Global Advanced Research Journal of Engineering, Technology and Innovation (ISSN: 2315-5124) Vol. 1(9) pp. 030-041, January, 2013
Available online http://garj.org/garjeti/index.htm

Full Length Research Paper

Color Based Image Segmentation using Different Versions of K-Means in two Spaces.

Faten Abu Shmmala[#1], Wesam Ashour[#2]

[#1#2]Computer Department, Islamic University

Abstract 24 December 2012

In this paper color based image segmentation is done in two spaces. First in LAB color space and second in RGB space all that done using three versions of K-Means: K-Means, Weighted K-Means and Inverse Weighted K-Means clustering algorithms for different types of images: biological images (tissues and blood cells) and ordinary full colored images. Comparison and analysis are done between these three algorithms in order to differentiate between them.

Keywords: Image segmentation, LAB space, RGB space, K-Means, Weighted K-Mean and Inverse Weighted K-Mean.

INTRODUCTION

Image processing is very attractive field, mostly image segmentation. Image segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse (Linda and George, 2001). Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Other type of segmentation is color based segmentation which this paper interested in. Image segmentation has many application for example in medical imaging (Dzung et al., 2000): to locate tumors and other pathologies, measure tissue volumes, computer-guided surgery, diagnosis, treatment planning and study of anatomical structure or for locating objects in satellite images (roads, forests, etc.) and it can be used for face and fingerprint recognition, traffic control systems and brake light detection and machine vision. Several general-purpose algorithms and techniques have been developed for image segmentation. Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem for a problem domain. These methods used in image segmentation such as: Clustering methods, Histogram-based methods, Edge detection, Region growing methods, Level set methods, Graph partitioning methods, Watershed transformation, Model based segmentation, Multi-scale segmentation, Semi-automatic segmentation and Neural networks segmentation. In this particular paper clustering methods used to accomplish color based image segmentation for different types of images. Clustering (Barbakh et al., (2009) is an active area in data mining and machine

Corresponding author Email: faten_abu_shmmala@hotmail.com

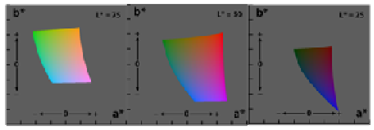

Figure 1. Lab Space

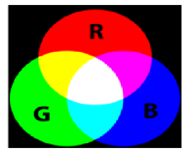

Figure 2. RGB space

Figure 3. MONOCYTE.

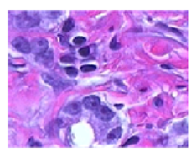

Figure 4.

Figure 5. This is an image of some vegetables

learning research. It can be considered as one of the most famous and important unsupervised learning problems. Clustering is a partitioning of a data set or objects into meaningful groups or clusters. The cluster can be defined as a collection of objects which resemble each other and are dissimilar or different to the objects. The K-Means algorithm (Hartigan and Wang, 1979) (Lloyd, 1957) (MacQueen, 1967) is one of the most frequently used investigatory algorithms in data analysis. The algorithm attempts to locate K prototypes or means throughout a data set in such a way that the K prototypes in some way best represents the data. It is an iterative algorithm in which K-Means are spread throughout the data and the data samples are allocated to the mean which is closest (often in Euclidean norm) to the sample. Then the K-Means are repositioned as the average of

Figure 6. Thermal image for a baby

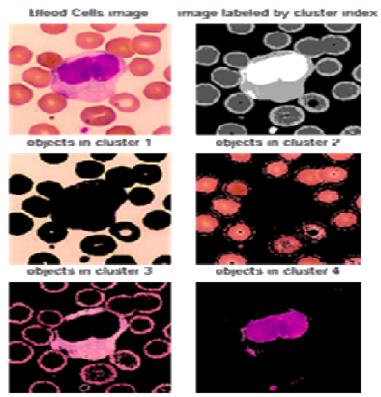

Figure 7. These are results form segmenting the Monocyte image using (K-Means for K=4), where each sub-image represents different cluster.

