Design and Development of a Shortwave near Infrared Spectroscopy using NIR LEDs and Regression Model

Kim Seng Chia, Yit Peng Tan
Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat,
Malaysia

Article Info

ABSTRACT

Near infrared (NIR) spectroscopic technology has been getting more attention in various fields. The development of a low cost NIR spectroscopy is crucial to reduce the financial barriers so that more NIR spectroscopic applications will be investigated and developed by means of the NIR spectroscopic technology. This study proposes an alternative to measure shortwave NIR spectrum using one collimating lens, two slits, one NIR transmission grating, one linear array sensor, and one microcontroller. Five high precision narrow bands NIR light emitting diodes (LEDs) were used to calibrate the proposed spectroscopy. The effects of the proposed two slits design, the distance between the grating and linear array sensor, and three different regression models were investigated. The accuracy of the proposed design was cross-validated using leave-one-out cross-validation. Results show that the proposed two slits design was able to eliminate unwanted signals substantially, and the cross-validation was able to estimate the best model with root mean squared error of cross-validation of 3.8932nm. Findings indicate that the cross-validation approach is a good approach to estimate the final model without over-fitting, and the proposed shortwave NIR spectroscopy was able to estimate the peak value of the acquired spectrum from NIR LEDs with RMSE of 1.1616nm.

Corresponding Author:

Kim Seng Chia,
Department of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia,
Batu Pahat, Malaysia.
Email: kschia@uthm.edu.my

1. INTRODUCTION

Near infrared spectroscopic technology has been getting more attention in various fields e.g. agriculture [1], Chinese medicine [2], [3], neuro-economic research [4], and food industry [5] to potentially replace conventional measurement approaches (e.g. sensory evaluation and chemical analysis) that are time-consuming, expensive per measurement, destructive, or inconvenient. Near infrared spectroscopic technology is promising to be an alternative measurement approach because the acquired reflected or the transmitted near infrared energy from a sample contains information that related to the composition of C-H, N-H, and O-H bonds of the sample. Since these composition are the common and basic chemical composition of many products e.g. agriculture products and food, many researches have been conducted to investigate the relationship between a component of interests e.g. pH, glucose, and soluble solids content and the near infrared energy.

Near Infrared Spectroscopic research aims to relate the component of interest to the absorption of the near-infrared light energy. The amount of the absorption is related to the chemical bond of molecule. In fact, the bond among atoms naturally will vibrate, the higher state of excitement requires more energy. Light

energy from Near Infrared LED provides more energy for chemical bond in atoms to vibrate. Different molecules contain different types of bonding among atoms, therefore, energy required by an atom to vibrate and excite to next level is different. Consequently, this variation energy can be seen in spectrum and provide related information to analyze the types of molecule that present in the samples.

There are three challenges in near infrared spectroscopic research. First, the modelling or the so-called chemo metrics. The modelling part focuses on extracting related information from these highly redundant and over-lapping near infrared spectral data, and then relating these information to a component of interests using a calibration model e.g. artificial neuron network and adaptive linear neuron [6]. The modelling part is crucial to ensure the near infrared spectroscopic technology is capable of achieving satisfied measurement accuracy in novel applications [7]. Second, the experimental design that aims to ensure the acquired near infrared spectral data has minimum effects from the surrounding e.g. temperature and moisture, and the acquired reference data are representative and accurate. A small deviation of temperature or moisture may ultimately cause the acquired data contain unwanted signals. Lastly, a market available near infrared spectroscopy is still relatively expensive and bulky. The development of a low cost and portable near infrared spectroscopy is crucial to reduce the financial barriers so that more applications will be developed for personal use in home by means of the near infrared spectroscopic technology [8].

Several attempts have been proposed to reduce the financial barrier of implementing spectroscopic analysis by replacing an expensive halogen lamp with wavelength selective light emitting diodes (LEDs) and a detector by photodiodes [9-13]. Even though the proposed alternative does not require accurate alignment of the components to guarantee high precision measurements as a commercial spectrometer needs [12], it is worth to highlight that only limited wavelengths can be measured because only wavelength selective light emitting diodes (LEDs) were used in the proposed measurement device. On the other hand, a recent work has presented a simple and inexpensive designs to acquire visible spectrum with acceptable accuracy [14]. This suggests that shortwave near infrared spectrum i.e. from 700–1000nm may be acquired using similar inexpensive setup that consists of a transmission grating, a slit, and a detector. Thus, this study aims to evaluate the feasibility of an alternative to measure shortwave near infrared spectrum.

