WATER-BODY SEGMENTATION IN SATELLITE IMAGERY APPLYING MODIFIED KERNEL K-MEANS

Paria Yousefi, Hamid A. Jalab, Rabha W. Ibrahim, Nurul F. Mohd Noor*,
Mohamad N. Ayub, Abdullah Gani

Department of Computer Systems & Technology, Faculty of Computer Science and Information Technology, University Malaya, 50603 Kuala Lumpur, Malaysia

Email: pr.usefi@gmail.com, hamidjalab@um.edu.my, rabhaibrahim@um.edu.my, fazmidar@um.edu.my, nizamayub@um.edu.my, abdullah@um.edu.my

ABSTRACT

The main purpose of k-Means clustering is partitioning patterns into various homogeneous clusters by minimizing cluster errors, but the modified solution of k-Means can be recovered with the guidance of Principal Component Analysis (PCA). In this paper, the linear Kernel PCA guides k-Means procedure using filter to modify images in situations where some parts are missing by k-Means classification. The proposed method consists of three steps: 1) transformation of the color space and using PCA to solve the eigenvalue problem pertaining to the covariance matrices of satellite image; 2) feature extraction from selected eigenvectors and are rearranged by applying the training map to extract the useful information as a set of new orthogonal variables called principal components; and 3) classification of the images based on the extracted features using k-Means clustering. The quantitative results obtained using the proposed method were compared with k-Means and k-Means PCA techniques in terms of accuracy in extraction. The contribution of this approach is the modification of PCA selection to achieve more accurate extraction of the water-body segmentation in satellite images.

Keywords: *PCA, k-Means clustering, statistical pattern recognition, water feature extraction, image segmentation, satellite images, contextual filter*

1.0   INTRODUCTION

In the real world, datasets generally express their variables relationship, and are statistically independent of the feature vectors. However, one can safely eliminate the least discriminative feature in any dataset. This is because data may contain either many redundant features which do not provide additional useful information or irrelevant features that have no effective informational value in any context than the currently chosen features. Using a single feature could express a combination of various types of information by a single value.

The Principal Component Analysis (PCA) is a versatile technique as it allows data to be recovered exactly as its original format. Assume the covariance matrix of the data $x_k, k = 1, \ldots, l, x_k \in R^N, \sum_{k=1}^{l} x_k = 0$, is defined as [1]:

$$C = \frac{1}{l} \sum_{j=1}^{l} x_j\, x_j^T \qquad (1)$$

The orthogonal linear transformation that transforms the data to the new coordinate system (onto the Eigenvectors), is called principle components. Assume a data matrix X is contained the n rows and the p columns; each rows represents a different repetition of the experiment, and each columns gives a particular kind of features. The transformation is dimensional vectors of weight that map each row vector to a new vector of principle component scores [2] $(t_{(i)})$: $t_{k(i)} =$

---

* corresponding author

143

Malaysian Journal of Computer Science. Vol. 31(2), 2018

$x_{(i)} \cdot w_{(k)}$. This transformation maps a data vector x from an original space to a new space which are uncorrelated over the dataset. This is useful for data reduction, noise reduction, visualization, and data compression. PCA is widely used in many applications to extract the most relevant dataset information such as for face recognition, object recognition, change detection [3], etc. The PCA technique was first developed in 1987 for human face recognition [4], and in 1991, it was used as Eigen faces for face detection and recognition [5]. In a face database, all face images are represented as very long vectors - not the usual matrix - and this takes up the entire image space, where each image is a point. It is important to consider, not all the principle components need to be kept. The truncated transformation makes by keeping only the first L principle components (producing by using only the first L loading vectors) [6]. The application of PCA for dimensionality reduction involves projecting the data onto the largest eigenvectors of it covariance matrix using independent components to reduce the dimensionality of the feature space. In order to minimize projection errors, the k largest eigenvectors are chosen for data projection so that, the total least squared regression or orthogonal regression is minimized in all directions.

Another technique - k-Means clustering - is used to disclose contextual structures of multivariate datasets. It is useful for extracting clusters with iterative optimization, but its drawback is the initialization of convergence for local minima [7], [8]. Ding and He [9] proposed the PCA-guided k-Means technique to find the optimal k-Means partition by data projections on principal directions. Honda [8] used a fuzzy PCA-guided method to robust k-Means clustering to perform k cluster-core identification, in which indicated clusters are calculated in a batch process by taking into consideration the weight of samples.

Accurate information about the distribution of open surface water is very important in many scientific disciplines. Pixel-based methodologies have been widely used to identify water features in images. These methods involve selecting a single band from multispectral images and defining the threshold value subjectively. Such a method, by itself, might lead to underestimation or overestimation of water area extraction [10]. An improved method - simple ratio method [11] - uses two bands from a multispectral image. The advantage of using different spectral responses from different land-covers is that it enhances the water features in the image. However, the extracted water features are usually mixed with the signal from built-up land [12].

