Cocoa bean quality assessment using closed range hyperspectral images

Oswaldo Bayona*, Daniel Ochoa*, Ronald Criollo*, Juan Cevallos-Cevallos* and Wenzhi Liao†
* Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo Km 30.5
Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
E-mail: {obayona, dochoa, rrcrioll, jmceva}@espol.edu.ec
† Department of Telecommunications and Information Processing, Ghent University,
Sint-Pietersnieuwstraat 41, Ghent, Belgium
E-mail: Wenzhi.Liao@UGent.be

***Abstract*—Farmers mix high and low quality cocoa beans to increase their income at the expense of chocolate flavor. We use closed range hyperspectral images to recognize two common varieties of cocoa beans at various fermentation stages. Several image calibration issues are addressed in this paper to reduce the effect of the bean's shape in the reflectance image estimation and specular patches on the bean's surface. Fusion and feature extraction techniques were exploited for bean classification. From our experimental results, we noticed that bean's biochemical processes during fermentation of each bean type influences their spectral signatures enabling an increasingly better discrimination. We found that spectral indexes related to anthocyanin reflectance index yield a high discriminant rate, particularly at later fermentation stages. These findings suggest that bean classification is possible and could be adopted as the standard method for fast bean quality assessment.**

Keywords: Hyperspectral Imaging, Cocoa, calibration

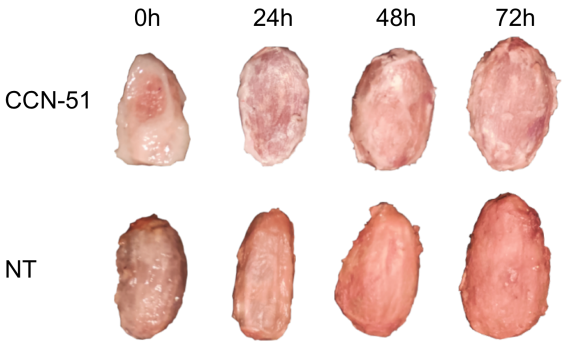


Fig. 1. National and CCN-51 beans at several fermentation stages.

I. Introduction

Cocoa beans are seeds come from the Theobroma Cacao Linnaeus tree that grows in the equatorial zones. Cocoa beans are fermented, dried and packaged for international markets. Cocoa pods contain between 30 and 40 beans wrapped in a mucilaginous pulp [1]. The seed has a coat and the embryo which is protected by two cotyledons. The chemical composition of cocoa beans includes: water, fat, proteins, polyphenols, starch, sucrose and purine bases such as caffeine and theobromine [1][2][3].

Among the polyphenols we found catechins, anthocyanins and proanthocyanins. During fermentation Fig. 1, the concentration of anthocyanins decreases [4][5][6], these pigments give its purple color to the cotyledon. The absence of anthocyanins makes cotyledons to turn brown. The concentration of (-)-epicatechin, (+)-catechin, theobromine, caffeine and moisture also decreases [7][6].

Varietes of cocoa include: Criollo, Forastero Amazonico and Trinitario trees. Cocoa quality is measured based on the mix of flavours it can provide. Low quality, bulk cocoa, is mostly obtained from Forasteros trees. Fine cocoa is obtained from Criollos and Trinitarios trees [8]. In Ecuador, there are two local varieties: Arriba National and Castro Naranjal Collection (CCN-51). Arriba National is a Criollo tree. Ecuador is the main fine cocoa producer, with a 50% of the world production [9]. The CCN-51 variety is a hibrid of Trinitario and Forastero trees. It is considered bulk cocoa, but it is more resistant to diseases and yields four times the number of pods than the National cocoa tree [10]. To get more money from each harvest, farmers mix high and low quality beans and sale them as National cocoa. This has caused problems in the international markets because the flavour of chocolate is difficult to reproduce. In practice, quality control is done by visual inspection in bean fermentation centers. Yet highly trained personnel is required to identify the two varieties in a fast and reliable way.

