Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com

JMCS
Journal of
Mathematics and
Computer Science

Image Segmentation with Improved Distance Measure in SOM and K Means Algorithms

Khosro Jalali[1], Mostafa Heydari[2], Asma Tanavar[3]

[1] Department of Electronic Engineering, Technical and Vocational Schools Mahmoud Abad
[2] Department of Computer Science and Engineering, Shomal University, Amol, Iran
Department of Computer Engineering ,Islamic Azad University ,Sari branch, Sari, Iran
Jalali47@gmail.com
baradaran5@yahoo.com
a_tanavar@yahoo.com

Abstract

This paper explains the task of segmenting image by improved distance measure in SOM and K means algorithms. Image segmentation, divides the image into its constituent regions. It can be said the most prominent features in segmenting is the image brightness for monochrome images and the color components of color images. Over all pixels of image analysis is difficult, Pixels with similar brightness, with the use of image segmentation are grouped together. To achieve higher accuracy of segmentation, we are used fit the soft computing techniques namely Fuzzy algorithms. Image segmentation in many cases (For example, the tumor area to help doctors detect tumor) only be used to assist human visual system. This paper compares segmentation-based methods, visual system and scoring is on him.

1. Introduction

   Partitioning of an image into several segments made up of sets of pixels is called image segmentation. Segmentation plays an important role in any automated image recognition system, because it is at this moment that one extracts the objects of interest, for further processing such as description or recognition. Segmentation of an image is the classification of each image pixel to any one of the image parts. Image segmentation has been the subject of considerable research activity over the last few decades. Many unsupervised algorithms have been developed for segmenting gray scale as well as colour images.

Mukherjee et al. [1] used Fuzzy C means classification algorithm for obtaining the meteorological information from Sodar facsimile records. The fuzzy k-means clustering (FKM) algorithm performs iteratively the partition step and new cluster representative generation step until convergence. The applications of FKM can be founded in reference [2]. [3] used Self-Organizing Map for determining prototypical segments in the images of the 101 object categories database. Chang and Teng [4] applied Self-Organizing Map for medical image segmentation.

At this paper, In Section 2 We will have a notion of image segmentation. In Section 3 clustering techniques discussed. And in Section 4 FKM algorithm has been discussed. About SOM method, we are talking in

Section 5. In Section 6 the proposed method focuses on the change in measure distance in K means and SOM algorithms in image Segmentation. And at the end, we will conclusions in Section 7.

2. Image segmentation

In computer vision, image segmentation is 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 analyze [5].

3. Clustering techniques

Clustering is the task of partitioning the data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. Clustering can also be thought of as a form of data compression, where a large number of samples are converted into a small number of representative prototypes or clusters. Depending on the data and the application, different types of similarity measures may be used to identify classes, where the similarity measure controls how the clusters are formed. Some examples of values that can be used as similarity measures include distance, connectivity, and intensity [6].

3.1. K means clustering

In data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.
The K means algorithm can be summarized in the following four steps:
  i.   *The initial values for the desired K class centers (randomly selected K pixels as centers of classes).*
  ii. *Assign each pixel of the image to the class that is closest to that class. Distance (x,y) pixel to the center of the class i is shown by μ_i the following relationship is obtained:*

$$d_i(x,y) = \left| I(x,y) - \mu_i \right| \tag{1}$$

  iii. *Computing centers classes. The average of all the pixels that belong to the class i the stage before (assignment) have been calculated as the center of the class i assume.*
  iv. *If we can show the average class i in p step (p repeat) is calculated with μ^p and the μ^p vector defined as follows:*

$$\mu^p = [\, \mu_1^p \;\; \mu_2^p \;\; \cdots \;\; \mu_k^p \,] \tag{2}$$

Then the averages (center) obtained in p+1 step amount is not much difference in p step, the algorithm is stopped and otherwise, return to step ii.

