International Journal of Hybrid Intelligent Systems 8 (2011) 25–30
DOI 10.3233/HIS-2011-0128
IOS Press

Rough colour quantisation

Gerald Schaefer[a,*], Huiyu Zhou[b], M. Emre Celebi[c] and Aboul Ella Hassanien[d]
[a]*Department of Computer Science, Loughborough University, Loughborough, UK*
[b]*Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast, UK*
[c]*Department of Computer Science, Louisiana State University in Shreveport, Shreveport, USA*
[d]*Information Technology Department, Cairo University, Giza, Egypt*

Abstract. Colour quantisation algorithms are essential for displaying true colour images using a limited palette of distinct colours. The choice of a good colour palette is crucial as it directly determines the quality of the resulting image. Colour quantisation can also be seen as a clustering problem where the task is to identify those clusters that best represent the colours in an image. In this paper, we use a rough c-means clustering algorithm for colour quantisation of images. Experimental results on a standard set of images show that this rough colour quantisation approach performs significantly better than other, purpose built colour reduction algorithms.

1. Introduction

Colour quantisation is a common image processing technique that allows the representation of true colour images using only a small number of colours. True colour images typically use 24 bits per pixel resulting overall in 2^{24}, i.e. more than 16 million different colours. Colour quantisation uses a colour palette that contains only a small number of distinct colours (usually between 8 and 256) and pixel data are then stored as indices to this palette. Clearly, the choice of the colours that make up the palette is of crucial importance for the quality of the quantised image. However, the selection of the optimal colour palette is known to be an np-hard problem [9]. In the image processing literature many different algorithms have been introduced that aim to find a palette that allows for good image quality of the quantised image. Soft computing techniques such as genetic algorithms have also been employed to extract a suitable palette [13,19].

Colour quantisation can also be seen as a clustering problem where the task is to identify those clusters that best represent the colours in an image. In this paper, we use a rough c-means clustering algorithm for colour quantisation of images; that is, we perform rough clustering and use the derived cluster centres as a colour palette. This rough-set based clustering approach utilises two sets for each cluster, a lower and an upper approximation. Through iterative refinement of the cluster centres, the algorithm converges towards a good colour palette. Experimental results on a standard set of images show that this rough image quantisation performs significantly better than other colour quantisation algorithms.

2. Related work

The process of colour quantisation is mainly comprised of two phases: palette design (the selection of a small set of colours that represents the original image colours) and pixel mapping (the assignment of each input pixel to one of the palette colours). The primary objective is to reduce the number of unique colours with minimal distortion. Since natural images often contain a large number of colours, faithful representation of these images with a limited size palette is a difficult problem.

Colour quantisation methods can be broadly classified into two categories: image-independent methods that determine a universal (fixed) palette without regard to any specific image, and image-dependent methods that determine a custom (adaptive) palette based on the colour distribution of the images. Despite being very fast, image-independent methods usually give poor re-

*Corresponding author. E-mail: gerald.schaefer@ieee.org.

1448-5869/11/$27.50

sults since they do not take into account the image contents. Therefore, most of the studies in the literature consider only image-dependent methods, which strive to achieve a better balance between computational efficiency and visual quality of the quantisation output.

Numerous image-dependent colour quantisation methods have been developed in the past three decades. These can be categorised into two families: pre-clustering methods and post-clustering methods [4]. Pre-clustering methods are mostly based on the statistical analysis of the colour distribution of the images. Divisive pre-clustering methods start with a single cluster that contains all N image pixels. This initial cluster is recursively subdivided until K clusters are obtained. Well-known divisive methods include median-cut [9], octree [8], variance-based quantisation [22], and greedy orthogonal bipartitioning [23]. On the other hand, agglomerative pre-clustering methods [1,3,20] start with N singleton clusters each of which contains one image pixel. These clusters are repeatedly merged until K clusters remain. In contrast to pre-clustering methods that compute the palette only once, post-clustering methods first determine an initial palette and then improve it iteratively. Essentially, any data clustering method can be used for this purpose. Since these methods involve iterative or stochastic optimisation, they can obtain higher quality results when compared to pre-clustering methods at the expense of increased computational time. Clustering algorithms adapted to colour quantisation include k-means [10,11], minmax [24], competitive learning [5–7,18,21], fuzzy c-means [14, 17], and BIRCH [2].