Statistical pattern recognition techniques have been used to extract land-cover features, including waterbody, and are different from traditional pixel-based methodologies. For instance, unsupervised Iterative Self-Organizing Data Analysis Technique (ISODATA) involves grouping multi-band spectral responses into spectrally pure clusters [13]. However, after processing the clusters must be assigned with actual land-covers (e.g. water) by using reference data or field collected data.

The integration of the pixel-based classification and object-based segmentation will give rise to other methods especially the integrated method. This method identifies waterbodies in remote sensing imagery based on the characteristic spectral reflectance curves [14]. However, this method is less efficient in handling water feature pixels that are mixed with other land-cover types such as built-up land and vegetation.

The objective of this study is to propose a new integrated method of pixel-based classification and statistical pattern recognition method for water features extraction in images. PCA is applied to reduce the number of features to classify by removing non-useful features. The newly selected features are then passed through a contextual filter, and finally, the water area (part of the river) is extracted by K-means clustering classification. The contribution of this approach is an improvement in PCA selection using K-means clustering to achieve the accurate extraction of the water-body.

The remaining sections of this paper are organized as follows: Section II describes the proposed approach, the flowchart of the proposed algorithm with details of each step. Section III presents analysis of the experimental results based on the satellite imagery of Terengganu River, and compares the results using the proposed method with the results of previous works. Section IV presents the conclusion and suggestion for future work.

144

Malaysian Journal of Computer Science. Vol. 31(2), 2018

2.0   PROPOSED APPROACH

The proposed approach includes four steps: 1) pre-processing; 2) PCA; 3) proposed contextual filter; and 4) k-Means clustering. A binary logo image will be used as the embedded watermark. Fig. 1 shows the flowchart of the proposed algorithm.

 In this study, we used the high resolution satellite images taken by National Oceanic and Atmospheric Administration (NOAA) and Digital Globe's QuickBird.
The study area covers the Kuala Terengganu district of Malaysia. This region is selected as a case study because of its complex geographical features. In order to support the proposed approach and the results of the study, the proposed approach was applied to two other areas: north-eastern Germany which adjoins the North Sea, and the Sea of Azov in south-eastern Europe.

2.1   Pre-Processing

The subject of statistics is based on the idea that having a dataset, we wish to analyze the relationships between the individual points or components of this dataset. In our study, our dataset consists of three-column values as per color image of *R, G* and *B*. When we want to use the red, green, and blue components of each pixel to classify an image, statistically, it presents a correlated feature problem. The *R, G, B* components of a pixel are statistically correlated because the image sensors that are most sensitive to red light also capture some blue and green light, and similarly for the blue and the green light sensors. Hence, removal of the red component from the feature vector, implicitly also eliminates the *G* and *B* channels. Therefore, before eliminating any features, the entire feature color space *(R, G* and *B)* must be transformed such that the underlying uncorrelated components are identified. In addition, our datasets also include urban areas with vegetation. Thus, to make the dimensions more distinctive, we rearrange the *RGB* values set to *HSV* color space in which its components and colorimetry are relative to the *RGB* color space. To process this statistically, we embed all dimensions into one matrix.
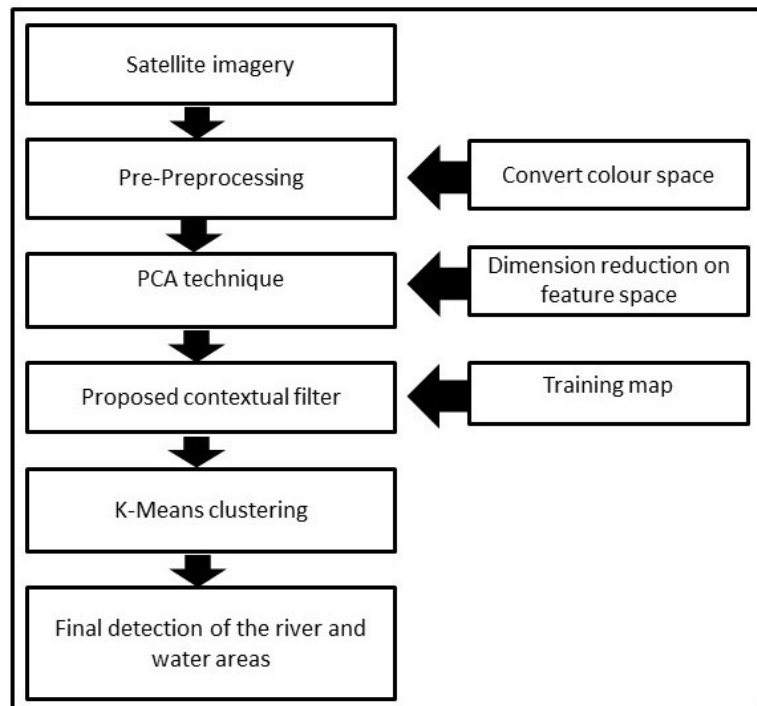

Fig. 1: Flowchart of the proposed algorithm

145

Malaysian Journal of Computer Science. Vol. 31(2), 2018

2.2   PCA

Feature selection is an essential step in image classification because it helps in removing redundancy and irrelevant input features to reduce learning time for computing and modifying and to improve accuracy in classification. There are three feature selection methods: 1) filter [15], [16]; 2) wrapper [17], [18]; and 3) hybrid methods [19], [20]. Selecting the feature subset using the filter methods involves a pre-processing step based on a chosen criterion of the classifier. In the wrapper method, the selection method directly optimizes the predictor performance. The hybrid method is a combination of the wrapper and filter methods to guide the classifier [21].