3.2. Fuzzy clustering

In non-fuzzy or hard clustering, data is divided into crisp clusters, where each data point belongs to exactly one cluster. In fuzzy clustering, the data points can belong to more than one cluster [6, 7], and associated with each of the points are membership grades that indicate the degree to which the data points belong to the different clusters. Fuzzy clustering belongs to the group of soft computing techniques (which include neural nets, fuzzy systems, and genetic algorithms). In real applications there is very often no sharp boundary between clusters so that fuzzy clustering is often better suited for the data. Membership degrees between zero and one are used in fuzzy clustering [6, 8]

4. Fuzzy K means

The importance and potential of this approach becomes clear that when we get to class centers. In the K means method, all pixels of the same class share the same calculation but out of the Fuzzy K means methods, the pixels used to determine and calculate the average value of the shares are not identical. FKM algorithm is exactly the same K means Algorithm process with this difference that in this algorithm, the membership probability is calculated. If ui be Register possibility a pixel to i class, obviously must [9, 10]:

$$\sum_{i=1}^{K} u_i = 1 \tag{3}$$

Because the pixels to be considered belongs one of the classes 1, 2, ..., K.

$u_i(x, y)$ representative of the membership of (x, y) pixel to class i and μ_i is average.

Therefore, the fuzzy k-means clustering algorithm is now presented as follows:

i. *Initialized and desired to join the possibility of all the pixels of the image. At this stage we need to consider for each (x, y) pixel, Registration possibility for the K class $(0 < u_i < 1)$.*

ii. *Assign each pixel of the image to class (central) that they have more than register possibility.*

iii. *Calculating the class centers by using of register possibility all the pixel and class them*

iv. *Calculation Register possibility of all image pixels in each K class.*

v. *If $u_i^{(p)}(x, y)$ be register possibility pixels (x, y) in i class Where it obtained by p iteration, if*

$$\left\| U^{(p+1)} - U^{(p)} \right\| = \sqrt{\sum_{i=1}^{K} \sum_{(x,y) \in C_i} (u_i^{(p+1)}(x, y) - u_i^{(p)}(x, y))^2} < \varepsilon \tag{4}$$

Algorithm stops and otherwise, it is repeated from step ii.

5. Image segmentation with using SOM neural network

SOM, which was originally introduced for the visual display of one- and two-dimensional data sets, has the same functional ideas as many other clustering algorithms. The SOM neural network is a topology-preserving map in which adjacent vectors in (^{n} are mapped to adjacent (or identical) cells in the array, and adjacent cells in the array have similar position vectors in (^{n}. The purpose of the self-organization process, as described by Kohonen [11, 12], is to find values for the position vectors such that the resultant mapping is topology- and distribution-preserving1. Dekker [13] presented the use of SOM network for quantization of colour graphics images. By adjusting a quality factor, the network is shown experimentally to produce images of much greater quality with longer running times, or slightly better quality with shorter running times than existing methods.

6. Distance measure improvements in SOM and K means Algorithms

To achieve more accurate segmentation, particularly in noisy images try:

1. *Reduce the impact of noise on the corrected weighting coefficients.*

2. *Each pixel (each input) based on the membership of the output neuron consider to correct weighting coefficients.*

If we can to define a new measure to calculate the distance, paragraphs 1 and 2 can be realized largely. By defining a new distance measure, we can also be improved segmentation using the K means Algorithms. In this method the same SOM method, only the illumination of pixels is used for image segmentation and it should be promoted. So that the illumination of pixels,

1. *The effect of noise on the measured distance will be reduced.*

2. *All pixels have the same amount of computing centres.*

with look at the K means and SOM algorithms we understand this methods are based on pixel illumination. If we forge the above paragraphs, it is necessary in addition to illumination, we use the additional features. Such as mean or variance of their neighbouring pixels. If we use the characteristic variation, to be avoided blurring especially in boundaries. The characteristics of medical images, such as images of the tumour, the position is very important.