3. Rough colour quantisation

As mentioned, colour quantisation can be seen as a clustering problem where the task is to identify those clusters that best represent the colours in an image. In this paper, we employ a rough c-means clustering algorithm for this purpose.

Lingras et al. [12] introduced a rough set inspired clustering algorithm based on the well known c-means algorithm. In their rough c-means approach, each cluster c_k is described not only by its centre m_k, but also contains additional information, in particular its lower approximation $\underline{c_k}$, its upper approximation $\overline{c_k}$, and its boundary area $c_k^b = \overline{c_k} - \underline{c_k}$. Lingras et al.'s algorithm proceeds in the following steps:

Step 1: *Initialisation:* Each data sample is randomly assigned to one lower approximation. As the lower approximation of a cluster is a subset of its upper approximation, this also automatically assigns the sample to the upper approximation of the same cluster.

Step 2: *Cluster centre calculation:* The cluster centres are updated as

$$m_k = \begin{cases} \omega_l \sum_{x_i \in \underline{c_k}} \frac{x_i}{|\underline{c_k}|} + \omega_b \sum_{x_i \in c_k^b} \frac{x_i}{|c_k^b|} \\ \quad \text{if } c_k^b \neq \{\} \\ \omega_l \sum_{x_i \in \underline{c_k}} \frac{x_i}{|\underline{c_k}|} \quad \text{otherwise} \end{cases} \tag{1}$$

The cluster centres are hence determined as a weighted average of the samples belonging to the lower approximation and the boundary area, where the weights ω_l and ω_b define the relative importance of the two sets.

Step 3: *Sample assignment:* For each data sample the closest cluster centre is determined and the sample assigned to its upper approximation. Then, all clusters that are at most ϵ further away than the closest cluster are determined. If such clusters exist, the sample will also be assigned to their upper approximations. If no such cluster exist, the sample is assigned also to the lower approximation of the closest cluster.

Step 4: *Termination:* If the algorithm has converged (i.e., if the cluster centres do not change any more, or after a pre-set number of iterations), terminate, otherwise go to Step 2.

Strictly speaking, this algorithm does not implement all properties set out for rough sets [15], and hence belongs to the reduced interpretation of rough sets as lower and upper approximations of data [25].

Peters [16] noticed some potential pitfalls of the algorithm as proposed by Lingras et al. in terms of objective function and numerical stability, and suggested some improvements to overcome these. Equation (1) is revised to

$$m_k = \omega_l \sum_{x_i \in \underline{c_k}} \frac{x_i}{|\underline{c_k}|} + \omega_u \sum_{x_i \in \overline{c_k}} \frac{x_i}{|\overline{c_k}|} \tag{2}$$

with $\omega_l + \omega_u = 1$, i.e. as a convex combination of lower and upper approximation means. In order to overcome the possibility of situations with empty lower approximations, Peters suggests two possible ways of addressing this, either by modifying the calculation of cluster centres so that for empty lower approximations the cluster centre is calculated as the average of samples in the upper approximation, or by ensuring that each lower approximation has at least one member. In our approach we choose the latter by assigning the data sample closest to the cluster centre to its lower approximation.

Fig. 1. The six test images used in the experiments: *Lenna, Peppers, Mandrill, Sailboat, Pool, and Airplane* (from left to right, top to bottom).

In addition, we perform a different initialisation procedure than Lingras et al. and Peters. Rather than randomly assigning samples to clusters, we generate random cluster centres first and then proceed with Steps 3, 2 and 4 (i.e., steps 2 and 3 reversed) of the algorithm.

4. Experimental results

For our experiments we used six standard images commonly used in the colour quantisation literature (*Lenna, Peppers, Mandrill, Sailboat, Airplane, and Pool* – see Fig. 1) and applied our rough c-means colour quantisation algorithm to generate quantised images with a palette of 16 colours.

