Study of the Color Space Contribution in the Satellite Image Clustering

* BEKKOUCHE Ibtissem[1] and FIZAZI Hadria[2]
[1] *Department of Computer Sciences, University of Science and Technology of Oran, Algeria.*
[2] *Department of Computer Sciences, University of Science and Technology of Oran, Algeria.*
[1] *ibtissem.bekkouche@univ-usto.dz*, [2] *hadria.fizazi@univ-usto.dz*

Abstract

Satellite images are accompanied by data and noise. The classification is one of the most important processes in satellite image processing because it enables interpretation, analysis and extraction of meaningful information efficiently.
In clustering of satellite images, the colored compositions make it possible to highlight the different types of surfaces on the multispectral images in the RGB colorimetric space.
Experimental results applied to different satellite images reveal that color space influences the efficiency and accuracy of clustering methods, whether hard or soft, to extract the necessary information.

Keywords: *Clustering, satellite image, color space.*

1. Introduction

In remote sensing, the classification consists in carrying out the correspondence between the elements of an image scene generally materialized by their radiometric values and clusters known or not by a user [1]. It is recognized as an important process of exploration and extraction data.
Satellite image clustering is generally an unsupervised learning technique consisting of grouping objects into a finite set of categories or classes or clusters, according to one or more similarity criteria using a classifier. The objects to be classified in an image can be punctual (pixels) in the case of medium and high spatial resolution images or structural (segments) in the case of very high spatial resolution images. The similarity criteria (color, texture, shape) depend on the nature of the objects to be classified and the intended application.
There are two types of clustering, soft overlay clustering (fuzzy) and an exclusive clustering (hard). In the hard classification, the goal would be to partition the dataset into partitions according to the number of clusters without overlap, whereas in fuzzy classification the goal is to repartition the dataset into partitions which allows the data to belong to a particular degree for each fuzzy cluster [2]. Fuzzy clustering is more natural and concrete than hard clustering. As methods that include soft clustering we have Fuzzy k- means, Fuzzy c-means, rough c- means, and similarly in hard clustering we have K-means, K-medoids et C-means [3, 4].
In digital image processing, color is primarily used to distinguish the different objects present in the images and thus facilitates the interpretation of the images. In satellite images one can work with colored compositions, either true or false colors, by combining three spectral bands but this technique is done in the RGB space. But if we change color

space, we can manipulate a very large number of colors. The colorimetry is a vast and complex world because it allows color information to be represented as three or four different color components and there are two types of color spaces: device dependent (Relative Colorimetric) like RGB, CMY, CMYK, YIQ, YUV and YCbCr or device independent (Absolute Colorimetric) like CIE XYZ, CIE $L*U*V*$, CIE $L*a*b*$ [5, 6, 7]. This wide color space is particularly useful for classifying satellite images.
The aim of this study is to show the influence of color space in the satellite images clustering. So, in this article, we are going to start with describes the concept of the color space as well as giving a briefly description of the two clustering algorithms uses K-means and Fuzzy c-means. Then, explain the methodology followed on this research as well as the study areas chosen. And finally, the comparative study between the different color spaces chosen due to the experiments made. The results show that the colorimetric space influences the clustering methods, whether hard or fuzzy, in the efficiency and accuracy as well in the exploration and extraction of relevant information in satellite images.

2. Colorimetric Space

Colorimetric space, also called color space or chromatic space, is a three -dimensional mathematical model to represent color information in the form of three or four different color components. Each color it contains is thus associated with coordinates determining a precise point and corresponding, for example, to values such as luminance, saturation and hue.
Color space explains how colors are represented and specifies the components of color space precisely to know what each color spectrum looks like. There are two types of color spaces: device dependent and device independent (see figure 1) [5, 6, 7].