With increasing the parameters and elements of decisions, no longer be the relations (1) and also measure the distancecan in SOM algorithm be used. This mode can be used to define the Euclidean distance.

According to the relations (1) and also measure the distance in SOM algorithm will be as follows:

$$d_i(x, y) = \|I(x, y) - \mu_i\|$$
$$= \sqrt{(I_1(x, y) - \mu_{1,i})^2 + (I_2(x, y) - \mu_{2,i})^2} \tag{5}$$

$I_1(x, y)$ and $I_2(x, y)$ are respectively the first and second eigenvalues. $\mu_{1,i}(x, y)$ and $\mu_{2,i}(x, y)$ are respectively the first and second the mean characteristics in the (x, y) points. We can calculate the mean and variance, respectively of the (1a) and (1b) filters. The filters are:

$$W_5$$
$$= (W_1 + W_2 + W_3 + W_4 + W_5 + W_6 + W_7 + W_8 + W_9)/9 \tag{6}$$

$$W_5$$
$$= 8W_5 - W_1 - W_2 - W_3 - W_4 - W_6 - W_7 - W_8 - W_9 \tag{7}$$

In equation (5), Norm of two vectors is used to calculate distances. However, this equation can be expanded using p norm at It can be defined the types of distances (depending on application).

$$d_i(x, y) = \|I(x, y) - \mu_i\|_p$$
$$= ((I_1(x, y) - \mu_{1,i})^p + (I_2(x, y) - \mu_{2,i})^p)^{1/p} \tag{8}$$

We can improved this relationship by adding a new parameter named weights (weighted coefficient average and variance of weighted coefficient):

$$d_i(x, y) = \|I(x, y) - \mu_i\|_p$$
$$= (W_1(I_1(x, y) - \mu_{1,i})^p + W_2(I_2(x, y) - \mu_{2,i})^p)^{1/p} \tag{9}$$

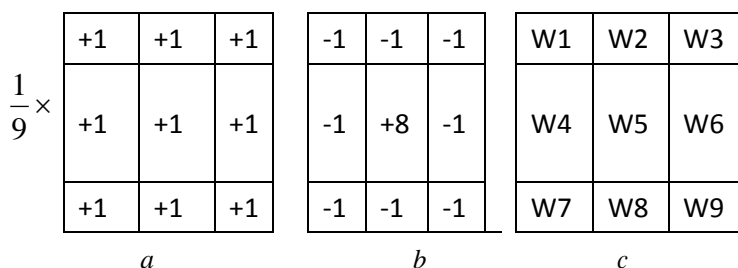|										
$\frac{1}{9} \times$	+1	+1	+1		-1	-1	-1		W1	W2	W3
	+1	+1	+1		-1	+8	-1		W4	W5	W6
	+1	+1	+1		-1	-1	-1		W7	W8	W9
		a				*b*				*c*	

Fig. 1: mean filter (a), standard deviation filter (b) and named pixels (c).

P - Norm Weighted distance names have been considered for the proposed method. Figure 2 shows the dependence the method of noise images.

(a)   (b)

(c)   (d)

