An Efficient Clustering Algorithm for Image Colour Feature Extraction

[1]Wirag Pushkar Ranjan ,[2] Mukesh Kumar, [3] Pritaj Yadav
[1]M.Tech Scholar , [2] Associate. Professor ,[3] Associate. Professor
[1]Department of Computer Science & Engineering, [1]RNTU, BHOPAL,INDIA

Abstract

The Colour separation or colour feature extraction operations allow different 'bands' for instance from a digital color photo or scanned images as if using color filters when taking the image or picture. After the color extraction, user can perform the normal Image Processing operations like image Filtering, Classification and Clustering etc. on these colour bands. The map that have a picture/image domain or 24 bit Color domain store for each pixel values like; Red, Green and Blue. These are the basic colour in nature. The Color separation function or operation allows the user to retrieve for each pixel either the Red, Green or Blue value and store these in a separate colour map. The person can also retrieve Cyan, Yellow, Magenta or any combination of colour like Gray values or Hue, Saturation or Intensity values for each colour pixel.

1.0 Introduction

1.1 Types of Images & its Combination

The following description of image types shows the use of colour in different fields of digital area:

Bitmap

A bitmap or pixmap is a type of memory organization or image file format used to store digital images in the computer graphics environment. The term bitmap simply means a “map of bits”. Many graphical user interfaces uses bitmap in their built-in graphics subsystems. The term bitmap implies one bit per pixel, while pix-map is used for images with multiple bits per pixel.

Duo-tone

Duo-tone is a halftone reproduction of an image using the super-imposition of one contrasting color halftone (usually black) over another color half-tone. This is most often used to bring out middle-tones and highlight an image. With the comparison, a fake duo-tone is shaped by printing a single solid color with a contrasting half-tone over it. This process generally loses a lot of contrast in the image, but it also forms a rich effect. Duo-tones, tri-tones and quad-tones are easily created by using image manipulation programs.

Grayscale

In the field of both photography and digital computing, a gray-scale digital image is an image in which the value of each pixel is a single sample which means it carries only intensity information. The images of this type also known as black & white image. These are composed exclusively of shades of gray, varying from black at the weakest intensity to white, at the strongest. Grayscale images are also called monochromatic, denoting the presence of only one colour.

Half-tone

The halftone is the “industry standard” reprographic technique. It simulates continuous tone imagery through the use of small dots, varying either in size, in shape or in spacing. The “halftone” is also used to denote specifically to the image that is produced by using this process. This reproduction technique relies on creating a basic optical illusion; that these tiny half-tone dots are blended into smooth tones by the human eye. The resolution of a halftone screen is measured in lines per inch.

2.0 Methodology

The K-means clustering implementation is a method through which a set of data points can be partitioned into several disjoint subsets where the points in each subset are deemed to be ‘close’ to each other (according to some metric). A common metric, at least when the points can be geometrically represented is the standard Euclidean distance function. The ‘k’ just refers to the number of subsets desired in the final output. It turns out that this approach is exactly what we need to divide our image into a set of colours.

Implementation of the original K-means clustering is very straightforward method. Suppose reduction to K colors; first, palette is established by selecting K random pixels from the input image. In second step, the K-means algorithm is used to modify this palette. Third, the input image is converted to output image.

2.1 Dominant color selection from image using k-means clustering

By using k-mean clustering, it is possible to detect dominant colors in an image. After going through sequence of web snippets and code playing, it is noticed excellent results using k-means colour clustering algorithm. The k-mean clustering can be implemented using a number of programming languages like scikit-learn, python, opencv etc.

Basically k-means is a clustering algorithm used in Machine Learning applications where a series of data points are to be categorized to 'k' groups. This algorithm works on simply distance calculation. The algorithm work as follows :

1. Randomly choose 'k' points from dataset.
2. For every data point, calculate distance from each of the 'k' clusters and assign it to least distant cluster.
3. Next every assignment recalculate the cluster centroid by averaging the distances of all the associated points.
4. In the further step, there is the process of distance calculation from each 'K' values
5. After distance calculation, form the specific clusters, calculate the mean value for each clusters which shows the particular colour
6. *If the data value is higher for a particular cluster, compute the mean value of that cluster. If the mean is greater than some threshold, split the cluster in half. Continue this process until all intensities are done, and we will have a more efficient clustering data points.*
7. Finally, separate the colours