PCA is a linear dimensionality reduction technique which searches for projection in the maximum variability directions. It computes the eigenvalues and eigenvectors of the data covariance matrix. Eigen vectors are sorted eigenvalues and the actual data is projected into the eigenvectors directions [22].

To center and scale the data, PCA will first standardize the data, to summarize it, and then the sample mean vector and standard deviation vector are calculated. Standardization of data means subtracting the sample mean from each observation, and dividing the product by the sample standard deviation.

For an N dimensional image sample, the covariance matrix of order $N \times N$ is calculated using the equation, below [22]:

$$\pounds = \frac{1}{N}\{(x - \bar{x})(x - \bar{x})^T\} \qquad (2)$$

where x is the given matrix with N dimension, and $\bar{x}$ is its mean vector.

In the next step, from the covariance matrix $\pounds$, eigenvectors V and diagonal elements of matrix D as eigenvalues, we will find [22],

$$V^{-1}\pounds V = D \qquad (3)$$

After sorting the eigenvectors, data is projected in the direction of the sorted eigenvector [22]. In the three dimensional set (hue, saturation and value), data can be projected onto the surface defined by the two largest eigenvectors to occupy a 2D feature space or it can be projected onto the largest eigenvector to obtain a 1D feature space. Two biggest eigenvectors corresponded with two biggest eigenvalues are chosen to reduce the image dimensions.

2.3   Proposed contextual filter

Filter feature selection is aimed at finding the relevant features set within the training map. The extra objects in the database which is connected to the water area mostly have similar color properties, and this makes it challenging to find the water pixel color range. K-means clustering alone is not enough to resolve this problem. PCA is used to select the most useful features and the proposed filter modifies the water-body segmentation process.

There is an advantage of using the training map for removing certain ranges of the hue. In the training map, the various hue values are considered in order to separate the river and the dam pixels from each other. We used the discrimination ratio (empirically fixed between 0 and 0.4) to separate the water from the dam. This delineation requires the pixel coordinate color hue information, and the old values of the hue are replaced. The other values are remained as they were.

2.4   K-means clustering

The basic task in data mining is clustering which involves grouping similar sets together and differentiating them from other sets. The clustering method is divided into four basic groups: 1) Connectively-based clustering [23]; 2) Centroid-based clustering [24]; 3) Distribution-based clustering [25]; and 4) Density-based clustering [26]. In connectively-based clustering, which is based on the core, different distances will form different clusters. In centroid-based clustering, the clusters are represented by a central vector, which may not necessarily be a member of the dataset. In this method, we assign objects to the nearest center of k clusters. To find the nearest cluster center, we can use squared distances, etc. In distribution-based clustering, the clustering objects belong to the same distribution. In density-based clustering, a cluster is defined based on the higher density area being grouped together, and the rest of the objects is assumed as noise. Since clustering is the grouping of similar objects, it is necessary to have some sort of measurement such as distance measurement, and similarity measurement to determine similarities/dissimilarities between two objects.

One of the centroid-based clustering techniques is K-means. It is an unsupervised algorithm which classifies P different features of similar objects in a feature vector to K clusters. The objective function is defined as the sum of within cluster errors expressed as follows:

146

Malaysian Journal of Computer Science. Vol. 31(2), 2018

$$L_{km} = \sum_{k=1}^{K} \sum_{i \in G_k} \|x_i - b_k\|^2 \qquad (4)$$

where K is the number of clusters, and b_k is the centroid of cluster G_k [8]. Certain centroid of pre-defined number of cluster K will be estimated including the condition to be as far as possible of each other, else it will give rise to different clusters. Sample assignment and centroid estimation are iterated until the solution is trapped at a local minimum.
The advantages of using this algorithm include ease of performance, rapid convergence and compatibility for sparse data [23]. However, the disadvantage is that it is sensitive to the number of cluster chosen.

3.0   EXPERIMENTAL RESULTS AND DISCUSSION

Satellite imagery segmentation enforces the combination of features information. A feature structure is a general purpose data structure which identifies and groups individual features, each of which associates a name with one or more values. Because of the generality of feature structures, they can be used to represent many different kinds of information. The statistical process helps to distinguish the main features from the other features. PCA of the data matrix will extract the dominant patterns in the matrix in terms of a complementary set of scores and loading plots. Statistical information is suitable to exploit spectral information and make a training map. The segmentation of the river in the proposed method is achieved by using k-Means clustering to classify the features that are related to each other.