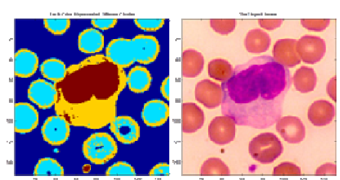

Figure 8. Same result with all clusters are in the same image but with different colors.

data points allocated to each mean. Perhaps it is the most widely used non-supervised technique. Undoubtedly, each approach has its own pros and cons. Many attempts have been made to improve the the sensitivity of the initial prototypes in order to solve this genetic algorithms (Krishna et al., 1997) (Krishna and Narasimha, 1999). Besides, many authors have improved the basic K-Means mapping through a neural network function (Bezdek, 1981); and Krishna suggested using (Hall et al., 1992). However, most of these improvements on the K-Means algorithm are computationally performance of the basic K-Means algorithm. The C-means, for instance, incorporates a fuzzy criterion demanding. Most famous disadvantage of the K-Means is problem Arthur and Vassilvitskii (Arthur and Vassilvitskii, 2006) improved the K-Means algorithm by substituting the random allocation of the prototypes with a seeding technique. They give experimental results that show the advantage of this algorithm in time and accuracy. In (Barbakh, 2007) (Barbakh et al., 2006) (Pena et al., 2007) new family of clustering algorithms that solve the problem of sensitivity to initial conditions in the K-Means algorithm driven, these algorithms are Weighted K-Means and Inverse Weighted K-Means. In this particular paper K-Means, Weighted K-Means And Inverse Weighted K-Means Are Used In Color-Based Image Segmentation To Differentiate Between Them As Much As Possible
.

K-MEANS CLUSTERING ALGORITHM

The K-Means algorithm, first developed four decades

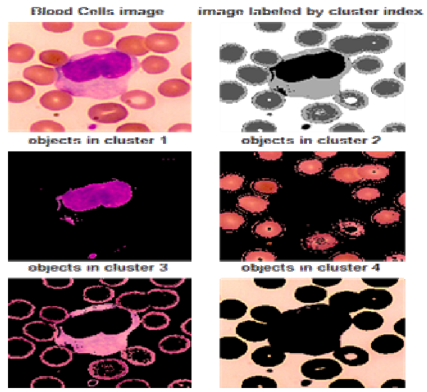

Figure 9. These are results form segmenting the Monocyte image using Weighted K-Means for K=4, where each sub-image represents different cluster

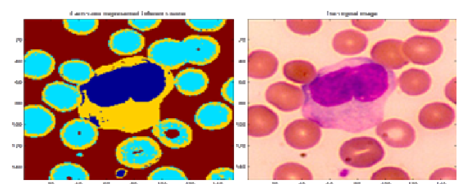

Figure 10. Same result with all clusters are in the same image but with different colors.

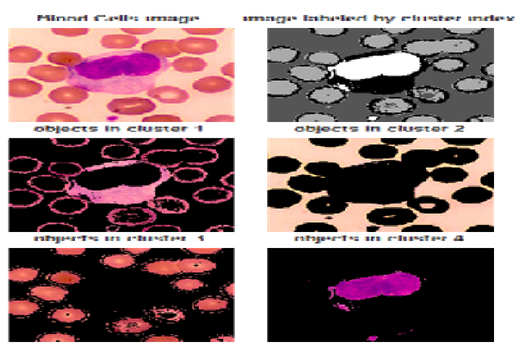

Figure 11. These are results form segmenting the Monocyte image, using Inverse Weighted K-Means for K=4, where each sub-image represents different cluster.

ago(MacQueen, 1967), is one of the most popular centre-based algorithms that attempts to find K clusters which minimize the mean squared quantization error , MSQE The algorithm tries to locate K prototypes (centroids) throughout a data set in such a way that the K prototypes in some way best represent the data. A summarization of the K-Means algorithm through the following steps Barbakh et al., (2009):

1. Initialization
a) Define the number of prototypes (K)
(b) Designate a prototype (a vector quantity that is of the same dimensionality as the data) for each cluster.
2. Assign each data point to the closest prototype. That data point is now a member of the class identified by that prototype.
3. Calculate the new position for each prototype (by calculating the mean of all the members of that class).
4. Observe the new prototypes' positions. If these values have not significantly changed over a certain number of iterations, exit the algorithm. If they have go back to step 2.

