N95- 23887

OPTIMAL BAND SELECTION FOR DIMENSIONALITY REDUCTION OF HYPERSPECTRAL IMAGERY

Stephen D. Stearns, Bruce E. Wilson and James R. Peterson

ESL Incorporated, a subsidiary of TRW
P.O. Box 3510
Sunnyvale, CA 94088-3510
steve_stearns@smtp.esl.com

1. INTRODUCTION

Hyperspectral images have many bands requiring significant computational power for machine interpretation. During image pre-processing, regions of interest that warrant full examination need to be identified quickly. One technique for speeding up the processing is to use only a small subset of bands to determine the "interesting" regions. The problem addressed here is how to determine the fewest bands required to achieve a specified performance goal for pixel classification. The band selection problem has been addressed previously by Chen et al. (1987, 1988, 1989), Ghassemian et al. (1988), Henderson et al. (1989), and Kim et al. (1990).

Some popular techniques for reducing the dimensionality of a feature space, such as principal components analysis, reduce dimensionality by computing new features that are linear combinations of the original features. However, such approaches require measuring and processing all the available bands before the dimensionality is reduced. Our approach, adapted from previous multidimensional signal analysis research, is simpler and achieves dimensionality reduction by selecting bands. Feature selection algorithms are used to determine which combination of bands has the lowest probability of pixel misclassification. Two elements required by this approach are a choice of objective function and a choice of search strategy.

2. OBJECTIVE FUNCTIONS

A variety of objective functions have been proposed for feature selection optimization, including the Shannon equivocation $H(\Omega \mid X)$, the Shannon mutual information $I(\Omega; X)$ that the feature vector X gives about the class Ω, the Bhattacharyya distance $B(\Omega, X)$, and the divergence $J(\Omega, X)$. The latter two quantities are defined for classes taken in pairs.

The two-class Bayes error probability P_e^* is bounded above and below in terms of these quantities.

$$P_e^* \le \frac{1}{2}H(\Omega \mid X) \le \frac{1}{2}\log M - \frac{1}{2}I(\Omega; X) \le \frac{1}{2}\exp(-B),$$

$$\frac{1}{4}\exp(-J/2) \le \frac{1}{4}\exp(-2B) \le \frac{1}{2}\left(1 - \sqrt{1-\exp(-2B)}\right) \le P_e^*.$$

The M-class Bayes error probability is upper bounded by a weighted sum of the two-class Bayes error probabilities between all pairs of classes.

3. FEATURE SELECTION PARADOXES AND ALGORITHMS

The theory of feature selection is a history of the discovery of paradoxes and of increasingly sophisticated algorithms designed to overcome these paradoxes (Cover, 1974; Cover et al., 1977; Narendra et al, 1977).

The (m, n) feature selection algorithm was developed as a means to handle a large number of candidate features (Stearns, 1976). This technique avoids having to evaluate all possible combinations of 200 or more features. It is resistant to the feature selection paradoxes, although not immune. At present, it is one of the most powerful and practical methods for selecting near-optimal subsets of features from a large set of candidates, e.g. automatic selection of hyperspectral bands.

Recently, a variant of the (m, n) feature selection algorithm, called the Greedy (m, n) algorithm, was applied to the problem of determining minimal band sets for hyperspectral imagery. Experimental results are summarized here. A companion paper provides theoretical discussion (Stearns et al., 1993).

4. EXPERIMENTAL RESULTS

An experiment was performed to compare three algorithms for automatically selecting subsets of bands for pixel classification. The three algorithms compared were the Best Individual Features algorithm, the Forward Sequential or (1, 0) algorithm, and the Greedy (2, 1) algorithm. The data for the experiment was a 224-band AVIRIS scene of the southern San Francisco peninsula. Regions of the scene were selected and used to form a library of 224-element feature vectors for the six classes: open water, evaporation ponds, marsh, green grass, brown grasslands, and urban. Portions of the scene not belonging to these six classes were not used.

The objective function used by the three feature selection algorithms was the minimum Bhattacharyya distance between any two of the six classes. The minimum Bhattacharyya distances were converted to upper bounds on the Bayes error probability according to

$$P_e^* \leq \frac{1}{M} \sum_{i>k} \exp(-B_{i,k}) \leq \frac{M-1}{2} \exp(-\min\{B_{i,k}\}) .$$

The three feature selection algorithms determined sets of one through nine bands. The results are shown in Tables 1 - 3. The Best Individual Features algorithm did poorly. This algorithm does not consider feature interactions. Consequently, the bands selected by this algorithm are grouped around the best individual band, 139. Each additional band

Table 1. Bands Selected by the Best Individual Features Algorithm.

Number of Bands	Bands	Upper Bound on P_e^*
1	139	0.688
2	137, 139	0.623
3	137, 138, 139	0.598
4	137, 138, 139, 140	0.575
5	137, 138, 139, 140, 143	0.558
6	136, 137, 138, 139, 140, 143	0.452
7	136, 137, 138, 139, 140, 141, 143	0.417
8	136, 137, 138, 139, 140, 141, 143, 144	0.405
9	136, 137, 138, 139, 140, 141, 142, 143, 144	0.393

conveys little or no new information over that provided by the bands selected already. The class separability does not increase much as the number of bands increases.

The Greedy (2, 1) algorithm produces identical subsets of bands as the Forward Sequential algorithm for the first eight sets, but the search paths start to diverge at nine bands. The Greedy (2, 1) algorithm is expected to yield a superior band set for subset sizes greater than nine.

