Communications

Automated Detection of Subpixel Hyperspectral Targets With Adaptive Multichannel Discrete Wavelet Transform

Lori Mann Bruce, Jiang Li, and Yan Huang

***Abstract*—This paper introduces the use of adaptive multichannel discrete wavelet transforms (AMDWTs), allowing for a customized design of optimum mother wavelets, for the detection of a constituent absorption band within a hyperspectral curve. Even when the target's amplitude is only 3% of the background signal's amplitude, the AMDWT approach produces target detection rates of 90%.**

I. Introduction

With the use of hyperspectral imagery, the ability to reliably detect subpixel targets through remote sensing is becoming a reality. A major step toward this goal is the ability to detect the constituent absorption bands that comprise a hyperspectral reflectance curve. Various methods for analyzing hyperspectral curves are being developed and a number of these methods are based on multiresolutional derivative analysis [1]–[4]. More recently, the methods using the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT) were investigated as a means of systematically analyzing hyperspectral curves via windows of varying width [5]. By appropriately selecting the mother wavelet, the method could be used to efficiently implement the smoothing and derivative operations in a multiresolutional manner. The wavelet transforms were utilized as a preprocessing tool for feature extraction in an automated hyperspectral target detection system, where the target was a low-amplitude constituent absorption band. Derivative-of-Gaussian mother wavelets were utilized with the CWT and a variety of standard, commonly used mother wavelets were investigated for the DWT. It was shown that the CWT method outperformed the DWT method in terms of target detection accuracy, but the DWT method was computationally more efficient. Also, the mother wavelets investigated for the DWT system were not optimized; they were simply drawn from a pool of commonly used mother wavelets. One cannot say that the DWT results were optimum since there are infinitely many mother wavelets from which to choose. Thus, for this application, as well as many others, there is a great need for DWT implementations where the mother wavelet is optimized for the specific task at hand.

This paper investigates the use of an adaptive multichannel discrete wavelet transform (AMDWT) for the detection of the same subpixel targets as reported in [5]. For a wavelet transform to be "adaptive," either: 1) the mother wavelet is varied in order to minimize an error criterion or 2) the basis formed from a preselected mother wavelet, such as which resolutions or scales are utilized, is varied in order to minimize an error criterion. In this study, case 1) is investigated. Similar to the previous work in [5], the AMDWT is used as a means of feature extraction and traditional statistical methods are used to classify the input data as "target present" or "no target present." The methods are tested on a database of Hyperspectral Digital Image Collection Experiment (HYDICE) curves. The results from using the AMDWT are compared to those of the CWT- and DWT-based methods.

Manuscript received June 1, 2001; revised December 13, 2001.

L. M. Bruce and J. Li are with the Electrical and Computer Engineering Department, Mississippi State University, Mississippi State, MS 39762–9571 USA (e-mail: bruce@ee.msstate.edu).

Y. Huang is with Veritas DGC, Inc., Houston, TX 77072 USA.

Publisher Item Identifier S 0196-2892(02)04603-X.

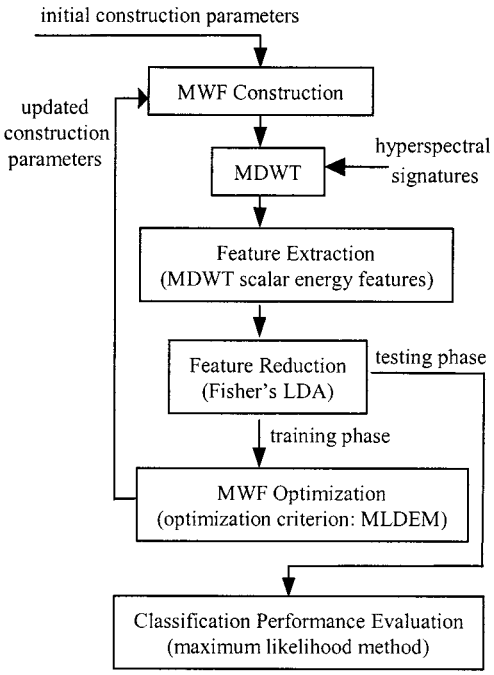

Fig. 1. AMDWT-based automated detection system block diagram.

