Forensic Science International 311 (2020) 110194

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE

Binu Melit Devassy*, Sony George

Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway

ARTICLE INFO

Article history:
Received 21 August 2019
Received in revised form 23 December 2019
Accepted 10 February 2020
Available online 12 February 2020

Keywords:
Dimensionality reduction
Hyperspectral imaging
Ink analysis
t-SNE
Visualisation

ABSTRACT

Ink analysis is an important tool in forensic science and document analysis. Hyperspectral imaging (HSI) captures large number of narrowband images across the electromagnetic spectrum. HSI is one of the non-invasive tools used in forensic document analysis, especially for ink analysis. The substantial information from multiple bands in HSI images empowers us to make non-destructive diagnosis and identification of forensic evidence in questioned documents. The presence of numerous band information in HSI data makes processing and storing becomes a computationally challenging task. Therefore, dimensionality reduction and visualization play a vital role in HSI data processing to achieve efficient processing and effortless understanding of the data. In this paper, an advanced approach known as t-Distributed Stochastic Neighbor embedding (t-SNE) algorithm is introduced into the ink analysis problem. t-SNE extracts the non-linear similarity features between spectra to scale them into a lower dimension. This capability of the t-SNE algorithm for ink spectral data is verified visually and quantitatively, the two-dimensional data generated by the t-SNE showed a better visualization and a greater improvement in clustering quality in comparison with Principal Component Analysis (PCA).

1. Introduction

Hyperspectral imaging (HSI) is a technology that captures the images in hundreds of narrow optical bands across the electromagnetic spectrum. HSI systems evolved as a result of advances in optical detector array fabrication that occurred in the late 1970s [1, pp. 18–19]. HSI imaging technique was initially invented to use in satellite imaging and crops monitoring [1, pp. 20–21] and later find applications in a variety of areas like food quality [2], medical imaging [3], material science [4], cultural heritage digitisation [5] and forensics [6].

In document forgery investigations, ink analysis plays an important role in the identification of ink type and age, which can be vital evidence in criminal prosecutions. While analyzing the types of ink in a questioned document, the presence of multiple inks may give an indication of some manipulations taken place. To get this evidence, a forensic expert needs to distinguish different inks used and this is difficult to perform visually. The human eye can detect different colors [7]; however, it may not differentiate colors that appear visually similar but spectrally different and this property is known as metamerism. More technically, it can be defined as; if two color stimuli have different spectral radiant power distributions but they possess a match in color for a given observer [8, p. 184]. The HSI images can vanquish this snag in the human visual system by extracting details from the abundant spectral components.

The recent trends in document analysis show a deviation towards non-destructive methods than destructive methods, which can preserve the evidence even after the analysis. Chromatography [9] and its variations [10] are the major technologies used in destructive document analysis. Rather than hyperspectral technology, there exist other non-contact techniques, such as colorimetric methods, absorption spectrum analysis, an examination by ultraviolet radiation, infrared radiation detection or infrared absorption [11]. In addition, there exist some more complex spectral based systems, such as FTIR (Fourier Transform Infra-Red) [12] and Raman Spectroscopy [13]

In this work, we are exploring the capability of t-Distributed Stochastic Neighbor Embedding (t-SNE) [14] algorithm in dimensionality reduction and visualization of hyperspectral images of ink. In HSI, the majority of applications require post-processing methods, which should achieve two basic and important goals. The first goal is to identify and classify the

* Corresponding author.
E-mail addresses: binu.m.devassy@ntnu.no (B. Melit Devassy), sony.george@ntnu.no (S. George).

http://dx.doi.org/10.1016/j.forsciint.2020.110194

materials for each pixel in the scene. The second goal is to lower the data volume or dimensionality, without dropping predominant information [15]. The idea behind dimensionality reduction and visualization is to achieve efficient processing of the data and easy to assimilate by a human analyst.

This report is focused on the second goal, the dimensionality reduction and visualization. The t-Distributed Stochastic Neighbor Embedding (t-SNE) [14] is selected as a candidate for this experiment and applied to various ink spectra to estimate the dimensionality reduction capability. To compare the performance, Principal Component Analysis (PCA) [16] is used as a standard reference. Apart from visual comparison, quantitative methods were used to get the clustering quality of both methods.

The major algorithms for achieving dimensionality reduction in hyperspectral imaging are, Principal Component Analysis (PCA) [17], Independent component analysis (ICA) [18] and Linear Discriminant Analysis (LDA) [19]. In this work, we explored t-SNE because of the following reasons. The first and important fact is that t-SNE is one among the few algorithms capable of retaining both local and global structure of the data simultaneously. Another important quality of t-SNE that attracts our attention is; it calculates the probability similarity of points in high dimensional space as well as in low dimensional space. Finally, t-SNE will work better for both linear and non-linear data rather than PCA, which struggle with non-linear data set. Because of the fact that PCA extracts a global linear model of the data, by projecting the data (n-dimensional) onto an m-dimensional ($m < n$) linear subspace defined by the leading eigenvectors of the original data's covariance matrix [20]. These are the few important experiments that used PCA as a tool for dimensionality reduction [17,21,22] also found attempts using few other algorithms such as Independent Component Analysis (ICA) [18] and few linear and statistical methods [23]. As this work is mainly concentrated on t-SNE based HSI classification, we found less number of relevant works in this stream. One of the important attempt done by *Emeline et al.* [24], they used t-SNE as dimensionality reduction method for hyperspectral data of paint pigments. A recent study was done by *Weijing et al.* [25] used an improved version of t-SNE for dimensionality reduction on remote sensing data. Both studies mentioned above revealed the advantage of t-SNE in hyperspectral dimensionality reduction. We also found a few more reports with hyperspectral data and t-SNE [26]; however, the specific focus was not on dimensionality reduction.