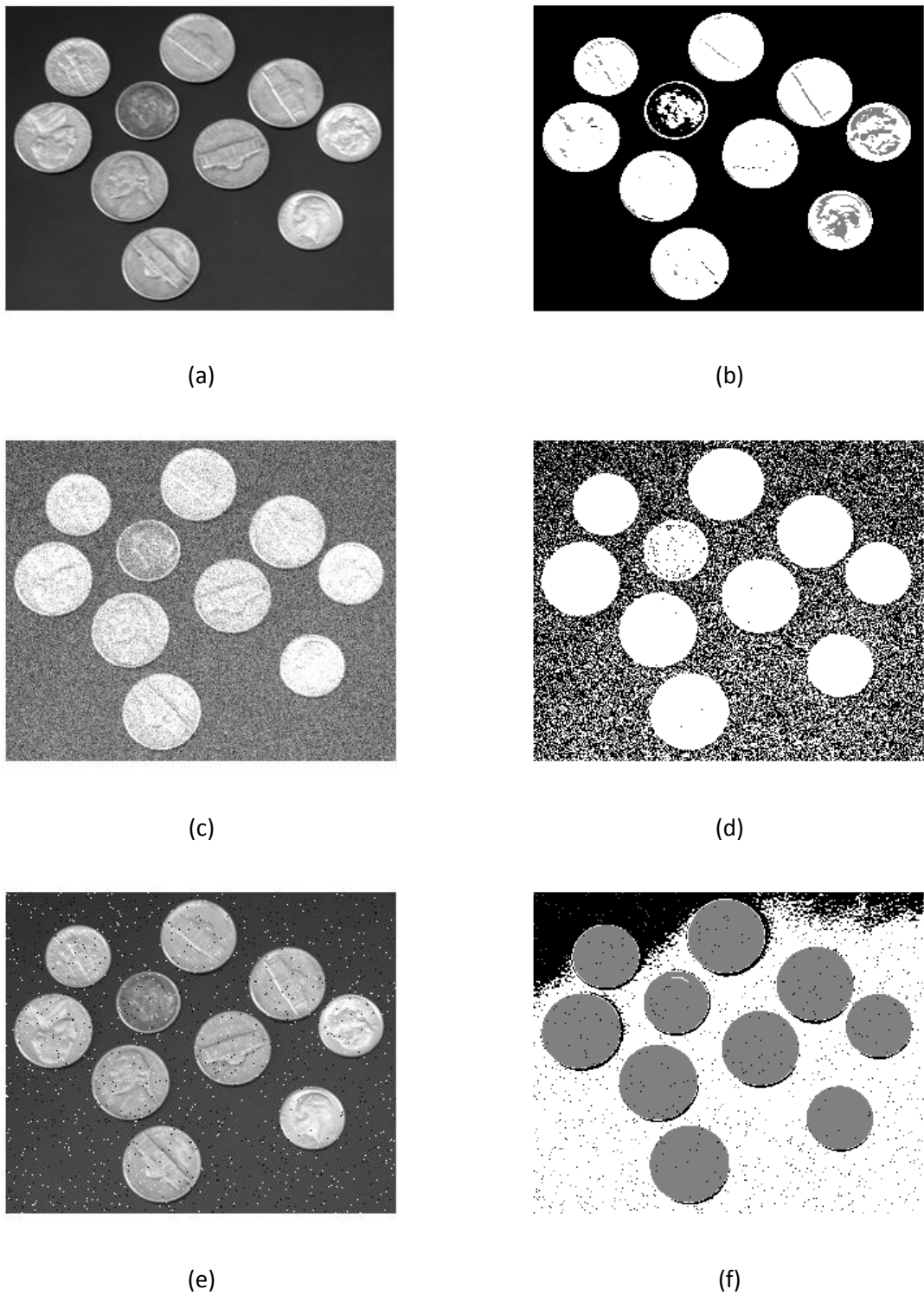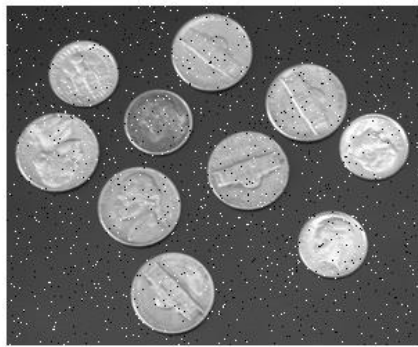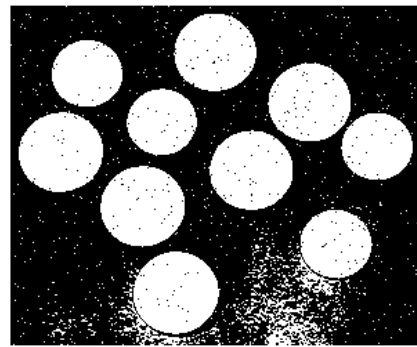(e)   (f)

