International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 10 Ver IV || Oct 2018 || PP 37-43

Experimental Study of Image Segmentation Using K-Means Clustering Algorithm with Image Quality Metrics

Shameem Fatima[1], Dr. M. Seshashayee[2]
[1]Research Scholar, Department of Computer Science, GITAM University, Visakhapatnam, India
[2]Assistant Professor, Department of Computer Science, GITAM University, Visakhapatnam, India
Corresponding Author: Shameem Fatima

Abstract: *Image Segmentation has been considered as one of the significant tool for image analysis. The main objective of image analysis is to understand the component of the image and extract useful information using image features. Over the past two decades, there have been various studies on a variety of well-known probabilistic approach for image segmentation. In this paper we have describe some probabilistic approach towards image segmentation, along with an experimental study of image segmentation using K-Means clustering Algorithm. Initialization of the cluster k is done by obtaining the image region using hierarchical clustering. Image Quality Metrics such Average Difference(AD), Maximum Difference(MD),Image Fidelity(IF),Peak Mean Square Error(PMSE),Signal to Noise Ratio (PSNR) are evaluated to observe the performance and presented. The summary of the segmented model and their most distinguishing features is then presented in the table at the end of this paper.*

Keywords - *Image Segmentation, K-Means Clustering Algorithm, Image Quality Metrics.*

Date of Submission: 17-10-2018 Date of acceptance: 03-11-2018

I. Introduction

Color Image segmentation has been an important and difficult task in processing many image, video and computer vision applications. It is one of the key tools for image analysis [1]. Segmentation has become the prominent objective in image analysis that uses the most common image feature such as intensity, color, texture, etc. In Literature much work has been studied based on statistical and structural method for image analysis for gray scale images. The statistical method make use of the probability distribution function of pixels and regions to characterize the image [2] and the structural, which analyzes the image in terms of organization and relationship of pixels and regions by specified relations [3].

For Color image segmentation, we have various technologies available among which clustering is widely chosen [4]. In Literature different clustering algorithms has been discussed and classified into four types: Hard Clustering, Fuzzy Clustering, Hierarchical Clustering and Probabilistic Clustering [7]. K-Means is the widely accepted hard clustering algorithm which is used for color image segmentation [5][6].

This paper presents an image segmentation approach using K-Means clustering technique based on intensity and color feature. In K-Means to calculate the distance between each point of the dataset to every centroid initialized, we make use of City Block distance metrics from the different technique available such as Euclidean distance, City Block distance (Manhattan), Cosine distance and Correlation distance. To obtain the number of image regions K, hierarchical clustering has been used. After that we go through experiment for implementing K-Means algorithm in Matlab. The performance of the segmented algorithm is evaluated by computing the Image Quality Metrics.

The rest of the paper is organized as follows. Section II presents a brief overview of the Related Work Section III describes the K-Means Clustering Algorithm. In section IV the Image Quality Metrics are presented and discussed. Section V describes the Methodology used. Section VI demonstrates the Experimental results and finally section VII Concludes with some remarks.

II. Related Work

Reddy *et al* (2007)[9] developed an image segmentation method based on Finite Generalized Gaussian Mixture Model using EM and K-Means algorithm. The methodology involves obtaining the number of image region and the initial estimates of the model parameter using k-means and through EM algorithm the final estimates of the parameters are obtained. The initial value of K (image region) is determined using histogram of the pixel intensities of the entire image. The segmentation algorithm is developed based on pixel allocation to the region which maximizes the component likelihood function. The performance of the developed algorithm is compared by evaluating different Image Quality Metrics such as Mean Square Error, Signal to Noise Ratio and

Quality Index. The Experimental results showed that the proposed algorithm has better retrieval capability compared to the Finite Gaussian Mixture Model Method. It has been observed that the developed Method performs much superior to the existing algorithms with respect to the image Quality Metrics. The performance of the Image Segmentation Model is also studies through classifier accuracy by computing the misclassification rate.

