# Comparison of Ink Classification Capabilities of Classic Hyperspectral Similarity Features

Binu Melit Devassy
Department of Computer Science
Norwegian University of Science and Technology
Gjøvik, Norway
e-mail: binu.m.devassy@ntnu.no

Sony George
Department of Computer Science
Norwegian University of Science and Technology
Gjøvik, Norway
e-mail: sony.george@ntnu.no

Jon Y. Hardeberg
Department of Computer Science
Norwegian University of Science and Technology
Gjøvik, Norway
e-mail: jon.hardeberg@ntnu.no

***Abstract*—Ink classification is an active topic in historic and forensic document analysis. In this work, we compared the ink classification capabilities of five commonly used and well-proven similarity measures for classification of hyperspectral imaging (HSI). They are Spectral Angle Mapper (SAM), Spectral Correlation Mapper (SCM), Euclidean Distance (ED), Spectral Information Divergence (SID) and Binary Encoding (BE). These techniques were well explored in different fields of HSI; however, they are not investigated in the field of ink classification. This study reveals the ink classification capabilities of these similarity measures. A combination of different types and colors of inks from different manufactures were used to create sample text. The SAM obtained higher accuracy compared to other methods and also identified that, inks that have nearly similar spectral signatures, cause a decline in accuracies due to misclassification between similar classes.**

***Keywords-component; Hyperspectral imaging, Ink classification, Spectral classification***

## I. INTRODUCTION

Hyperspectral imaging (HSI) is an emerging tool for processing documents; it has more capabilities in revealing the underlying information than conventional 3-channel imaging. Ink analysis is an important part of forensic analysis to identify document forgery [1]. The conventional ways of ink analysis are manual, like visual investigation of the document by using microscopy [2]. However, visual inspection is inadequate for distinguishing inks of similar colors, due to the limitation of the human visual system. Human visual system can identify colors in the visible region of the electromagnetic spectrum. However, it may not make reliable judgement of objects that appears similar under the visible spectrum but having different appearances under ultraviolet (UV) or infrared (IR) regions. Here, the HSI made its impact; it can detect objects through a larger spectrum.

HSI dataset contains a large number (typically a few tens to several hundreds) of contiguous spectral bands. Inks that look similar in visible spectrum might be different in wavelength beyond the visible range. The important technologies used so far in this area are, normal RGB camera[3], Raman spectroscopy[4][5], liquid chromatography[6] and infrared spectroscopy[7]. Using HSI techniques, the paper from Khan *et al*.[8] describes the ink mismatch detection. Another important work made by Reed *et al*.[2], employed HSI imaging to analyze gel pen inks. Wang *et al*.[9] executed a study to identify tampered handwriting from hyperspectral images and Livia *et al*.[10] used NIR (Near Infra-Red) HSI and chemo-metrics for analyzing ink crossings. Most recently Muhammad *et al*.[11] showed the potential of HSI for ink analysis by analyzing gel pen inks.

In this study, we selected five-similarity features; they are Spectral Angle Mapper (SAM), Spectral Correlation Mapper (SCM), Euclidean Distance (ED), the Spectral Information Divergence (SID) and Binary Encoding (BE). These algorithms were well explored for hyperspectral classification in different fields[12-15]. However, in the field of ink classification, there exists little effort to explore their capabilities in discriminating inks [16]. In this work, we used five similarity measures for classifying ink and compared their ink classification accuracies.

## II. MATERIALS AND METHODS

### *A. Hyperspectral Image Acquisition*

We used HySpex VNIR-1800 [17] camera from Norsk Elektro Optikk AS for HSI acquisition, and illumination technologies 3900e light box [18] as light source. The camera was positioned perpendicular to the paper (scan target) and the light source was at an angle of 45 degrees from the camera. This camera acquires images in a spectral range from visible to near infrared region, starting from 400 to 1000 nm with 186 spectral bands. The VNIR-1800 has a spectral sampling of 3.7 nm and have a spatial resolution of 1800 pixels across the field of view. The acquisition distance was 21cm from the camera with 10 cm field of view. A Spectralon reference target was kept in the scene in order to use it later to calculate the normalized reflectance [19]. Radiometric calibration was performed using the software provided along with the camera. This software processes the raw images into the sensor absolute radiance values. It also corrects the non-uniformity and dark current factors during the processing.

