

A Robust Color Image Quantization Algorithm Based on Knowledge Reuse of K -Means Clustering Ensemble

Yuchou Chang¹, Dah-Jye Lee¹, Yi Hong², James Archibald¹, and Dong Liang³

¹Department of Electrical and Computer Engineering, Brigham Young University, Provo, Utah, 84602 USA

Emails: ycchang@et.byu.edu, djlee@ee.byu.edu, jka@ee.byu.edu

²Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

yihong@cityu.edu.hk

³Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong

dliang@eee.hku.hk

Abstract— This paper presents a novel color image quantization algorithm. This algorithm improves color image quantization stability and accuracy using clustering ensemble. In our approach, we firstly adopt manifold single k -means clusterings for the color image to form a preliminary ensemble committee. Then, in order to avoid inexplicit correspondence among clustering groups, we use the original color values of each clustering centroid directly to construct a final ensemble committee. A mixture model based on the expectation-maximization (EM) algorithm is used as a consensus function to combine the clustering groups of the final ensemble committee to obtain color quantization results. Experimental results reveal that the proposed color quantization algorithm is more stable and accurate than k -means clustering. The preprocessing step of the algorithm, k -means clustering, can be implemented and executed in parallel to improve processing speed.

Index Terms—color image quantization, clustering ensemble, knowledge reuse, mixture model, expectation maximization algorithm.

I. INTRODUCTION

Color has been recognized as an important visual cue for image and scene analysis. Much work in color analysis has focused on color image formation, color quantization, human visual perception, image segmentation, color-based object recognition, and image retrieval. Color image quantization is the process used to reduce the number of colors needed to represent a digital color image. As a fundamental process in image processing, color quantization plays an important role in many subfields mentioned above. Generally speaking, color image quantization is divided into four phases [1]: (1) sampling the original image for color statistics; (2) choosing a color map based on the color statistics; (3) mapping original colors to their nearest neighbors in the color map; and (4) quantizing and rendering the original image.

Several factors such as color distortion, algorithm complexity, and the characteristics of hue, lightness, and

saturation (aka HVS) must be considered when developing a new color quantization method. Many color quantization methods are found in the literature. Notable examples include the median-cut algorithm [1], the uniform algorithm [1], the octree algorithm [2], the center-cut algorithm [3], PGF [4], the window-based algorithm [5], clustering-based algorithms such as k -means [6, 7], the adaptive clustering algorithm [8], SOM-based clustering algorithms [9], and an ant colony clustering algorithm [10]. Since these methods focus on analysis in different aspects of the quantization process, they have unique advantages and disadvantages depending on the color data sets encountered. For example, the uniform algorithm [1] is adequate for quantizing color images with homogeneous color distribution, but the algorithm greatly distorts image appearance if the image has complex color variations.

Clustering algorithms including k -means have been widely used in color quantization problems. A Lloyd's k -means clustering algorithm called the filtering algorithm [6] was implemented and applied to color quantization. A local k -means algorithm [7] was presented for efficient color quantization. In this paper, we adopt k -means as a base clustering method to study the color quantization problem because it is an unstable "weak" clustering algorithm that is able to obtain a population of diverse clustering solutions of a data set from different perspectives [11, 13, 20]. We use "ensemble" approach to combine multiple quantization results based on k -means to achieve the optimal result.

K -means methods need to address three problems: (1) determining the optimal number of clusters to be created; (2) choosing the initial clustering centers; and (3) assigning the number of points in different clusters. Different initial clustering centers will lead to different local optima and different partitions. Points lying far from the clustering centers can significantly shift the location of the center that is assigned to them. The result of this shift is that some trivial colors may be included in the cluster and distort the quantization result. For these

reasons, k -means algorithms may be unstable for quantizing the same color image many times, or they may provide inconsistent results depending on the image. Therefore, for images with widely varying color variation, the choice of quantization method is an important research problem, particularly if stable and accurate results are required.

In this paper, we propose a color quantization algorithm based on clustering ensemble [11-14]. In our approach, the positive characteristics of clustering ensemble overcome the stability and accuracy problems typically observed in k -means clustering, thus yielding better results. Our approach is based on the observation that k -means clustering can characterize the shape and density of data clusters from different perspectives, and hence can achieve the essential arbitrary shape and size of distributive color clusters. In the proposed scheme, the color image is quantized by k -means clustering with randomly chosen initial clustering centroids in an iterative fashion. Instead of using clustering labels for clustering ensemble [9], we use the original color data to obtain the final clustering result by using an expectation maximization (EM) algorithm.

This paper is organized as follows: In Section 2, we provide the background of clustering ensemble and analysis. In Section 3, we propose the new color quantization algorithm. Experimental results and analysis are given in Section 4. Section 5 includes our discussion and conclusion.

