

Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan¹ and Lubna mhammed Bader²

^{1,2}Department of Computer Science- University of Technology, Baghdad

¹surorh@yahoo.com

²al.ahammed@yahoo.com

Received: 10 October 2016

Accepted: 24 April 2017

Abstract

Dealing with satellites Images, interpretation, improvement and modification area important issues of researchers concerning, because of the great benefits derived from them. GIS on the other hand is making to build disaster planning, crisis management, and alarm systems effective decision. In this paper a method for digitization gray scale satellite image to a digital map will propose. The proposed method depends on slicing the image into multi-layer, and preprocessing steps such as noise removal and contrast adjustment then use approach K-mean clustering unsupervised Classification . At the last step; the edges of segments (objects) are interpolated for softening the sharp edges using cubic spline interpolation, and these parts will be rendered with desired color that gets from original image. Layers results could be considered as one map for roads, buildings, green zones, etc. more than one layer could be merged in new map and may append information on it. The produced digital map will be saved as SVG vector file format.

Keywords: Raster image, vector image, image classification, interpolation, SVG.

صور الاقمار الصناعية المتجهة على اساس طرق الاشتقاق المكاني

رحاب فليح حسن¹ و لبنى محمد بدر²

^{1,2} قسم علوم الحاسوب - الجامعة التكنولوجية

الخلاصة

لتعامل مع الصور الفضائية كعمليات التحسين والاستيفاء والتعديل تعتبر من القضايا المهمة للباحثين ، بسبب الفوائد الكبيرة المستمدة منها. ومن ناحية أخرى، تقوم نظم المعلومات الجغرافية ببناء نظام تخطيط فعال للكوارث، وإدارة الأزمات، ونظم الإنذار. في هذه الورقة سوف تقترح طريقة لتحويل الصورة الفضائية الرمادية إلى المتجهات الخرائط الرقمية. تعتمد الطريقة المقترحة على تقطيع الصورة إلى طبقات متعددة، وخطوات المعالجة المسبقة مثل إزالة الضوضاء وتعديل التباين ثم استخدام خوارزميه التصنيف (K-mean) لتوليد الطبقات. في الخطوة الأخيرة تم استخدام احدى طرق الاستيفاء cubic (interpolation) spline لتخفيف الحواف الحادة للطبقات بعد ذلك يتم الحصول على تلوين هذا الطبقات من الصورة الاصلية. ويمكن اعتبار نتائج الطبقات كخريطة واحدة للطرق والمباني والمناطق الخضراء وما إلى ذلك. ويمكن دمج أكثر من طبقة واحدة في خريطة جديدة ويمكن إلحاق معلومات عنها. سيتم حفظ الخريطة الرقمية المنتجة بصيغه (SVG) الكلمات المفتاحية: صور الاقمار الصناعية، المتجهات، تصنيف الصور، الاشتقاق المكاني.

Introduction

Each pixel of an image data sensed from a distance exemplifies a spot at a particular place of the planet Earth. These images have a large number of purposes which contain meteorology, cartography & military intelligence. The images of a satellite could be one of the following: vapor of water, infrared & visible light as shown in Figure 1[1]. Image is the easiest way to obtain geographic information.



Figure1: Satellite Image

Satellite Images Vectorization Based on Clustering and Interpolation Technique**Rehab F. Hassan and Lubna mhammed Bader**

The important point here is whether the data which are obtained from satellite image will give required accuracy in GIS or not. Image digitization is like a conversion of the state of signals from alternating into digital by the way of taking specimen and quantization process. On a space of display, any x,y grid of coordinates can be called raster. The raster file is sometimes referred to as a bitmap because it contains information that is directly mapped to the display grid. A raster file is usually difficult to modify without loss of information. Examples of raster image file types are: BMP, TIFF, GIF, and JPEG files [3]. A vector graphic is a set of instructions to draw lines, curves, and lettering. The advantage of vector graphics is that, with a relatively small file size, they can be rendered sharply at any size and resolution. As a more abstract & straightforward representation of a data than raster image is supplied by structure of vector data, a conversion from raster to vector, the automatic one of course, is not simple to be done, or what we name "the process of vectorization", yet the vector-raster conversion is rather common. Tracing of line, recognition of shape, pre-processing, creation of topology, attribute assignment and acquisition of image are all contained in the process of raster – vector complete conversion. The basic idea of any vectorization method is the classification process, where if a pixel satisfies a certain set of criteria then that pixel is assigned to the class that corresponds to those criteria. The methods of classification of image may be assorted into two major categories depending on two methods, the pixel based and the object based. The methods of pixel based have the ability to assort single pixels and they have not got to rely on the neighborhood or the pixel's spatial information. The methods of object (Region) based can deal with imagery that has a high resolution, and that could lead the process of classification of most of the methods of pixel based to be irritated [6]. Classes are going to be particularized with the ground's common features relying on the information kind taken from the main data. An individual pixel can symbolize class covers' mix, within class variableness or other patterns of complicated surface cover in remote sensing imagery, which one class cannot depict appropriately. Utilizing this image the demanded specifications are taken from. The varied specifications out of this like used land, forest and unused land and vegetation indices are particularized. The algorithms of classification, right after taken the specifications out of the image, can be applied to obtain varied groups of classification. The classification of an