### *B. Inks and Sample Script*

HSI dataset for the study was created by using VNIR-1800 hyperspectral system. To generalize the test samples, used ten different inks including ballpoint, gel and ink pens from different manufactures. Also included inks, which has similar spectral signatures and used them closely while writing text to confuse the classifiers. Table 1 shows the details of pens used in the experiment and Figure 1 shows a sample text used. In this report, the inks are called classes and named from C1 to C10. Also, assigned different colors for each class for a better visualization and is given in the Table 1.

| Class | Brand | Model | Type | Color | Sample | Classification Color |
|---|---|---|---|---|---|---|
| C1 | Hub Pen | 17r | Ball point | Blue | | |
| C2 | Cello | Pin Point | Ball point | Blue | | |
| C3 | Beifa | RX302 6-00 | Ink | Black | | |
| C4 | BIC | Atlantis | Ball point | Blue | | |
| C5 | Pilot | G-knock | Gel | Red | | |
| C6 | Pilot | G-knock | Gel | Black | | |
| C7 | Pilot | V Ball grip | Ink | Blue | | |
| C8 | Beifa | RX302 6-00 | Ink | Blue | | |
| C9 | Stick | Eazy | Ball Point | Black | | |
| C10 | Pilot | Super Grip F | Ball Point | Blue | | |

Table 1. *Details of ink classes used in this experiment*

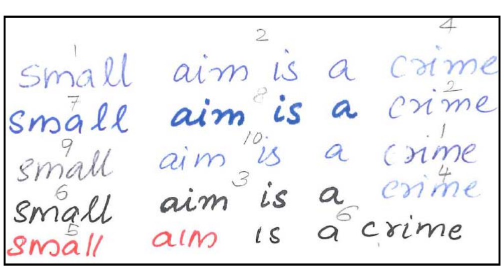


Figure 1. *Sample from the texts used for study, their class numbers are written on top of each text*

### *C. Spectral Angle Mapper (SAM)*

The SAM is a similarity measure for measuring the shape similarity between the reference and test spectra. The algorithm calculates the spectral matching between two spectra by estimating the "angle" between the two spectra, treating them as vectors in a space with dimensionality equal to the number of bands (*nb*) [20].

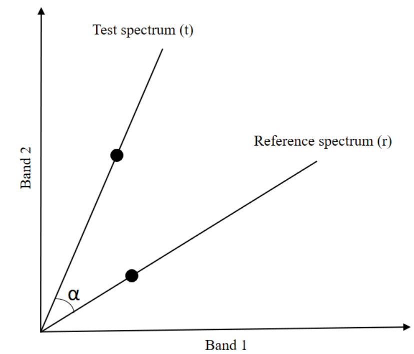


Figure 2. *Plot shows a two band image, alpha(α) represents the angle between the vectors corresponding to test and reference spectrum.*

The angle alpha (α) defines the similarity between test and reference spectrum, alpha can be calculated using the below equation (Equation 1), where r and t are the reference and target spectrums.

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{nb} t_i r_i}{\left(\sum_{i=1}^{nb} t_i^2\right)^{1/2}\left(\sum_{i=1}^{nb} r_i^2\right)^{1/2}}\right) \quad (1)$$

### *D. Spectral Correlation Mapper (SCM)*

In this method, a cross-correlation is calculated between the reference and target spectra. The SCM values varies from –1 to 1 and SAM values varies from 0 to 1[21]. The below equation (Equation 2) shows, how to calculate SCM value, where r and t be the reference and target spectrums and n is the number of spectral channels.

$$SCM\ (t,r) = \frac{n\sum_i^n t_i r_i - \sum_1^n r_i \sum_1^n t_i}{\sqrt{\left[n\sum_1^n r_i^2 - \left(\sum_1^n r_i\right)^2\right]\left[n\sum_1^n t_i^2 - \left(\sum_1^n t_i\right)^2\right]}} \quad (2)$$