The main problem of the K-Means algorithm (Barbakh et al., 2009) is its dependency on the prototypes' initialization. If the initial prototypes are not chosen

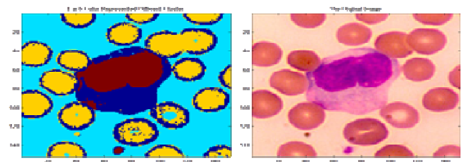

Figure 12. Same result with all clusters are in same image but with different colors

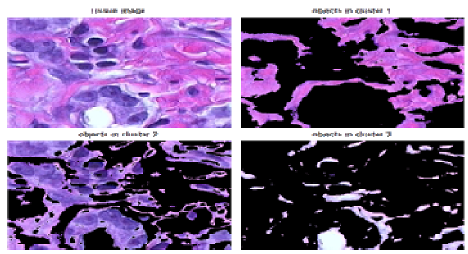

Figure 13. This the result form segmenting the Tissue image using K-Means for K=3, where each sub-image represent different cluster

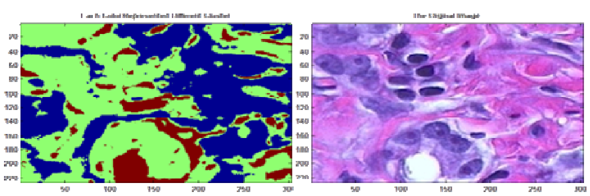

Figure 14. Same result with all clusters are in same image but will different colors

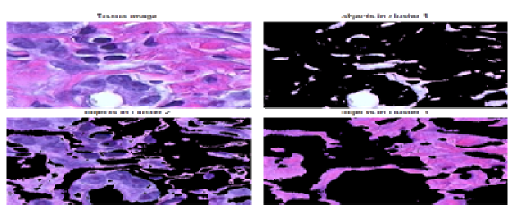

Figure 15. These are result form segmenting the Tissue image using Weighted K-Means for K=3, where each sub-image represent different cluster

carefully the computation will run the chance of converging to a local minimum rather than the global minimum solution. Thus initializing prototypes appropriately can have a big effect on K-Means. The performance function for K-Means may be written as

$$J_{Km} = \sum_{i=1}^{N} \min_{j=1}^{K} \| \mathbf{x}_i - \mathbf{m}_j \|^2 \quad (1)$$

WEIGHTED K-MEANS CLUSTRING ALGORITHM

A natural extension of the K-Means problem allows us to include some more information, namely, a set of *weights* associated with the data points. These might represent a measure of importance, a frequency count, or some other information. The intent is that a point with a weight of 5.0 is twice as "important" as a point with a weight of 2.5, for instance. This gives rise to the "Weighted" K-Means problem. Weighted K-Means performance function have the following properties (Barbakh et al., 2009):

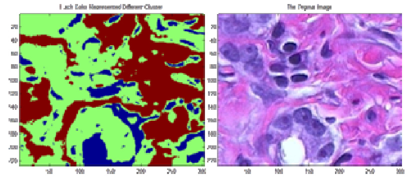

Figure 16. Same result with all clusters are in same image but will different colors

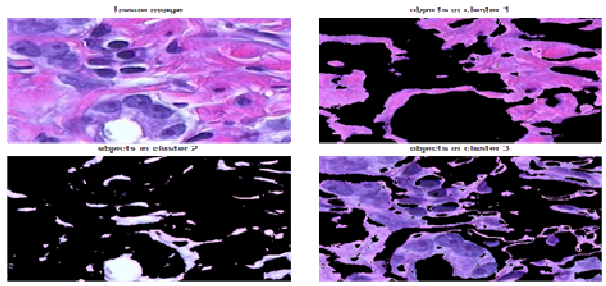

Figure 17. These results form segmenting the Tissue image using Inverse Weighted K-Means for K=3, where each sub-image represent different cluster

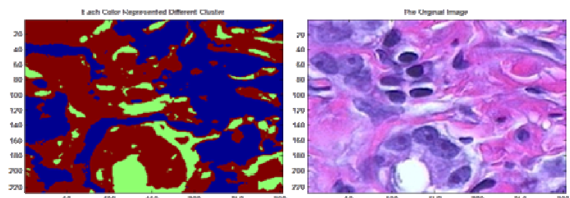

Figure 18. Same result with all clusters are in same image but will different colors

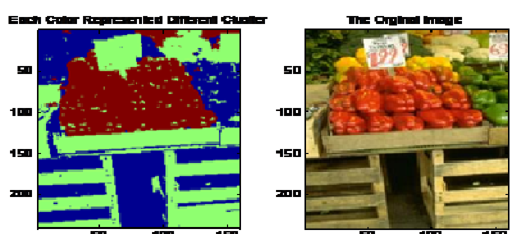

Figure 19. These results form segmenting the Vegetables image using K-Means for K=3, where each color represent different cluster

1-Minimum performance gives an intuitively `good' clustering.

