include discrete variables. This form can be expressed as

$$FBN^2 = (\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{Z}}, \tilde{\mathbf{W}}, \mathbf{L}^2, \mathbf{P}^2) \tag{14}$$

where Z is the fictitious discrete variable of the subset of continuous variables $\tilde{\mathbf{Z}} \subseteq \tilde{\mathbf{Y}}$, and $\mathbf{W}$ is the remaining other continuous variable

$$\tilde{\mathbf{W}} = \tilde{\mathbf{Y}}/\tilde{\mathbf{Z}} \tag{15}$$

whose descendants do not include discrete variables. $\mathbf{L}^2$ denotes the set of directed links whose contents should be clearly understood from the above definition: $\mathbf{X}$ and $\mathbf{Z}$ are discrete variables and they do not have continuous parents, and $\tilde{\mathbf{Z}}$ and $\tilde{\mathbf{W}}$ are still continuous variables; but each $Z \in \tilde{\mathbf{Z}}$ has only one discrete parent Z and is a leaf node, and each $W \in \tilde{\mathbf{W}}$ still keeps its original discrete parents $\Gamma_{\tilde{W}}^+$ and continuous parents $\Psi_{\tilde{W}}^+$ and may have continuous descendants. $\mathbf{P}^2$ denotes the set of conditional probability distributions

$$\mathbf{P}^2 = \mathbf{P_X} \cup \mathbf{P_Z} \cup \mathbf{P_{\tilde{Z}}} \cup \mathbf{P_{\tilde{W}}} \tag{16}$$

where $\mathbf{P_X}$ is defined in the same way as (8), and

$$\mathbf{P_Z} = \{P(Z|\Gamma_Z^+), \quad Z \in \mathbf{Z}\} \tag{17}$$

$$\mathbf{P_{\tilde{Z}}} = \{P(\tilde{Z}|Z), \quad \tilde{Z} \in \tilde{\mathbf{Z}}\} \tag{18}$$

$$\mathbf{P_{\tilde{W}}} = \{P(\tilde{W}|\Gamma_{\tilde{W}}^+, \Psi_{\tilde{W}}^+), \quad \tilde{W} \in \tilde{\mathbf{W}}\}. \tag{19}$$

The conditional distribution $P(X|\Gamma_X^+)$ for $X \in \mathbf{X}$ and $P(Z|\Gamma_Z^+)$ for $Z \in \mathbf{Z}$ can be assumed to be multinomial. The conditional distribution $P(\tilde{Z}|Z)$ for $\tilde{Z} \in \tilde{\mathbf{Z}}$ is simply a conditional Gaussian (CG)

$$P(\tilde{Z} = \tilde{z}|Z = z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left\{-\frac{(\tilde{z} - \mu_z)^2}{2\sigma_z}\right\} \tag{20}$$

where z denotes a discrete state of Z and $\tilde{z}$ denotes any continuous value of $\tilde{Z}$ and there is no one-to-one correspondence between z and $\tilde{z}$. But now the conditional distribution $P(\tilde{W}|\Gamma_{\tilde{W}}^+, \Psi_{\tilde{W}}^+)$ is approximated by a conditional Gaussian regression (CGR) model

$$P(\tilde{w}|\gamma_{\tilde{W}}^+, \psi_{\tilde{W}}^+) = \frac{1}{\sqrt{2\pi}\sigma_\gamma} \exp\left\{-\frac{(\tilde{w} - \mu_\gamma - \alpha_{\gamma\psi})^2}{2\sigma_\gamma}\right\} \tag{21}$$
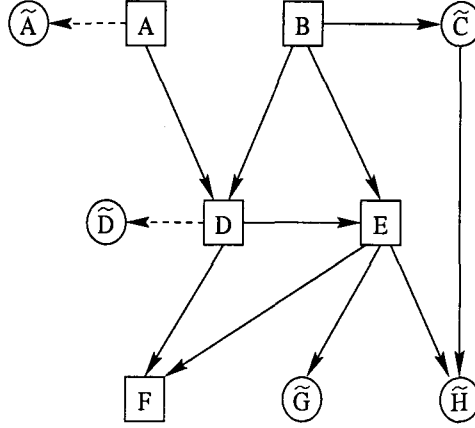
Fig. 5.   Form-II of fuzzy Bayesian network for the hybrid Bayesian network.

where γ on the right side of this equation abbreviates $\gamma_{\tilde{W}}^{+}$ referring to a configuration of $\tilde{W}$'s discrete parents, ψ on the right side abbreviates $\psi_{\tilde{W}}^{+}$ referring to a continuous value vector of $\tilde{W}$'s continuous parents, and α_{γ} is a vector of regression parameters for the given discrete configuration $\gamma_{\tilde{W}}^{+}$ of $\tilde{W}$'s discrete parents. Similar to (12) and (13), it is easy to see that the original conditional distribution of $\tilde{Z}$ for $\tilde{Z} \in \tilde{\mathbf{Z}}$ given its original discrete and continuous parents in HBN and $\tilde{Z}$'s marginal distribution are thus approximated by conditional Gaussian mixture (CGM) models, and the original conditional distribution of $\tilde{W}$ for $\tilde{W} \in \tilde{\mathbf{W}}$ given its original discrete and continuous parents in HBN and $\tilde{W}$'s marginal distribution are approximated by conditional Gaussian regression mixture (CGRM) models.

If we can assume the interaction between $\tilde{W}$ and its continuous parents that can indeed be approximated by regression models α_{γ}, then obviously the second form FBN^{2} provides a finer approximation to HBN than FBN^{1}. However, this assumption puts a limitation on the generality of the FBN^{2} as a price paid for the accuracy in the subclass of applications assumed. Figure 5 shows the second form of the fuzzy Bayesian network for the hybrid Bayesian network of Fig. 3.

3. A HYBRID JUNCTION TREE INFERENCE ALGORITHM FOR THE FORM-I FUZZY BAYESIAN NETWORKS

For a hybrid Bayesian network HBN defined by Eq. (1) representing the joint probability distribution $P(X, \tilde{Y})$, the form-I fuzzy Bayesian network FBN^{1} can be constructed by the definition (4) which provides a general approximation to HBN,

$$P(\mathbf{X}, \tilde{\mathbf{Y}}) \approx P(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}} | FBN^{1}) \,. \tag{22}$$

In the following, we shall only consider $P(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}} | FBN^{-1})$, which shall be written $P(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}})$ as the existence of FBN^{1} is implicitly assumed.

The local Markov property of FBN^1 leads to the following factorization

$$P(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}) = \prod_{X \in \mathbf{X}} P(X|\Gamma_X) \prod_{y \in \mathbf{Y}} P(Y|\Gamma_Y) \prod_{\tilde{Y} \in \tilde{\mathbf{Y}}} P(\tilde{Y}|Y)$$

$$= P(\mathbf{X}, \mathbf{Y})P(\tilde{\mathbf{Y}}|\mathbf{Y}) \tag{23}$$

where $P(\tilde{\mathbf{X}}, \mathbf{Y})$ defines a purely discrete Bayesian network BN^1 of discrete variables $\mathbf{T} = \mathbf{X} \cup \mathbf{Y}$:

$$P(\mathbf{X}, \mathbf{Y}) = \prod_{X \in \mathbf{X}} P(X|\Gamma_X) \prod_{Y \in \mathbf{Y}} P(Y|\Gamma_Y)$$

$$= \prod_{T \in \mathbf{T}} P(T|\Gamma_T)$$

$$= P(\mathbf{T}) \tag{24}$$

and $P(\tilde{\mathbf{Y}}|\mathbf{Y})$ defines a fuzzy peripheral star-like tree whose root node is the whole discrete network BN^1, and in fact, the topology is as simple as

$$P(\tilde{\mathbf{Y}}|\mathbf{Y}) = \prod_{\tilde{Y} \in \tilde{\mathbf{Y}}} P(\tilde{Y}|Y). \tag{25}$$

We shall also call BN^1 the discrete body of FBN^1 and the subgraphs defined by $\tilde{\mathbf{Y}}, \mathbf{Y}$ the fuzzy peripherals. Clearly, we can apply a general-purpose inference algorithm to BN^1 and the fuzzy peripherals can be handled separately. Assume we choose to use the junction tree algorithm.[8,15] A junction tree JT can be constructed by graph triangulation and associated transformations from the discrete body BN^1

$$JT = (\mathbf{C}, \mathbf{S}) \tag{26}$$

where $\mathbf{C}$ is the set of clusters each is a clique of variables from $\mathbf{T}$

$$\mathbf{C} = (C1, C2, \ldots, C_m), \quad C_k \subseteq \mathbf{T}, \quad k = 1, 2, \ldots, m \tag{27}$$

and $\mathbf{S}$ is the set of separators each separating two adjacent clusters in the junction tree JT

$$\mathbf{S} = \{S_{ij} = C_i \cap C_j, \forall C_i, C_j \in \mathbf{C}, Ci \neq C_j\}. \tag{28}$$

In the junction tree JT, for every fictitious discrete node $Y \in \mathbf{Y}$, there is always one cluster (at least one) C_Y which contains Y

$$Y \in C_Y. \tag{29}$$

We can create a new hybrid cluster $\bar{C}_Y$ which contains the continuous node $\tilde{Y}$ and its discrete partner Y,

$$\bar{C}_Y = \{Y, \tilde{Y}\} \tag{30}$$

which can be linked to C_Y through a new junction defined by a new discrete separator $\bar{S}_Y$

$$\bar{S}_Y = \{Y\} = C_Y \cap \bar{C}_Y. \tag{31}$$

This results in a hybrid junction tree HJT

$$HJT = (JT, \bar{\mathbf{C}}, \bar{\mathbf{S}}) = (\mathbf{C}, \bar{\mathbf{C}}, \mathbf{S}, \bar{\mathbf{S}}) \tag{32}$$

where

$$\bar{\mathbf{C}} = \{\bar{C}\} \quad \text{and} \quad \bar{\mathbf{S}} = \{\bar{S}\} . \tag{33}$$

With the hybrid junction tree HJT so obtained, we have

$$\Phi(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}) = \frac{\prod_{C \in \mathbf{C}} \Phi(C)}{\prod_{S \in \mathbf{S}} \Phi(S)} \frac{\prod_{\bar{C} \in \bar{\mathbf{C}}} \Phi(\tilde{C})}{\prod_{\tilde{S} \in \bar{\mathbf{S}}} \Phi(\tilde{S})} \tag{34}$$

where $\Phi()$ denotes the potential which is a nonvanishing nonnegative proportional of probability. The inference in the hybrid junction tree HJT can still be done using the message passing mechanism. The evidence propagation between two discrete clusters is still the same as a standard message passing[8,9] Here we only need to consider how to propagate evidence between a pair of discrete cluster C_Y and hybrid cluster $\bar{C}_Y$ through its separator $\bar{S}_Y$.

Suppose a set of evidence $\mathbf{E}$ is distributed to the hybrid Bayesian network HBN. Let $\Phi^*(A)$ denote the updated potential for a clique A (cluster or separator)

$$\Phi^*(A) = \Phi(A|\mathbf{E}) . \tag{35}$$

To propagate evidence from C_Y to $\bar{C}_Y$ through $\bar{S}_Y$, we first need to update the potential of the separator $\bar{S}_Y$

$$\Phi^*(\tilde{S}_Y) = \Phi^*(Y) = \sum_{A \in C_Y \backslash \{Y\}} \Phi(C_Y) \tag{36}$$

we can then update the potential of the hybrid cluster $\bar{C}_Y$

$$\Phi^*(\bar{C}_Y) = \Phi(\bar{C}_Y)\frac{\Phi^*(\bar{S}_Y)}{\Phi(\bar{S}_Y)} = \Phi(\bar{C}_Y)\frac{\Phi^*(Y)}{\Phi(Y)} . \tag{37}$$

According to the CGM model of (11), we have

$$\Phi(\bar{C}_Y = (\tilde{y}, y)) = \Phi(\tilde{y}, y) = \frac{\Phi(y)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y} - \mu_y)^2}{2\sigma_y}\right\} . \tag{38}$$

Applying (38) into (37), we obtain

$$\Phi^*(\bar{C}_Y|Y = y) = \Phi^*(\tilde{y}, y) = \frac{\Phi^*(y)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y} - \mu_y)^2}{2\sigma_y}\right\} . \tag{39}$$

The marginal potential of the continuous variable $\tilde{Y}$ can be computed by marginalization

$$\Phi^*(\tilde{y}) = \sum_{y \in \Omega_Y} \frac{\Phi^*(y)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y} - \mu_y)^2}{2\sigma_y}\right\} \tag{40}$$

where Ω_Y is the discrete frame of Y.

Defuzzification

The defuzzification from the discrete states of Y to a crisp value $\tilde{y}_0$ of $\tilde{Y}$ can be obtained by maximization

$$\tilde{y}_0 = \arg \max_{\tilde{y}} \Phi^*(\tilde{y}) . \tag{41}$$

This is useful or often required in fuzzy system control.