This section discusses the experimental results to demonstrate the characteristic features of our proposed algorithm. The performance of the proposed method is compared with that of K-means clustering and the PCA k-Means techniques with respect to their accuracy in water extraction. The static information in k-Means clustering using context and spectral information lacks accuracy in water extraction, especially in urban area. K-Means selects the cluster randomly and then normalizes the result. K-Means PCA technique needs to be improved in this aspect, too. The results of our experiment are shown in Fig. 2, and the three techniques were compared with respect to their ability to segregate the water from the dam area. Fig. 2(1-b) shows the results obtained by applying k-Means PCA to the original image in Fig. 2(1-a). It shows that certain part of the water has been removed by using k-Means clustering technique based on the definition of number of clusters to better distinguish the water area from the dam. However, the results shows some effects on some parts of the water because of the reflection of the dam color on the water. The boundary of the water remains, as it is part of the dam area, which will be reflected in the extraction result when using the k-Means PCA technique. In contrast, our proposed method can overcome this problem by using the modified filter, which exploits the training map in the chosen eigenvectors to indicate separation of the water from the dam.

147

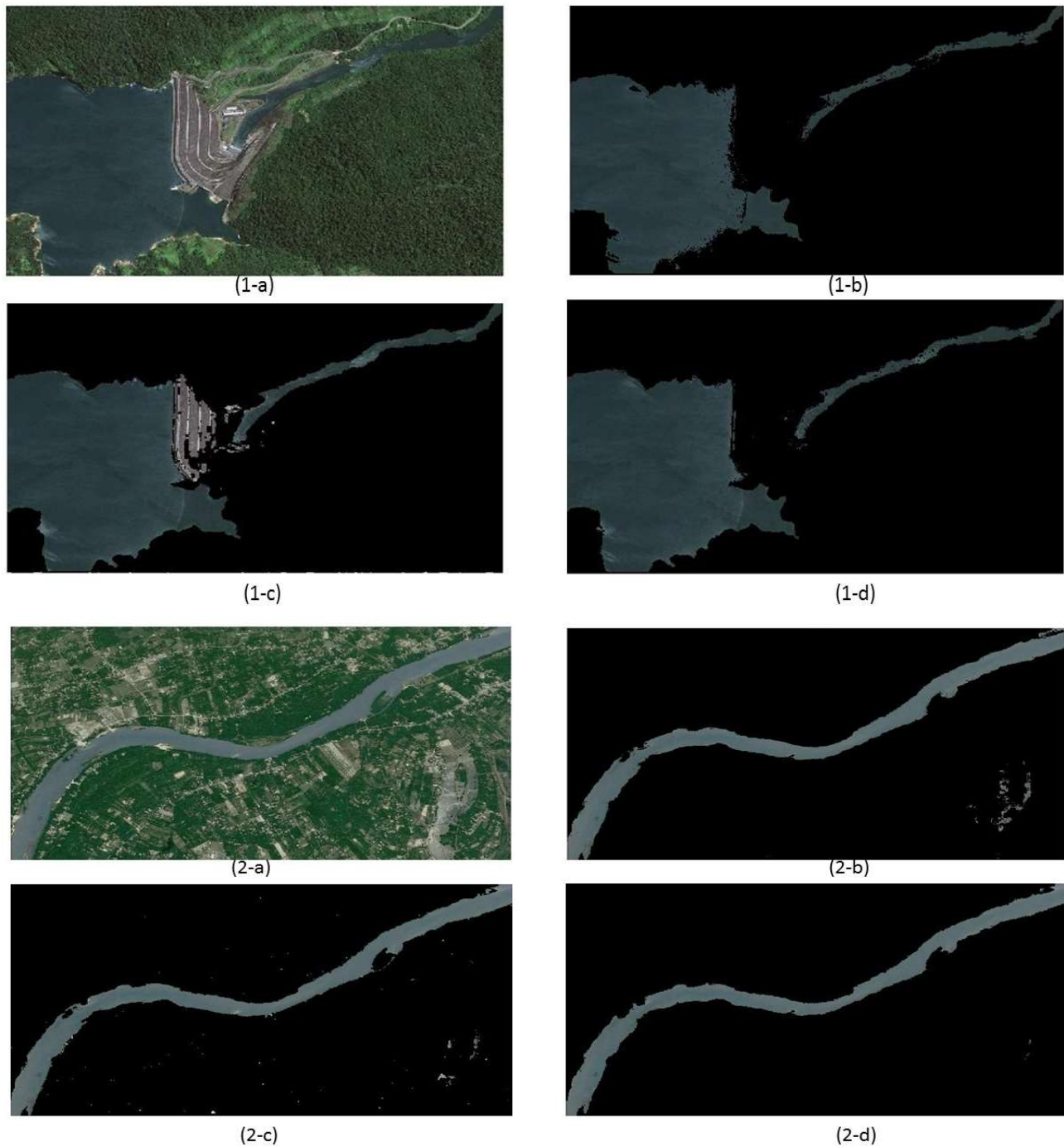Malaysian Journal of Computer Science. Vol. 31(2), 2018

Fig. 2: (1-a) and (2-a) Cropped original image of Terengganu River (2013). (1-b) and (2-b) K-Means clustering result. (1-c) and (2-c) Water extraction using k-Means PCA. (1-d) and (2-d) Resulting image using the proposed algorithm.