To put the results we obtain into context we have also implemented ten popular colour quantisation algorithms to generate corresponding quantised images with palette size 16. The algorithms we have tested were: popularity algorithm [9], median cut quantisation [9], octree quantisation [8], variance-based quantisation [22], Neuquant [7], modified minmax [24], split & merge [3], fuzzy c-means [14], fuzzy c-means (FCM) with partition index maximization (PIM) [14], and stepwidth adaptive simulated annealing (SWASA) [13].

For our rough c-means approach, we adopt the changes proposed by Peters with parameters $\omega_l = 0.7$, $\omega_u = 0.3$, $\epsilon = 0.001$ (image pixel values are normalised to $[0;1]^3$). For all algorithms, pixels in the quantised images were assigned to their nearest neighbours in the colour palette to provide the best possible image quality.

The results are listed in Table 1, expressed in terms of peak signal to noise ratio (PSNR) defined as

$$\mathrm{PSNR}(I_1, I_2) = 10 \log_{10} \frac{255^2}{MSE(I_1, I_2)} \tag{3}$$

with MSE (the mean-squared error) calculated as

$$\begin{aligned} \mathrm{MSE}(I_1, I_2) = \frac{1}{3nm} \sum_{i=1}^{n} \sum_{j=1}^{m} [(R_1(i,j) - R_2(i,j))^2 \\ +(G_1(i,j) - G_2(i,j))^2 + (B_1(i,j) - B_2(i,j))^2] \end{aligned} \tag{4}$$

where $R(i,j)$, $G(i,j)$, and $B(i,j)$ are the red, green, and blue pixel values at location (i,j), and n and m are the dimensions of the images.

From Table 1 we can see that of the early approaches, the popularity and median cut algorithms, perform relatively poorly compared to all other algorithms. Octree, Neuquant, variance-based quantisation, modified minmax, and split & merge perform much better but are also clearly inferior to the more recent clustering (fuzzy c-means) and optimisation (simulated annealing) algorithms. For our rough c-means approach, we ran the algorithm 10 times (randomly initialising the cluster centres) on each image and report both the average and the highest PSNR of these 10 runs in Table 1. Looking

Table 1
Quantisation results, given in terms of PSNR [dB]

	Lenna	Peppers	Mandrill	Sailboat	Pool	Airplane	average
Popularity algorithm [9]	22.24	18.56	18.00	8.73	19.87	15.91	17.22
Median cut [9]	23.79	24.10	21.52	22.01	24.57	24.32	23.39
Octree [8]	27.45	25.80	24.21	26.04	29.39	28.77	26.94
Variance-based [22]	27.80	25.78	24.01	26.49	29.59	28.35	27.00
Neuquant [7]	27.82	26.04	24.59	26.81	27.08	28.24	26.73
Modified minmax [24]	27.59	26.34	23.46	25.92	28.56	28.26	26.69
Split & merge [3]	26.20	26.24	23.97	25.12	29.38	26.49	26.23
Fuzzy c-means [14]	28.70	26.55	24.60	27.47	28.05	30.54	27.65
FCM w/ PIM [14]	28.89	26.79	25.00	27.55	26.74	30.20	27.53
SWASA [13]	27.79	26.16	24.46	26.69	29.84	29.43	27.40
Rough c-means (mean)	28.63	26.67	25.02	27.62	29.40	30.50	27.98
Rough c-means (max)	28.77	26.81	25.10	27.82	30.17	31.03	28.28