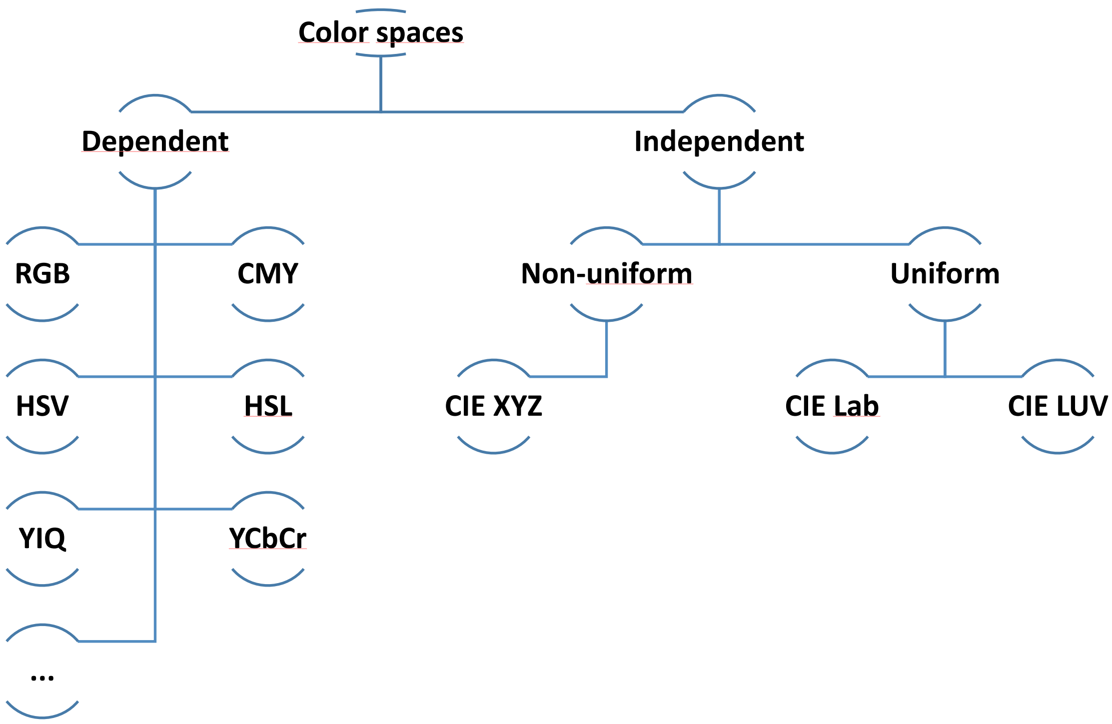


Figure 1. Dependent and Independent Color Space

2.1. Dependent Color Space

Dependent color spaces, called relative colorimetry's, only describe the characteristics of the corresponding device.

They put in correspondence exactly matched colors at a time to the color gamut of the input device and that of the output device (for example, between a monitor and a printer) to ensure that the colors produced depends only on the equipment used for display and all parameters.

Works best for monochrome images or images with a limited number of colors because colors outside the two gamuts can be mapped to a single color [5, 6, 7].

Example of device-dependent color scheme [8]:

- **RGB (Red Green Blue):** Additive color system based on trichromatic theory, used by CRT (Cathode Ray Tube) displays where proportions of excitation of red, green and blue emitting phosphors produce colors when visually fused. Easy to implement, nonlinear, device dependent, unintuitive, common and used in television cameras, computer graphic...
- **CMY(K) (Cyan Magenta Yellow (Black))**: Subtractive color. Used in printing and photography. Printers often include the fourth component, black ink, to improve the colour gamut (by increasing the density range), improving blacks, saving money and speeding drying (less ink to dry). Fairly easy to implement, difficult to transfer properly from RGB, device dependent, non-linear, unintuitive.
- **HSL (Hue Saturation and Lightness)**: designed in the 1970s to know the way human vision perceives color. This represents a wealth of similar color spaces, alternatives include HSI (Intensity), HSV (Value), HCI (Chroma / Colorfulness), HVC, TSD (Hue Saturation And Darkness) …. Most of these color spaces are linear transforms from RGB and are thus, device dependent, non-linear but very intuitive. In addition, the separation of the luminance component has advantages in image processing and other applications.
- **YIQ, YUV, YCbCr, YCC (Luminance - Chrominance)**: These are the television transmission colour spaces, also known as transmission primaries. YIQ (luminance (Y) and chrominance (I and Q)) and YUV are analogue for NTSC (National Television System Committee) and PAL (Phase Alternating Line), and YCbCr is digital (luminance (Y) and chrominance (Cb and Cr)). They separate luminance from chrominance (lightness from color) and are useful in compression and image processing applications. They are device dependent and unintuitive. Kodaks PhotoCD system uses a type of YCC color space, PhotoYCC, which is a device calibrated color space.

2.2. Independent Color Space

Independent color spaces, also called absolute colorimetry, describing a set of visible colors without referring to a particular device in the graphics chain.

They perfectly reproduce the available gamut colors and approximate the reproducible hue nearest the colors held out of gamut (at the expense of saturation). In the device-independent color space, a set of parameters will produce the same color regardless of the equipment used. This means that the color model is not affected by system or device properties. Thus, in this space, the coordinate used for the specified color allows the same color everywhere it applies.