2. MATERIALS AND METHODS
2.1. Spectrometer Development

Figure 1 illustrates the top view of the proposed shortwave near infrared spectroscopy that consists of a poly (methyl methacrylate) collimating lens(with 4.5mm beam aperture diameter for 200nm-1100nm), two slits, one near infrared transmission grating with 300G/mm and 24.5 degree (ThorLabs), one 1x256 linear array sensor, and one microcontroller (Arduino Nano). The collimating lens was used to make the acquired light travels parallel into the spectrometer. This will ensure the position of each wavelength is independent on acquisition setup so that the proposed spectrometer can accurately estimate the unknown wavelengths. Two slits were used, instead of one slits in a general spectrometer, to ensure only desired light can be pass through toward the transmission grating. The distance between first slit and second slit is fixed at 1.3cm. The transmission grating would disperse the light on the linear array sensor. The linear array sensor that consists of 256 photodiodes with a dimension of 63.5μm ×55.5μm each and with 8-μm spacing between photodiodes was used to measure the intensity of the dispersed light.
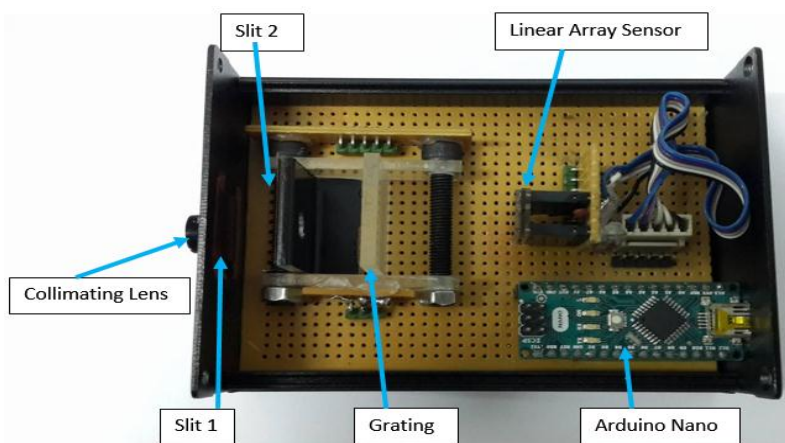

Figure 1. The top view of the proposed shortwave near infrared spectroscopy where the top cover was opened

Next, the microcontroller was used to interface with the linear array sensor, and then to transmit the data from the linear array sensor to a computer using a USB cable for record and analysis. All of these setup are inside an aluminum black box with a dimension of 100mm×76mm×35mm. The distance between the grating and the linear array sensor was adjusted until the dispersed light can be measured by the linear array sensor. Lastly, Matlab (R2009b) software was used to interface with the microcontroller for record and analyze the acquired data.

2.2. Wavelength Calibration

Wavelength calibration aims to ensure that the proposed spectroscopy is capable of identifying the wavelength of the acquired spectrum. Five high precision narrow bands near infrared LEDs (ThorLabs) with different center wavelengths of 780nm, 851nm, 870nm, 910nm, and 940nm were used as the references for calibration and validation. Since there were five data, leave-one-out cross-validation was used to validate the accuracy of the designed shortwave near infrared spectroscopy. The calibration model aims to estimate the wavelength when the unknown light is measured based on the following relationship, i.e. $y = mX$, in which, y is the estimated wavelength, X is the pixel position of the linear array sensor with value between 1 and 255 (i.e. from first pixel to the last pixel of the linear array sensor), and the m is the coefficient value that relates the X and y. One of the challenges was to ensure the acquired light would be successfully dispersed on the linear array sensor using the grating. The preliminary test was to ensure the peak of the acquired spectrum would be in the logic sequence. To avoid an extrapolation prediction, the prediction of both 780 and 940nm were excluded during the cross-validation. Three different regression models i.e. linear, quadratic, cubic regression models would be investigated to avoid both under- and over-fitting issues. Root mean squared error of cross-validation (RMSECV) was used to determine the best model. The architectures of the best models during the cross-validation would be averaged and used as the final model for the proposed sensor to predict the wavelength of the acquired spectrum.

3. RESULTS AND ANALYSIS

3.1. The Effects of the Second Slit

Figure 2 illustrates that the proposed two slits design was able to remove unwanted signals and the peak of the spectrum can easily be observed. These unwanted signals may be due to the present design did not consist of various focus lens to guide the light travels to the desired path. The proposed solution that used an extra slit to block unwanted light from other directions successfully filtered the noises. The first maximum of the peak of the 780nm was allocated at the 111st pixel of the linear array sensor. The second maximum of the peak was observed at 45th and 189th pixels. The similar findings were obtained for the rest of high narrow bands LEDs i.e. 851nm, 870nm, 910nm, and 940nm.