### *E. Euclidean Distance (ED)*

Euclidean distance is a quantitative measure of similarities between reference and target spectral vectors. If *r* and t will be the reference and target spectrums and *n* be the number of spectral channels, then ED can be calculated as follows [22]:

$$ED = \sqrt{\sum_1^n (r_i - t_i)^2} \quad (3)$$

### *F. Spectral Information Divergence (SID)*

SID is designed to consider each pixel in the spectrum as a random variable and measures the similarity between reference and target spectra by finding the discrepancy in probabilistic behaviors of two spectra[13]. In this method, first assume that the values in the input spectra (assume $x_l$ be the values) are non-negative, because of the nature of reflectance. Then normalize the $x_l$ to the range of [0,1], by defining $p = \frac{x_j}{\sum_{i=1}^{L} x_i}$. Here x is the given hyperspectral pixel vector $x = (x_1,---,x_L)^T$, then define a normalized vector $p= \{p\}_{i=1}^{L}$. To use this information for comparing spectrums, we need a reference spectrum and a test spectrum. Here, it is assumed that $x_l$ as reference spectrum and another hyperspectral vector $y = (y_1,---,y_L)^T$ as test spectrum with probability distribution $q= \{q\}_{i=1}^{L}$. Using p and q SID can be define as[13]

$$SID(x,y) = D(x \parallel y) + D(y \parallel x)$$

Where

$$D(x \parallel y) = \sum_{i=1}^{L} p_i \log\left({p_i}/{q_i}\right) \text{ and}$$

$$D(y \parallel x) = \sum_{i=1}^{L} q_i \log\left({q_i}/{p_i}\right)$$

### G. *Binary Encoding (BE)*

The BE algorithm is performed in two steps, in the initial step it encodes the input spectra into zeros and ones, by using a threshold. Here used spectrum means as threshold, this step can be expressed as below, T will be the threshold, $x_i$ be the input spectra and I will be the binary output.

$$I = \begin{cases} 1 & if\ x_i > T \\ 0 & else \end{cases}$$

In the next step the BE algorithm executes an Exclusive OR (XOR) between the binarised reference and target spectra[23]. The sum of this XOR result will give the error.

### H. *Processing and Accuracy Measurement*

The camera software applies some preprocessing on the hyperspectral image acquired. From this preprocessed data, the normalized reflectance data will be calculated with the help of standard reference present in the scene. To execute any similarity criteria, we need a reference spectrum; the reference spectrum for all inks were created by taking an average of 5x5 pixels from a random point of each ink. Then performed the similarity algorithms to get their classification accuracy. The accuracy was calculated as a ratio between numbers of correctly classified ink pixels against total number ink pixels present.

## III. RESULTS AND DISCUSSION

The reference spectrum obtained for different ink classes are given in the Figure 3. From this figure, we can observe that the spectra of all ball point blue inks looks nearly similar in shape but varies only in magnitude, they are C1, C2, C4 and C10.

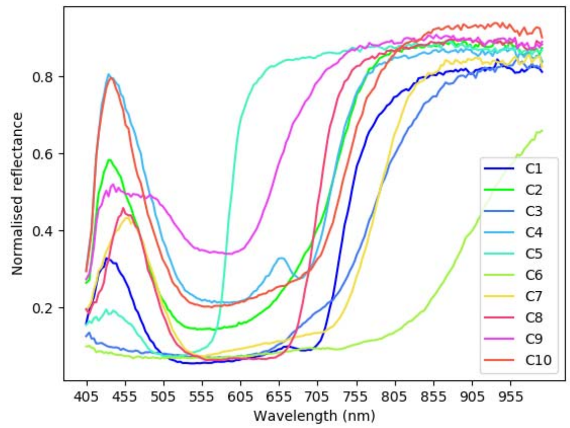


Figure 3 *Spectra of all ink reference classes used in the present study*

The results for all classification methods are given in Figures 4,5,6,7,8 and Figure 9 shows the overall accuracy between all the methods used. From Figure 9, it is clear that the accuracy of SAM classifier outperformed all other methods. SID and BE possess nearly similar accuracies but lower than that of SAM. The accuracies of SCM and ED seems to have the lowest accuracy values.

From the results, it is clear that the accuracy levels are good for all classes except for class 1(C1). The reason behind this can be explained based on the spectral characteristics of C1 and C2 inks. From Figure 10, it is clear that reference spectrum of C2 look almost similar to C1 sample. Hence, the SAM misclassified C1 as C2 in the classification process. From our observation, it is happening in the sample text, where the thickness of the script varies. They already have a similar shape and while varying the writing thickness; it will affect the magnitude. If the magnitude values were close to the C2 reference, that pixel might be misclassified as C1. Figure 11 illustrates this misclassification and this behavior is observed in all five classifiers used.