Satellite Images Vectorization Based on Clustering and Interpolation Technique**Rehab F. Hassan and Lubna mhammed Bader**

image can be considered as a complicated operation, and sure, more than one factor can be effective on it. There have been many non-parametric classifiers which are considered as significant approaches for the classification of multi-source data such as decision tree classifier, knowledge based classification and fuzzy logic. There is another point which is very significant in the operation of vectorization, that is the mathematical process of interpolation. [3][8].

Literature Review

There are many vectorization approaches developed in various domains for different purposes. In the following, some proposals have been drawn from the scientific literature:

- **Saleh, M.& Mohammed, K. & Ahmed, M.,** in 2012 [17] developed a conversion model of vectorization for the images of quad tree satellite to the format of vector, so that model may be utilized in the projects of GIS. Results showed that the model is acceptable for mapping purposes, and this quality leads it to be a well model utilized to get the quad tree vectorization issue settled. And for accuracy purposes, this model must be checked and enhanced. There are two ways of accuracy to be considered, which are location and size.
- **Sumner, R. & Noris, G ,A Hornung and Gross, M.,** in 2013[18] proposed - for the purpose of the drawings of clean line which get the topology of drawings analyzed to get by the junction obscurity – an algorithm of vectorization. We have two mechanisms contained in the approach, a mechanism assists obscure hard cases, which is called clustering of gradient – based pixel, and another uses the topology of drawing to make educated picks as treating with junctions in order to enhance the existing mechanisms result which supply the topology of drawing, and this one called procedure of reverse drawing.
- **Suleyman, E. & Ahmet, S.,** in 2015 [11] created – for transferring an image of a Landsat satellite from some machine to another one throughout web services – a scenario. It is demonstrated by experimental results that the reduction of the size of file belongs to data of an original raster map is feasible, to achieve the drastically transmitting and to the quality of scalability to be achieved on the application. Anyway, the hardness of algorithms causes time consumption for the operation of simplifying.

Satellite Images Vectorization Based on Clustering and Interpolation Technique**Rehab F. Hassan and Lubna mhammed Bader**

- **B. Yuan, S.C. Liew, L.K. Kwoh**, in 2010 [21] have used Line-Segment Vectorization on High-Resolution Satellite Images: This paper presents a procedure to extract and vectorize the mentioned line-segments. Scan part utilizes the continuing of edges and correlating of intensity to get line-segments recognized domestically. The incomes evolved by experiment utilizing images of satellite have high-resolution maintain that the suggested procedure of vectorizing line-segment can be both influential and competent for the applications in reality like manual/automatic establishing top extracting from the images of stereo satellite that have a high resolution.

3. Image Classification

- Image classification is a complex process that may be affected by many factors. Effective use of multiple features of remotely sensed data and the selection of a suitable classification method are especially significant for improving classification accuracy. Non-parametric classifiers such as fuzzy logic, neural network, decision tree classifier, and knowledge-based classification have increasingly become important approaches for multisource data classification. More research, however, is needed to identify and reduce uncertainties in the image-processing chain to improve classification accuracy. In general, image classification approaches can be grouped as supervised and unsupervised, or parametric and nonparametric, or hard and soft (fuzzy) classification, or per-pixel, sub-pixel, and per-field. Based on whether training samples used or not used image classification can be of two types (Supervised Classification) and (Unsupervised Classification), in this method use the K-Means Clustering was chosen to classify each pixel in the image input as well as to label each classification. In this way, the layers of the workbook were created according to the work of this algorithm, which depends mainly on the value centroids

- **The algorithm of K-Means Clustering**

This algorithm can be considered as one of the algorithms of unsupervised learning to settle the problem of well-established clustering or classification. It obeys a way quite simple and modest to get a given data set classified throughout a given number of clusters [12]. Defining k number of centroids, each one should be for a cluster. And those centroids

Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

better sited in a cunning way due to the fact that says different location causes different result. For that, arranging them very far from one another would be the ideal solution. Then we ought to take each point of a given data set and link it up to the closest centroids. As all points are waiting to settle, the first step is considered to be finished and an early age of group is accomplished [13]. At this end we would have to add up k new centroids again a center of the clusters produced from the preceding step. After obtaining those k new centroids, a new binding has to be achieved be among the points the same data set and the closest new centroid. And there has been a loop. As a result of that one might take notice that the k centroids alter their places slowly till there are no more alterations. That would mean a state of quietness happen to the centroids. Eventually, the object of the said algorithm is minimizing an objective function, a squared error function in such a situation [14].

$$J = \sum_{j=1}^k \sum_{i=1}^x \|X_i^{(j)} - c_j\|^2 \dots 1$$

Where C is centroid of cluster x is pixels in image and j is Euclidian distance. The the following algorithm is in detail:

Algorithm: K-Mean Clustering	
Input:	Gray scale Image
Output:	Multi layers as 2D matrix
Steps:	
1:	Begin
2:	select N number of clusters (layers)
3:	convert the pre-processed image to one dimension array V (vector)
4:	Select random N points as centroid of each clusters
5:	for each point in V find Euclidean distance from each centroid depend on objective function ()
6:	specify a minimum distance from N distance to be member of clusters
7:	find mean of each clusters as new centroids
8:	Apply a previous step for a new centroids until there are no changing on new centroids.

Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

- ```

9: convert V to 2D array represent the same dimension of image with layer label
 layer Matrix
10: End Algorithm

```

- Algorithm of vectorization

Vectorization, (Raster to Vector) or simply written R2V, is a process for converting raster image to a vector image. The history of vectorization process begins from scanned-line art black-white images and continues to evolve to cartoon drawings and finally realistic images. Various approaches have been proposed to solve the same problem. The most of these works focus on line drawings, fonts, texts and sketches which are in fact binary images with sparse set of pixel in the region of interest. On the other hand, the Vectorization (raster-to-vector conversion) consists of analyzing a raster image to convert its pixels representation into a vector representation. The basic assumption is that such a vector representation is more suitable for further interpretation of the image [20].

## Interpolation using Cubic Spline

There are many Interpolations method used to estimate data or points from other collection of them, the estimation values have a relation with the other known values. In this paper the cubic Spline used for interpolated data that will be used in vector graphics which is explained in figure 2[19].

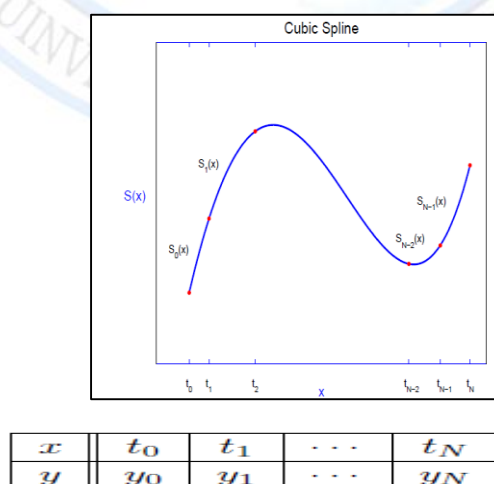


Figure (2) Spline interpolation

## Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

The cubic Spline procedure as follow:

Cubic Spline interpolation is a third method used for interpolation. As previous two methods, the cubic Spline interpolation is used to uniform or smooth the extracted layers borders of objects. It depends on polynomial equation to find intermediate point between sharp border points. The method is smoother and has smaller error than those interpolation where the total steps in this interpolation method are explain in algorithm.

| Algorithm: Cubic Spline Interpolation |                                                                          |
|---------------------------------------|--------------------------------------------------------------------------|
| Input:                                | LI Layer Image {layer image}, Sh, Sw insertion point in height and width |
| Output:                               | LE Enhanced Layer Image                                                  |
| Steps:                                |                                                                          |
| 1:                                    | Begin                                                                    |
| 2:                                    | Set Ho=height of LI, Set Wo=width of LI                                  |
| 3:                                    | for r = 1 to Ho                                                          |
| 4:                                    | Set n=0                                                                  |
| 5:                                    | for k=1 to Sw                                                            |
| 6:                                    | If k= 1 then $LE(r,c+n)=LE(r,c)$                                         |
| 7:                                    | Else Set $T=((LI(r,c+1)-LI(r,c))/(Sw-k))$                                |
| 8:                                    | Set $LE(r, c+n)=LI(r,c)+T$                                               |
| 9:                                    | End if                                                                   |
| 10:                                   | n=n+1                                                                    |
| 11:                                   | End for k                                                                |
| 12:                                   | End for c                                                                |
| 13:                                   | End for r                                                                |
| 14:                                   | for c=1 to $W_v \times Sw -1$                                            |
| 15:                                   | for r = 1 to Ho-1                                                        |
| 16:                                   | m=0                                                                      |
| 17:                                   | for k=1 to Sh                                                            |

## Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

```

18: If $k = 1$ then $LE(r+m,c)=LI(r,c)$
19: Else
20: Set $T=((LE(r+1,c)-LI(r,c))/(Sh-k))$
21: Set $LE(r+m,c)=LI(r,c)+T$
22: End if
23: $m=m+1$
24: End for k
25: End for r
26: End for c
27: Return LE {Enhanced Layer Image}
28: End Algorithm

```

**The Proposed method**

The proposed method is consists of stages start with pre-processing for satellite image. Each stage has specific functions; the main stages of the proposed method are explained in Figure3. Vectorization of Satellite images is the goal of this work based on interpolation methods. The basic method can have a proper definition by some relationships of mathematics that are among paths that connects points for image description & the points themselves. Image clustering, object extraction, border interpolation and object filling is the last stage in proposed method.

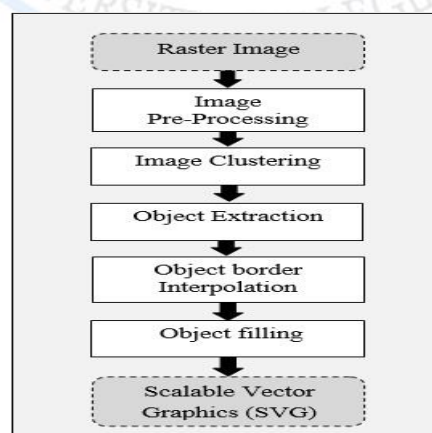


Figure (3): Proposed Method Framework

### 1. Image Pre-Processing

The pre-processing stage that should be applied on satellite image explained in figure 4 is begin by converting the image to gray scale image and then remove noise by applying median filter. The third step is image sharpening or image contrast adjustment, sharpening of image is useful to explain some details that are not clear and increase image contrast. Essentially, it contrasts to get edges enhanced. The step is grey scale image histogram equalization. Histogram equalization improves images contrast by getting the values of an image pixel value converted, or getting the ones in an index image colour map converted, so the output images histogram is nearly quantified. The range of histogram value is in  $[0, 255]$ . The histogram of the new image (histogram equalization) will be better in changing the intensity image pixels. Figure 5 explains the equalization of gray scale image.