As the CIE colorimetric space (see figure 2 [9]), determined in 1931 by the International Commission on Illumination, is still today the international standard for colorimetry. This independent color space represents all the colors perceptible to the human eye (about 8

million different shades) and therefore serves as a reference thanks to its different graphic variations [5].
Example of a device independent color model:

CIE XYZ is at the root of all colorimetry, make in 1931. It is defined such that all visible colors can be defined using only positive values, and the Y value is the luminance. Therefore, the colors of the XYZ primaries themselves are not visible. The chromaticity diagram is highly nonlinear, in that a unit-magnitude vector representing the difference between two chromaticities is not uniformly visible.
Nonlinear and uniform descriptions of the same space are made in 1976, CIELAB, for the characterization of surfaces, and CIELUV, for the characterization of the light sources of screens [8].
Ces modèles sont bien adaptés aux applications où les informations sont transmises via différents périphériques matériels.

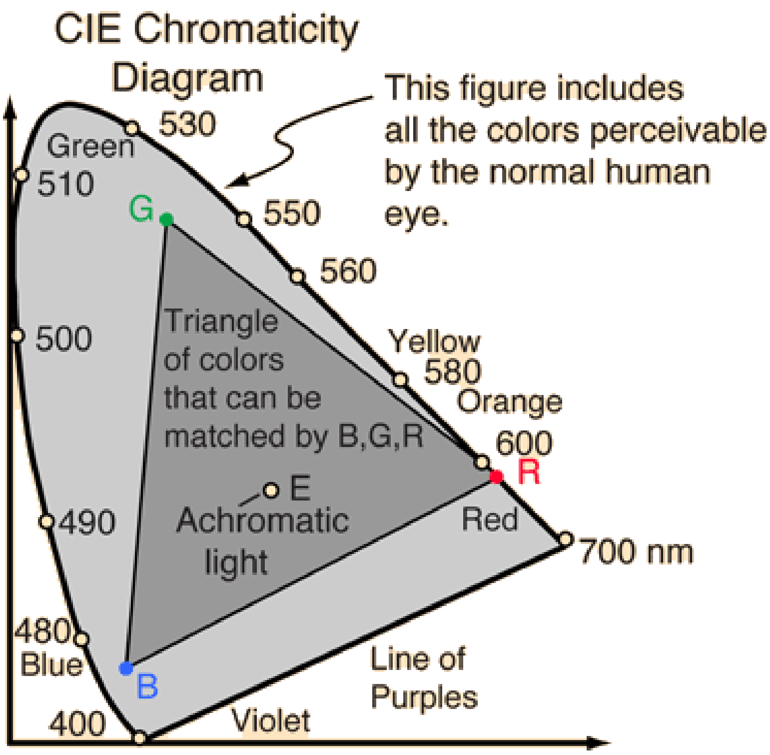


Figure 2. C.I.E. Color Space

3. Clustering

Image classification is generally an unsupervised learning technique for grouping similar data points (pixels) according to some measure of similarity that maximizes intra-class similarity and minimizes inter-class similarity [10].
Two types of classification algorithms are available: Soft Clustering (superimposed clustering) and Hard Clustering (exclusive clustering). In hard classification, the goal would be to partition the dataset X into partitions C1, ..., Cc without overlap and without vacuum, while in soft classification the goal is to divide the data set X into partitions which allows the data to belong to a particular degree between 0 and 1 for each fuzzy class [2].
Fuzzy C-Means (FCM) is a very popular soft clustering technique based on logic fuzziness, and likewise K-means is an important hard clustering technique.

3.1. K-Means

Among the most popular unsupervised classification algorithms in satellite image processing, we recommend the k-means algorithm, known as k-means proposed by MacQueen in 1967 [11].

K-means, called the moving center algorithm, is one of the simplest unsupervised learning algorithms that solves the classification problem due to its simplicity of implementation. It is an algorithm for classifying or grouping objects based on attributes/characteristics in K number of classes, in which the objects inside each class are as close as possible to each other and as far possible away from objects of other classes. Each class in the partition is defined by its objects and its centroid [12].
It is based on the method of centroids (or centers of gravity), we start by giving *k* arbitrary centers $c_1, c_2, \ldots, c_k$ where each ci represents the center of a class c_i. Each class c_i is represented by a set of individuals closer to c_i than to any other center.
After this initialization, a second partition is made by grouping the individuals around the m_j which then take the place of the c_j (m_j is the centroid of the class c_j, calculated from the new classes obtained) [13]. This algorithm aims at minimizing the following objective function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i - c_j \right\|^2 \qquad (1)$$

Where $\mu_{ij}^{m} \left\| x_i - c_j \right\|^2$ is a chosen distance measure between a data point x_i and the cluster centre c_j, it is an indicator of the distance of the *n* data points from their respective cluster centers.
The process is thus repeated until a state of stability is reached where no improvement is possible.