2-It creates a relationship between all data points and all prototypes. a performance function such as:

$$J_{WK} = \sum_{i=1}^{N} \left[\sum_{j=1}^{K} \| \mathbf{x}_i - \mathbf{m}_j \| \right] \min_{k=1}^{K} \| \mathbf{x}_i - \mathbf{m}_k \|^2 \quad (2)$$

The rationale behind this performance function is that tries to utilise the minimum distance in the learning algorithm but retain the global interaction which is necessary to ensure all prototypes play a part in the clustering. The resulting algorithm called Weighted versions [3] of K-Means. This gives a solution of:

$$\mathbf{m}_r(t+1) = \frac{\sum_{i \in V_r} \mathbf{x}_i a_{ir} + \sum_{i \in V_j, j \neq r} \mathbf{x}_i b_{ir}}{\sum_{i \in V_r} a_{ir} + \sum_{i \in V_j, j \neq r} b_{ir}} \quad (3)$$

where *Vr* contains the indices of data points that are closest to m*r*, *Vj* contains the indices of all the other points and

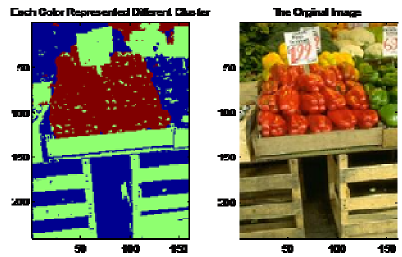

Figure 20. These results form segmenting the Vegetables image using Weighted K-Means for K=3, where each color represent different cluster.

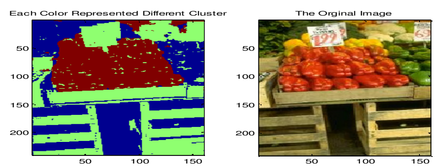

Figure 21. These results form segmenting the Vegetables image using Inverse Weighted K-Means for K=3, where each color represent different cluster

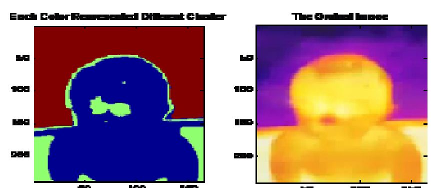

Figure 22. These results form segmenting the Thermal image using K-Means for K=3, where each color represent different cluster.

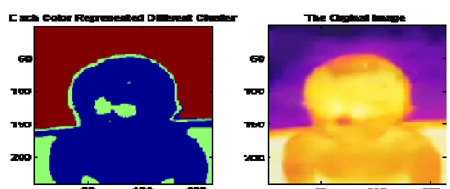

Figure 23. These results form segmenting the Thermal image using Weighted K-Means for K=3, where each color represent different cluster.

$$a_{ir} = \| \mathbf{x}_i - \mathbf{m}_r(t) \| + 2\sum_{j=1}^{K} \| \mathbf{x}_i - \mathbf{m}_j \|$$

$$b_{ir} = \frac{\| \mathbf{x}_i - \mathbf{m}_{k*} \|^2}{\| \mathbf{x}_i - \mathbf{m}_r(t) \|} \quad (4)$$

where again $k^* = \arg\min_k \| X_i - m_j \|$. However there can be some potential prototypes which are not sufficiently responsive to the data and so never move to identify a cluster. In fact, these points move to a (Weighted) centre of the data set. According to all of that a better version of WK clustering algorithm has been derived (Barbakh et al., 2009) is called Inverse Weighted K-Means