II. Methodologies

An adaptive multichannel wavelet-based automated detection system is designed and tested for detecting low-amplitude constituent bands in hyperspectral curves. The adaptive system customizes a set of parameterized multichannel wavelet filters (MWFs) through minimizing a user-defined error criterion during the system training phase. The results are a set of optimum MWFs, based on the training data. During the system testing phase, the optimum MWF is used to implement the multichannel discrete wavelet transform (MDWT) of the hyperspectral curves and wavelet coefficient scalar energy features are extracted. The performance of the wavelet-based features is evaluated via the maximum-likelihood classification accuracy. A block diagram of the overall system is shown in Fig. 1. The system is trained and tested using a jackknife procedure on a database of 1000 HYDICE signals where half contain a subpixel target, or additive Gaussian constituent band. The amplitude, or abundance, of the target is varied between 1% and 10% of the amplitude of the background hyperspectral curve. The databases of HYDICE signals are reported in more detail in [5].

A. MWF Construction and MDWT

The first step of the adaptive system is to construct parameterized mother wavelets (or equivalently wavelet filter banks [6], [7]). Fundamental theories and methods of designing M-channel wavelet filters, with arbitrary M, have been investigated in [8]–[11]. The applications of M-channel wavelet filters to feature extraction and classification problems have been reported in [12] and [13]. For this study, the methods presented in [11] and [13] are utilized for

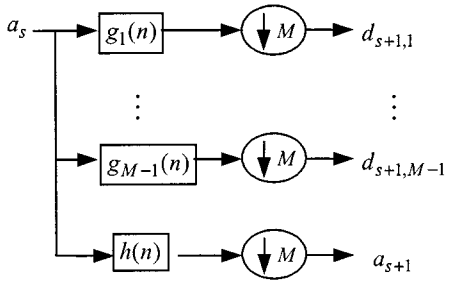

Fig. 2. Filter tree implementation of MDWT.

constructing MWF. The basic idea behind the construction of MWF is a process of shift product and factorization of wavelet matrices. The result is an orthogonal MWF matrix, which is a function of a set of parameter vectors, $\vec{u}_i$ and $\vec{v}$. The first row of an MWF matrix corresponds to a low-pass filter impulse response $h(n)$ and the remaining rows of the MWF matrix correspond to a series of band-pass and high-pass filters, $g_j(n)$, respectively, where $n = 0, 1, \ldots, N-1$ and $j = 1, 2, \ldots, M-1$. The variables M and N represent the number of wavelet filter channels and the filter length, respectively. Therefore, given a set of parameter vectors $\vec{u}_i$ and $\vec{v}$, a set of M-channel wavelet filter banks, $h(n)$ and $g_j(n)$, are determined. In general, there exists a relationship among the number of $\vec{u}_i$ parameters q and variables M and N, such that $N = M \cdot (q+1)$.

With the M-channel filter bank $h(n)$ and $g_j(n)$, an M-channel discrete wavelet transform can be implemented in a recursive manner as shown in Fig. 2. At each decomposition level (or scale), the filters are followed by a decimation with factor M. As a result, a set of wavelet detail coefficients $d_{s,j}$ and wavelet approximation coefficients a_s are obtained. Note that when the level $s = 0$, $a_s = a_0$ represents the original signal. Mathematically, these operations can be described as [13]

$$a_{s+1}(k) = \sum_{n=0}^{N-1} h(n) \cdot a_s(M \cdot k + n)$$
$$d_{s+1,j}(k) = \sum_{n=0}^{N-1} g_j(n) \cdot a_s(M \cdot k + n). \quad (1)$$

Notice that, when the number of channels is two, $M = 2$, the recursive operations reduce to the case of the standard dyadic DWT. For background information on wavelets and filter banks, the reader is referred to [7], [9], and [14].