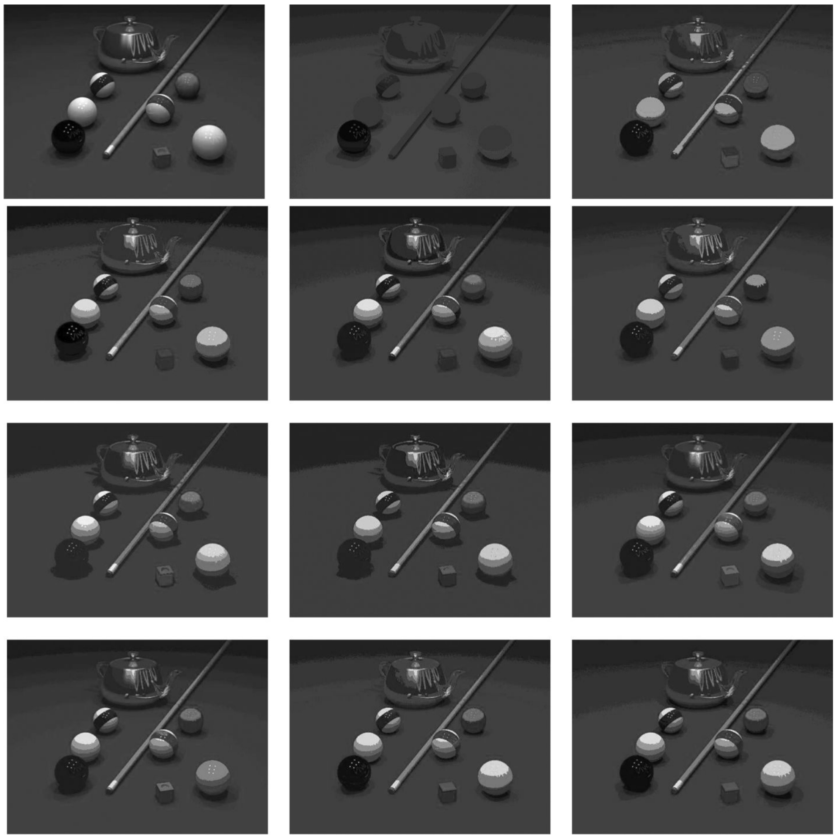

Fig. 2. Results of colour quantisation algorithms applied to the *Pool* image (top left image). Then, from left to right, top to bottom, we show the outputs of: popularity algorithm [9], median cut quantisation [9], octree quantisation [8], variance-based quantisation [22], Neuquant [7], modified minmax [24], split & merge [3], fuzzy c-means [14], fuzzy c-means with partition index maximization [14], stepwidth adaptive simulated annealing (SWASA) [13], and the proposed rough colour quantisation algorithm.

at the results, it can be seen that the rough c-means approach achieves better image quality than all other tested algorithms, which is quite remarkable. Interestingly, it also outperforms other clustering quantisation approaches, namely those of fuzzy c-means and fuzzy c-means with partition index maximization.

The results shown in Table 1 are further illustrated in Fig. 2 which show all colour quantised versions for each of the *Pool* test images. It is clear that the popularity algorithm performs poorly on this image and assigns virtually all of the colours in the palette to green and achromatic colours. Median cut is better but still