II. RELATED WORK AND ANALYSIS

The basic idea behind median-cut algorithm [1] is that every entry in the color map denotes the same number of pixels in the original image. It divides the color space based on distribution of the original colors. On the other hand, uniform quantization [1] partitions each axis into equal segments instead of considering color distribution. Original colors are mapped to regions of the divided color space. Octree algorithm [2] stores every color of the image in an octree of a depth of 8. A limit of K leaves is placed on the tree to obtain the final K quantized colors. Center-cut algorithm repeatedly splits the color set whose bounding box has the longest side until K sets are generated. The centers of the K sets are used as palette colors.

Clustering feature (CF) [5] is a triple summarizing information about a cluster. Based on CF tree, colors can be clustered to form compact and reduced color set. In [8], a superposed 3D histogram is calculated first. Then, the sorted histogram list is fed into an adaptive clustering algorithm to extract the palette colors in the image. A destined pixel-mapping algorithm is applied to classify pixels into their corresponding palette colors. Kohonen Neural Networks [9] with 256 neurons self-organize through learning to match the distribution of colors of an image. Position in RGB space of each neuron generates a color map which is used to quantize the image. According to the picking up-dropping theory, a promoted ant algorithm [10] is applied to group colors into a certain

clusters in RGB space. It finishes color quantization after colors mapping of every pixel.

Methods based on clustering ensemble have been shown to be effective in improving the robustness and stability of a clustering algorithm [11-14]. Classical clustering ensemble methods take multiple clusters into consideration by employing the following steps: First, a population of clusters is obtained by executing different clustering algorithms on the same data set. Second, an ensemble committee is constructed from all resulting clusters. Third, a consensus function is adopted to combine all clusters of the ensemble committee.

Figure 1 shows the framework of the classical clustering ensemble method. By leveraging the consensus across multiple clusters, clustering ensemble gives a generic knowledge framework for combining multiple clusters. Two factors crucial to the success of any clustering ensemble are the following:

- The construction of an accurate and diverse ensemble committee of diverse clusters.
- The design of an appropriate consensus function to combine the results of the ensemble committee.

Strehl and Ghosh [11] introduced the clustering ensemble problem and provided three effective and efficient algorithms to solve the problem: the Cluster-based Similarity Partitioning Algorithm (CSPA), the HyperGraph Partitioning Algorithm (HGPA) and the Meta-Clustering Algorithm (MCLA). In order to benefit from the clustering ensemble approach, objects can be represented using different features. The number and/or location of initial cluster centers in iterative algorithms such as k -means can be varied. The order of data presentation in on-line methods such as BIRCH can be varied. A portfolio of very different clustering algorithms can be jointly used. The experiments of Strehl and Ghosh showed that clustering ensemble can be used to obtain robust, super-linear clustering algorithms and to dramatically improve sets of subspace clusterings for different research domains.

Topchy et al. [12] extended the clustering ensemble research in several aspects. They introduced a unified representation for multiple clusterings and formulated the corresponding categorical clustering problem. They proposed a probabilistic model of the consensus function using a finite mixture of multinomial distributions in a space of clusterings. They also demonstrated the efficiency of combining partitions generated by weak clustering algorithms that use data projections and random data splits. Fred and Jain [13], based on the idea of evidence accumulation, considered that each partition is viewed as an independent evidence of data organization. Individual data partitions are combined based on a voting mechanism to generate a new $n \times n$ similarity matrix for n patterns. The final data partition of the n patterns is obtained by applying a hierarchical agglomerative clustering algorithm on this matrix. Kuncheva and Vetrov [14] used standard k -means started from a random initialization to evaluate the stability of a clustering ensemble. From the experimental results, they

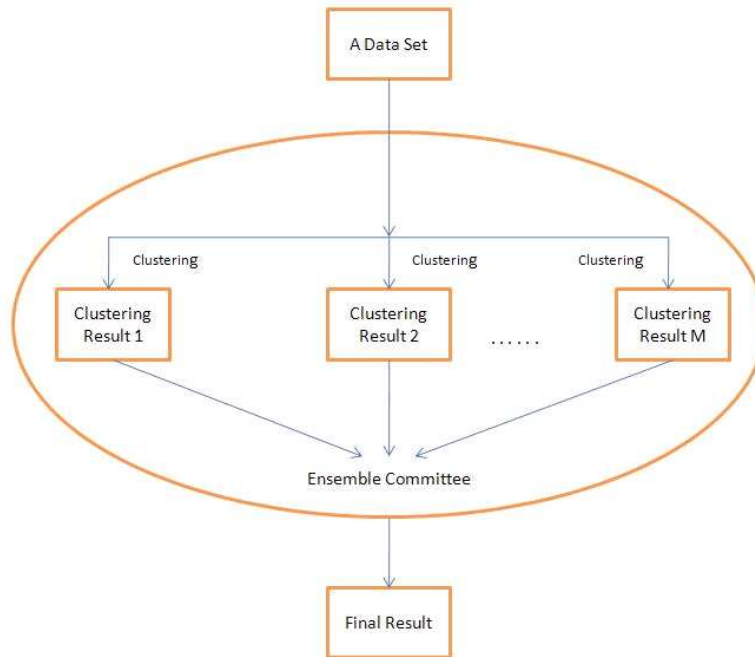


Figure 1 Framework of classical clustering ensembles.

concluded that an ensemble is generally more stable than single component clustering.

