Color Identification Uses the K-Means Clustering Method

I Gede Made Karma
Programmed of Engineering Science, Faculty of Engineering, Udayana University
Politeknik Negeri Bali
Bukit Jimbaran, Badung, Bali, Indonesia
igmkarma@pnb.ac.id

I Ketut Gede Darma Putra
Information Technology, Faculty of Engineering
Udayana University
Bukit Jimbaran, Badung, Bali, Indonesia
ikgdarmaputra@unud.ac.id

Made Sudarma
Electrical Engineering, Faculty of Engineering
Udayana University
Bukit Jimbaran, Badung, Bali, Indonesia
msudarma@unud.ac.id

Linawati
Electrical Engineering, Faculty of Engineering
Udayana University
Bukit Jimbaran, Badung, Bali, Indonesia
linawati@unud.ac.id

***Abstract* – Theoretically there are millions of colors that can be created and identified based on their values. The variety of colors and the number of color names that are available will certainly make it difficult to remember and recognize them. The possibility of wrong color identification is very common. Colors that look different are very likely to be identified by the same name. Likewise the opposite, by mentioning the same color name, actually it is very possible the color in question is different. This happens because of the limitations possessed by the eye, the human sense of sight. It takes effort to do color clustering so that it is easier to identify colors. Color clustering will be carried out using the K-Means Clustering method based on RGB color values with the number of clusters as many as 9, 16 and 21. These clustering results are then used to determine the color identity based on the central value of each cluster. From 746 color data, this study produces color clusters that are not much different. Colors with different RGB values with slightly different sightings will be identified the same as the cluster center value where the color is incorporated.**

Keywords – color, clustering, K-Means, identification, center value

I. INTRODUCTION

Color turns out to play a very important role in everyday human life. For school-age children it turns out that color has a very significant role in the introduction of the roadway in the school environment [1]. Color can be very effective in learning, education, marketing, communication, or sports [2]. Color is also believed to be the most important visual experience for humans. It serves as a powerful information channel to the human cognitive system and has been found to play an important role in improving memory performance [3].

Besides using form, color is also an important component in identifying and recognizing an object [4]. Based on the color of an object, an object can be distinguished from other objects. Through the color differences found in an object, the shape or features of an object can also be known [5, 6].

Theoretically, there are millions of colors that can be created and identified through RGB (Red, Green, Blue), Hexadecimal, HSL (Hue, Saturation, Lightness), HWB (Hue, Whiteness, Blackness), HSI (Hue, Saturation, Intensity) or other values. This value is then used as a model to represent color information [7]. The problem then is that not all colors can be clearly distinguished by the human eye. Two colors that have different values are very likely to be identified. This happens because of the limitations possessed by the eye, the human sense of sight, in distinguishing colors [8]. The similarity in seeing two colors or more certainly will make it difficult to identify a color [9]. On the other hand, perceptions and limitations of color knowledge make it possible to cause errors in identifying colors. Colors that are clearly visible are different, but are identified with the same name [10].

Various types of colors and the number of color names that exist, of course it will be difficult to be able to remember and recognize these colors. The possibility of wrong color identification is commonplace. Lighting changes can result in the same color being displayed in different brightness levels [11]. One way to solve this is to do color clustering through the process of clustering data. Unfortunately, in this data clustering there is no one perfect solution. Each algorithm tries to minimize certain mathematical criteria, which vary between existing algorithms [12, 13]. Data clustering problems often appear in various fields, such as data mining, knowledge discovery, data compression, vector quantization and so on [14]. The purpose of data clustering, also known as cluster analysis, is to find groups naturally from a set of patterns, points, or objects. Cluster analysis is a statistical classification technique to find out whether an individual population belongs to a different group by making quantitative comparisons of several characteristics [15]. Clustering is used to search for efficient elements in a data set. Clustering is very effective in multi-dimensional data that may be difficult to manage in an effective way [16].

Clustering is the process of separating data sets into groups called clusters. The process uses an algorithm, which until now has not been good, let alone perfect and generally applicable. For multidimensional datasets, clusters are very likely to be in different subspaces. This of course will make it difficult to identify these clusters [17]. The alternative chosen later is to combine different techniques, parameters and initialization values. The goal is to get better results [18]. In many cases, manual interaction is always needed to make the choice of an effective cluster center [19].

There are two main problems in the application of clustering algorithms. The first relates to the alternative use of non-convex optimization methods to find cluster solutions.