INVERSE WEIGHTED K-MEANS CLUSTERING ALGORITHM

Inverse Weighted K-Means (Gonzalez and Woods, 2002) considered an extension of Weighted K-Means, it is performance function :

$$J_{IWK} = \sum_{i=1}^{N} \left[\sum_{j=1}^{K} \frac{1}{\| \mathbf{x}_i - \mathbf{m}_j \|^p} \right] \min_{k=1}^{K} \| \mathbf{x}_i - \mathbf{m}_k \|^n \quad (5)$$

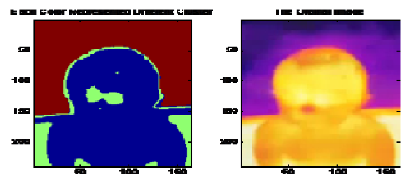

Figure 24. These results form segmenting the Thermal image using Inverse Weighted K-Means for K=3, where each color represent different cluster

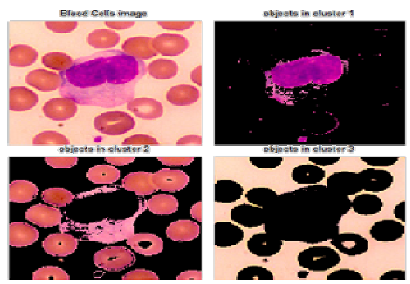

Figure 25. These results form segmenting the Monocyte image using K-Means for K=3, where each sub-image represent different cluster.

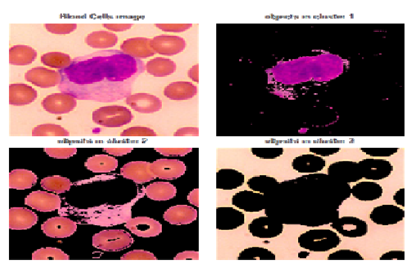

Figure 26. These results form segmenting the Monocyte image using Weighted K-Means for K=3, where each sub-image represent different cluster.

WITH :

$$\mathbf{m}_r(t+1) = \frac{\sum_{i \in V_r} \mathbf{x}_i a_{ir} + \sum_{i \in V_j, j \neq r} \mathbf{x}_i b_{ir}}{\sum_{i \in V_r} a_{ir} + \sum_{i \in V_j, j \neq r} b_{ir}} \quad (6)$$

Where *Vr* contains the indices of data points that are closest to *mr*. *Vj* contains the indices of all the other points and:

$$a_{ir} = -(n-p) \parallel \mathbf{x}_i - \mathbf{m}_r(t) \parallel^{n-p-2} - n \parallel \mathbf{x}_i - \mathbf{m}_r(t) \parallel^{n-2} \sum_{j \neq k*} \frac{1}{\parallel \mathbf{x}_i - \mathbf{m}_j \parallel^p}$$

$$b_{ir} = p \frac{\parallel \mathbf{x}_i - \mathbf{m}_{k*} \parallel^n}{\parallel \mathbf{x}_i - \mathbf{m}_r(t) \parallel^{p+2}} \quad (7)$$

N ≥ P IF THE DIRECTION OF THE FIRST TERM IS TO BE CORRECT AND N ≤ P +2 TO ENSURE STABILITY IN ALL PARTS OF THAT EQUATION.

METHODOLOGY

In this paper images are segmented in two spaces LAB and RGB space.

LAB Space

LAB color space is a color-opponent space with dimension ***L*** for lightness and ***A*** and ***B*** for the color-opponent dimensions, a space which can be computed

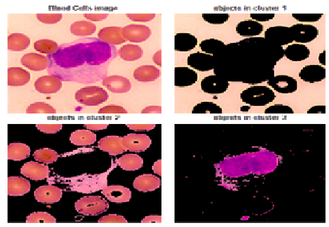

Figure 27. These results form segmenting the Monocyte image using Inverse Weighted K-Means for K=3, where each sub-image represent different cluster.

Figure 28. This is the noisy Vegetable image with speckle noise (variance = 0.06)

via simple formulas from the *XYZ* space, but is more perceptually uniform than *XYZ* (Marques, et al., 2005). *Perceptually uniform* means that a change of the same amount in a color value should produce a change of about the same visual importance. When storing colors in limited precision values.