Rajkumar *et al* (2011)[10]developed an image segmentation algorithm based on Finite doubly truncated bivariate Gaussian mixture distribution. It uses two important characteristics considered as feature vector of the color image are Hue and Saturation. The number of image regions in the whole image is determined using the hierarchical clustering algorithm. The bivariate feature vector (consisting of Hue angle and Saturation) of each pixel in the image region follows a doubly truncated bivariate Gaussian distribution. The model parameters are estimated using EM Algorithm; the updated equations of EM-Algorithm for a finite mixture of doubly truncated Gaussian distribution are derived. The performance of the developed algorithm is evaluated computing the image segmentation metrics by Probabilistic Rand Index (PRI), Global Consistency Error (GCE), and Variation of Information (VOI).

Seshashayee *et al* (2011)[8]proposed an unsupervised image segmentation algorithm based on finite new symmetric mixture model with K-means clustering. The methodology involves obtaining the number of image regions using K-means algorithm. The number of segments in each image is determined by the histogram of pixel intensities. The model parameters for each image region are estimated by deriving the updated equations of the Expectation Maximization (EM) algorithm. The segmentation of the image is done by maximizing the component likelihood. The performance of the segmentation technique is evaluated on (i) Probabilistic Rand Index (PRI), (ii) Variation Of Information (VOI) and (iii) Global Consistence Error (GCE).The experimental results observed that the developed algorithm performs better with respect to the image segmentation metrics and the image quality metrics. It is also observed that the algorithm performs well even when the kurtosis parameter of each component of the model is zero.

Seshashayee *et al* (2013)[11] proposed an image segmentation algorithm based on new symmetric mixture model with K-means. The methodology involves obtaining the image region using Hierarchical Clustering Algorithm. The updated equations of the model parameter are derived through EM algorithm under Bayesian framework. The probability density function of the pixel intensities of the image is estimated. The performance is studied by obtaining the image segmentation performance measures namely, Probabilistic Rand Index (PRI), Global Consistency Error (GCE) and the Variation Of Information (VOI) and image quality metrics such as average difference, maximum distance, image fidelity, mean square error, and signal to noise ratio and image quality index. The results shows that the k-means algorithm prove more suitable for the images having platy-kurtic image regions. The new symmetric mixture model is capable of characterizing several natural images with kurtosis close to 2.52.

Kirati and Tlili (2014)[7]presented a new segmentation model based on probabilistic clustering. The unsupervised segmentation method based on non-parametric clustering able to deal with two issues. Firstly the methods tend to estimate the region membership probabilities for each pixel of the image and secondly the segmentation results depend strongly on the initialization of these regions and the selection of the appropriate number of segments. The experiments show that, in absence of saturation, the proposed method may assign pixels of different regions to the same class when their hues are too close.

Jyothirmayi *et al* (2015)[12] developed an image segmentation method based on generalized Laplace Mixture Model integrated with hierarchical clustering method. The methodology used involves obtain the number of image regions and initial estimates of the model parameters using hierarchical clustering algorithm. The pixel intensities in the whole image region are characterized by a k-component mixture of generalized Laplace distribution where k is the number of regions. The model parameters are obtained through EM Algorithm. The image segmentation performance measures such as GCE, PRI, and VOI are computed. It is also observed from the image quality metrics for the image retrievals computed with this five images in image reconstruction outperforms existing image segmentation algorithms when pixel intensities in each image region are having platy kurtic distribution.

Krishna and Srinivas(2015)[14] presented an approach of image segmentation based on Bivariate Log normal Distribution. The concept involves color image segmentation using YIQ color model. The YIQ features are extracted where Y refers to the luminance which signifies the gray level component. I and Q signify the color chrominance. The methodology involves giving input images to the log normal mixture model. The pixel of each pdf (probability density function) of each pixel are extracted Y and I are considered .Each image is segmented using these color components. The number of image region are obtained by k-,means algorithm for which the initial value of number of components is identified by no of peaks of the image histogram. The performance is evaluated using metrics like Image Fidelity (IF),Average Distance(AD),Maximum Distance(MD) and Peak Signal to Noise Ratio(PSNR).

Jyothirmayi *et al* (2017)[13] proposed an image segmentation method based on doubly truncated generalized Laplace distribution mixture model and hierarchical clustering. The methodology involves obtaining the number of image segments and initial estimates of model parameters using hierarchical clustering algorithm. The feature vector of image region is characterized by truncated generalized Laplace distribution which improves Laplace and generalized Laplace distributions as limiting cases. The model parameters are estimated by deriving updated equations of the scale and location parameters. The shape parameter is estimated using sample kurtosis. The experiment was carried on five images from Berkeley data set and the performance was evaluated on PRI, GCE and VOI. The results showed that proposed algorithm perform better due to the effect of truncation used for modeling feature vector. It was also stated the proposed algorithm is much useful for analyzing images arising at several domains of applications. It is possible to extend this image segmentation method for color images considering a 3-dimensional feature vector.