To propagate an evidence $\tilde{Y} = \tilde{y}_0$ from $\bar{C}_Y$ to C_Y through $\bar{S}_Y$, we first update the potential of the separator

$$\Phi^*(\bar{S}_Y) = \Phi^*(Y = y) = \int_{-\infty}^{\infty} \Phi^*(\tilde{y}, y) d\tilde{y} = \frac{\Phi(y)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y}_0 - \mu_y)^2}{2\sigma_y}\right\} \tag{42}$$

and then we can update the potential for the cluster C_Y

$$\Phi^*(C_Y|Y = y) = \Phi(C_Y|Y = y)\frac{\Phi^*(\bar{S}_Y)}{\Phi(\bar{S}_Y)} = \frac{\Phi(C_Y|Y = y)}{\sqrt{2\pi\sigma_y}} \exp\left\{-\frac{(\tilde{y}_0 - \mu_y)^2}{2\sigma_y}\right\} . \tag{43}$$

4. AN AUGMENTED HYBRID JUNCTION TREE INFERENCE ALGORITHM FOR THE FORM-II FUZZY BAYESIAN NETWORKS

For a hybrid Bayesian network HBN defined by Eq. (1) representing the joint probability distribution $P(\mathbf{X}, \tilde{\mathbf{Y}})$, the form-II fuzzy Bayesian network FBN^2 can be constructed by the definition (14) which provides a general approximation to HBN if the interactions between adjacent continuous variables can be modeled by CGR models,

$$P(\mathbf{X}, \tilde{\mathbf{Y}}) \approx P(\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{Z}}, \tilde{\mathbf{W}}|FBN^2) . \tag{44}$$

In the following, we shall only consider $P(\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{Z}}, \tilde{\mathbf{W}}|FBN^2)$ and which shall be written $P(\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{Z}}, \tilde{\mathbf{W}})$ as the existence of FBN^2 is implicitly assumed. Apparently, the fuzzification transformation (6) does not change the decomposability of the original HBN.

   The local Markov property of FBN^2 leads to the following factorization

$$P(\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{Z}}, \tilde{\mathbf{W}}) = \prod_{X \in \mathbf{X}} P(X|\Gamma_X) \prod_{Z \in \mathbf{Z}} P(Z|\Gamma_Z) \prod_{\tilde{Z} \in \tilde{\mathbf{Z}}} P(\tilde{Z}|Z) \prod_{\tilde{W} \in \tilde{\mathbf{W}}} P(\tilde{W}|\Gamma_{\tilde{W}}^+, \Psi_{\tilde{W}}^+)$$

$$= P(\mathbf{X}, \mathbf{Z})P(\tilde{\mathbf{Z}}|\mathbf{Z})P(\tilde{\mathbf{W}}|\mathbf{X}) \tag{45}$$

$$= P(\mathbf{X}, \mathbf{Z}, \mathbf{W})P(\tilde{\mathbf{Z}}|\mathbf{Z}) \tag{46}$$

where $P(\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{W}})$ defines a hybrid Bayesian network HBN^2 of discrete variables $\mathbf{X}, \mathbf{Z}$ and continuous variables $\tilde{\mathbf{W}}$:

$$P(\mathbf{X}, \mathbf{Z}, \tilde{\mathbf{W}}) = \prod_{X \in \mathbf{X}} P(X|\Gamma_X) \prod_{Z \in \mathbf{Z}} P(Z|\Gamma_Z) \prod_{\tilde{W} \in \tilde{\mathbf{W}}} P(\tilde{W}|\Gamma_{\tilde{W}}^+, \Psi_{\tilde{W}}^+)$$

$$= P(\mathbf{X}, \mathbf{Z})P(\tilde{\mathbf{W}}|\mathbf{X}) \tag{47}$$

and $P(\tilde{\mathbf{Z}}|\mathbf{Z})$ defines a fuzzy peripheral star-like tree whose root node is HBN^2, and in fact, the topology is as simple as

$$P(\tilde{\mathbf{Z}}|\mathbf{Z}) = \prod_{\tilde{z} \in \tilde{\mathbf{Z}}} P(\tilde{z}|Z). \tag{48}$$

If a junction tree can be constructed from HBN^2, then $\tilde{\mathbf{Z}}$ can be attached to this tree with hybrid clusters $\{\bar{C}_Z = \{\tilde{Z}, Z\}\}$ through discrete separators $\{S_Z = \{Z\}\}$. The evidence propagation between a cluster C_Z which contains Z and a hybrid cluster $\bar{C}_Z$ through a separator S_Z can be done with a mechanism similar to the hybrid junction tree inference algorithm presented in the previous section. Therefore, for the augmented hybrid junction tree inference algorithm discussed here, we only need to consider the hybrid network HBN^2 of variables $\mathbf{X}, \mathbf{Z}$ and $\tilde{\mathbf{W}}$.

A hybrid junction tree HJT can be constructed by graph triangulation and associated transformations from the hybrid network HBN^2

$$HJT = (\mathbf{C}, \bar{\mathbf{C}}, \tilde{\mathbf{C}}, \mathbf{S}, \bar{\mathbf{S}}, \tilde{\mathbf{S}}) \tag{49}$$

where $\mathbf{C}$ is a set of discrete clusters of purely discrete variables, $\bar{\mathbf{C}}$ is a set of hybrid clusters of discrete and continuous variables, and $\tilde{\mathbf{C}}$ is a set of continuous clusters of purely continuous variables. Similarly, $\mathbf{S}, \bar{\mathbf{S}}$ and $\tilde{\mathbf{S}}$ are sets of discrete separators, hybrid separators and continuous separators respectively. Let $X - S - Z$ denote that clusters X, Z are adjuncted through separator S, and for $C, C' \in \mathbf{C}, C \neq C'; \bar{C}, \bar{C}' \in \bar{\mathbf{C}}, \bar{C} \neq \bar{C}'; \tilde{C}, \tilde{C}' \in \tilde{\mathbf{C}}, \tilde{C} \neq \tilde{C}'$; and $S \in \mathbf{S}, \bar{S} \in \bar{\mathbf{S}}, \tilde{S} \in \tilde{\mathbf{S}}$, each junction in the hybrid junction tree must belong to one and only one of the seven classes

$$C - S - C' \tag{50}$$

$$C - S - \bar{C}' \tag{51}$$

$$\bar{C} - S - \bar{C}' \tag{52}$$

$$\bar{C} - \bar{S} - \bar{C}' \tag{53}$$

$$\bar{C} - \tilde{S} - \bar{C}' \tag{54}$$

$$\bar{C} - \tilde{S} - \tilde{C}' \tag{55}$$

$$\tilde{C} - \tilde{S} - \tilde{C}'. \tag{56}$$

These classes exhaust all the possible junction types. We can see that a discrete cluster C is never adjacent to a continuous cluster $\tilde{C}$ and in the path between them there must be a hybrid cluster $\bar{C}$. We also see that we can always find a variable ordering consistent with FBN^2 in which all the discrete variables precede all the continuous variables. Consequently, we can always find an ordering of all the clusters in which the first cluster is a discrete one. A hybrid junction tree HJT can be constructed from this cluster ordering such that:

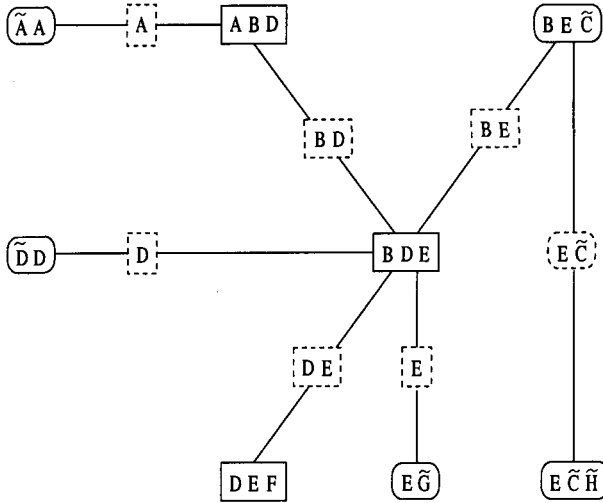1. We can take a discrete cluster in the tree as the root of HJT.

Fig. 6.   A hybrid junction tree HJT for the form-II fuzzy Bayesian network of Fig. 5. Solid boxes show discrete clusters, rounded solid boxes show hybrid clusters, dashed boxes show discrete separators, dashed rounded boxes show hybrid separators. There is no continuous cluster or separator in this example.

2. If any branch of HJT from the root to a leaf contains continuous clusters, then these continuous clusters must be farthest away from the discrete root and there must be at least one hybrid cluster between the continuous clusters to the discrete clusters in that branch.

Lauritzen called such a root as a *strong root*. Statement (iii) of Theorem 2 of Ref. 17 ensures that *the clusters of a decomposable form-II fuzzy Bayesian network can be organized in a hybrid junction tree with at least one strong root*. In the following discussions, we henceforth assume that a hybrid junction tree HJT with a discrete strong root has been constructed from FBN^2. Figure 6 shows a hybrid junction tree corresponding to the form-II fuzzy Bayesian network FBN^2 of Fig. 5.

Each of the seven junction classes should be treated differently for the message passage. The junction class (50) is the same as in an ordinary discrete junction tree handled by the standard junction tree inference algorithm.[8] The classes (51), (53), (54), and (56) are symmetric but the classes (51) and (55) are asymmetric and their bidirectional message passings should be treated differently. However, the junction class (53) can be considered a super class which contains all other classes. Therefore, it suffices just to show the message passing scheme with this class of junctions while different classes may be implemented differently for the sake of computational efficiency.

Consider two adjacent hybrid clusters $\bar{A}, \bar{B}$ that are separated by a hybrid separator $\bar{X}$:

$$\bar{A} = \mathbf{K}_A \cup \tilde{\mathbf{Y}}_A \tag{57}$$

$$\bar{B} = \mathbf{K}_A \cup \tilde{\mathbf{Y}}_B \tag{58}$$

$$\bar{X} = \mathbf{K}_X \cup \tilde{\mathbf{K}}_X, \mathbf{K}_X = \mathbf{K}_A \cap \mathbf{K}_B, \tilde{\mathbf{Y}}_X = \tilde{\mathbf{Y}}_A \cap \tilde{\mathbf{Y}}_B \tag{59}$$

where $\mathbf{K}_A$ and $\tilde{\mathbf{Y}}_A$ denote the set of discrete and continuous variables contained in the cluster $\bar{A}$ respectively. Similarly for $\bar{B}$ and $\bar{X}$.

Now assume a set of evidence $\mathbf{E}$ is distributed to FBN^2, and the potential of $\bar{A}$ is updated from $\Phi(\bar{A})$ to $\Phi^*(\bar{A})$

$$\Phi^*(\bar{A}) = \Phi(\bar{A}|\mathbf{E}) \tag{60}$$

we now need to propagate the evidence from $\bar{A}$ to $\bar{B}$ through separator $\bar{X}$.

First, the potential of separator $\bar{X}$ should be updated by

$$\Phi^*(\bar{X}) = \int_{-\infty}^{\infty} \cdots \tilde{Y} \in \tilde{\mathbf{Y}}_A \int_{-\infty}^{\infty} \left\{ \sum \cdots K \in \mathbf{K}_A \sum \Phi(\mathbf{K}_A, \tilde{\mathbf{Y}}_A) \right\} \prod_{\tilde{Y} \in \tilde{\mathbf{Y}}_A} d\tilde{Y} \tag{61}$$

where K enumerates K_A's discrete elements, and $\tilde{Y}$ enumerates $\tilde{\mathbf{Y}}_A$'s continuous elements.

We then update the potential of hybrid cluster $\bar{B}$

$$\Phi^*(\bar{B}) = \phi(\mathbf{K}_B, \tilde{\mathbf{Y}}_B) \frac{\Phi^*(\mathbf{K}_X, \tilde{\mathbf{Y}}_X)}{\Phi(\mathbf{K}_X, \tilde{\mathbf{Y}}_X)} . \tag{62}$$

Clearly, in order to reach an explicit computable formalism for a general hybrid message passing, we need to have a general conditional Gaussian regression (CGR) model for the potential of a hybrid clique (cluster or separator) and associated operations of general CGR models, including instantiation of discrete and/or continuous evidence, frame extension, marginalization, multiplication, and special division.