The second relates to the number of data clusters that must be known in advance, as the input parameters of the algorithm. As a result, often the number of clusters was guessed, which resulted in obtaining unsatisfactory results [20, 21]. The results obtained with these various clustering algorithms must be evaluated to determine the correctness of the results, especially in determining the appropriate number of clusters [22].

The k-means algorithm is a popular method for automatically classifying vector-based data. A set of data is divided into separate parts, which are then called clusters. Each cluster is identified through the cluster center, which is the average of the data included in the cluster. Cluster centers are also referred to as centroids, barycenters, or centers of mass [14, 23]. The existence of a sample can be quickly predicted based on its proximity to the existing cluster center. There is a cluster center, therefore the method is called K-Means Clustering [24]. These are used in a variety of applications such as vector quantization, density estimation, characterization of workload behavior, image compression, topic identification, and many others [25].

Because the parameter settings and initial selection of different cluster centers, the data clustering algorithm can produce different clusters for a dataset. Clustering validity index (CVI) is a method that can be used to evaluate the results of a clustering algorithm. CVI can be directly used to determine the optimal number of proposed clusters [26]. Unfortunately this method has very complex calculations, very low time efficiency and very few available applications. To improve the quality of the K-Means algorithm, an alternative that can be done is to improve the method of determining the cluster center as the initial parameter, by not making random assignments [27].

II. Related Works

Various algorithms have been developed in an attempt to group data. With its advantages and disadvantages, these algorithms are very helpful in identifying groups of data that have similar or similar pattern relationships.

Partition-based Algorithms

K-means algorithm [28, 29] and SOM (Self Organizing Map) [30] are two partition-based data clustering algorithms, which are very popular and widely used. Although widely used, this algorithm turns out to have several weaknesses. The first weakness is that it requires the user to provide the number of clusters as parameters. Another disadvantage is that this algorithm forces each data to be part of a cluster, which makes partition-based approaches sensitive to outliers [31, 32].

There are several weaknesses of this partition-based data clustering algorithm, inviting the interest of researchers to develop. The focus of the discussion was the issue of outliers and the number of clusters. One of the developments carried out was the application of the "corrupted clique graph" idea and the CAST (Cluster Affinity Search Technique) algorithm based on existing data models [33]. For cluster identification two steps of the Adapt procedure are introduced, namely by estimating the cluster center and continuing with estimating the radius of the cluster [32]. Unfortunately the clustering results remain very sensitive to different parameter settings and assumptions about cluster structure do not always apply. CAST is not very effective for embedded clusters, whereas Adapt to clusters intersect [34].

Hierarchical Clustering

The hierarchical clustering algorithm does not produce a set of clusters that are released, but a clustered cluster that can be described as a tree called a dendrogram. Based on the hierarchical decomposition method, clustering algorithms can be further divided into algorithms with a bottom-up approach [35, 36] and algorithms with a top-down approach [37].

This method also has several weaknesses. The first weakness is the difficulty in cutting dendrogram to determine the cluster. Second, it is difficult for the structure in the cluster to know objects that are medoid or outside of the cluster. And lastly, hierarchical methods are considered less reliable and less unique, and are very sensitive to input sequences and slight data disturbances [38].

Color Clustering vs Image Segmentation

Color clustering is the initial process in image segmentation or object identification. Various techniques are developed to obtain the results that are expected. Image segmentation is a crucial problem in image processing and computer vision. The majority of segmentation algorithms are only associated with gray-scale images. Some use color images with RGB value representation. Although it is convenient for display devices, this is considered inappropriate with the vision of eye psychology associated with a strong correlation between the three components. Because color representation with HSI values is considered more compatible with the vision of human psychology and the three components are relatively independent, the algorithm is applied by applying the HSI color space [39].

Several new techniques have been proposed for color clustering. One technique that is considered to show good performance is to use an average shift analysis to estimate cluster centers. This technique uses spheres of a predetermined size to find the center of the color cluster in the color space [40]. Another technique is to approach partition graphics. Color pixels are used as graphic nodes. The graphic end contains a pair of vertices with the same weight and spatial distance weights [41]. Charts are sorted by optimizing the separation criteria based on the calculation of eigenvalues [42].

In the face detection algorithm, information about color and color clustering related to skin color was able to provide better results [43]. The implementation is by clustering colors in certain areas into a number of color references. Furthermore, a feature description consisting of color representation and percentage in the region is given. Each will have a measure of similarity and distance [44].