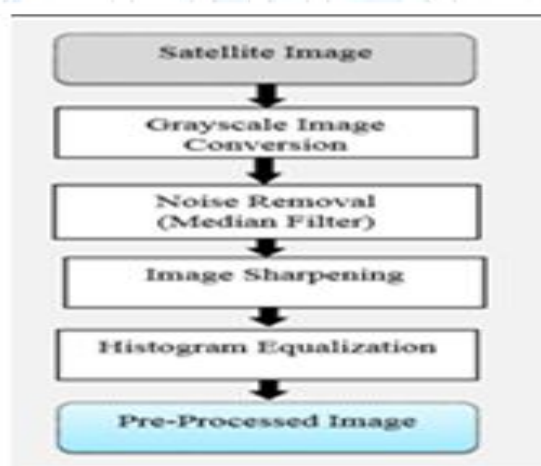


Figure (4): Image Pre-Processing Steps

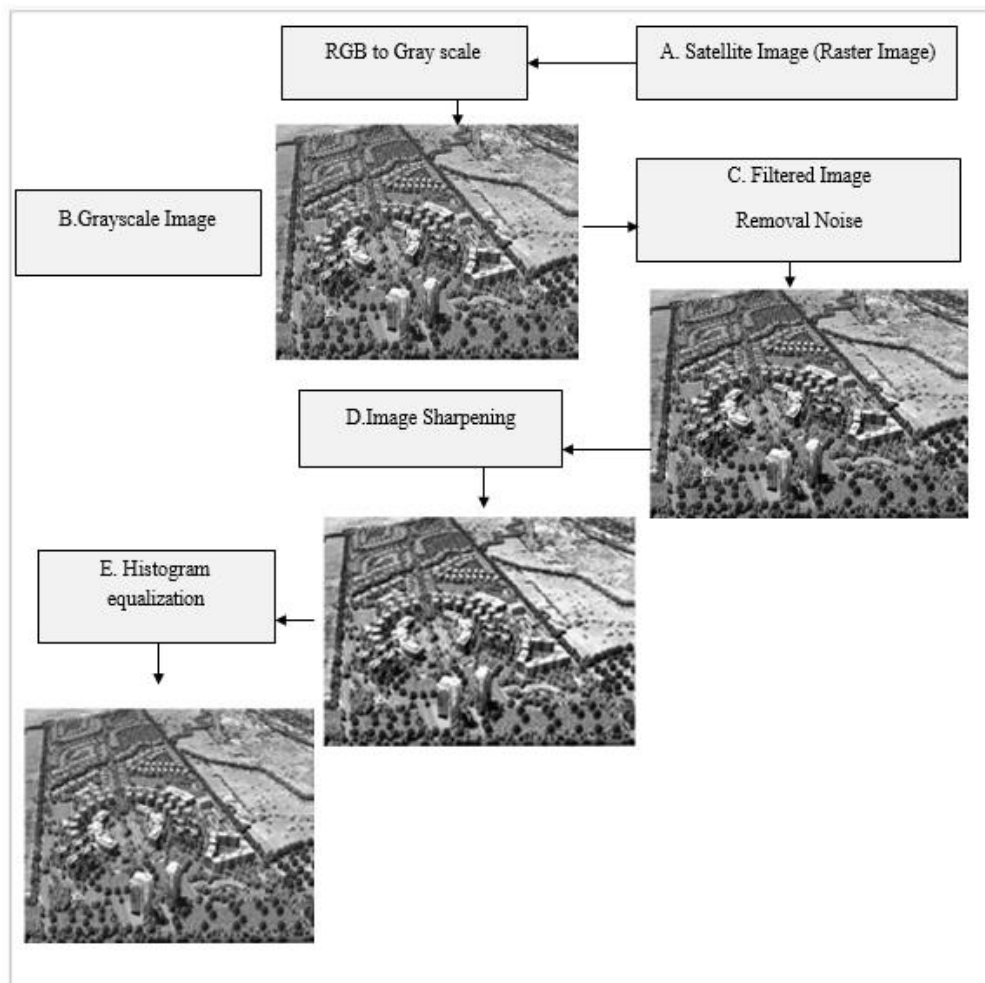


Figure (5): Image Histogram Equalization

## 2. Image Clustering

The k-mean clustering approach is used in this proposal for image clustering. In this stage the number of layers requested or clusters should be specified, and the main steps of clustering process will follow the following steps as explained in figure 6.

**Step1:** select N number of clusters (layers)

**Step 2:** convert the pre-processed image to one dimension array V (vector)

**Step 3:** Select random N points as centroid of each clusters

**Step 4:** for each point in V find Euclidean distance from each centroid depend on objective function ( )

## Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

**Step 5:** specify a minimum distance from N

distance to be member of clusters

**Step 6:** find the mean of each clusters as new centroids

**Step 7:** apply a previous step for a new centroids until there are no changing on new centroids.

**Step 8:** convert V to 2D array represent the same dimension of image with class labelClassMatrix



**Figure 6: image Clustering**

### 3. Object Extraction

The object extraction depend on the previous step by applying segmentation of each layer (cluster) alone, the segmentation of layers shown in figure 7, and follow the following steps:

**Step 1:** for each layer construct zero matrix  $ZM_1, ZM_2, \dots, ZM_n$

**Step 2:** for each layer from 1 to N apply step 3

**Step 3:** if ClassMatrix equal to layer number then set the layer matrix = color in pre-processed image otherwise set layer matrix to zero

**Step 4:** each layer will be an object

Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

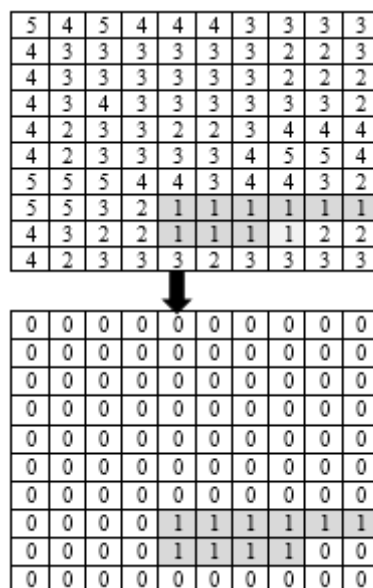


Figure 7: splitting object

Figure 8 shows an example of multi-layer extraction for a single satellite image:

## Satellite Images Vectorization Based on Clustering and Interpolation Technique

Rehab F. Hassan and Lubna mhammed Bader

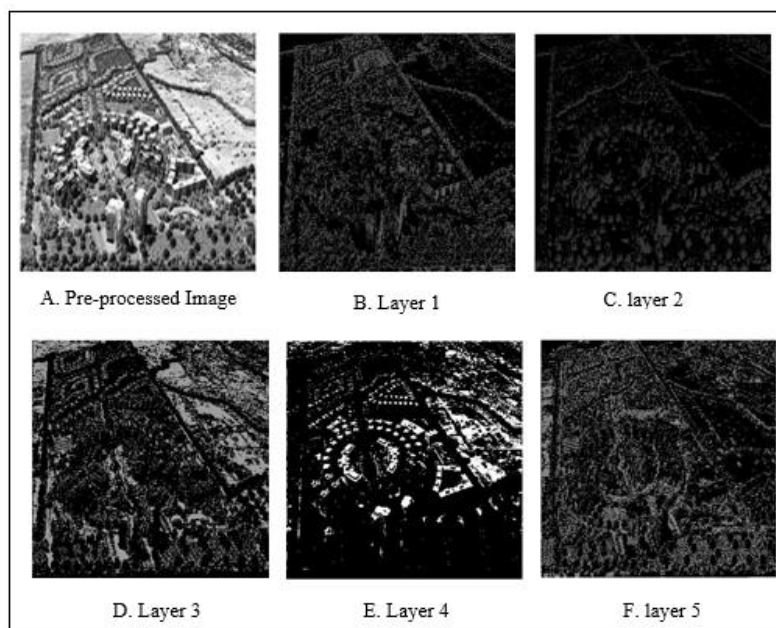


Figure 8 Layers Extraction

### 4. Object border Interpolation

Cubic Spline interpolation applied on each layer extracted from the previous step to smooth the borders of objects. An interpolating polynomial is gained by the said method, and that polynomial will be softer and has less errors. The interpolation applied on each layer separately for later rendering it. Figure 12 shows an example for this step.

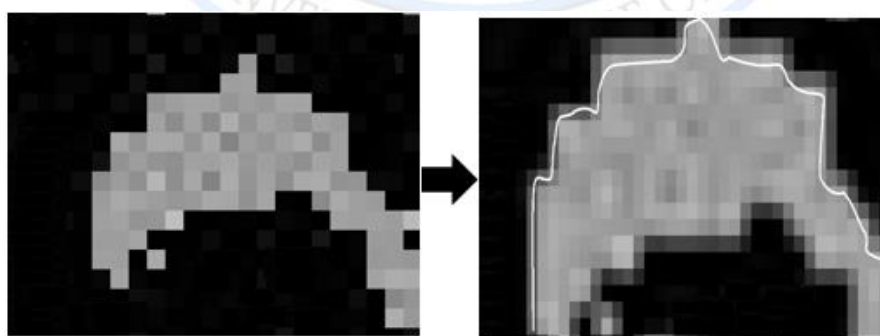


Figure (9): Cubic Interpolation for Different layers

## 5. Object filling

The rendering (filling) with color is a last step in the proposed method that recognize each layer, see figure 13, the selection of color depend on calculating the histogram of each layer in original image and take the maximum frequency in red color, in this a maximum green is

found and a maximum blue is also found by the following procedure:

**Step 1:** for each object (layer) obtained from previous apply

**Step 2:** depend on objects segment before split three band of original image

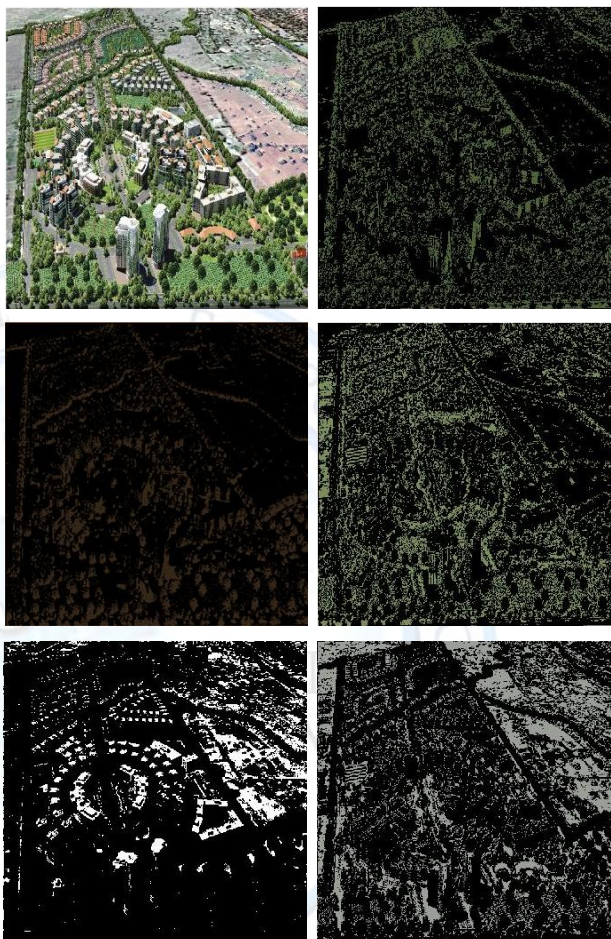
**Step 3:** find the histogram of red band of object

