Impact of Vector Quantization Compression on Hyperspectral Data in the Retrieval Accuracies of Crop Chlorophyll Content for Precision Agriculture

Baoxin Hu[1,2], Shen-en Qian[2], Driss Haboudane[3], John R. Miller[1,3], Allan Hollinger[2], and Nicolas Tremblay[4]

[1]Dept. of Physics and Astronomy, York University, 4700 Keele St. Toronto, ON M3J 1P3, Canada
[2]Canadian Space Agency, 6767 route l' Aeroport, St Hubert, QC J3Y 8Y9, Canada
[3]Center for Research in Earth and Space Science, York University, 4700 Keele St. Toronto, ON M3J 1P3, Canada
[4]Agriculture and Agi-Food Canada Research Branch, Horticulture Research and Development Center, 430 Blvd. Gouin, St. Jean-sur-Richelieu, QC J3B 3E6, Canada

Abstract--**In this study, impacts of vector quantization compression on prediction of leaf chlorophyll content of crops for the application to precision agriculture were evaluated. The compression algorithm tested in this paper is called successive approximation multi-stage vector quantization (SAMVQ). The hyperspectral data used were acquired by CASI over corn fields at L' Acadie experimental farm (Agriculture and Agri-Food Canada) during the summer of 2000. Nine zones in the corn fields with different fertilization levels (no fertilization, intermediate fertilization, and over-fertilization) were used to evaluate the difference between the leaf chlorophyll contents obtained from original and reconstructed reflectance data cubes. The root mean square errors (RMSEs) and the correlations between the chlorophyll content derived from the original data cube and that derived from the reconstructed data cubes were calculated in the nine zones. The spatial variability of chlorophyll content in the nine zones was also examined for the images of chlorophyll content created from the original and reconstructed reflectance data cubes. The results show that the chlorophyll content image created from the reconstructed reflectance data cube corresponding to SAMVQ with a compression ratio of 20 maintains good agreement with that derived from the original reflectance data cube in the nine zones in terms of estimated pigment mean and spatial variability. As a result, SAMVQ with a compression ratio of 20 is considered acceptable for the retrieval of crop chlorophyll content from CASI hyperspectral data for agriculture corn crops.**

I. INTRODUCTION

The development and applications of hyperspectral remote sensing present a big challenge to remote sensing data transmission and processing, due to the large volume and complexity of data obtained by the hyperspectral sensors. Data compression is one of the solutions to reduce data volume while preserving enough information needed in remote sensing applications. Data compression techniques using vector quantization (VQ) [1] are promising for remote sensing data compression, because of their high compression ratio and relatively simple structure [2,3]. The Canadian Space Agency has recently developed an innovative vector quantization compression algorithm, Successive Approximation Multi-stage Vector Quantization (SAMVQ) [4]. It is under patent pending. The SAMVQ compresses a hyperspectral data cube using extremely small codebooks in a manner of successive approximation. The long codebook-training time in the conventional VQ compression [1] is no longer a problem, since the codebook size in the SAMVQ is over two orders of magnitude smaller than that in the conventional VQ. Due to its unique characters, such as simple, fast, and near lossless with a high compression ratio, the SAMVQ can be used for on-ground and on-board data compression of hyperspectral imagery. However, due to its lossy attribute, it is essential to evaluate the users' acceptability for the hyperspectral data compressed/decompressed by the SAMVQ in terms of their end products or applications, before it can be operationally used. In this study, impacts of the SAMVQ on retrieval of crop chlorophyll content for application to precision agriculture were evaluated using a CASI (Compact Airborne Spectrographic Imager) data set.