III. Methodology

Identifying the colors of an object will be done using clustering techniques, which is by clustering a number of colors into a certain number of clusters. Colors are then identified based on the cluster identity where the color is incorporated. Color clustering into a number of clusters will use the K-Means Clustering algorithm, an algorithm that is very popular for data clustering. Cluster determination is done by repeated calculations [45].

The reason for choosing the K-Means Clustering algorithm to do color clustering compared to other methods is to consider the purpose of this color clustering. The purpose of this color clustering is to collect a number of similar colors into the same cluster. Color similarity will be seen from the proximity of the color to the center of the cluster. Color clusters will be identified based on the cluster center. The number of clusters can be determined in advance. In this study the number of clusters will be determined as 9, 16 and 21. The choice of the number of clusters is based solely on the number of rainbow colors plus white and black (7 + 2) and many commonly known colors (16 and 21). With this clustering, the colors will be more easily identified by the number of commonly known variations.

The K-Means algorithm stage begins by determining the number of groups / clusters (k values) that will be formed from the existing color data sets, namely 9, 16 and 16. For each cluster, the center value is then determined, which can be determined randomly. Referring to the existing cluster center value, color data is then grouped into one cluster center. Determination of the center of the cluster is done by considering the closest distance of the color data with the existing cluster center. After all color data is grouped, then the calculation of the new cluster center is based on the color data in the group. The value of the new cluster center is then used as a reference to reclassify color data. This step is repeated until finally the cluster center value is not changed again [46-49]. Repetition will be limited to a maximum of 100 times.

This color data clustering result is then used as a reference to determine the color of the color data that is owned by an object. Color is determined based on the center of the cluster where the color data is incorporated.

IV. Experiment Result

This study uses a list of colors available on the site https://www.colorhexa.com/color-names. There are 746 types of colors with the names and RGB values of each. This whole color data is then processed by using the K-Means Clustering method based on the RGB value of the color. The results of the clustering process are carried out by the number of clusters 9, 16 and 21 and the iteration process is a maximum of 100 times, as shown in Table I.

This clustering results are then used to identify the color of the sample. In this study 27 samples of color were used, with RGB values being a combination of values 0, 128 and 255. Each color sample was then calculated the distance to the center value of each cluster from the previous clustering process. Distance calculation is done using the Euclidean Distance formula [50, 51] as formula (1).

$$D(a,b) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2} \quad (1)$$

TABLE I. RESULTS OF THE CLUSTERING PROCESS

Number of Cluster: 9					
Cluster	Centroids				n
	R	G	B	Color	
1	208	169	112	1	94
2	13	84	80	2	75
3	112	76	71	3	98
4	44	150	193	4	83
5	222	224	214	5	120
6	213	46	45	6	102
7	203	137	199	7	63
8	214	36	182	8	47
9	221	203	19	9	64
Total					746
Distance (lower is better)				45188.52635	
GAP				0.352889143	

Number of Cluster: 16					
Cluster	Centroids				n
	R	G	B	Color	
1	19	92	56	1	51
2	232	133	206	2	38
3	25	175	201	3	47
4	155	202	204	4	51
5	114	52	45	5	69
6	238	221	115	6	28
7	223	43	135	7	40
8	238	201	17	8	53
9	217	45	38	9	77
10	158	42	232	10	27
11	88	214	61	11	25
12	242	236	238	12	47
13	217	137	102	13	64
14	125	119	131	14	51
15	241	210	190	15	42
16	16	51	159	16	36
Total					746
Distance (lower is better)				35545.29818	
GAP				0.209348371	

Number of Cluster: 21					
Cluster	Centroids				n
	R	G	B	Color	
1	238	113	205	1	26
2	207	116	25	2	42
3	93	57	61	3	52
4	12	70	164	4	44
5	245	240	235	5	42
6	162	27	27	6	36
7	223	177	208	7	37
8	223	42	136	8	34
9	239	216	14	9	40
10	232	28	50	10	43
11	171	11	242	11	17
12	127	207	123	12	24
13	162	212	224	13	34
14	220	138	113	14	60
15	128	112	113	15	39
16	246	224	178	16	28
17	12	88	51	17	43
18	63	231	20	18	12
19	241	216	105	19	26
20	20	199	204	20	32
21	115	124	197	21	35
Total					746
Distance (lower is better)				31370.96438	
GAP				0.152984352	