The remaining part of this report is organized into three parts; the first part will explain the HSI acquisition, sample preparation, algorithms and evaluation methods. The next part will discuss results, followed by a conclusion and possible future works.

2. Materials and methods

2.1. Hyperspectral acquisition

The hyperspectral camera HySpex VNIR-1800 [27] from Norsk Elektro Optikk AS is used for the preparation of HSI dataset. This pushbroom camera operated in visible and the near infrared region (VNIR) of the electromagnetic spectrum, between 0.4 μm and 1.0 μm, with a spectral sampling interval of 3.18 nm and produces 186 spectral bands. VNIR 1800 camera has a spatial resolution of 1800 pixels across the field of view, which is approximately 10 cm in length for the lens with a 30 cm focal length. The sample paper with or where the handwritten text was placed on a moving translator with the camera focusing on the sample from the top. The two halogen light sources were illuminate the sample at an angle of 45 degrees from the camera as shown in Fig. 1. A Contrast Multi-Step Target [28] was used as a reference, which was present in the scene. The software along with the VNIR-1800 camera

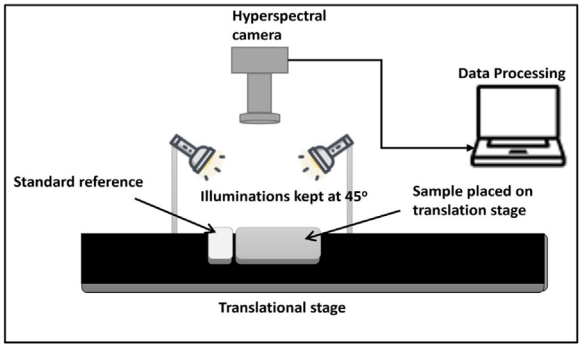


Fig. 1. Hyperspectral acquisition setup.

known as HySpex RAD performs radiometric calibrations. The HySpex RAD software converts the raw images into the sensor absolute radiance values, it also capable of correcting the non-uniformity and dark current factors during the processing.

2.2. Data

Data consist of hyperspectral images of handwritten texts from three different colored inks; they are blue, black and red from different manufactures and having different ink types like gel ink, ballpoint and liquid ink. Table 1 will give the details of pens and brands used. Most of the pens used in the present study are available internationally; this will help to reproduce the data set for ensuring the repeatability of the experiment. The samples were allowed to dry for 24 h before the acquisition. Fig. 2 shows a few sample texts from the original data set. The pens were marked with labels from one to twenty-five for each color, for example Pen 1, Pen 2, etc. the same labels are used throughout this paper. The popular saying “All roads lead to Rome” written by the same person is used as a sample text for all pens.

2.3. t-Distributed Stochastic Neighbor Embedding (t-SNE)

The t-SNE algorithm was introduced by van der *Maaten* and *Hinton* in 2008 [14] as an innovative tool for multidimensional scaling. This technique is now very popular in the machine learning community due to its remarkable ability to scale high dimensional data to lower dimensions. The algorithm starts by applying SNE (Stochastic Neighbor Embedding) to the data points, which converts the high-dimensional Euclidean distances between data points into conditional probabilities that represent similarities [14]. The similarity of data point xj to data point x_i is expressed by the conditional probability $p_{j|i}$, defined as in the below equation

$$P_{j|}i = \frac{\exp\left(\frac{-||x_i - x_j||^2}{2\sigma_i^2}\right)}{\sum_{k \neq i}\exp\left(\frac{-||x_i - x_k||^2}{2\sigma_i^2}\right)} \quad (1)$$

Then the probabilities in the original space are defined mathematically as in below equation

$$P_{i,j} = \frac{(P_{i|j} + P_{j|i})}{2n} \quad (2)$$

Where n represents the size of the data set. The t-SNE algorithm accepts an input parameter known as “perplexity” and it can be defined as the smooth measure of an effective number of neighbors [29]. Mathematically it can be expressed as

$$Perp(P_i) = 2^{H(P_i)}$$

Table 1
Details of pens used in this experiment.

	BLUE PENS		BLACK PENS		RED PENS	
ID	Brand	Model	Brand	Model	Brand	Model
1	Flair	Sunny	Pilot	Drawing Pen	BIC	Cristal Medium Red
2	Pilot	Frixion	Beifa Rx3026-00	Liquidly	Cello	Tri-Mate
3	Epoca	Svart	Unknown	Marketing Pen For: Sunvolden	Pilot	G-Knock
4	Uni	Jetstream	BIC	Cristal Medium Black	Pilot	Acroball
5	Papermate	Ink Joy	Schneider	Xpress	Pilot	G-2
6	Pentel	Energel	Pilot	Fixion	Pilot	V Ball Grip
7	Prodir	Z + F Zoller + Fröhlich	Pilot	G-Tec-C Maica	Pilot	G-Tec-C4
8	Pilot	Hi Tecpoint	Rorito	Fasrite – Think Faster, Write Faster	Pilot	V Sign Pen
9	Schneider	Xpress	Pental	Energel	Pilot	V Razor Point- Extra Fine
10	Cello	Papersoft	Pilot	Dr.Grip	Pilot	Frixion
11	Schneider	Tops 505 M	Pilot	V Sign Pen	Pilot	Super Grip-F
12	Pilot	G-Knock	Pilot	Feed-Gp4	Beifa Rx3026-00	Liquidly
13	Unknown	Marketing Pen For: Arc; www.He-Arc.Ch	TOTEM	DLX 18 C	Linc	Ocean Gel
14	Beifa Rx3026-00	Liquidly	Uni	Jetstream	Staedtler	Triplus Broadliner – Dry Safe
15	Flair	UNKNOWN	LINC	Pentonic	Reynolds	Racer Gel
16	Pentel	Energel			Schneider	Xpress
17	Pilot	Hi-Techpoint V5 Rt			Uni	Jetstream
18	BIC	Unknown			Reynolds	Brite Trendz
19	Unknown	Marketing Pen For: Nordic Choice Club - Choice.No/Se			Cello	Gel Tech NEON
20	Unknown	Marketing Pen For: Cartoon Network			Cello	Tri-Mate
21	Frixion	Ball Slim 038				
22	BIC	Cristal Medium Blue				
23	Pilot	Frixion				
24	Pilot	Rexgrip				
25	Uni	Power Tank				