As mentioned above, demonstrated by a large number of experiments on real and synthetic data, approaches based on clustering ensemble are able to improve clustering stability and accuracy. However, unlike classification problems where labels of data items are known beforehand, data items in unsupervised clustering problems are not labeled. Therefore, there is no explicit correspondence among the results provided by different clusters. For example, the following two clustering results:

$$L^{(1)} = (1, 1, 2, 2, 2, 3, 3, 5, 4) \text{ and}$$

$$L^{(2)} = (2, 2, 3, 3, 3, 1, 1, 4, 5)$$

are logically identical. To solve the problem of inconsistent clustering, a similarity matrix was adopted to create relationships between clustering results [13].

Although optimization algorithms such as nearest-neighbor matrix can be used to reduce space and time requirements, both space and time costs are still high. In our approach, we use the original data for the clustering ensemble in order to avoid the confusion of meaningless clustered labels as shown in $L^{(1)}$ and $L^{(2)}$, in which labels reflect only group identity rather than essential attributes of the clustering sets.

The Expectation-Maximization (EM) algorithm [19] is used to determine the maximum likelihood parameters of a mixture of N Gaussians in the feature space. The EM algorithm is an iterative method to obtain an optimal solution. Given the current estimation of the parameter set, each iteration of EM algorithm reestimates the parameter set according to expectation and maximization. It has been widely used for data clustering in machine learning and computer vision.

III. ALGORITHM

A. Algorithm Framework

We propose a new color quantization algorithm framework based on clustering ensemble. Figure 2 shows the flowchart of the complete algorithm. The color image usually represents each color component with an 8-bit integer, and each pixel requires 24 bits to completely and accurately specify its color. Of the many color spaces in existence – RGB, CMY, YIQ, YUV, CIE Lab, HSV, etc. – none can be considered universal because color can be interpreted and modeled differently [15]. In our experiments, we employ the commonly used RGB color space.

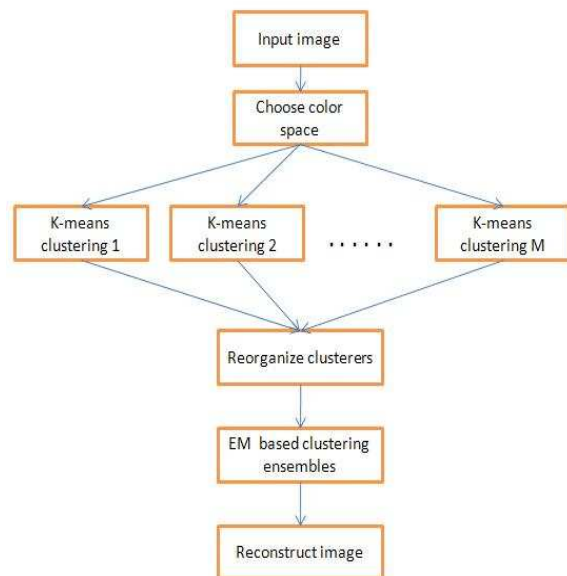


Figure 2 Framework of the proposed algorithm.

The main focus of this paper is a technique for stable and accurate color quantization using clustering ensemble. For each single k -means clustering, we randomly choose the initial centroids. Different initial centroids often lead to different quantization results. According to each pixel's clustering label from different k -means clusterings, we reorganize them into a single vector. Then, an EM algorithm is used to combine the vector of each pixel into a final partition, based on which, the image is reconstructed to obtain a quantized result. More details of implementation will be shown in the following subsections.

Before further illustrations about clustering ensemble-based color image quantization, we define notations used throughout the remainder of this paper. Let $I = \{P_1, P_2, \dots, P_{m \times n}\}$ denote a color image having a set of $m \times n$ pixels without labels. A clustering result of the data set I can be represented as a label vector $L \in N^{m \times n}$, where L_i is the label of the pixel P_i . Let $C = \{L^{(1)}, L^{(2)}, \dots, L^{(M)}\}$ be a set of M clustering results of the same image I , where each $L^{(i)}$ is a label vector of $\{L_1^{(i)}, L_2^{(i)}, \dots, L_{m \times n}^{(i)}\}$. Each of M clustering results has $K_i, i=1, 2, \dots, M$ number of centroids. Our algorithm combines the multiple clustering results in a clustering ensemble $C = \{L^{(1)}, L^{(2)}, \dots, L^{(M)}\}$ into a single consensus clustering result L^{final} .