**Step 4:** find the maximum value in histogram that not equal to zero

**Step 5:** by fixing the maximum red find maximum green that not equal to zero

**Step 6:** by fixing the maximum red value with **Figure (10)):** color object extraction maximum green find maximum blue that not equal to zero

**Step 7:** render the object by the obtained color



### Implementation and Testing Results

The Proposed method have been applied on several satellite images, the number of layer selected is 5 layers. The total test images and layers shown in figure 14. The number of layers are variable and selected by the user depend on the approach used it for.

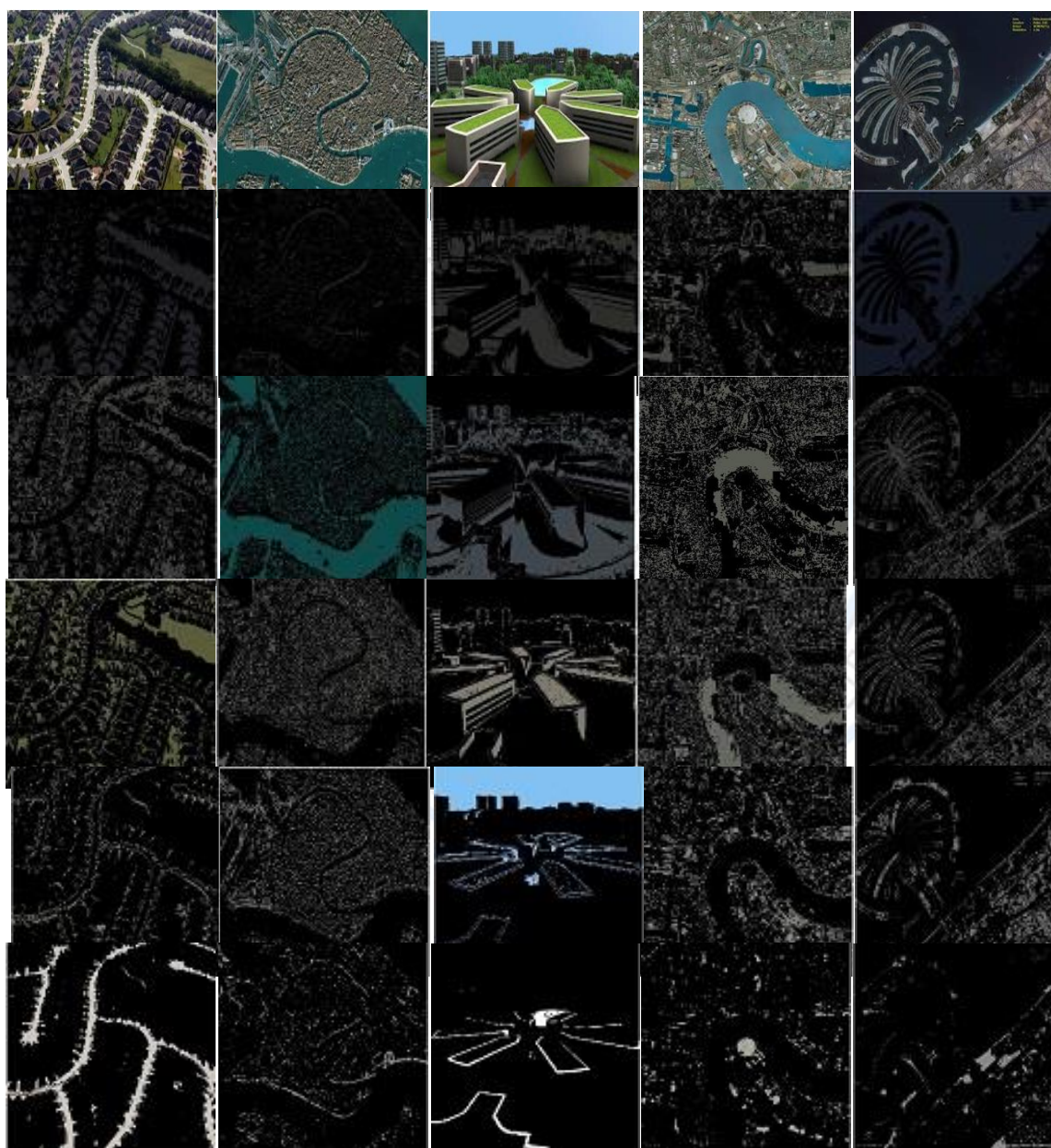


Figure (11): Samples of satellite image

**Satellite Images Vectorization Based on Clustering and Interpolation Technique****Rehab F. Hassan and Lubna mhammed Bader**