Unlike the RGB, LAB color is designed to approximate human vision. It aspires to perceptual uniformity, and its L component closely matches human perception of lightness. It can thus be used to make accurate color balance corrections by modifying output curves in the A and B components, or to adjust the lightness contrast using the L component which model the output of physical devices rather than human visual perception, these transformations can only be done with the help of appropriate blend modes in the editing application.

RGB Space

The RGB color model is an additive color model in which red, green, and blue light are added together in various ways to reproduce a broad array of colors. The name of the model comes from the initials of the three additive primary colors, red, green, and blue. The main purpose of the RGB color model is for the sensing, representation, and display of images in electronic systems, such as televisions and computers, though it has also been used in conventional photography.

Before the electronic age, the RGB color model already had a solid theory behind it, based in human perception of colors.

In these both spaces the three clustering techniques are used to segment biological and ordinary images with different number of K (number of segments) and in the worst scenarios of prototypes locations, for example all prototypes location are in the center or at the corners of an image ..etc. In RGB space the features are r, g, b, x, y, z where r, g and b values along with coordination play strong part in clusters determination, but in LAB space the location (x ,y) does not matter.

RESULTS

Some of the tested images in LAB space are :

Fig3. MONOCYTE. This cell is the largest of the leukocytes and is agranular. The nucleus is most often "U" or kidney bean shaped; the cytoplasm is abundant and light blue (more blue than this micrograph illustrates). These cells leave the blood stream (diapedesis) to become macrophages. As a monocyte or macrophage, these cells are phagocytic and defend the body against viruses and bacteria. These cells account for 3-9% of all leukocytes. In people with malaria, endocarditis, typhoid fever, and Rocky Mountain spotted fever, monocytes increase in number.

Fig4. This image of tissue stained with hemotoxylin and eosin (H&E). This staining method helps pathologists distinguish different tissue types.

Some of the tested images in RGB space are :

The result from using random prototypes and noise free images are: first in LAB space

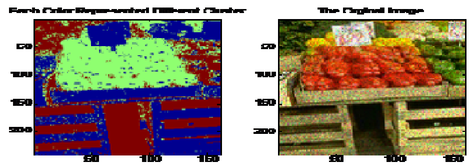

Figure 29. This is the noisy image clustered by K-Means with K=3.

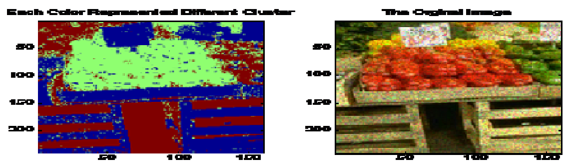

Figure 30. This is the noisy image clustered by Weighted K-Means with K=3.

Fig7 and Fig8 are the result from using K-Means with the Monocyte image. Fig9 and Fig10 using Weighted K-Means. Fig11 and Fig12 using Inverse Weighted K-Means.

As you will notice that Fig7, Fig9 and Fig11 are similar, the only difference is that which cluster contained which object. For example after using the K-Means (Fig7) at Monocyte image the result is four clusters, the first cluster is the pink area which is the background, the second cluster represents the red cells, the third cluster with a lighter pink color which gives the leukocytes and some of the edges of the red cells and the fourth one is the purple area which is the blue nuclei of the leukocytes, like wise Fig 9 and Fig11. In Fig8 the Monocyte segment image using the K-Means, but all clusters are in the same image, where each color represent a different cluster. You can see the leukocytes (Fig8) and the red cells edges in yellow where they all together considered one cluster and that is not good, like wise Fig10 and 12. To be able to segment the red cells alone with its edges it must be clustered into three parts not four,

Fig13 and Fig14 are the result from using K-Means with the Tissue image. Fig15 and Fig16 for the same image using Weighted K-Means. Fig17 and Fig18 for the same image using Inverse Weighted K-Means. Fig13 is the result from using the K-Means at Tissue image the result is three clusters, the first cluster gives the pink tissues, the second cluster represents the blue tissues and the third cluster gives the grey tissues, like wise Fig15 and Fig17.

In Fig14 the Tissue segment image using the K-Means, but all clusters are in the same image, where each color represent a different cluster (different tissue), like wise Fig16 and Fig18.

Second in RGB space

Fig19 is the result from using K-Means with the Vegetables image. Fig20 using Weighted K-Means. Fig21 using Inverse Weighted K-Means on the same image.