Conditional Gaussian Regression Potential (CGRP) Distribution
First of all, a general CGR model for a hybrid clique potential is a generalization of a CGR model of (21). Consider a hybrid clique $\bar{A}$ which contains discrete variables $\mathbf{K}$ and continuous variables $\tilde{\mathbf{Y}}$. Let $\bar{a} = (\mathbf{k}, \tilde{\mathbf{y}})$ be a configuration of $\bar{A}$, where k is a configuration of $\mathbf{K}$ and $\tilde{\mathbf{y}}$ is a vector of continuous values of $\tilde{\mathbf{Y}}$. The potential $\Phi(\bar{a})$ is a function of $\mathbf{k}$ and $\tilde{\mathbf{y}}$

$$\Phi(\bar{a}) = f(\mathbf{k}, \tilde{\mathbf{y}}) = \frac{\Phi(\mathbf{K})}{\sqrt{2\pi|\sum_{\mathbf{k}}|}} \exp\left\{ -\frac{1}{2}(\alpha_{\mathbf{k}}\tilde{\mathbf{y}} - \mu_{\mathbf{k}})' \sum_{\mathbf{K}}^{-1} (\alpha_{\mathbf{k}}\tilde{\mathbf{y}} - \mu_{\mathbf{k}}) \right\} \tag{63}$$

where $\alpha_{\mathbf{k}}$ and $\mu_{\mathbf{k}}$ are two vectors of constant continuous values, $\sum_{\mathbf{k}}$ is the covariance matrix of $\mathbf{y}$, all depending on $\mathbf{k}$. Let

$$g_{\mathbf{k}} = -\frac{1}{2}\mu_{\mathbf{k}}'\mu_{\mathbf{k}} \tag{64}$$

$$\mathbf{h_k} = \mu'_\mathbf{k} \sum_\mathbf{k}^{-1} \alpha_\mathbf{k} \tag{65}$$

$$\Lambda_\mathbf{k} = \frac{1}{2} \alpha'_\mathbf{k} \sum_\mathbf{k}^{-1} \alpha_\mathbf{k} \tag{66}$$

$\Phi(\mathbf{k})$ is a potential of $\mathbf{k}$ which is proportional to itself, so Eq. (63) can be written as

$$f(\mathbf{k}, \tilde{\mathbf{y}}) = \Phi(\mathbf{k}) \exp\{g_\mathbf{k} + \mathbf{h_k}' \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \Lambda_\mathbf{k} \tilde{\mathbf{y}}\} . \tag{67}$$

Let

$$\mathbf{g} = \{g_\mathbf{k}\}, \quad \mathbf{h} = \{\mathbf{h_k}\}, \quad \Lambda = \{\Lambda_\mathbf{k}\}, \text{ for } \mathbf{k} \in \Omega_\mathbf{K} . \tag{68}$$

The triple $(\mathbf{g}, \mathbf{h}, \Lambda)$ which are defined only for $\Phi(\mathbf{k}) > 0$ constitutes the canonical characteristics of the potential distribution $\Phi(\bar{A})$. We shall call the form defined by (63)–(67) the *conditional Gaussian regression potential (CGRP)* model for the potential of a hybrid cluster. Basic operations on CGRP models are provided by Lauritzen.[12] Their properties are summarized as follows:

1. *Extension* of a CGRP model can be done by adjoining 0's to the characteristics such as to give them the desired dimensions.
2. *Multiplication and division* of two CGRP models defined on the same set of variables (after the frame extension) become addition and subtraction of their characteristics.
3. *Marginalization* over continuous variables are finite iff the corresponding parts of Λ are positive definite. *Marginalization* over discrete variables leads to a mixture of CGRP models. This is the major subtle problem with CGRP models. Therefore, we have to resort to approximation using a single CGRP model for a mixture of CGRP models. In-depth elaboration on the marginalization of CGRP models is beyond the scope of this paper.

5. SOME THOUGHTS ON LEARNING FUZZY BAYESIAN NETWORKS FROM DATA

Learning an underlying hybrid Bayesian network from a given data set adds non-trivial extra complexities to learning a discrete network. But the previous work on learning discrete networks still provides a basic ground from where we can start exploring the unknown area of learning hybrid networks. This section is a minor part of this paper as learning of hybrid networks using the formalism of fuzzy Bayesian network is a problem more complicated than the inference with hybrid networks. Thus we may only provide some general thoughts, a computable formalism for learning hybrid networks using the formalism of fuzzy Bayesian network is open to future investigation.

In the previous work on learning Bayesian networks involving continuous variables, there were two commonly adopted assumptions. One is that all the variables are continuous and normally distributed,[6] the approaches based on this assumption

are obviously limited to such a subclass of networks. The other is that all the variables in the network are treated discrete but some observations may be continuous, so we need to discretize the continuous observations to discrete data and then the discrete network learning algorithms may be applied. Recent development along the second assumption is to discretize multivariate continuous observations dynamically or simultaneous while learning is in progress.[5,10,20] But because the underlying network is assumed to be discrete, the so learned network may not be appropriate for continuous variable inference as we presented in the last two sections.

In the following, we first summarize an integrated approach for learning discrete Bayesian networks which is largely based on using existing proved component algorithms. This work is our ongoing effort.[21] We then show the difference of learning hybrid networks and the possible extension of the current approach.

5.1. An Integrated Approach for Learning Discrete Bayesian Networks

Consider a given data set $\mathbf{D} = \{D_k, k = 1, 2, \ldots, N\}$ defined over a discrete variable set $V = (V_1, V_2, \ldots, V_n)$. Each data case D_k is a full configuration of $\mathbf{V}$ if D_k is complete and hard. In case of incomplete but still hard data, D_k can be a partial configuration of $\mathbf{V}$. In case of soft data, D_k contains a likelihood vector for each observed variable.

Assume the data set $\mathbf{D}$ is generated by an underlying discrete Bayesian network, a DAG

$$BN = (\mathbf{V}, \mathbf{L}, \mathbf{P}) \tag{69}$$

where $\mathbf{L}$ is the set of directed links among which there is no directed cycle, $\mathbf{P}$ is a set of conditional probability tables $\mathbf{P} = \{P(V|\Gamma_V), V \in V\}$, where Γ_V is V's parent set. Once the set $\mathbf{V}$ is fixed, we say BN is characterized by its DAG structure $S = \mathbf{L}$ and the vector of parameters describing $\mathbf{P}$ given S

$$BN = (S, \Theta). \tag{70}$$

5.1.1. An integrated criterion

An integrated approach of learning BN from $\mathbf{D}$ is to hypothesize and select an optimal $BN_o = (S_o, \Theta_o)$ from all possible valid alternatives under the MAP and interchangeably MDL criteria

$$BN_o = \arg\max_{S,\Theta} P(\mathbf{D}|\Theta, S)P(\Theta|S)P(S)$$

$$= \arg\min_{S,\Theta} L(\mathbf{D}|\Theta, S) + L(\Theta|S) + L(S) \tag{71}$$

where $L(X)$ denotes the description length of X which relates to $P(X)$ by

$$L(X) = -\log_2 P(X). \tag{72}$$

If we only want to find the optimal structure S_o, we can integrate the scoring functions over Θ and obtain

$$S_o = \arg\max_S P(\mathbf{D}|S)P(S) = \arg\min_S L(\mathbf{D}|S) + L(S). \tag{73}$$

Cooper and Herskovits[2] showed that under the multinomial distribution and complete data, the data likelihood $P(\mathbf{D}|S)$ has a closed-form solution. In case of incomplete and/or soft data, it can be shown that there is no closed-form solution for $P(\mathbf{D}|S)$, and we have to use $P(\mathbf{D}|\Theta, S)$ which can be computed by an iterative EM algorithm.[13]

However, with the pure MAP criterion, there is no generally valid assumption about the prior $P(S)$. In contrast, the MDL principle can be used as a general criterion.[11] With MDL criterion, the prior $P(S)$ is replaced by the prior description length of the structure $L(S)$ which is computable. However, in a pure MDL-based approach, there is no generally valid expression for $L(\mathbf{D}|S)$ without using the multinomial model of $P(\mathbf{D}|S)$. Therefore, in our approach, we choose to use $P(\mathbf{D}|S)$ together with $L(S)$ in an integrated criterion

$$S_o = \arg\max_S P(\mathbf{D}|S)2^{-L(S)} = \arg\min_S[-\log_2 P(\mathbf{D}|S) + L(S)]. \qquad (74)$$

5.1.2. An integrated algorithm

In principle, we should use the integrated criterion (74) through a simultaneous optimization which tends to favor a simpler structure S and also the better fit of S to the data $\mathbf{D}$. However, there is a lower bound of $L(S)$: any two variables $X, Y \in V$ should be directly linked if X, Y are not d-separated by a d-separation variable set $Z \subseteq V$, denoted by $(X \perp Y|\mathbf{Z})$. Equivalently, we may use conditional independence test based on deviance or Chi-square statistics to test if $(X \perp Y|\mathbf{Z})$ is true. Therefore we may start with a fully connected undirected graph of $\mathbf{V}$, and we may iteratively prune each link (X, Y) if $(X \perp Y|\mathbf{Z})$ for $|\mathbf{Z}| = 0, 1, \ldots, T_p$, where T_p is a threshold which can be set to the maximum number of parents for each variable known *a priori* or to $n - 2$ if unknown. In fact, the conditional independence tests should be used in a relative sense, so we should mainly rely on tests with low-order conditioning sets. The ultimate criterion to decide if there should be a link between each pair of variables (X, Y) and how a link should be oriented, is still the integrated criterion (74). Briefly, this integrated algorithm consists of the following four components:

1. **The Potential Graph Algorithm**: This algorithm starts with the fully connected undirected graph of $\mathbf{V}$, then iteratively tries to prune each link (X, Y) by testing if $(X \perp Y|\mathbf{Z})$ for $|\mathbf{Z}| = 0, 1, \ldots, T_p$. The test statistics are deviance which is an equivalent of mutual information, but sensible to the sampling insufficiency. We mainly rely on tests of low-order conditioning sets (d-separation sets). This results in a sparse undirected graph which is called a possibly minimal *potential graph (PG)*.

2. **The Axiomatic Causality Discovery Algorithm**: On the PG so obtained, a subset of undirected links can be oriented using the constraints of d-separation for converging arrows and acyclicity. This results in a subset of directed links L_S.

3. **The Inductive Causality Discovery Algorithm**: The potential graph PG and the subset of directed links L_S will constrain the remaining structural search

space substantially. Then we evaluate each of the alternative structures using the integrated criterion (74). This result in an approximate optimal structure S.

4. **The Structure Refinement Algorithm:** Allowing link omission and commission errors in PG and L_S, we then try to evaluate the loss of each cut link and the worthiness of each determined link still using the integrated criterion (74).

This approach is not pure axiomatic but an integrated one. We are inspired by the work in Refs. 2, 7, 11, 14 and 25 and many others. We believe that as long as we rarely have sufficient data and we can only assume the underlying distribution not exactly modelable by a Bayesian network, this integrated solution is the choice. For more detailed description, see Ref. 21.

5.2. Possible Extensions to Learning Hybrid Bayesian Networks Using the Fuzzy Bayesian Network Formalism

When a subset of variables $\tilde{Y} \subseteq V$ are continuous, the conditional independence between two variables given a conditioning set $(X \perp Y | Z)$, where either X or Y or both can be continuous or discrete and Z can be hybrid, can still be tested. Reference 14, Chapter 6 provides detailed treatment on testing conditional independence of discrete variables, of continuous variables and of mixed variables. Following Lauritzen's guidelines, the Potential Graph Algorithm introduced in the last subsection can be extended to produce a possibly minimal potential hybrid graph (PHG).

The Axiomatic Causality Discovery Algorithm involves pure graph-theoretical operations, so it should also apply to the learning hybrid networks.

Central to the Inductive Causality Discovery Algorithm and the Structure Refinement Algorithm is the evaluation of a hypothetical structure S using the integrated criterion (74). It is possible to construct a computable formalism for $P(\mathbf{D}|S)$ or $P(\mathbf{D}|\Theta, S)$ and $L(S)$ for each of the two forms of the fuzzy Bayesian network. We shall leave these possibilities for future research.

A new class of problems involved in learning FBN's is the determination of the number of components in CG or CGR models and the parameters of these models. McLachlan[18] summarized the existing tools available from the statistical literature for automatic evaluation of finite Gaussian mixture models.

6. CONCLUSION

Hybrid Bayesian networks in which continuous variables and discrete ones may appear anywhere in a DAG are the most general form of Bayesian networks demanded by practical applications. This paper presented a formalism of fuzzy Bayesian networks (FBN) with two alternative forms. The first form of FBN fuzzifies each continuous variable through a fuzzification transformation which replaces the continuous variable with a discrete partner variable and adds a directed link from the discrete partner to the original continuous one. The mapping between the continuous variable and its discrete partner is approximated by a conditional Gaussian (CG), which can be considered as an inverse of a fuzzy set membership function, but no fuzzy logic formalism is used throughout. The second form of FBN only

fuzzifies those continuous variables whose descendants include discrete variables still through the same fuzzification transformation. The mapping between such a continuous variable and its discrete partner is still modeled by a CG distribution, but the dependence between a continuous variable and its hybrid parents is approximated by a conditional Gaussian regression (CGR) distribution. The second form is a finer but less general approximation. We have shown that there is a closed-form hybrid junction-tree inference algorithm for the form-I FBNs, but for the form-II, further approximation is required in an augmented hybrid junction-tree inference algorithm mainly due to marginalization of hybrid clique potential. We have also shown that learning hybrid Bayesian networks is more complicated than learning discrete networks, but it is possible to extend an integrated approach for discrete network learning to learning hybrid networks using the formalism of fuzzy Bayesian networks.

