"A Implementation of K Means Clustering and Segmentation Based Image Impulse Noise Reductiontechnique"

Vishnu Kant Chaurasia, Electrical Department, Jabalpur Engineering College Jabalpur, India
Guided By:- Mr. Hemant Amhia, Electrical Department, Jabalpur Engineering College Jabalpur, India

Abstract- The main object of our research work is to look at the execution of different division systems for shading pictures. Division by K-implies bunching and thresholding methods are thought about for their execution in division of shading pictures. Division of a picture involves the division or partition of the picture into areas of comparative property. The most essential characteristic for division of a picture is its luminance adequacy for a monochrome picture and shading segments for a shading picture. This paper proposes a shading based division strategy that utilizes K-implies grouping system. The k-implies calculation is an iterative method used to parcel a picture into k groups. The standard K-Means calculation produces exact division results just when connected to pictures characterized by homogenous areas as for surface and shading since no neighborhood requirements are connected to force spatial congruity. At first, the pixels are grouped dependent on their shading and spatial highlights, where the bunching procedure is cultivated.

Keywords- Image segmentation, k-means clustering, thresholding, MSR, PSNR

I. INTRODUCTION

The pictures ruined by motivation clamor are frequently happened by and by. This sort of commotion may show up in computerized pictures as a result of channel decoder harms, coloring down of flag in correspondence joins. Applying of great middle channel for evacuation of such kind of commotion gives generally great outcomes, which are appeared in reestablishing of brilliance drops, objects edges and neighborhood tops in clamor tainted pictures [4]. Yet, investigation of various sources committed to middle separating appears, that the exemplary middle channel has a lot of weaknesses - flag debilitating (item's counters and edges are obscured in picture); - influencing to non-tainted ("great") picture pixels. Diverse adjustments of middle channel have been proposed to wipe out these impediments of middle separating. Presently the exchanging plan pulls in a high enthusiasm of numerous inquiries about [5]. This methodology demonstrates its productivity for salt-and-pepper drive commotion expulsion from advanced pictures. The exchanging plan approach implies part of commotion expulsion method into two primary stages [6]:

1. To begin with identification of clamor defiled pixels of computerized picture.

2. Separating of commotion motivations which have been recognized in first phase of handling utilizing data about accumulated picture properties.

Depicted plan with motivation locator is utilized in many propelled alterations of middle channel. For instance, in [5] creators reason to utilize a locator dependent on K mean and division technique. Furthermore, in [6] a versatile channel dependent on K mean and division [7] had offered for reclamation of advanced pictures. On the off chance that the thickness of salt-and-pepper commotion extremely high (over 30%), a calculation depicted in [1, 2] is appeared productive. This calculation additionally created utilizing exchanging plan. It utilizes adequately complex iterative strategies both at clamor recognizing stage and amid sifting of debased picture. This improves rebuilding of pictures even in all respects incredibly defiled by motivation commotion. The new calculation has substantial execution in expulsion of motivation commotion from advanced pictures while having a generally low unpredictability.

II. NOISE MODELS

The foremost wellspring of commotion in advanced pictures emerges amid picture obtaining (digitization) and transmission. The execution of imaging sensors is influenced by an assortment of components, for example, natural conditions amid picture securing and by the nature of the detecting components themselves. For example, in gaining pictures with a CCD camera, light dimensions and sensor temperature are central point influencing the measure of commotion in the subsequent picture. Pictures are debased amid transmission chiefly because of obstruction in the channel utilized for transmission

A. Gaussian Noise

The PDF of a Gaussian random variable, z is given by

$$p(z) = \frac{1}{2\pi\sigma} e^{-(z-\mu)^2/2\sigma^2}$$

Where z speaks to dim dimension, μ is the mean of normal estimation of z, and σ is its standard deviation. The standard deviation squared, σ 2 is known as the change of z. Due to its numerical tractability in both the spatial and recurrence areas, Gaussian (likewise called typical) commotion models are oftentimes utilized by and by. Truth be told, this tractability is convenient to the point that it frequently results in Gaussian models being utilized in circumstances in which they are possibly appropriate, best case scenario.