B. Single K-means Clustering

K -means [16] is one of the simplest unsupervised learning algorithms for solving the clustering problem. The algorithm starts by partitioning the input points into k initial sets, either randomly or using a heuristic. It then calculates the mean point, or centroid, of each set and constructs a new partition by associating each point with its closest centroid. After this association, the centroids are recalculated to form new clusters. The algorithm repeats these two steps until convergence, when no data points switch clusters. The algorithm tries to minimize total intra-cluster variance,

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} Dist(x_j, \mu_i) \tag{1}$$

where S_i represents k clusters, $i=1, 2, \dots, k$ and μ_i represents centroids of all the points $x_j \in S_i$. $Dist(x_j, \mu_i)$ is a chosen distance measured between a data point x_j and the cluster centroid μ_i . $Dist(x_j, \mu_i)$ can be Manhattan distance, Euclidean distance, Hamming distance, or a similar measure.

In order to depict the essential color distribution as accurate as possible, we adopt random initial clustering centroids which cause different results. The procedure of applying the single k -means clustering algorithm to the color image consists of the following steps:

- (1) Determine the numbers of clusters K_1, K_2, \dots, K_M for M k -means clusterings to form M clustering results of the same image I .
- (2) For each single k -means clustering, randomly select $K_i, i=1, 2, \dots, M$ pixels in RGB color space as the initial clustering centroids.
- (3) For each pixel in the image, assign pixel to the group that has the closest centroid based on the predetermined distance measure.

- (4) When all pixels have been assigned, recalculate the positions of the current clustering $K_i, i=1, 2, \dots, M$ centroids.
- (5) Repeat steps (3) and (4) until the centroids no longer move, then go to step (6).
- (6) Repeat steps (2) to (5) until M k -means clusterings are completed.

C. Clustering ensemble based on EM

After each single k -means clustering is completed, we obtain all label vectors $\{L_1^{(i)}, L_2^{(i)}, \dots, L_{m \times n}^{(i)}\}, i=1, 2, \dots, M$ from these M clusterings. Each pixel in the objective color image is labeled by M separate indices, each specifying a group within its own clustering environment. As previously noted, each index is meaningful only within its corresponding clustering, and no explicit correspondence of results spans the different single k -means clusterings. Pixel S may be labeled as $L_S^{(1)}, L_S^{(2)}, \dots, L_S^{(M)}$. Although the corresponding RGB colors to these labels may be similar to each other, the label indexes may be different from each other to a large extent, since the label index denotes sorted rather than original data values as mentioned in Section 2. Therefore, if we directly combined label indices to form a final partition, some pixel values would be distorted significantly, and the combined results would be even worse than those of a single k -means clustering. To overcome this challenge, we use the original data in computing a clustering ensemble in order to avoid color distortion.

For pixel S of the test color image, the R, G, and B values corresponding to label $L_S^{(i)}, i=1, 2, \dots, M$ are denoted as $R_S^{(i)}, G_S^{(i)}, B_S^{(i)}$. We construct a new vector L_S as follows:

$$L_S = \{L_S^{(1)}, L_S^{(2)}, \dots, L_S^{(M)}\}^T, \text{ and} \tag{2}$$

$$L_S^{(i)} = \{R_S^{(i)}, G_S^{(i)}, B_S^{(i)}\}, i=1, 2, \dots, M$$

where L_S is a vector consisting of $M \times 3$ elements. Based on this new vector label, we see that pixel S can be typified by M sets of color values grouped from M different clustering environments. These heterogeneous clustering environments discover the color distribution in the test image from different perspectives, and the resulting ensemble can approach the essential shape distribution of colors in the pixel data.

In order to combine all pixel labels to obtain the final color quantization result, we use a statistical mixture model [17], which can be resolved by an EM technique [18]. In this paper, the finite mixture model for the probability of the final cluster labels $y = \pi(x)$ of the pixel x is used [12]. The main assumption is that labels y_i are modeled as random variables drawn from a probability distribution described as a mixture model of multivariate component densities:

$$P(y_i | \theta) = \sum_{s=1}^S \alpha_s P_s(y_i | \theta_s) \tag{3}$$

where each component is parameterized by θ_s . Complete details about combining clusters to form an ensemble committee and obtaining a final partition from the ensemble committee can be found in [12].

We input each pixel's labels L_1, L_2, \dots, L_M from M heterogeneous k -means clusterings into the EM solution-based finite mixture model to obtain the final partitions L^{final} with a total of S groups. All pixels which have been grouped into final color quantization results in the whole objective color image are denoted as $L^{final} = \{L_{p_1}^{final}, L_{p_2}^{final}, \dots, L_{p_{max}}^{final}\}$. In order to reflect the original color distribution as closely as possible, we directly reconstruct the image from original pixel values rather than from M heterogeneous clusterings. For each group, we extract the average color values directly from the original image according to the final labels of each pixel. We obtain final quantized colors as follows:

$$[\bar{R}_i, \bar{G}_i, \bar{B}_i] = \frac{\sum_{j=1}^{S_i} [R_j, G_j, B_j]}{S_i}, i = 1, 2, \dots, S \quad (4)$$

where S_i is the number of the clustered groups the i^{th} pixel belonged to, $[\bar{R}_i, \bar{G}_i, \bar{B}_i]$ is the average color value of i^{th} pixel, and $[R_j, G_j, B_j]$ is the color value of the i^{th} pixel belonging to the j^{th} group. Hence, we use the resulting $[\bar{R}_i, \bar{G}_i, \bar{B}_i], i = 1, 2, \dots, S$ to reconstruct the color image using a total of S unique pixel values.