Table 2. Bands Selected by the Forward Sequential (1, 0) Algorithm.

Number of Bands	Bands	Upper Bound on P_e^*
1	139	0.688
2	120, 139	3.86×10^{-2}
3	33, 120, 139	1.43×10^{-2}
4	33, 120, 139, 154	5.74×10^{-4}
5	33, 120, 139, 154, 174	2.07×10^{-4}
6	33, 108, 120, 139, 154, 174	2.72×10^{-5}
7	33, 108, 120, 139, 154, 174, 217	6.90×10^{-6}
8	25, 33, 108, 120, 139, 154, 174, 217	7.20×10^{-7}
9	25, 33, 40, 108, 120, 139, 154, 174, 217	3.95×10^{-7}

Table 3. Bands Selected by the Greedy (2, 1) Algorithm.

Number of Bands	Bands	Upper Bound on P_e^*
1	139	0.688
2	120, 139	3.86×10^{-2}
3	33, 120, 139	1.43×10^{-2}
4	33, 120, 139, 154	5.74×10^{-4}
5	33, 120, 139, 154, 174	2.07×10^{-4}
6	33, 108, 120, 139, 154, 174	2.72×10^{-5}
7	33, 108, 120, 139, 154, 174, 217	6.90×10^{-6}
8	25, 33, 108, 120, 139, 154, 174, 217	7.20×10^{-7}
9	33, 40, 77, 108, 120, 139, 154, 174, 217	3.20×10^{-7}

5. SUMMARY AND CONCLUSIONS

Band selection has been shown here and elsewhere to be a practical method of data reduction for hyperspectral image data. Moreover, band selection has a number of advantages over linear band combining for reducing the dimensionality of high-dimensional data. Band selection eliminates the requirement that all bands be measured before data dimensionality is reduced. Bands that are uninformative about pixel classification need not be measured or communicated. Band sets can be tailored to specific classification goals (classes, error rates, etc.). Band selection reduces data link requirements, yet retains a tunable capability to collect as many bands as required for a specific application. Feature selection algorithms developed for statistical pattern classifier design can be used to perform band selection. The Greedy (2, 1) feature selection algorithm has been shown to be a practical means of selecting bands. In addition this algorithm has theoretical advantages over the Forward Sequential algorithm, making it the method of choice for hyperspectral applications.

6. ACKNOWLEDGEMENTS

The authors wish to thank David A. Landgrebe and Carl Salvaggio for helpful discussions that contributed to this paper.

7. REFERENCES

Cover T.M., 1974, "The best two independent measurements are not the two best," *IEEE Trans. Syst., Man, and Cybernetics*, (correspondence), vol. SMC-4, no. 1, pp. 116-117.

Cover T.M. and J.M. van Campenhout, 1977, "On the possible orderings in the measurement selection problem," *IEEE Trans. Syst., Man, and Cybernetics*, vol. SMC-7, no. 9., pp. 657-661.

Chen C.-C.T. and D.A. Landgrebe, 1987, "Spectral feature design for data compression in high dimensional multispectral data," *Proc. IEEE Int'l Geoscience and Remote Sensing Symp. (IGARSS'87)*, Ann Arbor, MI, pp. 685-690.

Chen C.-C.T. and D.A. Landgrebe, 1988, *Spectral feature design in high dimensional multispectral data*, Ph.D. thesis and Tech. Rep. TR-EE 88-35, School of Electrical Engineering, Purdue University, West Lafayette, IN, 140 pp.

Chen C.-C.T. and D.A. Landgrebe, 1988, "A spectral feature design system for high dimensional multispectral data," *Proc. IEEE Int'l Geoscience and Remote Sensing Symp. (IGARSS'88)*, Edinburgh, Scotland, pp. 891-894.

Chen C.-C.T. and D.A. Landgrebe, 1989, "A spectral feature design system for the HIRIS/MODIS era," *IEEE Trans. Geoscience and Remote Sensing*, vol. 27, no. 6, pp. 681-686.

Ghassemian H. and D.A. Landgrebe, 1988, "An unsupervised feature extraction method for high dimensional image data compaction," *Proc. IEEE Int'l Conf. on Syst., Man, and Cybernetics*, George Mason Univ., Alexandria, VA, October 20-23, 1987, pp. 540-544, also appeared in *IEEE Control Systems Magazine*, vol. 8, no. 3, pp. 42-48.

Henderson T.L., A. Szilagyi, M.F. Baumgardner, C.-C.T. Chen, and D. Landgrebe, 1989, "Spectral band selection for classification of soil organic matter content," *J. Soil Science Society of America*, vol. 53, no. 6, pp. 1778-1784.

Kim B. and D.A. Landgrebe, 1990, "Prediction of optimal number of features," *Proc. IEEE Int'l Geoscience and Remote Sensing Symp., (IGARSS'90)*, Washington, D.C., pp. 2393-2396.

Narendra P.M. and K. Fukunaga, 1977, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Computers*, vol. C-26, no. 9, pp. 917-922.

Stearns, S.D., 1976, "On selecting features for pattern classifiers," *Proc. Third Int'l Joint Conf. on Pattern Recognition*, pp. 71-75.

Stearns, S.D., B.E. Wilson, and J.R. Peterson, 1993, "Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery," *Proc. SPIE 38th Int'l Symp. on Optical Applied Science and Engineering*, San Diego.