II. DATA SET USED AND THE STUDY AREA

The hyperspectral remote sensing data set used in this study was obtained by CASI over corn fields at the L' Acadie experimental farm (Agriculture and Agri-Food Canada) during the summer of 2000. CASI has 72 spectral bands with a spectral resolution of 7.6 nm in the range 0.408 to 0.947 μm. The spatial resolution of this data set is 2m by 2m. The top panel of Fig. 1 is the CASI color composite image over the study area with Band 53 (0.775 μm) printed as red, Band 33 (0.631 μm) as green and Band 20 (0.538 μm) as blue. The study area includes three adjacent fields (Fields 1, 2, and 3.). Three major treatments have been supplied: A, no fertilization; B&C, intermediate treatment; and D, over-fertilization [5,6] (see the bottom panel for details). Nine zones, as shown in the bottom panel of Fig. 1, were used to analyze the effects of SAMVQ on the CASI hyperspectral data. From Fig.1, we can see that the difference between the plots. For example, the plots with treatment A is less red (less biomass as result of low nitrogen in the soil) than the plots with other treatments.

Compression and reconstruction were performed on the CASI radiance data cube. The SAMVQ produced three

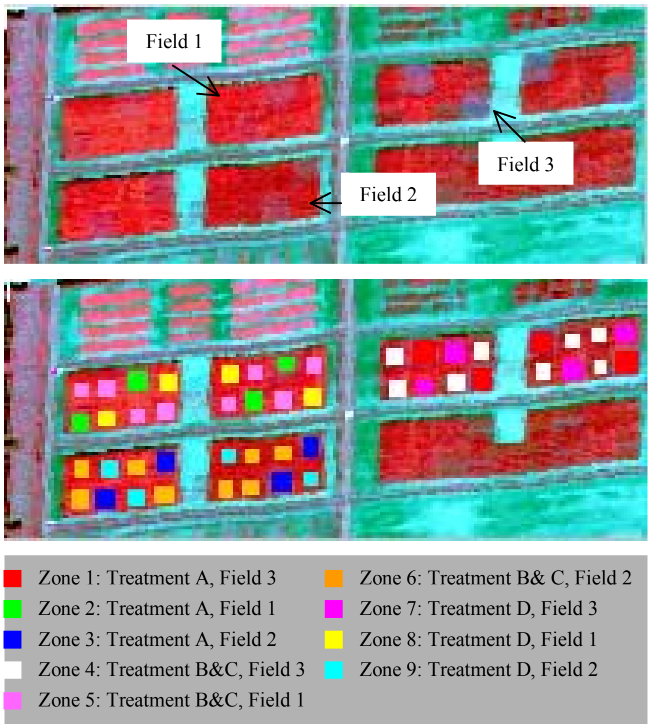

Fig. 1: The study area. It includes 48 plots with 3 treatments in three fields. Treatment A: no fertilization; Treatment B&C: intermediate fertilization; and Treatment D: over-fertilization.

compression ratios of 20, 30, and 50, respectively. The three compression cases are simply called SAMVQ-20, SAMVQ-30, and SAMVQ-50 in this paper. The original and reconstructed radiance data cubes were transformed to at-ground reflectance data cubes (simply called original and reconstructed reflectance data cubes) by the CAM5S atmospheric correction model [7]. The leaf chlorophyll content images were then created from the reflectance data cubes using the algorithm described in [5]. This algorithm has been validated using laboratory chlorophyll content and field measurements [5,6].

III. EVALUATION METHODS AND RESULTS

To evaluate the changes in the estimated leaf chlorophyll content in the nine zones as a result of lossy data compression, we calculated the root mean square errors (RMSEs) in the nine zones between the leaf chlorophyll content derived from the original reflectance data cube and that obtained from the reconstructed data cube for each of the compression cases investigated. The results are summarized in Table I. To get a sense if a certain RMSE value in a given zone is large or small, we compared the RMSE with the standard deviation of this zone in the leaf chlorophyll content derived from the original reflectance data cube. The standard deviations of the nine zones are 3.8, 4.3, 4.2, 4.1, 3.3, 3.6, 3.9, 4.3, 4.6 μg/cm^2, respectively. As shown in Table I, the RMSE in leaf chlorophyll content caused by SAMVQ-20

TABLE I: RMSE (μg/cm^2) IN THE NINE ZONES BETWEEN THE LEAF CHLOROPHYLL CONTENT DERIVED FROM THE ORIGINAL AND RECONSTRUCTED REFLECTANCE DATA CUBE.