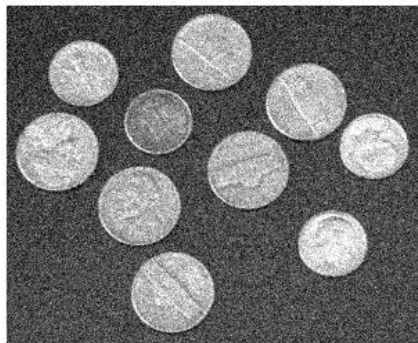Fig. 2: Dependent the SOM method on noise of image.

Similar to Figure 2, FKM method has been evaluated in Figure 3 per Noise images. Can be observed this method can not be useful for images noise (high noise). The reason for this can be sought in relation (5).
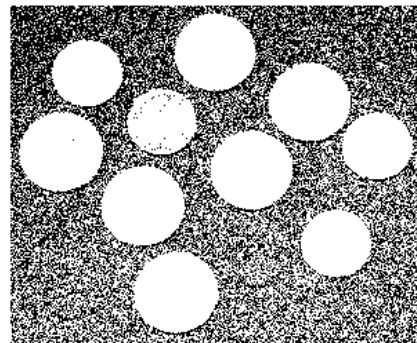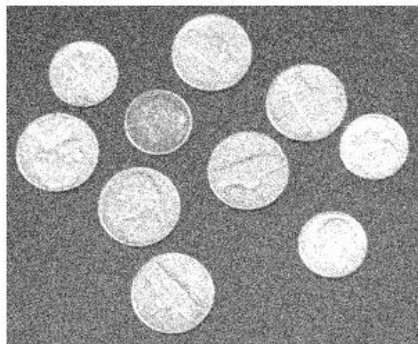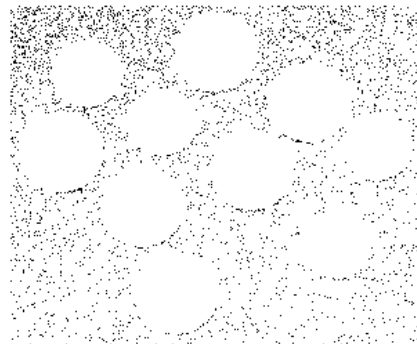
Fig. 3: Dependent the FKM method on noise of image. Image noise with Salt and pepper noise (a), Gaussian noise with variance 0.2 (c), Gaussian noise with variance 0.02 (e).

Figures 4 to 6, the K means and SOM methods are compared. Per new definition of distance, with their previous form. As the form is returned, the proposed method can improve image noise far better than previous methods. Meanwhile, in the proposed method edges of the image are preserved to a large extent and this would be very useful in medical images.