III. K-Means Clustering

The Image Processing using clustering is an efficient method. Clustering is defined as a process of grouping similar objects into different group or partitioning data into subsets so that the data in each subset is defined accordingly to some distance measure [15].Clustering technique is one of the common techniques for statistical data analysis which has been used in various fields such as pattern recognition, image analysis, machine learning, data mining and bioinformatics.

K-Means is an unsupervised clustering technique to find the partition of the data which minimizes the squared error or the sum of squared distance between all the points and their respective cluster centers [16]. K-means or Lloyd algorithm [17] is an iterative data-partitioning algorithm that takes k as the input parameter, the number of cluster and partitions a set of n objects into k clusters. The k-means clustering algorithm mainly depends on the calculation between pixels and centroid and thereby trying to allocate a pixel to a cluster or(segment)based on minimum distance criteria ,so the main goal of the algorithm is to minimize the mean of square distance between the pixel and the cluster centroid[18].

3.1 K-Means function

We make use of “k-means” function to perform k-means clustering technique in matlab[19]. The function k-means IDX= kmeans(x,k) perform kmeans clustering, using an iterative algorithm which assign objects to cluster so that the sum of distance from each object to its cluster centers[4]. The kmeans return an n by 1 vector IDX containing the cluster indices of each point where x is a vector, kmeans treats it as n by 1 data matric irrespective of the orientation. [IDX,c]= kmeans(x,k) return the k cluster centroid location.

The Formal Algorithm for K-means clustering is as follows [20]:

1. Select the number of cluster k and center.
2. Calculate the distance of each data item to its centroid
3. Generate k cluster by assigning all points to the closet centroid.
4. Recompute the centroid of each cluster until the centroid does not change.

IV. Image Quality Metrics

The advancement in digital image technologies for various applications such as medical imaging, pattern Recognition, traffic monitoring system, aerospace, Satellite imaging etc has emphasized for accurate quality assessment methods. The quality of images can be effect by various processes such as compression, transmission, display, and acquisition [21]. The quality measures included in our evaluation are Average Difference(AD), Maximum Difference(MD), Image Fidelity(IF),Peak Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR)[22].They provide some measure of closeness between two digital images by exploiting the differences in the statistical distributions of pixel values. Org(i, j) and Deg(i, j) denote the samples of original and degraded image fields.

4.1 Average Difference (AD)

The average difference refers to the pixel difference between the original image and its corresponding degraded image. This measure is applicable to any image processing applications where we find the average difference between two images. Larger value of the AD, specifies the poor quality of the image [21].

AverageDifference=

$$\sum_{i=1}^{M} \sum_{j=1}^{N} [Org(i,j) - Deg(i,j)]/MN$$

Where Org (i, j) is the Original image at i and j co-ordinates and Deg(i, j) is the Degraded image at i and j co-ordinates, M represents the number of rows of pixels of the images and N represents the number of columns of pixels of the image.

4.2 Maximum Difference (MD)

It is the absolute difference between original and Degraded image. Higher the value of Maximum Difference indicates that the image is poor quality [21].

Maximum Difference $= Max\{|Org(i,j) - Deg(i,j)|\}$

Where Org (i, j) is the Original image at i and j co-ordinates and Deg(i, j) is the Degraded image at i and j co-ordinates.

4.3 Image Fidelity

The fidelity measures the closeness of an image to its ideal image.

Fidelity $= 1 - (\sum_{i=1}^{M} \sum_{j=1}^{N} [Org(i,j) - Deg(i,j)]\,2 / \sum_{i=1}^{M} \sum_{j=1}^{N} [Deg(i,j)]\,2)$

Where Org (i, j) is the Original image at i and j co-ordinates and Deg(i, j) is the Degraded image at i and j co-ordinates, M represents the number of rows of pixels of the images and N represents the number of columns of pixels of the image.