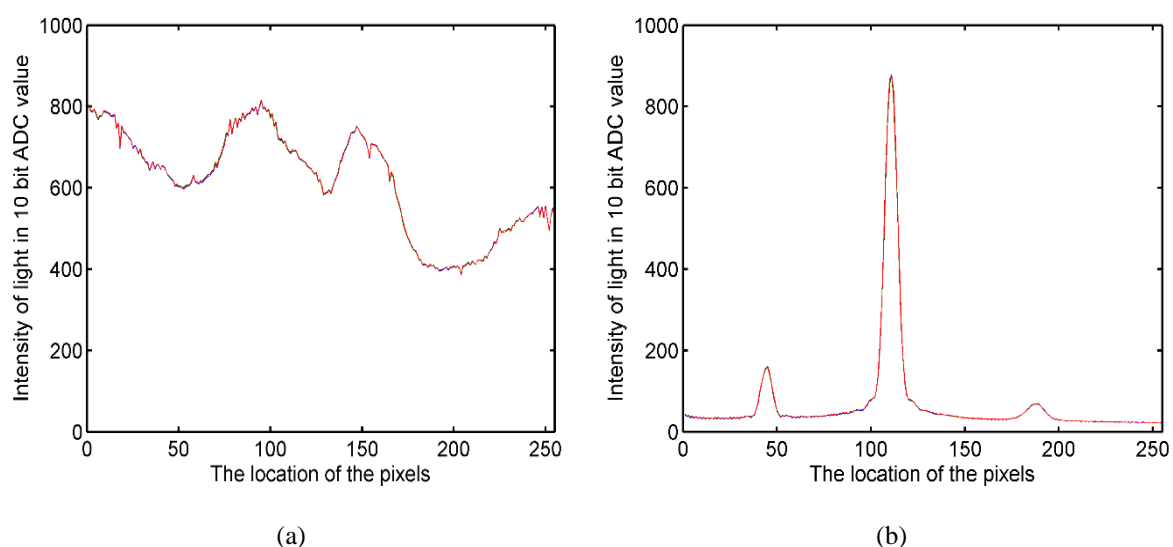|	
(a)	(b)

Figure 2. The intensity of the acquired light from the LED with peak at 780nm against the location of the pixels in the linear array sensor when the proposed design used: (a) single slit, and (b) two slits, respectively

3.2. The Distance between the Grating and Linear Array Sensor

Figure 3 depicts that the position of the peak would be shifted to left when the distance between the grating and the linear array sensor was increased. The same effect was observed for the rest of high narrow bands LEDs i.e. 851nm, 870nm, 910nm, and 940nm. In order to identify the most suitable distance between the grating and the linear array sensor, the total pixels between the peaks of different wavelengths were analyzed. Results indicate that distances of 2.5cm gave a better resolution compare to that of 1.7cm and 2.1cm. Thus, the distance of 2.5cm was chosen as the most suitable distance for the proposed sensor.

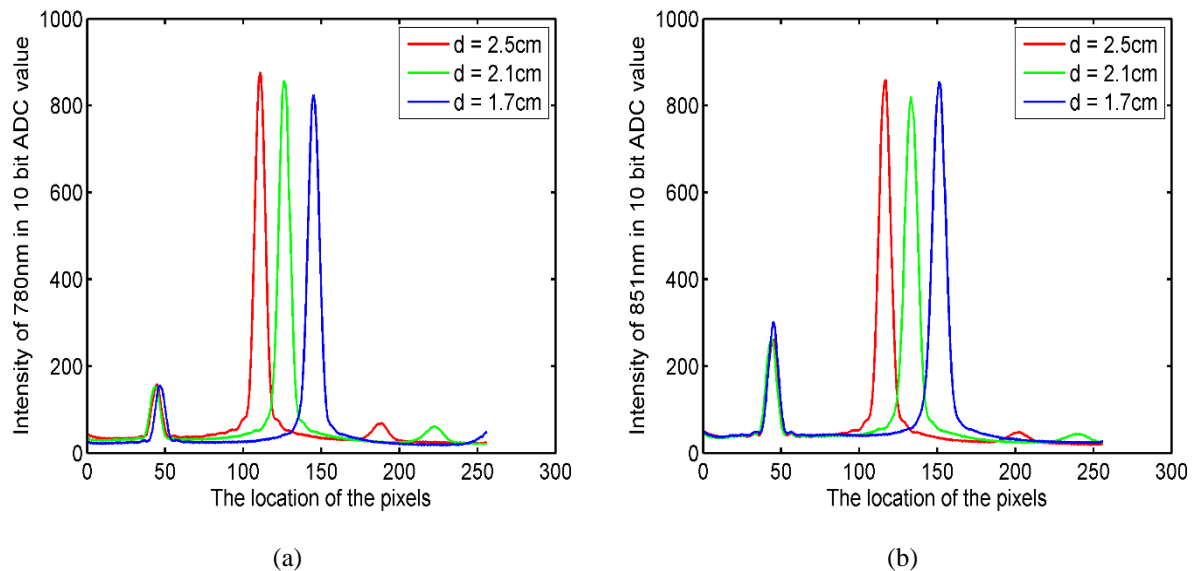(a)   (b)