Fig19 is the result from using the K-Means at the Vegetables image the result is three clusters, but all clusters are in the same image, where each color represent different cluster (different object), like wise Fig20 and Fig21. The result are so similar between the three algorithms.

Fig22 is the result from using K-Means with the Baby image. Fig23 using Weighted K-Means. Fig24 using Inverse Weighted K-Means on the same image.

Fig22 is the result from using the K-Means at the Vegetables image the result is three clusters, but all clusters are in the same image, where each cluster represent different temperature.

As you will notice that the results are so similar thus, we can say it is identical among the three algorithms and that can be explained, since the three algorithms are actually one that has been modified to overcome there shortcoming (disadvantages). So, the real question is how we can test the three algorithms to give real assessments? We have tried that by the following three ways: First any clustering algorithm will give good result at high K (number of partitions), so we tried to decrease the number of K for example, **Monocyte** image in LAB space:

This previous image was tested with higher number of K and gives similar results, so the reality decreasing K will not give real disparity between the three algorithms.

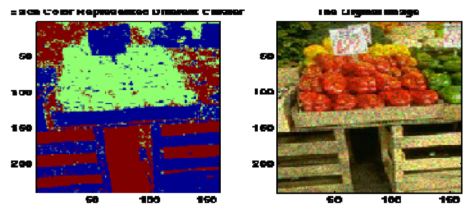

Figure 31. This is the noisy image clustered by Inverse Weighted K-Means with K=3.

Fig25 is the result from using K-Means with the Monocyte image. Fig26 using Weighted K-Means. Fig27 using Inverse Weighted K-Means on the same image. In all of the previous images the clustering algorithm is done with the initial prototypes are chosen randomly, so if we make it little bit harder by choosing the initial prototypes from one of the corners of the images in the all three clustering algorithms, by extensive testing, the results show that: Inverse Weighted K-Means is the only clustering algorithm that converge and give reasonably good results, which are the same if the prototypes are randomly chosen. K-Means give empty cluster and Weighted K-Means will not converge correctly. So as a result we could say that Inverse Weighted K-Means give better results than other clustering algorithms especially in hard circumstances such as bad initials centroids of the clusters.

The third scenario with noisy images. Noise has been defined in (Gonzalez and Woods, 2002) (McAndrew, 2004) to be any degradation in the image signal, caused by external disturbance. Most common types of noise are: Salt and Pepper noise, Gaussian noise and Speckle noise. Salt and pepper noise also called impulse noise, shot noise, or binary noise. This degradation can be caused by sharp, sudden disturbances in the image signal; its appearance is randomly scattered white or black (or both) pixels over the image. Gaussian noise is an idealized form of white noise, which is caused by random fluctuations in the signal. We can observe white noise by watching a television which is slightly mistuned to a particular channel. Gaussian noise is white noise which is normally distributed. Speckle noise: where as Gaussian noise can be modelled by random values added to an image; speckle noise (or more simply just speckle) can be modelled by random values multiplied by pixel values, hence it is also called multiplicative noise. Speckle noise is a major problem in some radar applications.

K-Means and its mentioned versions are highly sensitive to noise, by adding a standard amount of Pepper and Salt or Gaussian noise to an image, all K-Means failed, but by adding speckle noise with mean 0 and variance ≤0.3, all K-Means work well. On the other hand, Inverse Weighted K-Means gives the best results in that case. In case of adding speckle noise with higher values of variance >0.3 all K-Means failed.

Fig29 is the result from using K-Means with the Vegetables image. Fig30 using Weighted K-Means. Fig31 using Inverse Weighted K-Means on the same image.

As you will notice from the result that Inverse Weighed K-Means give the least noisy groups of clusters, but at high amount of noise all the mentioned K-Means versions give similar results which indicate the huge sensitivity of K-Means to noise. Weighted K-Means algorithm (Abras and Ballarin, 2005) prevents characteristics with very little presence from having significant incidence on the determination of classes, like wise Inverse Weighted K-Means.