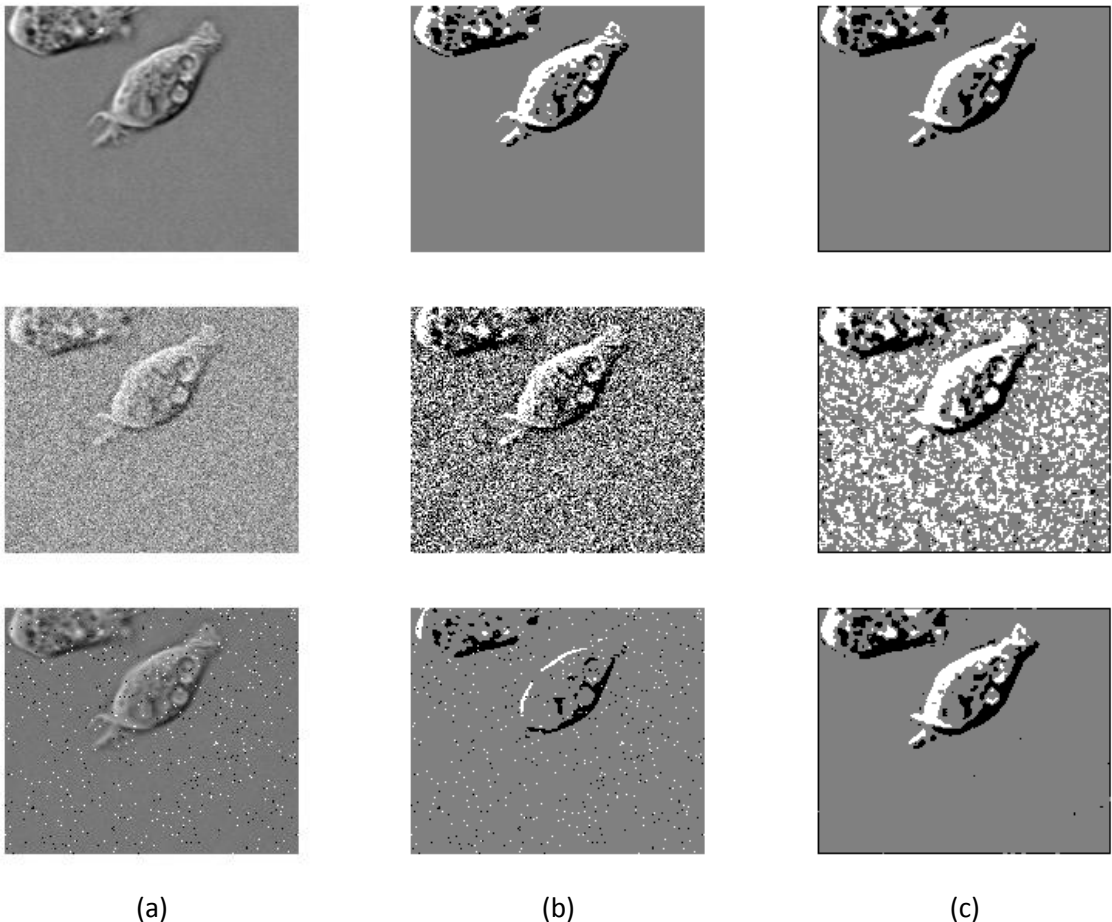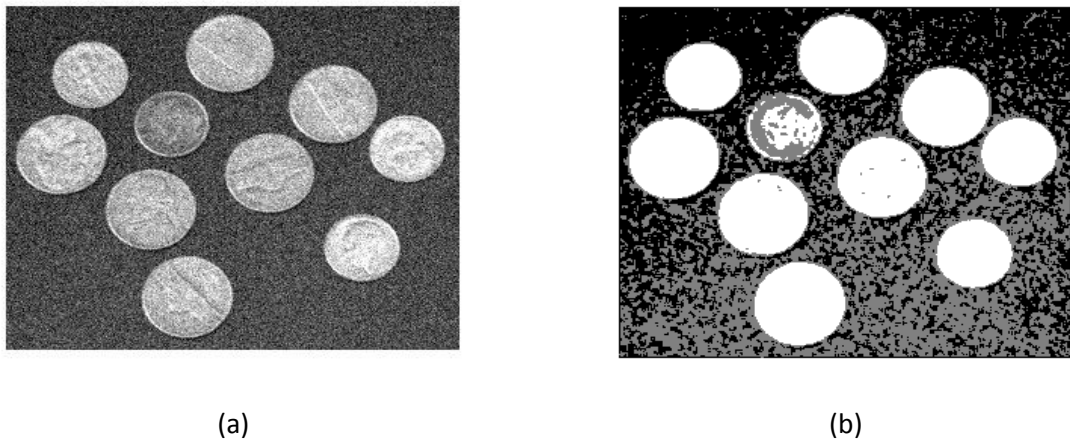(a)   (b)   (c)
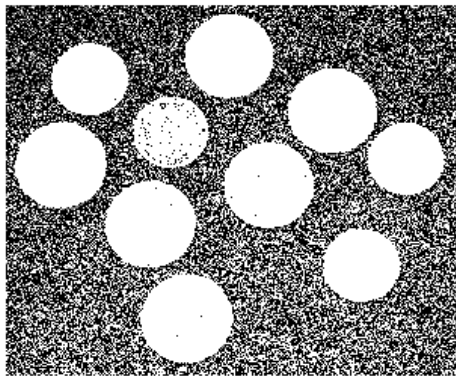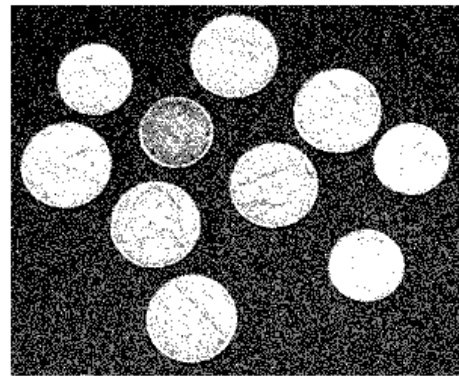
Fig. 4: A clean image, Gaussian noise, salt and pepper noise (a), segmentation SOM methods (b), and proposed segmentation P - norm weighted SOM method (c) with the mean characteristics of illumination and W1 = W2 = 1.