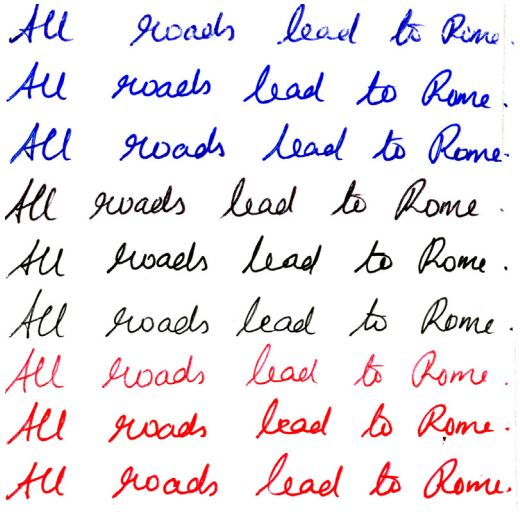


Fig. 2. Samples of visually similar pens used in this experiment.

Where $H(P_i)$ is the Shannon entropy, P_i measured in bits.

$$H(P_i) = -\sum_j P_{j|i} \log_2 P_{j|i}$$

Based on the pairwise distances of the points, this method automatically determines the variance σ_i, such that the effective number of neighbors coincide with the user-provided perplexity [30].

To avoid overcrowding, t-SNE employs the Student t-distribution with a single degree of freedom. Using this distribution the probability at low dimension q_{ij}, can be defined mathematically as in below equation

$$q_{ij} = \frac{(1 + \left||y_i - y_j\right||^2)^{-1}}{\sum_{k \neq l} (1 + ||y_k - y_l||^2)^{-1}}$$

Then the t-SNE algorithm finds the projections of the input data x_i in lower dimension as y_i, such that to lower the Kullback–Leibler [31] divergence between p_{ij} and q_{ij}, t-SNE employs a gradient-based technique for achieving this.

2.4. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a multivariate analysis technique and its goal is to extract the principal or important information from the input data, into a set of new orthogonal variables called principal components [33]. The first principal component defines the most variability of the input, and the second explains the next most variability, and so on. PCA is one of the well-established technique for dimensionality reduction and referred in several scientific papers in various domains [32,34], hence we selected PCA as a standard reference for the comparison.

2.5. Clustering performance evaluation

Clustering performance evaluation of an algorithm using visual inspection may not give the actual performance of the algorithm, a score is needed to define the quality of the clustering. This experiment used four different clustering evaluation methods based on the different aspects of the clustering. These methods require clustering information rather than visual information, hence the K-Means clustering algorithm [35] is used to predict the clusters from PCA and t-SNE processed data. As we are benchmarking the performance of t-SNE for ink analysis, we used known labeled data for getting performance matrix.

The *Silhouette Index* (SI) [36] is used as the first parameter since it defines how similar an object is to its own cluster (tightness) compared to other clusters (separation). The best value for SI is 1.0, the worst value is −1.0, and the values near 0.0 indicate overlapping between clusters. The second method is *Normalized Mutual Information* (NMI) [37], which provides mutual information between clusters. In the clustering context, mutual information is

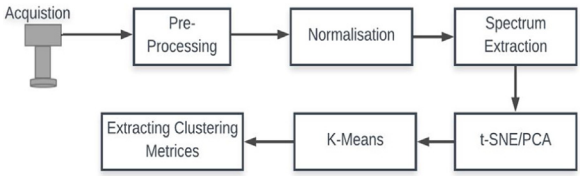


Fig. 3. Processing pipeline.

an estimate that measures the overlapping between clusters. NMI score varies between 0.0 and 1.0, 1.0 means the best clustering with respect to the given labels. These two criteria will define the goodness of a clustering algorithm; however, to get a more intuitive metric about the clustering quality, two more parameters are used. They are *Homogeneity Index* (HI) and Completeness Index (CI) [38]. Homogeneity defines whether each cluster contains only the data points from a single class or not. The completeness score tells us that, whether all data points having the same labels are assigned to the same cluster or not. Both CI and HI scores are bounded between 0.0 and 1.0, higher values represent better clustering.

2.6. Data processing

Fig. 3 will give a brief idea about the data processing pipeline. The image acquisition part contains the hyperspectral imaging system and the preprocessing performed by the camera software. The preprocessing includes sensor corrections and dark offset correction of image data. In the normalization part of the workflow, we calculate the absolute reflectance using the reference target [28] of known reflectance acquired along with the object. The ink spectra will be extracted from the specific position in normalized reflectance data. This spectra will feed into PCA and t-SNE algorithms for dimensionality reduction. To estimate the clustering performance, K-Means clustering was applied to the output data