Previous studies include: [11] in which near-infrared spectroscopy and chemometrics techniques are used to identify five kinds of cocoa beans, they achieved an optimal performance in their model; in [8] proton nuclear magnetic resonance spectroscopy (^{1}H-NMR) and chemometrics techniques were used to discriminate between different fermentation levels in various kinds of cocoa, including Arriba National, and they obtained a classification rate of 85.2%. These works got good results, but their methodologies are destructive, time-consuming, and rather expensive for real environments. In [12] recognition, between CCN-51 and National beans, is made in a non-destructive way using Raman spectroscopy and chemometrics approaches getting a 91.8% rate of accurate

detection. In this paper, we explore the possibility of using automatic classification of National and CCN-51 beans based on hyper-spectral (HS) data.

II. Methodology

A. Image Acquisition

The image acquisition system used in our experiments is described in detail in [13]. The system consists of: a monochrome 12 bit CCD camera 1500M-GE Thorlabs with NIR sensitivity, a spectrograph SPECIM Imspector V10 with spectral sensitivity in the [364,1031] nm. range and spectral resolution of 1.2 nm. and a lighting system with two 50 W. halogen lamps placed on the sides of the camera. These elements are mounted on a motorized slider, that allows a linear scanning of a fixed target. For each scan an hyperspectral cube of 520 images is obtained. Additionally, we used a high-dynamic range CCD camera to capture RGB images.

B. Preprocessing

To reduce the spectral distortion produced by changes in quantum efficiency of the CCD and the geometric configuration of the lightning system, the spectral data was initially calibrated using equation (1). Where R_λ is the reflectance in the λ wavelength, I_λ is the intensity of the sample, W_λ is the intensity of a white standard reference and D_λ is the intensity measured when the sensor does not receive light.

$$R_\lambda = \frac{I_\lambda - D_\lambda}{W_\lambda - D_\lambda} \tag{1}$$

Spectral curves were smoothed using the Savitzky-Golay (SG) filter with a window size of 11. The SG filter is commonly used to reduce the random noise derived from read-out times of the camera, the data transfer, and analog to digital conversion[14]. The first 100 and last 20 wavelengths were discarded as signals are quite noisy in those wavelengths. The effective spectral range was between 500 and 955 nm. When the sample is non-flat the distance form sample's points to the sensor and the relative angle of reflected light rays also have undesirable effects in the estimation of the reflectance spectrum. To compensate these effects, geometric normalizations based on Standard Normal Variance (SNV) transformation, linear optical models [15] and photometric invariance [16] can be applied. We chose the photometric invariance approach of the equation (2) [17].

$$X_\lambda = \frac{R_\lambda}{\sum_\lambda R_\lambda} \tag{2}$$

Fig. 2(a) shows a RGB image of a cocoa bean, Fig. 2(b) is a pseudo-color image (generated by three images of the HS cube) of a cocoa bean marked with 5 regions. Fig. 2(c) shows the average spectrums of the marked regions without geometric normalization, it can be observed that intensity of reflectance increase while a region is closer to the light. Finally in Fig. 2(d) can be observed the aligned spectrums after applying the normalization.

The image fusion technique in [18] was applied. It combines a high resolution RGB image and a low resolution HS cube to produce a fused HS cube with the same spectral dimension and enhanced spatial dimension. First, the algorithm obtains the principal components (PC) of the HS cube, then the PCs are up-sampled using the bicubic filter. The fusion is achieved transferring the spacial structures of the morphological profile of the high resolution image to the first *k* PC images using the guided filter.

Fusion algorithm requires the alignment of HS low-resolution and RGB high resolution images. We aligned the RGB image to the HS cube using GeFolki non-rigid registration algorithm [19]. GeFolki returns the optical flow at every pixel position than in turn is used to guide the registration step. Finally, the RGB images were segmented using a blob detector to remove the specular patches produced by the moisture of the mucilage of the bean. Fig. 3 shows the segmentation of a cocoa bean.

III. Experiments

For our experiments 30 beans of each variety at 0h, 24h, 48h and 72h fermentation stages were randomly selected from beans batch in a fermentation center. In total 240 HS cubes were obtained and processed using the approach described in section II. As input for classification well-known spectral features used previously in other types of samples were extracted. To discard possible outliers, we used the truncated mean removing the lowest and highest 12.5% of the data for each wavelength. The first feature set is simply the average spectrum of each bean pixels. Additional feature sets were computed from the output of several bands selection methods. Such methods keep the most significant bands according to a certain criterion, the 10 most significant bands were selected. We tested: the First Spectral Derivative ($D_1(\lambda)$), the Second Spectral Derivative ($D_2(\lambda)$), Information Entropy ($H(\lambda)$) and PCA ranking (PCAr). A detailed description of each methods can be found in [20].