REFERENCES

1. S. Alag and Q. M. Agogino, "Inference using message propagation and topology transformation in vector Gaussian continuous networks," *Proc. 12th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1996.
2. G. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Mach. Learn.* **9** (1992) 309–347.
3. E. Driver and D. Morrel, "Implementation of continuous Bayesian networks using sums of weighted Gaussians," *Proc. 11th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1995, pp. 134–140.
4. D. E. Edwards, "Hierarchical interaction models," *J. R. Stat. Soc.* **B52**, 1 (1990) 3–20.
5. N. Friedman and M. Goldszmidt, "Discretization of continuous attributes while learning Bayesian networks," *Proc. 13th Int. Conf. Machine Learning*, ed. L. Saitta, 1996.
6. Geiger and Hecherman, "Learning Gaussian networks," *Proc. 10th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1994.
7. D. Heckerman, D. Geiger and D. M. Chickering, "Learning Bayesian networks: the combination of knowledge and statistical data," *Proc. 10th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1994.
8. F. V. Jensen, S. L. Lauritzen and K. G. Olesen, "Bayesian updating in causal probabilistic networks by local computations," *Comput. Stat. Quart.* **4** (1990) 269–282.
9. F. V. Jensen, *An Introduction to Bayesian Networks*, Springer, 1996.
10. A. V. Kozlov and D. Koller, "Nonuniform dynamic discretization in hybrid networks," *Proc. 13th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1998.
11. W. Lam and F. Bacchus, "Learning Bayesian belief networks: an approach based on the MDL principle," *Comput. Intell.* **10** (1994) 269–293.
12. S. L. Lauritzen, "Propagation of probabilities, means and variances in mixed graphical association models," *J. Am. Stat. Assoc.* **87**, 420 (1992) 1098–1108.
13. S. L. Lauritzen, "The EM algorithm for graphical association models with missing data," *Comput. Stat. Data Anal.* **19** (1995) 191–201.
14. S. L. Lauritzen, *Graphical Models*, Clarendon Press, Oxford, 1996.
15. S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems," *J. R. Stat. Soc.* **B50**, 2 (1988) 157–224.
16. S. L. Lauritzen and N. Wermuth, "Graphical models for associations between variables, some of which are qualitative and some quantitative," *J. Ann. Statist.* **17** (1989) 31–57.

17. H. -G. Leimer, "Triangulated graphs with marked vertices," *Graph Theory in Memory of G. A. Dirac. Annals of Discrete Mathematics* **41**, eds. L. D. Andersen *et al.*, 1989, pp. 311–324.
18. McLachlan, *Finite Mixture Models* (Wiley, New York, 1999).
19. D. W. McMichael, L. Liu and H. P. Pan, "Estimating the parameters of mixed Bayesian networks from incomplete data," *Proc. 1999 Information, Decision and Control (IDC'99)*, February, Adelaide, Australia, IEEE Catalog Number 99EX251, 1999.
20. S. Monti and G. F. Cooper, "Learning hybrid Bayesian networks from data," *Learning and Inference in Graphical Models*, ed. M. Jordan, Springer-Verlag, NY, 1998.
21. H. P. Pan, L. Liu and D. W. McMichael, "An integrated approach for learning Bayesian networks under MAP/MDL criteria," CSSIP Commercial Report CR 8/99 (100 pp.), March 1999 (to be unclassified).
22. H. P. Pan and D. W. McMichael, "Fuzzy causal probabilistic networks — a new ideal and practical inference engine," *Proc. 1998 Int. Conf. Multisource-Multisensor Information Fusion*, July, Las Vegas, 1998.
23. J. Pearl and T. Verma, "A formal theory of inductive causation," Technical Report R-155, Cognitive Systems Laboratory, Computer Science Department, UCLA, 1990.
24. P. Spirtes, C. Glymour and R. Scheines, "Causality from probability," *Evolving knowledge in Natural and Artificial Intelligence*, ed. G. McKee, Pitman, London.
25. P. Spirtes, C. Glymour and R. Scheines, *Causation, Prediction and Search*, Lecture Notes in Statistics 81, Springer-Verlag, NY, 1993.

Heping Pan currently is a Distinguished Professor of photogrammetry, remote sensing and Director of Digital Earth Research Centre, Wuhan Technical University of Surveying and Mapping, Wuhan, China. He gained a Ph.D. in computer vision from the University of Twente, The Netherlands, 1990. He was a Research Fellow in the Institute of Photogrammetry, University of Bonn, Germany, for four years and a Senior Investigator and Information Fusion Project Leader at Australian Cooperative Research Centre for Sensor Signal and Information Processing, Adelaide for five years until the end of 1999. He has published extensively in stereo photogrammetry and computer vision, data and information fusion, as well as Bayesian networks. His recent work has been on hybrid Bayesian networks, learning Bayesian networks from data, stereo image matching, etc.

Lin Liu received the B.Eng. and M.Eng. degrees in automatic control theory from Xidian University, Xi'an, P.R. China in 1991 and 1994, respectively. She was a lecturer in the School of Electronic Engineering of Xidian University from 1994 to 1998. From 1998 to 1999, she was a visiting researcher in the Cooperative Research Centre for Sensor Signal and Information Processing, Adelaide, Australia, where she joined the Information Fusion Project and worked on the Bayesian network learning theory.

Her research interests include multitarget tracking, Bayesian networks and Petri nets. Currently, she is a Ph.D. student in the University of South Australia, working on modeling and analysis of distributed systems with Petri nets.

EXTRACTION OF BIBLIOGRAPHY INFORMATION BASED ON IMAGE OF BOOK COVER

HUA YANG, MASAAKI KASHIMURA,
NORIKADU ONDA AND SHINJI OZAWA

Department of Electrical Engineering Keio University
Hiyoshi 3-14-1, Kohoku Ward, Yokohama City, Japan 223-8522
E-mail: {yanghua, kassy, ozawa}@0zawa.ics.keio.ac.jp

This paper describes a new system for extracting and classifying bibliography regions from the color image of a book cover. The system consists of three major components: preprocessing, color space segmentation and text region extraction and classification.

Preprocessing extracts the edge lines of the book and geometrically corrects and segments the input image, into the parts of front cover, spine and back cover.

The same as all color image processing researches, the segmentation of color space is an essential and important step here. Instead of RGB color space, HSI color space is used in this system. The color space is segmented into achromatic and chromatic regions *first;* and both the achromatic and chromatic regions are segmented further to complete the color space segmentation.

Then text region extraction and classification follow. After detecting fundamental features (stroke width and local label width) text regions are determined. By comparing the text regions on front cover with those on spine, all extracted text regions are classified into suitable bibliography categories: author, title, publisher and other information, without applying OCR.

Keywords: Color space segmentation; HSI color space; achromatic region; chromatic region; text region.

1. INTRODUCTION

Today, it has become possible to search for bibliographies of books, indexes of journals and abstracts of papers online in many libraries or on the Internet. However, most of the original information is still extracted and input manually, so that building those index databases would require extensive human labor. On the other hand, rapid increase in the range and volume of publications makes manual work more and more difficult. Thereby automatic information extraction and input techniques are demanded,[6,14] and increasing emphasis is being placed on the realization of computer based systems which are able to analyze printed documents automatically. However, in document analysis research, much attention is paid to text parts;[2,6,14] while the research in relation to bibliography information extraction are carried out less frequently.

Our research is centered around the idea of automatic extraction and classification of bibliography information from the color image of a book cover. Here, we use Leaf CatchLight (Fig. 1), which is a digital camera that can produce high-quality color images, to get the cover image of an opened book as an input image (Fig. 2). The bibliography information region of the book will be extracted from the input
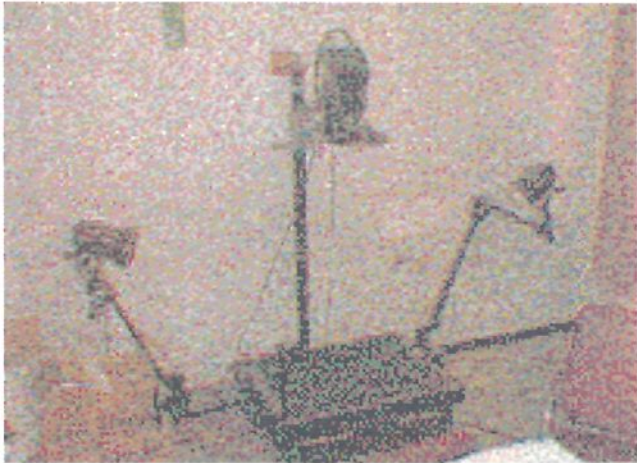
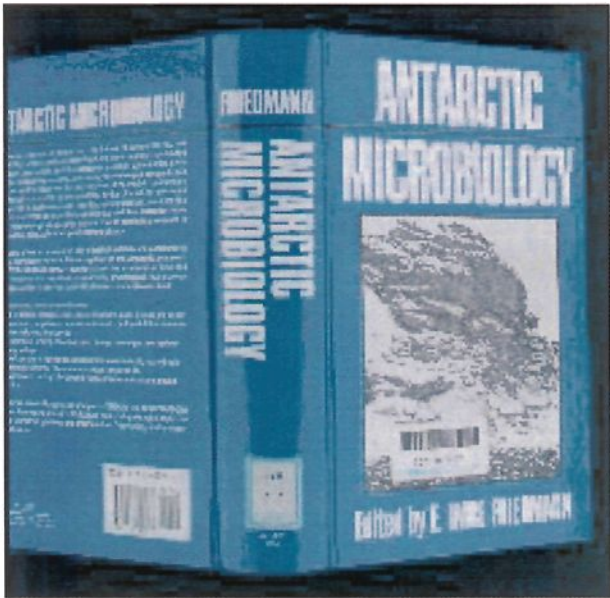Fig. 1.   Leaf CatchLight digital camera.

Fig. 2.   An input image.

image and the image will be created based on the extraction result. In this paper, we suggest to accomplish the task according to the flow shown in Fig. 3.

This paper is organized as follows: In Secs. 2–4, methods of preprocessing, color space segmentation and text region extraction and classification are addressed in order, with examples. Sections 5 and 6 provide experiments, conclusion and discussion.
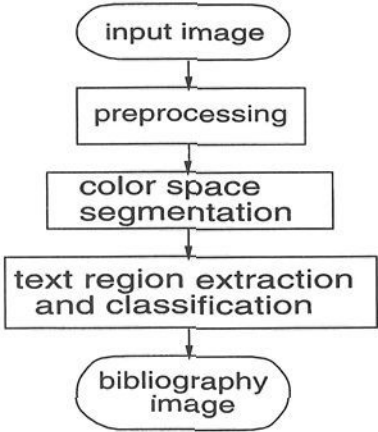
Fig. 3.   The system flow.

2. PREPROCESSING

A book cover consists of three parts: front cover, spine and back cover. When an opened book is photographed into an image plane face-down, the top and bottom edges of the book will appear convex broken lines (Fig. 4) in the image; and the convex broken lines break exactly at the points that correspond to the vertices of spine. Therefore the image of a book can be split into three parts, each of which
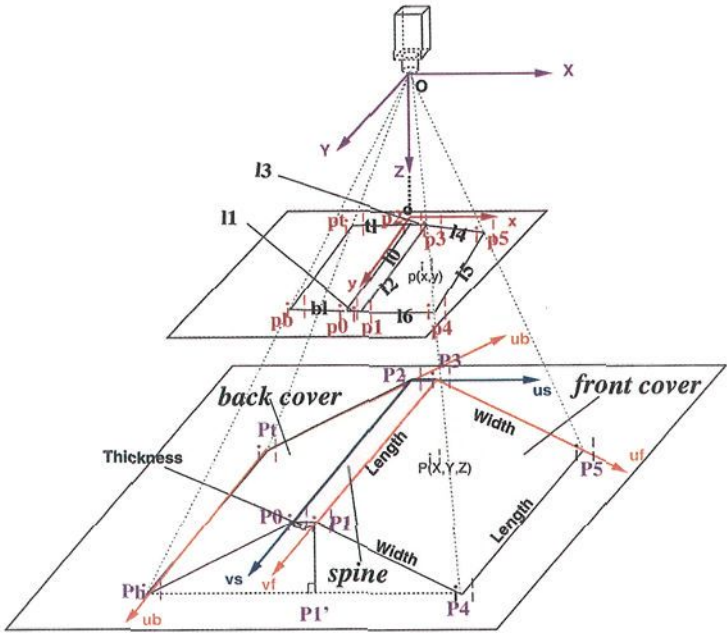
Fig. 4.   Definition of coordinate systems.
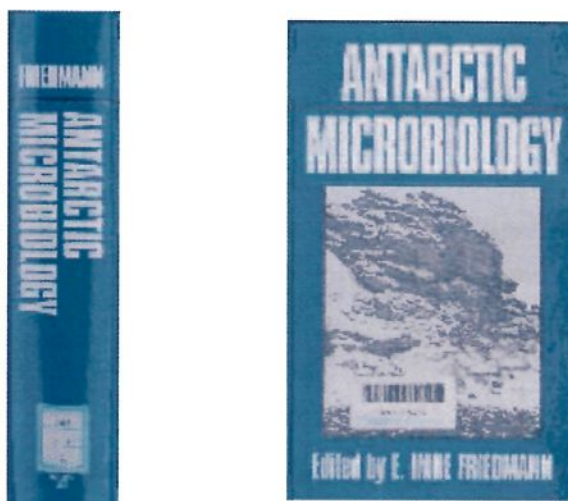
Fig. 5.   Geometrical segmentation result.

corresponds to one of the three parts of the book cover. In other words, as long as the vertices of spine are detected, the input image can be geometrically segmented into the parts of front cover, spine and back cover, along the lines between correspondent vertices.