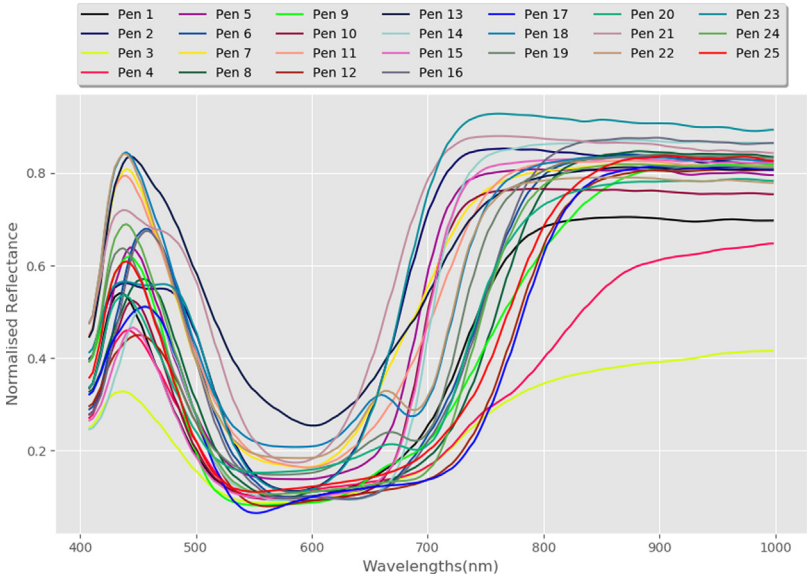


Fig. 4. Mean spectra of blue pens (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

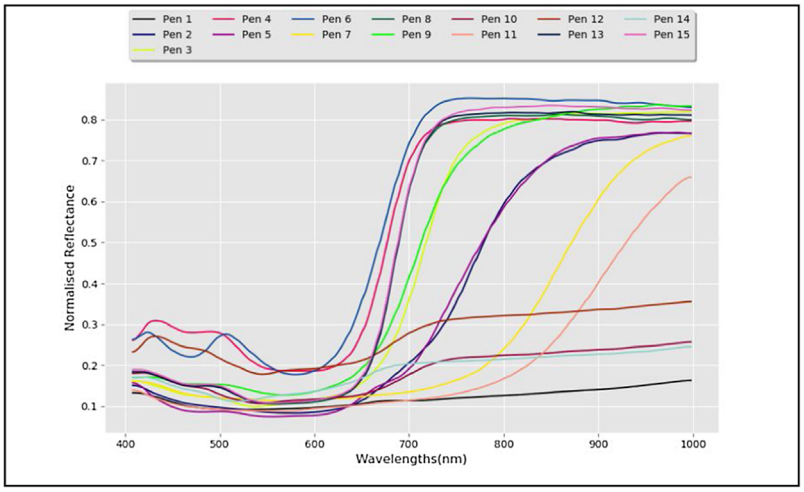


Fig. 5. Mean spectra of black pens.

from PCA and t-SNE. The clustering performance matrices will be calculated from the K-Means results using the known labels. This experiment used 25 blue, 20 red and 15 black pens and used 50 spectra from each pen. The Figs. 4–6 shows the mean spectra of these pens.

3. Results and discussions

The processed HSI images with reduced dimension (2-Dimensional) are shown in Figs. 7–9. From visual inspection, it is easy to identify that t-SNE provides better visualization of

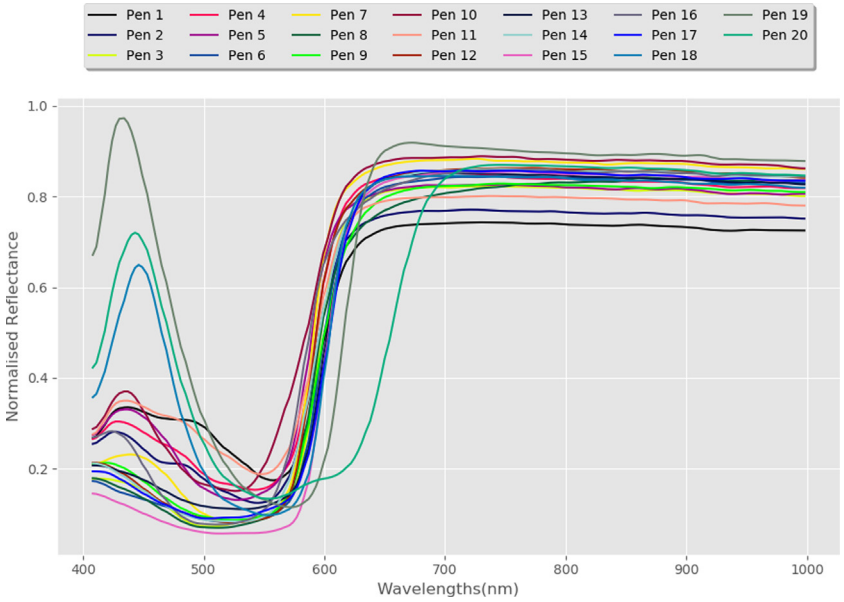


Fig. 6. Mean spectra of red pens (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

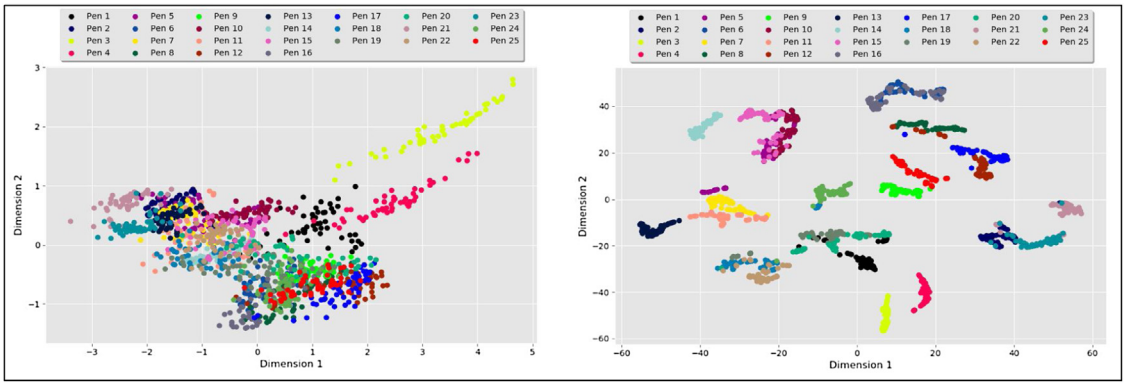


Fig. 7. Blue pens, left image shows PCA result and the right image shows t-SNE result (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