Although this modification could indicate more distinctly the separation of water in urban area, it might remove part of the water since the modified filter classification has been applied to some images, as shown in second part of the Fig. 2. In this context, there are guidelines pertaining to the use of various filters and separation of the water to avoid producing any false positive. The second part of the Fig. 2 shows the false positive classification produced using the k-Means PCA and k-Means techniques. The k-Means technique produces more false positives when compared with other techniques,

148

as shown in Fig. 2. (2-b). However, Fig. 2 (2-c) shows more pepper-and-salt false positives in the image including less amount of false positives on the right side of the image. This indicates that k-Means PCA performs better in keeping the water boundary when compared with k-Means clustering. The proposed method has removed almost all the false positives, but removed some parts of the water boundary, only.

Fig. 3 shows an urban area which is full of similar spectral color with water, and this poses a big challenge for extraction. Fig. 3 (d) shows the image following modification using the proposed algorithm almost no false negative within the extraction of the water boundary. This indicates that k-Means PCA has improved water classification but still produces some false positives in the final result, as shown in Fig. 3(c). However, the river seems wider in some parts when using the k-Means clustering segmentation technique directly with k-Means PCA and the proposed algorithm. The marked improvement achieved when using the proposed technique is due to its ability to totally eliminate the 'pepper-and-salt' noise seen in the second part of Fig. 2 and Fig. 3. Hence, areas of water feature have become homogeneous.
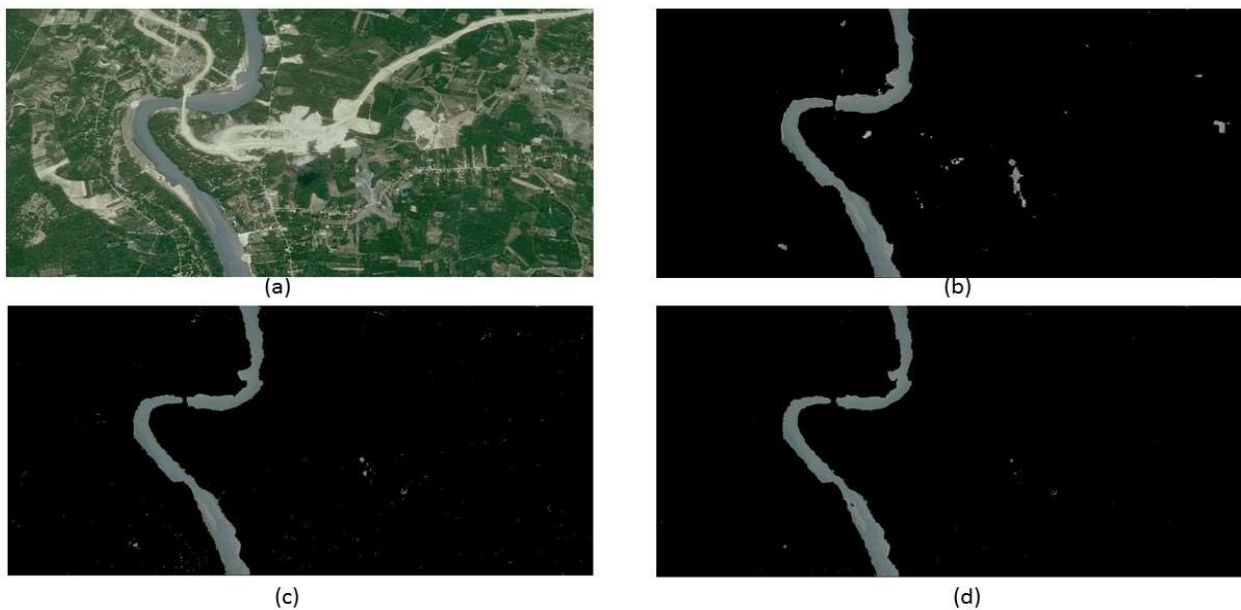

Fig. 3: (a) Cropped original image of an urban area. (b) K-Means clustering result. (c) Water extraction using k-Means PCA. (d) Resulting image produced the proposed algorithm.

Fig. 4(a) and 4(b) show the images before and after the application of the proposed approach, respectively. Fig. 4(a) shows the original image of a part of the North Sea which includes the vegetation area of north-eastern Germany. This image poses complex challenges for water-body extraction because of the different color intensities, and the similar dried-up water-body, as seen on the right side of the image. The image shown in Fig. 4(b) clearly indicates that the algorithm has solved the first problem by controlling the number of clusters. The second problem has also been resolved because of the characteristics of the proposed algorithm.