Features associated to the reflectance properties of the chemical compounds were also studied. Some works [7][5][6] suggest that the concentration of anthocyanin, theobromine, caffeine and epicatechine can be good indicators for beans variety identification. To measure the anthocyanin concentration, the vegetation index Anthocyanin reflectance index 2 (ARI2) [21] were computed, using the equation (3). Where ρ800, ρ550 and ρ700 are ranges of 20 wavelengths of the spectrum that are centered in the wavelengths 800, 550 and 700 nm. This index was used in [4] as a quality indicator of the fermentation process in cut beans.

$$ARI2 = \rho_{800}[\frac{1}{\rho_{550}} - \frac{1}{\rho_{700}}] \tag{3}$$

Theobromine, caffeine and epicatechine present peaks in the ultraviolet and infrared range of the absorbance spectrum [200-300] nm. and [500-4000] nm^{-1}. Unfortunately light at such wavelengths can not be captured by the current acquisition system, so no features were derived. The Water Index (WI)

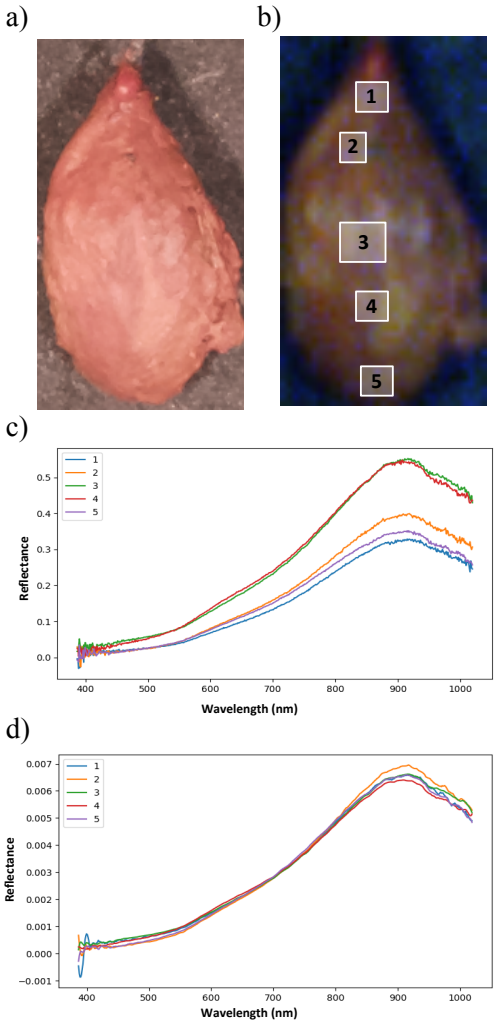


Fig. 2. Normalization of a cocoa bean spectrum

[22] is a vegetative index that measures moisture, see equation 4. The average WI of each bean was calculated to measure the water content changes during the fermentation and it could be a good indicator. The last feature that was used is ARI2 and WI together.

$$WI = \frac{\rho_{900}}{\rho_{970}} \tag{4}$$

For supervised classification, K-Nearest Neighbors (KNN) algorithm with k=8 was performed, and the average accuracy was used as performance metric of the classification process.

IV. Results

Fig. 4 shows the average spectrums of both varieties of cocoa for the four fermentation stages in low resolution and fused HS cube. It can be observed that in the visible range

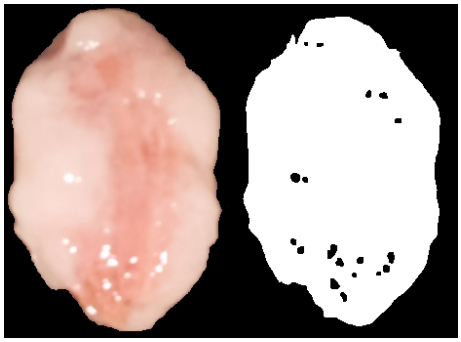


Fig. 3. Segmentation of a cocoa bean. The background and the regions with specular reflection were removed.

the first stages reflects more light, while in the NIR range the lasts stages reflects more light.