3.2. Fuzzy C-Means

Fuzzy C-Means (FCM), is an unsupervised fuzzy classification algorithm. Issued from the hard C-means clustering algorithm, developed by Dunn in 1973 [14] and improved by Bezdek in 1981 [15], it introduced the notion of fuzzy set in the definition of classes: each point in the data set belongs to each cluster with a certain degree, and all clusters are characterized by their center of gravity [16].
The goal of Fuzzy C-Means clustering is to find the minimum of the following function:

$$J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2, \quad 1 \leq m \leq \infty \qquad (2)$$

$$c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^{m} x_i}{\sum_{i=1}^{D} \mu_{ij}^{m}} \text{ And } \mu_{ij} = \sum_{k=1}^{c} \left\{ \frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right\}^{-2/m-1} \qquad (3)$$

Where *D* is the number of data points, *N* is the number of clusters, *m* is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap. Fuzzy overlap refers to how fuzzy the boundaries between clusters are, that is the number of data points that have significant membership in more than one cluster, x_i is the *i*th data point, c_j is the center of the *j*th cluster and μ_{ij} is the degree of membership of x_i in the *j*th cluster. For a given data point, x_i, the sum of the membership values for all clusters is one [17, 18].
FCM performs the following steps during clustering:

1. Randomly initialize the cluster membership values, μ_{ij}.

2. Calculate the cluster centers c_j
3. Update μ_{ij}
4. Calculate the objective function, J_m.
5. Repeat steps 2 to 4 until J_m improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

4. Images Used

Figure 3 shows the different areas of the satellite images of northwestern Algeria used in this study [19].

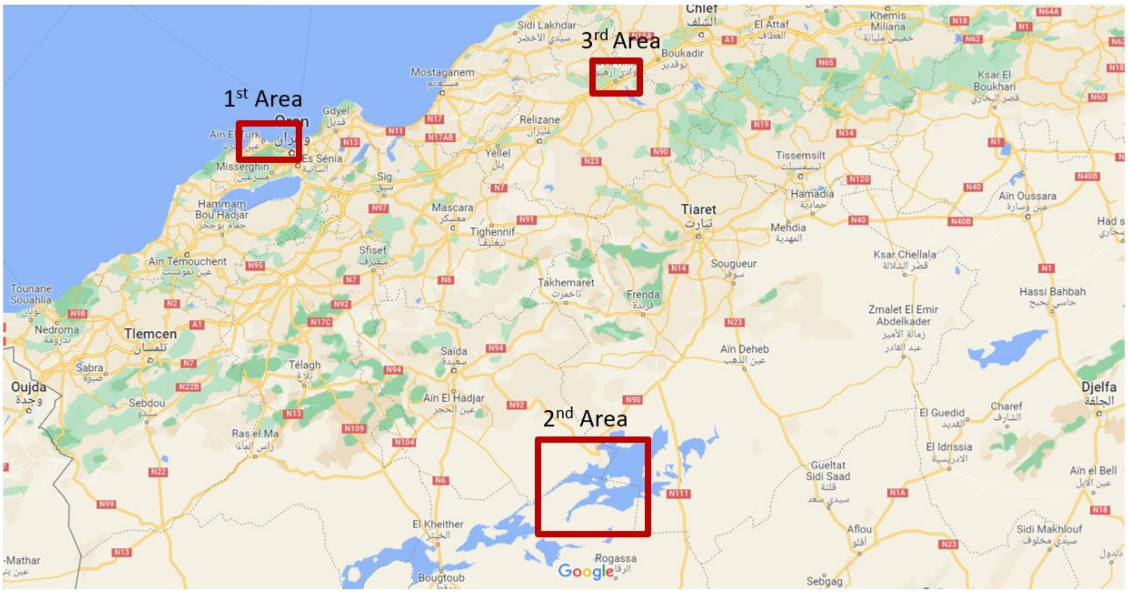
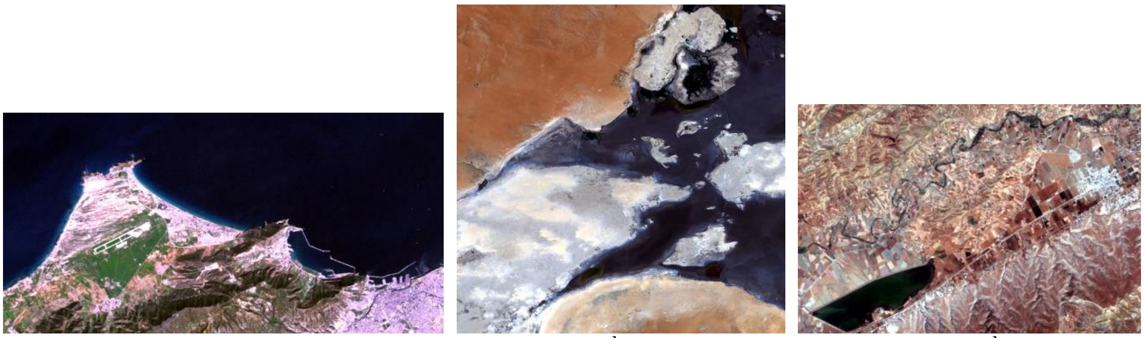