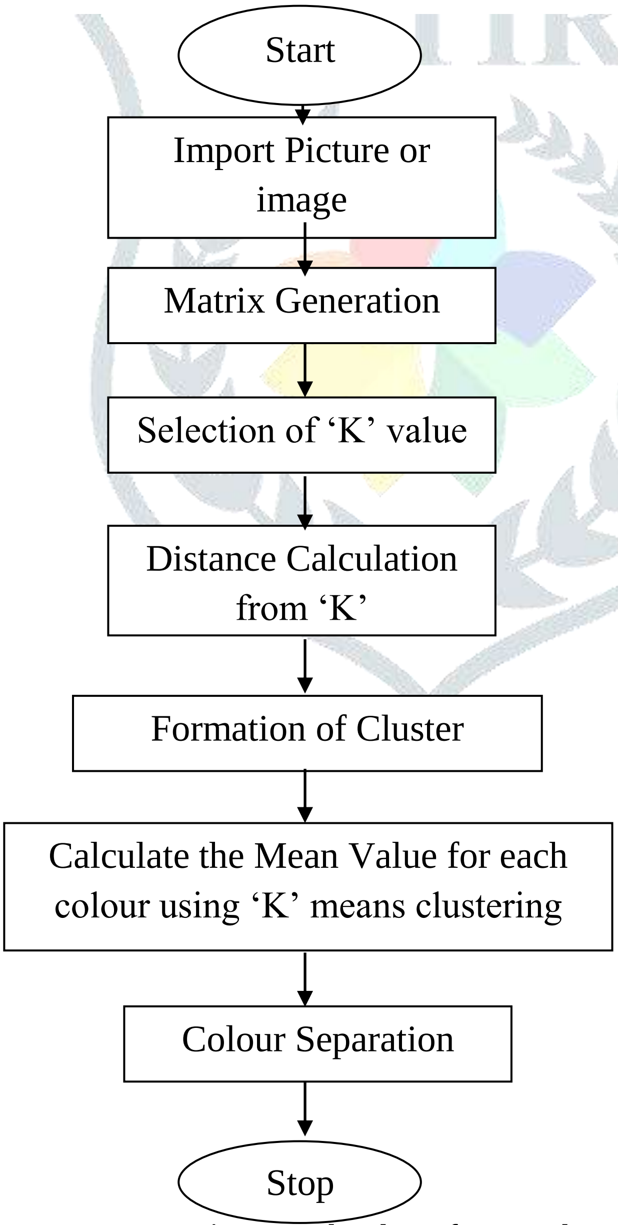

Figure 1 : Flowchart of proposed system

3.0 Result and Discussion

The *k*-means algorithm searches for a pre-determined number of clusters within an unlabeled multidimensional datasets. It accomplishes this using a simple formation of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers.

The k-mean algorithm is implemented in MATLAB 2016a for different types of images. The algorithm shows almost successful colour separation from images. Following figure shows the result obtained for four clusters C1, C2, C3 and C4. The percentage of each colour is depicted clearly.