III. FILTERS

A. Arithmetic mean filters:

This is the least complex of the mean channels. Let Sxy speaks to the arrangement of directions in a rectangular sub picture window of size m × n focused at point (x,y). The number-crunching mean sifting process figures the normal estimation of the undermined picture g(x , y)in the territory characterized by Sxy. The estimation of the reestablished picture f anytime (x,y) is essentially the number juggling mean processed utilizing the pixels in the area characterized by Sxy .at the end of the day,

$$\hat{f}(x,y) = \frac{1}{mn} \sum_{(s,t)\in S_{xy}} g(s,t)$$

This task can be actualized utilizing a convolution cover in which all coefficients have esteem 1/mn. A mean channel basically smoothes nearby varieties in a picture. Clamor is decreased because of obscuring [5].

B. Median filter

The best known request measurements channel is the middle channel, which replaces the estimation of a pixel by the middle of the dim dimensions in the area of that pixel

$$\hat{f}(x,y) = \underset{(s,t)\in S_{xy}}{\text{median}} \{g(s,t)\}$$

The first estimation of the pixel is incorporated into the calculation of the middle. Middle channels are very famous on the grounds that, for particular kinds of irregular commotion they give phenomenal clamor decrease abilities, with significantly less obscuring than straight smoothing channels of comparable size.

IV. PERFORMANCE MEASUREMENT

Exhibitions of calculations are estimated by computing PNSR (Peak flag to Noise Ratio) and SNRI (Signal to Noise Ratio Improvement).

Pinnacle flag to Noise Ratio (PNSR): It is estimated in decibel (dB) and for dim scale picture it is characterized as:

$$\text{MSE} = \frac{\sum_i \sum_j (Xij - Rij)^2}{(M \times N)}$$

$$\text{PSNR} = 10\log10 \times \frac{255x255}{\text{MSE}}$$

Where X - Original Image.

R - Restored Image

M x N - Size of Image.

MAE - Mean Absolute Error.

MSE - Mean Square Error.

PSNR - Peak Signal to Noise Ratio.

The higher the PNSR in the restored image the better is its quality.

Signal to Noise Ratio Improvement (SNR):

SNRI in dB is characterized as the distinction between the flag to clamor proportion (SNR) of the reestablished picture in dB and SNR of loud picture in dB.for example

SNRI (dB) = SNR of the reestablished picture - SNR of loud picture.

Where,

$$\text{SNR of restored image} = 10\log10 \times \frac{\sum_i \sum_j Xij^2}{\sum_i \sum_j (Xij - Rij)^2}$$

$$\text{SNR of Noisy image} = 10\log10 \times \frac{\sum_i \sum_j Xij^2}{\sum_i \sum_j (Xij - Nij)^2}$$

Where, Nij is Noisy Image Pixel. The higher value of SNRI reflects the better visual and restoration performance.

V. PROPOSED METHODOLOGY

Picture division is the characterization of a picture into various gatherings. Numerous inquires about have been done in the territory of picture division utilizing grouping. There are distinctive techniques and a standout amongst the most well known strategies is k-implies bunching calculation. K-implies bunching calculation is an unsupervised calculation and it is utilized to section the intrigue region from the foundation. In any case, before applying K-implies calculation, first halfway extending upgrade is connected to the picture to improve the nature of the picture. Subtractive grouping strategy is information bunching technique where it produces the centroid dependent on the potential estimation of the information focuses. So subtractive group is utilized to create the underlying focuses and these focuses are utilized in k-implies calculation for the division of picture. At that point at long last average channel is connected to the sectioned picture to expel any undesirable locale from the picture [8].