Furthermore, by geometrically correcting each part separately, the images of front cover, spine and back cover are created.

However, because bibliography information usually does not appear on back cover, processes described in the following sections operate on only front cover and spine (Fig. 5).

3. COLOR SPACE SEGMENTATION

Color space segmentation is an essential and important step in color image processing, because any error that occurs here might influence the following processing.[13] In previous research on document analysis, text is often assumed to be printed in black on a white background, so that it can be extracted by thresholding the gray-scale image.[17] On the other hand, although research on multicolor document are getting more and more popular in recent years, much of the attention is paid to text parts. The process presented in this paper pays attention to a book cover, where bibliography text could be printed in any font, size and on any background, which makes the task much more complicated; and the color is not known *a priori*.

3.1. Achromatic and Chromatic Regions Segmentation

Although there are various color spaces, *HSI* space is the one that closely corresponds to the human visual interpretation of color, which can be described in terms of *H*ue, *S*aturation and *I*ntensity.[9,16] Among the three attributes of *HSI* space,

hue is the most reliable one, because it is independent of the intensity attribute. However, there is still one major problem associated with hue: hue is unstable at too low or too high saturations and intensities.[1,8,12,13] To solve this problem, it is necessary to define the effective ranges of hue, i.e. to define achromatic and chromatic areas in HSI color space, first.

Achromatic colors are defined as colors which are completely desaturated and whose values depend on intensity only:[12] while chromatic colors are defined as colors with stable hue values and whose values depend on all the three attributes. From the concepts mentioned above and the characteristics of color display systems, an image can be divided into achromatic and chromatic regions following the next steps [Figs. 6(a)–7(b)]:

1. Thresholding global intensity into $I_0 \ldots I_{n-1}$ categories and saturation histogram into $S_0 \ldots S_{m-1}$ categories.
2. A color is defined as achromatic color if its intensity belongs to intensity category I_0 or I_{n-1}, or its saturation belongs to saturation category S_0 or S_{m-1} (in Fig. 6, $I_0 = [0, 101]$; $I_{n-1} = [153, 256)$; $S_0 = [0, 128]$; $S_{m-1} = [192, 256)$).
3. Thresholding global intensity into n categories named as $I_0 \ldots I_{n-1}$ and saturation into m categories named as $S_0 \ldots S_{m-1}$.
4. A color is defined as achromatic color if its intensity belongs to intensity category I_0 or I_{n-1}, or its saturation belongs to saturation category S_0 or S_{m-1}.
5. The other colors are defined as chromatic colors.

3.2. Further Segmentation

Because achromatic colors only depend on their intensity values, the achromatic region can be segmented further into i categories at the thresholds $TA_0 = 0, TA_1, \ldots,$ and $TA_i = 255$. Meanwhile, the chromatic region can be segmented further according to all the hue, intensity and saturation histogram thresholding following the steps:

1. Segment the global hue histogram into h categories at thresholds $Th_0 = 0$, $Th_1, \ldots,$ and $Th_h = 255$.
2. Within the region where *hue* $\in [Th_j, Th_{j+1}]$, segment its local intensity histogram into l_j categories at thresholds $TI_{j0} = 0, TI_{j1}, \ldots$ and $TI_{jl_j} = 255$.
3. Furthermore, within the region where *hue* $\in [Th_j, Th_{j+1}]$ and *intensity* $\in [TI_{jk}, TI_{j(k+1)}]$, segment its local saturation histogram into s_{jk} at thresholds $TS_{jk0} = 0, TS_{jkl}, \ldots$ and $TS_{jks_{jk}}$.

So now the chromatic region is segmented further into $\sum_{j=0}^{h-1} \sum_{k=0}^{l_j-1} s_{jk}$ categories [Fig. 6(c) and 7(c)]. All the thresholds are calculated automatically using the fourth central moment method.[5]

Thus the color space segmentation is completed and text regions will be extracted from binary images each of which is created corresponding to one of the segments.
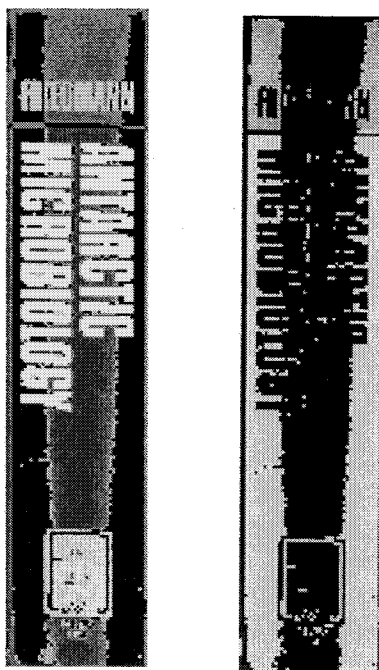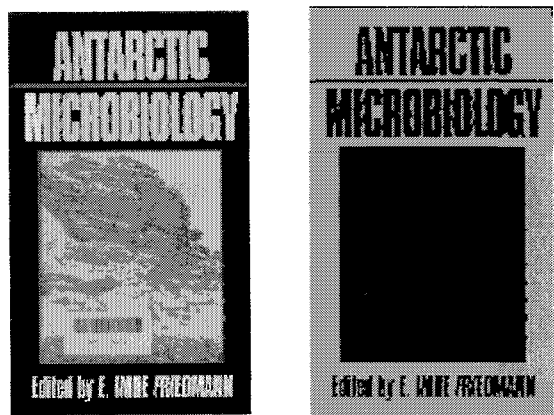
(a) Achromatic region   (b) Chromatic region
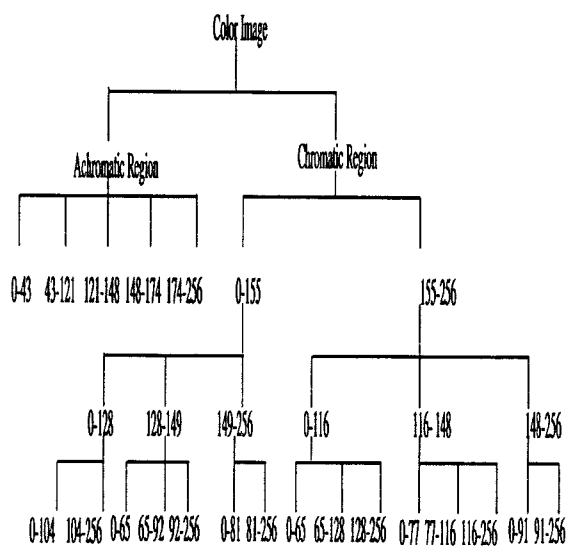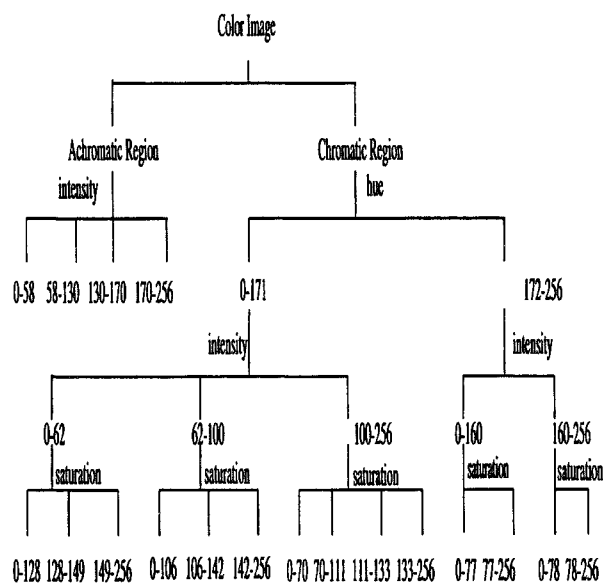

(c) Further Segmentation Result

Fig. 6.   Result of color space segmentation (spine).

(a) Achromatic region   (b) Chromatic region

(c) Further Segmentation Result

Fig. 7.   Result of color space segmentation(front cover).

4. TEXT REGION EXTRACTION AND CLASSIFICATION

Although a book cover structurally consists of rectangular blocks, it is a complex document containing text regions and graph regions, that probably appear in any position. Our objective in this section is to (1) extract text regions from given binary images, which are obtained from the previous process, to create a text-only image; and (2) classify the text regions into suitable bibliography categories.

4.1. Text Region Extraction

4.1.1. Feature Detection

1. Stroke Width

   There is an important and fundamental attribute of document images: text regions contain highly structured stroke units. Also we found that the feature of stroke width can be detected from run-length histogram. The stroke width set SW is defined as the run-length with high frequencies in the run-length histogram, except the run-length whose length is one.[7,15] This feature reflects the average width of strokes in a text image, and the number of the element (SN) of SW reflects the number of font size (Fig. 8). In other words, it can be considered that there is no text region in an image if $SN = 0$; a stroke width set must be able to detect from an image in which the text region exists.

2. Label

   By labeling a binary image, the connected pixels which belong to the same component are given the same label. So label is another important feature reflecting connected components, or pixels. In an ideal case, one label corresponds to one letter.

(a) a binary image   (b) run-length histogram
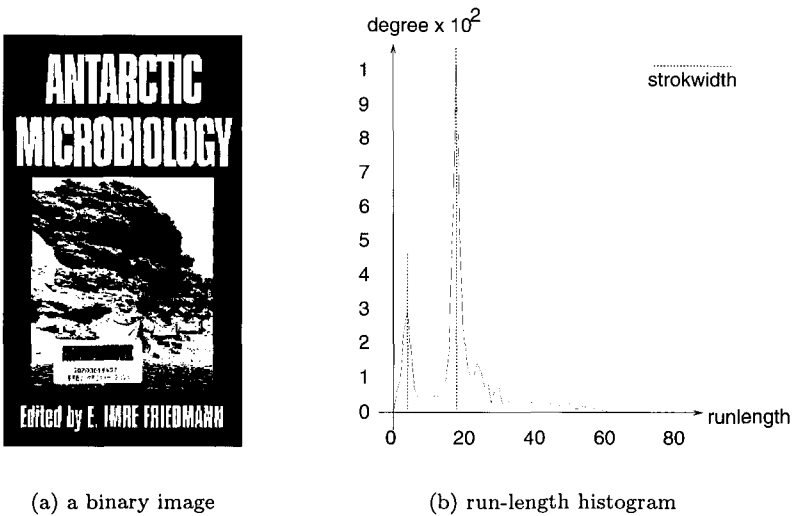
Fig. 8.   An example of stroke width feature detection.

4.1.2. Block Segmentation

A candidate image is composed of several blocks, each of which represents a coherent component of the image, one text line or one graph region. The blocks are obtained by smearing the image, so that adjacent connected components are merged into one block if the distance between them is less than the smearing threshold.[4,15] Because it is a rule that the space between letters is narrower than the width of a letter,[10]

the smearing threshold varies dynamically with the local average label width.

4.1.3. Text Region Determination

If a block is a text block, it should meet the following criteria:

1. Local stroke width set (sw) can be detected. Moreover, sw is a subset of SW.
2. The area of each label in the block is neither too large nor too small.
3. The ratio of the length to the width of a label is proper, generally for most formal printed letters it would be about 1.0, however, there are some exceptions such as the letter "l", "i", "h", "t", etc. whose length–width ratio may be 2.0–4.0.
4. When there are more than three labels in one block, the centroids of those labels are on the same straight line, assuming to be in a horizontal direction line.[3]

Based on the above criteria, it can be determined whether a block is a text block or not. After all the object binary images are operated on, an image in which only text regions remain is available by merging all the text blocks (Fig. 9).