B. Feature Extraction and Reduction

In this study, the wavelet transform coefficient scalar energy is computed and utilized as classification features. The orthogonal wavelet transform can be viewed as an energy partitioning process of the original digital signal [15]. The energy of the wavelet coefficients can be utilized as a feature that indicates how the original signal's energy is partitioned according to scale or resolution. For this study, we form a feature vector, $\vec{f}$, based on the calculation of MDWT coefficient scalar energies

$$\vec{f} = [E(a_L), E(d_{s,j})]^T \quad (2)$$

where $E(\cdot)$ is an operator of energy calculation, the superscript T represents the vector transpose, a_L are the MDWT approximation coefficients at the final decomposition level and $d_{s,j}$ are the MDWT detail coefficients at each decomposition level ($s = 1, 2, \ldots, L$) and in each filter channel ($j = 1, 2, \ldots, M-1$).

For hyperspectral data analysis, the feature extraction process is also a dimensionality reduction process. The HYDICE signatures used in this study consist of 210 spectral bands, i.e., the HYDICE data have a dimensionality of 210. After the MDWT-based feature extraction, according to (2), the dimensionality is reduced to $1 + L \cdot (M-1)$. For example, the dimensionality is reduced to 10 when using a four-channel MWF ($M = 4$) and a three-level decomposition ($L = 3$). The dimensionality of the feature vector could be reduced further using Fisher's linear discriminant analysis (LDA) method. The output from the LDA is an optimal linear weight matrix, W, in the sense of maximizing the interclass variance and minimizing the intraclass variance [16]. The weight matrix has a size of $D \times (C-1)$, where $D = 1 + L \cdot (M-1)$ is the dimensionality of the feature vector $\vec{f}$ and C is the number of classes. With the weight matrix, the reduced feature vector $\vec{f}_r$ is computed as an optimal linear combination of elements in the original feature vector $\vec{f}$ as

$$\vec{f}_r = W^T \cdot \vec{f}. \quad (3)$$

Note that the reduced feature vector $\vec{f}_r$ has a size of $(C-1) \times 1$. For a two-class problem, the feature vector can be reduced to a scalar value.

For the CWT and the DWT cases in [5], the features were extracted from a region of interest (ROI) within the hyperspectral curve, as well as from all of the hyperspectral curve. The "all" case represented the scenario where the system had no *a priori* information as to potential spectral location of the target's absorption band. For this study of adaptive wavelet transforms, only the "all" case was investigated, since it represented the most difficult scenario and was the case for which the CWT and the DWT methods resulted in the lowest target detection accuracies.

C. System Training and Testing

Based on the reduced MDWT features, an optimization operation or classification operation is performed when the system is being trained or tested, respectively. When conducting the optimization operation, the modified linear discriminant error measure (MLDEM) is utilized as an optimization criterion. The authors designed this measure based on the Fisher's linear discriminant analysis method. The MLDEM is defined as a ratio

$$MLDEM = \frac{\sigma_1^2 + \sigma_2^2}{(\mu_1 - \mu_2)^2} \quad (4)$$

where μ_1, σ_1^2, μ_2, and σ_2^2 are the means and variances of Class 1 and Class 2, respectively. The optimization operation is to minimize the value of the MLDEM, which implies maximizing the interclass variance and minimizing the intraclass variance. During each optimization iteration, the MLDEM is computed and minimized and at the same time the parameter vectors constructing the MWF, $\vec{u}_i$ and $\vec{v}$, are updated. The adaptive iteration process continues until the MLDEM is minimized. Various standard optimization methods [17] exist for implementing the minimization. In this study, a quasi-Newton algorithm incorporated by a mixed quadratic and cubic line search algorithm is used for MLDEM minimization. Note that this is a nonlinear optimization problem, thus it is possible to result in a local minimum, rather than a global minimum. This problem could be solved in an experimental manner. One way is to test several sets of initial parameters to try to obtain a solution that is reasonably close to the global minimum. Another way is to initially set a satisfactory criterion for the system performance, e.g., if the target detection accuracy (or the improvement of the detection accuracy) is satisfactory, the solution is regarded as optimal. The second method is a more application-dependent approach and is used for this study.