IV. EXPERIMENTAL RESULTS

We used 24-bit images for experiments and mean squared error (MSE) [9] to quantify the difference between the original image and the quantized image as a measure of the effectiveness and performance of the proposed color quantization algorithm. The MSE is defined as following:

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N d(c[i, j], q(c[i, j])) \quad (5)$$

where $c[i, j]$ and $q(c[i, j])$ are values of original pixels and quantized pixels, respectively; $M \times N$ is the number of pixels in the image; and $d(x, y)$ is the Euclidean distance between two colors. Smaller MSE values imply better image reconstruction and color quantization.

Our first test image is a 24-bit color "parrots" image with size 384×256. For the results presented, Euclidean distance was used as the distance measure of the single k -means clustering, and M (the number of clusters used in the ensemble) was set to 6. The cluster number of each of $K_i, i=1, 2, \dots, 6$ was set to 20. All 6 single k -means clusterings were initialized by random centroids. The

number of quantized colors was alternately set to 6, 7, 8, 9, and 10. For each number of quantized colors, 5 experiments were conducted for the clustering ensemble and k -means clustering respectively. For comparative k -means clustering, we also used a random initialization strategy. The number of final quantized colors was also set to 6, 7, 8, 9, and 10.

Table 1 shows color quantization results of clustering ensemble and standard k -means clustering measured by MSE for each of the 5 experiments. As the number of quantized colors varies, the color quantization results of the clustering ensemble were more stable than those of the k -means single clustering. For instance, when the number of the final quantized colors was 7, the span of MSE values based on clustering ensemble method was 6.8 (ranging from a minimum value of 125.8 to a maximum value of 132.6). In contrast, the corresponding span of MSE values for k -means single clustering was 22.1 (from a minimum of 133.1 to a maximum of 153.5). Because of randomly initialized centroids, the single k -means clustering exhibits chaotic quantization performance. The clustering ensemble combines clusters from different perspectives, each obtained from random initial points in single k -means algorithms. As a group, the ensemble depicts intrinsic clustering shape distribution from different viewpoints, and thus achieves a more stable quantization. Furthermore, the quantization resulting from the clustering ensemble was generally more accurate than that of single k -means clustering, since the former considers multiple clustering shapes. Figure 3 shows the "parrots" quantization results based on clustering ensemble and k -means clustering with 6 quantized colors.

A 24-bit color "sky" image was used to test the impact of varying the number of centroids $K_i, i=1, 2, \dots, M$ on final quantization results with 6 quantized colors. We constructed 5 clustering ensembles, each of which has 6 ($M=6$) single clusterings. 6 single clusterings within the clustering ensemble are set to have the same number of centroids, which are 20, 30, 40, 50, and 60 respectively. We conducted five separate experiments using clustering ensembles and k -means clustering with 6 colors and evaluated the performance of the resulting quantization using MSE.

From Table 2, we can see that the final quantization results based on clustering ensemble are more stable than those of k -means clustering. Overall, the quantization results using the ensemble are typically an improvement

TABLE 1: MSE MEASURES OF CLUSTERING ENSEMBLE AND K-MEANS CLUSTERING ON "PARROTS" IMAGE USING EUCLIDEAN DISTANCE MEASURE AND RANDOM INITIALIZATION TO OBTAIN DIFFERENT NUMBER OF FINAL QUANTIZED COLORS.

Number of final quantized colors	MSE (Clustering ensemble)					MSE (K -means clustering)				
	1 st	2 nd	3 rd	4 th	5 th	1 st	2 nd	3 rd	4 th	5 th
6	144.8	149.8	138.7	142.1	145.9	151.4	137.6	163.9	152.7	171.2
7	132.6	125.8	127.2	130.3	126.7	153.5	145.2	133.1	155.2	147.4
8	122.9	121.8	122.9	123.1	125.1	147.1	151.5	124.7	147.4	133.9
9	127.0	117.7	123.2	126.9	118.0	128.1	133.1	127.0	139.6	133.2
10	128.1	121.5	117.2	123.4	120.8	129.2	134.0	125.6	126.1	128.3

TABLE 2: MSE MEASURES OF CLUSTERING ENSEMBLE AND K-MEANS CLUSTERING ON “SKY” IMAGE USING EUCLIDEAN DISTANCE MEASURE AND RANDOM INITIALIZATION FOR SINGLE CLUSTERING. EVERY SINGLE COMPONENT CLUSTERING WITHIN EACH CLUSTERING ENSEMBLE IS SET TO 20, 30, 40, 50, 60 CENTROIDS RESPECTIVELY.