4.2. Text Region Classification

The design of book cover varies with book, however, there is a style guide for spine design: the top three-quarters of the spine, comprises the bibliographic identification area; and the remaining quarter is shared by the library identification area (location for the library classification number) above the publishers identification area;[11] and front cover bears stamped or printed materials, such as author and title, etc. Therefore by comparing front cover to spine, the text regions appearing on the cover can be classified. The text region classification strategy is as follows.

(a)   (b)   (c)   (d)
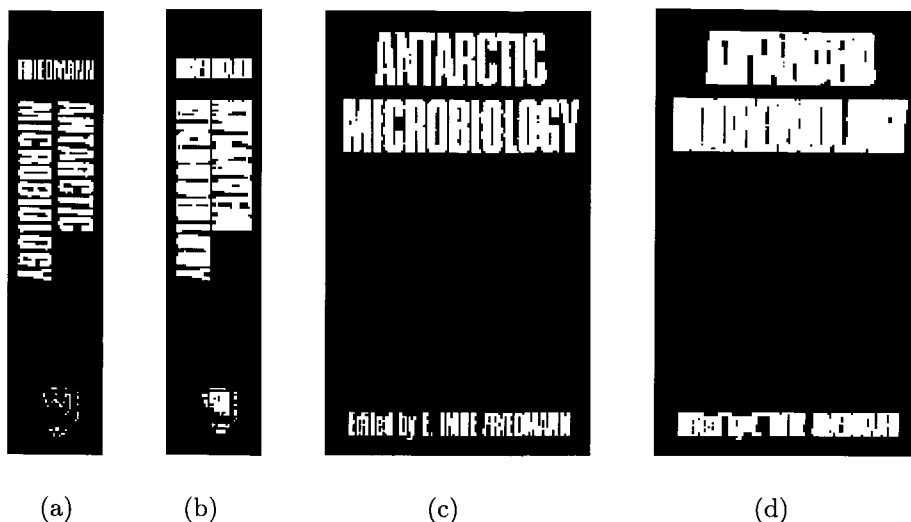
Fig. 9.   Results of text regions extraction and block segmentation. (a) and (b) results of spine; (c) and (d) results of front cover.

Table 1. Result of block matching.

1. Block segmentation with the dynamic smearing thresholds of local black-pixel-runlength (Fig. 9).
2. Normalization.
3. Match each block on front cover to those on spine: image matching and features matching. The block on front cover which contains components within a block on spine is considered to belong to the same region as that block (Table 1).
4. Those blocks on front cover which cannot find the corresponding blocks on spine are classified into *other information* region.

After classifying the text regions, a bibliography image and an *other information* image, if it so exists, are created, as a result (Fig. 10).
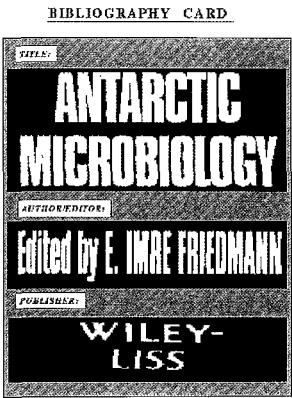
Fig. 10.   Final result.

5. EXPERIMENT AND RESULTS

The suggested system has been tested on different object books. In this section, we demonstrate the effectiveness of the system by presenting one more example, shown in Figs. 11–16 and Table 2. This system was worked on Sun Unix Workstation with SunOS 4.1, it cost 300 seconds, 10 seconds, 10 seconds in preprocessing, color space segmentation and text region extraction and classification sections respectively. Now the authors are porting the system from Unix application to windows application, the calculation time is expected to be reduced.
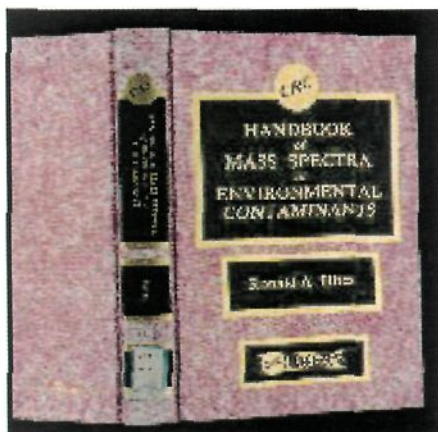
Fig. 11.   Input image.
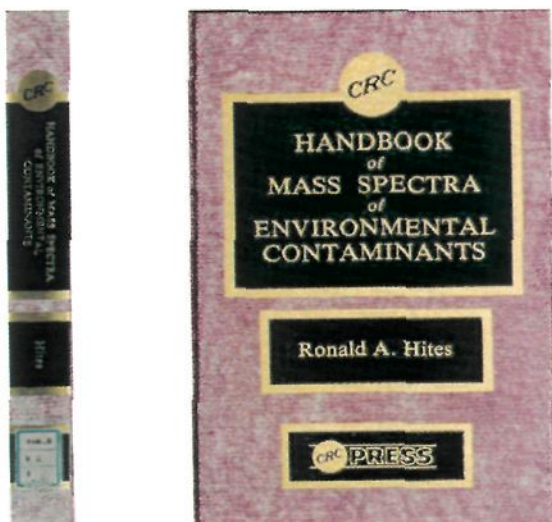
Fig. 12.   Results of preprocessing.
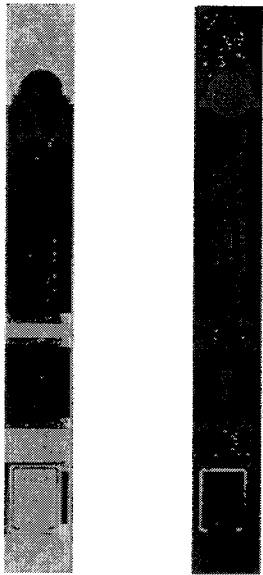
Fig. 13.   Color space segmentation results of spine.
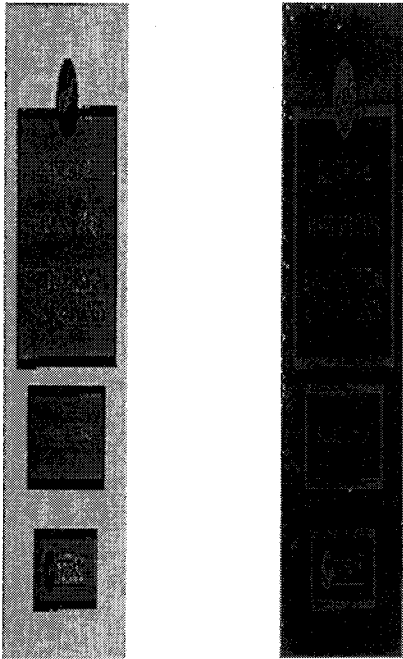
(a) Achromatic region   (b) Chromatic region

Fig. 14.   Color space segmentation results of front cover.

Fig. 15. Text regions extraction results.
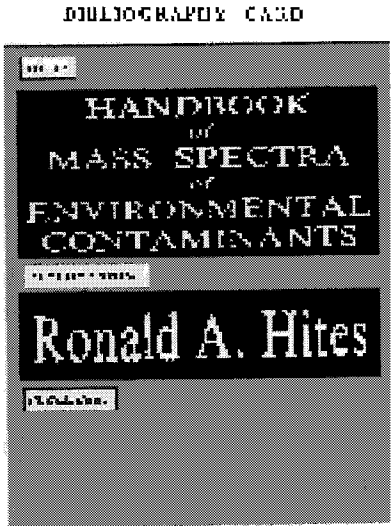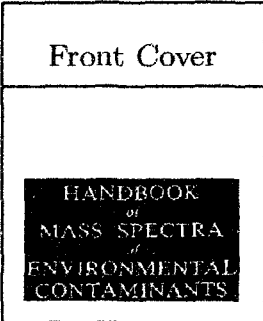
Fig. 16. Final result.

6. CONCLUSION AND DISCUSSION

6.1. Conclusion

In this paper, we proposed a new system for extracting bibliography regions from the color image of a book cover. It is a new application among document layout analysis. The images are taken just once so that it would be convenient for users and the object books would be protected from being damaged. The simple prepro-cessing, text extraction method ensured the executive speed. Moreover, OCR is not necessary for text region classification in our research, though it is necessary when creating the final information database.

Table 2. Result of block matching

Front Cover	Spine

6.2. Discussion

Although we have got the results we had expected, there is more to be done for our future project.

- For the current stage, our object books are all hardcover ones, which simplified the processes somehow. In the near future, we will make some attempts to apply our system to softcover books, which should be much more challenging.
- In our research, bibliography regions have been extracted and classified properly. However it would be more useful if the bibliography database could be created and input automatically afterwards. This will be our work in the future.

ACKNOWLEDGMENT

REFERENCES

1. J. Durrett, *Color and The Computer*, Academic Press, 1987.
2. L. A. Fletcher and R. Kasturi, "A robust algorithm for text string separation from mixed text/graphics images," *IEEE Trans. Patt. Anal. Mach. Intell.* **10**, 6 (1988).
3. B. Gatos, N. Papamarkos and C. Chamzas, "Skew detection and text line position determination in digitized documents," *Patt. Recogn.* **30**, 9 (1997) 1505–1519.
4. H. Goto and H. Aso, "Robust and fast text-line extraction using local linearity of the text-line," *Syst. Comput. Japan* **26**, 13 (1995) 21–31.
5. S. Guo and S. Ozawa, "A new criterion for automatic multi-level thresholding," *Int. Workshop Series on Image Analysis and Information Fusion (IAIF' 97)*, Nov. 1997, p. 85.
6. K. Iwane, M. Yamaoka and O. Iwaki, "A functional classification approach to layout analysis of document images," *Proc. Second Int. Conf. Document Analysis and Recognition*, Oct. 20–22, 1993, pp. 774–777.

7. Y. Liu and S. N. Srihari, "Document image binarization based on texture features," *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, 5 (1997) 540–544.

8. A. Moghaddamzadeh and N. Bourbakis, "A fuzzy region growing approach for segmentation of color images," *Patt. Recogn.* **30**, 6 (1997) 867–881.

9. I. Pitas and P. Kiniklis, "Multichannel techniques in color image enhancement and modeling," *IEEE Trans. Imag. Process.* **5** (1996) 168–171.

10. A. Simon, J. Christophe and A. P. Johnson, "A fast algorithm for bottom-up document layout analysis," *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, 3 (1997) 273–277.

11. Style Manual Committee, *Scientific Style And Format — The CBE Manual for Authors, Editors, Publishers*, Cambridge University Press, 1994.

12. M. Takagi and H. Shimoda, *Handbook of Image Analysis*, Tokyo University Press, 1991.

13. D. C. Tseng and C.-H. Chang, "Color segmentation using *UCS* perceptual attributes," *Proc. National Science Council*, Republic of China, Part A (Physical Science and Engineering) **18**, 3, pp. 305–313.

14. S. Tsujimoto and H. Asada, "Major components of a complete text reading system," *Proc. IEEE* **80**, 7 (1992).

15. D. Wang and S. N. Srihari, "Classification of newspaper image blocks using texture analysis," *Comput. Vis. Graph. Imag. Process.* **47** (1989) 327–352.

16. A. R. Weeks, C. E. Felix and H. R. Myler, "Edge detection of color image using the HSL color space," *Proc. SPIE*, Vol. 2424, 1995, pp. 291–297.

17. Y. Zhong, K. Karu and A. K. Jain, "Locating text in complex color images," *Patt. Recogn.* **28**, 10 (1995) 1523–1535.

Hua Yang received the B.S. degree in computer science from Nankai University, China in 1994 and the M.E. degree in electric engineering from Keio University, Japan in 1998, respectively. She is currently a Ph.D student in the Department of Electrical Engineering at Keio University, Japan.

Her research interests include computer vision, image processing and pattern recognition.
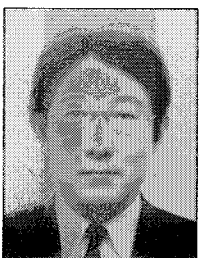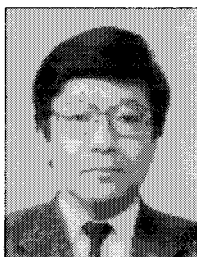
Norikadu Onda received B.E., M.E. degrees and Ph.D in electric engineering from Keio University, Japan, in 1980, 1982 and 1989, respectively. He has been a lecturer of Keio University, and is currently Associate Professor at Shobi University, Japan.

His research interests include pattern recognition and image coding. He is a member of IEICE (Japan) and IEEE (USA).
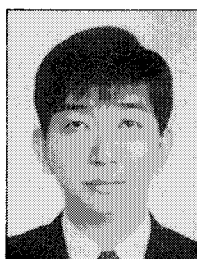
Masaaki Kashimura received the B.E. and M.E. degrees in the Department of Electrical Engineering from Keio University, Japan in 1988 and 1990 respectively. He has been one of the research staff of the HUMI Project in Keio University since 1997.

His research interests include image processing, image coding and digital archiving.