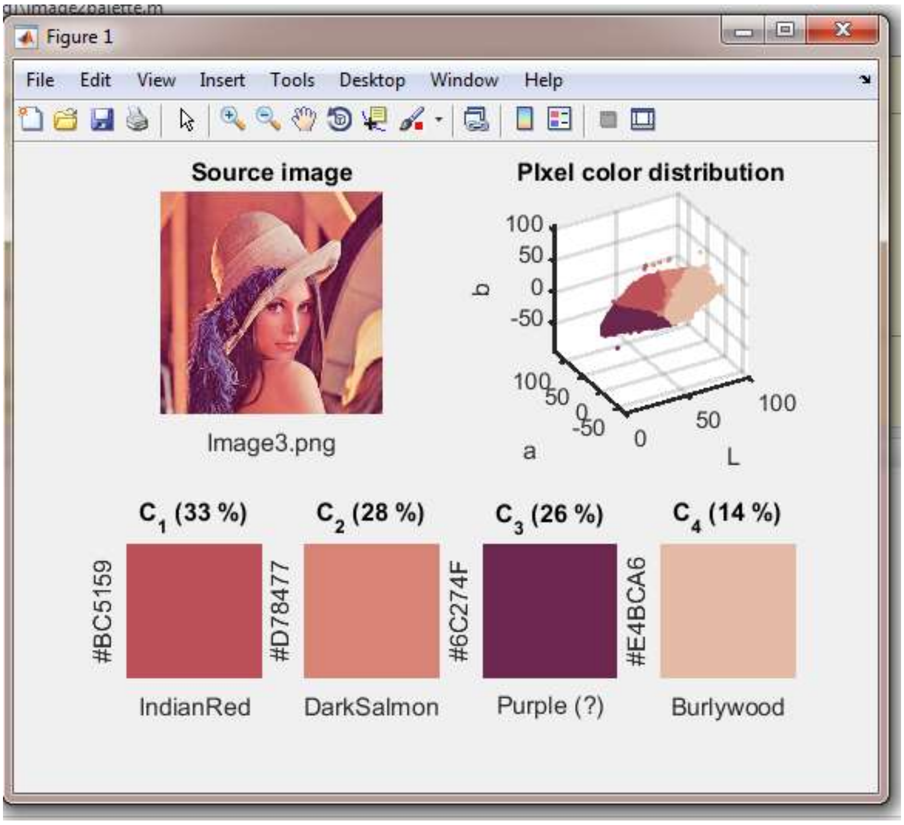

Figure 2 : Leena Image Colour Separation for .png file

Implemented K-means clustering for image [Image3.png]
#1 Lab = [49,43,17], IndianRed, 33%
#2 Lab = [64,30,20], DarkSalmon, 28%
#3 Lab = [27,35,-9], Purple, 26%
#4 Lab = [79,11,17], Burlywood, 14%

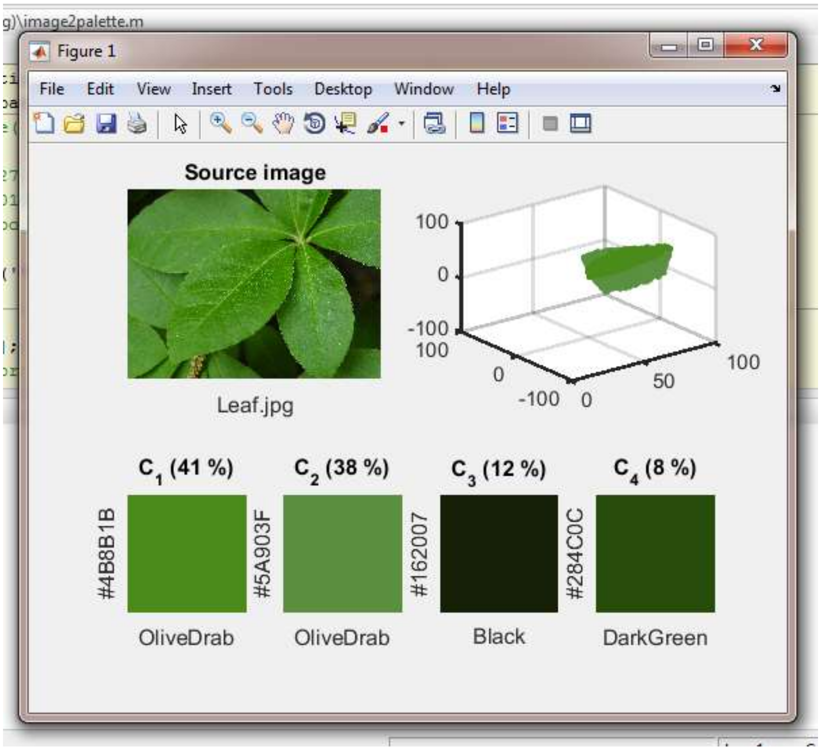

Figure 3 : Leaf Image Colour Separation for .jpg file

Implemented K-means clustering for image [leaf.jpg]
#1 Lab = [49,43,17], OliveDrab, 41%
#2 Lab = [20,30,20], OliveDrab, 38%
#3 Lab = [27,30,-9], MixYellow, 12%
#4 Lab = [19,11,17], OliveDrab, 8%