TABLE II. COLOR IDENTIFICATION RESULTS BASED ON COLOR CLUSTERS

Sample					Result base on cluster			ΔE sample vs center of cluster		
No	R	G	B	Color	9	16	21	9	16	21
1	0	0	0		2	1	17	203.79	203.00	191.34
2	0	0	128		2	16	4	182.59	114.43	150.41
3	0	0	255		4	16	4	321.71	172.06	191.30
4	0	128	0		2	1	17	145.09	111.98	109.70
5	0	128	128		2	1	4	113.40	128.98	128.37
6	0	128	255		4	3	20	124.23	128.62	162.99
7	0	255	0		2	11	18	360.93	193.39	122.52
8	0	255	128		4	3	20	241.57	195.28	159.05
9	0	255	255		4	3	20	239.99	182.50	137.64
10	128	0	0		3	5	6	184.27	124.31	88.56
11	128	0	128		3	5	3	174.27	158.69	155.28
12	128	0	255		8	10	11	195.01	103.99	79.81
13	128	128	0		3	5	2	147.19	166.56	143.35
14	128	128	128		3	14	15	134.47	19.21	38.39
15	128	128	255		7	4	21	153.20	171.15	85.43
16	128	255	0		9	11	18	193.61	137.72	125.61
17	128	255	128		1	11	12	222.03	143.19	96.28
18	128	255	255		5	4	13	183.61	136.47	113.08
19	255	0	0		6	9	10	133.44	525.78	98.61
20	255	0	128		8	7	8	126.72	412.75	101.27
21	255	0	255		8	10	11	144.52	190.63	148.29
22	255	128	0		9	8	2	163.37	150.87	93.48
23	255	128	128		1	13	14	117.74	77.51	67.27
24	255	128	255		7	2	1	121.28	80.55	82.26
25	255	255	0		9	8	9	122.50	114.49	85.11
26	255	255	128		5	6	19	148.00	76.35	87.92
27	255	255	255		5	12	5	102.34	50.29	44.72
Average								174.11	158.18	114.37

Based on the closest distance value, each sample color is then set on the cluster where the color of the sample is incorporated. The results are presented in Table II. In plain view, there is clearly a difference between the color of the sample and the color of the center of the cluster. Interestingly, even though it looks different, there are similarities between them. So it will be very possible these colors will be identified with the same name.

To find out how far the color of the sample is from the center of the cluster, Delta E is also calculated between the color of the sample and the color of the cluster center. Delta E (ΔE) is a metric for understanding how the human eye feels different colors [52]. The calculation is done by using the RGB color value with the following formula (2) [8]:

$$AE_{RGB} = \sqrt{3(\Delta R)^2 + 4(\Delta G)^2 + 2(\Delta B)^2} \tag{2}$$

Delta E is a standard calculation metric that connects human visual assessment of the differences between two colors. This standard quantifies the differences in the two colors and is used to calculate deviations from standard color benchmarks. In general, the lower the Delta E number, the closer the two colors look [53]. The results of the calculation of ΔE between the colors of the sample and each cluster center are presented in Table II.

The calculation results of ΔE in Table II show that even though the color of the sample and the center of the cluster appear to be quite similar, it has a relatively large ΔE value. On the number 9 cluster, the range of ΔE is 102.34 - 360.93, with a difference of 258.59. On the number of clusters of 16 pieces, the range is between 19.21 - 525.78, with a difference of 506.57. At 21 clusters, the range is from 38.39 - 191.34, with a difference of 152.94. While the mean ΔE for each was 174.11, 158.18 and 114.37. Based on the difference in range and mean ΔE, clustering with cluster number 21 is relatively better than the number of clusters 9 and 16. Although in the number of clusters 16 there are those who give ΔE which is relatively small compared to others, which is 19.21, but the range is very large, namely 506.57.

Based on the distance and gap values of each cluster number as presented in Table I, indicates the same results. The distance and gap values of the number of clusters 9, 16 and 21 are (45188.526, 0.353), (35545.298, 0.209) and (31370.964, 0.153), respectively. The number of cluster 21 turned out to have the smallest distance and gap value. This means that the clustering results are better compared to the number of clusters 9 and 16.

V. Conclusion and Future Work

Color clustering is relatively good enough to eliminate the number and variety of colors, especially if it is related to human ability in differentiating and identifying so many different colors. Clustering with number 21 clusters gives relatively better results compared to the number of clusters 9 and 16. There are indications that the more clusters will give better results.

Color clustering automatically limits the color range as much as the cluster center. Different colors are very likely to be incorporated into one cluster, because the similarities are determined based on their proximity to the cluster center. Different colors will be identified as the same color, because they are in the same cluster. The clustering results are believed to be very helpful in the process of color-based image segmentation.