Shinji Ozawa received B.E. and M.E. degrees and Ph.D in electric engineering from Keio University, Tokyo, Japan, in 1967, 1969 and 1972, respectively. He has been a Research Associate, Assistant Professor and Associate Professor from 1970 in the Department of Electric Engineering of Keio University. He is now a professor in the same department. In 1984, he had been a Visiting Professor of University of Maryland, USA. He has been working on Digital Communication and Signal Processing of Image and Voice. He is a member of IEICE (Japan), IPSJ (Japan) and IEEE (USA).

RADAR TARGET RECOGNITION BASED ON PARAMETERIZED HIGH RESOLUTION RANGE PROFILES*

XUEJUN LIAO and ZHENG BAO

Key Lab. For Radar Signal Processing
Xidian University, Xi'an 710071, P. R. China
E-mail: xjliao@rsp.xidian.edu.cn

A new scheme of radar target recognition based on parameterized high resolution range profiles (PHRRP) is presented in this paper. A novel criterion called generalized-weighted-normalized correlation (GWNC) is proposed for measuring the similarity between PHRRP's. By properly choosing the parameter of the mainlobe width in GWNC, aspect sensitivity of PHRRP's can be reduced without sacrificing their discriminative power. Performance of the scheme is evaluated using a dataset of three scaled aircraft models. The experimental results show that by using GWNC, only a small number of most dominant scatterers can achieve the same recognition rates as HRRP's, thus leading to a significant data reduction for the recognition system.

Keywords: Radar target recognition; high resolution range profiles (HRRP); dominant scatterers; parameterized HRRP (PHRRP); generalized-weighted-normalized correlation (GWNC).

1. INTRODUCTION

Radar target recognition based on high resolution range profiles (HRRP) has received much attention in recent years.[2,5] A HRRP is basically the distribution of the target's scatterers along the radar line of sight and is usually obtained via Fourier transform of the radar return recorded in the frequency domain. Since scatterers occur only at the peaks of a HRRP and the frequency domain data are usually zero padded in order to locate all significant scatterers correctly, information redundancy is quite severe in HRRP's. Based on the scatterers' model, parameterizations of HRRP's have been proposed to achieve data reduction using the modified Prony algorithm[1] and the least squares fitting method.[3] Since the distribution of scatterers is target-dependent and usually nonuniform, the parameterized HRRP's (PHRRP) cannot be regarded as discrete signals and the normalized correlation (NC) based recognition scheme can no longer be used to classify them. In this paper, we present a new scheme of radar target recognition based on PHRRP's. We first use the RELAX algorithm[4] to extract PHRRP's from the frequency domain measurements. Then we propose a novel criterion called generalized-weighted-normalized correlation (GWNC) for measuring the similarity of PHRRP's, and use it to classify the extracted PHRRP's.

The rest of the paper is organized as follows. Section 2 gives a brief introduction of the data model and the RELAX algorithm. Section 3 establishes the GWNC criterion. Section 4 presents experimental results. Finally Sec. 5 concludes the paper.

2. DATA MODEL AND EXTRACTION OF PHRRP'S

2.1. Data Model

The frequency domain radar return can be modeled as

$$X(f) = \sum_{k=1}^{K} \sigma_k \exp\left(j2\pi \frac{2r_k}{c}(f + f_0)\right) \tag{1}$$

where r_k and σ_k, respectively, are the down range and scattering strength of the kth scatterer, K is the number of scatterers on the target, f_0 is the center frequency, and c is the speed of light. Equation (1) can be interpreted as the superimposition of K complex sinusoids with complex amplitudes $\sigma_k \exp(j2\pi \frac{2r_k}{c} f_0)$ and frequencies $\frac{2r_k}{c}$, $k = 1, 2, \ldots K$. Hence the HRRP, i.e. Fourier transform of $X(f)$, is in fact a band limited and discrete spectrum, each spectral line corresponding to a scatterer on the target. This observation leads us naturally to using parametric spectrum estimation techniques[1,3,5] to extract the parameters, i.e. $\{\sigma_k \exp(j2\pi \frac{2r_k}{c} f_0), r_k\}_{k=1}^{K}$, in Eq. (1). Usually we are not concerned with the phase of the complex amplitudes. Therefore the parameters that serve for later recognition purpose are represented by $\{|\sigma_k|^2, r_k\}_{k=1}^{K}$, which are referred to as the parameterized HRRP (PHRRP). Various methods are available for extraction of PHRRP's, including the modified Prony algorithm[1] and the least squares fitting method.[3] The more recently proposed RELAX algorithm[4] gives the estimates of spectrum parameters directly and has been demonstrated to be robust to noise. For this reason, it shall be used in this paper to extract the PHRRP's from the frequency domain measurements.

2.2. Extraction of PHRRP's

Using matrix notations, the discrete version of Eq. (1) can be written as

$$\mathbf{x} = \mathbf{\Omega}\mathbf{a} \tag{2}$$

where $\mathbf{a}$ is column vector representing $\sigma_k \exp(j2\pi \frac{2r_k}{c} f_0)$ for $k = 1, 2, \ldots K$, and $\mathbf{\Omega}$ and $\mathbf{x}$ are matrix and column representations of $\exp(j2\pi \frac{2r_k}{c} f)$ and $X(f)$, respectively, at discrete samples of f. The parameter estimation of Eq. (2) is to estimate $\mathbf{\Omega}$ and $\mathbf{a}$ from $\mathbf{x}$, which is the frequency domain measurements recorded by the radar. This can be achieved by minimizing the square error $e^2 = \|\mathbf{x} - \mathbf{\Omega}\mathbf{a}\|^2$. The optimization can proceed in two steps. First suppose $\mathbf{\Omega}$ is known, then the square error is minimized by orthogonally projecting $\mathbf{x}$ onto the column space of $\mathbf{\Omega}$, i.e. setting $\mathbf{a} = \mathbf{\Omega}^+\mathbf{x}$, where $\mathbf{\Omega}^+ = (\mathbf{\Omega}^H\mathbf{\Omega})^{-1}\mathbf{\Omega}^H$ is the pseudo inverse of $\mathbf{\Omega}$. Next, we optimize $\mathbf{\Omega}$ to minimize $e^2 = \|\mathbf{x} - \mathbf{\Omega}\mathbf{\Omega}^+\mathbf{x}\|^2$, which is equivalent to maximizing $\|\mathbf{\Omega}\mathbf{\Omega}^+\mathbf{x}\|^2$ because $\mathbf{\Omega}\mathbf{\Omega}^+$ is the orthogonal projector on the column space of $\mathbf{\Omega}$. When $K = 1$, $\mathbf{\Omega} = [1 \exp(j2\pi \frac{2r_1}{c}\Delta f) \cdots \exp(j2\pi \frac{2r_1}{c}\Delta f(N-1))]^T$, with Δf

being the sampling interval of $X(f)$ and N the number of sampling points, and $\|\Omega\Omega^{+}\mathbf{x}\|^2 = \frac{1}{N}\|\Omega^H\mathbf{x}\|^2$, which can be implemented by FFT.

The RELAX algorithm[4] is an algorithm that iteratively finds the strongest complex sinusoid present in $\mathbf{x}$ using the above methods and then cleans the complex sinusoid form $\mathbf{x}$. At each iteration when the kth strongest complex sinusoid is found, the 1st–kth strongest complex sinusoids are reestimated, again iteratively, until the relative change of the residue energy of $\mathbf{x}$ is smaller than satisfied. This finely tunes the parameters to their real values. For the case, when the number of complex sinusoids K is unknown, the RELAX algorithm stops when the residue energy of $\mathbf{x}$ is smaller than desired.

3. GENERALIZED-WEIGHTED-NORMALIZED CORRELATION (GWNC)

The model of frequency domain radar return of Eq. (1) is rewritten as

$$X(f) = \sum_{k=1}^{K^x} \sigma_k^x \exp\left(j2\pi\frac{2r_k^x}{c}(f + f_0) \right) \tag{3}$$

where the superscript x is used to denote the associated data record. Taking Fourier transform of each side of Eq. (3) yields the range domain signal

$$x(r) = \sum_{k=1}^{K^x} A_k^x w(r - r_k^x) \tag{4}$$

where $x(r) = x_1(2r/c)$ with $x_1(r)$ being Fourier transform of $X(f)$, $A_k^x = \sigma_k^x \exp(j4\pi f_0 r_k^x/c)$, and $w(r) = w_1(2r/c)$ with $w_1(r)$ being Fourier transform of the frequency domain window function. Multiplying $x(r)$ by its conjugate and dropping the cross-term yields

$$\tilde{x}(r) \overset{\text{Def}}{=} |x(r)|^2 - \sum_{k=1}^{K^x}\sum_{\substack{l=1 \\ l\neq k}}^{K^x} A_k^x A_l^{x*} w(r - r_k^x)w^*(r - r_l^x)$$

$$= \sum_{k=1}^{K^x} \tilde{\sigma}_k^x |w(r - r_k^x)|^2 \tag{5}$$

where $\tilde{x}(r)$ is the cross-term free HRRP, * denotes the complex conjugate, and $\tilde{\sigma}_k^x = |A_k^x|^2 = |\sigma_k^x|^2$. For two cross-term free HRRP's $\tilde{x}(r)$ and $\tilde{y}(r)$, we define

$$C(x,y) \overset{\text{Def}}{=} \sum_{p=1}^{K^x} \tilde{x}(r_p^x)\tilde{y}(r_p^x) \Big/ |w(0)|^2$$

$$- \sum_{p=1}^{K^x}\sum_{\substack{l=1 \\ l\neq p}}^{K^x}\sum_{k=1}^{K^y} \tilde{\sigma}_l^k \tilde{\sigma}_k^y |w(r_p^x - r_l^x)|^2 |w(r_p^x - r_k^y)|^2 \Big/ |w(0)|^2$$

$$= \sum_{l=1}^{K^x}\sum_{k=1}^{K^y} \tilde{\sigma}_l^x \tilde{\sigma}_k^y |w(r_l^x - r_k^y)|^2. \tag{6}$$

Note that the cross-terms are again dropped in Eq. (6). Obviously, $C(y,x) = C(x,y)$. Define

$$\rho(x,y) \stackrel{\text{Def}}{=} \frac{C(x,y)}{\sqrt{C(x,x)C(y,y)}} \tag{7}$$

$\rho(x,y)$ is referred to as generalized-weighted-normalized correlation (GWNC), with the weights $|w(r_{x,l} - r_{y,k})|^2$ being a function of relative distance between scatterers. Just as normalized correlation (NC) is used to measure the similarity of HRRP's, GWNC can be used to measure the similarity of PHRRP's, each of which is designated as $\{\tilde{\sigma}_k^x, r_k^x\}_{k=1}^{K^x}$, with σ_k^x, r_k^x and K^x defined to be the same as in Eqs. (3) and (5). It should be noted that the choice of a proper mainlobe width for $w(r)$ is important in the application of GWNC. With a wide mainlobe, GWNC is robust to the azimuth variation. However, GWNC will lose its discriminating power if the mainlobe is too wide. From our experiences, the mainlobe width can be chosen to be 2–5 times the width of a range cell achieved by DFT .

It is easily seen from Eqs. (6) and (7) that GWNC is efficient in computation when the number of scatterers is not too large. For PHRRP templates $\{\tilde{\sigma}_k^x, r_k^x\}_{k=1}^{K^x}$, $C(x,x)$ can be computed offline and stored in memory. For an unknown PHRRP $\{\tilde{\sigma}_k^y, r_k^y\}_{k=1}^{K^y}$, $C(y,y)$ needs to be computed only once before it is matched against all templates. Thus the computation of GWNC can be further reduced. The storage efficiency of GWNC is obvious, as there are only $2K$ data points for K scatterers, and this is much smaller than the data points of a HRRP.