A quantitative methodology is measure sure highlights of the items, state level of milk and others, and items with high level of milk would be assembled together. All in all, we have n information indicates xi,i=1...n that have be apportioned in k bunches. The objective is to allocate a bunch to every datum point. K-implies is a grouping technique that means to discover the positions μi,i=1...k of the bunches that limit the separation from the information focuses to the group. K-implies grouping illuminates

$$\arg\min_{\mathbf{c}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in c_i} d(\mathbf{x}, \mu_i) = \arg\min_{\mathbf{c}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in c_i} \|\mathbf{x} - \mu_i\|_2^2$$

Where

Ci is the arrangement of focuses that have a place with bunch I. The K-implies bunching utilizes the square of the Euclidean separation d(x,μi)=‖x−μi‖22. This issue isn't inconsequential (in truth it is NP-hard), so the K-implies calculation just plans to locate the worldwide least, conceivably stalling out in an alternate arrangement.

The M-step is processing the centroid of each group. The following is a breakdown of how we can unravel it scientifically (don't hesitate to skip it).

The target work is:

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} \|x^i - \mu_k\|^2$$

Wherewik=1 for data point xi if it belongs to cluster *k*; otherwise, wik=0. Also, μk is the centroid of xi's cluster.

Mathematical Formulation for K-means Algorithm:

D= {**x1,x2,…,xi,…,xm**} à data set of m records

xi= (xi1,xi2,…,xin) à each record is an n-dimensional vector

$$C_j = \text{Cluster}(X_i) = \arg_j \min \| X_i - \mu_j \|^2$$

$$\text{Distortion} = \sum_{i=1}^{m} (x_i - c_i)^2 = \sum_{j=1}^{k} \sum_{i \in \text{OwnedBy}(\mu_j)} (X_i - \mu_j)^2$$

(within cluster sum of squares)

Finding Cluster Centers that Minimize Distortion:

Solution can be found by setting the partial derivative of Distortion w.r.t. each cluster center to zero.

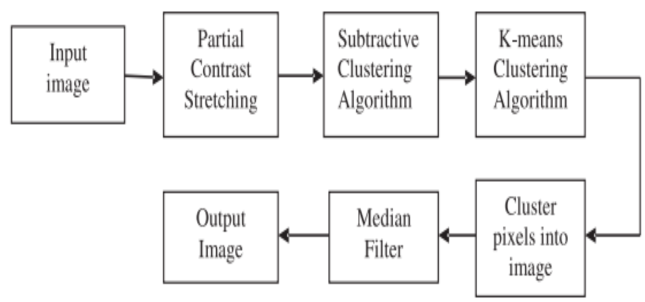
$$\frac{\partial\, \textbf{\textit{Distortion}}}{\partial\, \mu_j} = \frac{\partial}{\partial\, \mu_j} \sum_{i \in \text{OwnedBy}(\mu_j)} (x_i - \mu_j)^2 = -2 \sum_{i \in \text{OwnedBy}(\mu_j)} (x_i - \mu_j) = 0 \ \textbf{\textit{(for minimum)}}$$

$$\Rightarrow \mu_j = \frac{1}{|\text{OwnedBy}(\mu_j)|} \sum_{i \in \text{OwnedBy}(\mu_j)} x_i$$

For any k groups, the estimation of k ought to be with the end goal that regardless of whether we increment the estimation of k from after a few dimensions of bunching the contortion stays consistent. The accomplished point is known as the "Elbow".

Our proposed Algorithm Steps:

1.Load the picture to be portioned.
2. Apply fractional difference extending. Introduce number of group, k.
3. Use condition (1) to ascertain the potential for each pixel estimation of the picture.
4. Discover most extreme potential in stage 3 and set that point be first focus bunch and its comparing potential as greatest potential.
5. Use condition (2) to refresh the potential estimation of other residual pixels dependent on the principal group focus.
6. Again locate the most extreme potential in the stage 4 and let that point be second point.
7. Proceed with the procedure until it finds the k number of group.
8. Utilized k focus as beginning focus in the k-implies bunching calculation.
9. Locate the Euclidean separation of every centroid from each pixel of the picture utilizing the connection (3).
10. Allocate the pixel with least separation as for centroid to its particular group of the centroid. 11. Recalculate the new focus area by utilizing the condition (4).
12. Rehash the means 10– 12, until it fulfills the resistance or mistake esteem.
13. Reshape the bunch into picture.
14. Middle channel is connected to the divided picture to evacuate any undesirable commotion or district.