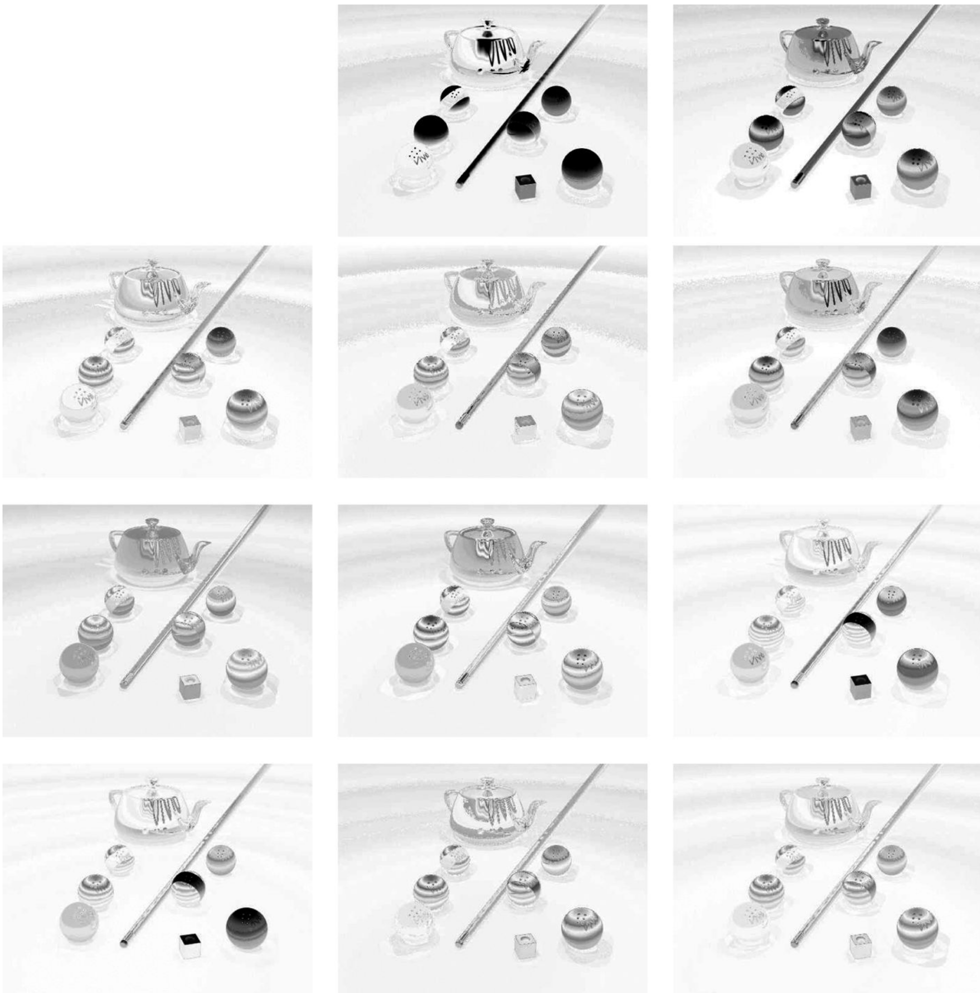

Fig. 3. Error images, corresponding to and laid out in the same fashion as the colour quantised images of Fig. 2.

provides fairly poor colour reproduction; most of the colours in the quantised image are fairly different from the original. A better result is achieved by the octree algorithm, although here also the red is not very accurate and the colour of the cue is greenish instead of brown. While it did not do too well on the other test images, on this particular image the variance-based method produces very good results. Still, some artefacts can be seen on the billard cue which turns partly red. For Neuquant, the most obvious shortcoming is the absence of an appropriate red colour in the colour palette. The modified minmax algorithm fails to include a clear black colour in the palette, while the split & merge algorithm exhibits a similar deficiency. The fuzzy c-means based techniques also do not produce a good black colour, more obvious however is the absence of a blue colour in the colourmap. The simulated annealing optimisation approach produces a reasonable good image. Still, clearly better image quality is maintained by applying our rough colour quantisation technique. Although the colour palette has only 16 entries all colours of the original image are accurately presented including the red ball, the colour of the billiard cue, the green cloth and the reflection on the kettle.

Error images (or image distortion maps) are commonly employed for judging the difference between images or the performance of competing algorithms [26]. In Fig. 3, we therefore provide for each quantised image an error image that represents the difference between the original and the palettised image. For this, we calculate the squared error at each pixel location, sum it up over the three colour channels and invert the resulting image. Again, it is apparent that the version obtained from our rough clustering algorithm provides the highest colour fidelity and hence the smoothest error images.

5. Conclusions

In this paper, we have proposed a rough c-means based colour quantisation algorithm. Rough c-means is applied to extract cluster centres corresponding to palette entries of colour quantised images. Experimental results obtained on a set of common test images have demonstrated that this approach can not only be effectively employed but clearly outperforms other colour quantisation algorithms.

References

[1] R. Balasubramaian and J. Allebach, A new approach to palette selection for color images, *Journal of Imaging Technology* **17**(6) (1991), 284–290.

[2] Z. Bing, S. Junyi and P. Qinke, An adjustable algorithm for color quantization, *Pattern Recognition Letters* **25** (2004), 1787–1797.