Figure 3. Different Areas of the Satellite Images of North West Algeria

The three RGB satellite images of the areas are shown in Figure 4. The first satellite image of Ain El Turck, Oran which was taken in 2003 by Landsat 5, the second satellite image of Rogassa, El Bayadh was taken in 2002 by Landsat 7 and the latest satellite image of Oued Rhiou, Relizane was taken in 2014 by Landsat 7. These areas were chosen for their varied landscapes which could be of interest to our study.



1st area 2nd area 3rd area

Figure 4. RGB Satellite Images of the Areas

5. Methodology

The following figure illustrates and summarizes the organization of the processing steps adopted in order to study the influence of color spaces on the clustering of satellite images, whether hard or soft clustering, and this according to the principle of satellite image classification.

The implementation scheme is made up of several stapes (see figure 5):

1st step: This is a data acquisition and preparation step, for this we must:

- We begin by collecting and studying satellite images.
- Next, we pass to pre-processing, for example the image cutting, if necessary.
- Afterwards, we proceed to the assembly of satellite data into various types of colorimetric space. For this on a chosen in the dependent colorimetric space: RGB (natural and false color composition) which is the basis of the spaces, HSV, YIQ and YCbCr. Likewise in independent space on a chosen one: CIE XYZ (non-uniform) and CIE Lab (uniform)

2nd step: It is a satellite image processing step by clustering and for this we will use for the first group of tests 'kmeans' which is a hard clustering technique. Afterwards for the second group of tests Fuzzy C-Means (FCM) that is a soft clustering technique.

3rd step: It is a generalization of final product obtained, which is a thematic map (classified image) and which will be evaluated against the ground reality.

4th step: Once the image is classified, the last step is to analyze, interpret and compare the different results. There are various evaluation measures that can be used and chosen wisely, since the choice of measure can influence the way performance is designated and interpreted. To assess the influence of colorimetric spaces in this study, we used visual analysis according to the ground truth available in the area. The use of the human visual system as a quality judgment tool is not to be neglected but necessary to check the quality of the images obtained either by a hard classifier or even for a soft classifier [20].