VI. IMAGE SEGMENTATION TECHNIQUE

A. Thresholding Method

Thresholding techniques are the easiest strategies for picture division. These strategies isolate the picture pixels as for their force level. These techniques are utilized over pictures having lighter items than foundation. The determination of these techniques can be manual or programmed for example can be founded on earlier learning or data of picture highlights. There are essentially three kinds of Thresholding [6] [2]:

1) Global Thresholding:
This is finished by utilizing any suitable edge esteem/T. This estimation of T will be steady for entire picture q(x, y). Based on T the yield picture can be acquired from unique picture p(x,y)as:-

$$q(x,y) = \begin{Bmatrix} 1, if\ p(x,y) > T \\ 0, if\ p(x,y) \leq T \end{Bmatrix}$$

2) Variable:-
Thresholding: In this sort of thresholding, the estimation of T can change over the picture. This can additionally be of two kinds:

• Local Threshold: In this the estimation of T relies on the area of x and y.

• Adaptive Threshold: The estimation of T is a component of x and y.

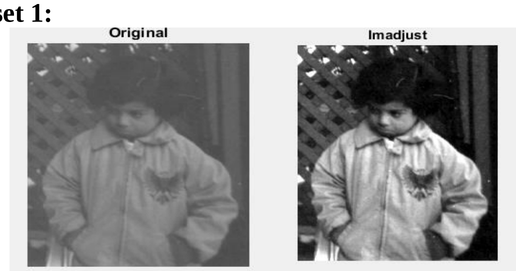
3) Multiple Thresholding: In this kind of thresholding, there are various limit esteems like T0 and T1. By utilizing these yield picture can be figured as:

$$q(x,y) = \begin{Bmatrix} m, if\ p(x,y) > T1 \\ n, if\ p(x,y) \leq T1 \\ o, if\ p(x,y) \leq T0 \end{Bmatrix}$$

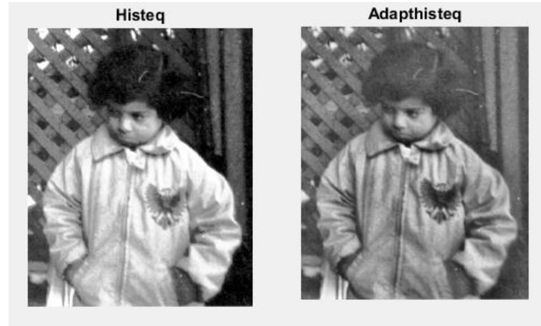
B. Clustering Based Segmentation Method:
The bunching based methods are the systems, which portion the picture into groups having pixels with comparable qualities. Information bunching is the technique that partitions the information components into groups to such an extent that components in same bunch are more like each other than others. There are two fundamental classifications of bunching techniques: Hierarchical strategy and Partition based technique. The various leveled strategies depend on the idea of trees. In this the base of the tree speaks to the entire database and the inside hubs speak to the groups. On the opposite side the segment based strategies use enhancement techniques iteratively to limit a goal work. In the middle of these two techniques there are different calculations to discover bunches [9-10].

VII. RESULT AND SIMULATION

Data set 1:

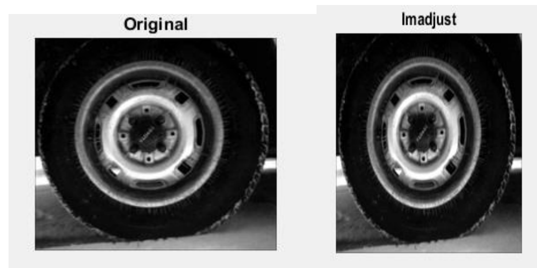


(1)(2)
Fig.1: (1) Original image, (2) Filtered Image for adjustment.