4.4 Mean Square Error (MSE), Peak Mean Square Error (PMSE)

The mean-squared-error (MSE) is the simplest, and the most widely used image quality measurement. When MSE value reaches zero then pixel by pixel matching of images becomes perfect[21].

MSE=

$$1/MN \sum_{i=1}^{M} \sum_{j=1}^{N} [Org(i,j) - Deg(i,j)]\,2$$

Peak Mean Square Error is given by

PMSE =

$$1/MN \sum_{i=1}^{M} \sum_{j=1}^{N} [Org(i,j) - Deg(i,j)]\,2 / [Max\{Deg(i,j)\}]2$$

Where Org (i, j) is the Original image at i and j co-ordinates and Deg(i, j) is the Degraded image at i and j co-ordinates, M represents the number of rows of pixels of the images and N represents the number of columns of pixels of the image.

4.5 Peak Signal to Noise Ratio (PSNR)

The PSNR is inversely proportional to the Mean Squared Error. When comparing the two images, PSNR is calculated by taking the Mean Squared Error (MSE) between the pixel intensities and taking the ratio of the maximum possible intensity to the result of the calculation.

PSNR $= 10 * \log_{10} MaxI^2 / MSE$

V. Methodology

Step 1: Load the image to be segmented.
Step 2: Resize Input Image.
Step 3: Convert color image into gray image.
Step 4: Apply adaptive histogram equalization to enhance the image
Step 5: Evaluate the histogram for the image.
Step 6: Obtain the number of image regions K using hierarchical clustering.
Step 7: Apply the K-Means Clustering algorithm.
Step 8: Display the final Segmented Image.

VI. Experimental Results

To demonstrate the image segmentation algorithm using K-means experiment is conducted with four images taken from Berkeley image data set https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html [23]. The pixel intensity and color of the image is taken as the image feature. The number of image region of each image considered for experiment is determined by Hierarchical Clustering Algorithm. After developing the image segmentation method, performance can be evaluated by subjective image quality testing or objective image quality testing. Few image quality measure has been taken for this experiment given by [22] The Experimental Results along with the Image quality metrics are computed and presented.

Table 1

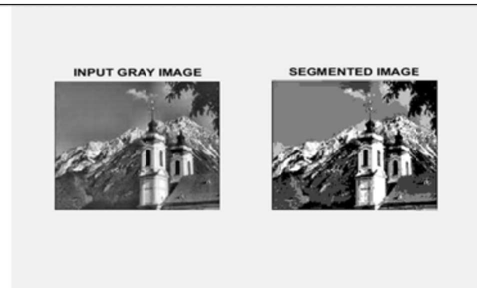
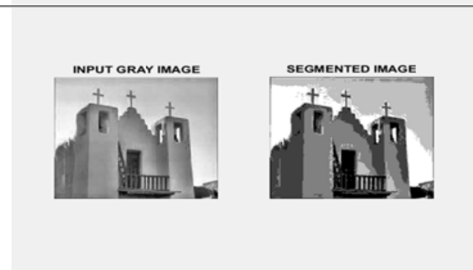
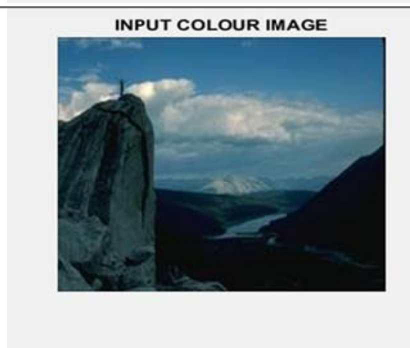
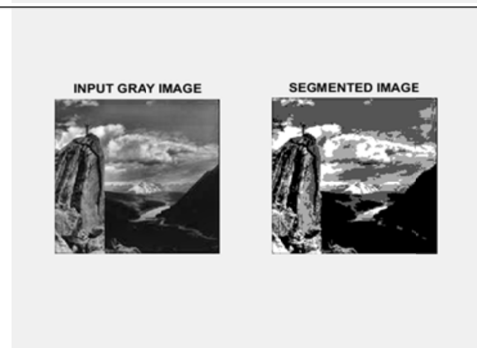
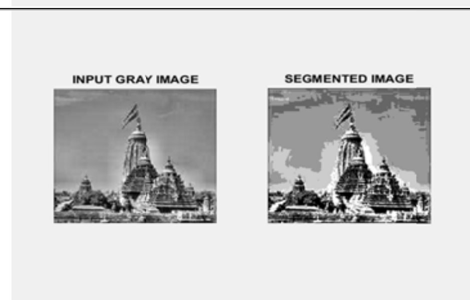
Original Image	Segmented Image
	