References

[1] Helvacıog, E. and N. Olguntürk, "Colour contribution to children's wayfinding in school environments," Optics & Laser Technology, 43(2), 2011, pp. 410-419.

[2] Dzulkifli, M.A. and M.F. Mustafar, "The influence of colour on memory performance: A review," The Malaysian journal of medical sciences: MJMS, 20(2), 2013, pp. 3.

[3] Wichmann, F.A., L.T. Sharpe, and K.R. Gegenfurtner, "The contributions of color to recognition memory for natural scenes," Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(3), 2002, pp. 509.

[4] Lloyd-Jones, T.J. and K. Nakabayashi, "Independent effects of colour on object identification and memory," The Quarterly Journal of Experimental Psychology, 62(2), 2009, pp. 310-322.

[5] Bramão, I., et al., "The role of color information on object recognition: A review and meta-analysis," Acta psychologica, 138(1), 2011, pp. 244-253.

[6] Shevell, S.K. and F.A. Kingdom, "Color in complex scenes," Annu. Rev. Psychol., 59, 2008, pp. 143-166.

[7] Agrawal, S. and D.K. Xaxa, "Survey on Image Segmentation Techniques and Color Models," Savita Agrawal et al,/(IJCSIT) International Journal of Computer Science and Information Technologies, 5(3), 2014, pp. 3025-3030.

[8] Mokrzycki, W. and M. Tatol, "Colour difference ΔE-A survey," Machine Graphics & Vision, 20(4), 2011, pp. 383-411.

[9] Abdul-Nasir, A.S., M.Y. Mashor, and Z. Mohamed, "Colour image segmentation approach for detection of malaria parasites using various colour models and k-means clustering," WSEAS transactions on biology and biomedicine, 10(1), 2013, pp. 41-55.

[10] Wallin, M. and R. Meganck, "Mapping Color," 2014.

[11] Recky, M. and F. Leberl, "Windows detection using k-means in cie-lab color space." in *Pattern Recognition (ICPR), 2010 20th International Conference on*. 2010. IEEE.

[12] Hamerly, G. and C. Elkan, "Alternatives to the k-means algorithm that find better clusterings." in *Proceedings of the eleventh international conference on Information and knowledge management*. 2002. ACM.

[13] Jumb, V., M. Sohani, and A. Shrivas, "Color image segmentation using K-means clustering and Otsu's adaptive thresholding," International Journal of Innovative Technology and Exploring Engineering (IJITEE), 3(9), 2014, pp. 72-76.

[14] Kanungo, T., et al., "An efficient k-means clustering algorithm: Analysis and implementation," IEEE Transactions on Pattern Analysis & Machine Intelligence, (7), 2002, pp. 881-892.

[15] Jain, A.K., "Data clustering: 50 years beyond K-means," Pattern Recognition Letters, 31(8), 2010, pp. 651-666.

[16] Bhatia, S.K., "Adaptive K-Means Clustering." in *FLAIRS conference*. 2004.

[17] Zhu, Y., K.M. Ting, and M.J. Carman, "Grouping Points by Shared Subspaces for Effective Subspace Clustering," Pattern Recognition, 2018.

[18] Cornuejols, A., et al., "Collaborative clustering: Why, when, what and how," Information Fusion, 39, 2018, pp. 81-95.

[19] Wang, G., Y. Wei, and P. Tse, "Clustering by defining and merging candidates of cluster centers via independence and affinity," Neurocomputing, 2018.

[20] Masud, M.A., et al., "I-nice: A new approach for identifying the number of clusters and initial cluster centres," Information Sciences, 466, 2018, pp. 129-151.

[21] Van Pham, N., et al., "A new cluster tendency assessment method for fuzzy co-clustering in hyperspectral image analysis," Neurocomputing, 307, 2018, pp. 213-226.

[22] Lee, S.-H., et al., "A new clustering validity index for arbitrary shape of clusters," Pattern Recognition Letters, 112, 2018, pp. 263-269.

[23] Arthur, D. and S. Vassilvitskii, "How slow is the k-means method?." in *Proceedings of the twenty-second annual symposium on Computational geometry*. 2006. ACM.

[24] Drake, J., Faster k-means clustering, 2013.

[25] Hamerly, G., "Making k-means even faster." in Proceedings of the 2010 SIAM international conference on data mining. 2010. SIAM.

[26] Zhou, S. and Z. Xu, "A novel internal validity index based on the cluster centre and the nearest neighbour cluster," Applied Soft Computing, 71, 2018, pp. 78-88.