The optimization process results in a set of optimum parameter vectors, $\vec{u}_i$ and $\vec{v}$, as well as the optimum weight matrix, W, for feature reduction. Utilizing $\vec{u}_i$ and $\vec{v}$, a set of optimum MWF can be constructed,

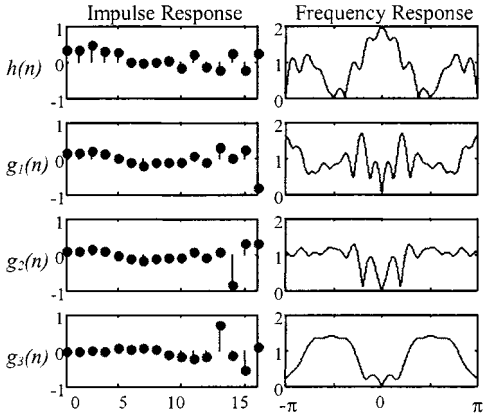

Fig. 3. Filter impulse response and amplitude of the frequency response of optimal $mwf-4-16$ for DM1, 3% positive target amplitude.

and, using W, the system can be used for classification. The performance of the AMDWT-based automated detection system is evaluated using maximum likelihood classification accuracies, which is the same method as reported in [5]. Thus, one can directly compare the system performance of using the CWT, DWT, and AMDWT features.

III. Results and Discussion

The HYDICE data matrices (DMs) utilized in this study are the same as those reported in [5]. They include the following: DM1: positive amplitude target with width $\sigma=\sqrt{7}$ centered at $\lambda_0=750$ nm; DM2: negative amplitude target with width $\sigma=\sqrt{7}$ centered at $\lambda_0=750$ nm; DM3: negative amplitude target with varying width centered at $\lambda_0=750$ nm. The signal-to-noise ratios (SNRs) for these databases range from -51 dB to -98 dB. This shows how insignificant the target is compared to the background clutter of the hyperspectral curve. Six different MWFs are adaptively constructed. The name of the MWF is denoted as $mwf-M-N$, where M is the number of wavelet filter channels and N is the wavelet filter length. As an example, Fig. 3 shows both the impulse response and the amplitude of the frequency response of $mwf-4-16$ for the DM1 with a positive 3% amplitude subpixel target.

The maximum likelihood classification accuracies resulting from the AMDWT are summarized in Table I. For the purpose of comparing system classification performance, pertinent results from the CWT and DWT [5] are also included in Table I. These results are: 1) the best result form the CWT using the seventh Gaussian-derivative mother wavelet (Gauss-7); 2) one of the best results from the DWT using Haar mother wavelet; and 3) results from the DWT using Daubechies-3 (Daub-3) and Daubechies-5 (Daub-5) mother wavelets, which have the same filter length as $mwf-2-6$ and $mwf-2-10$, respectively.

General observation indicates that using the AMDWT leads to a sizable improvement in classification accuracy as compared to the DWT and CWT. In general, at least a 10% increase in classification accuracy can be gained from using the AMDWT, regardless of the target's amplitude or variance. This is especially true when the MWFs have more than two channels, i.e., the multichannel wavelet analysis plays a key role in improving the target detection performance. This is reasonable and understandable since the MDWT conducts multiresolution analysis in a more refined and comprehensive way.

Moreover, the adaptation also plays a key role in improving the system performance. This is verified by comparing results from the two-channel MDWT and standard DWT when using the same order filters, e.g., $mwf-2-6$ versus Daub-3 and $mwf-2-10$ versus Daub-5. Taking the DM1 with a positive 3% amplitude subpixel target as an example, an improvement of 12%–30% in the classification accuracy is obtained. Similar observations can be made for various amplitude and variance subpixel targets in DM1, DM2, and DM3.