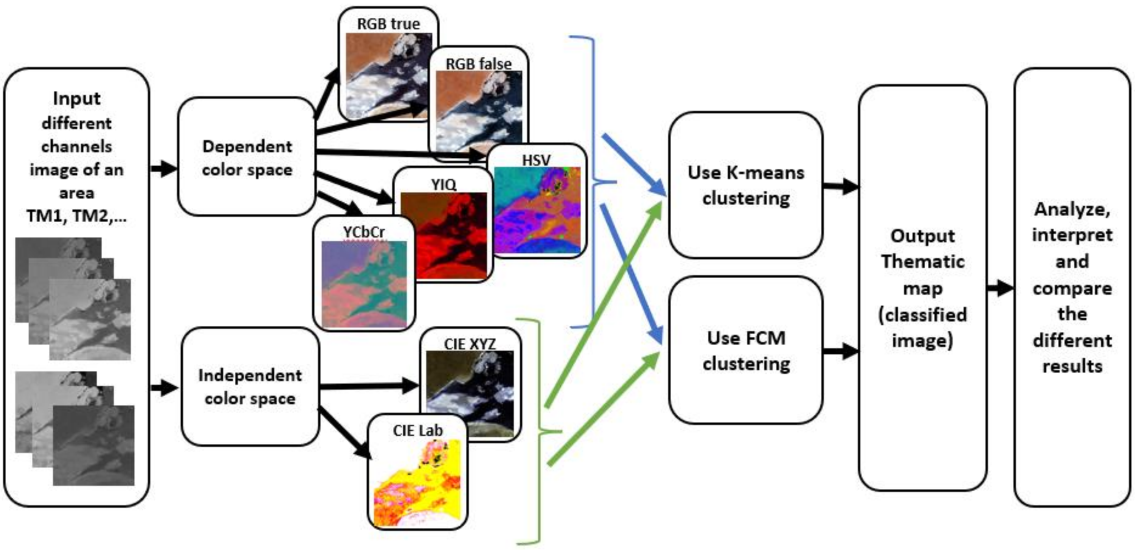


Figure 5. General Scheme of the Proposed System

6. Experiment Results and Discussion

For our comparison study between the selected color spaces (mentioned in section 4), we start by applying them to the different satellite images (presented in section 3), then the images obtained are processed by the K-mean and FCM clustering methods. Even if the

FCM method takes more time for the calculation than that of K-Means but our goal is to show the influence of the color spaces in the hard and soft clustering.
Table 1 shows the best clustering result for the different color spaces in terms of the objective function value for the two clustering methods. For the clustering of the three satellite images, the number of clusters for the satellite image of area 1 we set to eight (8), for the second image we set to five (5) and the last image we set to seven (7). The number of clusters was fixed after studying the different classes that these areas can contain.

Table 1. Comparison between different color spaces using K-Means and FCM clustering

	K-Mean			FCM		
	Image 1st area	Image 2nd area	Image 3rd area	Image 1st area	Image 2nd area	Image 3rd area
RGB True	6.441667 e+07	2.950472 e+08	4.557535 e+07	7.796843 e+06	6.695892 e+07	6.716426 e+06
RGB false	8.155762 e+07	3.123040 e+08	5.651386 e+07	9.568925 e+06	7.019525 e+07	6.537607 e+06
HSV	1.062498 e+08	5.003240 e+08	7.314579 e+07	1.061698 e+07	7.953334 e+07	9.882752 e+06
YIQ	6.111656 e+06	5.993873 e+07	7.608133 e+06	1.071177 e+06	1.385385 e+07	1.235229 e+06
YCbCr	1.703777 e+07	7.871735 e+07	1.534905 e+07	1.671859 e+06	1.514333 e+07	1.586842 e+06
CIE XYZ	1.342056 e+07	9.876126 e+07	1.523655 e+07	1.353088 e+06	1.568143 e+07	1.572651 e+06
CIE Lab	5.472461 e+06	7.618088 e+07	9.086443 e+06	5.696693 e+05	1.140594 e+07	9.080578 e+05

