DIMENSIONALITY REDUCTION BASED FAST VECTOR QUANTIZATION ALGORITHM FOR HYPERSPECTRAL IMAGE COMPRESSION

Yushi Chen[1,2], Zhixin Zhou[1], Ye Zhang[2]

[1]Institute of Image and Information technology
[2]Dept. of Information Engineering, Harbin Institute of Technology

1. INTRODUCTION

HyperSpectral Image (HSI) can get a lot of spectral information, which is important for many applications. At the same time, HSI has huge data volume. It is necessary to develop efficient compression algorithm. Vector Quantization(VQ) based methods yield good results for reducing the amount of the data, while have the shortcoming of computing complexity. The research area in developing fast VQ based algorithms for HSI compression is active. In this paper, dimensionality reduction based VQ algorithm for HSI compression is proposed. It's a new direction of fast VQ based HSI compression. Using the new idea, a lot of new algorithms can be proposed. The new algorithms make use of the high dimensional space of HSI is mostly empty, so it can be represented by a lower dimensional subspace. The computing complexity can be reduced dramatically in the lower dimensional subspace at almost no loss of quality.

2. DEMENSIONALITY REDUCTION BASED VQ

VQ based HSI compression method can get good compression result, at the same time, it requires large computational resource. The procedure of VQ based methods has two steps: codebook training and codevector matching. Due to its simplicity and relatively good fidelity, Generalized Lloyd Algorithm (GLA) is the most widely used VQ method to generate the codebook. In GLA, a training sequence, $X_i(i=1,2,...,M)$ with M vectors(the dimensionality is K) is used to generate a codebook, $Y_j(j=1,2,...,N)$ with N codewords. The computational complexity of the GLA is $O(MNK)$.

A big challenge when compressing HSI using the GLA is that huge computation resource is required. M , as the size of a training sequence (e.g., 512 × 614 vectors for AVIRIS image) is very large. The dimensionality of the training vectors, K , up to a few hundred (e.g., 224). The size of a codebook, N , to be generated can be up to a few thousand in order to attain better spectral information preservation. So a significant drawback to VQ based algorithms is that they require large computation, particularly for the codebook generation step. In many cases, VQ based methods are too large computation to be applied.

This work is supported by National Natural Science Foundation of China(60972143, 60972144) and China Postdoctoral Science Foundation(20090451490)

Yushi Chen(1978-), Ph. D., lecture, interested in: Hyperspectral image processing and data compression

HSI has the spatial and spectral information simultaneously and HSI has several hundred bands. So the dimensionality of HSI is more than one hundred. We know a lot of the characteristics of lower dimensional space, but there is a conceptual barrier to have the intuition of the properties of high dimensional space. Higher dimensional space is quite different from the three dimensional space. It's proved that the high dimensional space is mostly empty, which implies that multivariate data in R^d can be represented by a lower dimensional structure without losing significant information. So we can find a low dimension structure without losing significant information, which is good enough to represent the original high dimension.

The band-to-band correlation of HSI is generally very high. Those bands contain a lot of redundancy information and can be removed without loss of crucial information. The main VQ computing task executes in the low dimensionality, so the VQ computing time can be dramatically reduced.

There are two kinds of methods of HSI dimensionality reduction: feature extraction and feature transformation. DCT is one of feature transformation methods. DCT can get fine result while it has fast algorithm. We choose DCT as feature transformation method to analysis the computation complexity. After the DCT based linear transformation, we choose the most important several components, which is the good representation of the original data. Then the reduced date can be used in VQ algorithm. Figure 1 shows the scheme of feature transformation based VQ.

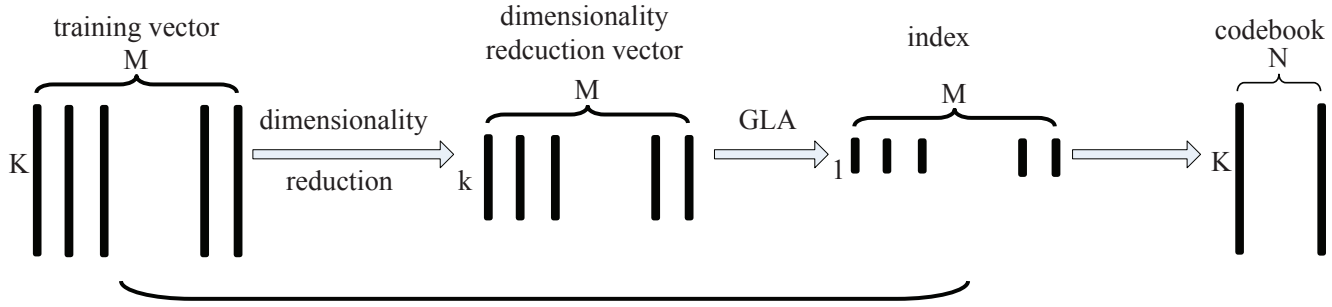

Figure 1 The scheme of demensionality reduction based VQ

To get the efficiency of the proposed method, the computational complexity is analyzed. The computational complexity of original data is $O(MNK)$. The computational complexity of reduced data has two parts: dimensionality reduction(DCT) and dimensionality reduction GLA. The DCT computational complexity of $X_i(i=1,2,...,M)$ is $O(MK\log K)$, while the feature selection based GLA is $O(MNk)$. For a group of typical parameters: M = 314368, N = 1024, K = 224, k = 5. The computational complexity ratio may be 36.2237:1. So using the fast algorithm can dramatically reduce the computing time.