TABLE I
Maximum Likelihood Classification Accuracy in Percentage Using AMDWT, DWT, and CWT. DM1: Positive Amplitude Target With $\sigma=\sqrt{7}$ and $\lambda_0=750$ nm; DM2: Negative Amplitude Target With $\sigma=\sqrt{7}$ and $\lambda_0=750$ nm; DM3: Negative Amplitude Target With Varying Width and $\lambda_0=750$ nm

		Target Amplitude			
		10%	5%	3%	1%
DM1	mwf-4-16	99.6	97.6	92.2	65.6
	mwf-4-8	97.0	91.8	83.0	65.2
	mwf-3-15	98.4	95.6	78.8	59.2
	mwf-3-9	96.8	95.0	71.2	57.8
	mwf-2-10	97.6	93.2	79.6	59.2
	mwf-2-6	91.2	82.4	67.6	51.0
	Daub-5	76.8	58.8	57.6	48.6
	Daub-3	84.4	67.2	56.4	48.4
	Haar	75.8	67.8	55.4	50.2
	Gauss-7	94.2	86.4	79.0	57.4
DM2	mwf-4-16	97.4	96.0	88.2	66.4
	mwf-4-8	94.6	91.0	85.0	56.4
	mwf-3-15	97.0	92.0	87.6	56.4
	mwf-3-9	96.6	91.4	85.0	60.6
	mwf-2-10	95.8	89.4	84.6	61.6
	mwf-2-6	89.6	81.2	67.8	55.2
	Daub-5	53.6	52.6	50.6	50.8
	Daub-3	86.6	70.6	55.4	50.0
	Haar	80.4	66.2	56.0	49.4
	Gauss-7	91.6	88.6	79.6	55.4
DM3	mwf-4-16	98.8	97.0	92.2	64.2
	mwf-4-8	96.6	85.4	68.4	58.4
	mwf-3-15	98.2	93.4	85.2	57.2
	mwf-3-9	97.4	81.8	67.6	54.8
	mwf-2-10	96.6	86.4	77.4	59.6
	mwf-2-6	93.8	77.2	65.0	57.8
	Daub-5	48.8	50.8	50.8	51.4
	Daub-3	87.2	66.8	52.6	50.2
	Haar	81.8	66.8	55.0	50.6
	Gauss-7	93.0	86.2	76.8	55.6

The classification results vary among the six different MWF investigated. General observation shows that the number of channels and the length of the filters are proportional to the classification accuracies. For example, $mwf-4-16$ generally outperforms $mwf-3-15$ and $mwf-3-9$ and $mwf-3-15$ generally outperforms $mwf-2-10$ and $mwf-2-6$. This trend could be resulting from the fact that more channels and longer filters provide a larger freedom, or larger dimensionality of search space, to adaptively adjust and optimize the MWF. Recall that the training and testing data are jackknifed such that they are mutually exclusive. Also, the data is drawn from a highly variable HYDICE scene containing a variety of natural and manmade objects. This was done to safeguard against overtraining in the adaptation process.

IV. Conclusion

In this paper, the authors extend the use of wavelet transforms for hyperspectral feature extraction and subpixel target detection to include adaptive wavelet transforms, specifically the adaptive multichannel discrete wavelet transform (AMDWT). The AMDWT consistently outperformed the DWT and the CWT methods used in [5], increasing the target detection accuracies by at least 10% regardless of target amplitude or width. For the cases where the target

amplitude was only 3%, as compared to the background hyperspectral curve, the AMDWT produced classification accuracies $> 90\%$. This result is particularly impressive when considering the hyperspectral curve's SNR is ≈ -75 dB.

The increased accuracy of the target detection system does not come without a price, however. The training of the system, as compared to the DWT and CWT methods, is much more computationally expensive. The process of adaptively constructing the mother wavelet is time-consuming. However, once the optimum mother wavelet is designed, the implementation of the AMDWT is very similar to the DWT and the computational expense is comparable. For example, assume the AMDWT system were to be used for the detection of a particular target. With the use of training data, the AMDWT system could be used to construct the optimum mother wavelet beforehand. Then during the actual usage of the system, the MDWT could be implemented with a subband filter bank either in software or specialized hardware, so the additional computations required for the AMDWT could be completed offline beforehand. With this in mind, the AMDWT method is considered to be superior to the CWT and the DWT methods due to its much increased target detection accuracies.

REFERENCES

[1] F. Tsai and W. Philpot, "Derivative analysis of hyperspectral data," *Remote Sens. Environ.*, vol. 66, pp. 41–51, Oct. 1998.

[2] Savitzky and N. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1638, July 1964.