[27] Zhu, E. and R. Ma, "An effective partitional clustering algorithm based on new clustering validity index," Applied Soft Computing, 71, 2018, pp. 608-621.

[28] Jing, L., M.K. Ng, and J.Z. Huang, "An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data," IEEE Transactions on knowledge and data engineering, 19(8), 2007.

[29] Krishna, K. and M.N. Murty, "Genetic K-means algorithm," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(3), 1999, pp. 433-439.

[30] Van Hulle, M.M., Self-organizing maps, in *Handbook of Natural Computing*2012, Springer. pp. 585-622.

[31] Sherlock, G., "Analysis of large-scale gene expression data," Briefings in bioinformatics, 2(4), 2001, pp. 350-362.

[32] De Smet, F., et al., "Adaptive quality-based clustering of gene expression profiles," Bioinformatics, 18(5), 2002, pp. 735-746.

[33] Ben-Dor, A. and Z. Yakhini, "Clustering gene expression patterns." in Proceedings of the third annual international conference on Computational molecular biology. 1999. ACM.

[34] Jiang, D., J. Pei, and A. Zhang, "DHC: a density-based hierarchical clustering method for time series gene expression data." in Bioinformatics and Bioengineering, 2003. Proceedings. Third IEEE Symposium on. 2003. IEEE.

[35] Bandyopadhyay, S. and E.J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks." in INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies. 2003. IEEE.

[36] Wang, L. and L. Khan, A New Hierarchical Approach for Image Clustering, in *Multimedia Data Mining and Knowledge Discovery*, V.A. Petrushin and L. Khan, Editors. 2007, Springer London: London. pp. 41-57.

[37] Shamir, R. and R. Sharan, "Click: A clustering algorithm for gene expression analysis." in *Proceedings of ISMB'00*. 2000.

[38] Suzuki, R. and H. Shimodaira, "Pvclust: an R package for assessing the uncertainty in hierarchical clustering," Bioinformatics, 22(12), 2006, pp. 1540-1542.

[39] Zhang, C. and P. Wang, "A new method of color image segmentation based on intensity and hue clustering." in Pattern Recognition, 2000. Proceedings. 15th International Conference on. 2000. IEEE.

[40] Comaniciu, D. and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 2002, pp. 603-619.

[41] Shi, J. and J. Malik, "Normalized cuts and image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 2000, pp. 888-905.

[42] Dong, G. and M. Xie, "Color clustering and learning for image segmentation based on neural networks," IEEE transactions on neural networks, 16(4), 2005, pp. 925-936.

[43] Kovac, J., P. Peer, and F. Solina, "Human skin color clustering for face detection." in *EUROCON 2003. Computer as a Tool. The IEEE Region 8*. 2003. IEEE.

[44] Deng, Y., et al., "An efficient color representation for image retrieval," IEEE Transactions on Image Processing, 10(1), 2001, pp. 140-147.

[45] Subbiah, B. and S.C. Christopher, "Image classification through integrated K-Means algorithm," International Journal of Computer Science Issues (IJCSI), 9(2), 2012, pp. 518.

[46] Jindal, M. and N. Kharb, "K-means clustering technique on search engine dataset using data mining tool," International Journal of Information and Computation Technology, 3(6), 2013, pp. 505-510.

[47] Kannan, R., S. Vempala, and A. Vetta, "On clusterings: Good, bad and spectral," Journal of the ACM (JACM), 51(3), 2004, pp. 497-515.

[48] Juang, L.-H. and M.-N. Wu, "MRI brain lesion image detection based on color-converted K-means clustering segmentation," Measurement, 43(7), 2010, pp. 941-949.

[49] Ray, S. and R.H. Turi, "Determination of number of clusters in k-means clustering and application in colour image segmentation." in *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*. 1999. Calcutta, India.

[50] Wang, L., Y. Zhang, and J. Feng, "On the Euclidean distance of images," IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 2005, pp. 1334-1339.

[51] Shih, F.Y. and Y.-T. Wu, "Fast Euclidean distance transformation in two scans using a 3× 3 neighborhood," Computer Vision and Image Understanding, 93(2), 2004, pp. 195-205.

[52] Schuessler, Z. "Delta E 101". 2016 [cited 2018 9/8/2018]; Available from: http://zschuessler.github.io/DeltaE/learn/.

[53] ViewSonic. "Technical Introduction". 2010 [cited 2018 9/8/2018]; Available from: https://www.viewsonic.com/uk/products/lcd/pdf/ti_delta_E.pdf.