149

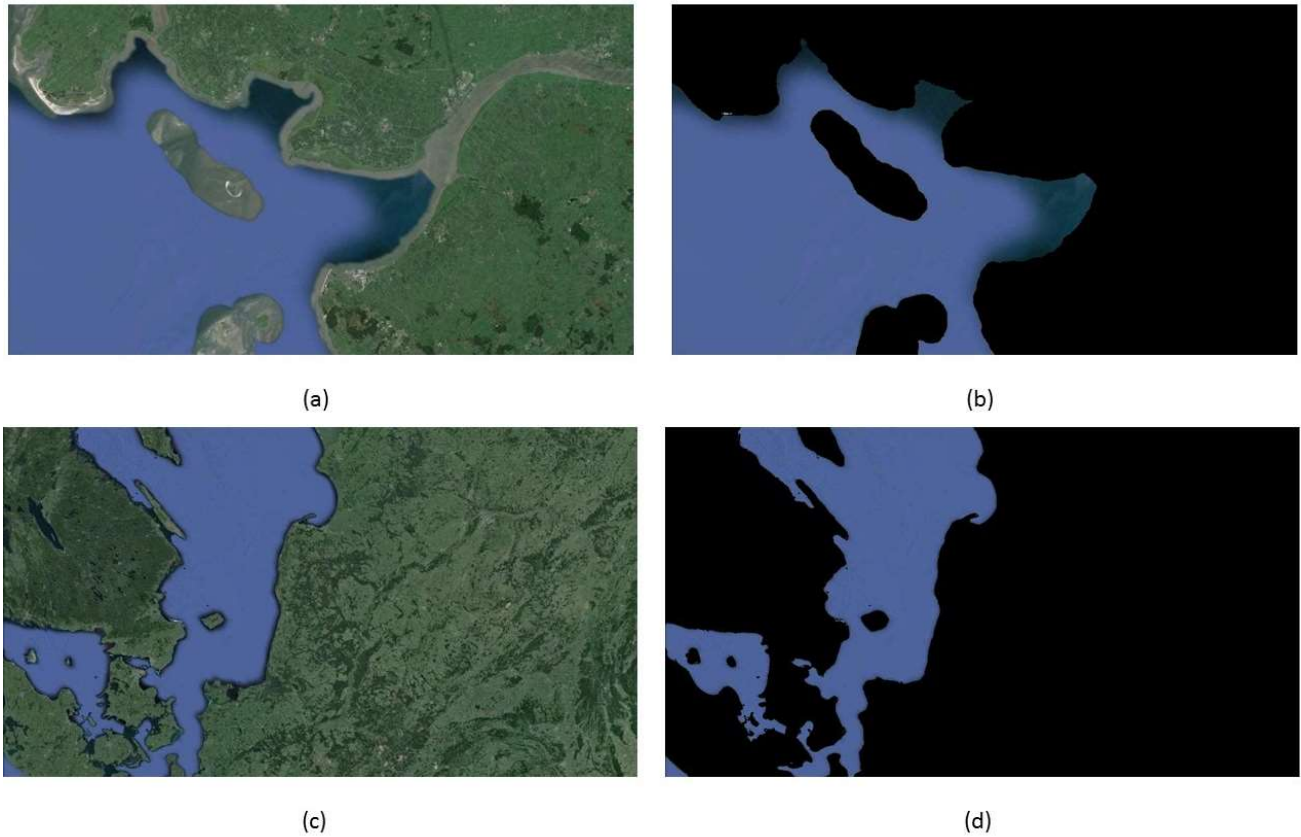Malaysian Journal of Computer Science. Vol. 31(2), 2018

Fig. 4: (a) Cropped original image of a part of the North Sea. (b) Resulting image produced by the proposed algorithm. (c) Cropped original image of north-eastern Germany. (d) Resulting image produced by the proposed algorithm

The results obtained by the proposed approach are shown in Fig. 4(c and d) and Fig. 5. The results clearly show that the algorithm is able to recognize the water-body areas from satellite images of different area, or landscapes.
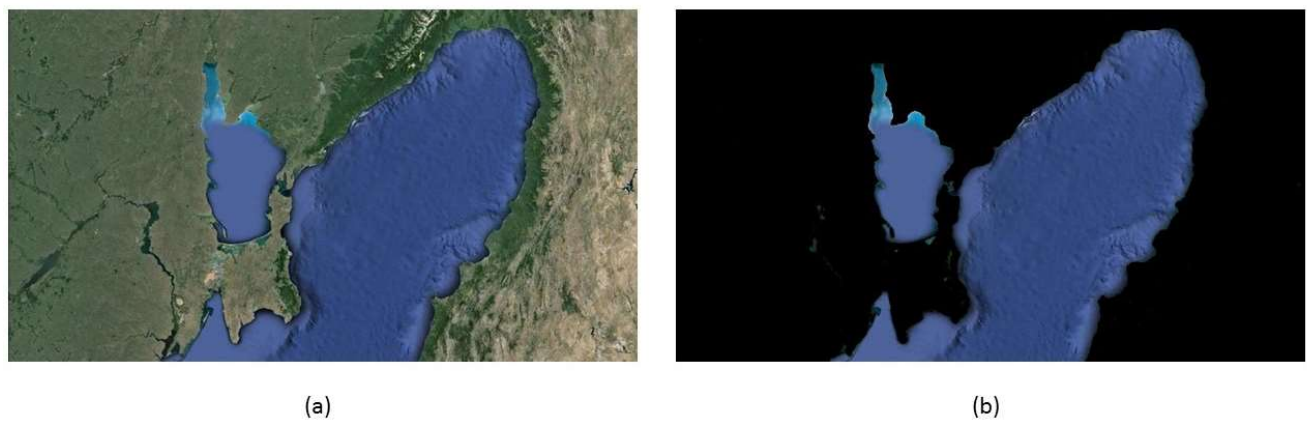