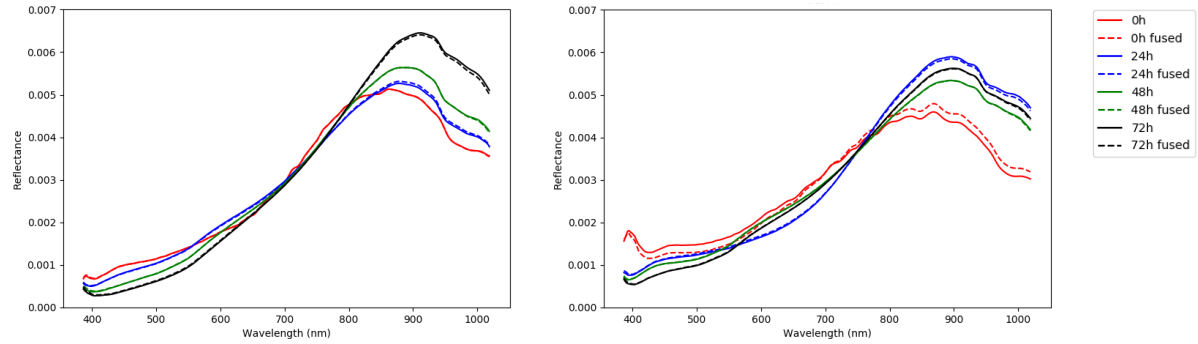
Table I shows the average accuracy of the classification performed on two classes: National and CCN-51, at each fermentation stage, using the low resolution and fused HS cube. The best results were obtained with the set of wavelengths extracted from the low-resolution cubes using $H(\lambda)$, because for 24h and 72h the accuracy values are the highest. Other outstanding result is that for the last fermentation stage in which 95% of accuracy is obtained using $H(\lambda)$, ARI2 and ARI2 + WI. However, the ARI2 + WI feature is not useful for the classification considering that the accuracy on 48h and 72h stages are the same using just the ARI2 feature and only on the first stages are higher than ARI2 but lower than PCAr and $H(\lambda)$. Additionally, comparing the accuracy values between low resolution and fused HS cubes, the image fusion technique improves the accuracy values of most of the features on 0h, 24h, and 48h stages, but no in the same way on 72h stage.

The result of ARI2 seems to indicate that the concentration of anthocyanin in the coat at 72h of fermentation is a good feature. Fig. 5 shows the distribution of ARI2. It can be observed that as the fermentation progresses, the differences in anthocyanin concentrations of both varieties increase. The studies report that the anthocyanin concentration of the cotyledons is reduced during fermentation, but the anthocyanin behavior in the pulp has not been reported.

V. Conclusions

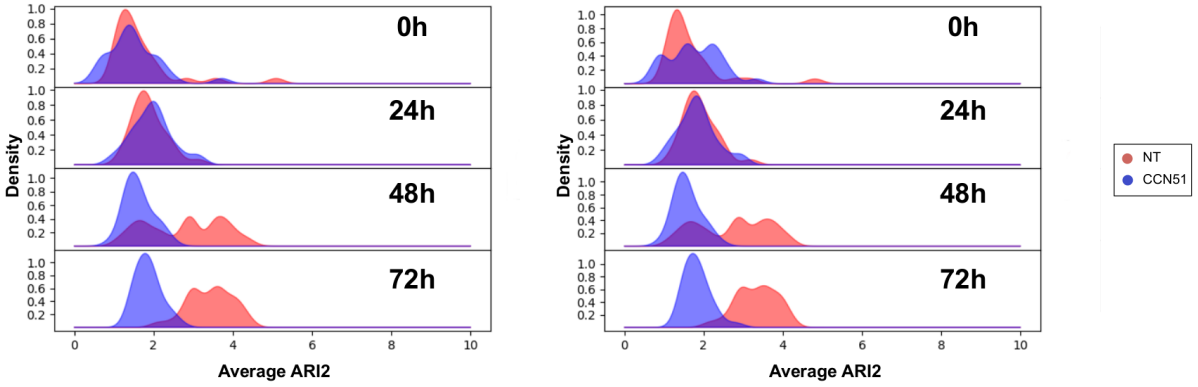
Our evaluation demonstrates that in order to discriminate between the two varieties of cocoa National and CCN-51 in a non-destructive way, the use of hyperspectral images with $H(\lambda)$ is a reliable technology, due to it has the higher accuracy rates for beans at 24h and 72h of fermentation. It is possible to build an optimized acquisition system to measure only those wavelengths that $H(\lambda)$ obtained in the visible spectrum and near infrared. Further, the resulting cubes need less storage space, so that the computational requirements to process the data is reduced.