Zone	1	2	3	4	5	6	7	8	9
CS1	1.7	1.5	1.9	1.3	1.2	1.6	1.3	1.3	1.8
CS2	4.0	2.3	3.0	2.6	2.2	2.7	2.3	2.8	2.8
CS3	9.9	4.2	6.8	3.5	2.0	3.0	3.0	2.3	4.6

CS1: SAMVQ-20; CS2: SAMVQ-30; and CS3: SAMVQ-50.

smaller than the standard deviation for each zone, and for other compression cases, the RMSEs are larger than the standard deviations for at least one zone, such as Zone 1 for SAMVQ-30 and Zones 1 and 3 for SAMVQ-50.

In this study, we also calculated the correlations between the leaf chlorophyll content derived from the original reflectance data cube and that derived from each of the reflectance data cubes corresponding to the three compression cases. The R^2 values are 0.96, 0.86, and 0.58 for SAMVQ-20, SAMVQ-30, and SAMVQ-50, respectively.

To evaluate the impacts of the SAMVQ on the retrieval of leaf chlorophyll content, we also examined the spatial variability of leaf chlorophyll content in the images created from the original and reconstructed reflectance data cubes. From Fig.1, we know that no nitrogen fertilizer (treatment A) is for Zones 1, 2, and 3, intermediate fertilizer (treatment B&C) in Zones 4, 5, and 6, and over-fertilizer (treatment D) in Zones 7, 8, and 9. We also know from Fig.1 that Zones 1, 4, and 7 are within Field 3, Zones 2, 5, 8 within Field 1, and Fields 3, 6, and 9 within Field 2. Soils in Fields 1 and 2 have more nitrogen (52 and 51 kg/ha, respectively) than in Field 3 (37kg/ha) [5]. From the image of leaf chlorophyll content derived from the original reflectance data cube, we calculated the mean leaf chlorophyll contents of the nine zones. They are 27.4, 41.0, 40.4, 46.5, 47.2, 48.4, 47.8, 47.7 and 49.8 μg/cm^2, respectively. From these values, we can see that the mean leaf chlorophyll content in Zone 1 (27.4) is smaller than those (41.0 and 40.4) in Zones 2 and 3. The difference is mainly caused by the nitrogen level of the fields before seeding (the nitrogen concentration in soil). The between-field differences in leaf chlorophyll content are very small for treatments B&C and D. As a result, we analyzed the spatial variability of chlorophyll content in the nine zones by examining the intra-treatment variability for treatment A, and inter-treatment variability among treatments, A, B&C, and D, and show the results in Table II. From Table II, we can notice that the mean leaf chlorophyll content derived from the original reflectance data cube is the lowest for zones with treatment A and the highest for the zones with treatment D, and for zones with treatment A, it is lower in Field 3 than in Fields 1 and 2. The same spatial patterns are preserved for all of the cases (SAMVQ-20, SAMVQ-30, and SAMVQ-50). From Table II, we can also see that the spatial variability (described by coefficient of variation) of leaf chlorophyll

TABLE II: MEAN AND COEFFICIENT OF VARIATION OF LEAF CHLOROPHYLL CONTENT IN ZONES WITH DIFFERENT TREATMENTS AND WITHIN DIFFERENT FIELDS DERIVED FROM THE ORIGINAL AND RECONSTRUCTED REFLECTANCE DATA CUBES.

	Nitrogen treatment	Original	SAMVQ 20	SAMVQ 30	SAMVQ 50
Mean ($\mu g/cm^2$)	A (field 3)	27.41	27.8	31.1	33.2
	A (fields 1,2)	40.67	39.6	43.1	39.9
	BC	47.41	47.5	47.8	41.0
	D	48.4	48.9	48.2	42.2
CV (%)	A (field 3)	14.08	11.9	13.8	17.0
	A (fields 1,2)	9.05	9.56	11.2	6.97
	BC	6.10	6.73	6.41	7.61
	D	5.83	6.17	6.51	9.99

content derived from the original reflectance data cube decreases with the increasing of nitrogen concentration (from treatment A to treatment D). This agrees with the expected spatial variability resulting from fertilization differences that large nitrogen quantities induce an increase in the homogeneity of the crop canopies, and thus decrease the variation of the crop chlorophyll concentration. [5]. The same trend is only kept in the image of leaf chlorophyll content derived from the reconstructed reflectance data cube corresponding to SAMVQ-20.