Fig. 5: (a) Cropped original image of the Sea of Azov. (b) Resulting image produced by the proposed algorithm.

150

Besides qualitative analysis, quantitative measurement was also performed to support the qualitative observation of the resulting images. In the conventional method, the peak signal-to-noise ratio (PSNR) is measured as it operates directly on the intensity of the image, based on the following formula:

$$PSNR = 20\,log_{10}\left(MAX_f\middle/\sqrt{MSE}\right) \qquad (5)$$

where MAX_f is the maximum signal value in our original image.

Table1 shows the results of the quantitative analysis of the PSNR per image using different segmentation methods. It shows that the PSNR in the proposed method produces higher signal values when compared with the other two methods. The results obtained using the proposed method show more similarity to the original image, which also has higher PSNR values.

Table 1: Quantitative comparison of the PSNR values produced by the different methods

Case	Proposed algorithm	Case	k-Means clustering	Case	k-Means PCA
PSNR for Fig 2(1-d)	24.4079	PSNR for Fig 2(1-b)	23.6865	PSNR for Fig 2(1-c)	21.7040
PSNR for Fig 2(2-d)	28.8587	PSNR for Fig 2(2-b)	27.4149	PSNR for Fig 2(2-c)	28.7899
PSNR for Fig 3(d)	27.6645	PSNR for Fig 3(b)	27.5518	PSNR for Fig 3(c)	27.6361

Much efforts had been made on quantitative analyses of the human visual system (HVS) and its sensitivity to different visual signals such as luminance, contrast, frequency, and the interaction among the different signal components [27, 28]. Based on the results obtained, the second method of analysis was chosen. The feature Similarity (FSIMc) [29] index for Image Quality Assessment (IQA) is based on the HVS, and is calculated using the formula, below:

$$FSIMc = \frac{\sum_{x\in\delta} S_L(x).\,[S_C(x)]^\lambda.\,PC_m(x)}{\sum_{x\in\delta} PC_m(x)} \qquad (6)$$

where > 0 , and is used to rank the importance of the chromatic components, and S_C is the chrominance similarity measure. The similarity of the two pictures based on gradient magnitude (GM) and phase congruency (PC) is shown by S_L. Also, PC is the phase congruency, thus, $PC_m(x) = \max(PC_1(x).\,PC_2(x))$; δ means the whole image spatial domain.

Table 2 shows the results of the quantitative analysis after applying FSIMc on the other two methods. It shows that the FSIMc values by the proposed method have achieved higher results comparing with the k-Means clustering, and k-Means PCA methods. Tables 1-2 indicate that the proposed algorithm is in agreement with the results from the qualitative analysis.

151

Malaysian Journal of Computer Science. Vol. 31(2), 2018

Table 2: Comparison of the FSIMc values produced by the different methods

Case	Proposed algorithm	Case	k-Means clustering	Case	
FSIMc for Fig 2(1-d)	0.9207	FSIMc for Fig 2(1-b)	0.9134	FSIMc for Fig 2(1-c)	0.9006
FSIMc for Fig 2(2-d)	0.9744	FSIMc for Fig 2(2-b)	0.9570	FSIMc for Fig 2(2-c)	0.9570
FSIMc for Fig 3(d)	0.9325	FSIMc for Fig 3(b)	0.8109	FSIMc for Fig 3(c)	0.8511

4.0   CONCLUSION

The proposed enhanced technique has been developed to achieve better accuracy in extraction of the water-body segmentation from satellite images. The statistical PCA technique was used for unsupervised dimension reduction, while k-Means clustering was used for performing unsupervised learning tasks. In this study, the PCA was applied to guide the modified filter to achieve better accuracy in extraction of the water-body segmentation in satellite images.
The study has achieved its objectives: (i) produced the eigenvectors pertaining to the eigenvalues using the PCA technique; (ii) rearranged the selected eigenvectors by using a training map; after modifying the filter, K-means clustering was used to classify the satellite images, and the river was extracted. Based on the results of the experiments, the proposed method has produced better image segmentation for the water-body areas in satellite images as compared with the other techniques such as k-Means clustering and PCA, and this is supported by the higher PSNR values and FSIMc values obtained from quantitative measurements. Further research in this area could focus on the improvement of segmentation technique of extraction of water-body from satellite images using other classification techniques based on the PCA approach.