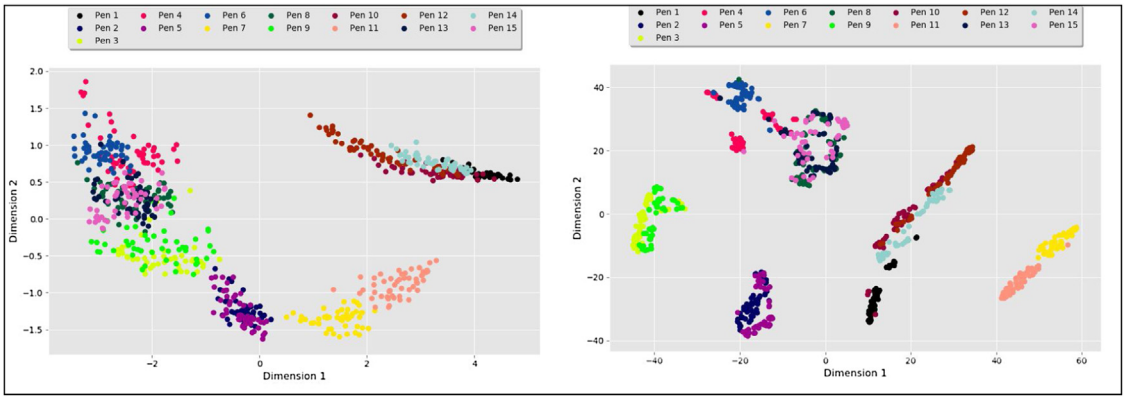


Fig. 8. Black pens, left image shows PCA result and the right image shows t-SNE result.

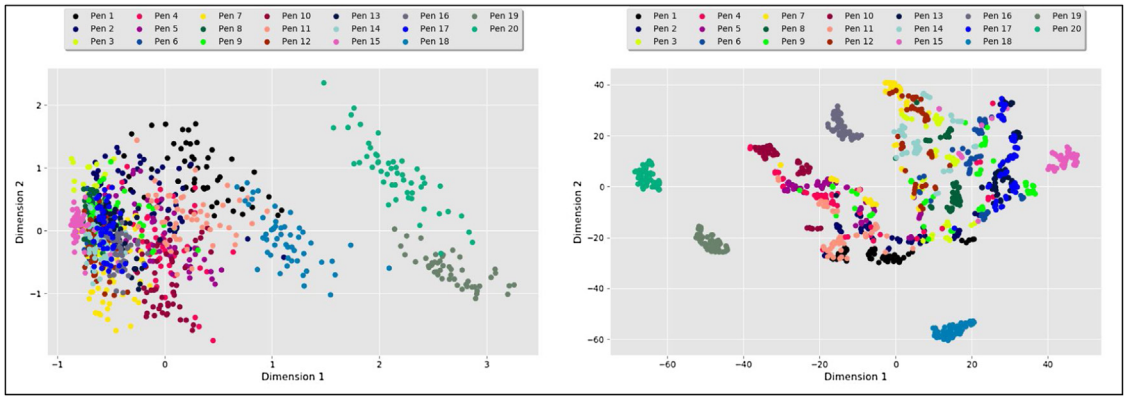


Fig. 9. Red pens, left image shows PCA result and the right image shows t-SNE result (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

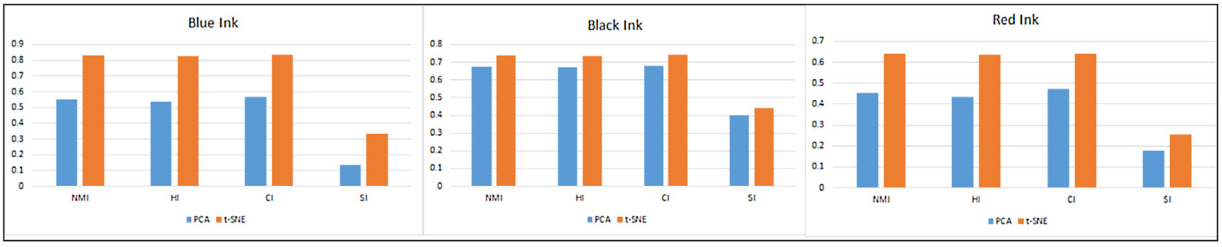


Fig. 10. Clustering indexes comparison between PCA and t-SNE.

Table 2
Average values of clustering indexes.

	Blue Ink		Black Ink		Red Ink	
	PCA	t-SNE	PCA	t-SNE	PCA	t-SNE
NMI	0.55	0.83	0.68	0.74	0.45	0.64
HI	0.53	0.83	0.67	0.73	0.44	0.64
CI	0.56	0.84	0.68	0.74	0.47	0.64
SI	0.13	0.34	0.40	0.44	0.17	0.25

clusters. To ensure the findings from visual inspection, we checked the clustering indexes calculated for all three inks. The t-SNE outperforms PCA in all four performance indexes used, as illustrated in Fig. 10. From the clustering indexes, it revealed that the t-SNE outperforms PCA for this specific task. The PCA always tries to find a linear relationship between data points; however, t-SNE extracts the non-linear relationship, which enables t-SNE to produce a better clustering.

Table 2 shows the average values of clustering indexes obtained for all pens. For blue inks, the NMI, HI and CI values showing above 80 % on average, which indicates a very good clustering score comparing to less than 60 % that of PCA. In addition to that, the SI index of t-SNE registered as significantly higher than that of PCA for all colors. The NMI, HI and CI values for blue inks t-SNE looks higher than other black and red, this can be explained the basis of the mean spectra of the inks shown in Figs. 4–6. The blue pens possess more spectra with a discrete signature than that of black and red and this behavior can be confirmed from the t-SNE results shown in Figs. 7–9.