(a)   (b)

(c)   (d)

Fig. 5: Comparison proposed method P - norm weigthed K means (b) with K means method (c) and SOM (d).

W1=1 , W2=0.25

W1=1 , W2=0.5

W1=1 , W2=1

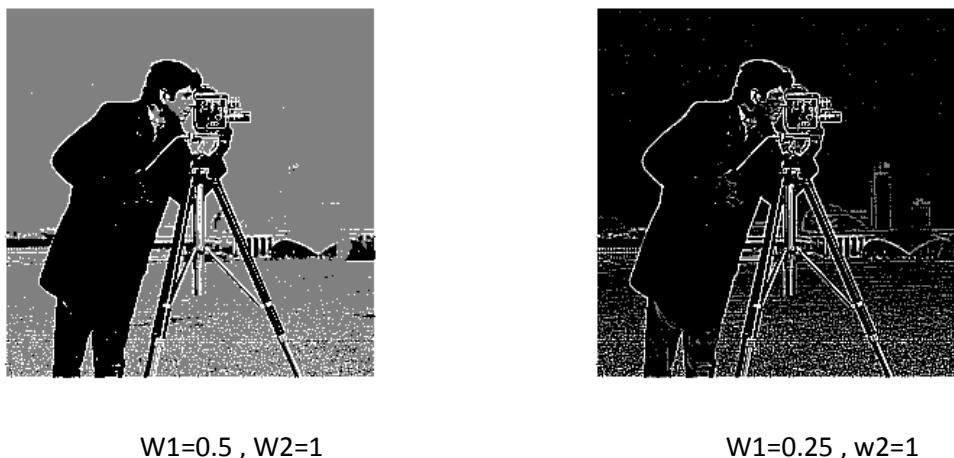W1=0.5 , W2=1   W1=0.25 , w2=1

Fig. 6: Comparison of the proposed method P - norm weigthed SOM by different W.