Clustering ensemble	Five clustering centroids					K-means clustering
	20	30	40	50	60	
1 st	67.7	78.1	76.0	77.2	75.7	81.1
2 nd	72.5	77.2	76.3	76.0	76.8	78.0
3 rd	66.4	67.9	77.5	66.7	78.0	77.9
4 th	73.2	66.6	76.0	73.7	66.3	65.6
5 th	71.9	78.7	76.9	73.5	64.9	80.8

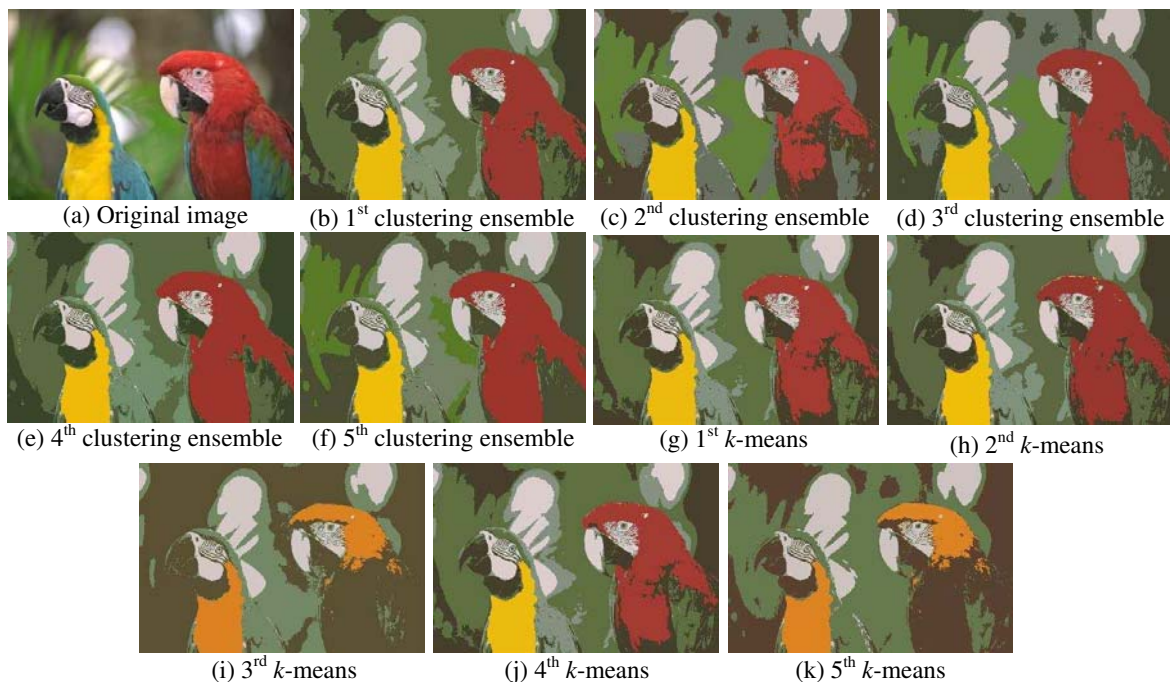


Figure 3 Quantization results of “parrots” image using 20 clustering centroids as the single clustering and keeping 6 final quantized colors.

over the results of *k*-means clustering. While the best quantization result of *k*-means clustering (65.6) is very close to the best result obtained from the clustering ensemble (64.9), the average results from the clustering ensemble are a significant improvement over the typical results of *k*-means clustering. Moreover, although 6 single clusterings within each clustering ensemble were set to different centroids numbers from 20 to 60, the quantization results were similar over that range. Figure 4 shows five quantized images of 6 colors based on clustering ensemble at $K_i=20, i=1, 2, \dots, M$, and five quantized images based on *k*-means clustering. The images demonstrate that the clustering ensemble method can discriminate between differences in the sky, but *k*-means clustering could not keep the visual integrity of the whole image. For instance, the two *k*-means images in Figure 4(g) and (k) do not differentiate the sky into more than two colors, as do all quantized images resulting from the clustering ensemble method. Because *k*-means clustering is initialized using random points, it sometimes finds local rather than global optima, as shown in Figure

4(g) and (k). In contrast, clustering ensemble-based quantization produces relatively consistent results. For example, when $K_i=20, i=1, 2, \dots, M$, the results using a clustering ensemble range from 66.6 to 73.2. Comparable results using *k*-means clustering vary over a range more than twice that large, from 65.6 to 81.1. The proposed method maintains a better balance between different parts of the objective color image and produces more stable and accurate results.