In the last fermentation stage, the ARI2 is potentially a helpful feature to classify, because 95% of accuracy is obtained



(a) National (b) CCN-51

Fig. 4. Average Reflectance Spectral curves of Low Resolution and Fused HS cubes.



(a) Low Resolution (b) Fused

Fig. 5. Scatter plot of ARI2 for each fermentation stage

TABLE I
AVERAGE ACCURACY OF THE CLASSIFICATION OF THE TWO VARIETIES OF COCOA FOR EACH FERMENTATION STAGE USING DIFFERENT FEATURES.

	0h		24h		48h		72h	
Features	**LR**	**Fused**	**LR**	**Fused**	**LR**	**Fused**	**LR**	**Fused**
Avg. Spectrum	0.60	0.73	0.87	0.88	0.78	0.78	0.90	0.90
$D_1(\lambda)$	0.63	0.70	0.82	0.87	0.82	0.87	0.87	0.85
$D_2(\lambda)$	0.70	0.80	0.78	0.78	0.55	0.58	0.82	0.82
$H(\lambda)$	0.67	0.72	0.88	0.75	0.80	0.63	0.95	0.85
PCAr	0.72	0.67	0.75	0.77	0.70	0.68	0.85	0.85
ARI2	0.42	0.65	0.50	0.43	0.78	0.80	0.95	0.95
WI	0.58	0.73	0.73	0.75	0.72	0.65	0.53	0.45
ARI2 + WI	0.45	0.68	0.62	0.52	0.78	0.80	0.95	0.95

and this suggests that if it is used with other spectral index it could obtain good classification rates.

The accuracy of the classification of cocoa beans without fermentation (0h stage), is very low for both types of cubes. The reason is that the pulp spectral signatures for both types of beans are similar.

To corroborate the measures performed with ARI2, we have to correlate them with anthocyanin concentration values in the bean's coat and include features related to theobromine, caffeine and epicatechine.

The main effect of adding spatial information on classification can be observed at early fermentation stages, nevertheless classifications rates are lower compared to later fermentation stages. An alternative would be to used spatial information to extract pulp free regions.

Our sample beans were fermented without any kind of mixing. In a future experiment we plan to mix National and CCN-51 beans to evaluate the effect of mass exchange between beans on hyperspectral signatures.

VI. Acknowledgment

This work was financially supported by the Flemish Interuniversity Council (VLIR).

References

[1] J. E. Kongor, M. Hinneh, D. V. de Walle, E. O. Afoakwa, P. Boeckx, and K. Dewettinck, "Factors influencing quality variation in cocoa (theobroma cacao) bean flavour profile a review," *Food Research International*, vol. 82, no. Supplement C, pp. 44 – 52, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0963996916300163

[2] A. Recalde, "Evaluacin del efecto del presecado y tiempo de fermentacin, en los contenidos de polifenoles totales, alcaloides y cidos voltiles en dos genotipos de cacao," Ph.D. dissertation, Universidad Central del Ecuador, 2007.

[3] D. Kafow, J. Bohlmann, P. Wilberth, and R. Lieberei, "Identification of main fine or flavour components in two genotypes of the cocoa tree (theobroma cacao l.)," *Applied Botany and Food Quality*, no. 86, pp. 90–98, 2013.

[4] J. Reyes, J. Soto, and I. William, "Hyperspectral analysis based anthocyanin index (ARI2) during cocoa bean fermentation process," vol. 00, pp. 169–172, 2015.