	
	
	

Table 2

Original Image	Image Metrics
	Average Difference: 106.7074 Maximum Difference: 250 Image Fidelity: 0.04863 Peak Mean Square Error: 813.0508 Signal to Noise Ratio: 43.8173
	Average Difference : 153.8512 Maximum Difference : 250 Image Fidelity : 0.040216 Peak Mean Square Error : 1264.8898 Signal to Noise Ratio : 39.3979
	Average Difference: 89.6288 Maximum Difference: 251 Image Fidelity: 0.044654 Peak Mean Square Error: 1023.9814 Signal to Noise Ratio: 41.5107
	Average Difference: 130.3428 Maximum Difference: 247 Image Fidelity: 0.054375 Peak Mean Square Error: 844.3485 Signal to Noise Ratio: 43.4396

VII. Conclusion

In this paper image segmentation technique using probabilistic distribution has been studied and listed in a table along with an experimental study for image segmentation using k-means clustering algorithm. The experiment was successfully implemented using four images from Berkeley image dataset, to obtain image region Hierarchical Clustering method and to calculate the distance city block distance (Manhattan) is used instead of Euclidean distance. The implementation was performed using Matlab. The results reveals that the image segmentation method outperform in segmenting the image with respect to quality metrics such as Image fidelity and signal to noise ratio. The Image quality metrics are computed and presented. The performance evaluation has always been a challenging issue and considering few metrics alone, we cannot evaluate the quality of the image. Other metrics such as NAE (normalized absolute error), SC (structural content) and SSIM (structural similarity index) plays a vital role to access the quality of the Image. It is possible to extend this research with different image feature and distance metrics

Table 3

Sr. No	Ref	Segmentation model	Image region Obtained	Model Parameter	Performance Measure and Image Quality metrics Used	Distribution
1	Jyothirmayi, *et al* (2017)	Doubly truncated generalized Laplace distribution mixture model and hierarchical clustering.	HC	EM Algorithm	Probabilistic Rand Index (PRI), Global Consistence Error (GCE) and Variation of Information (VOI)	Laplace Distributions
2	Krishna and Srinivas (2015)	Bivariate Log normal Distribution and (Principal component analysis)PCA	k-means		Image fidelity (IF),average distance(AD),maximum distance(MD) and peak signal to noise ratio(PSNR)	Bivariate Log normal Distribution
3	Jyothirmayi *et al* (2015)	Generalized Laplace Mixture Model integrated with hierarchical	HC	EM Algorithm	Global Consistence Error (GCE), Probabilistic Rand Index (PRI), and (Variation of Information)VOI	Laplace Distribution

		clustering method				
4	Rao *et al* (2014).	Finite mixture of Pearsonian Type IV distribution	k-means &HC	EM with Bayesian framework	Probabilistic Rand Index PRI, Global Consistence Error (GCE) and Variation Of Information (VOI)	Pearson Distribution
5	Seshashayee *et al* (2013)	Generalized new symmetric mixture distribution hierarchical clustering	k-means	EM With Bayesian framework .	Probabilistic Rand Index (PRI), Variation Of Information (VOI) and Global Consistence Error (GCE)	Gaussian Distribution
6	Rajkumar *et al* (2011)	Finite doubly truncated bivariate Gaussian mixture distribution	HC	EM Algorithm	Probabilistic Rand Index (PRI), Global Consistency Error (GCE), and Variation of Information (VOI).	Gaussian Distribution
7	Seshashayee *et al*(2011)	New symmetric mixture model with K-means clustering	k-means	EM Algorithm	Probabilistic Rand Index (PRI), Variation Of Information (VOI) and Global Consistence Error (GCE)	Gaussian Distribution
8	Reddy et al (2007)	Finite Generalized Gaussian Mixture Model using EM and K-Means algorithm	K-means	EM Algorithm	Mean Square Error, Signal to Noise Ratio and Quality Index.	Gaussian Distribution

References

[1]. K. S. Rao, P. C. Sekhar, and P. S. Rao, Image Segmentation for Animal Images using Finite Mixture of Pearson Type VI Distribution. Global Journal of Computer Science and Technology, 2014.