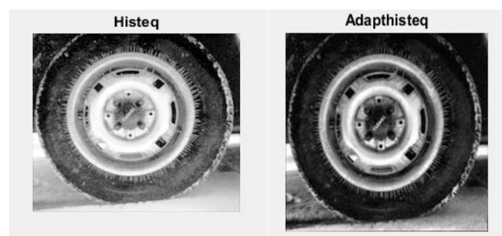


(3) (4)
Fig.2: (3) Histogram image, (4) Average image Filtered Image for adjustment.

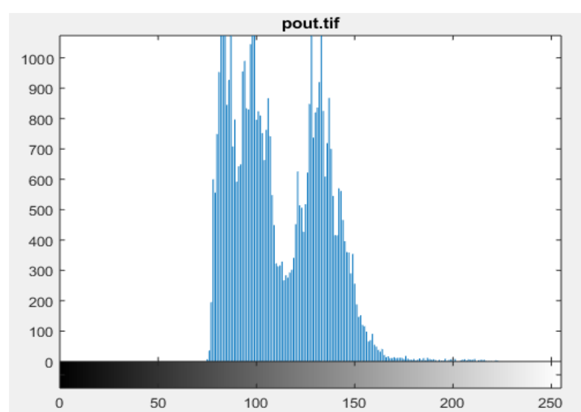
Data sets 2



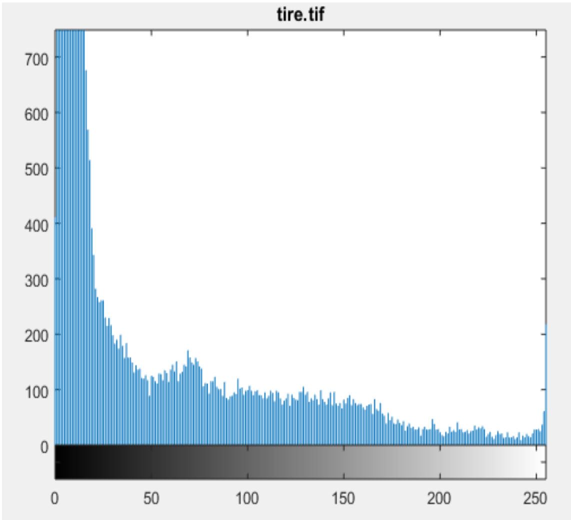
(1) (2)
Fig.3: (1) Original image,(2) Filtered Image for adjustment.



(3) (4)
Fig.4: (3) Histogram image, (4) Average image Filtered Image for adjustment.



(5)



(6)

Fig.5and6: (5-6) Histogram of two data sets.

Sl.no.	Image for analysis	Gaussian Filter	Median filter	Denoising algorithm
1	Image 1(a)	18.168	19.554	30.609
2	Image 2(a)	21.522	27.677	33.16
3	Image 3(a)	26.571	23.16	35.44
4	Image 4(a)	29.028	27.231	42.044
5	Image 5(a)	24.786	36.239	37.62
6	Image 6(a)	20.058	25.019	34.63
7	Image 7(a)	23.608	33.405	36.036
8	Image 8(a)	19.898	26.118	34.448
9	Image 9(a)	24.031	27.822	35.327
10	Image 10(a)	17.653	20.861	30.9

Fig.7: Base paper result.

Our Proposed method result:

Image of analysis	**[16]**	**MSE (Without filter)**	**PSNR**	**(MSE with filter)**
Data set 1	42	475.1254	213	656.37
Data set 2	35.1	396.5684	81.2	2.5113

Show by in table MSE price is reduced in our planned algorithms. From observation we tend to knew that the thresholding technique offers associate degree output of 2 segments. However in K-means technique we will get numerous segments in keeping with our cluster size. Performance will increase in keeping with cluster sizes. A lot ofis that the cluster size a lot of is that the accuracy proportion.

VIII. FUTURE WORK

The development channel has demonstrated that it is proficient for arbitrary esteemed motivation commotion in light of the fact that scientifically clamor isn't uniform over the direct in any perspective. We have used the idea of most extreme and least limit to recognize the two edges and uproarious piece of picture. It delivers great PSNR and diminished MSE for very adulterated pictures, particularly for over half clamor thickness. This technique has the accompanying focal points: The principle preferred standpoint of our strategy that is two edges utilized and the edge esteems can mindfully refresh as indicated by the commotion thickness of sifting window.