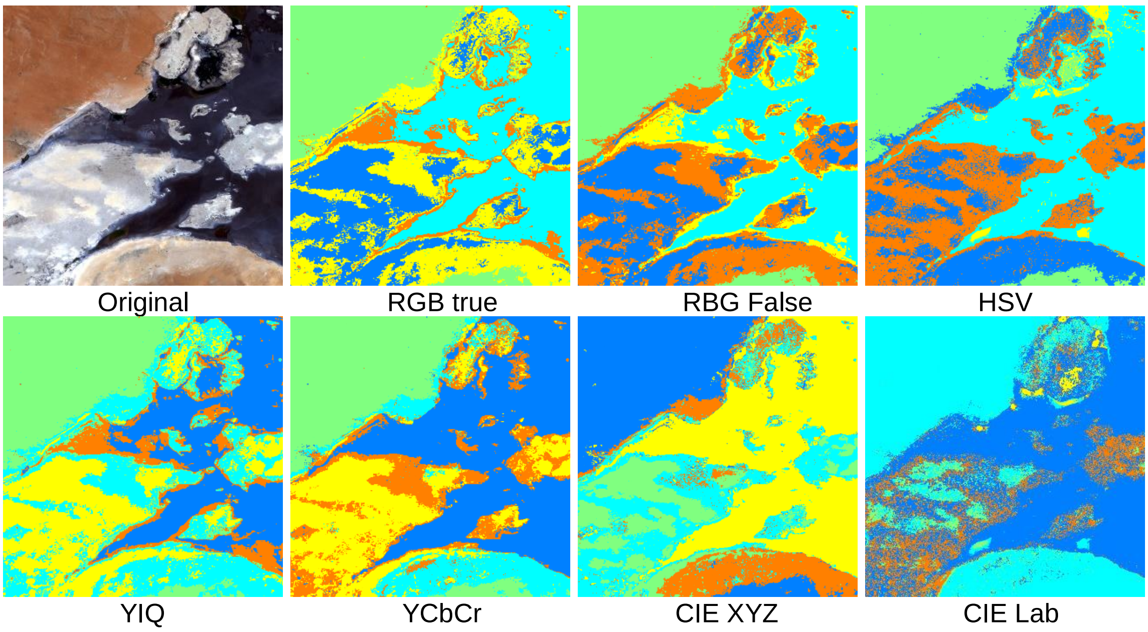


Figure 6. Clustering by K-Means

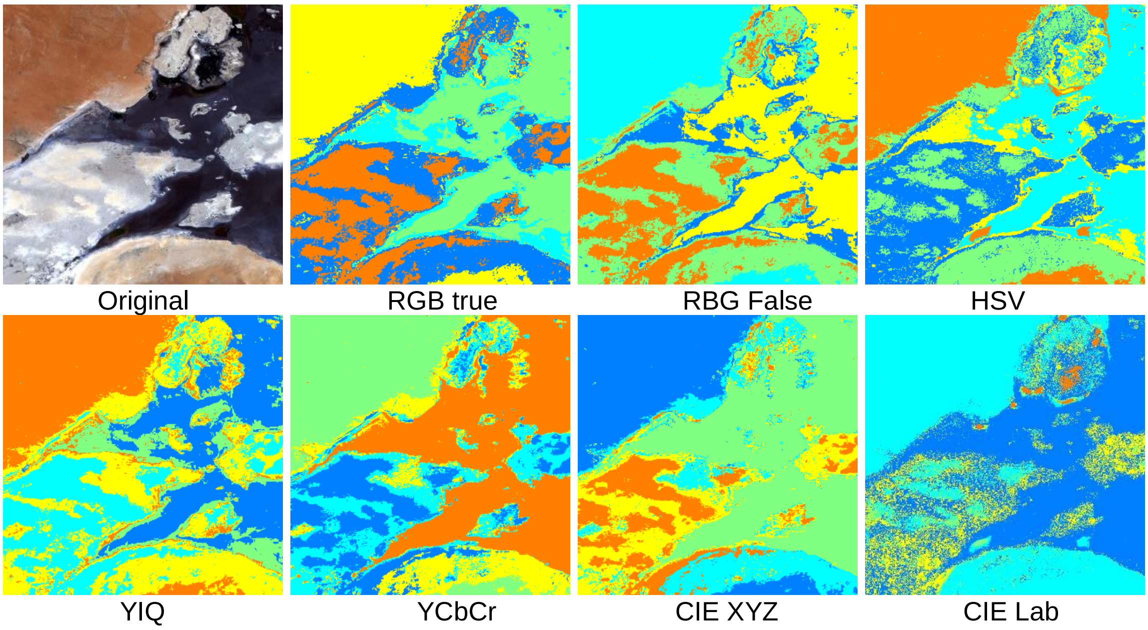


Figure 7. Clustering by FCM

If we follow the results of the optimal values presented in table 1, we see that the CIE Lab colorimetric space gave the best result when we use by the FCM method for the three images and also by the K-Means method for the first image and for the other two image k-means gave a better result with YIQ space.
The results presented in Table 1 are insufficient to conclude which color spaces are useful for the clustering of satellite images. For this, visual interpretation was also used. In figures 6 and 7, we present the results of the clustering in images just for the image of zone 2.
From table 1 and figures 6 and 7, we can conclude that the best clustering result for image 3 is given by the use of the YCbCr space according to the comparison between the different results by the visual entanglement and the calculations made by the two clustering's methods K-Means and FCM. And this is also implied for the other images (image zone1 and zone2), the use of the YCbCr space gave the best result.
So, our study and the experimental results show the usefulness of choosing the color space in the clustering of satellite images.

7. Conclusion

Clustering is an important process in the processing of satellite images because it allows to collect a huge amount of information. Usually, the color space used for multispectral images is the RGB color space. Even if the RGB space is a basic space of the majority of the spaces, but it is nonlinear with visual perception therefore it is not effective for the recognition of the majority of the colors knowing that a satellite image is a mixture of chrominance and luminance components. So, it is necessary to transform the satellite image in the RGB colorimetric space into other colorimetric spaces.
With our study, we have shown the usefulness of the colorimetric space in the clustering of satellite images. Experimentation and analysis show that the performance of both K-Means and FCM clustering methods gives better results with YCbCr color space compared to other RGB, HSV, YIQ, CIE XYZ and CIE Lab color spaces. The YCbCr color space is widely used to process the digital image because the luminance and chrominance information is stored separately.