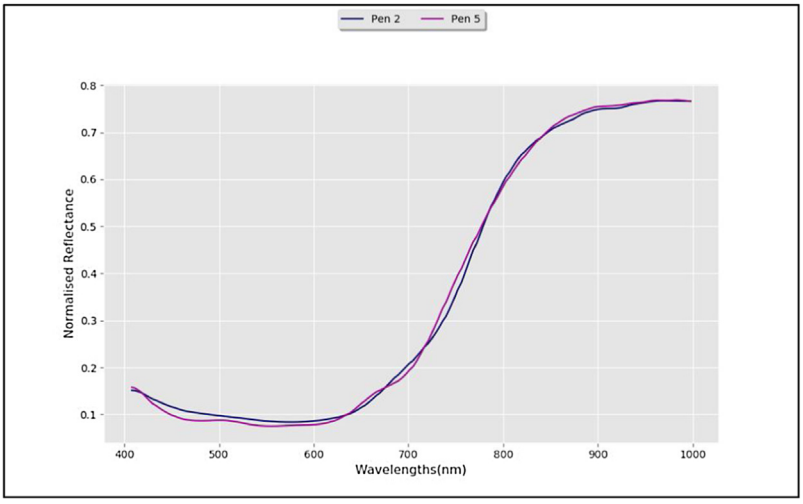


Fig. 11. Nearly identical spectra from black pens.

Table 3
Average values of clustering indexes after removing nearly identical spectra.

	Blue Ink		Black Ink		Red Ink	
	PCA	t-SNE	PCA	t-SNE	PCA	t-SNE
NMI	0.73	0.94	0.78	0.87	0.71	0.97
HI	0.7	0.94	0.78	0.87	0.7	0.97
CI	0.73	0.94	0.79	0.87	0.73	0.97
SI	0.34	0.49	0.45	0.49	0.33	0.43

After analyzing the spectra and combined the information from this report [39], we identified that the pens having nearly identical spectral signature causing a decline in clustering quality. Fig. 11 shows 2 sample spectra from black pens, the pen 2 and 5 have almost similar spectra in the wavelength range used. To substantiate our finding, we manually removed the spectra having nearly similar shapes and executed the cluster quality test again. The obtained results are given in Table 3, which shows a significant improvement in clustering quality for both PCA and t-SNE. These results are aligned with our findings, with t-SNE has an upper hand over PCA, the details of the pens, mean spectra and their clustering graphs are given in Appendix A.

In addition to these advantages, t-SNE is computationally more expensive compared to PCA, this might be considered as a possible drawback of t-SNE. In our case, for 50 samples from 10 different pens, t-SNE consumes 1.7 s in average while PCA consumed 0.05 s. The performance was measured in an Intel Core i7 8650U CPU with 16 GB of RAM, the PCA performed almost 34 times faster than that of the t-SNE. We can tune the parameter perplexity and learning rate for t-SNE algorithm, compared to the direct processing of PCA. In some scenarios, this tuning might be considered as extra processing comparing to PCA.

4. Conclusion and future work

In this work, we introduced t-SNE dimensionality reduction into the application area of ink analysis. Efforts are also made to estimate the performance of t-SNE using different indexes and the results are compared against PCA. According to the results obtained the t-SNE outperformed PCA for dimensionality reduction of ink spectral data. In addition, we found that the non-linear dimensionality reduction method is more suitable for ink spectral data than a linear method. To accept the previous statement completely, the presented research needs to be extended to more linear and non-linear methods. Another important outcome derived from this research is the importance of considering computational complexity while developing dimensionality reduction algorithms.

CRediT authorship contribution statement

Binu Melit Devassy: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation. **Sony George:** Supervision, Conceptualization, Methodology, Data curation, Writing - review & editing.

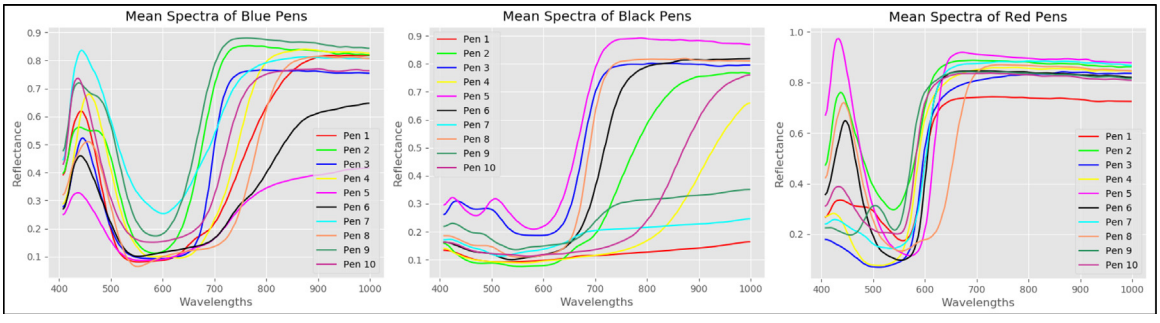
Acknowledgment

The authors would like to offer their thanks to Norsk Elektro Optikk, Norway for their support in conducting this research.