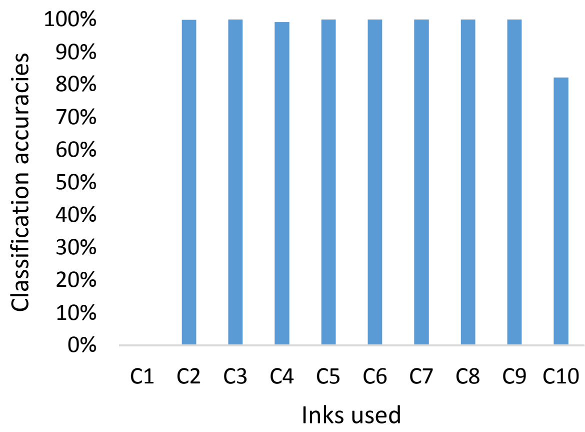


Figure 4 *SAM classification accuracy*

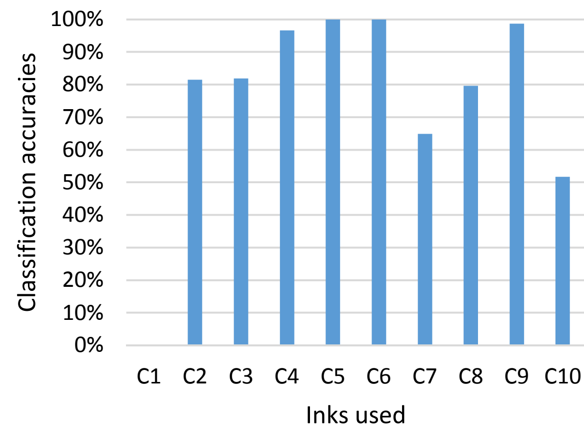


Figure 5 *SCM classification accuracy*

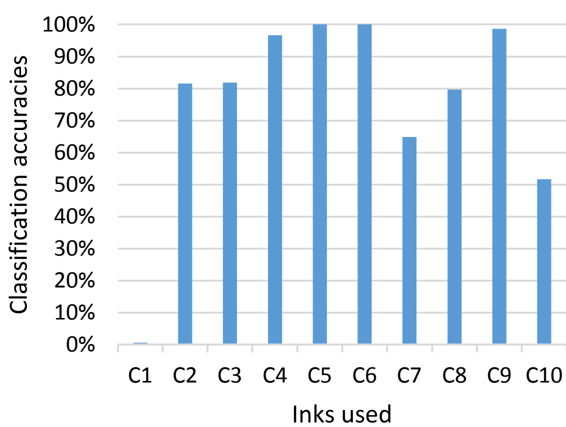


Figure 6 *ED classification accuracy*

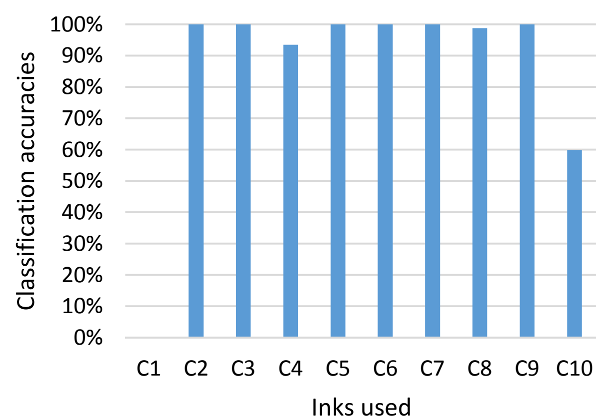


Figure 7 *SID classification accuracy*

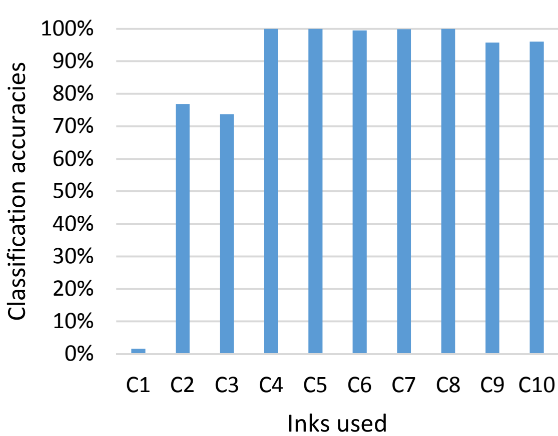


Figure 8 *BE classification accuracy*

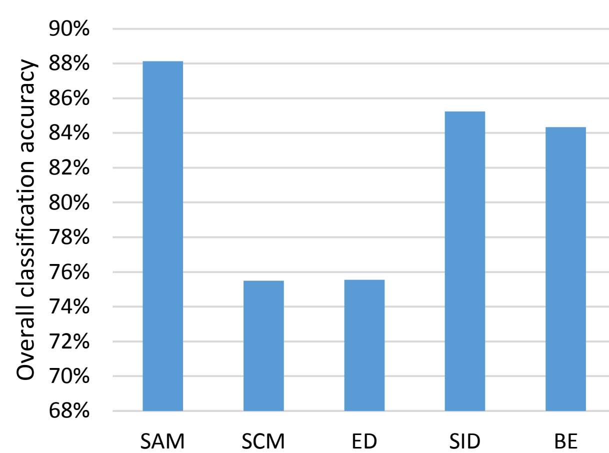


Figure 9 *Overall classification accuracy*

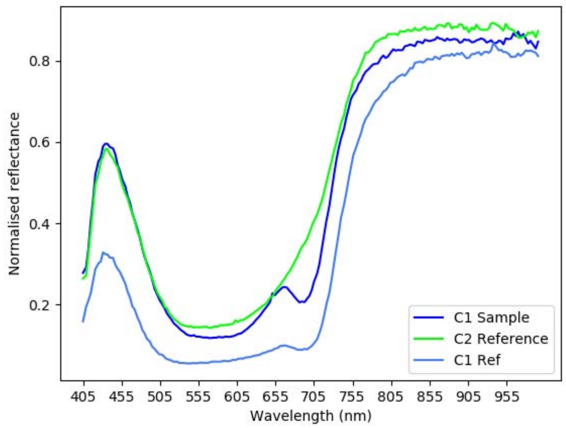


Figure 10 *Spectral similarity between C1 and C2 inks*

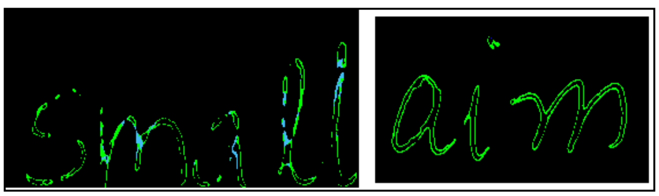

Figure 11 *Result of SAM classification, Left image C1 and right is C2. Ideally, C1 should be filled with blue, but due to spectral similarity it was mostly filled with green*

From Figure 9, it is clear that, SAM has the highest accuracy level compared to other four classification algorithms; this is because SAM is designed to compare the shape rather than magnitude. Another important observation was the decrease in accuracy for most of the ballpoint blue inks used (C1, C2, C4 and C10). This can be explained based on their spectrums shown in Figure 3; from the spectrum, it is clear that inks has nearly similar values in most parts of their spectra. Due to this similarity, they were misclassified between the classes. All classifiers obtained 100% accuracy for red ink (C5), because only a single red ink was used to prepare the sample text.

## IV. Conclusion

In this work, we investigated the ink classification capabilities of five hyperspectral similarity measures. The hyperspectral image dataset used in this study includes different types of inks as well as different colors. From the classification results, it is clear that SAM classification technique outperforms other similarity measures tested. In addition, it is observed that the usage of ink, which has similar spectral signatures, affected the accuracy of all methods. Finding a suitable method to improve the classification between inks, which have nearly similar spectral signature, might be considered as a future work of this experiment.

## V. References


[1] A. Braz, M. López-López, and C. García-Ruiz, "Raman spectroscopy for forensic analysis of inks in questioned documents," *Forensic Science International*, vol. 232, no. 1–3. pp. 206–212, 2013.

[2] G. Reed, K. Savage, D. Edwards, and N. Nic Daeid, "Hyperspectral imaging of gel pen inks: An emerging tool in document analysis," *Sci. Justice*, vol. 54, no. 1, pp. 71–80, 2014.