[2]. R.C. Dubes and A.K..Jain, “Random field models in image analy sis”, Journal of App lied Statistics,1989, Vol.16, No.2.

[3]. J. K. Udupa, and S. Samarasekera, Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. Graphical models and image processing, 1996, 58(3), 246-261.

[4]. M.Bora, D.Jyoti, D.Gupta, and A.Kumar,“Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab” International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2501-2506, ISSN: 0975-9646

[5]. S.Saraswathi, andL. M. I. Sheela, Lung image segmentation using K-means clustering algorithm with novel distance metric. Int. J. Recent Trends Eng. Res., 2016 2,236-245.

[6]. C.Chandhok, S.Chaturvedi, and A. A .Khurshid, An approach to image segmentation using K-means clustering algorithm. International Journal of Information Technology (IJIT), 2012, 1(1), 11-17.

[7]. I.Kirati, and Y. Tlili, Unsupervised Color Image Segmentation Based on Non Parametric Clustering. Journal of computing and information technology, 2014, 21(4), 247-254.

[8]. M. Seshashayee, K. S.Rao,and P. S. Rao, Image Segmentation Based on a Finite Generalized New Symmetric Mixture Model with K–Means , 2011.

[9]. P. P. Reddy, K. S. Rao, andS. Yarramalle,Unsupervised image segmentation method based on finite generalized gaussian distribution with EM and K-means algorithm. Proceedings of International Journal of Computer Science and Network Security, 2007, 7(4), 317-321.

[10]. GVS, Rajkumar, Srinivasa Rao, and Srinivasa Rao. "Image Segmentation and Retrievals based on finite doubly truncated bivariate Gaussian mixture model and K-eans." Image 25.4 (2011).

[11]. M.SESHASHAYEE, K. S. RAO, P. S.RAO, and C.SATYANARAYANA, IMAGE SEGMENTAION METHOD BASED ON GENERALIZED NEW SYMMETRIC MIXTURE DISTRIBUTION WITH HIERARCHICAL CLUSTERING. 2013. IMAGE, 6(6).

[12]. T.Jyothirmayi, K. S.Rao, P. S. Rao, and C. Satyanarayana, Studied on Image Segmentation Integrating Generalized Laplace Mixture Model and Hierarchical Clustering. International Journal of Computer Applications, 2015,128(12).

[13]. T.Jyothirmayi, K. S.Rao, P. S.Rao, and C. Satyanarayana,Performance Evaluation of Image Segmentation Method based on Doubly Truncated Generalized Laplace Mixture Model and Hierarchical Clustering. International Journal of Image, Graphics and Signal Processing, 2017.9(1), 41.

[14]. C. M.Krishna, andY. Srinivas, Unsupervised Color Image Segmentation Using Bivariate Features and Dimensionality Reduction Methodology. International Journal of Modern Computer Science and Applications (IJMCSA), Volume, (3).

[15]. T. S. Madhulatha, An overview on clustering methods. arXiv preprint arXiv:1205.1117, 2012.

[16]. H. Turi, Rose Cluster Based Image Segmentation. Diss. phd Thesis, Monash University, Australia, 2001.

[17]. S. Lloyd, "Least Squares Quantization in PCM." IEEE Transactions on Information Theory. Vol. 28, 1982, pp. 129–137.

[18]. D. J.Bora, Performance Comparison of K-Means Algorithm and FCM Algorithm with Respect to Color Image Segmentation. International Journal of Emerging Technology and Advanced Engineering, 2017. 7(8), 460-470.

[19]. https://in.mathworks.com/help/stats/k-means-clustering.html

[20]. S.Ayesha, T.Mustafa, A.R.Sattar, and M.I. Khan, Data mining model for higher education system. European Journal of Scientific Research, 2010, 43(1), pp.24-29.

[21]. S.Rajkumar,,and G. Malathi,A comparative analysis on image quality assessment for real time satellite images. Indian Journal of Science and Technology, 2016,9(34).

[22]. A. M.Eskicioglu, and P. S. Fisher, Image quality measures and their performance. IEEE Transactions on communications, 1995,43(12), 2959-2965.

[23]. https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html.