IV. CONCLUSIONS

In this study, we investigated the effects of the SAMVQ with compression ratios of 20 (SAMVQ-20), 30 (SAMVQ-30), and 50 (SAMVQ-50) on the retrieval of leaf chlorophyll content of corn from CASI hyperspectral data. The analysis was carried out in nine zones. The RMSEs and correlations between the leaf chlorophyll contents derived from the original and reconstructed reflectance data cubes were calculated for the nine zones. The spatial variability of leaf chlorophyll content in the nine zones was also examined for the images of leaf chlorophyll content derived from the original and reconstructed data cubes. The results show that the RMSEs of the estimated leaf chlorophyll content associated with SAMVQ-20 compression are smaller than 1.9 $\mu g/cm^2$ for the nine zones, which are smaller than the corresponding standard deviations (derived from the image of leaf chlorophyll content created from the original reflectance data cube) of the nine zones. The RMSEs caused by SAMVQ-30 and SAMVQ-50 are larger than their standard deviations for at least one out of the nine zones. The results also show that the leaf chlorophyll content images derived from the reconstructed reflectance data cubes corresponding to SAMVQ-20 and SAMVQ-30 show a good agreement with that derived from the original reflectance data cube, with R^2 values of 0.96 and 0.86, respectively. The intra-treatment and inter-treatment variations exhibited in the image of leaf chlorophyll content derived from the original reflectance data cube are only preserved in the chlorophyll content image derived from the reconstructed reflectance data cube corresponding to SAMVQ-20. As a result, we can conclude that the effects of the SAMVQ with a compression ratio 20 on the retrieval of leaf chlorophyll content of corn from CASI hyperspectral data are acceptable for this study area.

ACKNOWLEDGEMENTS

This work is supported in part by GEOIDE (GEOmatics for Informed Decisions), part of the Canadian NCE Programme and Canadian Space Agency. The authors would like to thank James R. Freemantle of CRESTech for assistance with processing of the CASI data.

REFERENCES

[1] R. M. Gray, "Vector quantization", *IEEE Acoust. Speech, and Signal processing Mag.*, col. 1, pp 4-29, 1984.

[2] S.-E. Qian, A. B. Hollinger, D. Williams, and D. Manak, "Fast 3D data compression of hyperspectral imagery using vector quantization with spectral-feature-based binary coding," *Optical Engineering*, vol. 35, no. 11, pp. 3242-3249, Nov. 1996.

[3] S.-E. Qian and A. B. Hollinger, "Lossy data compression of 3-dimensional hyperspectral imagery," one chapter of book *Information Processing of Remote Sensing* published by World Scientific Publishing Co. Pte. Ltd., 1999.

[4] S.-E. Qian, B. Hu, M. Bergeron, A. B. Hollinger, and P. Oswald, "Quantitative evaluation of hyperspectral data compressed by near lossless onboard compression techniques", submitted to *IEEE Int. Geoscience and Remote Sensing Symp.*, Toronto, June, 2002.

[5] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, L. Dextraze, and A. A. Viau, "Heterogeneity of CASI-estimated leaf chlorophyll in corn: assessment and comparison with ground truth from L'Acadie GEOIDE experimental site" 23rd Canadian Symposium on Remote Sensing, Quebec, Canada, August 21-24, 2001.

[6] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze, "Integrated narrow-band vegetation indices for a prediction of crop chlorophyll content for application to precision agriculture", *Remote Sens. Environn.* In press.

[7] O'Neill, N.T., Zagolski, F., Bergeron, M., Royer, A., Miller, R.J. and Freemantle, J., " Atmospheric correction validation of CASI images acquired over the BOREAS southern study area", *Can. J. of Remote Sensing.* 23: pp. 143-162, 1997.