3. EXPERIMENTS AND RESULTS

In order to test the usefulness of the proposed algorithm for HSI, two HSI cubes are tested. One of them comes from a mixed agriculture/forestry. Another contains an airport. Figure 2 shows the original HSI (The left image is

agriculture/forestry and the right image is airport). The Computation Time (CT), SNR and PSNR are used to measure the performance of our method.

Table 1-2 show the training and compression results for the two HSIs with codebook size N = 16 to 1024. The iteration distortion threshold ε of the codebook training process was set to 0.005 in the experiments in order to attain better codebook fidelity. The proposed algorithm improves the computation time by a factor of 1.6 to 27.5. The larger the codebook size, the more the time saving can be achieved. Figure 3 and figure 4 show the results of different components (the codebook size is 256). The solid line shows the result of original GLA and the dash line shows the result of the proposed algorithm. We can get better results with the increase of the number of components.

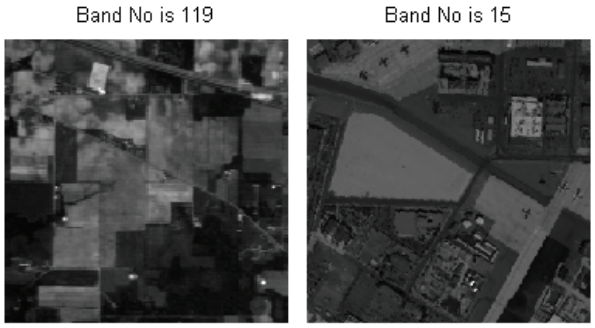

Figure 2: Two HSIs

Table 1 Codebook Training and compression results for the AVIRIS agriculture/forestry

Codebook size (N)	CR	GLA (Full bands)			Proposed algorithm (5 components)		
		CT (sec)	SNR (dB)	PSNR (dB)	CT (sec)	SNR (dB)	PSNR (dB)
64	172.97	58.954	33.017	42.790	20.625	32.857	42.630
256	55.172	397.74	35.845	45.618	32.953	35.757	45.530
1024	15.238	2221.4	38.194	47.968	80.844	38.152	47.925

Table 2 Codebook Training and compression results for the AVIRIS airport

Codebook size (N)	CR	GLA (Full bands)			Proposed algorithm (5 components)		
		CT (sec)	SNR (dB)	PSNR (dB)	CT (sec)	SNR (dB)	PSNR (dB)
64	252.99	167.26	25.252	38.418	85.438	24.16	37.325
256	126.99	576.8	27.194	40.36	132.64	27.182	40.347
1024	48.578	5575.4	30.594	43.76	331.31	30.591	43.757

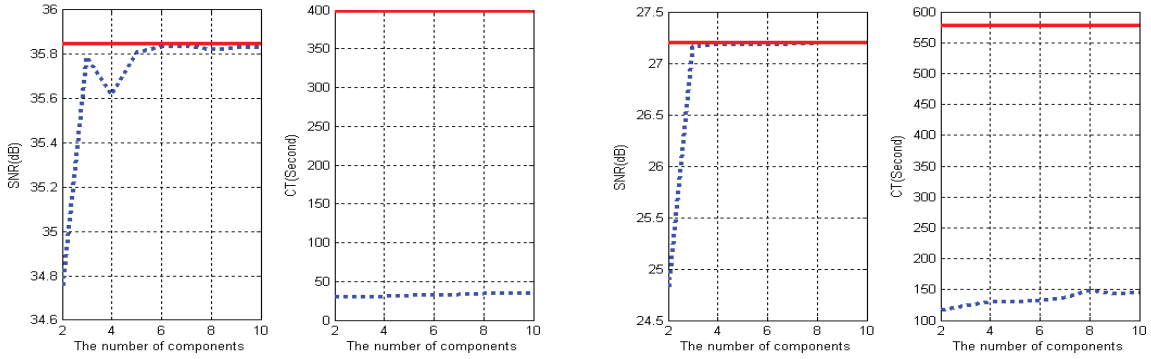

Figure 3 The results of different components (agriculture/forestry) Figure 4 The results of different components (airport)

4. CONCLUSION

A new direction of fast algorithm for VQ based HSI compression has been proposed. It makes use of the fact that high dimensional space is mostly empty, so VQ based compression can be done at the lower dimensional subspace to reduce the computational complexity. The algorithm can be combined with other fast VQ algorithm to achieve less computational complexity.

5. REFERENCES

[1] Emmanuel Christophe, Dominique Leger, Corinne Mailhes. "Quality criteria benchmark for hyperspectral imagery," IEEE Transactions on Geoscience and Remote Sensing. 2005, v43(9): 2103-2114

[2] Shen-En Qian. Hyperspectral data compression using a fast vector quantization algorithm.. IEEE Transactions on Geoscience and Remote Sensing, 2004, v42(8): 1791-1798