Satellite image vectorization method has ability to provide a solution, which is often easier to interpretation process of zooming with keeping a high resolution of satellite images. The layers obtained from proposed method may be used in different kind of applications. It are possible to use other interpolation method for softening the edges as well as it possible to choose other methods in image classification. The morphological operation may be used to fill or remove small point from layers such open or close that used dilation and erosion. The vectorising images may cause extra size in file but at all it useful in many approaches.

**References**

1. K. Jacobsen, "High Resolution Satellite Imaging Systems - Overview", University of Hannover, 2004.
2. Chris Solomon and Toby Breckon, "Fundamentals of Digital Image Processing A Practical Approach with Examples in Matlab", WELLY- BLACKWELL, 2011.
3. HU Wei, LIU Yong-xue, LI Man-chun, CHEN Jie-li, MAO Kun, " Research on ETM+ Image Restoration Method Based on Local Correlation Analysis" Science, School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210093, China), 2011.
4. Tuia D, Volpi M., "A survey of active learning algorithm for supervised remote sensing image classification", IEEE Journal of Selected Topics in Signal Processing. 2011. p. 606–17.
5. Xu S, Wu Y., "An algorithm for remote sensing image classification based on artificial immune B-Cell network", Remote Sensing and Spatial Information Sciences. 2008. p. 107–10.
6. Li Y, Cheng B., "An improved K-Nearest neighbor algorithm and its application to high resolution remote sensing image classification", IEEE Geo informatics. 2009.
7. Blanzieri E, Melgani F., "Nearest neighbor classification of remote sensing images with the maximum margin principle". IEEE Transaction on Geoscience and remote sensing. 2008. p. 1804–11.
8. Santos AB, De S, Arayo A, Menotti D., "Feature selection for classification of remote sensing hyper spectral images A filter approach using genetic algorithm and cluster validity", -world computer proceedings. 2012. p. 1–7.

**Satellite Images Vectorization Based on Clustering and Interpolation Technique**

**Rehab F. Hassan and Lubna mhammed Bader**

9. Melgani F, Brizzzone L. "Classification of hyperspectral remote sensing images with support vector machines", IEEE Transaction of Geoscience and Remote Sensing. 2004. p. 1778–90.
10. Gen QZ, Ling ZC, Qiong L, Heri XX, Zhang G. "High efficient classification on remote sensing images based on SVM", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2008. p. 1011–4.
11. Uma Shankar B., Meher K, Ghosh A, Bruzzono L., "Remote sensing image classification A neuro fuzzy mcs approach", ICVGIP, Springer-Verlag Berlin Acidelberg. 2006 p. 128– 39.
12. Bardossy A, Samariago L., "Fuzzy Rule based classification of remotely sensed imagery", IEEE Transaction on Geoscience and Remote Sensing. 2002. p. 362–74.
13. Zhang J, Foody GM., "A fuzzy classification of sub urban land cover from remotely sensed imagery", International Journal of Remote Sensing. 1998. p. 2721–38.
14. Sathishkumar, Saravanan G., "Fuzzy-Based Sub-Pixel Classification of Satellite Imagery", Pattern recognition letters. 2000. p. 385–97
15. Lillesand TM, Kiefer RW. Remote Sensing and Digital Interpretation. Wiley. 2000. p. 724.
16. Christianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press; 2000.
17. Saleh Mesbah, Mohamed Kholief, Ahmed Mahdy" A Gis vectorizatint model for quad-tree satellite IMAGES" ICCTA, 2012,
18. oris, G, Hornung and Sumner" Topology-Driven Vectorization of Clean Line Drawings" ACM, Vol 32, 2013
19. E. J. M. DELHEZ" A Spline Interpolation Technique that Preserves Mass Budgets" AML, 2002
20. irija Dharmaraj" Algorithms for Automatic Vectorization of Scanned Maps" UCGE, Vol 20226, 2005
21. Hang, J., Yanming Feng, Bofeng Li" Road Network Extraction with New Vectorization and Pruning from High-resolution RS Images" IEEE , 2008