IX. CONCLUSION

A similar investigation of two division procedures has been performed in this examination. The K-implies bunching and thresholding systems were picked for division. Utilizing these two strategies, the execution for various pictures were portioned by utilizing the parameters like MSE, PSNR, and SNR. From perceptions appeared in this paper, one can infer that the thresholding method gives a yield of two portions. Be that as it may, in k-implies systems, yield is of different portions as per group estimate. Execution improves as per group sizes. More is the group measure more is the exactness rate. MSR and PSNR are utilized to gauge the nature of remaking. PSNR estimation of four pictures in k-implies grouping is higher than thresholding and MSE esteem is lower.

X. REFERENCES

[1]. Muthukannan. K, Merlin Moses. M," Color Image Segmentation using K-means Clustering and Optimal Fuzzy C-Means Clustering", Proceedings of the International Conference on Communication and Computational Intelligence – 2010,Kongu Engineering College, Perundurai, Erode, T.N.,India.27 – 29 December,2010.pp.229-234.

[2]. Ms.ChinkiChandhok, Mrs.SoniChaturvedi, Dr.A.AKhurshid," An Approach to Image Segmentation using K means Clustering Algorithm", International Journal of Information Technology (IJIT), Volume – 1, Issue – 1, August 2012.PP.11-17.

[3]. Digital image processing 3rd Edition (DIP/3e), by Gonzalez and Woods

[4]. F. Destrempes, J.-F. Angers, and M. Mignotte, "Fusion of hidden Markov random field models and its Bayesian estimation," IEEE Trans. Image Process., vol. 15, no. 10,pp. 2920–2935, Oct. 2006.

[5]. J.A Hartigan "Clustering Algorithms", New York Wiley 1975.

[6]. S. P. Lloyd, ―"Least squares quantization in PCM," IEEE Trans. Inf.Theory, vol. IT-28, no. 2, pp. 129–136, Mar.1982.

[7]. J. Besag, "On the statistical analysis of dirty pictures,"J. Roy. Statist.Soc. B, vol. 48, pp. 259–302, 1986.

[8]. S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image

segmentation," IEEE Trans. Pattern Anal. Mach.Intell., vol. 18, no. 9, pp. 884–900, Sep.1996.

[9]. S.MaryPraveena, Dr.IlaVennila," Optimization Fusion Approach for Image Segmentation Using K-Means Algorithm", International Journal of Computer Applications (0975 – 8887)Volume 2 – No.7, June 2010

[10]. Seber, G. A. F. Multivariate Observations. Hoboken, NJ: John Wiley & Sons, Inc., 1984.

[11]. E. Abreu and S. Mitra, ―A signal-dependent rank ordered mean (SDROM) filter-a new approach for removal of impulses from highly corrupted images,‖ in ICASSP, vol. 4, pp. 2371 –2374, 1995 IEEE.

[12]. T.-C. Lin, P.-T. Yu, , A new adaptive center weighted median filter for suppressing impulsive noise in images, Information Sciences 177 (2007)1073–1087

[13]. James C. Church, Yixin Chen, and Stephen V. Rice A Spatial Median Filter for Noise Removal in Digital Images, Southeastcon, 2008

[14]. WANG Chang-you, YANG Fu-ping, GONG Hui, A new kind of adaptive weighted median filter algorithm, 2010 I.J.C on Computer Application and System Modeling (ICCASM 2010).

[15]. Zhu Youlian, Huang Cheng, An Improved Median Filtering Algorithm Combined with Average Filtering, IEEE, 2011 Third International Conference on Measuring Technology and Mechatronics Automation, 6-7 Jan. 2011, pp.420-423

[16]. Madhura J, "An Effective Hybrid Filter for the Removal of Gaussian-Impulsive Noise in Computed Tomography Images", 978-1-5090-6367-3/17/$31.00 ©2017 IEEE