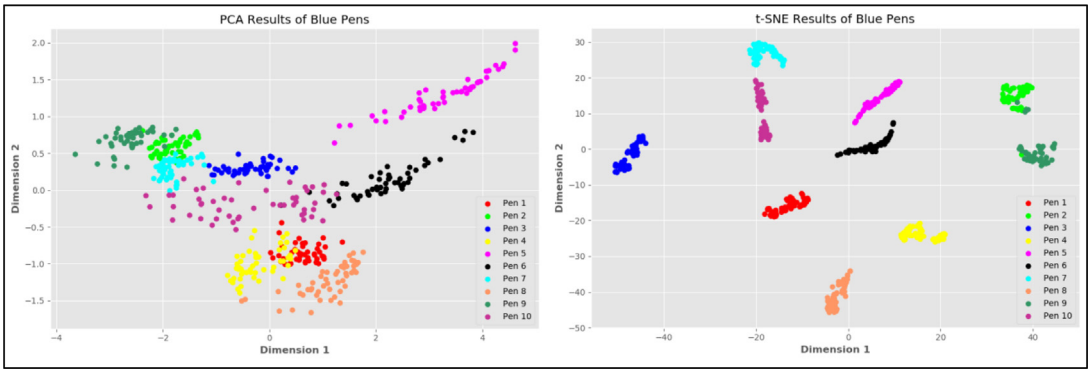
Appendix A. Details experiment after removing pens having nearly identical spectra

	BLUE PENS		BLACK PENS		RED PENS	
ID	BRAND	MODEL	BRAND	MODEL	BRAND	MODEL
1	Schneider	Xpress	Pilot	Drawing Pen	BIC	Cristal Medium
2	Pilot	Frixion	Schneider	Xpress	Pilot	Frixion
3	Cello	Paper Soft	BIC	Cristal Medium	Pilot	G-Tech C4
4	Pentel	Energel	Pilot	V-Sign Pen	Schneider	Xpress
5	Epoca	Svart	Pilot	Frixion	Cello	Fel Tech Neon
6	UNI	Jet Stream	Sunvolden	Not Available	Reynolds	Brite Trendz 39/18
7	Arc	Not Available	UNI	Jet Stream	B-Len	Not Available
8	Pilot	Hi-Techpoint v5 RT	Totem	DLX 18C	Cello	Trimate
9	Frixion	Ball Slim 038	Pilot	Acro Ball	Reynolds	Brite Trendz 35/18
10	Paper Mate	Replay	Pilot	G-Tec-C	Zebra	Tapli Clip

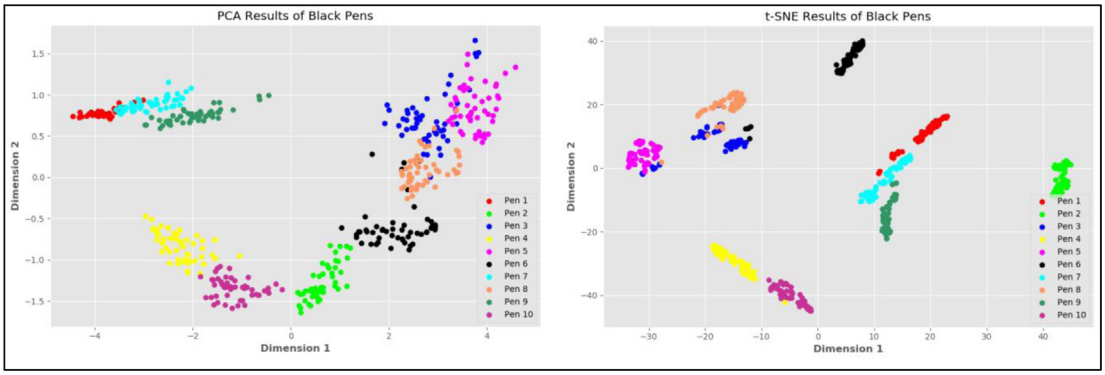
Pens list used after removing nearly identical spectra.



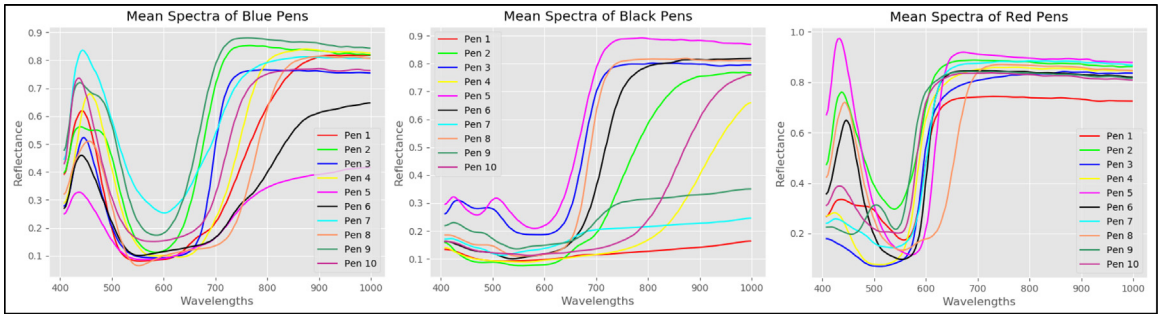
Mean Spectra of the three inks after removing nearly identical spectra.



PCA vs t-SNE results of blue pens indexes after removing nearly identical spectra.



PCA vs t-SNE results of black pens indexes after removing nearly identical spectra.



PCA vs t-SNE results of red pens indexes after removing nearly identical spectra.

References

[1] C.I. Chang, Hyperspectral Data Exploitation: Theory and Applications, (2006) .
[2] G. ElMasry, M. Kamruzzaman, D.W. Sun, P. Allen, Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review, Crit. Rev. Food Sci. Nutr. 52 (11) (2012) 999–1023.
[3] G. Lu, B. Fei, Medical hyperspectral imaging: a review, J. Biomed. Opt. 19 (1) (2014) 10901.
[4] P. Tatzer, M. Wolf, T. Panner, Industrial application for inline material sorting using hyperspectral imaging in the NIR range, Real Time Imaging 11 (2) (2005) 99–107.
[5] C. Fischer, I. Kakoulli, Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications, Estud. Conserv. E Restauro 51 (sup1) (2014) 3–16.
[6] G.J. Edelman, E. Gaston, T.G. van Leeuwen, P.J. Cullen, M.C.G. Aalders, Hyperspectral imaging for non-contact analysis of forensic traces, Forensic Sci. Int. 223 (1–3) (2012) 28–39.
[7] Lightness and retinex theory, J. Opt. Soc. Am. 61 (1) (1971) 1–11.
[8] W.S. Wyszecki, Günter, Stiles, Color science: concepts and methods, quantitative data and formulae, Color Res. Appl. (2000).