CONCLUSIONS

K-Means in general is both fast and efficient to extract regions with different colors in images, and the segmentation results are close to human perceptions. For future developments, spatial and texture information will be integrated into pixel features for clustering, and the segmented images will be used for region based image retrieval. The previous result prove that the Inverse Weighted K-means algorithm offers the best image segmentation in the worst scenarios. The K-Means algorithm in general worked as expected. It was fairly simple to implement and the results for image segmentation are impressive. As the results show, the number of partitions which was used in the segmentation has a very little effect on the output when it compares K-Means with its others mentioned versions. By using more partitions, the algorithms also run quickly enough that real-time image segmentation could be done with the K-Means algorithm and the other two versions. The results also indicate the huge sensitivity of the K-Means and it is mentioned versions to noise, while small amount of noise Inverse Weighed K-Means give the best results, but at medium amount all mentioned versions give similar results, while at larger amount K-Means and its mentioned versions have failed.

REFERENCES

Abras GN, Ballarin VL (2005): " A Weighted K-means Algorithm applied to Brain Tissue Classification ", Signal Processing Laboratory, School of Engineering, Universidad Nacional de Mar del Plata, October 2005

Arthur D, Vassilvitskii S (2006). K-means++: "The advantages of careful seeding". In Bay Area Theory Symposium, BATS 06, 2006. http://www.stanford.edu/˜sergeiv/papers/kMeansPP-soda.pdf.

Barbakh W (2007). "The family of inverse exponential k-means algorithms. Computing and Information Systems", 11(1):1–10, February 2007. ISSN 1352-9404.

Barbakh w, Crowe M, Fyfe C (2006).. "A family of novel clustering algorithms. In 7th international conference on intelligent data engineering and automated learning", IDEAL2006, pages 283–290, University of Burgos, Spain, September 2006. Springer. ISSN 0302-9743 ISBN-13 978-3-540-45485-4.

Barbakh W, Ying W, Colin F (2009).. "Non-standard parameter adaptation for exploratory data analysis" (2009), University of the west of Scotland, Scotland ISBN: 978-3-642-04004-7

Bezdek JC (1981). "Pattern Recognition with fuzzy objective function algorithms," New York: Plenum Press, 1981. Krishna K, Ramakrishnan KR, M.A.L. Thathachar (1997). "Vector quantization using genetic K-Means algorithm for image compression," Proceedings of the International Conference on Information, Communications and Signal Processing, ,Vol. 3, 1997, pp. 1585-1587.

Dzung LP, Chenyang X, Jerry LP (2000). "Current Methods in Medical Image Segmentation", Annual Review of Biomedical Engineering, volume 2, pp 315-337.

Gonzalez RC, Woods RE (2002). " Digital Image Processing", Prentice Hall, Upper Saddle River, New Jersey 07458.2002, ISBN: 0-201-18075-8.

Hall LO, Bensaid AM, Clarke LP, Velthuizen RP, Silbiger MS, Bezdek JC (1992). "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," IEEE Trans. On Neural Networks, Vol. 3, No. 5, 1992, pp. 672- 682.

Hartigan J, Wang M (1979). "A k-means clustering algorithm. Applied Statistics", 28:100–108, 1979.

Krishna K, Narasimha MM (1999). "Genetic K-means algorithm," IEEE Trans. on Systems, Man and Cybernetics, Vol. 29,1999, pp. 3.

Linda GS, George CS (2001). "Computer Vision", pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3.

Lloyd SP (1957.: "Least squares quantization in pcm." Technical note, Bell Laboratories, 1957. Published in 1982 in IEEE Transactions on Information Theory 28, 128-137.

MacQueen J (1967). "Some methods for classification and analysis of multivariate observations". Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, pages 281–297, 1967.

Marques, Blanca NP, Pina P (2005). "Pattern Recognition and Image Analysis" Singh, S.; Singh, M.; Apte, C.; Perner, P. Estoril, Portugal, (2005), ISBN: 978-3-540-28833-6.

McAndrew A (2004). "An Introduction to Digital Image Processing with Matlab Notes for SCM2511 Image Processing 1", School of Computer Science and Mathematics, Victoria University of Technology, 2004.

Pena M, Barbakh W, Fyfe C (2007).. "Principal Manifolds for Data Visualization and Dimension Reduction", chapter Topology-Preserving Mappings for Data Visualization, pages 132–152. Springer, 2007. ISBN 978-3-540-73749-0.