[3] R. Shiradkar, L. Shen, G. Landon, S. H. Ong, and P. Tan, "A new perspective on material classification and ink identification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2275–2282, 2014.

[4] M. Hoehse, A. Paul, I. Gornushkin, and U. Panne, "Multivariate classification of pigments and inks using combined Raman spectroscopy and LIBS," *Anal. Bioanal. Chem.*, vol. 402, no. 4, pp. 1443–1450, 2012.

[5] C. A. Teixeira and R. J. Poppi, "Discriminating blue ballpoint pens inks in questioned documents by Raman imaging and mean-field approach independent component analysis (MF-ICA)," *Microchem. J.*, vol. 144, pp. 411–418, 2019.

[6] Y. Z. Liu, J. Yu, M. X. Xie, Y. Liu, J. Han, and T. T. Jing, "Classification and dating of black gel pen ink by ion-pairing high-performance liquid chromatography," *J. Chromatogr. A*, vol. 1135, no. 1, pp. 57–64, 2006.

[7] C. S. Silva, F. de S. L. Borba, M. F. Pimentel, M. J. C. Pontes, R. S. Honorato, and C. Pasquini, "Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis," *Microchem. J.*, vol. 109, pp. 122–127, 2013.

[8] Z. Khan, F. Shafait, and A. Mian, "Hyperspectral imaging for ink mismatch detection," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 877–881, 2013.

[9] W. Wang, L. Zhang, D. Wei, Y. Zhao, and J. Wang, "The principle and application of hyperspectral imaging technology in detection of handwriting," *2017 9th Int. Conf. Adv. Infocomm Technol. ICAIT 2017*, pp. 345–349, 2018.

[10] L. Rodrigues e Brito, A. Braz, R. Saldanha Honorato, M. F. Pimentel, and C. Pasquini, "Evaluating the potential of near infrared hyperspectral imaging associated with multivariate data analysis for examining crossing ink lines," *Forensic Sci. Int.*, vol. 298, pp. 169–176, 2019.

[11] M. N. M. Asri, N. A. M. Noor, W. N. S. M. Desa, and D. Ismail, "Forensic Analysis of Gel Pen Inks using Hyperspectral Imaging coupled with Chemometric

Procedures," *Arab J. Forensic Sci. Forensic Med.*, vol. 1, no. 9, pp. 1174–1184, 2019.

[12] L. Yan and D. P. Roy, "Large-area gap filling of landsat reflectance time series by spectral-angle-mapper based spatio-temporal similarity (SAMSTS)," *Remote Sens.*, vol. 10, no. 4, 2018.

[13] Chein-I Chang, "Spectral information divergence for hyperspectral image analysis," *IEEE 1999 Int. Geosci. Remote Sens. Symp. IGARSS'99 (Cat. No.99CH36293)*, vol. 1, no. 1, pp. 509–511, 2003.

[14] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Modeling data manifold geometry in hyperspectral imagery," in *IEEE International IEEE International IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings.*, vol. 5, pp. 3203–3206, 2004.

[15] K. Vongsy and M. J. Mendenhall, "A comparative study of spectral detectors," 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, 2011, pp. 1-4..

[16] S. George and J. Y. Hardeberg, "Ink classification and visualisation of historical manuscripts: Application of hyperspectral imaging," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2015–Novem, pp. 1131–1135, 2015.

[17] "HySpex VNIR 1800." [Online]. Available: https://www.hyspex.no [Accessed: 08-August-2019].

[18] "3900E DC Regulated Lightsource" [Online]. Available: https://illuminationtech.com [Accessed: 08-August-2019].

[19] Labsphere, "Spectralon Targets," 2013. [Online]. Available: http://www.labsphere.com [Accessed: 08-August-2019].

[20] F. A. Kruse *et al.*, "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, no. 2–3, pp. 145–163, 1993.

[21] O. . De Carvalho and P. R. Meneses, "Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM). In: Summaries of the 9th JPL airborne earth science workshop," no. equation 2, 2000.

[22] J. C. Gower, "Properties of Euclidean and non-Euclidean distance matrices," *Linear Algebra Appl.*, vol. 67, no. C, pp. 81–97, 1985.

[23] D. R. Scott, "Effects of binary encoding on pattern recognition and library matching of spectral data," *Chemom. Intell. Lab. Syst.*, vol. 4, no. 1, pp. 47–63, 1988.