[9] J.A. Zlotnick, F.P. Smith, Chromatographic and electrophoretic approaches in ink analysis, J. Chromatogr. B Biomed. Sci. Appl. 733 (1–2) (1999) 265–272.
[10] A. Kher, M. Mulholland, E. Green, B. Reedy, Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis, Vib. Spectrosc. 40 (2) (2006) 270–277.
[11] A. Morales, M.A. Ferrer, M. Diaz-Cabrera, C. Carmona, G.L. Thomas, The use of hyperspectral analysis for ink identification in handwritten documents, Proceedings - International Carnahan Conference on Security Technology (2014) vol. 2014–Octob, no. October.
[12] J. Zięba-Palus, M. Kunicki, Application of the micro-FTIR spectroscopy, Raman spectroscopy and XRF method examination of inks, Forensic Sci. Int. 158 (2–3) (2006) 164–172.
[13] A. Braz, M. López-López, C. García-Ruiz, Raman spectroscopy for forensic analysis of inks in questioned documents, Forensic Sci. Int. 232 (1–3) (2013) 206–212.
[14] L. Van Der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2625.
[15] J.C. Harsanyi, C.I. Chang, Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach, IEEE Trans. Geosci. Remote Sens. 32 (4) (1994) 779–785.
[16] M.E. Timmerman, Principal Component Analysis 98 (464) (2003).
[17] M.D. Farrell, R.M. Mersereau, On the impact of PCA dimension reduction for hyperspectral detection of difficult targets, IEEE Geosci. Remote Sens. Lett. 2 (2) (2005) 192–195.
[18] J. Wang, C.I. Chang, Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis, IEEE Trans. Geosci. Remote Sens. 44 (6) (2006) 1586–1600.
[19] T.V. Bandos, L. Bruzzone, G. Camps-Valls, Classification of hyperspectral images with regularized linear discriminant analysis, IEEE Trans. Geosci. Remote Sens. 47 (3) (2009) 862–873.
[20] N. Kambhatla, T.K. Leen, Fast non-linear dimension reduction, IEEE International Conference on Neural Networks – Conference Proceedings, Vol. 1993–Janua, 1993, pp. 1213–1218.
[21] E. Martel, et al., Implementation of the Principal Component Analysis onto high-performance computer facilities for hyperspectral dimensionality reduction: results and comparisons, Remote Sens. (Basel) 10 (6) (2018).
[22] W. Sun, Q. Du, Graph-regularized fast and robust principal component analysis for hyperspectral band selection, IEEE Trans. Geosci. Remote Sens. 56 (6) (2018) 3185–3195.
[23] N. Renard, S. Bourennane, J. Blanc-Talon, Denoising and dimensionality reduction using multilinear tools for hyperspectral images, IEEE Geosci. Remote Sens. Lett. 5 (2) (2008) 138–142.
[24] E. Pouyet, N. Rohani, A.K. Katsaggelos, O. Cossairt, M. Walton, Innovative data reduction and visualization strategy for hyperspectral imaging datasets using t-SNE approach, Pure Appl. Chem. 90 (3) (2018) 493–506.
[25] W. Song, L. Wang, P. Liu, K.K.R. Choo, Improved t-SNE based manifold dimensional reduction for remote sensing data processing, Multimed. Tools Appl. 78 (4) (2019) 4311–4326.
[26] J. Zhang, L. Chen, L. Zhuo, X. Liang, J. Li, An efficient hyperspectral image retrieval method: deep spectral-spatial feature extraction with DCGAN and dimensionality reduction using t-SNE-based NM hashing, Remote Sens. 10 (2) (2018).
[27] "VNIR 1800." [Online]. Available: https://www.hyspex.no. [Accessed: 20-Aug-2019].
[28] "Contrast Multi-Step Target." [Online]. Available: https://www.labspherestore.com/.
[29] C.R. García-Alonso, L.M. Pérez-Naranjo, J.C. Fernández-Caballero, Multi-objective evolutionary algorithms to identify highly autocorrelated areas: the case of spatial distribution in financially compromised farms, Ann. Oper. Res. 219 (1) (2014) 187–202.
[30] A. Gisbrecht, A. Schulz, B. Hammer, Parametric nonlinear dimensionality reduction using kernel t-SNE, Neurocomputing 147 (1) (2015) 71–82.
[31] J.F.C. Kingman, S. Kullback, Information Theory and Statistics vol. 54, no. 387, Dover Publications Inc, Mineola, New York, 2007.
[32] A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, M.A. Sid-Ahmed, Human face recognition based on multidimensional PCA and extreme learning machine, Pattern Recognit. 44 (10–11) (2011) 2588–2597.
[33] D.J. Bartholomew, Principal components analysis, Int. Encycl. Educ. (2010) 374–377.
[34] M.E. Timmerman, Principal component analysis, J. Am. Stat. Assoc. 98 (June (464)) (2003) 1082–1083.
[35] J.A. Hartigan, M.A. Wong, Algorithm AS 136: a K-Means clustering algorithm, Appl. Stat. 28 (1) (2006) 100.
[36] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (C) (1987) 53–65.
[37] A.F. McDaid, D. Greene, N. Hurley, Normalized Mutual Information to evaluate overlapping community finding algorithms, arXiv e-prints (October) (2011) p. arXiv:1110.2515.
[38] A. Rosenberg, J. Hirschberg, V-measure: a conditional entropy-based external cluster evaluation measure, Comput. Linguist. 1 (June) (2007) 410–420.
[39] B.M. Devassy, S. George, Ink classification using convolutional neural network, Norsk Informasjonssikkerhetskonferanse 2019, NISK, 2019.