Figure 3. The intensity of the acquired light against the location of the pixels in the linear array sensor from the LED with peaks of: (a) 780nm, and (b) 851nm, d = the distance between Distance between the linear array sensor and grating

3.3. The Accuracy of the Shortwave near Infrared Spectrum

Table 1 tabulates the position of the acquired wavelength via the proposed shortwave near infrared sensor. All the acquired spectrum were in the correct sequence. This indicates that the proposed design was able to detect the peak of a given spectrum at the correct sequence. The detail of the predicted value via the best model would be explained after we had discussed the outcomes of the cross-validation as follow.

Table 1. The position of the peaks that detected in the linear array sensor

Wavelength, nm	Position	Predicted value via the best model, nm
780	128th	780.3584
851	136th	852.8096
870	138th	868.7244
910	144th	911.1936
940	149th	940.5401

Table 2 tabulates the cross-validation performance of three different regression model of linear, quadratic, and cubic. Results depicts that the model was under-fitted when simple model of linear regression was used. On the other hand, the model was over-fitted when complex model of cubic regression was applied. Consequently, quadratic regression model achieved the best performance with root mean squared error of cross-validation of 3.8932nm. Finding implies that there was a nonlinear relationship between the spectrum detected by the proposed design and the respective wavelength.

Table 3 tabulates the coefficient of the quadratic regression models that were used for the cross-validation. The estimated final model that used to predict the wavelength of the acquired spectrum equals $-0.1099X^2 + 38.07X - 2292$ by averaging the coefficient values from the cross-validation. The final model was able to estimate the peak value of the acquired spectrum as that tabulated in Table 1. The root mean

square error of the final model was 1.1616nm that was smaller than the RMSECV. This shows that the cross-validation approach was able to estimate the optimal final model without over-fitting problem.

Table 2. Leave-one-out cross validation for the proposed shortwave near infrared sensor

Model	Cross-validation	R^2	Actual wavelength, nm	Predicted wavelength, nm	Error, nm	RMSECV, nm
Linear regression	1	0.9859	851	846.2717	-4.7283	
	2	0.9916	870	860.6977	-9.3023	10.7359
	3	0.9886	910	907.4761	-2.5239	
Quadratic regression	1	0.9995	851	853.2834	2.2834	
	2	1.0000	870	866.9901	-3.0099	3.8932
	3	0.9989	910	910.9402	0.9402	
Cubic regression	1	NaN	851	854.8528	3.8528	
	2	NaN	870	866.9437	-3.0563	10.1605
	3	NaN	910	918.8911	8.8911	

Table 3. The coefficient of the quadratic regression models

Cross-validation	X^2	X	Constant
1	-0.1193	40.64	-2467
2	-0.098	34.77	-2065
3	-0.1125	38.79	-2343
Average	-0.1099	38.07	-2292

4. CONCLUSION

The feasibilities of the proposed alternative to measure shortwave near infrared spectrum between 780 and 940nm were investigated. The proposed two slits design was able to eliminate unwanted signals substantially so that the peak of the wavelength could be visualized clearly. The accuracy of the proposed design was validated using cross-validation approach. The best model was obtained using quadratic regression model with root mean squared error of cross-validation of 3.8932nm. The equation of the final model was $y = -0.1099X^2 + 38.07X - 2292$ with root mean squared error of 1.1616nm. Finding suggests that shortwave near infrared spectrum can be measured using the proposed design. In other words, the proposed design is promising to encourage more shortwave near infrared spectroscopic analysis by reducing the financial barrier. In future, we will investigate the feasibility of the proposed shortwave near infrared spectroscopy in various near infrared spectroscopic analysis e.g. predicting the internal quality of local fruits.