[3] T. H. Demetriades-Shah, M. D. Steven, and A. C. Clark, "High resolution derivative spectra in remote sensing," *Remote Sens. Environ.*, vol. 33, pp. 55–64, July 1990.

[4] M. A. Piech and K. R. Piech, "Symbolic representation of hyperspectral data," *Appl. Opt.*, vol. 26, no. 18, pp. 4018–4026, Sept. 1987.

[5] L. Bruce, C. Morgan, and S. Larsen, "Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 2217–2226, Oct. 2001.

[6] M. Vetterli and C. Herley, "Wavelet and filer banks: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992.

[7] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Boston, MA: Wellesley-Cambridge Press, 1996.

[8] P. P. Vaidyanathan, "Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having the perfect-reconstruction property," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 476–492, 1987.

[9] ——, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[10] P. Steffen, P. Heller, R. Gopinath, and C. Burrus, "Theory of regular M-band wavelet bases," *IEEE Trans. Signal Processing*, vol. 41, pp. 3497–3511, Dec. 1993.

[11] R. Turcajova and J. Kautsky, "Shift products and factorization of wavelet matrices," *Numer. Alg.*, vol. 8, pp. 27–54, 1994.

[12] W. Qian, L. Clarke, H. Li, and R. Clark, "Digital mammography: M-channel quadrature mirror filters for microcalcification extraction," *Comput. Med. Imag. Graph.*, vol. 18, no. 5, pp. 301–314, 1994.

[13] Y. Mallet, D. Coomans, J. Kautsky, and O. De Vel, "Classification using adaptive wavelets for feature extraction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 1058–1066, Oct. 1997.

[14] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[15] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[16] A. Webb, *Statistical Pattern Recognition*. London, U.K.: Arnold, 1999.

[17] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.

A Novel Multifractal Estimation Method and Its Application to Remote Image Segmentation

G. Du and T. S. Yeo

***Abstract*—Based on the gliding-box and relative differential box-counting algorithms, a novel method that estimates accurately the multifractal exponents, a distinct characteristics of gray-scale digital images, is proposed. Four natural texture images are used to test the performance of the novel multifractal measure. Comparisons with published methods show that the proposed method can efficiently describe texture images and can provide accurate classification results.**

***Index Terms*—Image segmentation, multispectral estimation, synthetic aperture radar.**

I. INTRODUCTION

Multifractal analysis has gain popularity in image analysis during the recent years. The multifractal parameters of physical phenomena have been shown [1] to be distinct with respect to the texture contents of the images, and yet they are consistent across sensing platforms. Multifractal analysis has also been known to produce fairly accurate classification results even in cases where the image does not possess ideal fractal properties. In remote sensing, multifractal analysis has been successfully applied in several applications such as the analysis and classification of land [2] and sea-ice [3].

Several algorithms have been proposed for the computation of multifractal parameters. One of the commonly used methods was presented by Chaudhuri and Sarkar [4]. The method was based on the differential box-counting (DBC) algorithm and has been shown to produce fairly good results. In this paper, we propose a novel modification to [4], resulting in better accuracy and efficiency.

II. MULTIFRACTAL MEASURE

The classical multifractal estimation method proposed by Chaudhuri and Sarkar [4] can be described as follows. Consider an image of size $M \times M$ which has been scaled down to size $s \times s$, where s ranges between 2 to $M/2$. The image space is partitioned into multiple grids, each of size $s \times s$. On each grid there is a column of boxes of size $s \times s \times s$. Let the maximum and minimum pixel values in the (i, j)th grid fall in box number v and u, respectively. The relative box index n_r, a descretized version of the relative pixel strength, can be defined as

$$n_r(i, j) = v - u + 1. \tag{1}$$

Taking contribution from all grids (i.e., counting all boxes), we have

$$N_r = \sum_{i, j} n_r(i, j). \tag{2}$$

Manuscript received August 20, 2001; revised December 4, 2001.

The authors are with the Electrical and Computer Engineering Department, National University of Singapore, Singapore 119260 (e-mail: eleyeots@nus.edu.sg).

Publisher Item Identifier S 0196-2892(02)04623-5.