To compensate for the translational range shift, the numerator of Eq. (7) should be modified as

$$C(x,y) = \frac{\max}{\Delta r} \sum_{l=1}^{K^x} \sum_{k=1}^{K^y} \tilde{\sigma}_l^x \tilde{\sigma}_k^y |w(\Delta r + r_l^x - r_k^y)|^2 . \tag{8}$$

In practice, the search for Δr can be done in the neighborhood of $\overline{\Delta r}$, where

$$\overline{\Delta r} = \sum_{k=1}^{K^y} \tilde{\sigma}_k^y r_k^y \Big/ \sum_{k=1}^{K^y} \tilde{\sigma}_k^y - \sum_{l=1}^{K^x} \tilde{\sigma}_l^x r_l^x \Big/ \sum_{l=1}^{K^x} \tilde{\sigma}_l^x . \tag{9}$$

4. EXPERIMENTAL RESULTS

We evaluate the performance of GWNC using the dataset of three scaled aircraft models: B52 (scale 1:91), Q6 (scale 1:20) and Q7 (scale 1:15). The raw data are collected by placing the scaled models on a turntable in a microwave anechoic chamber and measuring the radar returns at stepped frequencies ranging from 12 GHz to 18 GHz with a 0.06 GHz increment. Azimuthal angles of the measurements are from 0° to 155° (0° is the nose-on azimuth) with an average increment of 0.43°. The elevation angles remain constant at 5°. The PHRRP's are obtained by extracting the parameters of K most dominant scatterers from the stepped-frequency measurements using the RELAX algorithm.[5] The corresponding HRRP's are obtained via FFT of the stepped-frequency measurements.
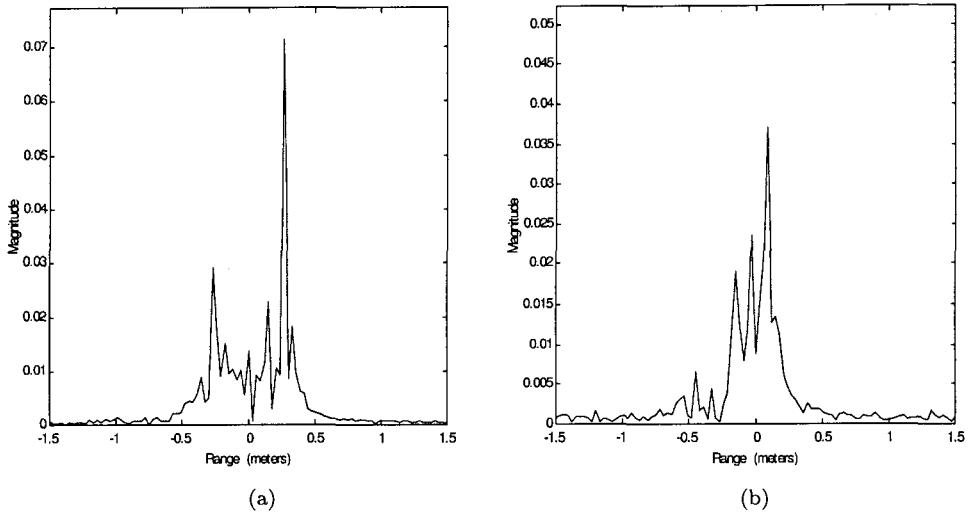
Fig. 1.   Examples of HRRP's of the aircraft models. (a) B52, at the azimuth of 48.5°; (b) Q6, at the azimuth of 59.5°.
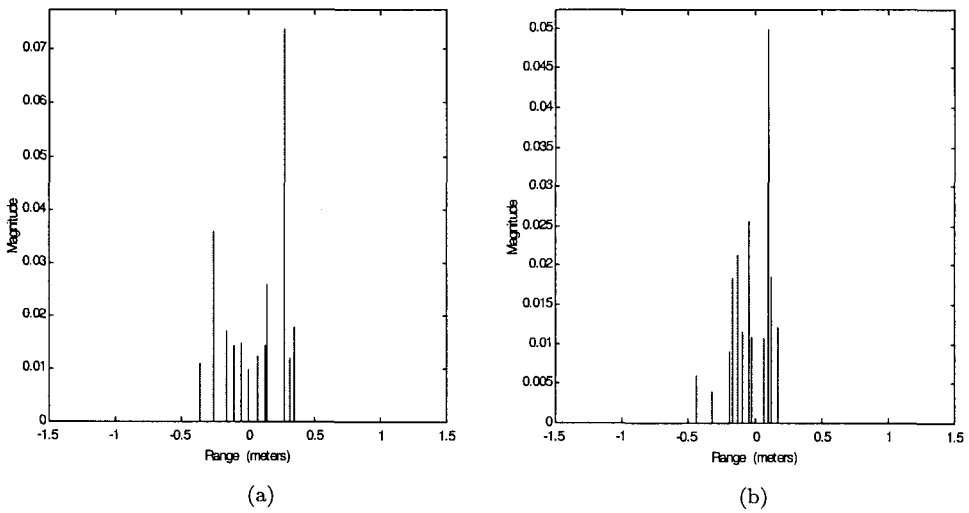

Fig. 2.   Examples of PHRRP's of the aircraft models. (a) B52, at the azimuth of 48.5°; (b) Q6, at the azimuth of 59.5°.

Figure 1 gives some examples of HRRP's of the three aircraft models and Fig. 2 gives the corresponding PHRRP's. The abscissas range in meters. It is seen that the PHRRP is a good representation of the dominant scatterers as indicated by the peaks in the corresponding HRRP.

Figure 3 gives the NC of HRRP's and GWNC of PHRRP's, respectively, of B52 for all three targets. The abscissas are azimuth variation in degrees. Figure 4 gives similar plots of the Q6 aircraft. In Figs. 3(b) and 4(b), the number of most
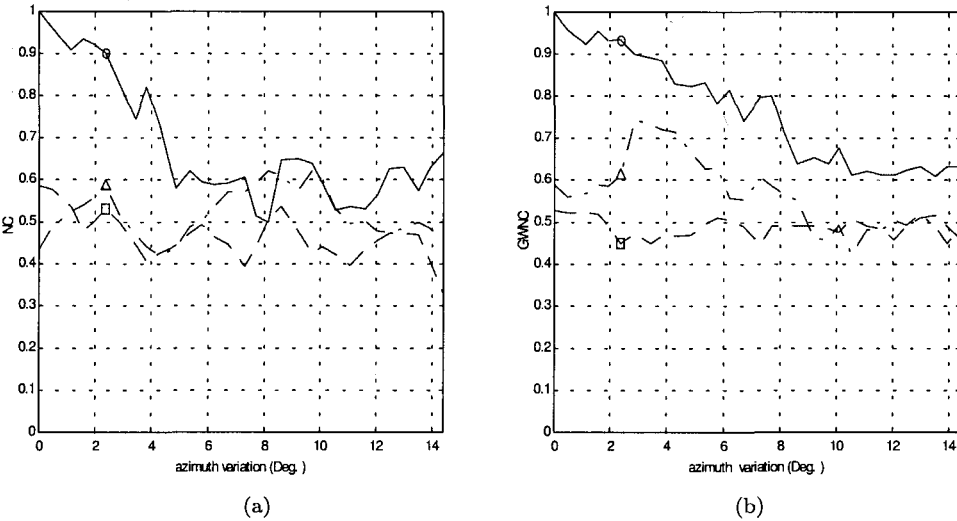
Fig. 3.   (a) NC between the HRRP's of B52 and those of all three targets.   (b) GWNC between the PHRRP's of B52 and those of all three targets — Solid with "o": B52, Dashed with "□": Q6, Dotted with "△": Q7. Reference azimuth: 48.5°.
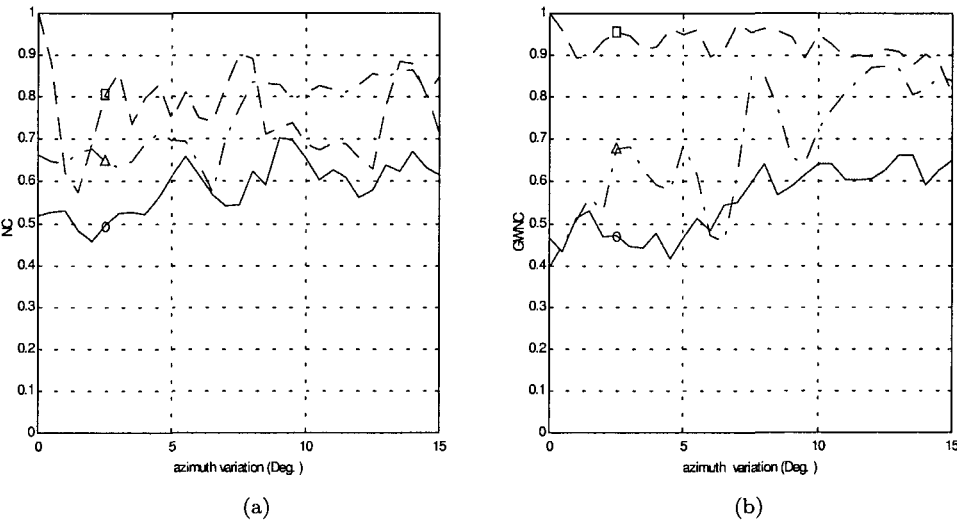

Fig. 4.   (a) NC between the HRRP's of Q6 and those of all three targets.   (b) GWNC between the PHRRP's of Q6 and those of all three targets — Solid with "o": B52, Dashed with "□": Q6, Dotted with "△": Q7. Reference azimuth: 59.5°.

dominant scatterers are all 12. Since translational range shifts do not occur for turntable targets, the GWNC is computed using Eqs. (6) and (7), with $w(r)$ is chosen as Fourier transform of the Hamming window function to reduce the effect of sidelobes. The mainlobe width of $w(r)$ is chosen to be three times the width of a

range cell achieved by DFT. It is seen from Figs. 3 and 4 that the azimuth sensitivity of PHRRP's can be reduced without sacrificing their discriminative power if the mainlobe width of $w(r)$ is chosen properly.

Now we evaluate the performance of GWNC via comparison of classification rates. We construct the template set for PHRRP (or HRRP) by fetching PHRRP's (or HRRP's) from every 4.7° azimuths starting from the 2° azimuth, thus obtaining a total of 99 templates for PHRRP (or HRRP) over the 0°–155° azimuth range for the three aircraft models. An unknown PHRRP (or HRRP) is classified by first computing its GWNC (or NC) with all PHRRP (or HRRP) templates and then assigning it to the class of which a template has the maximum GWNC (or NC) with the unknown PHRRP (or HRRP). In computation of GWNC, $w(r)$ is again chosen as Fourier transform of the Hamming window function with its mainlobe width chosen to be three times the width of a range cell achieved by DFT. The classification results are summarized in Table 1.

It is seen from Table 1 that the classification rates achieved by PHRRP's increase rapidly as an increased number of dominant scatterers are used and the classification rates achieved by only 12 dominant scatterers are already comparable to those achieved by HRRP's, thus demonstrating the effectiveness of GWNC. The number of data points for a HRRP is 101 and it can be reduced to 64 by discarding the range cells at each end, which carry little target features. On the other hand, there are only $2K$ data points for K most dominant scatterers. This results in a 62% data reduction for 12 dominant scatterers.

Table 1.   Correct classification rates in percentage.

|		Correct classification rates (%)			
		B52	Q6	Q7	Average
NC applied to HRRP's		88	89	91	89
GWNC	$K = 4$	77	80	79	79
applied to	$K = 8$	89	89	78	85
PHRRP's	$K = 12$	94	88	84	89

5. CONCLUSIONS

In this paper we have presented a new scheme for radar target recognition based on parameterized high resolution range profiles (PHRRP). We have proposed a novel criterion called generalized-weighted-normalized correlation (GWNC), which can be used to measure the similarity between PHRRP's. With a properly chosen mainlobe width of GWNC, the PHRRP's azimuth sensitivity can be reduced and yet their discriminative power does not suffer. The experimental results with the dataset

of three scaled aircraft models show that by using GWNC only a small number of most dominant scatterers can achieve the recognition rates as good as those of HRRP's, thus leading to a significant data reduction for the recognition system.

REFERENCES

1. R. Carriere and R. L. Moses, "High resolution radar target modeling using modified prony estimator," *IEEE Trans. Antennas and Propagation* **40**, 1 (1992) 13–18.
2. S. Hudson and D. Psaltis, "Correlation filters for aircraft identification from radar range profiles," *IEEE Trans. Aerosp. Electron. Syst.* **29**, 3 (1993) 741–748.
3. Q. Li, E. J. Rothwell, K. M. Chen and D. P. Nyquist, "Radar target discrimination schemes using time-domain and frequency-domain methods for reduced data storage," *IEEE Trans. Antennas and Propagation* **45**, 6 (1997) 995–1000.
4. J. Li and P. Stoica, "Efficient mixed-spectrum estimation with application to target feature extraction," *IEEE Trans. Sign. Process.* **44**, 2 (1996) 281–295.
5. H. J. Li and S. H. Yang, "Using range profiles as feature vectors to identify aerospace objects," *IEEE Trans. Antennas and Propagation* **41**, 3 (1993) 261–268.
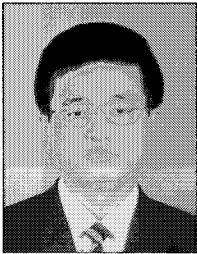
Xuejun Liao received the B.S. and M.S. degrees in electrical engineering, from Hunan University, China, in 1990 and 1993, respectively, and Ph.D. in electrical engineering from Xidian University, China, in 1999. He is currently with the National Key Lab for Radar Signal Processing, Xidian University.

His current research interests are in the area of radar imaging and target recognition, array signal processing and statistical signal processing.
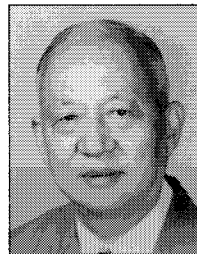
Zheng Bao graduated from the Communication Engineering Institute of China in 1953. Currently he is a professor at Xidian University and a member of the Chinese Academy of Science.

His research fields are radar systems, signal processing and circuit theory.