REFERENCES

[1]   L. Xuemei and L. Jianshe, "Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration," *Measurement,* vol. 46, pp. 3808-3814, 2013.
[2]   W.-l. Li, H.-f. Han, L. Zhang, Y. Zhang, and H.-b. Qu, "Manufacturer identification and storage time determination of "Dong'e Ejiao" using near infrared spectroscopy and chemometrics," *Journal of Zhejiang University SCIENCE B,* vol. 17, pp. 382-390, 2016.
[3]   Y. Luo, W.-l. Li, W.-h. Huang, X.-h. Liu, Y.-g. Song, and H.-b. Qu, "Rapid quantification of multi-components in alcohol precipitation liquid of Codonopsis Radix using near infrared spectroscopy (NIRS)," *Journal of Zhejiang University SCIENCE B,* vol. 18, pp. 383-392, 2017.
[4]   I. M. Kopton and P. Kenning, "Near-infrared spectroscopy (NIRS) as a new tool for neuroeconomic research," *Frontiers in Human Neuroscience,* vol. 8, 2014.
[5]   J. T. Alander, V. Bochko, B. Martinkauppi, S. Saranwong, and T. Mantere, "A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety," *International Journal of Spectroscopy,* vol. 2013, p. 36, 2013.
[6]   K. S. Chia, "Adaptive linear neuron in visible and near infrared spectroscopic analysis: predictive model and variable selection," *ARPN Journal of Engineering and Applied Sciences,* vol. 10, pp. 9055-9059, 2015.

[7] K. Wiesner, K. Fuchs, A. M. Gigler, and R. Pastusiak, "Trends in Near Infrared Spectroscopy and Multivariate Data Analysis From an Industrial Perspective," *Procedia Engineering,* vol. 87, pp. 867-870, 2014.

[8] J. Tenhunen, H. Kopola, and R. Myllylä, "Non-invasive glucose measurement based on selective near infrared absorption; requirements on instrumentation and spectral range," *Measurement,* vol. 24, pp. 173-177, 1998.

[9] V. Giovenzana, R. Civelli, R. Beghi, R. Oberti, and R. Guidetti, "Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (Vitis vinifera L.) for Franciacorta wine," *Talanta,* vol. 144, pp. 584-591, 2015.

[10] T. S. Yeh and S. S. Tseng, "A low cost LED based spectrometer," *Journal of the Chinese Chemical Society,* vol. 53, pp. 1067-1072, 2006.

[11] M. Degner, H. Ewald, and E. Lewis, *"LED based spectroscopy - A low cost solution for high resolution concentration measurements e.g. for gas monitoring applications,"* in 2011 Fifth International Conference on Sensing Technology, 2011, pp. 145-150.

[12] F. A. Gentilin, B. T. Kobo, L. de Souza Ribeiro, J. A. de França, A. L. de S.M. Felício, M. B. de M. França*, et al.*, "Development and testing of a hardware platform for measuring instruments based on near-infrared diffuse reflection," *Measurement,* vol. 86, pp. 14-25, 2016.

[13] K. S. Chia, N. A. S. Suarin, and S. F. Z. M. Fuzi, *"Prediction of Glucose Concentration using Near Infrared Light and Adaptive Linear Neuron,"* in IEEE 13th International Colloquium on Signal Processing & its Application (CSPA 2017), Penang, Malaysia, 2017.

[14] M. Rodrigues, M. B. Marques, and P. S. Carvalho, "How to build a low cost spectrometer with Tracker for teaching light spectra," *Physics Education,* vol. 51, p. 014002, 2016.

BIOGRAPHIES OF AUTHORS
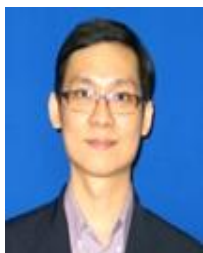
Kim Seng Chia obtained his PhD and BEng from Universiti Teknologi Malaysia in Feb. 2014 and Sept. 2010, respectively. He was an assistant professor at Southern University College, Skudai, Malaysia from March 2014 to August 2014. He has served as a senior lecturer in the Faculty ofElectrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Malaysia since September 2014. His research interests are Machine Learning, Artificial Intelligence, Microcontroller, Soft Modeling, and Near Infrared Spectroscopic Analysis (email: kschia@uthm.edu.my, url: https://sites.google.com/site/kschia86/).

Tan Yit Peng is currently pursuing the BEng in Mechatronic and Robotic Engineering in the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Malaysia. She was the team leader of UTHM Robotic Club 2016/2017. She was one of the winners in Innovate Malaysia Design Competiton 2017: Sime Darby Plantation Design Challenge. Her research interests are Artificial Intelligence and Near Infrared Spectroscopic Analysis (e-mail: yiping927.yt@gmail.com).