7. Result

   Proposed K means and SOM methods given that did not use the concept of grade of membership in the classification of pixels, they showed poor performance on noisy images. Using the fuzzy degree of membership has come down to pixel noise in a classroom and thereby Their intervention comes down in classification. As well, We were able to extract

images from the original images more accurately whit a new definition of distance measure. Which can be seen easily by the recovery to the eye especially in noisy images (for example, see section 6). Can be seen in Figure 7 block diagram of the improved method presented in Section 6.
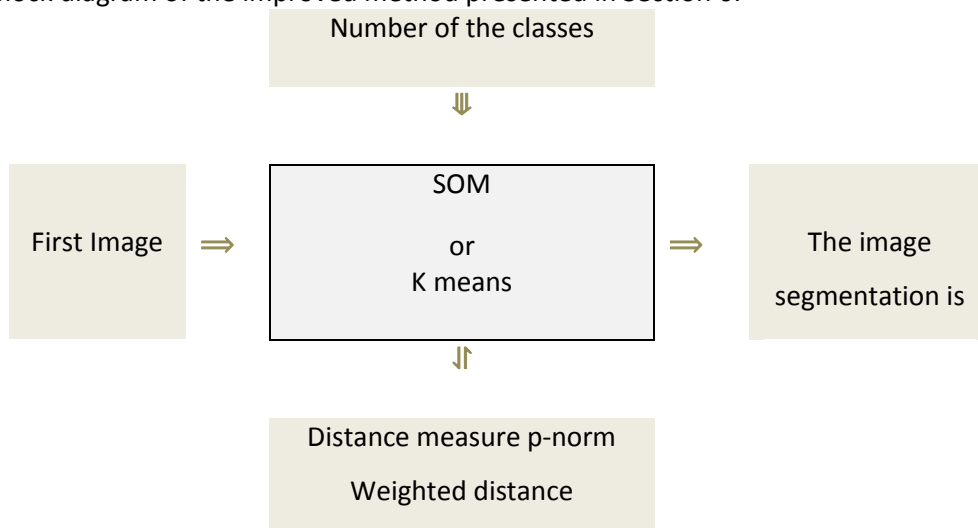

Fig. 7: Block diagram of the proposed method
p-norm Weighted distance.

References

[1] D.P. Mukherjee, P. Pal, J. Das, "Sodar image segmentation by fuzzy c-means," Signal Processing 54 (November (3)) (1996) 295–301.

[2]   C. Doring, M. J. Lesot, and R. Kruse, "Data analysis with fuzzy clustering methods, Computational Statistics and Data Analysis," 2006, pp. 192-214.

[3]   J. Laaksonen, V. Viitaniemi, M. Koskela, "Application of Self-Organizing Maps and automatic image segmentation to 101 object categories database," in: Fourth International Workshop on Content-Based Multimedia Indexing, Riga, Latvia, June 21–23, 2005.

[4]   P.-L. Chang, W.-G. Teng, "Exploiting the Self-Organizing Map for medical image segmentation," in: Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), 2007, pp. 281–288.

[5]   G. Linda Shapiro and C. George Stockman (2001): "Computer Vision," pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3.

[6]   B. Sowmya, B. Sheela Rani, "Colour image segmentation using fuzzy clustering techniques and competitive neural network," in: Applied Soft Computing, 2011, pp 3170-3178.

[7]   S. Chen, D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," IEEE Transactions on Systems, Man and Cybernetics 34 (August (4)) (2004) 1907–1916.

[8]   M.R. Rezaee, P.M.J. Van Der Zwet, B.P.E. Lelieveldt, R.J. Van Der Geest, J.H.C. Reiber, "A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering."

[9]   B. Sowmya, B. Sheela Rani, (2011). "Color image segmentation using fuzzy clustering techniques and competitive neural network. Applied Soft Computing," 11(3), 3170–3178. doi:10.1016/j.asoc.2010.12.019.

[10] Francisco de A.T. de Carvalho, "Fuzzy c-means clustering methods for symbolic interval data," Pattern Recognition Letters, Volume 28, Issue 4, 1 March 2007, Pages 423-437.

[11] T. Kohonen, "Self-Organizing Maps," second ed. Springer Series in Information Sciences, 1997.

[12] T. Kohonen, "The self-organizing map," IEEE Proceedings 78 (1990). 1464–1477.

[13] A.H. Dekker, "Kohonen neural networks for optimal color quantization," Network: Computation in Neural Systems 5 (1994) 351–367.