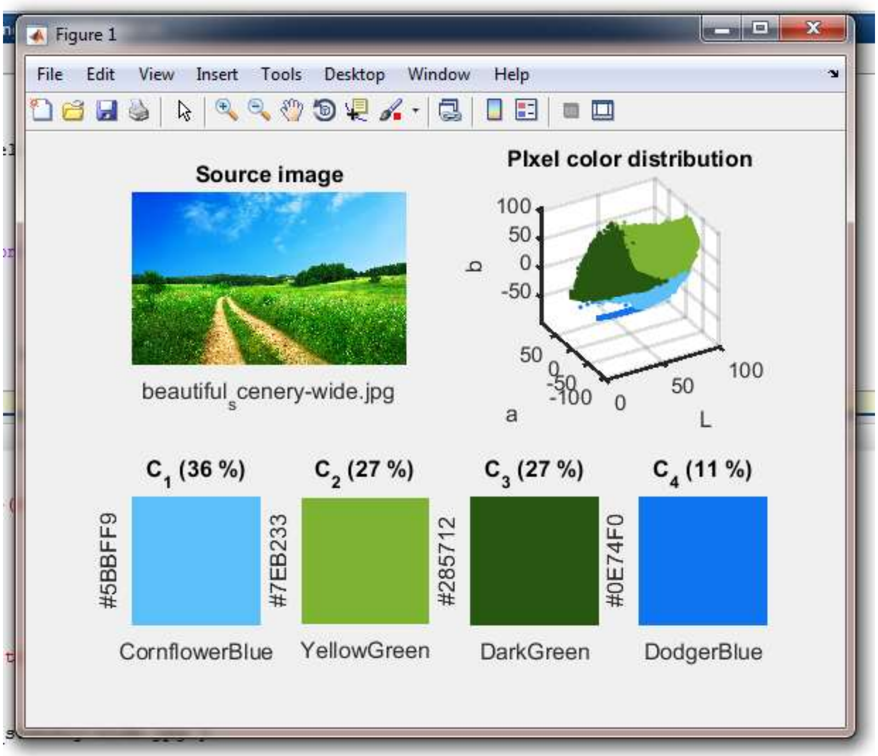

Figure 4 : Multi-colour Image Colour Separation for .jpg file

Implemented K-means clustering for image [beautiful_scenery-wide.jpg]
#1 Lab = [74,-12,-37], CornflowerBlue, 36%
#2 Lab = [67,-37,56], YellowGreen, 27%
#3 Lab = [32,-29,33], DarkGreen, 27%

#4 Lab = [50,21,-69], DodgerBlue, 11%

Following table shows the results obtain from k-mean and implemented k-mean algorithms.

Table 1 : Success rate of K-means & implemented K-mean clustering algorithm for different clusters

Picture	Cluster	K-mean clustering	Implemented K-mean clustering	Success Rate (%)
1	C1	30	33	100
	C2	24	28	
	C3	21	26	
	C4	16	14	
	Remaining	09	-	
2	C1	35	41	97
	C2	36	36	
	C3	10	12	
	C4	8	8	
	Remaining	11	03	
3	C1	30	36	99
	C2	27	27	
	C3	21	27	
	C4	10	10	
	Remaining	12	01	

The obtained result shows that the *k*-means algorithm (in the above simple case) assigns the points to clusters very similarly to how we might assign them by eye. It is notices that the k-mean algorithm finds these clusters so quickly and easily.

Here we can notice that the usage of k-means outperforms the usage of simple random selection. In fact, the images obtained with random selection with 0-255 colors are very similar to those obtained with k-means with all color ranges independently from the color space used. The drawback of using k-means is that it is obviously much slower especially if the training is performed with all the pixels of the original image. This can be mitigated by using only a random sample of the pixels.

On the other hand at first glance the results of k-means in RGB space and k-means in Lab space are similar. Anyway by taking a closer look at each couple of images, in some cases it is possible to notice some details that make that one is better than the other. In general for this particular image the face seems to be clearer and more defined when using RGB and this can be noticed especially in the first two couple of images. Overall it is noticed that when using k-means a color space is a better choice than the other, at least not for this particular image.

4.0 Conclusion

K-means algorithm seems to be very useful for color separation problem. Achieved results are better than those obtained by using other colour separation programs and methods. Modification of the initial phase of the algorithm was investigated with the modest improvement for separation from 16 colors to 256 colours. However, for some pictures shows much better improvements were achieved in case of RGB images. This may be further studied in the future.

In the research study, the k-mean algorithm is implemented in MATLAB 2016a for different types of images. The algorithm shows almost successful colour separation outcomes from selected images. The obtained results are based on different clusters like C1, C2, C3 and C4. The percentage of each colour is depicted clearly. The result shows almost 98 percentage successful results.