[5] R. Nazaruddin, L. Seng, O. Hassan, and M. Said, "Effect of pulp preconditioning on the content of polyphenols in cocoa beans (theobroma cacao) during fermentation," *Industrial Crops and Products*, vol. 24, no. 1, pp. 87 – 94, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0926669006000598

[6] P. Pelez, S. Guerra, and D. Contreras, "Changes in physical and chemical characteristics of fermented cocoa (theobroma cacao) beans with manual and semi-mechanized transfer, between fermentation boxes," *Scientia Agropecuaria*, 2016.

[7] H. Kim and P. Keeney, "(-)-epicatechin content in fermented and unfermented cocoa beans," *Pennsylvania Agricultural Experiment Station*, no. 6694, 1984.

[8] A. Caligiani, L. Palla, D. Acquotti, A. Marseglia, and G. Palla, "Application of ^{1}H NMR for the characterisation of cocoa beans of different geographical origins and fermentation levels," *Food Chemistry*, vol. 157, no. Supplement C, pp. 94 – 99, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0308814614001587

[9] L. Herrmann, I. Haase, M. Blauhut, and M. Fischer, "Dna-based differentiation of the ecuadorian cocoa types ccn-51 and arriba based on sequence differences in the chloroplast genome," *Agricultural and Food Chemistry*, 2014.

[10] C. J. Melo and G. M. Hollander, "Unsustainable development: Alternative food networks and the ecuadorian federation of cocoa producers, 19952010," *Journal of Rural Studies*, vol. 32, no. Supplement C, pp. 251 – 263, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0743016713000569

[11] E. Teye, J. Uhomoibhi, and W. Hui, "Nondestructive authentication of cocoa bean cultivars by ft-nir spectroscopy and multivariate techniques," *Focus on Sciences*, vol. 2, no. 3, pp. 1–5, 2016.

[12] P. Vargas, V. Ciobotă, W. Salinas, B. Kampe, P. Aponte, P. Rsch, J. Popp, and L. Ramos, "Distinction of ecuadorian varieties of fermented cocoa beans using raman spectroscopy," *Food Chemistry*, 2016.

[13] D. Ochoa, J. Cevallos, G. Vargas, R. Criollo, D. Romero, R. Castro, and O. Bayona, "Hyperspectral imaging system for disease scanning on banana plants," in *SPIE Commercial+ Scientific Sensing and Imaging*. International Society for Optics and Photonics, 2016, pp. 98 640M–98 640M.

[14] D.-W. Sun, *Hyperspectral imaging for food quality analysis and control*. Elsevier, 2010.

[15] M. Shahrimie, P. Mishra, S. Mertens, S. Dhondt, N. Wuyts, and P. Scheunders, "Modeling effects of illumination and plant geometry on leaf reflectance spectra in close-range hyperspectral imaging," 2016.

[16] H. Stokman and T. Gevers, "Deteccion and clasification of hyper-spectral edges," *The Tenth Britisch Machine vision Conference*, 1999.

[17] G. Polder, G. W.A.M. Van der Heijden, H. van der Voet, and I. Young, "Measuring surface distribution of carotenes and chlorophyll in ripening tomatoes using imaging spectrometry," 2004.

[18] D. Ochoa, R. Criollo, W. Liao, J. Cevallos, R. C. Castro, and O. Bayona, "Improving the detection of cocoa bean fermentation-related changes using image fusion," pp. 10 198 – 10 198 – 6, 2017. [Online]. Available: https://doi.org/10.1117/12.2262827

[19] G. Brigot *et al.*, "Adaptation and evaluation of an optical flow method applied to co-registration of forest remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, July 2016.

[20] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 7, pp. 793 – 802, 2004.

[21] A. A. Gitelson, M. N. Merzlyak, and O. B. Chivkunova, "Optical properties and nondestructive estimation of anthocyanin content in plant leaves," *Photochemistry and photobiology*, vol. 74, no. 1, pp. 38–45, 2001.

[22] J. Peuelas, J. Piol, R. Ogaya, and I. Filella, "Estimation of plant water concentration by the refectance water index wi (r900/r970)," *Int. J. Remote sensing*, vol. 18, no. 13, pp. 2869 – 2875, 1997.