As mentioned above, the proposed algorithm not only stabilizes color quantization result, but also increases the quantization accuracy to some extent. Clustering ensemble has been successfully used in data clustering [11-14] in recent years. From the proposed algorithm and experimental results, it can be seen that it also achieves better performance for color quantization problem in computer vision. Increased stabilization and accuracy of color quantization based on *k*-means algorithm will help enhance efficiency of image compression or transmission tasks.

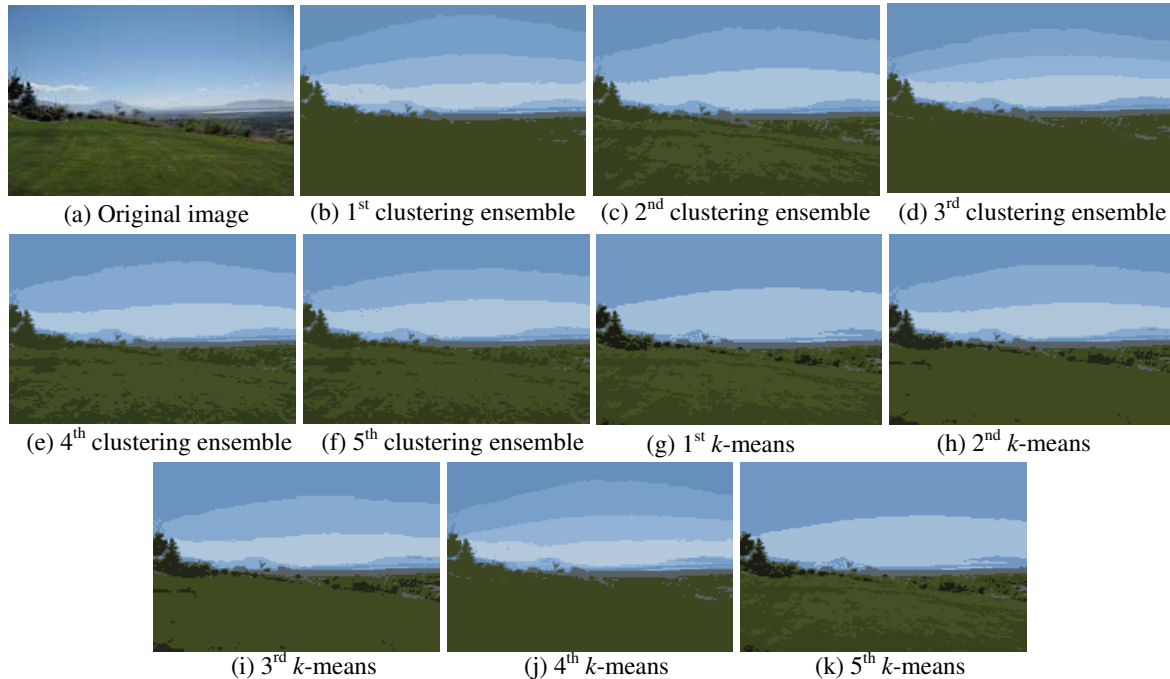


Figure 4 Quantization results of "sky" image using 20 clustering centroids as the single clustering and keeping 6 final quantized colors.

V. CONCLUSIONS

In order to solve the stability and accuracy problems in k -means clustering-based color quantization, we have proposed a color image quantization algorithm using a k -means clustering ensemble. We chose the RGB color space, and then applied single k -means clustering to the test color image to obtain heterogeneous clustering results. Considering the inexplicit correspondence among labels of clustering results, we adopted color values of the clustering centroids directly to construct a final ensemble committee. Finally, the mixture model solved by an expectation-maximization technique was used to combine all the heterogeneous clustering groups to form the final quantized image. Experimental results show that the proposed method is robust to real color images and more reliable than color image quantization based on single k -means clustering.

We applied the clustering ensemble, a popular research topic in unsupervised machine learning in recent years, to color image quantization. Although clustering ensemble processing is computationally expensive, with modern computing platforms, parallelism can be used to process all the single clusterings to improve processing speed. Our future work will focus on combining different clustering methods such as mean shift [19] for this novel color quantization algorithm.