References

[1] *J. Dreo, A. Pétrowski, P. Siarry and E. Taillard , "Métheuristiques pour l'optimisation difficile ", Editions Eyrolles, **(2003)**.*

[2] *O. M. Alia, R. Mandava, and M. E. Aziz, "A hybrid harmony search algorithm for MRI brain segmentation", Evolutinary Intelligence, Springer-Verlag. vol. 4, no. 1, **(2011)**, pp. 31-49.*

[3] *G. Kumar, P. P. Sarthi, P. Ranjan and R. Rajesh, "Performance of k-means based satellite image clustering in RGB and HSV color space", International Conference on Recent Trends in Information Technology, Chennai, India, **(2016)**, pp. 1-5.*

[4] *D. J. Bora and A. K. Gupta, "A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm", International Journal of Computer Trends and Technology. vol. 10, no. 2, **(2014)**, pp. 108-113.*

[5] *P. Nicoué, F. Numérique, "Qu'est-ce qu'un espace colorimetrique ?", **(2016)** Les numeriques, Photo, display in : https://www.lesnumeriques.com/photo/les-espaces-colorimetriques-pu102215.html **(**consulted on :**30/12/2022**).*

[6] *ACDSee Photo Manager, "Glossaire", display in: http://acdsee.helpmax.net/fr/glossaire/, **(**consulted on :**26/12/2022**).*

[7] *P. GANESAN, V. RAJINI, B. S. SATHISH, V. Kalist, and S. K. Basha, "Satellite image segmentation based on YCbCr color space", Indian Journal of Science and Technology. vol. 8, no. 1, **(2015)**, pp. 35-.41.*

[8] *Vorlesungen bis WS2004/05,"What colour space should I use?", Some definitions and explanations, display in: https://www5.in.tum.de/lehre/vorlesungen/graphik/info/csc/COL_9.htm#topic8 **(**consulted on :**30/01/2023**).*

[9] *HyperPhysics is hosted by the Department of Physics and Astronomy, "The C.I.E. Color Space", display in: http://hyperphysics.phy-astr.gsu.edu/hbase/vision/cie.html **(**consulted on :**30/01/2023**).*

[10] *O.M. Alia, R. Mandava, and M. E. Aziz, "A Novel mage Segmantation Algorithm Basd on Harmony Fuzzy Search Algorithm", International Conference of Soft Computing and Pattern Recognition SCPAR'09, IEEE Computer Society, **(2009)**, pp. 335-340.*

[11] *J.B. MacQueen, "Some methodes for classification and analysis of multivariate observations", Proccedings of fifth Berkeley symposium on mathematical statistics and probability, CA: university of California press. vol. 1, **(1967)**, pp. 281-297.*

[12] *H.S. Rahli, N. Benamrane, "Interprétation des images de mammographie par l'algorithme Search Harmony Annals ", Computer Science Series. vol. 10, no 2, **(2012)**, pp. 77-81.*

[13] *L. Laouamer, "*Approche exploratoire sur la classification appliquée aux images*", Memory Presentedt to the University of Uebec At Trois-Rivières, University Of Quebec, **(2006)**.*

[14] *J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Taylor & Francis, Journal of cybernetics. vol. 3, **(1973)**, pp. 32-57.*

[15] *J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy* c*-means clustering algorithm", Elsevier, Computers & Geosciences. vol. 10, **(1984)**, pp. 191-203.*

[16] *Z. Wang, N. Zhao, W. Wang, R. Tang, and S. Li, "A Fault Diagnosis Approach for Gas Turbine Exhaust Gas Temperature Based on Fuzzy C-Means Clustering and Support Vector Machine", Hindawi, Mathematical problems in Engieering. vol. 2015, **(2015)**.*

[17] *A. Maarir and E. S. Azougaghe, and B. Bouikhalene, "Automatic Change Detection Based on the Independent Component Analysis and Fuzzy C-Means Methods", In: International Conference on Business Intelligence, Springer, Cham,* ***(2022)****, pp. 178–187.*
[18] *M. H. Kesikoglu, U. H. Atasever, and C. Ozkana, "Unsupervised change detection in satellite images using fuzzy C-means clustering and principal component analysis", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 7, P.W2,* ***(2013).***
[19] *Google Maps,"Algeria", display in:* https://www.google.com/maps/@28.0948734,1.666194,5z ***(****consulted on :****31/03/2023****).*
[20] *C. Thomas, " Fusion d'images de résolutions spatiales différents ". Doctoral thesis, Mathematics, National School of Mines of Paris, France,* ***(2007)****.*