5.0   ACKNOWLEDGMENT

REFERENCES

[1]   B. Schölkopf, A. Smola, and K.-R. Müller. *Kernel principal component analysis*. in *International Conference on Artificial Neural Networks*, Springer, 1997.

[2]   I. Jolliffe, *Principal component analysis,* Wiley Online Library, 2001.

[3]   C. Munyati, "Use of principal component analysis (PCA) of remote sensing images in wetland change detection on the Kafue Flats, Zambia", *Geocarto International*, Vol. 19, No. 3, 2004, pp. 11-22.

[4]   M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 12, No. 1, 1990, pp. 103-108.

[5]   M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of cognitive neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.

[6]   Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives", *Pattern*

152

Malaysian Journal of Computer Science. Vol. 31(2), 2018

Analysis and Machine Intelligence, IEEE Transactions on, Vol. 35, No. 8, 2013, pp. 1798-1828.

[7]   J. MacQueen, "Some methods for classification and analysis of multivariate observations", in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *Oakland, USA* , Vol. 1, 1967, pp. 281-297.

[8]   K. Honda, A. Notsu, and H. Ichihashi, "Fuzzy PCA-Guided Robust-Means Clustering", *IEEE Transactions on Fuzzy Systems*. Vol. 18, No. 1, 2010, pp. 67-79.

[9]   C. H. Ding and X. He, "Principal Component Analysis and Effective K-Means Clustering", in *Proceedings of the 2004 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics, 2004, pp. 497-501.

[10]   S. K. McFeeters, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features.", *International journal of remote sensing*. Vol. 17, No. 7, 1996, pp. 1425-1432.

[11]   D. H. Boland, *Trophic classification of lakes using Landsat-1 (ERTS-1) multispectral scanner data*, US Environmental Protection Agency, Office of Research and Development, Corvallis Environmental Research Laboratory, Assessment and Criteria Development, 1976.

[12]   H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery", *International journal of remote sensing*. Vol. 27, 2006, pp. 3025-3033.

[13]   R. Sivanpillai and S. N. Miller, "Improvements in mapping water bodies using ASTER data", *Ecological Informatics*, Vol. 5, No. 1, 2010, pp. 73-78.

[14]   F. Sun, W. Sun, J. Chen, and P. Gong, "Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery", *International journal of remote sensing*, Vol. 33, No. 21, 2012, pp. 6854-6875.

[15]   T. M. Phuong, Z. Lin, and R. B. Altman, "Choosing SNPs using feature selection", *Journal of bioinformatics and computational biology*, Vol. 4, No. 2, 2006, pp. 241-257.

[16]   G. Roffo, S. Melzi, and M. Cristani, "Infinite Feature Selection*"*, in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4202-4210.

[17]   S. C. Shah and A. Kusiak, "Data mining and genetic algorithm based gene/SNP selection", *Artificial intelligence in medicine*, Vol. 31, No. 3, 2004, pp. 183-196.

[18]   P. Xuan, M. Guo, J. Wang, C. Wang, X. Liu, and Y. Liu, "Genetic algorithm-based efficient feature selection for classification of pre-miRNAs", *Genetetics and Molecular Res*earch, Vol. 10, 2011, pp. 588-603.

[19]   N. Long, D. Gianola, G. Rosa, and K. Weigel, "Dimension reduction and variable selection for genomic selection: application to predicting milk yield in Holsteins", *Journal of Animal Breeding and Genetics*, Vol. 128, No. 4, 2011, pp. 247-257.

[20]   E. B. Huerta, B. Duval, and J.-K. Hao, *A hybrid GA/SVM approach for gene selection and classification of microarray data*, in *Applications of Evolutionary Computing*, Springer, 2006, pp. 34-44.

[21]   E. De la Hoz, E. De La Hoz, A. Ortiz, J. Ortega, and B. Prieto, "PCA filtering and probabilistic SOM for network intrusion detection", *Neurocomputing*, Vol. 162, 2015, pp. 71-81.

[22]   D. Giri, U. R. Acharya, R. J. Martis, S. V. Sree, T.-C. Lim, T. Ahamed, and J. S. Suri, "Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform", *Knowledge-Based Systems*. Vol. 37, 2013, pp. 274-282.

153

Malaysian Journal of Computer Science. Vol. 31(2), 2018

[23] L. Rokach and O. Maimon, *Clustering methods*, in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 321-352.

[24] S. P. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, Vol. 28, No. 2, 1982, pp. 129-137.

[25] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity", *IEEE Transactions on Image Processing,* Vol. 21, No. 5, 2012, pp. 2481-2499.

[26] H. P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1, No. 3, 2011, pp. 231-240.

[27] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model", *IEEE Transactions on Image Processing*, Vol. 9, No. 4, 2000, pp. 636-650.

[28] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images", *IEEE Transactions on Image Processing*, Vol. 16, No. 9, 2007, pp. 2284-2298.

[29] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment", *IEEE Transactions on Image Processing*, Vol. 20, No. 8, 2011, pp. 2378-2386.

154

Malaysian Journal of Computer Science. Vol. 31(2), 2018