REFERENCES

- [1] P. Heckbert, Color Image Quantization for Frame Buffer Display, ACM Computer Graphics, vol. 16, no. 3, pp. 297–307, 1982.
- [2] M. Gervautz, and W. Purgathofer., A Simple Method for Color Quantization: Octree Quantization. Proceedings of CG International_88, pp. 219–231, 1988.
- [3] G. Joy and Z. Xiang, 1993, Center-Cut for Color Image Quantization, International Journal of Computer Graphics, vol. 10, no. 1, pp. 62–66.
- [4] Y. Deng, C. Kenney, M. S. Moore, and B. S. Manjunath, Peer Group Filtering and Perceptual Color Image Quantization, Proc. IEEE International Symposium on Circuits and Systems, vol.4, pp. 21–24, 1999.
- [5] Z. Tian, R. Raghu, and L. Miron, BIRCH: An Efficient Data Clustering Method for Very Large Databases. SIGMOD_96, Montreal, Canada, pp. 103–114, 1996.
- [6] T. Kanunqo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman and A.Y. Wu, An Efficient K-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881–892, 2002.
- [7] O. Verevka, Color Image Quantization in Window System with Local K-means Algorithm. Proc. Western Computer Graphics Symposium, pp. 74–79, 1995.
- [8] I.-S. Hsieh and K.-C. Fan, 2000, An Adaptive Clustering Algorithm for Color Quantization, Pattern Recognition Letters, vol. 21, pp. 337–346, 2000.
- [9] A. H. Dekker, Kohonen Neural Networks for Optimal Colour Quantization, Network Computation. Neural System. vol. 5, pp. 351–367, 1994.
- [10] X. Hu, T. Wang and D. Li, A New Approach of Color Quantization Based on Ant Colony Clustering Algorithm, International Conference on Information Technology: Coding and Computing, 2005.
- [11] A. Strehl and J. Ghosh, Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions. Journal of Machine Learning Research, no.3 pp.583–617, 2002.
- [12] A. Topchy, A.K. Jain and W. Punch, Clustering Ensembles: Models of Consensus and Weak Partitions. IEEE

Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 12, pp. 1866-1881, 2005.

- [13] A. Fred and A.K. Jain, Combining Multiple Clusterings Using Evidence Accumulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 835-850, 2005.
- [14] L.I. Kuncheva and D.P. Vetrov, Evaluation of Stability of K-Means Cluster Ensembles with Respect to Random Initialization. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1798-1808.
- [15] H. Stokman and T. Gevers, Selection and Fusion of Color Models for Feature Detection. IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 560-565, 2005.
- [16] J.B. MacQueen, Some Methods for Classification and Analysis of Multivariate Observations. Proc. of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1: 281-297.
- [17] G. McLachlan and K.E. Basford, Mixture Models. Marcel Dekker, Inc., Basel, NY, 1988.
- [18] A. Dempster, N. Laird and D. Rubin, Maximum Likelihood from the Incomplete Data Via the EM Algorithm. Journal of Royal Statistical Society, Series B, 39(1):1-38, 1977.
- [19] D. Comaniciu and P. Meer, Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603-619.
- [20] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data, Machine Learning, vol.52, pp.91-118, 2003.

Yuchou Chang was born in Hunan, China in 1980. He received his B.S. degree in automatic control department from Northwestern Polytechnical University, Xi'an, China, in 2003 and M.S. degree in institute of image processing and pattern recognition from Shanghai Jiao Tong University, Shanghai, China, in 2006. He is currently working toward his Ph.D. degree in the Robotic Vision Laboratory in the Electrical and Computer Engineering Department at Brigham Young University.

His research interest includes machine learning-assisted multimedia analysis, image segmentation, image and video semantic content description, content-based multimedia indexing and retrieval. He is an IEEE student member.

Dah-Jye Lee received his B.S. from National Taiwan University of Science and Technology in 1984, M.S. and Ph.D. degrees in electrical engineering from Texas Tech University in 1987 and 1990, respectively. He also received his MBA degree from Shenandoah University, Winchester, Virginia in 1999.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering at Brigham Young University. He worked in the machine vision industry for eleven years prior to joining BYU in 2001. His research work focuses on Medical informatics and imaging, shape-based pattern recognition, hardware implementation of real-time 3-D vision algorithms and machine vision applications.

Dr. Lee is a senior member of IEEE and a member of SPIE. He has actively served as a paper and proposal reviewer and conference organizer. He has served as the editor, general chair, and steering committee member of the IEEE International Symposium of Computer-based Medical Systems. He received the best faculty advisor award from Brigham Young University Student Association in 2005.

Yi Hong received his M.S. in Computer Science from Shanghai Jiao Tong University, P.R. China in 2006. He is now a full-time research assistant in the Department of Computer Science at the City University of Hong Kong. His research areas include pattern recognition, machine learning, data mining and evolutionary computation.

James K. Archibald received the B.S. degree in mathematics from Brigham Young University in 1981 and the M.S. and Ph.D. degrees in computer science from the University of Washington in 1983 and 1987, respectively. He has been with the Electrical and Computer Engineering Department at Brigham Young University since 1987. His research interests include robotics, multi-agent systems, and machine vision. Dr. Archibald is a member of the ACM and Phi Kappa Phi.

Dong Liang received his Ph.D. degree from Shanghai Jiaotong university of China in 2006. He then joined the University of Hong Kong as a research assistant and research associate in 2006. He is currently a postdoctoral fellow in the Department of Electronic Engineering at the University of Wisconsin-Milwaukee. His main research interests are image processing, image retrieval, pattern recognition, magnetic resonance imaging.