[3] L. Brun and M. Mokhtari, Two high speed color quantization algorithms, in: *Int Conference on Color in Graphics and Image Processing*, pages 116–121, 2000.

[4] L. Brun and A. Trémeau, Color quantization, in: *Digital Color Imaging Handbook*, G. Sharma, ed., CRC Press, 2002, pp. 589–638.

[5] M.E. Celebi, An effective color quantization method based on the competitive learning paradigm, in: *Workshop on Soft Computing in Image Processing and Computer Vision*, volume 2, pages 876–880, 2009.

[6] C.-H. Chang, P. Xu, R. Xiao and T. Srikanthan, New adaptive color quantization method based on self-organizing maps, *IEEE Trans Neural Networks* **16** (2005), 237–249.

[7] A.H. Dekker, Kohonen neural networks for optimal colour quantization, *Network: Computation in Neural Systems* **5** (1994), 351–367.

[8] M. Gervautz and W. Purgathofer, A simple method for color quantization: Octree quantization, in: *Graphics Gems*, A.S. Glassner, ed., 1990, pp. 287–293.

[9] P.S. Heckbert, Color image quantization for frame buffer display, *ACM Computer Graphics* (*ACM SIGGRAPH '82 Proceedings*) **16**(3) (1982), 297–307.

[10] Y.-C. Hu and B.-H. Su, Accelerated k-means clustering algorithm for colour image quantization, *Imaging Science Journal* **56**(1) (2008), 29–40.

[11] H. Kasuga, H. Yamamoto and M. Okamoto, Color quantization using the fast k-means algorithm, *Systems and Computers in Japan* **31** (2000), 33–40.

[12] P. Lingras and C. West, Interval set clustering of web users with rough k-means, *Journal Intell Inform Syst* **23** (2004), 5–16.

[13] L. Nolle and G. Schaefer, Color map design through optimization, *Engineering Optimization* **39**(3) (2007), 327–343.

[14] D. Ozdemir and L. Akarun, Fuzzy algorithm for color quantization of images, *Pattern Recognition* **35**(8) (2002), 1785–1791.

[15] Z. Pawlak, Rough sets, *Int Journal Inform Comput Sci* **11** (1982), 145–172.

[16] G. Peters, Some refinements of rough k-means clustering, *Pattern Recognition* **39** (2006), 1481–1491.

[17] G. Schaefer and H. Zhou, Fuzzy clustering for colour reduction in images, *Telecommunication Systems* **40**(1–2) (2009), 17–25.

[18] P. Scheunders, A comparison of clustering algorithms applied to color image quantization, *Pattern Recognition Letters* **18** (1997), 1379–1384.

[19] P. Scheunders, A genetic c-means clustering algorithm applied to color image quantization, *Pattern Recognition* **30**(6) (1997), 859–866.

[20] L. Velho, J. Gomez and M.V.R. Sobreiro, Color image quantization by pairwise clustering, in: *X Brazilian Symposium on Computer Graphics and Image Processing*, 1997, pp. 203–210.

[21] O. Verevka and J. Buchanan, Local k-means algorithm for color image quantization, in: *Graphics Interface*, 1995, pp. 128–135.

[22] S.J. Wan, P. Prusinkiewicz and S.K.M. Wong, Variance-based color image quantization for frame buffer display, *Color Research Applications* **15**(1) (1990), 52–58.

[23] X. Wu, Efficient statistical computations for optimal color quantization, in: *Graphics Gems II*, J. Arvo, ed., 1991, pp. 126–133.

[24] Z.G. Xiang, Color image quantization by minimizing the maximum intercluster distance, *ACM Trans on Graphics* **16**(3) (July 1997), 260–276.

[25] Y.Y. Yao, X. Li, T.Y. Lin and Q. Liu, Representation and classification of rough set models, in: *3rd Int Workshop on Rough Sets and Soft Computing*, 1994, pp. 630–637.

[26] X.M. Zhang and B.A. Wandell, Color image fidelity metrics evaluated using image distortion maps, *Signal Processing* **70**(3) (November 1998), 201–214.