5.0 References

1. Malwindersingh, Meenakshibansal, A Survey on Various K- Means algorithms for Clusteringî, International Journal of Computer Science and Network Security (IJCSNS), VOL.15 No.6, June 2016
2. Amandeep Kaur Mann, Navneet Kaur Mann, Review Paper On Clustering Techniquesî ,Global Journal of Computer Science And Technology Software & Data Engineering, VOL. 13 ,2015

3. K. A. Abdul Nazeer, M. P. Sebastian, Improving the Accuracy and Efficiency of thek-means Clustering Algorithmî,Proceedings of the World Congress on Engineering 2009 Vol I WCE 2014, July 1 - 3, 2014, London, U.K.
4. Y. S. Thakare, S. B. Bagal, Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics, International Journal of Computer Applications (0975 ñ 8887) Volume 110 ñ No. 11, January 2015
5. Soumi Ghosh, Sanjay Kumar Dubey, Comparative Analysis of K-Means and Fuzzy C-Means Algorithms, International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2015
6. Sakthi, M., Thanamani, A., An Enhanced K Means Clustering using Improved Hopfield Artificial Neural Network and Genetic Algorithm, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Vol-2, 2015
7. Shafeeq A., Hareesha,K, Dynamic Clustering of Data with Modified K-Means Algorithm International Conference on Information and Computer Networks, vol. 27, 2014
8. Libao ZHANG, Faming LU, An LIU, Pingping GUO, Cong LIU, Application of K-Means Clustering Algorithm for Classification of NBA Guardsi, International Journal of Science and Engineering Applications Volume 5 Issue 1, 2016, ISSN- 2319-7560
9. R. Amutha, Renuka. K, Different Data Mining Techniques And Clustering Algorithms, International Journal Of Technology Enhancements And Emerging Engineering Research, VOL 3, ISSUE 11, ISSN 2347-4289, 2014
10. Nidhi Singh, Divakar Singh, Performance Evaluation of K- Means and Hierarchal Clustering in Terms of Accuracy and Running Time, International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2014, 4119-4121
11. Yang Y, Huang S. Image segmentation by fuzzy C-means clustering algorithm with a novel penalty term. Comput In format. 2014; 26(1):17–31.
12. Sowmya B, Sheelarani B. Colour image segmentation using fuzzy clustering techniques and competitive neural network. Applied Soft Computing. 2015; 11(3):3170–8.
13. Ebrahim A, Gharpure D. Incorporating FCM and back propagation neural network for image segmentation. International Journal of Computer and Communication Technology. 2013; 2(8):121–6.
14. Deshmukh K, Shinde GN. An adaptive neuro-fuzzy system for color image segmentation. J Indian Inst Sci. 2013; 86(5):493–506.
15. Krishnan N, Babu CNK, Rajapandian VVJ, Devaraj NR. A fuzzy image segmentation using feed forward neural networks with supervised learning. Proceedings of the International Conference on Cognition and Recognition; 2014.
16. Mujica-Vargas D, Gallegos-Funes FJ, Rosales-Silva AJ. A fuzzy clustering algorithm with spatial robust estimation constraint for noisy color image segmentation. Advances in Pattern Recognition Methodology and Applications. 2013; 34(4):400–13.
17. Yambal M, Gupta H. Image segmentation using fuzzy C means clustering: a survey. International Journal of Advanced Research in Computer and Communication Engineering. 2013 Jul; 2(7).
18. Borji A, Hamidi M, moghadam AME. CLPSO-based fuzzy color image segmentation. Fuzzy Information Processing Society, 2007. NAFIPS ‘07. Annual Meeting of the North American; 2013. p. 508–13.
19. Kang W-X, Yang Q-Q, Liang R-P. The comparative research on image segmentation algorithms. First International Workshop on Education Technology and Computer Science; 2012.
20. Jaffar MA, Ishtiaq M, Hussain A, Mirza AM. Wavelet-based color image segmentation using self-organizing map neural network. IPCSIT; Singapore: IACSIT Press; 2012.
21. Ortiz A, Gorriz JM, Ramirez J, Salas-Gonzalez D, Llamas Elvira JM. Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies. Applied Soft Computing. 2013; 13:2668–82.
22. Harrabi R, Ben Braiek E. Color image segmentation based on a modified fuzzy c-means technique and statistical features. International Journal of Computational Engineering Research. 2(1):120–35. ISSN: 2250–3005.